

2023 IEEE/CVF International Conference on Computer Vision (ICCV 2023)

**Paris, France
1-6 October 2023**

Pages 1-453



**IEEE Catalog Number: CFP23198-POD
ISBN: 979-8-3503-0719-1**

**Copyright © 2023 by the Institute of Electrical and Electronics Engineers, Inc.
All Rights Reserved**

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved.

****** This is a print representation of what appears in the IEEE Digital Library. Some format issues inherent in the e-media version may also appear in this print version.***

IEEE Catalog Number:	CFP23198-POD
ISBN (Print-On-Demand):	979-8-3503-0719-1
ISBN (Online):	979-8-3503-0718-4
ISSN:	1550-5499

Additional Copies of This Publication Are Available From:

Curran Associates, Inc
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: (845) 758-0400
Fax: (845) 758-2633
E-mail: curran@proceedings.com
Web: www.proceedings.com

CURRAN ASSOCIATES INC.
proceedings
.com

2023 IEEE/CVF International Conference on Computer Vision (ICCV) ICCV 2023

Table of Contents

Organizers	ccxxxviii
Keynotes	ccxl

(O) 3D from Multi-view and Sensors 1

Multi-Modal Neural Radiance Field for Monocular Dense SLAM with a Light-Weight ToF Sensor 1	
<i>Xinyang Liu (Zhejiang University), Yijin Li (Zhejiang University), Yanbin Teng (Zhejiang University), Hujun Bao (Zhejiang University), Guofeng Zhang (Zhejiang University), Yinda Zhang (Google), and Zhaopeng Cui (Zhejiang University)</i>	
ScanNet++: A High-Fidelity Dataset of 3D Indoor Scenes 12	
<i>Chandan Yeshwanth (Technical University of Munich), Yueh-Cheng Liu (Technical University of Munich), Matthias Nießner (Technical University of Munich), and Angela Dai (Technical University of Munich)</i>	
Translating Images to Road Network: A Non-Autoregressive Sequence-to-Sequence Approach 23	
<i>Jiachen Lu (Fudan University), Renyuan Peng (Fudan University), Xinyue Cai (Huawei Noah's Ark Lab), Hang Xu (Huawei Noah's Ark Lab), Hongyang Li (Shanghai AI Lab / Shanghai Jiao Tong University), Feng Wen (Huawei Noah's Ark Lab), Wei Zhang (Noah's Ark Lab, Huawei Technologies), and Li Zhang (Fudan University)</i>	
Doppelgangers: Learning to Disambiguate Images of Similar Structures 34	
<i>Ruojin Cai (Cornell University), Joseph Tung (Cornell University), Qianqian Wang (Cornell), Hadar Averbuch-Elor (Tel Aviv University), Bharath Hariharan (Cornell University), and Noah Snavely (Cornell University and Google AI)</i>	
EgoLoc: Revisiting 3D Object Localization from Egocentric Videos with Visual Queries 45	
<i>Jinjie Mai (KAUST), Abdullah Hamdi (KAUST), Silvio Giancola (KAUST), Chen Zhao (KAUST), and Bernard Ghanem (KAUST)</i>	
ClothPose: A Real-world Benchmark for Visual Analysis of Garment Pose via An Indirect Recording Solution 58	
<i>Wenqiang Xu (Shanghai Jiao Tong University), Wenxin Du (Shanghai Jiao Tong University), Han Xue (Shanghai Jiao Tong University), Yutong Li (Shanghai Jiao Tong University), Ruolin Ye (Cornell University), Yan-Feng Wang (Cooperative medianet innovation center of Shanghai Jiao Tong University), and Cewu Lu (Shanghai Jiao Tong University)</i>	

EMR-MSF: Self-Supervised Recurrent Monocular Scene Flow Exploiting Ego-Motion Rigidity	69
<i>Zijie Jiang (Tokyo Institute of Technology) and Masatoshi Okutomi (Tokyo Institute of Technology)</i>	
ENVIDR: Implicit Differentiable Renderer with Neural Environment Lighting	79
<i>Ruofan Liang (University of Toronto), Huiting Chen (University of Toronto), Chunlin Li (University of Toronto), Fan Chen (University of Toronto), Selvakumar Panneer (Intel Cooperation), and Nandita Vijaykumar (University of Toronto)</i>	

(O) Adversarial Attack and Defense

Robust Mixture-of-Expert Training for Convolutional Neural Networks	90
<i>Yihua Zhang (Michigan State University), Ruisi Cai (The University of Texas at Austin), Tianlong Chen (University of Texas at Austin), Guanhua Zhang (Max Planck Institute for Intelligent Systems), Huan Zhang (CMU), Pin-Yu Chen (IBM Research), Shiyu Chang (UCSB), Zhangyang Wang (University of Texas at Austin), and Sijia Liu (Michigan State University)</i>	
Set-level Guidance Attack: Boosting Adversarial Transferability of Vision-Language Pre-training Models	102
<i>Dong Lu (Southern University of Science and Technology), Zhiqiang Wang (Southern University of Science and Technology), Teng Wang (The University of Hong Kong), Weili Guan (Monash University), Hongchang Gao (Temple University), and Feng Zheng (SUSTech)</i>	
CleanCLIP: Mitigating Data Poisoning Attacks in Multimodal Contrastive Learning	112
<i>Hritik Bansal (UCLA), Nishad Singhi (Max Planck Institute for Intelligent Systems, Tübingen), Yu Yang (University of California, Los Angeles), Fan Yin (UCLA), Aditya Grover (UCLA), and Kai-Wei Chang (University of California, Los Angeles)</i>	
CGBA: Curvature-aware Geometric Black-box Attack	124
<i>Md Farhamdur Reza (North Carolina State University), Ali Rahmati (NC State University), Tianfu Wu (NC State University), and Huaiyu Dai (NC State University)</i>	
Robust Evaluation of Diffusion-Based Adversarial Purification	134
<i>Minjong Lee (Postech) and Dongwoo Kim (POSTECH)</i>	
Advancing Example Exploitation Can Alleviate Critical Challenges in Adversarial Training	145
<i>Yao Ge (Nanjing University of Posts and Telecommunications), Yun Li (Nanjing University of Posts and Telecommunications), Keji Han (Nanjing University of Posts and Telecommunications), Junyi Zhu (Boston University), and Xianzhong Long (Nanjing University of Posts and Telecommunications)</i>	

The Victim and The Beneficiary: Exploiting a Poisoned Model to Train a Clean Model on Poisoned Data	155
<i>Zixuan Zhu (State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, School of Cyber Security, University of Chinese Academy of Sciences), Rui Wang (State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, School of Cyber Security, University of Chinese Academy of Sciences), Cong Zou (State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, School of Cyber Security, University of Chinese Academy of Sciences), and Lihua Jing (State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, School of Cyber Security, University of Chinese Academy of Sciences)</i>	
TIJO: Trigger Inversion with Joint Optimization for Defending Multimodal Backdoored Models....	165
<i>Indranil Sur (SRI International), Karan Sikka (SRI International), Matthew Walmer (University of Maryland), Kaushik Koneripalli (SRI International), Anirban Roy (SRI International), Xiao Lin (SRI International), Ajay Divakaran (SRI, USA), and Susmit Jha (SRI International)</i>	

Vision and Robotics

Simoun: Synergizing Interactive Motion-appearance Understanding for Vision-based Reinforcement Learning	176
<i>Yangru Huang (Peking University), Peixi Peng (Peking University), Yifan Zhao (Peking University), Yunpeng Zhai (Peking University), Haoran Xu (Sun Yat-Sen University), and Yonghong Tian (Peking University)</i>	
Among Us: Adversarially Robust Collaborative Perception by Consensus	186
<i>Yiming Li (New York University), Qi Fang (Beijing University of Posts and Telecommunications), Jiamu Bai (New York University), Siheng Chen (Shanghai Jiao Tong University, Shanghai AI Laboratory), Felix Juefei-Xu (Meta AI), and Chen Feng (New York University)</i>	
Walking Your LiDOG: A Journey Through Multiple Domains for LiDAR Semantic Segmentation ..	196
<i>Cristiano Saltori (University of Trento), Aljosa Osep (TUM Munich), Elisa Ricci (FBK - Technologies of Vision), and Laura Leal-Taixé (NVIDIA)</i>	
Stabilizing Visual Reinforcement Learning via Asymmetric Interactive Cooperation	207
<i>Yunpeng Zhai (Peking University), Peixi Peng (Peking University), Yifan Zhao (Peking University), Yangru Huang (Peking University), and Yonghong Tian (Peking University)</i>	
MAAL: Multimodality-Aware Autoencoder-Based Affordance Learning for 3D Articulated Objects.....	217
<i>Yuanzhi Liang (UTS), Xiaohan Wang (Zhejiang University), Linchao Zhu (Zhejiang University), and Yi Yang (Zhejiang University)</i>	

Rethinking Range View Representation for LiDAR Segmentation	228
<i>Lingdong Kong (National University of Singapore), Youquan Liu (Hochschule Bremerhaven), Runnan Chen (The University of Hong Kong), Yuexin Ma (ShanghaiTech University), Xinge Zhu (The Chinese University of Hong Kong), Yikang Li (Shanghai AI Lab), Yuenan Hou (Shanghai AI Lab), Yu Qiao (Shanghai AI Laboratory), and Ziwei Liu (Nanyang Technological University)</i>	
PourIt!: Weakly-Supervised Liquid Perception from a Single Image for Visual Closed-Loop Robotic Pouring	241
<i>Haitao Lin (Fudan University), Yanwei Fu (Fudan University), and Xiangyang Xue (Fudan University)</i>	
CROSSFIRE: Camera Relocalization On Self-Supervised Features from an Implicit Representation	252
<i>Arthur Moreau (Mines ParisTech - PSL University), Nathan Piasco (Huawei Technologies France), Moussab Bennehar (Huawei Paris Research Center), Dzmitry Tsishkou (Huawei Technologies), Bogdan Stanciulescu (Mines ParisTech), and Arnaud de La Fortelle (MINES ParisTech - PSL University)</i>	
Environment Agnostic Representation for Visual Reinforcement Learning	263
<i>Hyesong Choi (Ewha w. university), Hunsang Lee (Yonsei Univ.), Seongwon Jeong (Ewha Womans University), and Dongbo Min (Ewha Womans University)</i>	
Test-time Personalizable Forecasting of 3D Human Poses	274
<i>Qiongjie Cui (Nanjing University of Science and Technology), Huaijiang Sun (Nanjing University of Science and Technology), Jianfeng Lu (Nanjing University of Science and Technology), Weiqing Li (Nanjing University of Science and Technology), Bin Li (Tianjin AiForward Science and Technology), Hongwei Yi (Max Planck Institute for Intelligent Systems), and Haofan Wang (Carnegie Mellon University)</i>	
HM-ViT: Hetero-Modal Vehicle-to-Vehicle Cooperative Perception with Vision Transformer	284
<i>Hao Xiang (University of California, Los Angeles), Runsheng Xu (University of California, Los Angeles), and Jiaqi Ma (University of California, Los Angeles)</i>	

Vision and Graphics

Efficient Neural Supersampling on a Novel Gaming Dataset	296
<i>Antoine Mercier (Qualcomm Technologies Inc), Ruan Erasmus (Qualcomm), Yashesh Savani (Qualcomm Technologies Inc), Manik Dhingra (Qualcomm Technologies Inc), Fatih Porikli (Australian National University), and Guillaume Berger (Qualcomm Technologies Inc.)</i>	
Locally Stylized Neural Radiance Fields	307
<i>Hong-Wing Pang (The Hong Kong University of Science and Technology), Binh-Son Hua (Trinity College Dublin), and Sai-Kit Yeung (Hong Kong University of Science and Technology)</i>	
NEMTO: Neural Environment Matting for Novel View and Relighting Synthesis of Transparent Objects	317
<i>Dongqing Wang (EPFL), Tong Zhang (EPFL), and Sabine Ssstrunk (EPFL)</i>	

DDColor: Towards Photo-Realistic Image Colorization via Dual Decoders	328
<i>Xiaoyang Kang (Alibaba), Tao Yang (Alibaba), Wenqi Ouyang (Alibaba), Peiran Ren (Alibaba), Lingzhi Li (Alibaba Group), and Xuansong Xie (Alibaba)</i>	
IntrinsicNeRF: Learning Intrinsic Neural Radiance Fields for Editable Novel View Synthesis.....	339
<i>Weicai Ye (Zhejiang University), Shuo Chen (Zhejiang University), Chong Bao (Zhejiang University), Hujun Bao (Zhejiang University), Marc Pollefeys (ETH Zurich / Microsoft), Zhaopeng Cui (Zhejiang University), and Guofeng Zhang (Zhejiang University)</i>	
PARIS: Part-level Reconstruction and Motion Analysis for Articulated Objects	352
<i>Jiayi Liu (Simon Fraser University), Ali Mahdavi-Amiri (Simon Fraser University), and Manolis Savva (Simon Fraser University)</i>	
ReMoDiffuse: Retrieval-Augmented Motion Diffusion Model	364
<i>Mingyuan Zhang (Nanyang Technological University), Xinying Guo (Nanyang Technological University), Liang Pan (Nanyang Technological University), Zhongang Cai (SenseTime International Pte Ltd), Fangzhou Hong (Nanyang Technological University), Huirong Li (Nanyang Technological University), Lei Yang (Sensetime Group Limited), and Ziwei Liu (Nanyang Technological University)</i>	
DS-Fusion: Artistic Typography via Discriminated and Stylized Diffusion	374
<i>Maham Tanveer (Simon Fraser University), Yizhi Wang (Simon Fraser University), Ali Mahdavi-Amiri (Simon Fraser University), and Hao Zhang (Simon Fraser University)</i>	
Dynamic Mesh-Aware Radiance Fields	385
<i>Yi-Ling Qiao (University of Maryland, College Park), Alexander Gao (University of Maryland, College Park), Yiran Xu (University of Maryland, College Park), Yue Feng (University of Maryland), Jia-Bin Huang (University of Maryland College Park), and Ming C. Lin (UMD-CP & UNC-CH)</i>	
Neural Reconstruction of Relightable Human Model from Monocular Video	397
<i>Wenzhang Sun (Beijing Institute of Technology), Yunlong Che (AI² Robotics), Han Huang (AI² Robotics), and Yandong Guo (AI² Robotics)</i>	
Neural Microfacet Fields for Inverse Rendering	408
<i>Alexander Mai (University of California, San Diego), Dor Verbin (Google Research), Falko Kuester (UC San Diego), and Sara Fridovich-Keil (UC Berkeley)</i>	
A Theory of Topological Derivatives for Inverse Rendering of Geometry	419
<i>Ishit Mehta (University of California San Diego), Manmohan Chandraker (UC San Diego), and Ravi Ramamoorthi (University of California San Diego)</i>	
Vox-E: Text-Guided Voxel Editing of 3D Objects	430
<i>Etai Sella (Tel Aviv University), Gal Fiebelman (Tel Aviv University), Peter Hedman (Google Research), and Hadar Averbuch-Elor (Tel Aviv University)</i>	

StegaNeRF: Embedding Invisible Information within Neural Radiance Fields	441
<i>Chenxin Li (Xiamen University), Brandon Y. Feng (University of Maryland, College Park), Zhiwen Fan (University of Texas at Austin), Panwang Pan (ByteDance), and Zhangyang Wang (University of Texas at Austin)</i>	
GlobalMapper: Arbitrary-Shaped Urban Layout Generation	454
<i>Liu He (Purdue University) and Daniel Aliaga (Purdue University)</i>	
Urban Radiance Field Representation with Deformable Neural Mesh Primitives	465
<i>Fan Lu (Tongji University), Yan Xu (Chinese University of Hong Kong), Guang Chen (Tongji University), Hongsheng Li (The Chinese University of Hong Kong), Kwan-Yee Lin (SenseTime Research), and Changjun Jiang (Tongji University)</i>	
End2End Multi-View Feature Matching with Differentiable Pose Optimization	477
<i>Barbara Roessle (Technical University of Munich) and Matthias Nießner (Technical University of Munich)</i>	
Tree-Structured Shading Decomposition	488
<i>Chen Geng (Stanford University), Hong-Xing Yu (Stanford University), Sharon Zhang (Princeton University), Maneesh Agrawala (Stanford University), and Jiajun Wu (Stanford University)</i>	
Lens Parameter Estimation for Realistic Depth of Field Modeling	499
<i>Dominique Piché-Meunier (Université Laval), Yannick Hold-Geoffroy (Adobe Research), Jianming Zhang (Adobe Research), and Jean-François Lalonde (Université Laval)</i>	
AttT2M: Text-Driven Human Motion Generation with Multi-Perspective Attention Mechanism	509
<i>Chongyang Zhong (Institute of Computing Technology, Chinese Academy of Sciences), Lei Hu (Institute of Computing Technology, Chinese Academy of Sciences), Zihao Zhang (Institute of Computing Technology, Chinese Academy of Sciences), and Shihong Xia (institute of computing technology of the Chinese academy of sciences)</i>	
Cross-modal Latent Space Alignment for Image to Avatar Translation	520
<i>Manuel Ladron de Guevara (Carnegie Mellon University), Jose Echevarria (Adobe System Inc.), Yijun Li (Adobe Research), Yannick Hold-Geoffroy (Adobe Research), Cameron Smith (Adobe Research), and Daichi Ito (Adobe Inc.)</i>	
Computationally-Efficient Neural Image Compression with Shallow Decoders	530
<i>Yibo Yang (University of California, Irvine) and Stephan Mandt (University of California, Irvine)</i>	

Segmentation, Grouping and Shape Analysis

3D Instance Segmentation via Enhanced Spatial and Semantic Supervision	541
<i>Salwa Al Khatib (Mohamed Bin Zayed University of Artificial Intelligence), Mohamed El Amine Amine Boudjoghra (Mohamed Bin Zayed University of Artificial Intelligence), Jean Lahoud (MBZUAI), and Fahad Shahbaz Khan (MBZUAI)</i>	
Learning Neural Eigenfunctions for Unsupervised Semantic Segmentation	551
<i>Zhijie Deng (Shanghai Jiao Tong University) and Yucen Luo (Max Planck Institute for Intelligence Systems)</i>	

Divide and Conquer: 3D Point Cloud Instance Segmentation With Point-Wise Binarization	562
<i>Weiguang Zhao (Duke Kunshan University), Yuyao Yan (Xi'an Jiaotong-Liverpool University), Chaolong Yang (Duke Kunshan University), Jianan Ye (Xi'an Jiaotong-Liverpool Univ.), Xi Yang (Xi'an Jiaotong Liverpool University), and Kaizhu Huang (Duke Kunshan University)</i>	
Point2Mask: Point-supervised Panoptic Segmentation via Optimal Transport	572
<i>Wentong Li (Zhejiang University), Yuqian Yuan (Zhejiang University), Song Wang (Zhejiang University), Jianke Zhu (Zhejiang University), Jianshu Li (Ant Group), Jian Liu (Ant Group), and Lei Zhang (Hong Kong Polytechnic University, Hong Kong, China)</i>	
Handwritten and Printed Text Segmentation: A Signature Case Study	582
<i>Sina Gholamian (Thomson Reuters) and Ali Vahdat (Thomson Reuters)</i>	
Semantic-Aware Implicit Template Learning via Part Deformation Consistency	593
<i>Sihyeon Kim (Korea University), Minseok Joo (Korea University), Jaewon Lee (Korea University), Juyeon Ko (Korea University), Juhan Cha (Korea University), and Hyunwoo J. Kim (Korea University)</i>	
LeaF: Learning Frames for 4D Point Cloud Sequence Understanding	604
<i>Yunze Liu (Tsinghua University), Junyu Chen (Tsinghua University), Zekai Zhang (Tsinghua University), Jingwei Huang (Huawei), and Li Yi (Tsinghua University)</i>	
MARS: Model-agnostic Biased Object Removal without Additional Supervision for Weakly-Supervised Semantic Segmentation	614
<i>Sanghyun Jo (OGQ), In-Jae Yu (Visual Display Business, Samsung Electronics), and Kyungsu Kim (Massachusetts General Hospital and Harvard Medical School)</i>	
USAGE: A Unified Seed Area Generation Paradigm for Weakly Supervised Semantic Segmentation....	624
<i>Zelin Peng (Shanghai Jiao Tong University), Guanchun Wang (Xidian university), Lingxi Xie (Huawei Inc.), Dongsheng Jiang (Huawei Cloud & AI), Wei Shen (SJTU), and Qi Tian (Huawei Cloud BU)</i>	
XMem++: Production-level Video Segmentation From Few Annotated Frames	635
<i>Maksym Bekuzarov (MBZUAI), Ariana Bermudez (MBZUAI), Joon-Young Lee (Adobe Research), and Hao Li (Pinscreen / MBZUAI)</i>	
SIGMA: Scale-Invariant Global Sparse Shape Matching	645
<i>Maolin Gao (TUM), Paul Roetzer (Uni Bonn), Marvin Eisenberger (TU Munich), Zorah Löhner (University of Siegen), Michael Moeller (University of Siegen), Daniel Cremers (TU Munich), and Florian Bernard (University of Bonn)</i>	
Self-Calibrated Cross Attention Network for Few-Shot Segmentation	655
<i>Qianxiong Xu (Nanyang Technological University), Wenting Zhao (Nanjing University of Science and Technology), Guosheng Lin (Nanyang Technological University), and Cheng Long (Nanyang Technological University)</i>	

Multi-granularity Interaction Simulation for Unsupervised Interactive Segmentation	666
<i>Kehan Li (Peking University), Yian Zhao (Peking University Shenzhen Graduate School), Zhennan Wang (Peng Cheng Laboratory), Zesen Cheng (School of Electronic and Computer Engineering, Peking University), Peng Jin (Shenzhen Graduate School, Peking University), Xiangyang Ji (Tsinghua University), Li Yuan (Peking University), Chang Liu (Tsinghua University), and Jie Chen (Peking University)</i>	
Texture Learning Domain Randomization for Domain Generalized Segmentation	677
<i>Sunghwan Kim (Agency for Defense Development), Dae-hwan Kim (Agency for Defense Development), and Hoseong Kim (Agency for Defense Development)</i>	
Unsupervised Video Object Segmentation with Online Adversarial Self-Tuning	688
<i>Tiankang Su (Nanjing University of Information Science & Technology), Huihui Song (nuist), Dong Liu (Netflix Inc), Bo Liu (JD.com), and Qingshan Liu (Nanjing University of Information Science & Technology)</i>	
Exploring Open-Vocabulary Semantic Segmentation from CLIP Vision Encoder Distillation Only..	699
<i>Jun Chen (King Abdullah University of Science and Technology), Deyao Zhu (King Abdullah University of Science and Technology), Guocheng Qian (KAUST), Bernard Ghanem (KAUST), Zhicheng Yan (Meta Reality Labs), Chenchen Zhu (Meta), Fanyi Xiao (Meta), Sean Chang Culatana (Facebook), and Mohamed Elhoseiny (KAUST)</i>	
RbA: Segmenting Unknown Regions Rejected by All	711
<i>Nazir Nayal (Koç University), Misra Yavuz (Koç University), João F. Henriques (University of Oxford), and Fatma Güney (Koc University)</i>	
SEMPART: Self-supervised Multi-resolution Partitioning of Image Semantics	723
<i>Sriram Ravindran (Adobe) and Debraj Basu (Adobe)</i>	
Multi-Object Discovery by Low-Dimensional Object Motion	734
<i>Sadra Safadoust (Koc University) and Fatma Güney (Koc University)</i>	
MemorySeg: Online LiDAR Semantic Segmentation with a Latent Memory	745
<i>Enxu Li (University of Toronto), Sergio Casas (University of Toronto), and Raquel Urtasun (Waabi)</i>	
Treating Pseudo-labels Generation as Image Matting for Weakly Supervised Semantic Segmentation	755
<i>Changwei Wang (Institute of Automation, Chinese Academy of Sciences; School of Artificial Intelligence, University of Chinese Academy of Sciences), Rongtao Xu (Institute of Automation, Chinese Academy of Sciences; School of Artificial Intelligence, University of Chinese Academy of Sciences), Shibiao Xu (School of Artificial Intelligence, Beijing University of Posts and Telecommunications), Weiliang Meng (Institute of Automation, Chinese Academy of Sciences; School of Artificial Intelligence, University of Chinese Academy of Sciences), and Xiaopeng Zhang (Institute of Automation, Chinese Academy of Sciences; School of Artificial Intelligence, University of Chinese Academy of Sciences)</i>	
BoxSnake: Polygonal Instance Segmentation with Box Supervision	766
<i>Rui Yang (Tsinghua University), Lin Song (Tencent AI Lab), Yixiao Ge (Tencent), and Xiu Li (Tsinghua University)</i>	

Dynamic Token Pruning in Plain Vision Transformers for Semantic Segmentation	777
<i>Quan Tang (South China University of Technology), Bowen Zhang (University of Adelaide), Jiajun Liu (CSIRO), Fagui Liu (South China University of Technology), and Yifan Liu (University of Adelaide)</i>	
Instance Neural Radiance Field	787
<i>Yichen Liu (The Hong Kong University of Science and Technology), Benran Hu (The Hong Kong University of Science and Technology), Junkai Huang (The Hong Kong University of Science and Technology), Yu-Wing Tai (Kuaishou Technology / HKUST), and Chi-Keung Tang (Hong Kong University of Science and Technology)</i>	
Global Knowledge Calibration for Fast Open-Vocabulary Segmentation	797
<i>Kunyang Han (Beijing Jiaotong University), Yong Liu (Tsinghua University), Jun Hao Liew (ByteDance), Henghui Ding (Nanyang Technological University), Jiajun Liu (bytedance), Yitong Wang (Bytedance), Yansong Tang (Tsinghua University), Yujiu Yang (Tsinghua University), Jiashi Feng (ByteDance), Yao Zhao (Beijing Jiaotong University), and Yunchao Wei (UTS)</i>	
Diffusion-based Image Translation with Label Guidance for Domain Adaptive Semantic Segmentation	808
<i>Duo Peng (Singapore University of Technology and Design), Ping Hu (Boston University), Qihong Ke (Monash University), and Jun Liu (Singapore University of Technology and Design)</i>	
Boosting Semantic Segmentation from the Perspective of Explicit Class Embeddings	821
<i>Yuhe Liu (Beihang university), Chuanjian Liu (Huawei Noah's Ark Lab), Kai Han (Huawei Noah's Ark Lab), Quan Tang (South China University of Technology), and Zengchang Qin (Intelligent Computing & Machine Learning Lab, School of ASEE, Beihang University)</i>	
The Making and Breaking of Camouflage	832
<i>Hala Lamdouar (University of Oxford), Weidi Xie (Shanghai Jiao Tong University), and Andrew Zisserman (University of Oxford)</i>	
CoinSeg: Contrast Inter- and Intra- Class Representations for Incremental Segmentation	843
<i>Zekang Zhang (Beijing Institute of Technology), Guangyu Gao (Beijing Institute of Technology), Jianbo Jiao (University of Birmingham; University of Oxford), Chi Harold Liu (Beijing Institute of Technology), and Yunchao Wei (UTS)</i>	
Few-Shot Physically-Aware Articulated Mesh Generation via Hierarchical Deformation	854
<i>Xueyi Liu (Tsinghua University), Bin Wang (Beijing Institute for General Artificial Intelligence), He Wang (Peking University), and Li Yi (Tsinghua University)</i>	
HAL3D: Hierarchical Active Learning for Fine-Grained 3D Part Labeling	865
<i>Fenggen Yu (Simon Fraser University), Yiming Qian (Amazon), Francisca Gil-Ureta (Amazon), Brian Jackson (Amazon.com), Eric Bennett (Amazon), and Hao Zhang (Simon Fraser University)</i>	

FreeCOS: Self-Supervised Learning from Fractals and Unlabeled Images for Curvilinear Object Segmentation	876
<i>Tianyi Shi (Huazhong University of Science and Technology), Xiaohuan Ding (Huazhong University of Science and Technology), Liang Zhang (Huazhong University of science and technology), and Xin Yang (Huazhong University of Science and Technology)</i>	
MasQCLIP for Open-Vocabulary Universal Image Segmentation	887
<i>Xin Xu (Peking University), Tianyi Xiong (Tsinghua University), Zheng Ding (University of California, San Diego), and Zhuowen Tu (UC San Diego)</i>	
CTVIS: Consistent Training for Online Video Instance Segmentation	899
<i>Kaining Ying (Zhejiang University of Technology), Qing Zhong (The University of Adelaide), Weian Mao (University of Adelaide), Zhenhua Wang (Northwest A&F University), Hao Chen (Zhejiang University), Lin Yuanbo Wu (Swansea University), Yifan Liu (University of Adelaide), Chengxiang Fan (Zhejiang University), Yunzhi Zhuge (University of Adelaide), and Chunhua Shen (University of Adelaide, Australia)</i>	
A Generalist Framework for Panoptic Segmentation of Images and Videos	909
<i>Ting Chen (Google), Lala Li (Google), Saurabh Saxena (Google Research), Geoffrey Hinton (Google), and David J. Fleet (University of Toronto)</i>	
Spectrum-guided Multi-granularity Referring Video Object Segmentation	920
<i>Bo Miao (University of Western Australia), Mohammed Bennamoun (University of Western Australia), Yongsheng Gao (Griffith University), and Ajmal Mian (University of Western Australia)</i>	
Space Engage: Collaborative Space Supervision for Contrastive-Based Semi-Supervised Semantic Segmentation	931
<i>Changqi Wang (Northeastern University), Haoyu Xie (Northeastern University), Yuhui Yuan (Microsoft Research), Chong Fu (Northeastern University), and Xiangyu Yue (The Chinese University of Hong Kong)</i>	
Adaptive Superpixel for Active Learning in Semantic Segmentation	943
<i>Hoyoung Kim (POSTECH), Minhyeon Oh (POSTECH), Sehyun Hwang (POSTECH), Suha Kwak (POSTECH), and Jungseul Ok (POSTECH)</i>	
Multimodal Variational Auto-encoder based Audio-Visual Segmentation	954
<i>Yuxin Mao (Northwestern Polytechnical University), Jing Zhang (Australian National University), Mochu Xiang (Northwestern Polytechnical University), Yiran Zhong (Australian National University), and Yuchao Dai (Northwestern Polytechnical University)</i>	
Isomer: Isomeric Transformer for Zero-shot Video Object Segmentation	966
<i>Yichen Yuan (Dalian University of Technology), Yifan Wang (Dalian University of Technology), Lijun Wang (Dalian University of Technology), Xiaoqi Zhao (Dalian University of Technology), Huchuan Lu (Dalian University of Technology), Yu Wang (Guangdong OPPO Mobile Telecommunications Co., Ltd), Weibo Su (OPPO), and Lei Zhang (Hong Kong Polytechnic University, Hong Kong, China)</i>	

2D-3D Interlaced Transformer for Point Cloud Segmentation with Scene-Level Supervision	977
<i>Cheng-Kun Yang (National Taiwan University), Min-Hung Chen (NVIDIA), Yung-Yu Chuang (National Taiwan University), and Yen-Yu Lin (National Yang Ming Chiao Tung University)</i>	
Foreground-Background Separation through Concept Distillation from Generative Image Foundation Models	988
<i>Mischa Dombrowski (Friedrich-Alexander-Universität Erlangen-Nürnberg), Hadrien Reynaud (Imperial College London), Matthew Baugh (Imperial College London), and Bernhard Kainz (Imperial College London, FAU Erlangen-Nürnberg)</i>	
SegPrompt: Boosting Open-World Segmentation via Category-Level Prompt Learning	999
<i>Muzhi Zhu (Shanghai Jiao Tong University), Hengtao Li (Zhejiang University), Hao Chen (Zhejiang University), Chengxiang Fan (Zhejiang University), Weian Mao (University of Adelaide), Chenchen Jing (Zhejiang University), Yifan Liu (University of Adelaide), and Chunhua Shen (Zhejiang University)</i>	
Monte Carlo Linear Clustering with Single-Point Supervision is Enough for Infrared Small Target Detection	1009
<i>Boyang Li (National University of Defense Technology), Yingqian Wang (National University of Defense Technology), Longguang Wang (National University of Defense Technology), Fei Zhang (Shanghai Jiao Tong University), Ting Liu (National University of Defense Technology), Zaiping Lin (National University of Defense Technology), Wei An (National University of Defense Technology), and Yulan Guo (Sun Yat-sen University)</i>	
A Simple Framework for Open-Vocabulary Segmentation and Detection	1020
<i>Hao Zhang (hkust), Feng Li (Hong Kong University of Science and Technology), Xueyan Zou (university of wisconsin, madison), Shilong Liu (Tsinghua University), Chunyuan Li (Microsoft Research, Redmond), Jianwei Yang (Microsoft Research), and Lei Zhang (International Digital Economy Academy, IDEA)</i>	
Source-free Depth for Object Pop-out	1032
<i>Zongwei Wu (Univ. Bourgogne Franche-Comte, France), Danda Pani Paudel (ETH Zürich, INSAIT Sofia), Deng-Ping Fan (ETH Zurich), Jingjing Wang (AUST), Shuo Wang (ETH Zurich), Cédric Demonceaux (Université de Bourgogne), Radu Timofte (University of Wurzburg & ETH Zurich), and Luc Van Gool (ETH Zurich)</i>	
DynaMITE: Dynamic Query Bootstrapping for Multi-object Interactive Segmentation Transformer	1043
<i>Amit Kumar Rana (Computer Vision Group, RWTH Aachen University), Sabarinath Mahadevan (RWTH Aachen University), Alexander Hermans (RWTH Aachen University), and Bastian Leibe (RWTH Aachen University)</i>	
Atmospheric Transmission and Thermal Inertia Induced Blind Road Segmentation with a Large-Scale Dataset TBRSD	1053
<i>Junzhang Chen (beihang university) and Xiangzhi Bai (Beihang University)</i>	

Informative Data Mining for One-Shot Cross-Domain Semantic Segmentation	1064
<i>Yuxi Wang (CASIA), Jian Liang (CASIA), Jun Xiao (University of Chinese Academy of Sciences), Shuqi Mei (Tencent), Yuran Yang (Tencent), and Zhaoxiang Zhang (Chinese Academy of Sciences, China)</i>	
Homography Guided Temporal Fusion for Road Line and Marking Segmentation	1075
<i>Shan Wang (ANU, CSIRO), Chuong Nguyen (CSIRO Data61), Jiawei Liu (Australian National University), Kaihao Zhang (Australian National University), Wenhan Luo (Sun Yat-sen University), Yanhao Zhang (Australian National University), Sundaram Muthu (Csiro), Fahira Afzal Maken (The University of Sydney), and Hongdong Li (Australian National University, Australia)</i>	
Open-Vocabulary Semantic Segmentation with Decoupled One-Pass Network	1086
<i>Cong Han (Meituan), Yujie Zhong (University of Oxford), Dengjie Li (Meituan), Kai Han (The University of Hong Kong), and Lin Ma (Meituan)</i>	
TCOVIS: Temporally Consistent Online Video Instance Segmentation	1097
<i>Junlong Li (Tsinghua University), Bingyao Yu (Tsinghua University), Yongming Rao (Tsinghua University), Jie Zhou (Tsinghua University), and Jiwen Lu (Tsinghua University)</i>	
FPR: False Positive Rectification for Weakly Supervised Semantic Segmentation	1108
<i>Liyi Chen (PolyU), Chenyang Lei (CAIR, HKISI-CAS / Princeton University), Ruihuang Li (Hong Kong Polytechnic University), Shuai Li (The Hong Kong Polytechnic University), Zhaoxiang Zhang (Chinese Academy of Sciences, China), and Lei Zhang (Hong Kong Polytechnic University, Hong Kong, China)</i>	
Stochastic Segmentation with Conditional Categorical Diffusion Models	1119
<i>Lukas Zbinden (University of Bern), Lars Doorenbos (University of Bern), Theodoros Pissas (University of Bern), Adrian Thomas Huber (Department of Diagnostic, Interventional and Pediatric Radiology, Inselspital University Hospital, Bern, Switzerland), Raphael Sznitman (University of Bern), and Pablo Márquez-Neila (University of Bern)</i>	
SegGPT: Towards Segmenting Everything in Context	1130
<i>Xinlong Wang (Beijing Academy of Artificial Intelligence), Xiaosong Zhang (Beijing Academy of Artificial Intelligence), Yue Cao (Beijing Academy of Artificial Intelligence), Wen Wang (Zhejiang University), Chunhua Shen (University of Adelaide, Australia), and Tiejun Huang (Peking University)</i>	
Open-vocabulary Panoptic Segmentation with Embedding Modulation	1141
<i>Xi Chen (Hong Kong University), Shuang Li (MIT), Ser-Nam Lim (Facebook AI), Antonio Torralba (MIT), and Hengshuang Zhao (The University of Hong Kong)</i>	
Residual Pattern Learning for Pixel-Wise Out-of-Distribution Detection in Semantic Segmentation	1151
<i>Yuyuan Liu (University of Adelaide), Choubo Ding (University of Adelaide), Yu Tian (Harvard University), Guansong Pang (Singapore Management University), Vasileios Belagiannis (Friedrich-Alexander-Universität Erlangen-Nürnberg), Ian Reid (University of Adelaide, Australia), and Gustavo Carneiro (University of Surrey)</i>	

Zero-guidance Segmentation Using Zero Segment Labels	1162
<i>Pitchaporn Rewatbowornwong (Vistec), Nattanat Chatthee (Vidyasirimedhi Institute of Science and Technology), Ekapol Chuangsuwanich (Chulalongkorn University), and Supasorn Suwajanakorn (Vistec)</i>	
Model Calibration in Dense Classification with Adaptive Label Perturbation	1173
<i>Jiawei Liu (Australian National University), Changkun Ye (Australian National University), Shan Wang (ANU, CSIRO), Ruikai Cui (Australian National University), Jing Zhang (Australian National University), Kaihao Zhang (Australian National University), and Nick Barnes (ANU)</i>	
Enhanced Soft Label for Semi-Supervised Semantic Segmentation	1185
<i>Jie Ma (Sun Yat-Sen University), Chuan Wang (Face++, (Megvii), Yang Liu (Sun Yat-sen University), Liang Lin (Sun Yat-sen University), and Guanbin Li (Sun Yat-sen University)</i>	
MixReorg: Cross-Modal Mixed Patch Reorganization is a Good Mask Learner for Open-World Semantic Segmentation	1196
<i>Kaixin Cai (SUN YAT-SEN UNIVERSITY), Pengzhen Ren (Sun Yat-sen University), Yi Zhu (Huawei Noah's Ark Lab), Hang Xu (Huawei Noah's Ark Lab), Jianzhuang Liu (Huawei Noah's Ark Lab), Changlin Li (University of Technology Sydney), Guangrun Wang (University of Oxford), and Xiaodan Liang (Sun Yat-sen University)</i>	
DiffuMask: Synthesizing Images with Pixel-level Annotations for Semantic Segmentation Using Diffusion Models	1206
<i>Weijia Wu (Zhejiang University), Yuzhong Zhao (University of Chinese Academy of Sciences), Mike Zheng Shou (National University of Singapore), Hong Zhou (Zhejiang University), and Chunhua Shen (University of Adelaide, Australia)</i>	
Alignment Before Aggregation: Trajectory Memory Retrieval Network for Video Object Segmentation	1218
<i>Rui Sun (University of Science and Technology of China), Yuan Wang (University of Science and Technology of China), Huayu Mai (University of Science and Technology of China), Tianzhu Zhang (University of Science and Technology of China), and Feng Wu (University of Science and Technology of China)</i>	
Semi-Supervised Semantic Segmentation under Label Noise via Diverse Learning Groups	1229
<i>Peixia Li (The University of Sydney), Pulak Purkait (Amazon), Thalaiyasingam Ajanthan (Amazon), Majid Abdolshah (Amazon), Ravi Garg (University of Adelaide), Hisham Husain (Amazon), Chenchen Xu (The Australian National University), Stephen Gould (Australian National University, Australia), Wanli Ouyang (The University of Sydney), and Anton van den Hengel (University of Adelaide)</i>	
SUMMIT: Source-Free Adaptation of Uni-Modal Models to Multi-Modal Targets	1239
<i>Cody Simons (University of California, Riverside), Dripta S. Raychaudhuri (AWS AI Labs), Sk Miraj Ahmed (University of California Riverside), Suyu You (U.S. Army Research Laboratory), Konstantinos Karydis (University of California Riverside), and Amit K. Roy-Chowdhury (University of California, Riverside)</i>	

Class-incremental Continual Learning for Instance Segmentation with Image-level Weak Supervision	1250
<i>Yu-Hsing Hsieh (National Taiwan University), Guan-Sheng Chen (National Taiwan University), Shun-Xian Cai (National Taiwan University), Ting-Yun Wei (National Taiwan University), Hwei-Fang Yang (National Sun Yat-sen University), and Chu-Song Chen (National Taiwan University)</i>	
Coarse-to-Fine Amodal Segmentation with Shape Prior	1262
<i>Jianxiong Gao (Fudan University), Xuelin Qian (Fudan University), Yikai Wang (Fudan University), Tianjun Xiao (Amazon), Tong He (Amazon), Zheng Zhang (Amazon), and Yanwei Fu (Fudan University)</i>	
Rethinking Amodal Video Segmentation from Learning Supervised Signals with Object-centric Representation	1272
<i>Ke Fan (Fudan University), Jingshi Lei (Fudan University), Xuelin Qian (Fudan University), Miaopeng Yu (Fudan University), Tianjun Xiao (Amazon), Tong He (Amazon), Zheng Zhang (Amazon), and Yanwei Fu (Fudan University)</i>	
DVIS: Decoupled Video Instance Segmentation Framework	1282
<i>Tao Zhang (Wuhan University), Xingye Tian (kuaishou), Yu Wu (Wuhan University), Shunping Ji (Wuhan University), Xuebo Wang (kuaishou), Yuan Zhang (kuaishou), and Pengfei Wan (Kuaishou Technology)</i>	
3D Segmentation of Humans in Point Clouds with Synthetic Data	1292
<i>Ayça Takmaz (ETH Zurich), Jonas Schult (RWTH Aachen University), Irem Kaftan (ETH Zurich), Mertcan Akçay (ETH Zurich), Bastian Leibe (RWTH Aachen University), Robert Sumner (Disney Research Studios / ETH Zurich), Francis Engelmann (ETH Zurich), and Siyu Tang (ETH Zurich)</i>	
WaterMask: Instance Segmentation for Underwater Imagery	1305
<i>Shijie Lian (Hainan University), Hua Li (Hainan University), Runmin Cong (Shandong University), Suqi Li (Hainan University), Wei Zhang (Shandong University), and Sam Kwong (City University of Hong Kong)</i>	
Tracking Anything with Decoupled Video Segmentation	1316
<i>Ho Kei Cheng (University of Illinois Urbana-Champaign), Seoung Wug Oh (Adobe Research), Brian Price (Adobe Research), Alexander Schwing (University of Illinois Urbana-Champaign), and Joon-Young Lee (Adobe Research)</i>	

Recognition: Categorization

Cross Contrasting Feature Perturbation for Domain Generalization	1327
<i>Chenming Li (Southern University of Science and Technology), Daoan Zhang (University of Rochester), Wenjian Huang (Southern University of Science and Technology), and Jianguo Zhang (Southern University of Science and Technology)</i>	
Flexible Visual Recognition by Evidential Modeling of Confusion and Ignorance	1338
<i>Lei Fan (Northwestern University), Bo Liu (Wormpex AI Research), Haoxiang Li (Wormpex AI Research), Ying Wu (Northwestern University), and Gang Hua (Wormpex AI Research)</i>	

CDUL: CLIP-Driven Unsupervised Learning for Multi-Label Image Classification	1348
<i>Rabab Abdelfattah (University of South Carolina, University of Southern Mississippi), Qing Guo (A*STAR), Xiaoguang Li (University of South Carolina), Xiaofeng Wang (University of South Carolina), and Song Wang (University of South Carolina)</i>	
RankMixup: Ranking-Based Mixup Training for Network Calibration	1358
<i>Jongyoun Noh (Yonsei University), Hyekang Park (Yonsei University), Junghyup Lee (Yonsei University), and Bumsub Ham (Yonsei University)</i>	
Label-Noise Learning with Intrinsically Long-Tailed Data	1369
<i>Yang Lu (Xiamen University), Yiliang Zhang (Xiamen University), Bo Han (HKBU / RIKEN), Yiu-ming Cheung (Hong Kong Baptist University), and Hanzi Wang (Xiamen University)</i>	
Parallel Attention Interaction Network for Few-Shot Skeleton-Based Action Recognition	1379
<i>Xingyu Liu (Xi'an Jiaotong University), Sanping Zhou (Xi'an Jiaotong University), Le Wang (Xi'an Jiaotong University), and Gang Hua (Wormpex AI Research)</i>	
Rethinking Mobile Block for Efficient Attention-based Models	1389
<i>Jiangning Zhang (Zhejiang University), Xiangtai Li (Nanyang Technological University), Jian Li (Tencent Youtu), Liang Liu (Tencent), Zhucun Xue (WuHan university), Boshen Zhang (Tencent), Zhengkai Jiang (Tencent Youtu Lab), Tianxin Huang (Zhejiang University), Yabiao Wang (Tencent), and Chengjie Wang (Tencent, Shanghai Jiao Tong University)</i>	
Read-only Prompt Optimization for Vision-Language Few-shot Learning	1401
<i>Dongjun Lee (Korea University), Seokwon Song (Korea University), Jihee Suh (Korea University), Joonmyeong Choi (Korea University), Sanghyeok Lee (Korea University), and Hyunwoo J. Kim (Korea University)</i>	
Understanding Self-attention Mechanism via Dynamical System Perspective	1412
<i>Zhongzhan Huang (Sun Yat-Sen University), Mingfu Liang (Northwestern University), Jinghui Qin (Guangdong University of Technology), Shanshan Zhong (School of Computer Science and Engineering, Sun Yat-Sen University), and Liang Lin (Sun Yat-sen University)</i>	
Learning in Imperfect Environment: Multi-Label Classification with Long-Tailed Distribution and Partial Labels	1423
<i>Wenqiao Zhang (NUS), Changshuo Liu (National University of Singapore), Lingze Zeng (missing), Bengchin Ooi (NUS), Siliang Tang (Zhejiang University), and Yueting Zhuang (Zhejiang University)</i>	
What do neural networks learn in image classification? A frequency shortcut perspective	1433
<i>Shunxin Wang (University of Twente), Raymond Veldhuis (University of Twente), Christoph Brune (University of Twente), and Nicola Strisciuglio (University of Twente)</i>	
Inducing Neural Collapse to a Fixed Hierarchy-Aware Frame for Reducing Mistake Severity	1443
<i>Tong Liang (Ohio State University) and Jim Davis (Ohio State University)</i>	
Unified Out-Of-Distribution Detection: A Model-Specific Perspective	1453
<i>Reza Averly (The Ohio State University) and Wei-Lun Chao (The Ohio State University)</i>	

A Unified Framework for Robustness on Diverse Sampling Errors	1464
<i>Myeongho Jeon (Seoul National University), Myungjoo Kang (Seoul National University), and Joonseok Lee (Google Research & Seoul National University)</i>	
Scene-Aware Label Graph Learning for Multi-Label Image Classification	1473
<i>Xuelin Zhu (Southeast University), Jian Liu (Ant Group), Weijia Liu (Southeast University), Jiawei Ge (Southeast University), Bo Liu (Southeast University), and Jiuxin Cao (Southeast University)</i>	
Holistic Label Correction for Noisy Multi-Label Classification	1483
<i>Xiaobo Xia (The University of Sydney), Jiankang Deng (Imperial College London), Wei Bao (The University of Sydney), Yuxuan Du (University of Sydney), Bo Han (HKBU / RIKEN), Shiguang Shan (Institute of Computing Technology, Chinese Academy of Sciences), and Tongliang Liu (The University of Sydney)</i>	
Strip-MLP: Efficient Token Interaction for Vision MLP	1494
<i>Guiping Cao (Southern University of Science and Technology), Shengda Luo (Southern University of Science and Technology), Wenjian Huang (Southern University of Science and Technology), Xiangyuan Lan (Peng Cheng Laboratory), Dongmei Jiang (Pengcheng Lab), Yaowei Wang (PengCheng Laboratory), and Jianguo Zhang (Southern University of Science and Technology)</i>	
EQ-Net: Elastic Quantization Neural Networks	1505
<i>Ke Xu (Anhui University), Lei Han (Anhui University), Ye Tian (Anhui University), Shangshang Yang (Anhui University), and Xingyi Zhang (Anhui University)</i>	
Data-free Knowledge Distillation for Fine-grained Visual Categorization	1515
<i>Renrong Shao (East China Normal University), Wei Zhang (East China Normal University), Jianhua Yin (Shandong University), and Jun Wang (ECNU)</i>	
Shift from Texture-bias to Shape-bias: Edge Deformation-based Augmentation for Robust Object Recognition	1526
<i>Xilin He (Shenzhen University), Qinliang Lin (Shenzhen University), Cheng Luo (Shenzhen University), Weicheng Xie (Shenzhen University), Siyang Song (University of Cambridge), Feng Liu (Shenzhen University), and Linlin Shen (Shenzhen University)</i>	
Latent-OFFER: Detect, Mask, and Reconstruct with Latent Vectors for Occluded Facial Expression Recognition	1536
<i>Isack Lee (Chonnam National University), Eungi Lee (Chonnam National University), and Seok Bong Yoo (Chonnam National University)</i>	
DR-Tune: Improving Fine-tuning of Pretrained Visual Models by Distribution Regularization with Semantic Calibration	1547
<i>Nan Zhou (Beihang University), Jiabin Chen (Beihang University), and Di Huang (Beihang University, China)</i>	
Understanding the Feature Norm for Out-of-Distribution Detection	1557
<i>Jaewoo Park (Yonsei University), Jacky Chen Long Long Chai (Yonsei University), Jaeho Yoon (Yonsei University), and Andrew Beng Jin Jin Teoh (Yonsei University)</i>	

Multi-View Active Fine-Grained Visual Recognition	1568
<i>Ruoyi Du (Beijing University of Posts and Telecommunications), Wenqing Yu (Beijing University of Posts and Telecommunications), Heqing Wang (Beijing University of Posts and Telecommunications), Ting-En Lin (Alibaba Group), Dongliang Chang (Beijing University of Posts and Telecommunications), and Zhanyu Ma (Beijing University of Posts and Telecommunications)</i>	
DIFFGUARD: Semantic Mismatch-Guided Out-of-Distribution Detection Using Pre-Trained Diffusion Models	1579
<i>Ruiyuan Gao (The Chinese University of Hong Kong), Chenchen Zhao (The Chinese University of Hong Kong), Lanqing Hong (Huawei Noah's Ark Lab), and Qiang Xu (The Chinese University of Hong Kong)</i>	
Task-aware Adaptive Learning for Cross-domain Few-shot Learning	1590
<i>Yurong Guo (Beijing University of Posts and Telecommunications), Ruoyi Du (Beijing University of Posts and Telecommunications), Yuan Dong (Beijing University of Posts and Telecommunications), Timothy Hospedales (Edinburgh University), Yi-Zhe Song (University of Surrey), and Zhanyu Ma (Beijing University of Posts and Telecommunications)</i>	
Improving Adversarial Robustness of Masked Autoencoders via Test-time Frequency-domain Prompting	1600
<i>Qidong Huang (University of Science and Technology of China), Xiaoyi Dong (University of Science and Technology of China), Dongdong Chen (Microsoft Cloud AI), Yinpeng Chen (Microsoft), Lu Yuan (Microsoft), Gang Hua (Wormpex AI Research), Weiming Zhang (University of Science and Technology of China), and Nenghai Yu (University of Science and Technology of China)</i>	
Saliency Regularization for Self-Training with Partial Annotations	1611
<i>Shouwen Wang (Huazhong University of Science and Technology), Qian Wan (Wuhan Research Institute of Posts and Telecommunications), Xiang Xiang (Huazhong University of Science and Technology), and Zhigang Zeng (Huazhong University of Science and Technology)</i>	
Learning Gabor Texture Features for Fine-Grained Recognition	1621
<i>Lanyun Zhu (Singapore University of Technology and Design), Tianrun Chen (Zhejiang University), Jianxiong Yin (NVIDIA AI Tech Centre), Simon See (NVIDIA AI Tech Centre), and Jun Liu (Singapore University of Technology and Design)</i>	
UniFormerV2: Unlocking the Potential of Image ViTs for Video Understanding	1632
<i>Kunchang Li (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences), Yali Wang (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences), Yinan He (Shanghai AI Laboratory), Yizhuo Li (The University of Hong Kong), Yi Wang (Shanghai AI Laboratory), Limin Wang (Nanjing University), and Yu Qiao (Shanghai AI Laboratory)</i>	
RankMatch: Fostering Confidence and Consistency in Learning with Noisy Labels	1644
<i>Ziyi Zhang (Nanjing university), Weikai Chen (Tencent America), Chaowei Fang (Xidian University), Zhen Li (The Chinese University of Hong Kong, Shenzhen), Lechao Chen (Zhejiang Lab), Liang Lin (Sun Yat-sen University), and Guanbin Li (Sun Yat-sen University)</i>	

MetaGCD: Learning to Continually Learn in Generalized Category Discovery	1655
<i>Yanan Wu (Beijing Jiaotong University), Zhixiang Chi (UofT), Yang Wang (Concordia University), and Songhe Feng (School of Computer and Information Technology, Beijing Jiaotong University)</i>	
FerKD: Surgical Label Adaptation for Efficient Distillation	1666
<i>Zhiqiang Shen (MBZUAI)</i>	
Point-Query Quadtree for Crowd Counting, Localization, and More	1676
<i>Chengxin Liu (Huazhong University of Science and Technology), Hao Lu (Huazhong University of Science and Technology), Zhiguo Cao (Huazhong Univ. of Sci.&Tech.), and Tongliang Liu (The University of Sydney)</i>	
Nearest Neighbor Guidance for Out-of-Distribution Detection	1686
<i>Jaewoo Park (Yonsei University), Yoon Gyo Jung (Northeastern University), and Andrew Beng Jin Jin Teoh (Yonsei University)</i>	
Bayesian Optimization Meets Self-Distillation	1696
<i>HyunJae Lee (Lunit Inc.), Heon Song (Lunit Inc.), Hyeonsoo Lee (Lunit Inc.), Gi-hyeon Lee (Lunit), Suyeong Park (Lunit), and Donggeun Yoo (Lunit)</i>	
When Prompt-based Incremental Learning Does Not Meet Strong Pretraining	1706
<i>Yu-Ming Tang (Sun Yat-sen University, China), Yi-Xing Peng (Sun Yat-sen University, China), and Wei-Shi Zheng (Sun Yat-sen University, China)</i>	
When to Learn What: Model-Adaptive Data Augmentation Curriculum	1717
<i>Chengkai Hou (Jilin university), Jieyu Zhang (University of Washington), and Tianyi Zhou (University of Maryland, College Park)</i>	
Parametric Information Maximization for Generalized Category Discovery	1729
<i>Florent Chiaroni (ETS Montreal), Jose Dolz (ETS Montreal), Ziko Imtiaz Masud (Thales), Amar Mitiche (INRS), and Ismail Ben Ayed (ETS Montreal)</i>	
Boosting Few-shot Action Recognition with Graph-guided Hybrid Matching	1740
<i>Jiazheng Xing (Zhejiang University), Mengmeng Wang (Zhejiang University), Yudi Ruan (Zhejiang University), Bofan Chen (Zhejiang University), Yaowei Guo (Zhejiang University), Boyu Mu (Zhejiang University), Guang Dai (State Grid Corporation of China), Jingdong Wang (Baidu), and Yong Liu (Zhejiang University)</i>	
Domain Generalization via Rationale Invariance	1751
<i>Liang Chen (University of Adelaide), Yong Zhang (Tencent AI Lab), Yibing Song (Fudan University), Anton van den Hengel (University of Adelaide), and Lingqiao Liu (University of Adelaide)</i>	
Masked Spiking Transformer	1761
<i>Ziqing Wang (Sun Yat-sen University), Yuetong Fang (The Hong Kong University of Science and Technology, (Guangzhou)), Jiahang Cao (The Hong Kong University of Science and Technology, (Guangzhou)), Qiang Zhang (The Hong Kong University of Science and Technology, (Guangzhou)), Zhongrui Wang (The University of Hong Kong), and Renjing Xu (The Hong Kong University of Science and Technology, (Guangzhou))</i>	
Prototype Reminiscence and Augmented Asymmetric Knowledge Aggregation for Non-Exemplar Class-Incremental Learning	1772
<i>Wuxuan Shi (Wuhan University) and Mang Ye (Wuhan University)</i>	

Distilled Reverse Attention Network for Open-world Compositional Zero-Shot Learning	1782
<i>Yun Li (CSIRO DATA61), Zhe Liu (Jiangnan University), Saurav Jha (University of New South Wales), and Lina Yao (University of New South Wales)</i>	
Candidate-aware Selective Disambiguation Based On Normalized Entropy for Instance-dependent Partial-label Learning	1792
<i>Shuo He (University of Electronic Science and Technology of China), Guowu Yang (Big Data Research Center and School of Computer Science and Engineering, University of Electronic Science and Technology of China), and Lei Feng (Nanyang Technological University)</i>	
CLIPN for Zero-Shot OOD Detection: Teaching CLIP to Say No	1802
<i>Hualiang Wang (The Hong Kong University of Science and Technology), Yi Li (The Hong Kong University of Science and Technology), Huifeng Yao (The Hong Kong University of Science and Technology), and Xiaomeng Li (The Hong Kong University of Science and Technology)</i>	
Self-similarity Driven Scale-invariant Learning for Weakly Supervised Person Search	1813
<i>Benzhi Wang (CASIA), Yang Yang (Institute of Automation, Chinese Academy of Sciences), Jinlin Wu (Institute of Automation, Chinese Academy of Sciences, Beijing, China), Guo-jun Qi (OPPO Research / Westlake University), and Zhen Lei (NLPR, CASIA, China)</i>	
Sample-wise Label Confidence Incorporation for Learning with Noisy Labels	1823
<i>Chanho Ahn (Samsung Advanced Institute of Technology), Kikyung Kim (Seoul National University), Ji-won Baek (Samsung Advanced Institute of Technology), Jongin Lim (Seoul National University), and Seungju Han (Samsung Advanced Institute of Technology)</i>	
Combating Noisy Labels with Sample Selection by Mining High-Discrepancy Examples	1833
<i>Xiaobo Xia (The University of Sydney), Bo Han (HKBU / RIKEN), Yibing Zhan (JD Explore Academy), Jun Yu (University of Science and Technology of China), Mingming Gong (University of Melbourne), Chen Gong (Nanjing University of Science and Technology), and Tongliang Liu (The University of Sydney)</i>	
Spatial-Aware Token for Weakly Supervised Object Localization	1844
<i>Pingyu Wu (USTC), Wei Zhai (University of Science and Technology of China), Yang Cao (University of Science and Technology of China), Jiebo Luo (U. Rochester), and Zheng-Jun Zha (University of Science and Technology of China)</i>	

Explainable AI for CV

Towards Improved Input Masking for Convolutional Neural Networks	1855
<i>Sriram Balasubramanian (Microsoft Research) and Soheil Feizi (University of Maryland)</i>	
PDiscoNet: Semantically consistent part discovery for fine-grained recognition	1866
<i>Robert D van der Klis (INRAE), Stephan Alaniz (University of Tübingen), Massimiliano Mancini (University of Trento), Cassio F. Dantas (TETIS, INRAE, Univ Montpellier), Dino Ienco (INRAE), Zeynep Akata (University of Tübingen), and Diego Marcos (Inria)</i>	

Corrupting Neuron Explanations of Deep Visual Features	1877
<i>Divyansh Srivastava (UC San Diego), Tuomas Oikarinen (UC San Diego), and Tsui-Wei Weng (UCSD)</i>	
ICICLE: Interpretable Class Incremental Continual Learning	1887
<i>Dawid Rymarczyk (Jagiellonian University), Joost van de Weijer (Computer Vision Center), Bartosz Zieliński (Jagiellonian University), and Bartłomiej Twardowski (Computer Vision Center, UAB)</i>	
ProbVLM: Probabilistic Adapter for Frozen Vision-Language Models	1899
<i>Uddeshya Upadhyay (University of Tübingen), Shyamgopal Karthik (University of Tübingen), Massimiliano Mancini (University of Trento), and Zeynep Akata (University of Tübingen)</i>	
Out-of-Distribution Detection for Monocular Depth Estimation	1911
<i>Julia Hornauer (Ulm University), Adrian Holzbock (Ulm University), and Vasileios Belagiannis (Friedrich-Alexander-Universität Erlangen-Nürnberg)</i>	
Studying How to Efficiently and Effectively Guide Models with Explanations	1922
<i>Sukrut Rao (Max Planck Institute for Informatics), Moritz Böhle (Max-Planck-Institute for Informatics), Amin Parchami-Araghi (Max Planck Institute for Informatics), and Bernt Schiele (MPI Informatics)</i>	
Rosetta Neurons: Mining the Common Units in a Model Zoo	1934
<i>Amil Dravid (Northwestern University), Yossi Gandelsman (BAIR), Alexei A. Efros (UC Berkeley), and Assaf Shocher (Weizmann Institute of Science)</i>	
Prototype-based Dataset Comparison	1944
<i>Nanne van Noord (University of Amsterdam)</i>	
Learning to Identify Critical States for Reinforcement Learning from Videos	1955
<i>Haozhe Liu (King Abdullah University of Science and Technology), Mingchen Zhuge (Microsoft), Bing Li (KAUST), Yuhui Wang (King Abdullah University of Science and Technology), Francesco Faccio (The Swiss AI Lab IDSIA), Bernard Ghanem (KAUST), and Jürgen Schmidhuber (IDSIA - Lugano)</i>	
Leaping Into Memories: Space-Time Deep Feature Synthesis	1966
<i>Alexandros Stergiou (Vrije Universiteit Brussel) and Nikos Deligiannis (Vrije Universiteit Brussel - imec)</i>	
MAGI: Multi-Annotated Explanation-Guided Learning	1977
<i>Yifei Zhang (Emory University), Siyi Gu (Emory University), Yuyang Gao (Emory University), Bo Pan (Emory University), Xiaofeng Yang (Emory University), and Liang Zhao (Emory University)</i>	
SAFARI: Versatile and Efficient Evaluations for Robustness of Interpretability	1988
<i>Wei Huang (Purple Mountain Laboratories), Xingyu Zhao (University of Warwick), Gaojie Jin (University of Liverpool), and Xiaowei Huang (Liverpool University)</i>	
Do DALL-E and Flamingo Understand Each Other?	1999
<i>Hang Li (Ludwig Maximilian University of Munich and Siemens AG), Jindong Gu (University of Oxford), Rajat Koner (Ludwig Maximilian University of Munich), Sahand Sharifzadeh (DeepMind), and Volker Tresp (Siemens AG and Ludwig Maximilian University of Munich)</i>	

Evaluation and Improvement of Interpretability for Self-Explainable Part-Prototype Networks	2011
<i>Qihan Huang (Zhejiang University), Mengqi Xue (Zhejiang University), Wenqi Huang (China southern power grid), Haofei Zhang (Zhejiang University), Jie Song (Zhejiang University), Yongcheng Jing (The University of Sydney), and Mingli Song (Zhejiang University)</i>	
MoreauGrad: Sparse and Robust Interpretation of Neural Networks via Moreau Envelope	2021
<i>Jingwei Zhang (The Chinese University of Hong Kong) and Farzan Farnia (The Chinese University of Hong Kong)</i>	
Towards Understanding the Generalization of Deepfake Detectors from a Game-Theoretical View	2031
<i>Kelu Yao (Institute of Computing Technology, Chinese Academy of Sciences), Jin Wang (Institute of Computing Technology, Chinese Academy of Sciences), Boyu Diao (Institute of Computing Technology, Chinese Academy of Sciences), and Chao Li (zhejianglab)</i>	
Counterfactual-based Saliency Map: Towards Visual Contrastive Explanations for Neural Networks	2042
<i>Xue Wang (Wuhan University), Zhibo Wang (Zhejiang University), Haiqin Weng (Ant Group), Hengchang Guo (Wuhan University), Zhifei Zhang (Adobe Research), Lu Jin (Ant Group Co., Ltd.), Tao Wei (Ant Group), and Kui Ren (Zhejiang University)</i>	
Beyond Single Path Integrated Gradients for Reliable Input Attribution via Randomized Path Sampling	2052
<i>Giyoun Jeon (LG AI Research), Haedong Jeong (UNIST), and Jaesik Choi (KAIST/INEEJI)</i>	
Learning Support and Trivial Prototypes for Interpretable Image Classification	2062
<i>Chong Wang (Australian Institute for Machine Learning, The University of Adelaide), Yuyuan Liu (University of Adelaide), Yuanhong Chen (University of Adelaide), Fengbei Liu (University of Adelaide), Yu Tian (Harvard University), Davis McCarthy (St Vincent's Institute of Medical Research), Helen Frazer (St Vincent's Hospital Melbourne), and Gustavo Carneiro (University of Surrey)</i>	
Visual Explanations via Iterated Integrated Attributions	2073
<i>Oren Barkan (The Open University), Yehonatan Elisha (Open University Of Israel), Yuval Asher (Tel Aviv University), Amit Eshel (Tel-Aviv University), and Noam Koenigstein (Tel Aviv University)</i>	

Neural Generative Models

Unsupervised Compositional Concepts Discovery with Text-to-Image Generative Models	2085
<i>Nan Liu (University of Illinois at Urbana-Champaign), Yilun Du (MIT), Shuang Li (MIT), Joshua B. Tenenbaum (MIT), and Antonio Torralba (MIT)</i>	
Human Preference Score: Better Aligning Text-to-Image Models with Human Preference	2096
<i>Xiaoshi Wu (Chinese University of Hong Kong), Keqiang Sun (Chinese University of Hong Kong), Feng Zhu (University of Science and Technology of China), Rui Zhao (SenseTime Group Limited), and Hongsheng Li (The Chinese University of Hong Kong)</i>	

DLT: Conditioned layout generation with Joint Discrete-Continuous Diffusion Layout Transformer	2106
<i>Elad Levi (Sightful), Eli Brosh (Wix), Mykola Mykhailych (UCU), and Meir Perez (Wix)</i>	
Anti-DreamBooth: Protecting Users from Personalized Text-to-image Synthesis	2116
<i>Thanh Van Le (VinAI), Hao Phung (VinAI), Thuan Hoang Nguyen (VinAI), Quan Dao (VinAI), Ngoc N. Tran (VinAI Research), and Anh Tran (VinAI)</i>	
GECCO: Geometrically-Conditioned Point Diffusion Models	2128
<i>Michał J Tyszkiewicz (EPFL), Pascal Fua (EPFL, Switzerland), and Eduard Trulls (Google)</i>	
DiffDreamer: Towards Consistent Unsupervised Single-view Scene Extrapolation with Conditional Diffusion Models	2139
<i>Shengqu Cai (ETH Zurich), Eric Ryan Chan (Stanford University), Songyou Peng (ETH Zurich and MPI-IS), Mohamad Shahbazi (ETH Zurich), Anton Obukhov (ETH Zurich), Luc Van Gool (ETH Zurich), and Gordon Wetzstein (Stanford University)</i>	
Guided Motion Diffusion for Controllable Human Motion Synthesis	2151
<i>Korrawe Karunratanakul (ETH Zurich), Konpat Preechakul (Vidyasirimedhi Institute of Science and Technology), Supasorn Suwajanakorn (Vistec), and Siyu Tang (ETH Zurich)</i>	
COOP: Decoupling and Coupling of Whole-Body Grasping Pose Generation	2163
<i>Yanzhao Zheng (Alibaba DAMO Academy), Yunzhou Shi (Alibaba), Yuhao Cui (Alibaba DAMO Academy), Zhongzhou Zhao (Alibaba Group), Zhiling Luo (Alibaba Group), and Wei Zhou (Alibaba)</i>	
Zero-Shot Spatial Layout Conditioning for Text-to-Image Diffusion Models	2174
<i>Guillaume Couairon (Facebook AI Research), Marlène Careil (Facebook/Télécom Paris), Matthieu Cord (Sorbonne University), Stéphane Lathuilière (Telecom-Paris), and Jakob Verbeek (Meta)</i>	
StyleDomain: Efficient and Lightweight Parameterizations of StyleGAN for One-shot and Few-shot Domain Adaptation	2184
<i>Aibek Alanov (Artificial Intelligence Research Institute), Vadim Titov (AIRI), Maksim Nakhodnov (MSU), and Dmitry Vetrov (Higher School of Economics)</i>	
GRAM-HD: 3D-Consistent Image Generation at High Resolution with Generative Radiance Manifolds	2195
<i>Jianfeng Xiang (Microsoft Research Asia), Jiaolong Yang (Microsoft Research), Yu Deng (Xiaobing.AI), and Xin Tong (Microsoft)</i>	
Your Diffusion Model is Secretly a Zero-Shot Classifier	2206
<i>Alexander C. Li (Carnegie Mellon University), Mihir Prabhudesai (Carnegie Mellon University), Shivam Duggal (Carnegie Mellon University), Ellis Brown (Carnegie Mellon University), and Deepak Pathak (Carnegie Mellon University)</i>	
Learning Hierarchical Features with Joint Latent Space Energy-Based Prior	2218
<i>Jiali Cui (Stevens Institute of Technology), Ying Nian Wu (University of California, Los Angeles), and Tian Han (Stevens Institute of Technology)</i>	

ActFormer: A GAN-based Transformer towards General Action-Conditioned 3D Human Motion Generation	2228
<i>Liang Xu (Shanghai Jiao Tong University), Ziyang Song (The Hong Kong Polytechnic University), Dongliang Wang (SenseTime Group Limited), Jing Su (Fudan University), Zhicheng Fang (Baidu), Chenjing Ding (Sensetime Group Limited), Weihao Gan (SenseTime), Yichao Yan (AI Institute of Shanghai Jiao Tong University), Xin Jin (Eastern Institute for Advanced Study), Xiaokang Yang (Shanghai Jiao Tong University of China), Wenjun Zeng (EIT Institute for Advanced Study), and Wei Wu (SenseTime Group Limited)</i>	
Landscape Learning for Neural Network Inversion	2239
<i>Ruoshi Liu (Columbia University), Chengzhi Mao (Columbia University), Purva Tendulkar (Columbia University), Hao Wang (Rutgers University), and Carl Vondrick (Columbia University)</i>	
Diffusion in Style	2251
<i>Martin Nicolas Everaert (EPFL), Marco Bocchio (Largo.ai), Sami Arpa (Largo.ai), Sabine Süssstrunk (EPFL), and Radhakrishna Achanta (EPFL)</i>	
Diffusion-SDF: Conditional Generative Modeling of Signed Distance Functions	2262
<i>Gene Chou (Princeton University), Yuval Bahat (Princeton University), and Felix Heide (Princeton / Algolux)</i>	
GETAvatar: Generative Textured Meshes for Animatable Human Avatars	2273
<i>Xuanmeng Zhang (University of Technology Sydney), Jianfeng Zhang (NUS), Rohan Chacko (Carnegie Mellon University), Hongyi Xu (Bytedance), Guoxian Song (Bytedance), Yi Yang (Zhejiang University), and Jiashi Feng (ByteDance)</i>	
A-STAR: Test-time Attention Segregation and Retention for Text-to-image Synthesis	2283
<i>Aishwarya Agarwal (Adobe Research), Srikrishna Karanam (Adobe Research), K J Joseph (Adobe Research), Apoorv Saxena (Indian Institute of Science), Koustava Goswami (Insight Centre for Data Analytics, Data Science Institute, National University of Ireland Galway), and Balaji Vasani Srinivasan (Adobe Research)</i>	
TF-ICON: Diffusion-Based Training-Free Cross-Domain Image Composition	2294
<i>Shilin Lu (Nanyang Technological University), Yanzhu Liu (Centre for Frontier AI Research & Institute for Infocomm Research, Agency for Science, Technology and Research, 138632, Singapore), and Adams Wai-Kin Kong (Nanyang Technological University)</i>	
Breaking The Limits of Text-conditioned 3D Motion Synthesis with Elaborative Descriptions	2306
<i>Yijun Qian (Carnegie Mellon University), Jack Urbanek (Meta), Alexander G. Hauptmann (Carnegie Mellon University), and Jungdam Won (Meta AI)</i>	
BeLFusion: Latent Diffusion for Behavior-Driven Human Motion Prediction	2317
<i>German Barquero (Universitat de Barcelona), Sergio Escalera (Computer Vision Center, UAB), and Cristina Palmero (Universitat de Barcelona)</i>	
Delta Denoising Score	2328
<i>Amir Hertz (Tel Aviv University), Kfir Aberman (Google), and Daniel Cohen-Or (Tel Aviv University)</i>	

Mimic3D: Thriving 3D-Aware GANs via 3D-to-2D Imitation	2338
<i>Xingyu Chen (Xiaobing.AI), Yu Deng (Xiaobing.AI), and Baoyuan Wang (Xiaobing.AI)</i>	
DreamBooth3D: Subject-Driven Text-to-3D Generation	2349
<i>Amit Raj (Google), Srinivas Kaza (Google Inc), Ben Poole (Google Brain), Michael Niemeyer (Google), Nataniel Ruiz (Boston University), Ben Mildenhall (Google Research), Shiran Zada (Weizmann Institute of Science), Kfir Aberman (Google), Michael Rubinstein (Google), Jonathan Barron (Google Research), Yuanzhen Li (Google), and Varun Jampani (Google)</i>	
Feature Proliferation — the “Cancer” in StyleGAN and its Treatments	2360
<i>Shuang Song (Cardiff University), Yuanbang Liang (Cardiff University), Jing Wu (Cardiff University), Yu-Kun Lai (Cardiff University), and Yipeng Qin (Cardiff University)</i>	
Unsupervised Facial Performance Editing via Vector-Quantized StyleGAN Representations	2371
<i>Berkay Kicanaoglu (University of Amsterdam), Pablo Garrido (Flawless AI), and Gaurav Bharaj (Multiple)</i>	
3D-aware Image Generation using 2D Diffusion Models	2383
<i>Jianfeng Xiang (Microsoft Research Asia), Jiaolong Yang (Microsoft Research), Binbin Huang (ShanghaiTech University), and Xin Tong (Microsoft)</i>	
Neural Collage Transfer: Artistic Reconstruction via Material Manipulation	2394
<i>Ganghun Lee (Seoul National University), Minji Kim (Seoul National University), Yunsu Lee (Seoul National University), Minsu Lee (Seoul National University), and Byoung-Tak Zhang (Seoul National University)</i>	
Phasic Content Fusing Diffusion Model with Directional Distribution Consistency for Few-Shot Model Adaption	2406
<i>Teng Hu (Shanghai Jiao Tong University), Jiangning Zhang (Zhejiang University), Liang Liu (Tencent), Ran Yi (Shanghai Jiao Tong University), Siqi Kou (Shanghai Jiao Tong University), Haokun Zhu (Shanghai Jiaotong University), Xu Chen (Tencent), Yabiao Wang (Tencent), Chengjie Wang (Tencent, Shanghai Jiao Tong University), and Lizhuang Ma (Shanghai Jiao Tong University)</i>	
Single-Stage Diffusion NeRF: A Unified Approach to 3D Generation and Reconstruction	2416
<i>Hansheng Chen (Tongji University), Jiatao Gu (Apple), Anpei Chen (ETH Zürich), Wei Tian (Tongji University), Zhuowen Tu (UC San Diego), Lingjie Liu (University of Pennsylvania), and Hao Su (UCSD)</i>	
Erasing Concepts from Diffusion Models	2426
<i>Rohit Gandikota (Northeastern University), Joanna Materzynska (MIT), Jaden Fiotto-Kaufman (Northeastern University), and David Bau (Northeastern University)</i>	
Make Encoder Great Again in 3D GAN Inversion through Geometry and Occlusion-Aware Encoding..	2437
<i>Ziyang Yuan (Graduate school at ShenZhenTsinghua university), Yiming Zhu (Graduate school at ShenZhen, Tsinghua university), Yu Li (International Digital Economy Academy), Hongyu Liu (HKUST), and Chun Yuan (Graduate school at ShenZhen, Tsinghua university)</i>	

HairNeRF: Geometry-Aware Image Synthesis for Hairstyle Transfer	2448
<i>Seunggyu Chang (NAVER Cloud), Gihoon Kim (KAIST), and Hayeon Kim (UNIST)</i>	

Vision and Language 1

SMAUG: Sparse Masked Autoencoder for Efficient Video-Language Pre-Training	2459
<i>Yuanze Lin (University of Washington), Chen Wei (Johns Hopkins University), Huiyu Wang (Meta), Alan Yuille (Johns Hopkins University), and Cihang Xie (University of California, Santa Cruz)</i>	
DiffusionRet: Generative Text-Video Retrieval with Diffusion Model	2470
<i>Peng Jin (Shenzhen Graduate School, Peking University), Hao Li (Peking University), Zesen Cheng (School of Electronic and Computer Engineering, Peking University), Kehan Li (Peking University), Xiangyang Ji (Tsinghua University), Chang Liu (Tsinghua University), Li Yuan (Peking University), and Jie Chen (Peking University)</i>	
Explore and Tell: Embodied Visual Captioning in 3D Environments	2482
<i>Anwen Hu (Alibaba DAMO Academy), Shizhe Chen (INRIA), Liang Zhang (Renmin University of China), and Qin Jin (Renmin University of China)</i>	
Distilling Large Vision-Language Model with Out-of-Distribution Generalizability	2492
<i>Xuanlin Li (UCSD), Yunhao Fang (University of California, San Diego), Minghua Liu (UCSD), Zhan Ling (UC San Diego), Zhuowen Tu (UC San Diego), and Hao Su (UCSD)</i>	
Learning Trajectory-Word Alignments for Video-Language Tasks	2504
<i>Xu Yang (Southeast University), Zhangzikang Li (Southeast university), Haiyang Xu (Alibaba DAMO Academy), Hanwang Zhang (Nanyang Technological University), Qinghao Ye (Alibaba Group), Chenliang Li (Alibaba Group), Ming Yan (Alibaba Group), Yu Zhang (Southeast University), Fei Huang (Alibaba), and Songfang Huang (Alibaba Group)</i>	
Variational Causal Inference Network for Explanatory Visual Question Answering	2515
<i>Dizhan Xue (Institute of Automation, Chinese Academy of Sciences), Shengsheng Qian (institute of automation chinese academy of sciences), and Changsheng Xu (CASIA)</i>	
TextManiA: Enriching Visual Feature by Text-driven Manifold Augmentation	2526
<i>Moon Ye-Bin (POSTECH), Jisoo Kim (Ewha womens university), Hongyeob Kim (Nuvilab), Kilho Son (SRI international), and Tae-Hyun Oh (POSTECH)</i>	
Segment Every Reference Object in Spatial and Temporal Spaces	2538
<i>Jiannan Wu (The University of Hong Kong), Yi Jiang (Bytedance), Bin Yan (Dalian University of Technology), Huchuan Lu (Dalian University of Technology), Zehuan Yuan (Bytedance.Inc), and Ping Luo (The University of Hong Kong)</i>	

Gradient-Regulated Meta-Prompt Learning for Generalizable Vision-Language Models	2551
<i>Juncheng Li (Zhejiang University), Minghe Gao (Zhejiang University), Longhui Wei (University of Science and Technology of China), Siliang Tang (Zhejiang University), Wenqiao Zhang (NUS), Mengze Li (Zhejiang University), Wei Ji (National University of Singapore), Qi Tian (Huawei Cloud BU), Tat-Seng Chua (National university of Singapore), and Yueting Zhuang (Zhejiang University)</i>	
Misalign, Contrast then Distill: Rethinking Misalignments in Language-Image Pre-training	2563
<i>Bumsoo Kim (LG AI Research), Yeonsik Jo (LG AI Research), Jinhyung Kim (LG AI Research), and Seunghwan Kim (LG AI Research)</i>	
Toward Multi-Granularity Decision-Making: Explicit Visual Reasoning with Hierarchical Knowledge	2573
<i>Yifeng Zhang (University of Minnesota, Twin Cities), Shi Chen (University of Minnesota), and Qi Zhao (University of Minnesota)</i>	
VL-Match: Enhancing Vision-Language Pretraining with Token-Level and Instance-Level Matching	2584
<i>Junyu Bi (Institute of Computing Technology, Chinese Academy of Sciences), Daixuan Cheng (Microsoft), Ping Yao (Institute of Computing Technology, Chinese Academy of Sciences), Bochen Pang (Microsoft), Yuefeng Zhan (Microsoft), Chuanguang Yang (Institute of Computing Technology, Chinese Academy of Sciences), Yujing Wang (MSRA), Hao Sun (Microsoft), Weiwei Deng (Microsoft), and Qi Zhang (Microsoft)</i>	
Moment Detection in Long Tutorial Videos	2594
<i>Ioana Croitoru (Institute of Mathematics of the Romanian Academy), Simion-Vlad Bogolin (Institute of Mathematics of the Romanian Academy), Samuel Albanie (University of Cambridge), Yang Liu (Peking University), Zhaowen Wang (Adobe Research), Seunghyun Yoon (Adobe Research), Franck Dernoncourt (Adobe Research), Hailin Jin (Adobe Research), and Trung Bui (Adobe Research)</i>	
Not All Features Matter: Enhancing Few-shot CLIP with Adaptive Prior Refinement	2605
<i>Xiangyang Zhu (City University of Hong Kong), Renrui Zhang (Shanghai AI Lab), Bowei He (City University of Hong Kong), Aojun Zhou (The Chinese University of Hong Kong), Dong Wang (Shanghai AI Laboratory), Bin Zhao (Northwestern Polytechnical University), and Peng Gao (Chinese university of hong kong)</i>	
Breaking Common Sense: WHOOPS! A Vision-and-Language Benchmark of Synthetic and Compositional Images	2616
<i>Nitzan Bitton-Guetta (Ben-Gurion University of the Negev), Yonatan Bitton (The Hebrew University of Jerusalem), Jack Hessel (Allen Institute for AI), Ludwig Schmidt (University of Washington), Yuval Elovici (Ben-Gurion University of the Negev), Gabriel Stanovsky (Hebrew University), and Roy Schwartz (Hebrew University of Jerusalem)</i>	
Advancing Referring Expression Segmentation Beyond Single Image	2628
<i>Yixuan Wu (Zhejiang University), Zhao Zhang (SenseTime Research), Chi Xie (Tongji University), Feng Zhu (University of Science and Technology of China), and Rui Zhao (SenseTime Group Limited)</i>	

PointCLIP V2: Prompting CLIP and GPT for Powerful 3D Open-world Learning	2639
<i>Xiangyang Zhu (City University of Hong Kong), Renrui Zhang (Shanghai AI Lab), Bowei He (City University of Hong Kong), Ziyu Guo (Peking University), Ziyao Zeng (Yale University), Zipeng Qin (The Chinese University of Hong Kong), Shanghang Zhang (Peking University), and Peng Gao (Chinese university of hong kong)</i>	
Unsupervised Prompt Tuning for Text-Driven Object Detection	2651
<i>Weizhen He (Zhejiang University), Weijie Chen (Zhejiang University), Binbin Chen (Hikvision Research Institute), Shicai Yang (Hikvision Research Institute), Di Xie (Hikvision Research Institute), LuoJun Lin (Fuzhou University), Donglian Qi (Zhejiang University), and Yueting Zhuang (Zhejiang University)</i>	
Distilling Coarse-to-Fine Semantic Matching Knowledge for Weakly Supervised 3D Visual Grounding	2662
<i>Zehan Wang (zhejiang university), Haifeng Huang (Zhejiang University), Yang Zhao (Zhejiang University), Linjun Li (Zhejiang University), Xize Cheng (Zhejiang University), Yichen Zhu (ZheJiang University), Aoxiong Yin (Zhejiang University), and Zhou Zhao (Zhejiang University)</i>	
I Can't Believe There's No Images! Learning Visual Tasks Using only Language Supervision	2672
<i>Sophia Gu (Allen Institute for AI), Christopher Clark (Allen Institute for AI), and Aniruddha Kembhavi (Allen Institute for Artificial Intelligence)</i>	
Learning Cross-Modal Affinity for Referring Video Object Segmentation Targeting Limited Samples	2684
<i>Guanghui Li (Anhui University of Technology), Mingqi Gao (University of Warwick, SUSTech), Heng Liu (Anhui University of Technology), Xiantong Zhen (United Imaging Healthcare), and Feng Zheng (SUSTech)</i>	
MeViS: A Large-scale Benchmark for Video Segmentation with Motion Expressions	2694
<i>Henghui Ding (Nanyang Technological University), Chang Liu (Nanyang Technological University), Shuting He (Zhejiang University), Xudong Jiang (Nanyang Technological University), and Chen Change Loy (Nanyang Technological University)</i>	
Diverse Data Augmentation with Diffusions for Effective Test-time Prompt Tuning	2704
<i>Chun-Mei Feng (Institute of High Performance Computing, A*STAR), Kai Yu (Agency for Science, Technology and Research, (A*STAR), Yong Liu (IHPC), Salman Khan (MBZUAI), and Wangmeng Zuo (Harbin Institute of Technology, China)</i>	
ShapeScaffolder: Structure-Aware 3D Shape Generation from Text	2715
<i>Xi Tian (University of Bath), Yong-Liang Yang (University of Bath), and Qi Wu (University of Adelaide)</i>	
SuS-X: Training-Free Name-Only Transfer of Vision-Language Models	2725
<i>Vishaal Udandarao (University of Cambridge), Ankush Gupta (DeepMind), and Samuel Albanie (University of Cambridge)</i>	

X-Mesh: Towards Fast and Accurate Text-driven 3D Stylization via Dynamic Textual Guidance ..	2737
<i>Yiwei Ma (Xiamen University), Xiaoqing Zhang (Xiamen University), Xiaoshuai Sun (Xiamen University), Jiayi Ji (Xiamen University), Haowei Wang (Xiamen University), Guannan Jiang (CATL), Weilin Zhuang (Contemporary Amperex Technology Co., Limited), and Rongrong Ji (Xiamen University, China)</i>	
OnlineRefer: A Simple Online Baseline for Referring Video Object Segmentation	2749
<i>Dongming Wu (Beijing Institute of Technology), Tiancai Wang (Megvii Technology), Yuang Zhang (Shanghai Jiao Tong University), Xiangyu Zhang (Megvii Technology), and Jianbing Shen (University of Macau)</i>	
Attentive Mask CLIP	2759
<i>Yifan Yang (Microsoft), Weiquan Huang (Tongji University), Yixuan Wei (Tsinghua University), Houwen Peng (Microsoft Research), Xinyang Jiang (Microsoft Research Asia), Huiqiang Jiang (Microsoft Research Asia), Fangyun Wei (Microsoft Research Asia), Yin Wang (Tongji University), Han Hu (Microsoft Research Asia), Lili Qiu (The University of Texas at Austin), and Yuqing Yang (Microsoft Research)</i>	
Knowledge Proxy Intervention for Deconfounded Video Question Answering	2770
<i>Jiangtong Li (Shanghai Jiao Tong University), Li Niu (Shanghai Jiao Tong University), and Liqing Zhang (Shanghai Jiao Tong University)</i>	
UniVTG: Towards Unified Video-Language Temporal Grounding	2782
<i>Kevin Qinghong Lin (National University of Singapore), Pengchuan Zhang (Microsoft Research AI), Joya Chen (National University of Singapore), Shraman Pramanick (Johns Hopkins University), Difei Gao (NUS), Alex Jinpeng Wang (National University of Singapore), Rui Yan (Nanjing University), and Mike Zheng Shou (National University of Singapore)</i>	
Self-supervised Cross-view Representation Reconstruction for Change Captioning	2793
<i>Yunbin Tu (University of Chinese Academy of Sciences), Liang Li (Institute of Computing Technology, Chinese Academy of Sciences), Li Su (University of Chinese Academy of Sciences), Zheng-Jun Zha (University of Science and Technology of China), Chenggang Yan (Hangzhou Dianzi University), and Qingming Huang (University of Chinese Academy of Sciences)</i>	
Unified Coarse-to-Fine Alignment for Video-Text Retrieval	2804
<i>Ziyang Wang (UNC Chapel Hill), Yi-Lin Sung (UNC Chapel Hill), Feng Cheng (University of North Carolina at Chapel Hill), Gedas Bertasius (UNC Chapel Hill), and Mohit Bansal (University of North Carolina at Chapel Hill)</i>	
Confidence-aware Pseudo-label Learning for Weakly Supervised Visual Grounding	2816
<i>Yang Liu (Peking University), Jiahua Zhang (Peking University), Qingchao Chen (Peking University), and Yuxin Peng (Peking University)</i>	
TextPSG: Panoptic Scene Graph Generation from Textual Descriptions	2827
<i>Chengyang Zhao (Peking University), Yikang Shen (MIT-IBM Watson AI Lab), Zhenfang Chen (MIT-IBM Watson AI Lab), Mingyu Ding (UC Berkeley), and Chuang Gan (MIT-IBM Watson AI Lab)</i>	

MAth, eXpand and Improve: Unsupervised Finetuning for Zero-Shot Action Recognition with Language Knowledge	2839
<i>Wei Lin (Graz University of Technology), Leonid Karlinsky (IBM-Research), Nina Shvetsova (Goethe University Frankfurt), Horst Possegger (Graz University of Technology), Mateusz Kozinski (ICG TUGRAZ), Rameswar Panda (MIT-IBM Watson AI Lab), Rogerio Feris (MIT-IBM Watson AI Lab, IBM Research), Hilde Kuehne (University of Bonn), and Horst Bischof (Graz University of Technology)</i>	
Unify, Align and Refine: Multi-Level Semantic Alignment for Radiology Report Generation	2851
<i>Yaowei Li (Peking University), Bang Yang (Peking University), Xuxin Cheng (Peking University), Zhihong Zhu (Peking University), Hongxiang Li (Peking University), and Yuexian Zou (Peking University)</i>	
CLIPTrans: Transferring Visual Knowledge with Pre-trained Models for Multimodal Machine Translation	2863
<i>Devaansh Gupta (BITS Pilani), Siddhant Kharbanda (Microsoft), Jiawei Zhou (Harvard University), Wanhua Li (Harvard University), Hanspeter Pfister (Harvard University), and Donglai Wei (Boston College)</i>	
Learning Human-Human Interactions in Images from Weak Textual Supervision	2875
<i>Morris Alper (Tel Aviv University) and Hadar Averbuch-Elor (Tel Aviv University)</i>	
BUS: Efficient and Effective Vision-Language Pre-Training with Bottom-Up Patch Summarization.	2888
<i>Chaoya Jiang (Peking University), Haiyang Xu (Alibaba DAMO Academy), Wei Ye (Peking University), Qinghao Ye (Alibaba Group), Chenliang Li (Alibaba Group), Ming Yan (Alibaba Group), Bin Bi (Alibaba Group), Shikun Zhang (Peking University), Fei Huang (Alibaba), and Songfang Huang (Alibaba Group)</i>	
3D-VisTA: Pre-trained Transformer for 3D Vision and Text Alignment	2899
<i>Ziyu Zhu (Tsinghua University), Xiaojian Ma (University of California, Los Angeles), Yixin Chen (UCLA), Zhidong Deng (Tsinghua University), Siyuan Huang (Beijing Institute for General Artificial Intelligence), and Qing Li (BIGAI)</i>	
ALIP: Adaptive Language-Image Pre-Training with Synthetic Caption	2910
<i>Kaicheng Yang (DeepGlint), Jiankang Deng (Imperial College London), Xiang An (DeepGlint), Jiawei Li (deepglint), Ziyong Feng (DeepGlint), Jia Guo (InsightFace.ai), Jing Yang (University of Cambridge), and Tongliang Liu (The University of Sydney)</i>	
LoGoPrompt: Synthetic Text Images Can Be Good Visual Prompts for Vision-Language Models ..	2920
<i>Cheng Shi (ShanghaiTech University) and Sibe Yang (ShanghaiTech University)</i>	
Noise-Aware Learning from Web-Crawled Image-Text Data for Image Captioning	2930
<i>Wooyoung Kang (Kakao Brain), Jonghwan Mun (Kakao Brain), Sungjun Lee (kakaobrain), and Byungseok Roh (KakaoBrain)</i>	
Decouple Before Interact: Multi-Modal Prompt Learning for Continual Visual Question Answering	2941
<i>Zi Qian (Tsinghua University), Xin Wang (Tsinghua University), Xuguang Duan (Tsinghua University), Pengda Qin (Alibaba), Yuhong Li (Alibaba), and Wenwu Zhu (Tsinghua University)</i>	

PromptCap: Prompt-Guided Image Captioning for VQA with GPT-3	2951
<i>Yushi Hu (University of Washington), Hang Hua (University of Rochester), Zhengyuan Yang (Microsoft), Weijia Shi (UCLA), Noah A. Smith (University of Washington and Allen Institute for AI), and Jiebo Luo (U. Rochester)</i>	
Grounded Image Text Matching with Mismatched Relation Reasoning	2964
<i>Yu Wu (ShanghaiTech University), Yana Wei (ShanghaiTech University), Haozhe Wang (ShanghaiTech University), Yongfei Liu (ShanghaiTech), Sibe Yang (ShanghaiTech University), and Xuming He (ShanghaiTech University)</i>	
GePSAn: Generative Procedure Step Anticipation in Cooking Videos	2976
<i>Mohamed A. Abdelsalam (Samsung AI Center - Toronto), Samrudhdi B. Rangrej (Samsung AI Center - Toronto), Isma Hadji (Samsung AI Center - Toronto), Nikita Dvornik (Waabi), Konstantinos G. Derpanis (York University), and Afsaneh Fazly (SAIC Toronto)</i>	
LLM-Planner: Few-Shot Grounded Planning for Embodied Agents with Large Language Models	2986
<i>Chan Hee Song (The Ohio State University), Jiaman Wu (The Ohio State University), Clayton Washington (The Ohio State University), Brian M. Sadler (Army Research Laboratory, USA), Wei-Lun Chao (The Ohio State University), and Yu Su (OSU)</i>	
VL-PET: Vision-and-Language Parameter-Efficient Tuning via Granularity Control	2998
<i>Zi-Yuan Hu (The Chinese University of Hong Kong), Yanyang Li (The Chinese University of Hong Kong), Michael R. Lyu (The Chinese University of Hong Kong), and Liwei Wang (CUHK)</i>	
With a Little Help from Your Own Past: Prototypical Memory Networks for Image Captioning ...	3009
<i>Manuele Barraco (University of Modena and Reggio Emilia), Sara Sarto (University of Modena and Reggio Emilia), Marcella Cornia (University of Modena and Reggio Emilia), Lorenzo Baraldi (University of Modena and Reggio Emilia), and Rita Cucchiara (Università di Modena e Reggio Emilia)</i>	
DALL-Eval: Probing the Reasoning Skills and Social Biases of Text-to-Image Generation Models	3020
<i>Jaemin Cho (UNC Chapel Hill), Abhay Zala (UNC at Chapel Hill), and Mohit Bansal (University of North Carolina at Chapel Hill)</i>	
Learning Navigational Visual Representations with Semantic Map Supervision	3032
<i>Yicong Hong (Australian National University), Yang Zhou (Adobe Research), Ruiyi Zhang (Adobe Research), Franck Deroncourt (Adobe Research), Trung Bui (Adobe Research), Stephen Gould (Australian National University, Australia), and Hao Tan (UNC, Chapel Hill)</i>	
CoTDet: Affordance Knowledge Prompting for Task Driven Object Detection	3045
<i>Jiajin Tang (ShanghaiTech University), Ge Zheng (ShanghaiTech University), Jingyi Yu (Shanghai Tech University), and Sibe Yang (ShanghaiTech University)</i>	
Open Set Video HOI detection from Action-Centric Chain-of-Look Prompting	3056
<i>Nan Xi (State University of New York at Buffalo), Jingjing Meng (Nanyang Technological University), and Junsong Yuan (State University of New York at Buffalo, USA)</i>	

Learning Concise and Descriptive Attributes for Visual Recognition	3067
<i>An Yan (UC San Diego), Yu Wang (University of California, San Diego), Yiwu Zhong (University of Wisconsin-Madison), Chengyu Dong (University of California, San Diego), Zexue He (University of California, San Diego), Yujie Lu (University of California, Santa Barbara), William Yang Wang (UC Santa Barbara), Jingbo Shang (University of California, San Diego), and Julian McAuley (UCSD)</i>	
Open-vocabulary Video Question Answering: A New Benchmark for Evaluating the Generalizability of Video Question Answering Models	3078
<i>Dohwan Ko (Korea University), Ji Soo Lee (Korea University), Miso Choi (Korea University), Jaewon Chu (Korea University), Jihwan Park (Korea university), and Hyunwoo J. Kim (Korea University)</i>	
Encyclopedic VQA: Visual Questions About Detailed Properties of Fine-Grained Categories	3090
<i>Thomas Mensink (Google Research), Jasper Uijlings (Google Research), Lluís Castrejon (Google Research), Arushi Goel (University of Edinburgh), Felipe Cador (Universidade Federal de Minas Gerais), Howard Zhou (Google), Fei Sha (Google Research), André Araujo (Google), and Vittorio Ferrari (Google Research)</i>	
Story Visualization by Online Text Augmentation with Context Memory	3102
<i>Daechul Ahn (Yonsei university), Daneul Kim (GIST), Gwangmo Song (LG AI Research), Seung Hwan Kim (LG AI Research), Honglak Lee (LG AI Research / University of Michigan), Dongyeop Kang (University of Minnesota), and Jonghyun Choi (Yonsei University)</i>	
Transferable Decoding with Visual Entities for Zero-Shot Image Captioning	3113
<i>Junjie Fei (Xiamen University), Teng Wang (The University of Hong Kong), Jinrui Zhang (Southern University of Science and Technology), Zhenyu He (Harbin Institute of Technology, (Shenzhen), Chengjie Wang (Tencent; Shanghai Jiao Tong University), and Feng Zheng (SUSTech)</i>	
Too Large; Data Reduction for Vision-Language Pre-Training	3124
<i>Alex Jinpeng Wang (National University of Singapore), Kevin Qinghong Lin (National University of Singapore), David Junhao Zhang (National University of Singapore), Stan Weixian Lei (National University of Singapore), and Mike Zheng Shou (National University of Singapore)</i>	
ViLTA: Enhancing Vision-Language Pre-training through Textual Augmentation	3135
<i>Weihan Wang (Tsinghua University), Zhen Yang (Tsinghua University), Bin Xu (Tsinghua University), Juanzi Li (Tsinghua University), and Yankui Sun (Tsinghua University)</i>	
Teaching CLIP to Count to Ten	3147
<i>Roni Paiss (Tel Aviv University, Google), Ariel Ephrat (Google), Omer Tov (Google), Shiran Zada (Weizmann Institute of Science), Inbar Mosseri (Google), Michal Irani (Weizmann Institute, Israel), and Tali Dekel (Weizmann Institute of Science)</i>	

3D from Multi-view and Sensors 1

Learning a More Continuous Zero Level Set in Unsigned Distance Fields through Level Set Projection	3158
<i>Junsheng Zhou (Tsinghua University), Baorui Ma (Tsinghua university), Shujuan Li (Tsinghua University), Yu-Shen Liu (Tsinghua University), and Zhizhong Han (Wayne State University)</i>	
Enhancing NeRF akin to Enhancing LLMs: Generalizable NeRF Transformer with Mixture-of-View-Experts	3170
<i>Wenyan Cong (University of Texas at Austin), Hanxue Liang (University of Texas at Austin), Peihao Wang (University of Texas at Austin), Zhiwen Fan (University of Texas at Austin), Tianlong Chen (University of Texas at Austin), Mukund Varma (Indian Institute of Technology, Madras, India), Yi Wang (University of Texas at Austin), and Zhangyang Wang (University of Texas at Austin)</i>	
MatrixCity: A Large-scale City Dataset for City-scale Neural Rendering and Beyond	3182
<i>Yixuan Li (The Chinese University of Hong Kong), Lihan Jiang (Wuhan University), Linning Xu (CUHK), Yuanbo Xiangli (the Chinese University of Hong Kong), Zhenzhi Wang (Chinese University of Hong Kong), Dahua Lin (The Chinese University of Hong Kong), and Bo Dai (Shanghai AI Lab)</i>	
R3D3: Dense 3D Reconstruction of Dynamic Scenes from Multiple Cameras	3193
<i>Aron Schmied (ETH Zurich), Tobias Fischer (ETH Zurich), Martin Danelljan (ETH Zurich), Marc Pollefeys (ETH Zurich / Microsoft), and Fisher Yu (ETH Zurich)</i>	
ClimateNeRF: Extreme Weather Synthesis in Neural Radiance Field	3204
<i>Yuan Li (Zhejiang University), Zhi-Hao Lin (University of Illinois at Urbana-Champaign), David Forsyth (Univeristy of Illinois at Urbana-Champaign), Jia-Bin Huang (University of Maryland College Park), and Shenlong Wang (UIUC)</i>	
Rendering Humans from Object-Occluded Monocular Videos	3216
<i>Tiange Xiang (Stanford University), Adam Sun (Stanford University), Jiajun Wu (Stanford University), Ehsan Adeli (Stanford University), and Li Fei-Fei (Stanford University)</i>	
AssetField: Assets Mining and Reconfiguration in Ground Feature Plane Representation	3228
<i>Yuanbo Xiangli (the Chinese University of Hong Kong), Linning Xu (CUHK), Xingang Pan (Max Planck Institute for Informatics), Nanxuan Zhao (Adobe Research), Bo Dai (Shanghai AI Lab), and Dahua Lin (The Chinese University of Hong Kong)</i>	
PETRV2: A Unified Framework for 3D Perception from Multi-Camera Images	3239
<i>Yingfei Liu (Megvii Technology), Junjie Yan (Fudan University), Fan Jia (Tianjin University), Shuailin Li (Megvii Technology), Aqi Gao (Tianjin University), Tiancai Wang (Megvii Technology), and Xiangyu Zhang (Megvii Technology)</i>	
MIMO-NeRF: Fast Neural Rendering with Multi-input Multi-output Neural Radiance Fields	3250
<i>Takuhiro Kaneko (NTT Corporation)</i>	
Adaptive Positional Encoding for Bundle-Adjusting Neural Radiance Fields	3261
<i>Zelin Gao (Zhejiang University), Weichen Dai (Hangzhou Dianzi University), and Yu Zhang (ZheJiang University)</i>	

NeuS2: Fast Learning of Neural Implicit Surfaces for Multi-view Reconstruction	3272
<i>Yiming Wang (Peking University), Qin Han (Peking University), Marc Habermann (Max Planck Institute for Informatics), Kostas Daniilidis (University of Pennsylvania), Christian Theobalt (MPI Informatik), and Lingjie Liu (University of Pennsylvania)</i>	
Learning from Semantic Alignment between Unpaired Multiviews for Egocentric Video Recognition	3284
<i>Qitong Wang (University of Delaware), Long Zhao (Google Research), Liangzhe Yuan (Google Research), Ting Liu (Google Research), and Xi Peng (University of Delaware)</i>	
Uncertainty Guided Adaptive Warping for Robust and Efficient Stereo Matching	3295
<i>Junpeng Jing (Imperial College London), Jiankun Li (megvii), Pengfei Xiong (Shopee), Jiangyu Liu (megvii inc), Shuaicheng Liu (UESTC; Megvii), Yichen Guo (BUAA), Xin Deng (Beihang university), Mai Xu (BUAA), Lai Jiang (BUAA), and Leonid Sigal (University of British Columbia)</i>	
Compatibility of Fundamental Matrices for Complete Viewing Graphs	3305
<i>Martin Bråtelund (University of Oslo) and Felix Rydell (KTH)</i>	
ProtoTransfer: Cross-Modal Prototype Transfer for Point Cloud Segmentation	3314
<i>Pin Tang (Shanghai Jiao Tong University), Hai-Ming Xu (The University of Adelaide), and Chao Ma (Shanghai Jiao Tong University)</i>	
SA-BEV: Generating Semantic-Aware Bird's-Eye-View Feature for Multi-view 3D Object Detection	3325
<i>Jinqing Zhang (Beihang University), Yanan Zhang (Beihang University), Qingjie Liu (State Key Laboratory of Virtual Reality Technology and System, Beihang University, Beijing 100191, China), and Yunhong Wang (State Key Laboratory of Virtual Reality Technology and System, Beihang University, Beijing 100191, China)</i>	
GraphAlign: Enhancing Accurate Feature Alignment by Graph matching for Multi-Modal 3D Object Detection	3335
<i>Ziying Song (School of Computer and Information Technology, Beijing Jiaotong University), Haiyue Wei (School of Information Science and Engineering, Hebei University of Science and Technology), Lin Bai (Beijing Jiaotong University), Lei Yang (Tsinghua University), and Caiyan Jia (School of Computer and Information Technology & Beijing Key Lab of Traffic Data Analysis and Mining)</i>	
Tangent Sampson Error: Fast Approximate Two-view Reprojection Error for Central Camera Models	3347
<i>Mikhail Terekhov (ETH Zurich) and Viktor Larsson (Lund University)</i>	
Using a Waffle Iron for Automotive Point Cloud Semantic Segmentation	3356
<i>Gilles Puy (valeo.ai), Alexandre Boulch (valeo.ai), and Renaud Marlet (Ecole des Ponts ParisTech / Valeo)</i>	
Fast Globally Optimal Surface Normal Estimation from an Affine Correspondence	3367
<i>Levente Hajder (Eötvös Loránd University), Lajos Lóczi (Eötvös Loránd University), and Daniel Barath (ETH Zürich)</i>	

Preface: A Data-driven Volumetric Prior for Few-shot Ultra High-resolution Face Synthesis	3379
<i>Marcel C. Bühler (ETH Zürich), Kripasindhu Sarkar (Google), Tanmay Shah (Google), Gengyan Li (ETH Zürich/Google), Daoye Wang (Google), Leonhard Helminger (Google), Sergio Orts-Escolano (Google), Dmitry Lagun (Google), Otmar Hilliges (ETH Zurich), Thabo Beeler (Google), and Abhimitra Meka (Google)</i>	
Canonical Factors for Hybrid Neural Fields	3391
<i>Brent Yi (UC Berkeley), Weijia Zeng (UC Berkeley), Sam Buchanan (TTIC), and Yi Ma (UC Berkeley)</i>	
Center-Based Decoupled Point-cloud Registration for 6D Object Pose Estimation	3404
<i>Haobo Jiang (Nanjing University of Science and Technology), Zheng Dang (EPFL), Shuo Gu (Nanjing University of Science and Technology), Jin Xie (Nanjing University of Science and Technology), Mathieu Salzmann (EPFL), and Jian Yang (Nanjing University of Science and Technology)</i>	
Deep Geometry-Aware Camera Self-Calibration from Video	3415
<i>Annika Hagemann (Robert Bosch GmbH), Moritz Knorr (Robert Bosch GmbH), and Christoph Stiller (Institute of Measurement and Control Systems, Karlsruhe Institute of Technology, (KIT)</i>	
V-FUSE: Volumetric Depth Map Fusion with Long-Range Constraints	3426
<i>Nathaniel Burgdorfer (Stevens Institute of Technology) and Philippos Mordohai (Stevens Institute of Technology)</i>	
Consistent Depth Prediction for Transparent Object Reconstruction from RGB-D Camera	3436
<i>Yuxiang Cai (Nankai University), Yifan Zhu (Nankai University), Haiwei Zhang (Nankai University), and Bo Ren (Nankai University)</i>	
FaceCLIPNeRF: Text-driven 3D Face Manipulation using Deformable Neural Radiance Fields	3446
<i>Sungwon Hwang (KAIST), Junha Hyung (KAIST), Daejin Kim (Korea Advanced Institute of Science and Technology), Min-Jung Kim (KAIST), and Jaegul Choo (Korea Advanced Institute of Science and Technology)</i>	
HollowNeRF: Pruning Hashgrid-Based NeRFs with Trainable Collision Mitigation	3457
<i>Xiufeng Xie (OPPO), Riccardo Gherardi (Oppo), Zhihong Pan (OPPO), and Stephen Huang (OPPO)</i>	
ICE-NeRF: Interactive Color Editing of NeRFs via Decomposition-Aware Weight Optimization ...	3468
<i>Jae-Hyeok Lee (KAIST) and Dae-Shik Kim (KAIST)</i>	
FULLER: Unified Multi-modality Multi-task 3D Perception via Multi-level Gradient Calibration	3479
<i>Zhijian Huang (Sun-Yat-sen University), Sihao Lin (RMIT University), Guiyu Liu (Noah's Ark Lab, Huawei Inc.), Mukun Luo (Shanghai Jiao Tong University), Chaoqiang Ye (Huawei), Hang Xu (Huawei Noah's Ark Lab), Xiaojun Chang (University of Technology Sydney), and Xiaodan Liang (Sun Yat-sen University)</i>	
Neural Fields for Structured Lighting	3489
<i>Aarrushi Shandilya (Carnegie Mellon University), Benjamin Attal (Carnegie Mellon University), Christian Richardt (Meta), James Tompkin (Brown University), and Matthew O'Toole (Carnegie Mellon University)</i>	

CO-Net: Learning Multiple Point Cloud Tasks at Once with A Cohesive Network	3500
<i>Tao Xie (Harbin Institute of Technology), Ke Wang (Harbin Institute of Technology), Siyi Lu (China Coal Technology and Engineering Group Intelligent Storage and Loading Technology Co., Ltd.), Yukun Zhang (Beijing Institute of Technology), Kun Dai (Harbin Institute of Technology), Xiaoyu Li (Harbin Institute of Technology), Jie Xu (Harbin Institute of Technology), Li Wang (Tsinghua University), Lijun Zhao (Harbin Institute of Technology), Xinyu Zhang (Tsinghua University), and Ruifeng Li (Harbin Institute of Technology)</i>	
Pose-Free Neural Radiance Fields via Implicit Pose Regularization	3511
<i>Jiahui Zhang (Nanyang Technological University), Fangneng Zhan (Max Planck Institute for Informatics), Yingchen Yu (Nanyang Technological University), Kunhao Liu (Nanyang Technological University), Rongliang Wu (Nanyang Technological University), Xiaoqin Zhang (Wenzhou University), Ling Shao (UCAS-Terminus AI Lab, UCAS), and Shijian Lu (Nanyang Technological University)</i>	
TransHuman: A Transformer-based Human Representation for Generalizable Neural Human Rendering	3521
<i>Xiao Pan (Zhejiang University), Zongxin Yang (Zhejiang University), Jianxin Ma (Alibaba Group), Chang Zhou (Alibaba Group), and Yi Yang (Zhejiang University)</i>	
S-VolSDF: Sparse Multi-View Stereo Regularization of Neural Implicit Surfaces	3533
<i>Haoyu Wu (Stony Brook University), Alexandros Graikos (Stony Brook University), and Dimitris Samaras (Stony Brook University)</i>	
DPS-Net: Deep Polarimetric Stereo Depth Estimation	3546
<i>Chaoran Tian (Zhejiang University), Weihong Pan (Zhejiang University), Zimo Wang (Zhejiang University), Mao Mao (Zhejiang University), Guofeng Zhang (Zhejiang University), Hujun Bao (Zhejiang University), Ping Tan (Hong Kong University of Science and Technology), and Zhaopeng Cui (Zhejiang University)</i>	
3DPPE: 3D Point Positional Encoding for Transformer-based Multi-Camera 3D Object Detection..	3557
<i>Changyong Shu (Houmo.AI), Jiajun Deng (The University of Sydney), Fisher Yu (ETH Zurich), and Yifan Liu (University of Adelaide)</i>	
Deformable Neural Radiance Fields using RGB and Event Cameras	3567
<i>Qi Ma (ETH zurich), Danda Pani Paudel (ETH Zürich, INSAIT Sofia), Ajad Chhatkuli (ETH Zurich), and Luc Van Gool (ETH Zurich)</i>	
NeILF++: Inter-Reflectable Light Fields for Geometry and Material Estimation	3578
<i>Jingyang Zhang (Apple), Yao Yao (Nanjing University), Shiwei Li (Apple Inc.), Jingbo Liu (Apple Inc.), Tian Fang (Apple), David McKinnon (Apple), Yanghai Tsin (Apple Inc), and Long Quan (Apple)</i>	
Hierarchical Prior Mining for Non-local Multi-View Stereo	3588
<i>Chunlin Ren (Northwestern Polytechnical University), Qingshan Xu (Huazhong University of Science and Technology), Shikun Zhang (Northwestern Polytechnical University), and Jiaqi Yang (Northwestern Polytechnical University)</i>	

Exploring Object-Centric Temporal Modeling for Efficient Multi-View 3D Object Detection	3598
<i>Shihao Wang (Beijing Institute of Technology), Yingfei Liu (Megvii Technology), Tiancai Wang (Megvii Technology), Ying Li (Beijing Institute of Technology), and Xiangyu Zhang (Megvii Technology)</i>	
Re-ReND: Real-Time Rendering of NeRFs across Devices	3609
<i>Sara Rojas (KAUST), Jesus Zarzar (KAUST), Juan C. Perez (KAUST), Artsiom Sanakoyeu (Meta AI), Ali Thabet (Facebook), Albert Pumarola (Meta), and Bernard Ghanem (KAUST)</i>	
Learning Shape Primitives via Implicit Convexity Regularization	3619
<i>Xiaoyang Huang (Shanghai Jiao Tong University), Yi Zhang (Shanghai Jiao Tong University), Kai Chen (Shanghai AI Laboratory), Teng Li (Anhui University), Wenjun Zhang (Shanghai Jiao Tong University), and Bingbing Ni (Shanghai Jiao Tong University)</i>	
Geometry-guided Feature Learning and Fusion for Indoor Scene Reconstruction	3629
<i>Ruihong Yin (University of Amsterdam), Sezer Karaoglu (University of Amsterdam), and Theo Gevers (University of Amsterdam)</i>	
LiDAR-Camera Panoptic Segmentation via Geometry-Consistent and Semantic-Aware Alignment	3639
<i>Zhiwei Zhang (Shanghai Jiao Tong University), Zhizhong Zhang (East China Normal University), Qian Yu (East China Normal University), Ran Yi (Shanghai Jiao Tong University), Yuan Xie (East China Normal University), and Lizhuang Ma (Shanghai Jiao Tong University)</i>	
PivotNet: Vectorized Pivot Learning for End-to-end HD Map Construction	3649
<i>Wenjie Ding (Megvii), Limeng Qiao (Megvii Inc.), Xi Qiu (Megvii Inc.), and Chi Zhang (Megvii Inc.)</i>	
Sat2Density: Faithful Density Learning from Satellite-Ground Image Pairs	3660
<i>Ming Qian (Wuhan University), Jincheng Xiong (Wuhan University), Gui-Song Xia (Wuhan University), and Nan Xue (Wuhan University)</i>	
Mask-Attention-Free Transformer for 3D Instance Segmentation	3670
<i>Xin Lai (The Chinese University of Hong Kong), Yuhui Yuan (Microsoft Research), Ruihang Chu (The Chinese University of Hong Kong), Yukang Chen (The Chinese University of Hong Kong), Han Hu (Microsoft Research Asia), and Jiaya Jia (Chinese University of Hong Kong)</i>	
Scene-Aware Feature Matching	3681
<i>Xiaoyong Lu (Southeast University), Yaping Yan (Southeast University), Tong Wei (Southeast University), and Songlin Du (Southeast University)</i>	
Revisiting Domain-Adaptive 3D Object Detection by Reliable, Diverse and Class-balanced Pseudo-Labeling	3691
<i>Zhuoxiao Chen (University of Queensland), Yadan Luo (University of Queensland), Zheng Wang (UESTC), Mahsa Baktashmotlagh (University of Queensland), and Zi Huang (University of Queensland)</i>	
GO-SLAM: Global Optimization for Consistent 3D Instant Reconstruction	3704
<i>Youmin Zhang (University of Bologna), Fabio Tosi (University of Bologna), Stefano Mattocchia (University of Bologna), and Matteo Poggi (University of Bologna)</i>	
BANSAC: A Dynamic Bayesian Network for Adaptive Sample Consensus	3715
<i>Valter Piedade (Instituto Superior Técnico) and Pedro Miraldo (MERL Mitsubishi Electric Research Laboratories)</i>	

Theoretical and Numerical Analysis of 3D Reconstruction Using Point and Line Incidences	3725
<i>Felix Rydell (KTH), Elima Shehu (Osnabrück University and Max Planck Institute for Mathematics in the Sciences), and Angélica Torres (Centre de Recerca Matemàtica)</i>	
RealGraph: A Multiview Dataset for 4D Real-world Context Graph Generation	3735
<i>Haozhe Lin (Tsinghua University), Zequn Chen (Tsinghua University), Jinzhi Zhang (Tsinghua University), Bing Bai (Tsinghua University), Yu Wang (JD AI Research), Ruqi Huang (Tsinghua Shenzhen International Graduate School, (SIGS), and Lu Fang (Tsinghua University)</i>	
CL-MVSNet: Unsupervised Multi-View Stereo with Dual-Level Contrastive Learning	3746
<i>Kaiqiang Xiong (Peking University), Rui Peng (Peking University), Zhe Zhang (Peking University), Tianxing Feng (Peking University), Jianbo Jiao (University of Birmingham; University of Oxford), Feng Gao (Peking University), and Ronggang Wang (Peking University)</i>	
Temporal Enhanced Training of Multi-view 3D Object Detector via Historical Object Prediction	3758
<i>Zhuofan Zong (SenseTime), Dongzhi Jiang (The Chinese University of Hong Kong), Guanglu Song (Sensetime), Zeyue Xue (The University of Hong Kong), Jingyong Su (Harbin Institute of Technology, Shenzhen), Hongsheng Li (The Chinese University of Hong Kong), and Yu Liu (SenseTime Group LTD)</i>	
Object as Query: Lifting Any 2D Object Detector to 3D Detection	3768
<i>Zitian Wang (Beihang University), Zehao Huang (TuSimple), Jiahui Fu (Beihang University), Naiyan Wang (TuSimple), and Si Liu (Beihang University)</i>	
PARTNER: Level up the Polar Representation for LiDAR 3D Object Detection	3778
<i>Ming Nie (Fudan University), Yujing Xue (National University of Singapore), Chunwei Wang (Shanghai Jiao Tong University), Chaoqiang Ye (Huawei), Hang Xu (Huawei Noah's Ark Lab), Xinge Zhu (The Chinese University of Hong Kong), Qingqiu Huang (Huawei), Michael Bi Mi (Huawei International Pte Ltd), Xinchao Wang (National University of Singapore), and Li Zhang (Fudan University)</i>	
Not Every Side Is Equal: Localization Uncertainty Estimation for Semi-Supervised 3D Object Detection	3791
<i>Chuxin Wang (University of Science and Technology of China), Wenfei Yang (University of Science and Technology of China), and Tianzhu Zhang (University of Science and Technology of China)</i>	
QD-BEV : Quantization-aware View-guided Distillation for Multi-view 3D Object Detection	3802
<i>Yifan Zhang (Nanjing University), Zhen Dong (UC Berkeley), Huanrui Yang (UC Berkeley), Ming Lu (Intel Labs China), Cheng-Ching Tseng (Peking University), Yuan Du (Nanjing University), Kurt Keutzer (EECS, UC Berkeley), Li Du (Nanjing University), and Shanghang Zhang (Peking University)</i>	

(O) Vision, Graphics, and Robotics

Adding Conditional Control to Text-to-Image Diffusion Models	3813
<i>Lvin Zhang (Stanford University), Anyi Rao (Stanford University), and Maneesh Agrawala (Stanford University)</i>	
Factorized Inverse Path Tracing for Efficient and Accurate Material-Lighting Estimation	3825
<i>Liven Wu (UCSD), Rui Zhu (UCSD), Mustafa B. Yaldiz (University of California, San Diego), Yinhao Zhu (Qualcomm AI Research), Hong Cai (Qualcomm AI Research), Janarbek Matai (Qualcomm AI Research), Fatih Porikli (Australian National University), Tzu-Mao Li (University of California, San Diego), Manmohan Chandraker (UC San Diego), and Ravi Ramamoorthi (University of California San Diego)</i>	
Manipulate by Seeing: Creating Manipulation Controllers from Pre-Trained Representations	3836
<i>Jianren Wang (Carnegie Mellon University), Sudeep Dasari (Carnegie Mellon University), Mohan Kumar Srirama (Carnegie Mellon University), Shubham Tulsiani (Carnegie Mellon University), and Abhinav Gupta (Carnegie Mellon University Robotics Institute)</i>	
3D Implicit Transporter for Temporally Consistent Keypoint Discovery	3846
<i>Chengliang Zhong (Xi'an High-Tech Research Institution, Tsinghua University), Yuhang Zheng (Beihang University), Yupeng Zheng (Institute of Automation, Chinese Academy of Sciences), Hao Zhao (Tsinghua University), Li Yi (Tsinghua University), Xiaodong Mu (Xi'an High-Tech Research Institution), Ling Wang (Xi'an Research Institute of High-Tech), Pengfei Li (Tsinghua University), Guyue Zhou (Tsinghua University), Chao Yang (Shanghai AI Laboratory), Xinliang Zhang (), and Jian Zhao (Institute of North Electronic Equipment)</i>	
Chordal Averaging on Flag Manifolds and Its Applications	3858
<i>Nathan Mankovich (Colorado State University) and Tolga Birdal (Imperial College London)</i>	
UniDexGrasp++: Improving Dexterous Grasping Policy Learning via Geometry-Aware Curriculum and Iterative Generalist-Specialist Learning	3868
<i>Weikang Wan (Peking University), Haoran Geng (Peking University), Yun Liu (Tsinghua University), Zikang Shan (Peking University), Yaodong Yang (Institute for AI, Peking University), Li Yi (Tsinghua University), and He Wang (Peking University)</i>	
GameFormer: Game-theoretic Modeling and Learning of Transformer-based Interactive Prediction and Planning for Autonomous Driving	3880
<i>Zhiyu Huang (Nanyang Technological University), Haochen Liu (Nanyang Technological University), and Chen Lv (Nanyang Technological University)</i>	
PPR: Physically Plausible Reconstruction from Monocular Videos	3891
<i>Gengshan Yang (Carnegie Mellon University), Shuo Yang (Carnegie Mellon University), John Z. Zhang (Carnegie Mellon University), Zachary Manchester (Carnegie Mellon University), and Deva Ramanan (CMU)</i>	

(O) Privacy, Security, Fairness, and Explainability

Zolly: Zoom Focal Length Correctly for Perspective-Distorted Human Mesh Reconstruction	3902
<i>Wenjia Wang (The University of Hong Kong), Yongtao Ge (The University of Adelaide), Haiyi Mei (SenseTime Group Limited), Zhongang Cai (SenseTime International Pte Ltd), Qingping Sun (City University of Hong Kong), Yanjun Wang (Shanghai Jiao Tong University), Chunhua Shen (University of Adelaide, Australia), Lei Yang (SenseTime Group Limited), and Taku Komura (The University of Hong Kong)</i>	
ACLS: Adaptive and Conditional Label Smoothing for Network Calibration	3913
<i>Hyekang Park (Yonsei University), Jongyoun Noh (Yonsei University), Youngmin Oh (Yonsei University), Donghyeon Baek (Yonsei University), and Bumsub Ham (Yonsei University)</i>	
PGFed: Personalize Each Client's Global Objective for Federated Learning	3923
<i>Jun Luo (University of Pittsburgh), Matias Mendieta (University of Central Florida), Chen Chen (University of Central Florida), and Shandong Wu (Department of Radiology, Biomedical Informatics, and Bioengineering, University of Pittsburgh)</i>	
Overwriting Pretrained Bias with Finetuning Data	3934
<i>Angelina Wang (Princeton University) and Olga Russakovsky (Princeton University)</i>	
ITI-GEN: Inclusive Text-to-Image Generation	3946
<i>Cheng Zhang (Carnegie Mellon University), Xuanbai Chen (Carnegie Mellon University), Siqi Chai (Carnegie Mellon University), Chen Henry Wu (Carnegie Mellon University), Dmitry Lagun (Google), Thabo Beeler (Google), and Fernando de la Torre (Carnegie Mellon)</i>	
FunnyBirds: A Synthetic Vision Dataset for a Part-Based Analysis of Explainable AI Methods.....	3958
<i>Robin Hesse (Technische Universität Darmstadt), Simone Schaub-Meyer (TU Darmstadt), and Stefan Roth (TU Darmstadt)</i>	
X-VoE: Measuring eXplanatory Violation of Expectation in Physical Events	3969
<i>Bo Dai (Peking University), Linge Wang (Tsinghua University), Baoxiong Jia (UCLA), Zeyu Zhang (UCLA), Song-Chun Zhu (UCLA), Chi Zhang (University of California, Los Angeles), and Yixin Zhu (Peking University)</i>	
Adaptive Testing of Computer Vision Models	3980
<i>Irena Gao (Stanford University), Gabriel Ilharco (University of Washington), Scott Lundberg (Microsoft Research), and Marco Tulio Ribeiro (Microsoft Research)</i>	

(O) Recognition, Segmentation, and Shape Analysis

Segment Anything	3992
<i>Alexander Kirillov (Meta AI Research, FAIR), Eric Mintun (Meta AI Research, FAIR), Nikhila Ravi (Meta AI Research, FAIR), Hanzi Mao (Meta AI Research, FAIR), Chloe Rolland (Meta AI Research, FAIR), Laura Gustafson (Meta AI Research, FAIR), Tete Xiao (Meta AI Research, FAIR), Spencer Whitehead (Meta AI Research, FAIR), Alexander C. Berg (Meta AI Research, FAIR), Wan-Yen Lo (Meta AI Research, FAIR), Piotr Dollar (Meta AI Research, FAIR), and Ross Girshick (Meta AI Research, FAIR)</i>	

Shape Analysis of Euclidean Curves under Frenet-Serret Framework	4004
<i>Perrine Chassat (Université Paris Saclay), Juhyun Park (ENSIIE), and Nicolas Brunel (ENSIIE)</i>	
Unmasking Anomalies in Road-Scene Segmentation	4014
<i>Shyam Nandan Rai (Politecnico di Torino), Fabio Cermelli (Politecnico di Torino), Dario Fontanel (Politecnico di Torino), Carlo Masone (Politecnico di Torino), and Barbara Caputo (Politecnico di Torino)</i>	
High Quality Entity Segmentation	4024
<i>Lu Qi (The University of California, Merced), Jason Kuen (Adobe Research), Tiancheng Shen (The Chinese University of Hong Kong), Jiuxiang Gu (Adobe Research), Wenbo Li (The Chinese University of Hong Kong), Weidong Guo (Tencent), Jiaya Jia (Chinese University of Hong Kong), Zhe Lin (Adobe Research), and Ming-Hsuan Yang (University of California at Merced)</i>	
Towards Open-Vocabulary Video Instance Segmentation	4034
<i>Haochen Wang (UvA), Cilin Yan (Beihang University), Shuai Wang (University of Amsterdam), Xiaolong Jiang (Xiaohongshu Inc.), Xu Tang (Xiaohongshu Inc.), Yao Hu (Xiaohongshu), Weidi Xie (Shanghai Jiao Tong University), and Efstratios Gavves (University of Amsterdam)</i>	
Beyond One-to-One: Rethinking the Referring Image Segmentation	4044
<i>Yutao Hu (The University of Hong Kong), Qixiong Wang (Beihang), Wenqi Shao (The Chinese University of HongKong), Enze Xie (The University of Hong Kong), Zhenguo Li (Huawei Noah's Ark Lab), Jungong Han (University of Sheffield), and Ping Luo (The University of Hong Kong)</i>	
Multiple Instance Learning Framework with Masked Hard Instance Mining for Whole Slide Image Classification	4055
<i>Wenhao Tang (Chongqing University), Sheng Huang (Chongqing University), Xiaoxian Zhang (Chongqing University), Fengtao Zhou (The Hong Kong University of Science and Technology), Yi Zhang (Chongqing University), and Bo Liu (JD.com)</i>	
Scale-MAE: A Scale-Aware Masked Autoencoder for Multiscale Geospatial Representation Learning	4065
<i>Colorado J Reed (University of California, Berkeley), Ritwik Gupta (University of California, Berkeley), Shufan Li (University of California, Berkeley), Sarah Brockman (Kitware, Inc.), Christopher Funk (Kitware, Inc), Brian Clipp (Kitware), Kurt Keutzer (EECS, UC Berkeley), Salvatore Candido (MetaAI), Matt Uyttendaele (Meta), and Trevor Darrell (UC Berkeley)</i>	
Progressive Spatio-Temporal Prototype Matching for Text-Video Retrieval	4077
<i>Pandeng Li (University of Science and Technology of China), Chen-Wei Xie (Alibaba Group), Liming Zhao (Alibaba Group), Hongtao Xie (University of Science and Technology of China), Jiannan Ge (University of Science and Technology of China), Yun Zheng (Alibaba group), Deli Zhao (Alibaba Group), and Yongdong Zhang (University of Science and Technology of China)</i>	

Towards Deeply Unified Depth-aware Panoptic Segmentation with Bi-directional Guidance Learning	4088
<i>Junwen He (Dalian University of Technology), Yifan Wang (Dalian University of Technology), Lijun Wang (Dalian University of Technology), Huchuan Lu (Dalian University of Technology), Bin Luo (DAMO Academy, Alibaba Group), Jun-Yan He (DAMO Academy, Alibaba Group), Jin-Peng Lan (DAMO Academy, Alibaba Group), Yifeng Geng (Alibaba Group), and Xuansong Xie (Alibaba)</i>	
LogicSeg: Parsing Visual Semantics with Neural Logic Learning and Reasoning	4099
<i>Liulei Li (University of Technology Sydney), Wenguan Wang (Zhejiang University), and Yi Yang (Zhejiang University)</i>	
ASIC: Aligning Sparse in-the-wild Image Collections	4111
<i>Kamal Gupta (University of Maryland), Varun Jampani (Google), Carlos Esteves (Google Research), Abhinav Shrivastava (University of Maryland), Ameesh Makadia (Google Research), Noah Snavely (Cornell University and Google AI), and Abhishek Kar (Google)</i>	

(O) Generative AI 1

CLIPascene: Scene Sketching with Different Types and Levels of Abstraction	4123
<i>Yael Vinker (Tel Aviv University), Yuval Alaluf (Tel Aviv University), Daniel Cohen-Or (Tel Aviv University), and Ariel Shamir (The Interdisciplinary Center)</i>	
LD-ZNet: A Latent Diffusion Approach for Text-Based Image Segmentation	4134
<i>Koutilya PNVR (University of Maryland), Bharat Singh (vchar.ai), Pallabi Ghosh (Amazon), Behjat Siddiquie (Amazon), and David Jacobs (University of Maryland, USA)</i>	
TexFusion: Synthesizing 3D Textures with Text-Guided Image Diffusion Models	4146
<i>Tianshi Cao (University of Toronto, NVIDIA), Karsten Kreis (NVIDIA), Sanja Fidler (University of Toronto, NVIDIA), Nicholas Sharp (NVIDIA), and Kangxue Yin (NVIDIA)</i>	
NeuRBF: A Neural Fields Representation with Adaptive Radial Basis Functions	4159
<i>Zhang Chen (OPPO US Research Center, InnoPeak Technology, Inc.), Zhong Li (InnoPeak Technology, Inc.), Liangchen Song (University at Buffalo), Lele Chen (University of Rochester), Jingyi Yu (Shanghai Tech University), Junsong Yuan (State University of New York at Buffalo, USA), and Yi Xu (OPPO US Research Center)</i>	
Scalable Diffusion Models with Transformers	4172
<i>William Peebles (University of California, Berkeley) and Saining Xie (New York University)</i>	
Texture Generation on 3D Meshes with Point-UV Diffusion	4183
<i>Xin Yu (The University of Hong Kong), Peng Dai (The University of Hong Kong), Wenbo Li (The Chinese University of Hong Kong), Lan Ma (TCL Corporate Research), Zhengzhe Liu (The Chinese University of Hong Kong), and Xiaojuan Qi (The University of Hong Kong)</i>	

Generative Novel View Synthesis with 3D-Aware Diffusion Models	4194
<i>Eric R. Chan (Stanford University), Koki Nagano (NVIDIA), Matthew A. Chan (Stanford), Alexander W. Bergman (Stanford University), Jeong Joon Park (Stanford University), Axel Levy (Stanford University), Miika Aittala (NVIDIA), Shalini De Mello (NVIDIA Research), Tero Karras (NVIDIA), and Gordon Wetzstein (Stanford University)</i>	
DiffFit: Unlocking Transferability of Large Diffusion Models via Simple Parameter-efficient Fine-Tuning	4207
<i>Enze Xie (The University of Hong Kong), Lewei Yao (Huawei Noah's Ark Lab), Han Shi (The Hong Kong University of Science and Technology), Zhili Liu (Huawei Noah's Ark Lab), Daquan Zhou (National University of Singapore), Zhaoqiang Liu (Huawei), Jiawei Li (Huawei Noah's Ark Lab), and Zhenguo Li (Huawei Noah's Ark Lab)</i>	
VQ3D: Learning a 3D-Aware Generative Model on ImageNet	4217
<i>Kyle Sargent (Google), Jing Yu Koh (Carnegie Mellon University), Han Zhang (Google), Huiwen Chang (Google Research), Charles Herrmann (Cornell), Pratul Srinivasan (Google Research), Jiajun Wu (Stanford University), and Deqing Sun (Google)</i>	
Ref-NeuS: Ambiguity-Reduced Neural Implicit Surface Learning for Multi-View Reconstruction with Reflection	4228
<i>Wenhao Ge (Hong Kong University of Science and Technology, (Guangzhou), Tao Hu (Chinese University of Hong Kong), Haoyu Zhao (The Hong Kong University of Science and Technology), Shu Liu (SmartMore), and Ying-Cong Chen (Hong Kong University of Science and Technology, (Guangzhou)</i>	
A Complete Recipe for Diffusion Generative Models	4238
<i>Kushagra Pandey (University of California, Irvine) and Stephan Mandt (University of California, Irvine)</i>	
MMVP: Motion-Matrix-Based Video Prediction	4250
<i>Yiqi Zhong (University of Southern California), Luming Liang (Microsoft), Ilya Zharkov (Microsoft), and Ulrich Neumann (USC)</i>	

Adversarial Attack and Defense

SAGA: Spectral Adversarial Geometric Attack on 3D Meshes	4261
<i>Tomer Stolik (Tel Aviv University), Itai Lang (Tel Aviv University), and Shai Avidan (Tel Aviv University)</i>	
Benchmarking and Analyzing Robust Point Cloud Recognition: Bag of Tricks for Defending Adversarial Examples	4272
<i>Qiufan Ji (New Jersey Institute of Technology), Lin Wang (HKUST), Cong Shi (New Jersey Institute of Technology), Shengshan Hu (Huazhong University of Science and Technology), Yingying Chen (Rutgers University), and Lichao Sun (Lehigh University)</i>	

ACTIVE: Towards Highly Transferable 3D Physical Camouflage for Universal and Robust Vehicle Evasion	4282
<i>Naufal Suryanto (Pusan National University), Yongsu Kim (SmartM2M), Harashta Tatimma Larasati (Pusan National University, Institut Teknologi Bandung), Hyeoun Kang (Pusan National University), Thi-Thu-Huong Le (Pusan National University), Yoonyoung Hong (Pusan National University), Hunmin Yang (KAIST), Se-Yoon Oh (Agency for Defense Development), and Howon Kim (Pusan National University)</i>	
Frequency-aware GAN for Adversarial Manipulation Generation	4292
<i>Peifei Zhu (LINE Corporation), Genki Osada (LINE Corporation), Hirokatsu Kataoka (LINE Corporation), and Tsubasa Takahashi (LINE Corporation)</i>	
Breaking Temporal Consistency: Generating Video Universal Adversarial Perturbations Using Image Models	4302
<i>Hee-Seon Kim (KAIST), Minji Son (KAIST), Minbeom Kim (KAIST), Myung-Joon Kwon (KAIST), and Changick Kim (KAIST)</i>	
Tracing the Origin of Adversarial Attack for Forensic Investigation and Deterrence	4312
<i>Han Fang (National University of Singapore), Jiyi Zhang (National University of Singapore), Yupeng Qiu (National University of Singapore), Jiayang Liu (National University of Singapore), Ke Xu (Huawei International), Chengfang Fang (Huawei International), and Ee-Chien Chang (NUS)</i>	
Downstream-agnostic Adversarial Examples	4322
<i>Ziqi Zhou (Huazhong University of Science and Technology), Shengshan Hu (Huazhong University of Science and Technology), Ruizhi Zhao (Huazhong University of Science and Technology), Qian Wang (Wuhan University), Leo Yu Zhang (Griffith University), Junhui Hou (City University of Hong Kong, Hong Kong), and Hai Jin (Huazhong University of Science and Technology)</i>	
Hiding Visual Information via Obfuscating Adversarial Perturbations	4333
<i>Zhigang Su (Xidian University), Dawei Zhou (Xidian University), Nannan Wang (Xidian University), Decheng Liu (Xidian University), Zhen Wang (Zhejiang Lab), and Xinbo Gao (Chongqing University of Posts and Telecommunications)</i>	
An Embarrassingly Simple Backdoor Attack on Self-supervised Learning	4344
<i>Changjiang Li (Penn State), Ren Pang (The Pennsylvania State University), Zhaohan Xi (The Pennsylvania State University), Tianyu Du (Penn State University), Shouling Ji (Zhejiang University), Yuan Yao (Nanjing University), and Ting Wang (Penn State)</i>	
Efficient Decision-based Black-box Patch Attacks on Video Recognition	4356
<i>Kaixun Jiang (Fudan University), Zhaoyu Chen (Fudan University), Hao Huang (Amazon/Peking University), Jiafeng Wang (Fudan University), Dingkang Yang (Fudan University), Bo Li (Tencent), Yan Wang (Fudan University), and Wenqiang Zhang (Fudan University)</i>	

Adversarial Finetuning with Latent Representation Constraint to Mitigate Accuracy-Robustness Tradeoff	4367
<i>Satoshi Suzuki (NTT Computer and Data Science Laboratories / The University of Electro-Communications), Shin'ya Yamaguchi (NTT / Kyoto University), Shoichiro Takeda (NTT Corporation), Sekitoshi Kanai (NTT), Naoki Makishima (NTT), Atsushi Ando (NTT Corporation), and Ryo Masumura (NTT Corporation)</i>	
Towards Building More Robust Models with Frequency Bias	4379
<i>Qingwen Bu (Shanghai Jiao Tong University), Dong Huang (The University of Hong Kong), and Heming Cui (The University of Hong Kong)</i>	
Does Physical Adversarial Example Really Matter to Autonomous Driving? Towards System-Level Effect of Adversarial Object Evasion Attack	4389
<i>Ningfei Wang (University of California, Irvine), Yunpeng Luo (University of California, Irvine), Takami Sato (University of California, Irvine), Kaidi Xu (Drexel University), and Qi Alfred Chen (UC Irvine)</i>	
Improving Generalization of Adversarial Training via Robust Critical Fine-Tuning	4401
<i>Kaijie Zhu (Institute of Automation, Chinese Academy of Sciences), Xixu Hu (City University of Hong Kong), Jindong Wang (Microsoft Research), Xing Xie (Microsoft Research Asia), and Ge Yang (National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences)</i>	
Enhancing Generalization of Universal Adversarial Perturbation through Gradient Aggregation	4412
<i>Xuannan Liu (Beijing University of Posts and Telecommunications), Yaoyao Zhong (Beijing University of Posts and Telecommunications), Yuhang Zhang (Beijing University of Posts and Telecommunicates), Lixiong Qin (Beijing University of Posts and Telecommunications), and Weihong Deng (Beijing University of Posts and Telecommunications)</i>	
Unified Adversarial Patch for Cross-Modal Attacks in the Physical World	4422
<i>Xingxing Wei (Beihang University), Yao Huang (Beihang University), Yitong Sun (Beihang University), and Jie Yu (Beihang University)</i>	
RFLA: A Stealthy Reflected Light Adversarial Attack in the Physical World	4432
<i>Donghua Wang (Zhejiang University), Wen Yao (National Innovation Institute of Defense Technology, Chinese Academy of Military Science), Tingsong Jiang (Defense Innovation Institute, Chinese Academy of Military Science), Chao Li (Xidian University), and Xiaoqian Chen (National Innovation Institute of Defense Technology, Chinese Academic of Military Science)</i>	
Enhancing Fine-Tuning Based Backdoor Defense with Sharpness-Aware Minimization	4443
<i>Mingli Zhu (The Chinese University of Hong Kong, Shenzhen, Shenzhen Research Institute of Big Data), Shaokui Wei (Shenzhen Research Institute of Big Data, The Chinese University of Hong Kong, Shenzhen), Li Shen (JD Explore Academy), Yanbo Fan (Tencent AI Lab), and Baoyuan Wu (The Chinese University of Hong Kong, Shenzhen, Shenzhen Research Institute of Big Data)</i>	

Conditional 360-degree Image Synthesis for Immersive Indoor Scene Decoration	4455
<i>Ka Chun Shum (The Hong Kong University of Science and Technology), Hong-Wing Pang (The Hong Kong University of Science and Technology), Binh-Son Hua (Trinity College Dublin), Duc Thanh Nguyen (Deakin University, Australia), and Sai-Kit Yeung (Hong Kong University of Science and Technology)</i>	
An Adaptive Model Ensemble Adversarial Attack for Boosting Adversarial Transferability	4466
<i>Bin Chen (Fuzhou University), Jiali Yin (Fuzhou University), Shukai Chen (Fuzhou University), Bohao Chen (Yuan Ze University), and Ximeng Liu (Fuzhou University)</i>	
Mitigating Adversarial Vulnerability through Causal Parameter Estimation by Adversarial Double Machine Learning	4476
<i>Byung-Kwan Lee (KAIST), Junho Kim (KAIST), and Yong Man Ro (KAIST)</i>	
LEA2: A Lightweight Ensemble Adversarial Attack via Non-overlapping Vulnerable Frequency Regions	4487
<i>Yaguan Qian (Zhejiang University of Science and Technology), Shuke He (Zhejiang University of Science and Technology), Chenyu Zhao (Zhejiang University of Science and Technology), Jiaqiang Sha (Zhejiang University of Science and Technology), Wei Wang (Beijing Jiaotong University), and Bin Wang (Network and Information Security Laboratory of Hangzhou Hikvision Digital Technology Co.)</i>	
Explaining Adversarial Robustness of Neural Networks from Clustering Effect Perspective	4499
<i>Yulin Jin (Xidian University), Xiaoyu Zhang (Xidian University), Jian Lou (Zhejiang University), Xu Ma (Qufu Normal university), Zilong Wang (Xidian University), and Xiaofeng Chen (Xidian University)</i>	
VertexSerum: Poisoning Graph Neural Networks for Link Inference	4509
<i>Ruyi Ding (Northeastern University), Shijin Duan (Northeastern University), Xiaolin Xu (Northeastern University), and Yunsi Fei (NEU)</i>	
How to Choose your Best Allies for a Transferable Attack?	4519
<i>Thibault Maho (Inria), Seyed-Mohsen Moosavi-Dezfooli (Imperial College London), and Teddy Furon (Inria)</i>	
Enhancing Adversarial Robustness in Low-Label Regime via Adaptively Weighted Regularization and Knowledge Distillation	4529
<i>Dongyoon Yang (Seoul National University), Insung Kong (Seoul National University), and Yongdai Kim (Seoul National University)</i>	
AdvDiffuser: Natural Adversarial Example Synthesis with Diffusion Models	4539
<i>Xinquan Chen (Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences), Xitong Gao (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences), Juanjuan Zhao (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences), Kejiang Ye (Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences), and Cheng-Zhong Xu (University of Macau)</i>	
F&F Attack: Adversarial Attack against Multiple Object Trackers by Inducing False Negatives and False Positives	4550
<i>Tao Zhou (Zhejiang University), Qi Ye (Zhejiang Univeristy), Wenhan Luo (Sun Yat-sen University), Kaihao Zhang (Australian National University), Zhiguo Shi (Zhejiang University), and Jiming Chen (Zhejiang University)</i>	

Rickrolling the Artist: Injecting Backdoors into Text Encoders for Text-to-Image Synthesis.....	4561
<i>Lukas Struppek (Technical University of Darmstadt), Dominik Hintersdorf (Technical University of Darmstadt), and Kristian Kersting (TU Darmstadt)</i>	
Hard No-Box Adversarial Attack on Skeleton-Based Human Action Recognition with Skeleton-Motion-Informed Gradient	4574
<i>Zhengzhi Lu (Xi'an Jiaotong University), He Wang (University College London), Ziyi Chang (Durham University), Guoan Yang (Xi'an Jiaotong University), and Hubert P. H. Shum (Durham University)</i>	
Structure Invariant Transformation for better Adversarial Transferability	4584
<i>Xiaosen Wang (Huazhong University of Science and Technology), Zeliang Zhang (University of Rochester), and Jianping Zhang (The Chinese University of Hong Kong)</i>	
Beating Backdoor Attack at Its Own Game	4597
<i>Min Liu (Carnegie Mellon University), Alberto Sangiovanni-Vincentelli (University of California, Berkeley), and Xiangyu Yue (The Chinese University of Hong Kong)</i>	
Transferable Adversarial Attack for Both Vision Transformers and Convolutional Networks via Momentum Integrated Gradients	4607
<i>Wenshuo Ma (Tsinghua University), Yidong Li (Beijing Jiaotong University, China), Xiaofeng Jia (Beijing Big Data Centre), and Wei Xu (Tsinghua University)</i>	
REAP: A Large-Scale Realistic Adversarial Patch Benchmark	4617
<i>Nabeel Hingun (UC Berkeley), Chawin Sitawarin (UC Berkeley), Jerry Li (Microsoft), and David Wagner (UC Berkeley)</i>	
Multi-Metrics Adaptively Identifies Backdoors in Federated Learning	4629
<i>Siquan Huang (South China University of Technology), Yijiang Li (Johns Hopkins University), Chong Chen (South China University of Technology), Leyu Shi (South China University of Technology), and Ying Gao (south china university of technology)</i>	
Backpropagation Path Search On Adversarial Transferability	4640
<i>Zhuoer Xu (Nanjing University), Zhangxuan Gu (Ant Group), Jianping Zhang (The Chinese University of Hong Kong), Shiwen Cui (Ant Group), Changhua Meng (Ant Group), and Weiqiang Wang (Ant Group)</i>	
Rapid Network Adaptation: Learning to Adapt Neural Networks Using Test-Time Feedback	4651
<i>Teresa Yeo (EPFL), Oğuzhan Fatih Kar (EPFL), Zahra Sodagar (EPFL), and Amir Zamir (Swiss Federal Institute of Technology, (EPFL)</i>	
One-bit Flip is All You Need: When Bit-flip Attack Meets Model Training	4665
<i>Jianshuo Dong (Tsinghua University), Han Qiu (Tsinghua University), Yiming Li (Tsinghua University), Tianwei Zhang (Nanyang Technological University), Yuanjie Li (Tsinghua University), Zeqi Lai (Tsinghua University), Chao Zhang (Tsinghua University), and Shu-Tao Xia (Tsinghua University)</i>	
PolicyCleanse: Backdoor Detection and Mitigation for Competitive Reinforcement Learning	4676
<i>Junfeng Guo (The University of Texas at Dallas), Ang Li (angli.ai), Lixu Wang (Northwestern University), and Cong Liu (Univ. of California Riverside)</i>	

Towards Viewpoint-Invariant Visual Recognition via Adversarial Training	4686
<i>Shouwei Ruan (beihang university), Yinpeng Dong (Tsinghua University), Hang Su (Tsinghua University), Jianteng Peng (OPPO), Ning Chen (Tsinghua University), and Xingxing Wei (Beihang University)</i>	
Fast Adversarial Training with Smooth Convergence	4697
<i>Mengnan Zhao (Dalian University of Technology), Lihe Zhang (Dalian University of Technology), Yuqiu Kong (Dalian University of Technology), and Baocai Yin (Beijing University of Technology)</i>	
The Perils of Learning From Unlabeled Data: Backdoor Attacks on Semi-supervised Learning	4707
<i>Virat Shejwalkar (UMass Amherst), Lingjuan Lyu (Sony AI), and Amir Houmansadr (University of Massachusetts Amherst)</i>	
Boosting Adversarial Transferability via Gradient Relevance Attack	4718
<i>Hegui Zhu (Northeastern University), Yuchen Ren (Northeastern University), Xiaoyan Sui (Northeastern University), Lianping Yang (Northeastern University), and Wuming Jiang (Beijing Eyecool Technology Co., Ltd)</i>	
Towards Robust Model Watermark via Reducing Parametric Vulnerability	4728
<i>Guanhao Gan (Tsinghua Shenzhen International Graduate School), Yiming Li (Tsinghua University), Dongxian Wu (the University of Tokyo), and Shu-Tao Xia (Tsinghua University)</i>	
TRM-UAP: Enhancing the Transferability of Data-Free Universal Adversarial Perturbation via Truncated Ratio Maximization	4739
<i>Yiran Liu (Chongqing University of Technology), Xin Feng (Chongqing University of Technology), Yunlong Wang (IQVIA), Wu Yang (Chongqing University of Technology), and Di Ming (Chongqing University of Technology)</i>	

Fairness, Privacy, Ethics, Social-Good, Transparency, Accountability in Vision

Enhancing Privacy Preservation in Federated Learning via Learning Rate Perturbation	4749
<i>Guangnian Wan (Beijing University of Posts and Telecommunications), Haitao Du (China Mobile Company), Xuejing Yuan (), Jun Yang (Beijing University of Posts and Telecommunications), Meiling Chen (China Mobile), and Jie Xu (Beijing university of posts and telecommunications)</i>	
TARGET: Federated Class-Continual Learning via Exemplar-Free Distillation	4759
<i>Jie Zhang (Zhejiang University), Chen Chen (Sony AI), Weiming Zhuang (Nanyang Technological University), and Lingjuan Lyu (Sony AI)</i>	
FACTS: First Amplify Correlations and Then Slice to Discover Bias	4771
<i>Sriram Yenamandra (Georgia Institute of Technology), Pratik Ramesh (Georgia Institute of Technology), Viraj Prabhu (Georgia Institute of Technology), and Judy Hoffman (Georgia Institute of Technology)</i>	
Computation and Data Efficient Backdoor Attacks	4782
<i>Yutong Wu (Nanyang Technological University), Xingshuo Han (Nanyang Technological University), Han Qiu (Tsinghua University), and Tianwei Zhang (Nanyang Technological University)</i>	

Global Balanced Experts for Federated Long-Tailed Learning	4792
<i>Yaopei Zeng (The Chinese University of Hong Kong, Shenzhen), Lei Liu (The Chinese University of Hong Kong, Shenzhen), Li Liu (The Hong Kong University of Science and Technology, (Guangzhou), Li Shen (JD Explore Academy), Shaoguo Liu (Alibaba Group), and Baoyuan Wu (The Chinese University of Hong Kong, Shenzhen, Shenzhen Research Institute of Big Data)</i>	
Source-free Domain Adaptive Human Pose Estimation	4803
<i>Qucheng Peng (University of Central Florida), Ce Zheng (University of Central Florida), and Chen Chen (University of Central Florida)</i>	
Gender Artifacts in Visual Datasets	4814
<i>Nicole Meister (Princeton University), Dora Zhao (Sony AI), Angelina Wang (Princeton University), Vikram V. Ramaswamy (Princeton University), Ruth Fong (Princeton University), and Olga Russakovsky (Princeton University)</i>	
FRAug: Tackling Federated Learning with Non-IID Features via Representation Augmentation ..	4826
<i>Haokun Chen (Siemens), Ahmed Frikha (Siemens and LMU), Denis Krompass (Siemens), Jindong Gu (University of Oxford), and Volker Tresp (Siemens AG and Ludwig Maximilian University of Munich)</i>	
zPROBE: Zero Peek Robustness Checks for Federated Learning	4837
<i>Zahra Ghodsi (Purdue University), Mojan Javaheripi (UC San Diego), Nojan Sheybani (UC San Diego), Xinqiao Zhang (UC San Diego), Ke Huang (San Diego State University), and Farinaz Koushanfar (UCSD)</i>	
Practical Membership Inference Attacks Against Large-Scale Multi-Modal Models: A Pilot Study	4848
<i>Myeongseob Ko (Virginia Tech), Ming Jin (Virginia Tech), Chenguang Wang (Washington University), and Ruoxi Jia (Virginia Tech)</i>	
FedPD: Federated Open Set Recognition with Parameter Disentanglement	4859
<i>Chen Yang (City University of Hong Kong), Meilu Zhu (City University of Hong Kong), Yifan Liu (Chinese University of Hong Kong), and Yixuan Yuan (Chinese University of Hong Kong)</i>	
MUter: Machine Unlearning on Adversarially Trained Models	4869
<i>Junxu Liu (Renmin University of China), Mingsheng Xue (Xidian University), Jian Lou (Zhejiang University), Xiaoyu Zhang (Xidian University), Li Xiong (Emory University), and Zhan Qin (Zhejiang University)</i>	
Beyond Skin Tone: A Multidimensional Measure of Apparent Skin Color	4880
<i>William Thong (Sony AI), Przemyslaw Joniak (The University of Tokyo), and Alice Xiang (Partnership on AI)</i>	
A Multidimensional Analysis of Social Biases in Vision Transformers	4891
<i>Jannik Brinkmann (University of Mannheim), Paul Swoboda (MPI fuer Informatik, Saarbruecken), and Christian Bartelt (University of Mannheim)</i>	
Partition-And-Debias: Agnostic Biases Mitigation via a Mixture of Biases-Specific Experts	4901
<i>Jiaxuan Li (The University of Tokyo), Duc Minh Vo (The University of Tokyo), and Hideki Nakayama (The University of Tokyo)</i>	

Rethinking Data Distillation: Do Not Overlook Calibration	4912
<i>Dongyao Zhu (University of California San Diego), Bowen Lei (Texas A&M University), Jie Zhang (Zhejiang University), Yanbo Fang (Certik), Yiqun Xie (The University of Maryland), Ruqi Zhang (Purdue University), and Dongkuan Xu (North Carolina State University)</i>	
Mining bias-target Alignment from Voronoi Cells	4923
<i>Remi Nahon (Telecom Paris), Van-Tam Nguyen (Telecom Paris - Institut Polytechnique de Paris), and Enzo Tartaglione (Télécom Paris - Institut Polytechnique de Paris)</i>	
Better May Not Be Fairer: A Study on Subgroup Discrepancy in Image Classification	4933
<i>Ming-Chang Chiu (University of Southern California), Pin-Yu Chen (IBM Research), and Xuezhe Ma (University of Southern California)</i>	
GIFD: A Generative Gradient Inversion Method with Feature Domain Optimization	4944
<i>Hao Fang (Harbin Institute of Technology, Shenzhen), Bin Chen (Harbin Institute of Technology, Shenzhen), Xuan Wang (Harbin Institute of Technology), Zhi Wang (Tsinghua University), and Shu-Tao Xia (Tsinghua University)</i>	
Benchmarking Algorithmic Bias in Face Recognition: An Experimental Approach Using Synthetic Faces and Human Evaluation	4954
<i>Hao Liang (Rice University), Pietro Perona (California Institute of Technology), and Guha Balakrishnan (Rice University)</i>	
FedPerfix: Towards Partial Model Personalization of Vision Transformers in Federated Learning	4965
<i>Guangyu Sun (University of Central Florida), Matias Mendieta (University of Central Florida), Jun Luo (University of Pittsburgh), Shandong Wu (Department of Radiology, Biomedical Informatics, and Bioengineering, University of Pittsburgh), and Chen Chen (University of Central Florida)</i>	
Towards Attack-tolerant Federated Learning via Critical Parameter Analysis	4976
<i>Sungwon Han (KAIST), Sungwon Park (KAIST), Fangzhao Wu (MSRA), Sundong Kim (Gwangju Institute of Science and Technology), Bin Zhu (Microsoft), Xing Xie (Microsoft Research Asia), and Meeyoung Cha (KAIST & IBS)</i>	
What can Discriminator do? Towards Box-free Ownership Verification of Generative Adversarial Networks	4986
<i>Ziheng Huang (Wuhan University), Boheng Li (Wuhan University), Yan Cai (Wuhan University), Run Wang (Wuhan University), Shangwei Guo (Chongqing University), Liming Fang (Nanjing University of Aeronautics and Astronautics), Jing Chen (Wuhan University), and Lina Wang (Wuhan University)</i>	
Robust Heterogeneous Federated Learning under Data Corruption	4997
<i>Xiuwen Fang (Wuhan University), Mang Ye (Wuhan University), and Xiyuan Yang (Wuhan University)</i>	

Communication-efficient Federated Learning with Single-Step Synthetic Features Compressor for Faster Convergence	5008
<i>Yuhao Zhou (Sichuan University), Mingjia Shi (Sichuan University), Yuanxi Li (University of Illinois at Urbana-Champaign), Yanan Sun (Sichuan university), Qing Ye (Sichuan University), and Jiancheng Lv (Sichuan University)</i>	
GPFL: Simultaneously Learning Global and Personalized Feature Information for Personalized Federated Learning	5018
<i>Jianqing Zhang (Shanghai Jiao Tong University), Yang Hua (Queen's University Belfast), Hao Wang (Louisiana State University), Tao Song (Shanghai Jiao Tong University), Zhengui Xue (Queen's University Belfast), Ruhui Ma (Shanghai Jiao Tong University), Jian Cao (Shanghai Jiao Tong University), and Haibing Guan (Shanghai Jiao Tong University)</i>	
MPCViT: Searching for Accurate and Efficient MPC-Friendly Vision Transformer with Heterogeneous Attention	5029
<i>Wenxuan Zeng (Peking University), Meng Li (Peking University), Wenjie Xiong (Virginia Tech), Tong Tong (Peking University), Wen-jie Lu (Alibaba Group), Jin Tan (Antgroup), Runsheng Wang (Peking University), and Ru Huang (Peking University)</i>	
Identification of Systematic Errors of Image Classifiers on Rare Subgroups	5041
<i>Jan Hendrik Metzen (Bosch Center for Artificial Intelligence), Robin Hutmacher (Bosch Center for Artificial Intelligence), N. Grace Hua (Robert Bosch GmbH), Valentyn Boreiko (University of Tübingen), and Dan Zhang (Bosch Center for Artificial Intelligence)</i>	
Adaptive Image Anonymization in the Context of Image Classification with Neural Networks	5051
<i>Nadiya Shvai (CYCLOPE.AI), Arcadi Llanza Carmona (LISSI), and Amir Nakib (unviersité Paris Est)</i>	
When Do Curricula Work in Federated Learning?	5061
<i>Saeed Vahidian (University of California San Diego), Sreevatsank Kadaveru (University of California San Diego), Woonjoon Baek (University of California, San Diego), Weijia Wang (University of California San Diego), Vyacheslav Kungurtsev (Czech Technical University), Chen Chen (University of Central Florida), Mubarak Shah (University of Central Florida), and Bill Lin (University of California San Diego)</i>	
Domain Specified Optimization for Deployment Authorization	5072
<i>Haotian Wang (National University of Defense Technology), Haoang Chi (National University of Defense Technology), Wenjing Yang (National University of Defense Technology), Zhipeng Lin (National University of Defense Technology), Mingyang Geng (National University of Defense Technology), Long Lan (NUDT), Jing Zhang (The University of Sydney), and Dacheng Tao (The University of Sydney)</i>	

STPrivacy: Spatio-Temporal Privacy-Preserving Action Recognition	5083
<i>Ming Li (National University of Singapore), Xiangyu Xu (Sea AI Lab), Hehe Fan (ZJU), Pan Zhou (Sea AI Lab), Jun Liu (Singapore University of Technology and Design), Jia-Wei Liu (National University of Singapore), Jiahe Li (National University of Singapore), Jussi Keppo (National University of Singapore), Mike Zheng Shou (National University of Singapore), and Shuicheng Yan (National University of Singapore, Department of Electrical and Computer Engineering)</i>	
SAL-ViT: Towards Latency Efficient Private Inference on ViT using Selective Attention	
Search with a Learnable Softmax Approximation	5093
<i>Yuke Zhang (University of Southern California), Dake Chen (University of Southern California), Souvik Kundu (Intel Labs), Chenghao Li (University of Southern California), and Peter A. Beerel (University of Southern California)</i>	
Generative Gradient Inversion via Over-Parameterized Networks in Federated Learning	5103
<i>Chi Zhang (Institute of High Performance Computing, A*STAR), Zhang Xiaoman (Astar), Ekanut Sotthiwat (IHPC), Yanyu Xu (Institute of High Performance Computing, A*Star), Ping Liu (IHPC), Liangli Zhen (Institute of High Performance Computing, A*STAR), and Yong Liu (IHPC)</i>	
Inspecting the Geographical Representativeness of Images from Text-to-Image Models	5113
<i>Abhipsa Basu (Indian Institute of Science, Bangalore), R. Venkatesh Babu (Indian Institute of Science), and Danish Pruthi (Indian Institute of Science)</i>	
Divide and Conquer: a Two-Step Method for High Quality Face De-identification with Model Explainability	5125
<i>Yunqian Wen (Shanghai Jiao Tong University), Bo Liu (University of Technology Sydney, (UTS)), Jingyi Cao (Shanghai Jiao Tong University), Rong Xie (Shanghai Jiao Tong University), and Li Song (Shanghai Jiao Tong University)</i>	
Exploring the Benefits of Visual Prompting in Differential Privacy	5135
<i>Yizhe Li (Xi'an Jiaotong University), Yu-Lin Tsai (National Yang Ming Chiao Tung University), Chia-Mu Yu (National Yang Ming Chiao Tung University), Pin-Yu Chen (IBM Research), and Xuebin Ren (Xi'an Jiaotong University)</i>	
Towards Fairness-aware Adversarial Network Pruning	5145
<i>Lei Zhang (Zhejiang University), Zhibo Wang (Zhejiang University), Xiaowei Dong (Wuhan University), Yunhe Feng (University of North Texas), Xiaoyi Pang (Wuhan University), Zhifei Zhang (Adobe Research), and Kui Ren (Zhejiang University)</i>	
AutoReP: Automatic ReLU Replacement for Fast Private Network Inference	5155
<i>Hongwu Peng (University of Connecticut), Shaoyi Huang (University of Connecticut), Tong Zhou (Northeastern University), Yukui Luo (Northeastern University), Chenghong Wang (Duke University), Zigeng Wang (University of Connecticut), Jiahui Zhao (University of Connecticut), Xi Xie (University of Connecticut), Ang Li (Pacific Northwest National Laboratory), Tony Geng (University of Rochester), Kaleel Mahmood (University of Connecticut), Wujie Wen (NCSU), Xiaolin Xu (Northeastern University), and Caiwen Ding (University of Connecticut)</i>	

Flatness-Aware Minimization for Domain Generalization	5166
<i>Xingxuan Zhang (Tsinghua University), Renzhe Xu (Tsinghua University), Han Yu (Tsinghua University), Yancheng Dong (Tsinghua University), Pengfei Tian (Tsinghua University), and Peng Cui (Tsinghua University)</i>	
Communication-Efficient Vertical Federated Learning with Limited Overlapping Samples	5180
<i>Jingwei Sun (Duke University), Ziyue Xu (NVIDIA), Dong Yang (NVIDIA Corporation), Vishwesh Nath (NVIDIA), Wenqi Li (NVIDIA), Can Zhao (Nvidia), Daguang Xu (NVIDIA Corporation), Yiran Chen (Duke University), and Holger R. Roth (NVIDIA)</i>	

First Person (Egocentric) Vision

Multimodal Distillation for Egocentric Action Recognition	5190
<i>Gorjan Radevski (KU Leuven), Dusan Grujicic (KU Leuven), Matthew Blaschko (KU Leuven), Marie-Francine Moens (KU Leuven), and Tinne Tuytelaars (KU Leuven)</i>	
Self-Supervised Object Detection from Egocentric Videos	5202
<i>Peri Akiwa (Rutgers University), Jing Huang (Facebook), Kevin J Liang (Meta AI), Rama Kovvuri (Facebook), Xingyu Chen (Facebook AI), Matt Feiszli (FAIR at Meta), Kristin Dana (Rutgers University), and Tal Hassner (Facebook AI)</i>	
Multi-label Affordance Mapping from Egocentric Vision	5215
<i>Lorenzo Mur-Labadia (University of Zaragoza), Jose J. Guerrero (Universidad de Zaragoza), and Ruben Martinez-Cantin (University of Zaragoza)</i>	
Ego-Only: Egocentric Action Detection without Exocentric Transferring	5227
<i>Huiyu Wang (Meta AI), Mitesh Kumar Singh (Meta AI), and Lorenzo Torresani (Meta AI)</i>	
COPILOT: Human-Environment Collision Prediction and Localization from Egocentric Videos ...	5239
<i>Boxiao Pan (Stanford University), Bokui Shen (Stanford), Davis Rempe (Stanford University), Despoina Paschalidou (Stanford), Kaichun Mo (NVIDIA), Yanchao Yang (Stanford University), and Leonidas J. Guibas (Stanford University)</i>	
EgoPCA: A New Framework for Egocentric Hand-Object Interaction Understanding	5250
<i>Yue Xu (Shanghai Jiao Tong University), Yong-Lu Li (Shanghai Jiao Tong University), Zhemin Huang (SJTU), Michael Xu Liu (New Hope Investment Group), Cewu Lu (Shanghai Jiao Tong University), Yu-Wing Tai (Kuaishou Technology / HKUST), and Chi-Keung Tang (Hong Kong University of Science and Technology)</i>	
EgoVLPv2: Egocentric Video-Language Pre-training with Fusion in the Backbone	5262
<i>Shraman Pramanick (Johns Hopkins University), Yale Song (Facebook AI Research), Sayan Nag (University of Toronto), Kevin Qinghong Lin (National University of Singapore), Hardik Shah (Meta Inc), Mike Zheng Shou (National University of Singapore), Rama Chellappa (Johns Hopkins University), and Pengchuan Zhang (Microsoft Research AI)</i>	

Representation Learning

WDiscOOD: Out-of-Distribution Detection via Whitened Linear Discriminant Analysis	5275
<i>Yiye Chen (Georgia Institute of Technology), Yunzhi Lin (Georgia Institute of Technology), Ruinian Xu (Georgia Institute of Technology), and Patricio A. Vela (Georgia Institute of Technology)</i>	
Pairwise Similarity Learning is SimPLE	5285
<i>Yandong Wen (Max Planck Institute for Intelligent Systems), Weiyang Liu (University of Cambridge), Yao Feng (Max Planck Institute for Intelligent Systems), Bhiksha Raj (Carnegie Mellon University), Rita Singh (Carnegie Mellon University), Adrian Weller (University of Cambridge), Michael J. Black (Max Planck Institute for Intelligent Systems), and Bernhard Schölkopf (MPI for Intelligent Systems, Tübingen)</i>	
No Fear of Classifier Biases: Neural Collapse Inspired Federated Learning with Synthetic and Fixed Classifier	5296
<i>Zexi Li (Zhejiang University), Xinyi Shang (Xiamen University), Rui He (Zhejiang University), Tao Lin (Westlake University), and Chao Wu (Zhejiang University)</i>	
Generalizable Neural Fields as Partially Observed Neural Processes	5307
<i>Jeffrey Gu (Stanford University), Kuan-Chieh Wang (Stanford University), and Serena Yeung (Stanford University)</i>	
M2T: Masking Transformers Twice for Faster Decoding	5317
<i>Fabian Mentzer (Google), Eirikur Agustson (Google), and Michael Tschannen (Google)</i>	
Keep It SimPool: Who Said Supervised Transformers Suffer from Attention Deficit?	5327
<i>Bill Psomas (National Technical University of Athens), Ioannis Kakogeorgiou (National Technical University of Athens), Konstantinos Karantzalos (National Technical University of Athens), and Yannis Avrithis (IARAI)</i>	
Improving Pixel-based MIM by Reducing Wasted Modeling Capability	5338
<i>Yuan Liu (Shanghai AI Laboratory), Songyang Zhang (Shanghai AI Laboratory), Jiacheng Chen (Simon Fraser University), Zhaohui Yu (Shanghai AI Laboratory), Kai Chen (Shanghai AI Laboratory), and Dahua Lin (Shanghai AI Laboratory; The Chinese University of Hong Kong)</i>	
Learning Image-Adaptive Codebooks for Class-Agnostic Image Restoration	5350
<i>Kechun Liu (University of Washington), Yitong Jiang (The Chinese University of Hong Kong), Inchang Choi (SenseBrain), and Jinwei Gu (The Chinese University of Hong Kong)</i>	
Quality Diversity for Visual Pre-Training	5361
<i>Ruchika Chohan (University of Edinburgh), Henry Gouk (University of Edinburgh), Da Li (Samsung), and Timothy Hospedales (Edinburgh University)</i>	
Subclass-balancing Contrastive Learning for Long-tailed Recognition	5372
<i>Chengkai Hou (Jilin university), Jieyu Zhang (University of Washington), Haonan Wang (National University of Singapore), and Tianyi Zhou (University of Maryland, College Park)</i>	

Mastering Spatial Graph Prediction of Road Networks	5385
<i>Anagnostidis Sotirios (ETH Zurich), Aurelien Lucchi (ETH Zurich), and Thomas Hofmann (ETH Zurich)</i>	
Poincaré ResNet	5396
<i>Max van Spengler (University of Amsterdam), Erwin Berkhout (Academic Center for Dentistry Amsterdam, (ACTA), and Pascal Mettes (University of Amsterdam)</i>	
Exploring Model Transferability through the Lens of Potential Energy	5406
<i>Xiaotong Li (Peking University), Zixuan Hu (Peking University), Yixiao Ge (Tencent), Ying Shan (Tencent), and Ling-Yu Duan (Peking University)</i>	
Improving CLIP Fine-tuning Performance	5416
<i>Yixuan Wei (Tsinghua University), Han Hu (Microsoft Research Asia), Zhenda Xie (Tsinghua University), Ze Liu (USTC), Zheng Zhang (MSRA), Yue Cao (Beijing Academy of Artificial Intelligence), Jianmin Bao (Microsoft Research Asia), Dong Chen (Microsoft Research Asia), and Baining Guo (MSR Asia)</i>	
Unsupervised Manifold Linearizing and Clustering	5427
<i>Tianjiao Ding (Johns Hopkins University), Shengbang Tong (UC Berkeley), Kwan Ho Ryan Chan (Johns Hopkins University), Xili Dai (University of California, Berkeley), Yi Ma (UC Berkeley), and Benjamin D. Haeffele (Johns Hopkins University)</i>	
Generalized Sum Pooling for Metric Learning	5439
<i>Yeti Z. Gürbüz (Technische Universität Berlin), Ozan Sener (Intel Labs), and A. Aydin Alatan (Middle East Technical University, Turkey)</i>	
Partition Speeds Up Learning Implicit Neural Representations Based on Exponential-Increase Hypothesis	5451
<i>Ke Liu (Zhejiang University), Feng Liu (UTS/RIKEN), Haishuai Wang (Zhejiang University), Ning Ma (Zhejiang University), Jiajun Bu (Zhejiang University), and Bo Han (HKBU / RIKEN)</i>	
The Effectiveness of MAE Pre-Pretraining for Billion-Scale Pretraining	5461
<i>Mannat Singh (Meta AI), Quentin Duval (Meta AI), Kalyan Vasudev Alwala (Meta AI), Haoqi Fan (Meta AI), Vaibhav Aggarwal (Meta AI), Aaron Adcock (Meta AI), Armand Joulin (Meta AI), Piotr Dollar (Meta AI), Christoph Feichtenhofer (Meta AI), Ross Girshick (Meta AI), Rohit Girdhar (Meta AI), and Ishan Misra (Meta AI)</i>	
Token-Label Alignment for Vision Transformers	5472
<i>Han Xiao (Tsinghua University), Wenzhao Zheng (Tsinghua University), Zheng Zhu (Tsinghua University), Jie Zhou (Tsinghua University), and Jiwen Lu (Tsinghua University)</i>	
Efficiently Robustify Pre-Trained Models	5482
<i>Nishant Jain (iit roorkee), Harkirat Singh Behl (Microsoft Research Ltd), Yogesh Singh Rawat (University of Central Florida), and Vibhav Vineet (Microsoft Research)</i>	

OFVL-MS: Once for Visual Localization across Multiple Indoor Scenes	5493
<i>Tao Xie (Harbin Institute of Technology), Kun Dai (Harbin Institute of Technology), Siyi Lu (China Coal Technology and Engineering Group Intelligent Storage and Loading Technology Co., Ltd.), Ke Wang (Harbin Institute of Technology), Zhiqiang Jiang (Harbin Institute of Technology), Jinghan Gao (Harbin Institute of Technology), Dedong Liu (Harbin Institute of Technology), Jie Xu (Harbin Institute of Technology), Lijun Zhao (Harbin Institute of Technology), and Ruifeng Li (Harbin Institute of Technology)</i>	
Feature Prediction Diffusion Model for Video Anomaly Detection	5504
<i>Cheng Yan (Tianjin University), Shiyu Zhang (Tianjin University), Yang Liu (Zhejiang University), Guansong Pang (Singapore Management University), and Wenjun Wang (Tianjin University)</i>	
Joint Implicit Neural Representation for High-fidelity and Compact Vector Fonts	5515
<i>Chia-Hao Chen (Tsinghua University), Ying-Tian Liu (Tsinghua University), Zhifei Zhang (Adobe Research), Yuan-Chen Guo (Tsinghua University), and Song-Hai Zhang (Tsinghua University)</i>	
How Far Pre-trained Models Are from Neural Collapse on the Target Dataset Informs their Transferability	5526
<i>Zijian Wang (University of Queensland), Yadan Luo (University of Queensland), Liang Zheng (Australian National University), Zi Huang (University of Queensland), and Mahsa Baktashmotlagh (University of Queensland)</i>	
OPERA: Omni-Supervised Representation Learning with Hierarchical Supervisions	5536
<i>Chengkun Wang (Tsinghua University), Wenzhao Zheng (Tsinghua University), Zheng Zhu (Tsinghua University), Jie Zhou (Tsinghua University), and Jiwon Lu (Tsinghua University)</i>	
Perceptual Grouping in Contrastive Vision-Language Models	5548
<i>Kanchana Ranasinghe (Stony Brook University), Brandon McKinzie (Apple), Sachin Ravi (Princeton University), Yinfei Yang (Apple), Alexander Toshev (Apple), and Jonathon Shlens (Google)</i>	
Fully Attentional Networks with Self-emerging Token Labeling	5562
<i>Bingyin Zhao (Clemson University), Zhiding Yu (NVIDIA), Shiyi Lan (NVIDIA CORPORATION), Yutao Cheng (Fudan university), Anima Anandkumar (Caltech), Yingjie Lao (Clemson University), and Jose M. Alvarez (NVIDIA)</i>	
Instance and Category Supervision are Alternate Learners for Continual Learning	5573
<i>Xudong Tian (East China Normal University), Zhizhong Zhang (East China Normal University), Xin Tan (East China Normal University), Jun Liu (Tencent), Chengjie Wang (Tencent; Shanghai Jiao Tong University), Yanyun Qu (XMU), Guannan Jiang (CATL), and Yuan Xie (East China Normal University)</i>	
SkeletonMAE: Graph-based Masked Autoencoder for Skeleton Sequence Pre-training	5583
<i>Hong Yan (Sun Yat-sen University), Yang Liu (Sun Yat-sen University), Yushen Wei (Sun Yet-sen university), Zhen Li (The Chinese University of Hong Kong, Shenzhen), Guanbin Li (Sun Yat-sen University), and Liang Lin (Sun Yat-sen University)</i>	

Motion-Guided Masking for Spatiotemporal Representation Learning	5596
<i>David Fan (Amazon), Jue Wang (Amazon), Shuai Liao (Amazon), Yi Zhu (Amazon), Vimal Bhat (Amazon), Hector Santos-Villalobos (Amazon Inc), Rohith MV (Amazon Lab126), and Xinyu Li (Amazon)</i>	
Data Augmented Flatness-aware Gradient Projection for Continual Learning	5607
<i>Enneng Yang (Northeastern University), Li Shen (JD Explore Academy), Zhenyi Wang (University of Maryland, College Park), Shiwei Liu (UT Austin), Guibing Guo (Northeastern University), and Xingwei Wang (Northeastern University)</i>	
Take-A-Photo: 3D-to-2D Generative Pre-training of Point Cloud Models	5617
<i>Ziyi Wang (Tsinghua University), Xumin Yu (Tsinghua University), Yongming Rao (Tsinghua University), Jie Zhou (Tsinghua University), and Jiwen Lu (Tsinghua University)</i>	
BiViT: Extremely Compressed Binary Vision Transformers	5628
<i>Yefei He (Zhejiang University), Zhenyu Lou (Zhejiang University), Luoming Zhang (Zhejiang University), Jing Liu (Monash University), Weijia Wu (Zhejiang University), Hong Zhou (Zhejiang University), and Bohan Zhuang (Monash University)</i>	
Spatio-Temporal Crop Aggregation for Video Representation Learning	5641
<i>Sepehr Sameni (University of Bern), Simon Jenni (Adobe Research), and Paolo Favaro (University of Bern)</i>	
Hierarchical Visual Primitive Experts for Compositional Zero-Shot Learning	5652
<i>Hanjae Kim (Yonsei Univ.), Jiyoung Lee (NAVER AI Lab), Seongheon Park (Yonsei Univ.), and Kwanghoon Sohn (Yonsei Univ.)</i>	
Semantic Information in Contrastive Learning	5663
<i>Shengjiang Quan (The University of Tokyo), Masahiro Hirano (The University of Tokyo), and Yuji Yamakawa (The University of Tokyo)</i>	
Cross-Domain Product Representation Learning for Rich-Content E-Commerce	5674
<i>Xuehan Bai (kuaishou), Yan Li (Kuaishou), Yanhua Cheng (Kuaishou), Wenjie Yang (Kuaishou), Quan Chen (Kuaishou), and Han Li (Kuaishou)</i>	
Contrastive Continuity on Augmentation Stability Rehearsal for Continual Self-Supervised Learning	5684
<i>Haoyang Cheng (University of Electronic Science and Technology of China), Haitao Wen (University of Electronic Science and Technology of China), Xiaoliang Zhang (University of Electronic Science and Technology of China), Heqian Qiu (University of Electronic Science and Technology of China), Lanxiao Wang (University of Electronic Science and Technology of China), and Hongliang Li (University of Electronic Science and Technology of China)</i>	
HybridAugment++: Unified Frequency Spectra Perturbations for Model Robustness	5695
<i>Mehmet Kerim Yucel (Samsung R&D UK), Ramazan Gokberk Cinbis (METU), and Pinar Duygulu (Hacettepe University)</i>	
Unleashing Text-to-Image Diffusion Models for Visual Perception	5706
<i>Wenliang Zhao (Tsinghua University), Yongming Rao (Tsinghua University), Zuyan Liu (Tsinghua University), Benlin Liu (University of Washington), Jie Zhou (Tsinghua University), and Jiwen Lu (Tsinghua University)</i>	

Deep Learning Architectures

Efficient Controllable Multi-Task Architectures	5717
<i>Abhishek Aich (University of California, Riverside; NEC Labs America), Samuel Schuster (NEC Labs America), Amit K. Roy-Chowdhury (University of California, Riverside), Manmohan Chandraker (NEC Labs America; University of California, San Diego), and Yumin Suh (NEC Labs America)</i>	
ParCNetV2: Oversized Kernel with Enhanced Attention	5729
<i>Ruihan Xu (Peking University), Haokui Zhang (Lighthouse Co.Ltd), Wenzhe Hu (Lighthouse Co.Ltd), Shiliang Zhang (Peking University), and Xiaoyu Wang (Hong Kong University of Science and Technology, (HKUST)</i>	
Unleashing the Power of Gradient Signal-to-Noise Ratio for Zero-Shot NAS	5740
<i>Zihao Sun (Institute of Computing Technology, Chinese Academy of Sciences), Yu Sun (Academy of Mathematics and Systems Science, Chinese Academy of Sciences), Longxing Yang (Institute of Computing Technology, Chinese Academy of Sciences), Shun Lu (Institute of Computing Technology, Chinese Academy of Sciences), Jilin Mei (Institute of Computing Technology, Chinese Academy of Sciences), Wenxiao Zhao (Academy of Mathematics and Systems Science, Chinese Academy of Sciences), and Yu Hu (Institute of Computing Technology, Chinese Academy of Sciences)</i>	
MMST-ViT: Climate Change-aware Crop Yield Prediction via Multi-Modal Spatial-Temporal Vision Transformer	5751
<i>Fudong Lin (University of Delaware), Summer Crawford (University of Louisiana at Lafayette), Kaleb Guillot (University of Louisiana at Lafayette), Yihe Zhang (University of Louisiana at Lafayette), Yan Chen (University of Connecticut), Xu Yuan (University of Delaware), Li Chen (University of Louisiana at Lafayette), Shelby Williams (University of Louisiana at Lafayette), Robert Minvielle (University of Louisiana at Lafayette), Xiangming Xiao (University of Oklahoma), Drew Gholson (Mississippi State University), Nicolas Ashwell (Mississippi State University), Tri Setiyono (Louisiana State University), Brenda Tubana (LSU AgCenter), Lu Peng (Tulane University), Magdy Bayoumi (University of Louisiana at Lafayette), and Nian-Feng Tzeng (University of Louisiana at Lafayette)</i>	
FastViT: A Fast Hybrid Vision Transformer Using Structural Reparameterization	5762
<i>Pavan Kumar Anasosalu Vasu (Apple Inc.), James Gabriel (Apple.inc), Jeff Zhu (Apple, Inc.), Oncel Tuzel (Apple), and Anurag Ranjan (Apple)</i>	
IIEU: Rethinking Neural Feature Activation from Decision-Making	5773
<i>Sudong Cai (Kyoto University)</i>	
Scratching Visual Transformer's Back with Uniform Attention	5784
<i>Nam Hyeon-Woo (POSTECH), Kim Yu-Ji (POSTECH), Byeongho Heo (NAVER AI LAB), Dongyoon Han (NAVER AI Lab), Seong Joon Oh (Tübingen AI Center), and Tae-Hyun Oh (POSTECH)</i>	

SpaceEvo: Hardware-Friendly Search Space Design for Efficient INT8 Inference	5796
<i>Xudong Wang (Shanghai Jiao Tong University), Li Lyna Zhang (Microsoft Research), Jiahang Xu (Microsoft), Quanlu Zhang (Microsoft), Yujing Wang (MSRA), Yuqing Yang (Microsoft Research), Ningxin Zheng (Microsoft Research), Ting Cao (Microsoft Research), and Mao Yang (MSRA)</i>	
ElasticViT: Conflict-aware Supernet Training for Deploying Fast Vision Transformer on Diverse Mobile Devices	5806
<i>Chen Tang (Tsinghua University), Li Lyna Zhang (Microsoft Research), Huiqiang Jiang (Microsoft Research Asia), Jiahang Xu (Microsoft), Ting Cao (Microsoft Research), Quanlu Zhang (Microsoft), Yuqing Yang (Microsoft Research), Zhi Wang (Tsinghua University), and Mao Yang (MSRA)</i>	
Gramian Attention Heads are Strong yet Efficient Vision Learners	5818
<i>Jongbin Ryu (Ajou University), Dongyoon Han (NAVER AI Lab), and Jongwoo Lim (Hanyang University)</i>	
EfficientTrain: Exploring Generalized Curriculum Learning for Training Visual Backbones	5829
<i>Yulin Wang (Tsinghua University), Yang Yue (Tsinghua University), Rui Lu (Tsinghua University), Tianjiao Liu (Huawei), Zhao Zhong (HUAWEI), Shiji Song (Department of Automation, Tsinghua University), and Gao Huang (Tsinghua)</i>	
Ord2Seq: Regarding Ordinal Regression as Label Sequence Prediction	5842
<i>Jinhong Wang (Zhejiang University), Yi Cheng (Zhejiang University), Jintai Chen (UIUC), TingTing Chen (Zhejiang University), Danny Chen (University of Notre Dame), and Jian Wu (Zhejiang University)</i>	
Unified Data-Free Compression: Pruning and Quantization without Fine-Tuning	5853
<i>Shipeng Bai (Zhejiang University), Jun Chen (Zhejiang University), Xintian Shen (Zhejiang university), Yixuan Qian (Zhejiang University), and Yong Liu (Zhejiang University)</i>	
LaPE: Layer-adaptive Position Embedding for Vision Transformers with Independent Layer Normalization	5863
<i>Runyi Yu (Peking University), Zhennan Wang (Peng Cheng Laboratory), Yinhuai Wang (Peking University Shenzhen Graduate School), Kehan Li (Peking University), Chang Liu (Tsinghua University), Haoyi Duan (Zhejiang University), Xiangyang Ji (Tsinghua University), and Jie Chen (Peking University)</i>	
Exemplar-Free Continual Transformer with Convolutions	5874
<i>Anurag Roy (IIT Kharagpur), Vinay K. Verma (Amazon), Sravan Voonna (Microsoft), Kripabandhu Ghosh (Indian Institute of Science Education and Research Kolkata), Saptarshi Ghosh (Indian Institute of Technology Kharagpur), and Abir Das (IIT Kharagpur)</i>	
Building Vision Transformers with Hierarchy Aware Feature Aggregation	5885
<i>Yongjie Chen (University of Science and Technology Beijing), Hongmin Liu (School of Automation and Electrical Engineering, University of Science and Technology Beijing), Haoran Yin (Horizon Robotics), and Bin Fan (University of Science and Technology Beijing)</i>	

ShiftNAS: Improving One-shot NAS via Probability Shift	5896
<i>Mingyang Zhang (Zhejiang University of Technology), Xinyi Yu (Zhejiang University of Technology), Haodong Zhao (Zhejiang university of technology), and Linlin Ou (Zhejiang University of Technology)</i>	
DarSwin: Distortion Aware Radial Swin Transformer	5906
<i>Akshaya Athwale (University Laval), Arman Afrasiyabi (Yale University), Justin Lagüe (Université Laval), Ichrak Shili (Université Laval), Ola Ahmad (Thales Digital Solutions), and Jean-François Lalonde (Université Laval)</i>	
ROME: Robustifying Memory-Efficient NAS via Topology Disentanglement and Gradient Accumulation	5916
<i>Xiaoxing Wang (Shanghai Jiao Tong University), Xiangxiang Chu (Meituan), Yuda Fan (ETH Zurich), Zhexi Zhang (Shanghai Jiao Tong University), Bo Zhang (Meituan), Xiaokang Yang (Shanghai Jiao Tong University of China), and Junchi Yan (Shanghai Jiao Tong University)</i>	
FDViT: Improve the Hierarchical Architecture of Vision Transformer	5927
<i>Yixing Xu (AMD), Chao Li (AMD), Dong Li (Xilinx), Xiao Sheng (AMD), Fan Jiang (AMD), Lu Tian (AMD), and Ashish Sirasao (AMD Inc)</i>	
FLatten Transformer: Vision Transformer using Focused Linear Attention	5938
<i>Dongchen Han (Tsinghua University), Xuran Pan (Tsinghua University), Yizeng Han (Tsinghua University), Shiji Song (Department of Automation, Tsinghua University), and Gao Huang (Tsinghua)</i>	
MixPath: A Unified Approach for One-shot Neural Architecture Search	5949
<i>Xiangxiang Chu (Meituan), Shun Lu (Institute of Computing Technology, Chinese Academy of Sciences), Xudong Li (University of Chinese Academy of Sciences), and Bo Zhang (Meituan)</i>	
SSF: Accelerating Training of Spiking Neural Networks with Stabilized Spiking Flow	5959
<i>Wang Jingtao (CASIA), Zengjie Song (Institute of Automation, Chinese Academy of Sciences), Yuxi Wang (CASIA), Jun Xiao (University of Chinese Academy of Sciences), Yuran Yang (Tencent), Shuqi Mei (Tencent), and Zhaoxiang Zhang (Chinese Academy of Sciences, China)</i>	
Dynamic Perceiver for Efficient Visual Recognition	5969
<i>Yizeng Han (Tsinghua University), Dongchen Han (Tsinghua University), Zeyu Liu (Tsinghua University), Yulin Wang (Tsinghua University), Xuran Pan (Tsinghua University), Yifan Pu (Tsinghua University), Chao Deng (China Mobile Research Institute), Junlan Feng (China Mobile Research), Shiji Song (Department of Automation, Tsinghua University), and Gao Huang (Tsinghua)</i>	
SG-Former: Self-guided Transformer with Evolving Token Reallocation	5980
<i>Sucheng Ren (South China University of Technology), Xingyi Yang (National University of Singapore), Songhua Liu (National University of Singapore), and Xinchao Wang (National University of Singapore)</i>	
Scale-Aware Modulation Meet Transformer	5992
<i>Weifeng Lin (South China University of Technology), Ziheng Wu (Alibaba Group), Jiayu Chen (alibaba), Jun Huang (Alibaba Group), and Lianwen Jin (South China University of Technology)</i>	

Learning to Upsample by Learning to Sample	6004
<i>Wenze Liu (Huazhong University of Science and Technology), Hao Lu (Huazhong University of Science and Technology), Hongtao Fu (Huazhong university of Science and Technology), and Zhiguo Cao (Huazhong Univ. of Sci.&Tech.)</i>	
GET: Group Event Transformer for Event-Based Vision	6015
<i>Yansong Peng (University of Science and Technology of China), Yueyi Zhang (University of Science and Technology of China), Zhiwei Xiong (University of Science and Technology of China), Xiaoyan Sun (University of Science and Technology of China), and Feng Wu (University of Science and Technology of China)</i>	
Adaptive Frequency Filters As Efficient Global Token Mixers	6026
<i>Zhipeng Huang (University of Science and Technology of China), Zhizheng Zhang (Microsoft Research), Cuiling Lan (Microsoft Research), Zheng-Jun Zha (University of Science and Technology of China), Yan Lu (Microsoft Research Asia), and Baining Guo (MSR Asia)</i>	
Fcaformer: Forward Cross Attention in Hybrid Vision Transformer	6037
<i>Haokui Zhang (Lighthouse Co.Ltd), Wenze Hu (Lighthouse Co.Ltd), and Xiaoyu Wang (Hong Kong University of Science and Technology, (HKUST)</i>	
Dynamic Snake Convolution Based on Topological Geometric Constraints for Tubular Structure Segmentation	6047
<i>Yaolei Qi (Southeast University), Yuting He (Southeast University), Xiaoming Qi (Southeast University), Yuan Zhang (Southeast University), and Guanyu Yang (Southeast University)</i>	
Sentence Attention Blocks for Answer Grounding	6057
<i>Seyedalireza Khoshirat (University of Delaware) and Chandra Kambhamettu (University of Delaware)</i>	
MST-compression: Compressing and Accelerating Binary Neural Networks with Minimum Spanning Tree	6068
<i>Quang Hieu Vo (Kyung Hee University), Linh-Tam Tran (Kyung Hee University), Sung-Ho Bae (Kyung Hee University), Lok-Won Kim (KyungHee University), and Choong Seon Hong (Kyung Hee University)</i>	
EGformer: Equirectangular Geometry-biased Transformer for 360 Depth Estimation	6078
<i>Ilwi Yun (Seoul National University), Chanyong Shin (Inha University), Hyunku Lee (Seoul national university), Hyuk-Jae Lee (SNU), and Chae Eun Rhee (Inha University)</i>	
SPANet: Frequency-balancing Token Mixer using Spectral Pooling Aggregation Modulation	6090
<i>Guhnoo Yun (Korea Institute of Science and Technology), Juhan Yoo (semyung university), Kijung Kim (Korea University), Jeongho Lee (Korean Institute of Science and Technology), and Dong Hwan Kim (Korea Institute of Science and Technology)</i>	
ModelGiF: Gradient Fields for Model Functional Distance	6102
<i>Jie Song (Zhejiang University), Zhengqi Xu (Zhejiang University), Sai Wu (Zhejiang Univ), Gang Chen (Zhejiang University), and Mingli Song (Zhejiang University)</i>	

ClusT3: Information Invariant Test-Time Training	6113
<i>Gustavo A. Vargas Vargas Hakim (École de Technologie Supérieure), David Osowiecki (Ecole de Technologie Supérieure), Mehrdad Noori (École de Technologie Supérieure), Milad Cheraghlikhani (École de Technologie Supérieure), Ali Bahri (École de technologie supérieure ÉTS), Ismail Ben Ayed (ETS Montreal), and Christian Desrosiers (ETS, Canada)</i>	
Cumulative Spatial Knowledge Distillation for Vision Transformers	6123
<i>Borui Zhao (MEGVII Technology), Renjie Song (MEGVII Technology), and Jiajun Liang (MEGVII Technology)</i>	
Luminance-aware Color Transform for Multiple Exposure Correction	6133
<i>Jong-Hyeon Baek (ChungNam National University), DaeHyun Kim (Chungnam National University), Su-Min Choi (42dot), Hyo-jun Lee (ChungNam National university), Hanul Kim (Seoul National University of Science and Technology), and Yeong Jun Koh (Chungnam National University)</i>	
Towards Memory- and Time-Efficient Backpropagation for Training Spiking Neural Networks ...	6143
<i>Qingyan Meng (The Chinese University of Hong Kong, Shenzhen and Shenzhen Research Institute of Big Data), Mingqing Xiao (Peking University), Shen Yan (Peking University), Yisen Wang (Peking University), Zhouchen Lin (Peking University), and Zhi-Quan Luo (The Chinese University of Hong Kong, Shenzhen and Shenzhen Research Institute of Big Data)</i>	
Domain Generalization Guided by Gradient Signal to Noise Ratio of Parameters	6154
<i>Mateusz Michalkiewicz (University of Queensland), Masoud Faraki (NEC Labs), Xiang Yu (NEC Labs), Manmohan Chandraker (UC San Diego), and Mahsa Baktashmotlagh (University of Queensland)</i>	
DOT: A Distillation-Oriented Trainer	6166
<i>Borui Zhao (MEGVII Technology), Quan Cui (Waseda University), Renjie Song (MEGVII Technology), and Jiajun Liang (MEGVII Technology)</i>	
Extensible and Efficient Proxy for Neural Architecture Search	6176
<i>Yuhong Li (University of Illinois at Urbana and Champaign), Jiajie Li (University at Buffalo), Cong Hao (Georgia Institute of Technology), Pan Li (Georgia Tech.), Jinjun Xiong (University at Buffalo), and Deming Chen (University of Illinois at Urbana-Champaign)</i>	
Learning to Transform for Generalizable Instance-wise Invariance	6188
<i>Utkarsh Singhal (UC Berkeley), Carlos Esteves (Google Research), Ameesh Makadia (Google Research), and Stella X. Yu (University of Michigan)</i>	
Convolutional Networks with Oriented 1D Kernels	6199
<i>Alexandre Kirchmeyer (Carnegie Mellon University) and Jia Deng (Princeton University)</i>	

Recognition: Detection

Random Boxes Are Open-world Object Detectors	6210
<i>Yanghao Wang (Nanyang Technological University), Zhongqi Yue (Nanyang Technological University), Xian-Sheng Hua (Terminus Group), and Hanwang Zhang (Nanyang Technological University)</i>	

Unleashing Vanilla Vision Transformer with Masked Image Modeling for Object Detection	6221
<i>Yuxin Fang (Huazhong University of Science and Technology), Shusheng Yang (Huazhong University of Science and Technology), Shijie Wang (Huazhong University of Science and Technology), Yixiao Ge (Tencent), Ying Shan (Tencent), and Xinggang Wang (Huazhong University of Science and Technology)</i>	
CoIn: Contrastive Instance Feature Mining for Outdoor 3D Object Detection with Very Limited Annotations	6231
<i>Qiming Xia (Xiamen University), Jinhao Deng (Xiamen University), Chenglu Wen (Xiamen University), Hai Wu (Xiamen University), Shaoshuai Shi (MPI Informatics), Xin Li (Texas A&M University), and Cheng Wang (Xiamen University)</i>	
A Dynamic Dual-Processing Object Detection Framework Inspired by the Brain's Recognition Mechanism	6241
<i>Minying Zhang (Alibaba Group), Tianpeng Bu (Alibaba Group), and Lulu Hu (Alibaba Group)</i>	
Anchor-Intermediate Detector: Decoupling and Coupling Bounding Boxes for Accurate Object Detection	6252
<i>Yilong Lv (Xi'an Institute of High Technology), Min Li (Xi'an High-Tech Research Institute), Yujie He (EPGC), Shaopeng Li (Tsinghua University, Department of Automation), Zhuzhen He (National University of Defense Technology), and Aitao Yang (Xi'an Institute of High Technology)</i>	
Inter-Realization Channels: Unsupervised Anomaly Detection Beyond One-Class Classification...	6262
<i>Declan McIntosh (University Of Victoria) and Alexandra Branzan Albu (University of Victoria)</i>	
Deep Equilibrium Object Detection	6273
<i>Shuai Wang (nanjing university), Yao Teng (Nanjing University), and Limin Wang (Nanjing University)</i>	
RecursiveDet: End-to-End Region-Based Recursive Object Detection	6284
<i>Jing Zhao (East China Normal University), Li Sun (East China Normal University), and Qingli Li (East China Normal University)</i>	
Small Object Detection via Coarse-to-fine Proposal Generation and Imitation Learning	6294
<i>Xiang Yuan (Northwestern Polytechnical University), Gong Cheng (Northwestern Polytechnical University), Keping Yan (Northwestern Polytechnical University), Qinghua Zeng (Northwestern Polytechnical University), and Junwei Han (Northwestern Polytechnical University)</i>	
ASAG: Building Strong One-Decoder-Layer Sparse Detectors via Adaptive Sparse Anchor Generation	6305
<i>Shenghao Fu (Sun Yat-sen University), Junkai Yan (Sun Yat-sen University), Yipeng Gao (Sun Yat-sen University, China), Xiaohua Xie (Sun Yat-sen University), and Wei-Shi Zheng (Sun Yat-sen University, China)</i>	
COCO-O: A Benchmark for Object Detectors under Natural Distribution Shifts	6316
<i>Xiaofeng Mao (Alibaba Group), Yuefeng Chen (Alibaba Group), Yao Zhu (Zhejiang University), Da Chen (University of Bath), Hang Su (Tsinghua University), Rong Zhang (Alibaba), and Hui Xue (Alibaba)</i>	

Generative Prompt Model for Weakly Supervised Object Localization	6328
<i>Yuzhong Zhao (University of Chinese Academy of Sciences), Qixiang Ye (University of Chinese Academy of Sciences, China), Weijia Wu (Zhejiang University), Chunhua Shen (University of Adelaide, Australia), and Fang Wan (University of Chinese Academy of Sciences)</i>	
UniKD: Universal Knowledge Distillation for Mimicking Homogeneous or Heterogeneous Object Detectors	6339
<i>Shanshan Lao (Tsinghua University), Guanglu Song (Sensetime), Boxiao Liu (Institute of Computing Technology, Chinese Academy of Sciences), Yu Liu (SenseTime Group LTD), and Yujiu Yang (Tsinghua University)</i>	
PNI : Industrial Anomaly Detection using Position and Neighborhood Information	6350
<i>Jaehyeok Bae (Seoul National University), Jae-Han Lee (Gauss Labs Inc.), and Seyun Kim (Gauss Labs)</i>	
Masked Autoencoders Are Stronger Knowledge Distillers	6361
<i>Shanshan Lao (Tsinghua University), Guanglu Song (Sensetime), Boxiao Liu (Institute of Computing Technology, Chinese Academy of Sciences), Yu Liu (SenseTime Group LTD), and Yujiu Yang (Tsinghua University)</i>	
GPA-3D: Geometry-aware Prototype Alignment for Unsupervised Domain Adaptive 3D Object Detection from Point Clouds	6371
<i>Ziyu Li (Southeast University), Jingming Guo (Huawei Noah's Ark Lab), Tongtong Cao (Huawei Technologies Canada), Liu Bingbing (Huawei Noah's Ark Lab, Canada), and Wankou Yang (Southeast University)</i>	
ADNet: Lane Shape Prediction via Anchor Decomposition	6381
<i>Lingyu Xiao (Southeast University), Xiang Li (Nankai University), Sen Yang (Southeast University), and Wankou Yang (Southeast University)</i>	
Periodically Exchange Teacher-Student for Source-Free Object Detection	6391
<i>Qipeng Liu (FuZhou University), LuoJun Lin (Fuzhou University), Zhifeng Shen (Fuzhou University), and Zhifeng Yang (FuZhou University)</i>	
Towards Fair and Comprehensive Comparisons for Image-Based 3D Object Detection	6402
<i>Xinzhu Ma (The University of Sydney), Yongtao Wang (Dalian University of Technology), Yinmin Zhang (University of Sydney), Zhiyi Xia (Tsinghua University), Yuan Meng (Tsinghua University), Zhihui Wang (Dalian University of Technology), Haojie Li (Shandong University of Science and Technology), and Wanli Ouyang (The University of Sydney)</i>	
Monocular 3D Object Detection with Bounding Box Denoising in 3D by Perceiver	6413
<i>Xianpeng Liu (NC State University), Ce Zheng (University of Central Florida), Kelvin Cheng (North Carolina State University), Nan Xue (Wuhan University), Guo-Jun Qi (OPPO Research / Westlake University), and Tianfu Wu (NC State University)</i>	
Template-guided Hierarchical Feature Restoration for Anomaly Detection	6424
<i>Hewei Guo (Institute of Automation, Chinese Academy of Sciences), Liping Ren (missing), Jingjing Fu (Microsoft), Yuwang Wang (Tsinghua University), Zhizheng Zhang (Microsoft Research), Cuiling Lan (Microsoft Research), Haoqian Wang (Tsinghua Shenzhen International Graduate School, Tsinghua University), and Xinwen Hou (Institute of Automation, Chinese Academy of Sciences)</i>	

ALWOD: Active Learning for Weakly-Supervised Object Detection	6436
<i>Yuting Wang (Rutgers University), Velibor Ilic (The Institute for Artificial Intelligence Research and Development of Serbia), Jiatong Li (Rutgers University), Branislav Kisačanin (NVIDIA), and Vladimir Pavlovic (Rutgers University)</i>	
ProtoFL: Unsupervised Federated Learning via Prototypical Distillation	6447
<i>Hansol Kim (Kakaobank), Youngjun Kwak (Kakaobank), Minyoung Jung (KETI), Jinho Shin (Kakaobank), Youngsung Kim (Inha Univ.), and Changick Kim (KAIST)</i>	
Efficient Adaptive Human-Object Interaction Detection with Concept-guided Memory	6457
<i>Ting Lei (Peking University), Fabian Caba (Adobe Research), Qingchao Chen (Peking University), Hailin Jin (Adobe Research), Yuxin Peng (Peking University), and Yang Liu (Peking University)</i>	
Detection Transformer with Stable Matching	6468
<i>Shilong Liu (Tsinghua University), Tianhe Ren (IDEA), Jiayu Chen (alibaba), Zhaoyang Zeng (IDEA), Hao Zhang (hkust), Feng Li (Hong Kong University of Science and Technology), Hongyang Li (South China University of Technology), Jun Huang (Alibaba Group), Hang Su (Tsinghua University), Jun Zhu (Tsinghua University), and Lei Zhang (International Digital Economy Academy, IDEA)</i>	
Distilling DETR with Visual-Linguistic Knowledge for Open-Vocabulary Object Detection	6478
<i>Liangqi Li (Hikvision Research Institute), Jiaxu Miao (Zhejiang University), Dahu Shi (Hikvision Research Institute), Wenming Tan (Hikvision Research Institute), Ye Ren (Hikvision Research Institute), Yi Yang (Zhejiang University), and Shiliang Pu (Hikvision Research Institute)</i>	
Anomaly Detection Under Distribution Shift	6488
<i>Tri Cao (University of Science -Vietnam National University Ho Chi Minh City), Jiawen Zhu (Singapore Management University), and Guansong Pang (Singapore Management University)</i>	
Detecting Objects with Context-Likelihood Graphs and Graph Refinement	6501
<i>Aritra Bhowmik (University of Amsterdam), Yu Wang (TomTom), Nora Baka (Tomtom), Martin R. Oswald (ETH Zurich), and Cees G. M. Snoek (University of Amsterdam)</i>	
Unsupervised Object Localization with Representer Point Selection	6511
<i>Yeonghwan Song (GIST), Seokwoo Jang (GIST), Dina Katabi (Massachusetts Institute of Technology), and Jeany Son (GIST)</i>	
DETR Does Not Need Multi-Scale or Locality Design	6522
<i>Yutong Lin (Xi'an Jiaotong University), Yuhui Yuan (Microsoft Research), Zheng Zhang (MSRA), Chen Li (Xi'an Jiaotong University), Nanning Zheng (Xi'an Jiaotong University), and Han Hu (Microsoft Research Asia)</i>	
Deep Directly-Trained Spiking Neural Networks for Object Detection	6532
<i>Qiaoyi Su (University of Chinese Academy of Sciences), Yuhong Chou (Xi'an Jiaotong university), Yifan Hu (Tsinghua University), Jianing Li (Peking University), Shijie Mei (Tsinghua University), Ziyang Zhang (HUAWEI TECHNOLOGIES CO.LTD), and Guoqi Li (Tsinghua University)</i>	

GACE: Geometry Aware Confidence Enhancement for Black-Box 3D Object Detectors on LiDAR-Data	6543
<i>David Schinagl (Graz University of Technology), Georg Krispel (Graz University of Technology), Christian Fruhwirth-Reisinger (Graz University of Technology), Horst Possegger (Graz University of Technology), and Horst Bischof (Graz University of Technology)</i>	
StageInteractor: Query-based Object Detector with Cross-stage Interaction	6554
<i>Yao Teng (Nanjing University), Haisong Liu (Nanjing University), Sheng Guo (Ant Group), and Limin Wang (Nanjing University)</i>	
Adaptive Rotated Convolution for Rotated Object Detection	6566
<i>Yifan Pu (Tsinghua University), Yiru Wang (SenseTime Group Limited), Zhuofan Xia (Tsinghua University), Yizeng Han (Tsinghua University), Yulin Wang (Tsinghua University), Weihao Gan (SenseTime), Zidong Wang (tsinghua university), Shiji Song (Department of Automation, Tsinghua University), and Gao Huang (Tsinghua)</i>	
Decoupled DETR: Spatially Disentangling Localization and Classification for Improved End-to-End Object Detection	6578
<i>Manyuan Zhang (Sensetime), Guanglu Song (Sensetime), Yu Liu (SenseTime Group LTD), and Hongsheng Li (The Chinese University of Hong Kong)</i>	
Exploring Transformers for Open-world Instance Segmentation	6588
<i>Jiannan Wu (The University of Hong Kong), Yi Jiang (Bytedance), Bin Yan (Dalian University of Technology), Huchuan Lu (Dalian University of Technology), Zehuan Yuan (Bytedance.Inc), and Ping Luo (The University of Hong Kong)</i>	
DDG-Net: Discriminability-Driven Graph Network for Weakly-supervised Temporal Action Localization	6599
<i>Xiaojun Tang (Beijing University of Posts and Telecommunications), Junsong Fan (Chinese Academy of Sciences, China), Chuanchen Luo (CASIA), Zhaoxiang Zhang (Chinese Academy of Sciences, China), Man Zhang (Beijing University of Posts and Telecommunications), and Zongyuan Yang (Beijing University of Posts and Telecommunications)</i>	
Group DETR: Fast DETR Training with Group-Wise One-to-Many Assignment	6610
<i>Qiang Chen (Baidu Inc), Xiaokang Chen (Peking University), Jian Wang (Baidu), Shan Zhang (Australian National University), Kun Yao (Baidu Inc.), Haocheng Feng (Baidu Inc.), Junyu Han (Baidu Inc.), Errui Ding (Baidu Inc.), Gang Zeng (Peking University), and Jingdong Wang (Baidu)</i>	
Category-aware Allocation Transformer for Weakly Supervised Object Localization	6620
<i>Zhiwei Chen (Xiamen University), Jinren Ding (Xiamen university), Liujuan Cao (Xiamen University), Yunhang Shen (Tencent), Shengchuan Zhang (Xiamen University), Guannan Jiang (CATL), and Rongrong Ji (Xiamen University, China)</i>	
The Devil is in the Crack Orientation: A New Perspective for Crack Detection	6630
<i>Zhuangzhuang Chen (shenzhen university), Jin Zhang (Shenzhen University), Zhuonan Lai (Shenzhen University), Guanming Zhu (shenzhen univeristy), Zun Liu (Shenzhen University), Jie Chen (Shenzhen University), and Jianqiang Li (Shenzhen University)</i>	

Clusterformer: Cluster-based Transformer for 3D Object Detection in Point Clouds	6641
<i>Yu Pei (HIK), Xian Zhao (Hikvision), Hao Li (Hikvision Digital Technology Co. Ltd), Jingyuan Ma (HikVision Research Institute), Jingwei Zhang (Hikvision Digital Technology Co. Ltd), and Shiliang Pu (Hikvision Research Institute)</i>	
Less is More: Focus Attention for Efficient DETR	6651
<i>Dehua Zheng (Huazhong University of Science and Technology), Wenhui Dong (huawei), Hailin Hu (Huawei Noah's Ark Lab), Xinghao Chen (Huawei Noah's Ark Lab), and Yunhe Wang (Huawei Technologies)</i>	
DFA3D: 3D Deformable Attention For 2D-to-3D Feature Lifting	6661
<i>Hongyang Li (South China University of Technology), Hao Zhang (hkust), Zhaoyang Zeng (IDEA), Shilong Liu (Tsinghua University), Feng Li (Hong Kong University of Science and Technology), Tianhe Ren (IDEA), and Lei Zhang (International Digital Economy Academy, IDEA)</i>	
Multi-Label Self-Supervised Learning with Scene Images	6671
<i>Ke Zhu (Nanjing University), Minghao Fu (Nanjing University), and Jianxin Wu (Nanjing University)</i>	
Cascade-DETR: Delving into High-Quality Universal Object Detection	6681
<i>Mingqiao Ye (ETHz), Lei Ke (HKUST), Siyuan Li (ETH Zurich), Yu-Wing Tai (Kuaishou Technology / HKUST), Chi-Keung Tang (Hong Kong University of Science and Technology), Martin Danelljan (ETH Zurich), and Fisher Yu (ETH Zurich)</i>	
Representation Disparity-aware Distillation for 3D Object Detection	6692
<i>Yanjing Li (Beihang University), Sheng Xu (Beihang University), Mingbao Lin (Xiamen University, China), Jihao Yin (BUAA), Baochang Zhang (Beihang University), and Xianbin Cao (Beihang University, China)</i>	
FeatEnHancer: Enhancing Hierarchical Features for Object Detection and Beyond Under Low-Light Vision	6702
<i>Khurram Azeem Hashmi (Deutsches Forschungszentrum für Künstliche Intelligenz, DFKI, GmbH), Goutham Kallempudi (TU Kaiserslautern), Didier Stricker (DFKI), and Muhammad Zeshan Afzal (Deutsches Forschungszentrum für Künstliche Intelligenz, DFKI, GmbH)</i>	
DetZero: Rethinking Offboard 3D Object Detection with Long-term Sequential Point Clouds	6713
<i>Tao Ma (Chinese University of Hong Kong), Xuemeng Yang (Shanghai AI Lab), Hongbin Zhou (Shanghai AI Lab), Xin Li (East China Normal University), Botian Shi (Shanghai AI Laboratory), Junjie Liu (South China University of Technology), Yuchen Yang (Fudan University), Zhizheng Liu (ETH Zurich), Liang He (ECNU), Yu Qiao (Shanghai AI Laboraotry), Yikang Li (Shanghai AI Lab), and Hongsheng Li (The Chinese University of Hong Kong)</i>	
DETRs with Collaborative Hybrid Assignments Training	6725
<i>Zhuofan Zong (SenseTime), Guanglu Song (Sensetime), and Yu Liu (SenseTime Group LTD)</i>	

Open-Vocabulary Object Detection With an Open Corpus	6736
<i>Jiong Wang (ZJU), Huiming Zhang (Alibaba Group), Haiwen Hong (Alibaba), Xuan Jin (Alibaba Turing Lab, Alibaba Group), Yuan He (Alibaba Group), Hui Xue (Alibaba), and Zhou Zhao (Zhejiang University)</i>	
SparseDet: Improving Sparsely Annotated Object Detection with Pseudo-positive Mining	6747
<i>Saksham Suri (University of Maryland, College Park), Saketh Rambhatla (Meta AI Research), Rama Chellappa (Johns Hopkins University), and Abhinav Shrivastava (University of Maryland)</i>	
Unsupervised Surface Anomaly Detection with Diffusion Probabilistic Model	6759
<i>Xinyi Zhang (Tsinghua University), Naiqi Li (Tsinghua University), Jiawei Li (Tsinghua University), Tao Dai (Shenzhen University), Yong Jiang (Tsinghua University), and Shu-Tao Xia (Tsinghua University)</i>	
UniTR: A Unified and Efficient Multi-Modal Transformer for Bird's-Eye-View Representation	6769
<i>Haiyang Wang (Peking University), Hao Tang (Peking University), Shaoshuai Shi (MPI Informatics), Aoxue Li (Noah's Ark Lab), Zhenguo Li (Huawei Noah's Ark Lab), Bernt Schiele (MPI Informatics), and Liwei Wang (Peking University)</i>	
Focus the Discrepancy: Intra- and Inter-Correlation Learning for Image Anomaly Detection	6780
<i>Xincheng Yao (Shanghai Jiao Tong University), Ruoqi Li (SJTU), Zefeng Qian (Shanghai Jiao Tong University), Yan Luo (Shanghai Jiao Tong University), and Chongyang Zhang (Shanghai Jiao Tong University)</i>	
MonoNeRD: NeRF-like Representations for Monocular 3D Object Detection	6791
<i>Junkai Xu (Zhejiang University), Liang Peng (ZJU), Haoran Cheng (Zhejiang University), Hao Li (FABU & ZJU), Wei Qian (Fabu Inc.), Ke Li (Fullong Tech.), Wenxiao Wang (Zhejiang University), and Deng Cai (ZJU)</i>	
Integrally Migrating Pre-trained Transformer Encoder-decoders for Visual Object Detection	6802
<i>Feng Liu (University of Chinese Academy of Sciences), Xiaosong Zhang (Beijing Academy of Artificial Intelligence), Zhiliang Peng (University of Chinese Academy of Sciences), Zonghao Guo (University of Chinese Academy of Sciences), Fang Wan (University of Chinese Academy of Sciences), Xiangyang Ji (Tsinghua University), and Qixiang Ye (University of Chinese Academy of Sciences, China)</i>	
Generating Dynamic Kernels via Transformers for Lane Detection	6812
<i>Ziye Chen (the University of Melbourne), Yu Liu (Mach Drive), Mingming Gong (University of Melbourne), Bo Du (Wuhan University), Guoqi Qian (University of Melbourne), and Kate Smith-Miles (University of Melbourne)</i>	
Meta-ZSDETR: Zero-shot DETR with Meta-learning	6822
<i>Lu Zhang (Fudan University), Chenbo Zhang (Fudan University), Jiajia Zhao (Science and Technology on Complex System Control and Intelligent Agent Cooperation Laboratory), Jihong Guan (Tongji University), and Shuigeng Zhou (Fudan University)</i>	

Spatial Self-Distillation for Object Detection with Inaccurate Bounding Boxes	6832
<i>Di Wu (University of Chinese Academy of Sciences), Pengfei Chen (University of Chinese Academy of Sciences), Xuehui Yu (University of Chinese Academy of Sciences), Guorong Li (University of Chinese Academy of Sciences), Zhenjun Han (University of Chinese Academy of Sciences), and Jianbin Jiao (University of Chinese Academy of Sciences)</i>	
AlignDet: Aligning Pre-training and Fine-tuning in Object Detection	6843
<i>Ming Li (University of Central Florida), Jie Wu (ByteDance Inc), Xionghui Wang (Bytedance), Chen Chen (University of Central Florida), Jie Qin (School of Artificial Intelligence, University of Chinese Academy of Sciences, Institute of Automation, Chinese Academy of Sciences), Xuefeng Xiao (ByteDance Inc), Rui Wang (ByteDance), Min Zheng (ByteDance), and Xin Pan (ByteDance)</i>	
MULLER: Multilayer Laplacian Resizer for Vision	6854
<i>Zhengzhong Tu (Google), Peyman Milanfar (Google), and Hossein Talebi (Google)</i>	
Unilaterally Aggregated Contrastive Learning with Hierarchical Augmentation for Anomaly Detection	6854
<i>Guodong Wang (Beihang University), Yunhong Wang (State Key Laboratory of Virtual Reality Technology and System, Beihang University, Beijing 100191, China), Jie Qin (Nanjing University of Aeronautics and Astronautics), Dongming Zhang (State Key Laboratory of Communication Content Cognition, People's Daily Online), Xiuguo Bao (National Computer Network Emergency Response Technical Team/Coordination Center of China), and Di Huang (Beihang University, China)</i>	
DETRDistill: A Universal Knowledge Distillation Framework for DETR-families	6875
<i>Jiahao Chang (University of Science and Technology of China), Shuo Wang (University of Science and Technology of China), Hai-Ming Xu (The University of Adelaide), Zehui Chen (University of Science and Technology of China), Chenhongyi Yang (University of Edinburgh), and Feng Zhao (University of Science and Technology of China)</i>	
Delving into Motion-Aware Matching for Monocular 3D Object Tracking	6886
<i>Kuan-Chih Huang (National Taiwan University), Ming-Hsuan Yang (University of California at Merced), and Yi-Hsuan Tsai (Google)</i>	
FB-BEV: BEV Representation from Forward-Backward View Transformations	6896
<i>Zhiqi Li (Nanjing University), Zhiding Yu (NVIDIA), Wenhai Wang (Nanjing University), Anima Anandkumar (Caltech), Tong Lu (Nanjing University), and Jose M. Alvarez (NVIDIA)</i>	
Learning from Noisy Data for Semi-Supervised 3D Object Detection	6906
<i>Zehui Chen (University of Science and Technology of China), Zhenyu Li (Harbin Institute of Technology), Shuo Wang (University of Science and Technology of China), Dengpan Fu (NIO), and Feng Zhao (University of Science and Technology of China)</i>	
Boosting Long-tailed Object Detection via Step-wise Learning on Smooth-tail Data	6917
<i>Na Dong (Harbin Institute of Technology), Yongqiang Zhang (Harbin institute of Technology/KAUST), Mingli Ding (Harbin institute of Technology), and Gim Hee Lee (National University of Singapore)</i>	

Objects Do Not Disappear: Video Object Detection by Single-Frame Object Location Anticipation	6927
<i>Xin Liu (Delft University of Technology), Fatemeh Karimi Nejadasl (TomTom), Jan C. van Gemert (Delft University of Technology), Olaf Booi (TU Delft), and Siloia L. Pintea (TUDelft)</i>	
Unified Visual Relationship Detection with Vision and Language Models	6939
<i>Long Zhao (Google Research), Liangzhe Yuan (Google Research), Boqing Gong (Google), Yin Cui (Google), Florian Schroff (Google Inc.), Ming-Hsuan Yang (University of California at Merced), Hartwig Adam (Google), and Ting Liu (Google Research)</i>	
Universal Domain Adaptation via Compressive Attention Matching	6951
<i>Didi Zhu (Zhejiang University), Yinchuan Li (Huawei Noah's Ark Lab), Junkun Yuan (Zhejiang University), Zexi Li (Zhejiang University), Kun Kuang (Zhejiang University), and Chao Wu (Zhejiang University)</i>	
Unsupervised Domain Adaptive Detection with Network Stability Analysis	6963
<i>Wenzhang Zhou (University of Chinese Academy of Sciences), Heng Fan (University of North Texas), Tiejian Luo (University of Chinese Academy of Sciences), and Libo Zhang (Institute of Software Chinese Academy of Sciences)</i>	
ImGeoNet: Image-induced Geometry-aware Voxel Representation for Multi-view 3D Object Detection	6973
<i>Tao Tu (National Tsing Hua University), Shun-Po Chuang (National Taiwan University), Yu-Lun Liu (National Yang Ming Chiao Tung University), Cheng Sun (National Tsing Hua University), Ke Zhang (Amazon), Donna Roy (Amazon), Cheng-Hao Kuo (Amazon), and Min Sun (NTHU)</i>	
Cyclic-Bootstrap Labeling for Weakly Supervised Object Detection	6985
<i>Yufei Yin (University of Science and Technology of China), Jiajun Deng (The University of Sydney), Wengang Zhou (University of Science and Technology of China), Li Li (University of Science and Technology of China), and Houqiang Li (University of Science and Technology of China)</i>	

Image and Video Synthesis 1

Text-Driven Generative Domain Adaptation with Spectral Consistency Regularization	6996
<i>Zhenhuan Liu (Institute of Computing Technology, Chinese Academy of Sciences), Liang Li (Institute of Computing Technology, Chinese Academy of Sciences), Jiayu Xiao (Institute of Computing Technology, Chinese Academy of Sciences), Zheng-Jun Zha (University of Science and Technology of China), and Qingming Huang (University of Chinese Academy of Sciences)</i>	
MosaiQ: Quantum Generative Adversarial Networks for Image Generation on NISQ Computers .	7007
<i>Daniel Silver (Northeastern University), Tirthak Patel (Rice University), William Cutler (Northeastern University), Aditya Ranjan (Northeastern University), Harshitta Gandhi (Northeastern University), and Devesh Tiwari (Northeastern University)</i>	

Controllable Visual-Tactile Synthesis	7017
<i>Ruihan Gao (Carnegie Mellon University), Wenzhen Yuan (Carnegie Mellon University), and Jun-Yan Zhu (Carnegie Mellon University)</i>	
Editing Implicit Assumptions in Text-to-Image Diffusion Models	7030
<i>Hadas Orgad (Technion - Israel Institute of Technology), Bahjat Kawar (Technion), and Yonatan Belinkov (Technion)</i>	
DINAR: Diffusion Inpainting of Neural Textures for One-Shot Human Avatars	7039
<i>David Svitov (Samsung AI Center), Dmitrii Gudkov (Samsung AI Center Moscow), Renat Bashirov (Samsung), and Victor Lempitsky (missing)</i>	
Smoothness Similarity Regularization for Few-Shot GAN Adaptation	7050
<i>Vadim Sushko (Bosch Center for Artificial Intelligence), Ruyu Wang (Robert Bosch GmbH), and Juergen Gall (University of Bonn)</i>	
HSR-Diff: Hyperspectral Image Super-Resolution via Conditional Diffusion Models	7060
<i>Chanyue Wu (Northwestern Polytechnical University), Dong Wang (Northwestern Polytechnical University), Yunpeng Bai (Aberystwyth University), Hanyu Mao (nwpu), Ying Li (Northwestern Polytechnical University), and Qiang Shen (Aberystwyth University)</i>	
Long-Term Photometric Consistent Novel View Synthesis with Diffusion Models	7071
<i>Jason J. Yu (York University), Fereshteh Forghani (York University), Konstantinos G. Derpanis (York University), and Marcus A. Brubaker (York University)</i>	
AutoDiffusion: Training-Free Optimization of Time Steps and Architectures for Automated Diffusion Model Acceleration	7082
<i>Lijiang Li (Xiamen University), Huixia Li (ByteDance), Xiawu Zheng (Peng Cheng Laboratory), Jie Wu (ByteDance Inc), Xuefeng Xiao (ByteDance Inc), Rui Wang (ByteDance), Min Zheng (ByteDance), Xin Pan (ByteDance), Fei Chao (Xiamen University), and Rongrong Ji (Xiamen University, China)</i>	
Collecting The Puzzle Pieces: Disentangled Self-Driven Human Pose Transfer by Permuting Textures	7092
<i>Nannan Li (Boston University), Kevin J Shih (NVIDIA), and Bryan A. Plummer (Boston University)</i>	
Multi-Directional Subspace Editing in Style-Space	7104
<i>Chen Naveh (Reichman University)</i>	
HyperReenact: One-Shot Reenactment via Jointly Learning to Refine and Retarget Faces	7115
<i>Stella Bounareli (Kingston University of London), Christos Tzelepis (Queen Mary University of London), Vasileios Argyriou (Kingston University London), Ioannis Patras (Queen Mary University of London), and Georgios Tzimiropoulos (Queen Mary University of London)</i>	
Generating Realistic Images from In-the-wild Sounds	7126
<i>Taegyeong Lee (UNIST), Jeonghun Kang (UNIST), Hyeonyu Kim (UNIST), and Taehwan Kim (UNIST)</i>	
CC3D: Layout-Conditioned Generation of Compositional 3D Scenes	7137
<i>Sherwin Bahmani (TU Darmstadt), Jeong Joon Park (Stanford University), Despoina Paschalidou (Stanford), Xingguang Yan (Simon Fraser University), Gordon Wetzstein (Stanford University), Leonidas Guibas (Stanford University), and Andrea Tagliasacchi (Google Brain and University of Toronto)</i>	

UMFuse: Unified Multi View Fusion for Human Editing Applications	7148
<i>Rishabh Jain (Adobe), Mayur Hemani (Adobe Systems), Duygu Ceylan (Adobe Research), Krishna Kumar Singh (Adobe Research), Jingwan Lu (Adobe Research), Mausoom Sarkar (Adobe), and Balaji Krishnamurthy (MDSR Adobe)</i>	
Evaluating Data Attribution for Text-to-Image Models	7158
<i>Sheng-Yu Wang (Carnegie Mellon University), Alexei A. Efros (UC Berkeley), Jun-Yan Zhu (Carnegie Mellon University), and Richard Zhang (Adobe)</i>	
Neural Characteristic Function Learning for Conditional Image Generation	7170
<i>Shengxi Li (Beihang University), Jialu Zhang (Beihang University), Yifei Li (Beihang university), Mai Xu (BUAA), Xin Deng (Beihang university), and Li Li (Beihang university)</i>	
WaveIPT: Joint Attention and Flow Alignment in the Wavelet domain for Pose Transfer	7181
<i>Liyuan Ma (Zhejiang University), Tingwei Gao (Alibaba Group), Haitian Jiang (Zhejiang University), Haibin Shen (Zhejiang University), and Kejie Huang (Zhejiang University)</i>	
LayoutDiffusion: Improving Graphic Layout Generation by Discrete Diffusion Probabilistic Models	7192
<i>Junyi Zhang (Shanghai Jiao Tong University), Jiaqi Guo (Microsoft), Shizhao Sun (Microsoft Research), Jian-Guang Lou (Microsoft), and Dongmei Zhang (Microsoft Research Asia)</i>	
Human-Inspired Facial Sketch Synthesis with Dynamic Adaptation	7203
<i>Fei Gao (Hangzhou Institute of Technology, Xidian University), Yifan Zhu (Hangzhou Dianzi University), Chang Jiang (Hangzhou Dianzi University), and Nannan Wang (Xidian University)</i>	
Conceptual and Hierarchical Latent Space Decomposition for Face Editing	7214
<i>Savas Ozkan (Samsung Research UK), Mete Ozay (Samsung Research UK), and Tom Robinson (Samsung)</i>	
Improving Diversity in Zero-Shot GAN Adaptation with Semantic Variations	7224
<i>Seogkyu Jeon (Yonsei University), Bei Liu (Microsoft Research Asia), Pilhyeon Lee (Yonsei University), Kibeom Hong (Yonsei University; SwatchOn), Jianlong Fu (Microsoft Research Asia), and Hyeran Byun (Yonsei University)</i>	
BallGAN: 3D-aware Image Synthesis with a Spherical Background	7234
<i>Minjung Shin (Yonsei University), Yunji Seo (Yonsei University), Jeongmin Bae (Yonsei University), Young Sun Choi (Yonsei University), Hyunsu Kim (NAVER AI Lab), Hyeran Byun (Yonsei University), and Youngjung Uh (Yonsei University)</i>	
End-to-End Diffusion Latent Optimization Improves Classifier Guidance	7246
<i>Bram Wallace (Cornell University), Akash Gokul (University of California at Berkeley), Stefano Ermon (Stanford University), and Nikhil Naik (MIT)</i>	
Deep Geometrized Cartoon Line Inbetweening	7257
<i>Li Siyao (Nanyang Technological University), Tianpei Gu (UCLA), Weiye Xiao (Southeast University), Henghui Ding (Nanyang Technological University), Ziwei Liu (Nanyang Technological University), and Chen Change Loy (Nanyang Technological University)</i>	

UnitedHuman: Harnessing Multi-Source Data for High-Resolution Human Generation	7267
<i>Jianglin Fu (SenseTime), Shikai Li (SenseTime Research), Yuming Jiang (Nanyang Technological University), Kwan-Yee Lin (SenseTime Research), Wayne Wu (Shanghai AI Laboratory), and Ziwei Liu (Nanyang Technological University)</i>	
Towards Authentic Face Restoration with Iterative Diffusion Models and Beyond	7278
<i>Yang Zhao (University at Buffalo), Tingbo Hou (Google Research), Yu-Chuan Su (Google), Xuhui Jia (Google), Yandong Li (Google), and Matthias Grundmann (Google Research)</i>	
SVDiff: Compact Parameter Space for Diffusion Fine-Tuning	7289
<i>Ligong Han (Rutgers University), Yinxiao Li (Google), Han Zhang (Google), Peyman Milanfar (Google), Dimitris Metaxas (Rutgers), and Feng Yang (Google Research)</i>	
MI-GAN: A Simple Baseline for Image Inpainting on Mobile Devices	7301
<i>Andranik Sargsyan (Picsart Inc.), Shant Navasardyan (Picsart Inc.), Xingqian Xu (UIUC), and Humphrey Shi (U of Oregon; UIUC; PAIR)</i>	
Structure and Content-Guided Video Synthesis with Diffusion Models	7312
<i>Patrick Esser (Runway), Johnathan Chiu (Runway), Parmida Atighehchian (RunwayML), Jonathan Granskog (Runway), and Anastasis Germanidis (Runway AI, Inc.)</i>	
Scenimefy: Learning to Craft Anime Scene via Semi-Supervised Image-to-Image Translation	7323
<i>Yuxin Jiang (Nanyang Technological University), Liming Jiang (Nanyang Technological University), Shuai Yang (Nanyang Technological University), and Chen Change Loy (Nanyang Technological University)</i>	
Efficient-VQGAN: Towards High-Resolution Image Generation with Efficient Vision Transformers	7334
<i>Shiyue Cao (Institute of Automation, Chinese Academy of Sciences), Yueqin Yin (Institute of Automation, Chinese Academy of Sciences), Lianghua Huang (Alibaba Group), Yu Liu (Alibaba Group), Xin Zhao (Institute of Automation, Chinese Academy of Sciences), Deli Zhao (Alibaba Group), and Kaiqi Huang (Chinese Academy of Sciences, China)</i>	
A Latent Space of Stochastic Diffusion Models for Zero-Shot Image Editing and Guidance	7344
<i>Chen Henry Wu (Carnegie Mellon University) and Fernando De la Torre (Carnegie Mellon)</i>	
Generative Multiplane Neural Radiance for 3D-Aware Image Generation	7354
<i>Amandeep Kumar (MBZUAI), Ankan Kumar Bhunia (MBZUAI), Sanath Narayan (Technology Innovation Institute), Hisham Cholakkal (MBZUAI), Rao Muhammad Anwer (MBZUAI/AALTO), Salman Khan (MBZUAI), Ming-Hsuan Yang (University of California at Merced), and Fahad Shahbaz Khan (MBZUAI)</i>	
Parallax-Tolerant Unsupervised Deep Image Stitching	7365
<i>Lang Nie (Beijing Jiaotong University), Chunyu Lin (Beijing Jiaotong University), Kang Liao (Beijing Jiaotong University), Shuaicheng Liu (UESTC; Megvii), and Yao Zhao (Beijing Jiaotong University)</i>	
GAIT: Generating Aesthetic Indoor Tours with Deep Reinforcement Learning	7375
<i>Desai Xie (Stony Brook University), Ping Hu (Stony Brook University), Xin Sun (Adobe Research), Soren Pirk (Adobe), Jianming Zhang (Adobe Research), Radomir Mech (Adobe Systems Incorporated), and Arie E. Kaufman (Stony Brook University)</i>	

EverLight: Indoor-Outdoor Editable HDR Lighting Estimation	7386
<i>Mohammad Reza Karimi Karimi Dastjerdi (Université Laval), Jonathan Eisenmann (Adobe), Yannick Hold-Geoffroy (Adobe Research), and Jean-François Lalonde (Université Laval)</i>	
Prompt Tuning Inversion for Text-driven Image Editing Using Diffusion Models	7396
<i>Wenkai Dong (Institute of automation, university of chinese academy of sciences), Song Xue (Baidu), Xiaoyue Duan (Beihang University), and Shumin Han (Baidu Inc.)</i>	
Efficient Diffusion Training via Min-SNR Weighting Strategy	7407
<i>Tiankai Hang (Southeast University), Shuyang Gu (Microsoft Research Asia), Chen Li (Xi'an Jiaotong University), Jianmin Bao (Microsoft Research Asia), Dong Chen (Microsoft Research Asia), Han Hu (Microsoft Research Asia), Xin Geng (Southeast University), and Baining Guo (MSR Asia)</i>	
BoxDiff: Text-to-Image Synthesis with Training-Free Box-Constrained Diffusion	7418
<i>Jinheng Xie (National University of Singapore), Yuexiang Li (Jarvis Lab, Tencent), Yawen Huang (Tencent), Haozhe Liu (King Abdullah University of Science and Technology), Wentian Zhang (Shenzhen University), Yefeng Zheng (Tencent), and Mike Zheng Shou (National University of Singapore)</i>	
Improving Sample Quality of Diffusion Models Using Self-Attention Guidance	7428
<i>Susung Hong (Korea University), Gyuseong Lee (Korea University), Wooseok Jang (Korea University), and Seungryong Kim (Korea University)</i>	
Not All Steps are Created Equal: Selective Diffusion Distillation for Image Manipulation	7438
<i>Luozhou Wang (Hong Kong University of Science and Technology, (GUANGZHOU), Shuai Yang (Hong Kong University of Science and Technology, (Guangzhou), Shu Liu (SmartMore), and Ying-cong Chen (Hong Kong University of Science and Technology, (Guangzhou)</i>	
Deep Image Harmonization with Learnable Augmentation	7448
<i>Li Niu (Shanghai Jiao Tong University), Junyan Cao (Shanghai Jiao Tong University), Wenyan Cong (University of Texas at Austin), and Liqing Zhang (Shanghai Jiao Tong University)</i>	
Out-of-Domain GAN Inversion via Invertibility Decomposition for Photo-Realistic Human Face Manipulation	7458
<i>Xin Yang (The Hong Kong University of Science and Technology), Xiaogang Xu (The Chinese University of Hong Kong), and Yingcong Chen (Hong Kong University of Science and Technology, (Guangzhou)</i>	
Bidirectionally Deformable Motion Modulation For Video-based Human Pose Transfer	7468
<i>Wing-Yin Yu (City University of Hong Kong), Lai-Man Po (City University of Hong Kong), Ray C.C. Cheung (City University of Hong Kong), Yuzhi Zhao (City University of Hong Kong), Yu Xue (City University of Hong Kong), and Kun Li (City University of Hong Kong)</i>	
Size Does Matter: Size-aware Virtual Try-on via Clothing-oriented Transformation Try-on Network	7479
<i>Chieh-Yun Chen (None), Yi-Chung Chen (National Taiwan University), Hong-Han Shuai (National Yang Ming Chiao Tung University), and Wen-Huang Cheng (National Taiwan University)</i>	

VidStyleODE: Disentangled Video Editing via StyleGAN and NeuralODEs	7489
<i>Moayed Haji Ali (Koc University), Andrew Bond (Koç University), Tolga Birdal (Imperial College London), Duygu Ceylan (Adobe Research), Levent Karacan (Iskenderun Technical University), Erkut Erdem (Hacettepe University), and Aykut Erdem (Koc University)</i>	
Learning Global-aware Kernel for Image Harmonization	7501
<i>Xintian Shen (Zhejiang university), Jiangning Zhang (Zhejiang University), Jun Chen (Zhejiang University), Shipeng Bai (Zhejiang University), Yue Han (Zhejiang University), Yabiao Wang (Tencent), Chengjie Wang (Tencent Shanghai Jiao Tong University), and Yong Liu (Zhejiang University)</i>	
Expressive Text-to-Image Generation with Rich Text	7511
<i>Songwei Ge (University of Maryland, College Park), Taesung Park (Adobe Research), Jun-Yan Zhu (Carnegie Mellon University), and Jia-Bin Huang (University of Maryland College Park)</i>	
A Large-Scale Outdoor Multi-Modal Dataset and Benchmark for Novel View Synthesis and Implicit Scene Reconstruction	7523
<i>Chongshan Lu (Fudan University), Fukun Yin (Fudan University), Xin Chen (ShanghaiTech University), Wen Liu (ShanghaiTech University), Tao Chen (Fudan University), Gang Yu (Tencent), and Jiayuan Fan (Fudan University)</i>	
Efficient Region-Aware Neural Radiance Fields for High-Fidelity Talking Portrait Synthesis.....	7534
<i>Jiahe Li (Beihang University), Jiawei Zhang (Beihang University), Xiao Bai (Beihang University), Jun Zhou (Griffith University), and Lin Gu (RIKENAIP / The University of Tokyo)</i>	
Perceptual Artifacts Localization for Image Synthesis Tasks	7545
<i>Lingzhi Zhang (University of Pennsylvania), Zhengjie Xu (University of Pennsylvania), Connelly Barnes (Adobe), Yuqian Zhou (Adobe), Qing Liu (Adobe), He Zhang (Adobe), Sohrab Amirghodsi (Adobe Research), Zhe Lin (Adobe Research), Eli Shechtman (Adobe Research, US), and Jianbo Shi (University of Pennsylvania)</i>	
Learning to Generate Semantic Layouts for Higher Text-Image Correspondence in Text-to-Image Synthesis	7557
<i>Minho Park (Korea Advanced Institute of Science and Technology), Jooyeol Yun (Korea Advanced Institute of Science and Technology), Seunghwan Choi (Korea Advanced Institute of Science and Technology), and Jaegul Choo (Korea Advanced Institute of Science and Technology)</i>	
StylerDALLE: Language-Guided Style Transfer Using a Vector-Quantized Tokenizer of a Large-Scale Generative Model	7567
<i>Zipeng Xu (University of Trento), Enver Sangineto (University of Modena and Reggio Emilia), and Nicu Sebe (Trento, Italy)</i>	
Shortcut-V2V: Compression Framework for Video-to-Video Translation Based on Temporal Redundancy Reduction	7578
<i>Chaeyeon Chung (Korea Advanced Institute of Science and Technology), Yeojeong Park (KAIST), Seunghwan Choi (Korea Advanced Institute of Science and Technology), Munkhsoyol Ganbat (KAIST), and Jaegul Choo (Korea Advanced Institute of Science and Technology)</i>	

Tune-A-Video: One-Shot Tuning of Image Diffusion Models for Text-to-Video Generation	7589
<i>Jay Zhangjie Wu (National University of Singapore), Yixiao Ge (Tencent), Xintao Wang (Tencent), Stan Weixian Lei (National University of Singapore), Yuchao Gu (National University of Singapore), Yufei Shi (National University of Singapore), Wynne Hsu (National University of Singapore), Ying Shan (Tencent), Xiaohu Qie (Tencent), and Mike Zheng Shou (National University of Singapore)</i>	
BlendFace: Re-designing Identity Encoders for Face-Swapping	7600
<i>Kaede Shiohara (The University of Tokyo), Xingchao Yang (Cyberagent, Inc), and Takafumi Taketomi (Cyberagent)</i>	
Talking Head Generation with Probabilistic Audio-to-Visual Diffusion Priors	7611
<i>Zhentao Yu (XiaoBing), Zixin Yin (The Hong Kong University of Science and Technology), Deyu Zhou (HKUST(GZ), Duomin Wang (xiaobing.ai), Finn Wong (Xiaobing), and Baoyuan Wang (Xiaobing.AI)</i>	
LinkGAN: Linking GAN Latents to Pixels for Controllable Image Synthesis	7622
<i>Jiapeng Zhu (HKUST), Ceyuan Yang (Chinese University of Hong Kong), Yujun Shen (Dept. of IE, CUHK), Zifan Shi (HKUST), Bo Dai (Shanghai AI Lab), Deli Zhao (Alibaba Group), and Qifeng Chen (HKUST)</i>	
Open-vocabulary Object Segmentation with Diffusion Models	7633
<i>Ziyi Li (Shanghai Jiao Tong University), Qinye Zhou (Shanghai Jiao Tong University), Xiaoyun Zhang (Shanghai Jiao Tong University), Ya Zhang (Cooperative Medianet Innovation Center, Shanghai Jiao Tong University), Yanfeng Wang (Cooperative medianet innovation center of Shanghai Jiao Tong University), and Weidi Xie (Shanghai Jiao Tong University)</i>	
StyleDiffusion: Controllable Disentangled Style Transfer via Diffusion Models	7643
<i>Zhizhong Wang (Zhejiang University), Lei Zhao (Zhejiang University), and Wei Xing (Zhejiang University)</i>	
ToonTalker: Cross-Domain Face Reenactment	7656
<i>Yuan Gong (Tsinghua University), Yong Zhang (Tencent AI Lab), Xiaodong Cun (Tencent AI Lab), Fei Yin (Tsinghua University), Yanbo Fan (Tencent AI Lab), Xuan Wang (Ant Group), Baoyuan Wu (The Chinese University of Hong Kong, Shenzhen, Shenzhen Research Institute of Big Data), and Yujiu Yang (Tsinghua University)</i>	
Dense Text-to-Image Generation with Attention Modulation	7667
<i>Yunji Kim (NAVER AI Lab), Jiyoung Lee (NAVER AI Lab), Jin-Hwa Kim (NAVER AI Lab), Jung-Woo Ha (NAVER Cloud AI Lab), and Jun-Yan Zhu (Carnegie Mellon University)</i>	
Householder Projector for Unsupervised Latent Semantics Discovery	7678
<i>Yue Song (University of Trento), Jichao Zhang (University of Trento), Nicu Sebe (Trento, Italy), and Wei Wang (EPFL)</i>	
Deep Image Harmonization with Globally Guided Feature Transformation and Relation Distillation	7689
<i>Li Niu (Shanghai Jiao Tong University), Linfeng Tan (Shanghai Jiao Tong University), Xinhao Tao (Shanghai Jiao Tong University), Junyan Cao (Shanghai Jiao Tong University), Fengjun Guo (IntSig Information Co. Ltd), Teng Long (INTSIG), and Liqing Zhang (Shanghai Jiao Tong University)</i>	

One-Shot Generative Domain Adaptation	7699
<i>Ceyuan Yang (Chinese University of Hong Kong), Yujun Shen (Dept. of IE, CUHK), Zhiyi Zhang (ByteDance Inc.), Yinghao Xu (Chinese University of Hong Kong), Jiapeng Zhu (HKUST), Zhirong Wu (Microsoft Research), and Bolei Zhou (UCLA)</i>	
Hashing Neural Video Decomposition with Multiplicative Residuals in Space-Time	7709
<i>Cheng-Hung Chan (National Tsing Hua University), Cheng-Yang Yuan (National Tsing Hua University), Cheng Sun (National Tsing Hua University), and Hwann-Tzong Chen (National Tsing Hua University)</i>	
Versatile Diffusion: Text, Images and Variations All in One Diffusion Model	7720
<i>Xingqian Xu (UIUC), Zhangyang Wang (University of Texas at Austin), Gong Zhang (University of Oregon), Kai Wang (U of Oregon), and Humphrey Shi (U of Oregon; UIUC; PAIR)</i>	
Harnessing the Spatial-Temporal Attention of Diffusion Models for High-Fidelity Text-to-Image Synthesis	7732
<i>Qiucheng Wu (University of California, Santa Barbara), Yujian Liu (University of California, Santa Barbara), Handong Zhao (Adobe Research), Trung Bui (Adobe Research), Zhe Lin (Adobe Research), Yang Zhang (IBM T. J. Watson Research), and Shiyu Chang (UCSB)</i>	

Vision and Audio

Sound Source Localization is All about Cross-Modal Alignment	7743
<i>Arda Senocak (KAIST), Hyeonngon Ryu (KAIST), Junsik Kim (Harvard University), Tae-Hyun Oh (POSTECH), Hanspeter Pfister (Harvard University), and Joon Son Chung (KAIST)</i>	
Class-Incremental Grouping Network for Continual Audio-Visual Learning	7754
<i>Shentong Mo (Carnegie Mellon University), Weiguo Pian (University of Luxembourg), and Yapeng Tian (The University of Texas at Dallas)</i>	
Audio-Visual Class-Incremental Learning	7765
<i>Weiguo Pian (University of Luxembourg), Shentong Mo (Carnegie Mellon University), Yunhui Guo (University of Texas at Dallas), and Yapeng Tian (The University of Texas at Dallas)</i>	
DiffV2S: Diffusion-Based Video-to-Speech Synthesis with Vision-Guided Speaker Embedding	7778
<i>Jeongsoo Choi (KAIST), Joanna Hong (KAIST), and Yong Man Ro (KAIST)</i>	
The Power of Sound (TPoS): Audio Reactive Video Generation with Stable Diffusion	7788
<i>Yujin Jeong (Korea University), Wonjeong Ryoo (Korea University), Seunghyun Lee (Korea University), Dabin Seo (Korea University), Wonmin Byeon (NVIDIA Research), Sangpil Kim (Korea University), and Jinkyu Kim (Korea University)</i>	
SIDGAN: High-Resolution Dubbed Video Generation via Shift-Invariant Learning	7799
<i>Urwa Muaz (Amazon), Wondong Jang (Amazon Studios), Rohun Tripathi (Amazon), Santhosh Mani (Amazon.com), Wenbin Ouyang (Amazon), Ravi Teja Gadde (Amazon), Baris Gecer (Amazon Studios), Sergio Elizondo (Amazon), Reza Madad (Amazon), and Naveen Nair (Amazon)</i>	

On the Audio-visual Synchronization for Lip-to-Speech Synthesis	7809
<i>Zhe Niu (The Hong Kong University of Science and Technology) and Brian Mak (The Hong Kong University of Science and Technology)</i>	
Be Everywhere - Hear Everything (BEE): Audio Scene Reconstruction by Sparse Audio-Visual Samples	7819
<i>Mingfei Chen (University of Washington), Kun Su (Department of Electrical & Computer Engineering, University of Washington, Seattle), and Eli Shlizerman (Departments of Applied Mathematics and Electrical & Computer Engineering, University of Washington Seattle)</i>	
Dense 2D-3D Indoor Prediction with Sound via Aligned Cross-Modal Distillation	7829
<i>Heeseung Yun (Seoul National University), Joonil Na (Seoul National University), and Gunhee Kim (Seoul National University)</i>	
Hyperbolic Audio-visual Zero-shot Learning	7839
<i>Jie Hong (Australian National University), Zeeshan Hayder (CSIRO), Junlin Han (CSIRO), Pengfei Fang (The Australian National University), Mehrtash Harandi (Monash University), and Lars Petersson (Data61/CSIRO)</i>	
AdVerb: Visually Guided Audio Dereverberation	7850
<i>Sanjoy Chowdhury (University of Maryland, College Park), Sreyan Ghosh (University of Maryland, College Park), Subhrajyoti Dasgupta (Mila - Quebec AI Institute), Anton Ratnarajah (University of Maryland, College Park), Utkarsh Tyagi (University of Maryland, College Park), and Dinesh Manocha (University of Maryland at College Park)</i>	
Sound Localization from Motion: Jointly Learning Sound Direction and Camera Rotation	7863
<i>Ziyang Chen (University of Michigan), Shengyi Qian (University of Michigan), and Andrew Owens (University of Michigan)</i>	

(O) Humans, 3D Modeling, and Driving

Text2Room: Extracting Textured 3D Meshes from 2D Text-to-Image Models	7875
<i>Lukas Höllein (Technical University of Munich), Ang Cao (University of Michigan), Andrew Owens (U Michigan), Justin Johnson (University of Michigan), and Matthias Nießner (Technical University of Munich)</i>	
LivePose: Online 3D Reconstruction from Monocular Video with Dynamic Camera Poses	7887
<i>Noah Stier (University of California, Santa Barbara), Baptiste Angles (Apple), Liang Yang (Apple Inc), Yajie Yan (Apple), Alex Colburn (Apple Inc.), and Ming Chuang (Apple)</i>	
NDDepth: Normal-Distance Assisted Monocular Depth Estimation	7897
<i>Shuwei Shao (Beihang University), Zhongcai Pei (beihang university), Weihai Chen (Beihang University), Xingming Wu (Beihang University), and Zhengguo Li (Institute for Infocomm Research, A*STAR)</i>	
LATR: 3D Lane Detection from Monocular Images with Transformer	7907
<i>Yueru Luo (The Chinese University of Hong Kong, ShenZhen), Chaoda Zheng (The Chinese University of Hong Kong, Shen Zhen), Xu Yan (The Chinese University of Hong Kong, Shenzhen), Tang Kun (Tencent), Chao Zheng (Tencent), Shuguang Cui (The Chinese University of Hong Kong, Shenzhen), and Zhen Li (The Chinese University of Hong Kong, Shenzhen)</i>	

DriveAdapter: Breaking the Coupling Barrier of Perception and Planning in End-to-End Autonomous Driving	7919
<i>Xiaosong Jia (Shanghai Jiao Tong University), Yulu Gao (Beihang University), Li Chen (Shanghai AI Laboratory), Junchi Yan (Shanghai Jiao Tong University), Patrick Langechuan Liu (Xsense.ai), and Hongyang Li (Shanghai AI Lab / Shanghai Jiao Tong University)</i>	
Dynamic Point Fields	7930
<i>Sergey Prokudin (ETH Zurich), Qianli Ma (Max Planck Institute for Intelligent Systems), Maxime Raafat (ETH Zurich), Julien Valentin (Microsoft), and Siyu Tang (ETH Zurich)</i>	
Generalizing Neural Human Fitting to Unseen Poses With Articulated SE(3) Equivariance	7943
<i>Haiwen Feng (Max Planck Institute for Intelligent Systems), Peter Kulits (Max Planck Institute for Intelligent Systems), Shichen Liu (University of Southern California), Michael J. Black (Max Planck Institute for Intelligent Systems), and Victoria Fernandez Abrevaya (Max Planck Institute)</i>	
Probabilistic Human Mesh Recovery in 3D Scenes from Egocentric Views	7955
<i>Siwai Zhang (ETH Zurich), Qianli Ma (Max Planck Institute for Intelligent Systems), Yan Zhang (ETH Zurich), Sadegh Aliakbarian (Microsoft), Darren Cosker (Microsoft), and Siyu Tang (ETH Zurich)</i>	
DECO: Dense Estimation of 3D Human-Scene Contact In The Wild	7967
<i>Shashank Tripathi (Max Planck Institute for Intelligent Systems, Tübingen, Germany), Agniv Chatterjee (Max Planck Institute for Intelligent Systems, Tübingen, Germany), Jean-Claude Passy (Max Planck Institute for Intelligent Systems, Tübingen, Germany), Hongwei Yi (Max Planck Institute for Intelligent Systems, Tübingen, Germany), Dimitrios Tzionas (University of Amsterdam, the Netherlands), and Michael J. Black (Max Planck Institute for Intelligent Systems, Tübingen, Germany)</i>	
Decoupled Iterative Refinement Framework for Interacting Hands Reconstruction from a Single RGB Image	7980
<i>Pengfei Ren (Beijing University of Posts and Telecommunications), Chao Wen (Bytedance), Xiaozheng Zheng (ByteDance), Zhou Xue (Bytedance), Haifeng Sun (Beijing university of posts and telecommunications), Qi Qi (Beijing University of Posts and Telecommunications), Jingyu Wang (Beijing University of Posts and Telecommunications), and Jianxin Liao (beijing university of posts and telecommunications)</i>	
Chasing Clouds: Differentiable Volumetric Rasterisation of Point Clouds as a Highly Efficient and Accurate Loss for Large-Scale Deformable 3D Registration	7992
<i>Mattias P. Heinrich (Universitaet zu Luebeck), Alexander Bigalke (University of Lübeck), Christoph Großbröhmer (University of Lübeck), and Lasse Hansen (EchoScout GmbH)</i>	
Rehearsal-Free Domain Continual Face Anti-Spoofing: Generalize More and Forget Less	8003
<i>Rizhao Cai (NTU Singapore), Yawen Cui (University of Oulu), Zhi Li (Nanyang Technological University), Zitong Yu (Great Bay University), Haoliang Li (CityU), Yongjian Hu (South China University of Technology), and Alex Kot (Nanyang Technological University)</i>	

(O) Low-Level Vision and Theory

A 5-Point Minimal Solver for Event Camera Relative Motion Estimation	8015
<i>Ling Gao (ShanghaiTech University), Hang Su (ShanghaiTech University), Daniel Gehrig (University of Zurich and ETH Zurich), Marco Cannici (University of Zurich), Davide Scaramuzza (University of Zurich, Switzerland), and Laurent Kneip (ShanghaiTech University)</i>	
General Planar Motion from a Pair of 3D Correspondences	8026
<i>Juan Carlos Dibene (Stevens Institute of Technology), Zhixiang Min (Stevens Institute of Technology), and Enrique Dunn (Stevens Institute of Technology)</i>	
Beyond the Pixel: a Photometrically Calibrated HDR Dataset for Luminance and Color Prediction	8037
<i>Christophe Bolduc (Université Laval), Justine Giroux (Université Laval), Marc Hébert (Laval University), Claude Demers (Université Laval), and Jean-François Lalonde (Université Laval)</i>	
DDFM: Denoising Diffusion Model for Multi-Modality Image Fusion	8048
<i>Zixiang Zhao (Xi'an Jiaotong University), Haowen Bai (Xi'an Jiaotong University), Yuanzhi Zhu (ETH Zurich), Jianshe Zhang (Xi'an Jiaotong University), Shuang Xu (Northwestern Polytechnical University), Yulun Zhang (ETH Zurich), Kai Zhang (ETH, Zurich), Deyu Meng (Xi'an Jiaotong University), Radu Timofte (University of Wurzburg & ETH Zurich), and Luc Van Gool (ETH Zurich)</i>	
Iterative Prompt Learning for Unsupervised Backlit Image Enhancement	8060
<i>Zhexin Liang (Nanyang Technological University), Chongyi Li (Nanyang Technological University), Shangchen Zhou (Nanyang Technological University), Ruicheng Feng (Nanyang Technological University), and Chen Change Loy (Nanyang Technological University)</i>	
Similarity Min-Max: Zero-Shot Day-Night Domain Adaptation	8070
<i>Rundong Luo (Peking University), Wenjing Wang (Peking University), Wenhan Yang (Peng Cheng Laboratory), and Jiaying Liu (Peking University)</i>	
Multi-interactive Feature Learning and a Full-time Multi-modality Benchmark for Image Fusion and Segmentation	8081
<i>Jinyuan Liu (Dalian University of Technology), Zhu Liu (Dalian University of Technology), Guanyao Wu (Dalian University of Technology), Long Ma (Dalian University of Technology), Risheng Liu (Dalian University of Technology), Wei Zhong (Dalian University of Technology), Zhongxuan Luo (DALIAN UNIVERSITY OF TECHNOLOGY), and Xin Fan (Dalian University of Technology)</i>	
Computational 3D Imaging with Position Sensors	8091
<i>Jeremy Klotz (Columbia University), Mohit Gupta (University of Wisconsin-Madison, USA), and Aswin Sankaranarayanan (Carnegie Mellon University)</i>	
Passive Ultra-Wideband Single-Photon Imaging	8101
<i>Mian Wei (University of Toronto), Sotiris Nousias (University of Toronto), Rahul Gulve (University of Toronto), David B. Lindell (University of Toronto), and Kiriakos N. Kutulakos (University of Toronto)</i>	

Viewing Graph Solvability in Practice	8113
<i>Federica Arrigoni (Politecnico di Milano), Tomas Pajdla (Czech Technical University in Prague), and Andrea Fusiello (UNIUD)</i>	
Minimal Solutions to Generalized Three-View Relative Pose Problem	8122
<i>Yaqing Ding (Lund University), Chiang-Heng Chien (Brown University), Viktor Larsson (Lund University), Karl Åström (Lund University), and Benjamin Kimia (USA)</i>	
SoDaCam: Software-defined Cameras via Single-Photon Imaging	8131
<i>Varun Sundar (University of Wisconsin Madison), Andrei Ardelean (EPFL), Tristan Swedish (Ubicept Inc), Claudio Bruschini (EPFL), Edoardo Charbon (EPFL), and Mohit Gupta (University of Wisconsin-Madison, USA)</i>	

Navigation and Autonomous Driving

Robust Monocular Depth Estimation under Challenging Conditions	8143
<i>Stefano Gasperini (Technical University of Munich), Nils Morbitzer (Technical University of Munich), HyunJun Jung (Technical University of Munich), Nassir Navab (TU Munich, Germany), and Federico Tombari (Google, TU Munich)</i>	
UMC: A Unified Bandwidth-efficient and Multi-resolution based Collaborative Perception Framework	8153
<i>Tianhang Wang (Tongji University), Guang Chen (Tongji University), Kai Chen (Tongji University), Zhengfa Liu (Tongji University), Bo Zhang (westwell), Alois Knoll (Robotics and Embedded Systems), and Changjun Jiang (Tongji University)</i>	
View Consistent Purification for Accurate Cross-View Localization	8163
<i>Shan Wang (ANU, CSIRO), Yanhao Zhang (Australian National University), Akhil Perincherry (Ford Motor Company), Ankit Vora (Ford Motor Company), and Hongdong Li (Australian National University, Australia)</i>	
Semi-supervised Semantics-guided Adversarial Training for Robust Trajectory Prediction	8173
<i>Ruo Chen Jiao (Northwestern University), Xiangguo Liu (Northwestern University), Takami Sato (University of California, Irvine), Qi Alfred Chen (UC Irvine), and Qi Zhu (Northwestern University)</i>	
NeRF-LOAM: Neural Implicit Representation for Large-Scale Incremental LiDAR Odometry and Mapping	8184
<i>Junyuan Deng (Shanghai Jiao Tong University), Qi Wu (Shanghai Jiao Tong University), Xieyuanli Chen (the College of Intelligence Science and Technology, National University of Defense Technology), Songpengcheng Xia (Shanghai Jiao Tong University), Zhen Sun (Shanghai Jiao Tong University), Guoqing Liu (Shanghai Jiao Tong University), Wenxian Yu (Shanghai Jiao Tong university), and Ling Pei (Shanghai Jiao Tong University)</i>	
MapPrior: Bird's-Eye View Map Layout Estimation with Generative Models	8194
<i>Xiyue Zhu (University of Illinois Urbana Champaign), Vlas Zyrianov (University of Illinois Urbana Champaign), Zhijian Liu (MIT), and Shenlong Wang (UIUC)</i>	

Hidden Biases of End-to-End Driving Models	8206
<i>Bernhard Jaeger (University of Tübingen), Kashyap Chitta (University of Tübingen), and Andreas Geiger (University of Tübingen)</i>	
Search for or Navigate to? Dual Adaptive Thinking for Object Navigation	8216
<i>Ronghao Dang (Tongji University), Liuyi Wang (Tongji University), Zongtao He (Tongji University), Shuai Su (Tongji University), Jiagui Tang (Tongji University), Chengju Liu (Tongji University), and Qijun Chen (Tongji University)</i>	
BiFF: Bi-level Future Fusion with Polyline-based Coordinate for Interactive Trajectory Prediction	8226
<i>Yiyao Zhu (HKUST), Di Luan (HKUST), and Shaojie Shen (HKUST)</i>	
Towards Zero Domain Gap: A Comprehensive Study of Realistic LiDAR Simulation for Autonomy Testing	8238
<i>Sivabalan Manivasagam (University of Toronto), Ioan Andrei Bârsan (Uber ATG, University of Toronto), Jingkang Wang (University of Toronto), Ze Yang (University of Toronto), and Raquel Urtasun (Waabi)</i>	
Clustering based Point Cloud Representation Learning for 3D Analysis	8249
<i>Tuo Feng (University of Technology Sydney), Wenguan Wang (Zhejiang University), Xiaohan Wang (Zhejiang University), Yi Yang (UTS), and Qinghua Zheng (Xi'an Jiaotong University)</i>	
ADAPT: Efficient Multi-Agent Trajectory Prediction with Adaptation	8261
<i>Görkay Aydemir (Codeway), Adil Kaan Akan (Koc University), and Fatma Güney (Koc University)</i>	
MV-DeepSDF: Implicit Modeling with Multi-Sweep Point Clouds for 3D Vehicle Reconstruction in Autonomous Driving	8272
<i>Yibo Liu (York University), Kelly Zhu (University of Toronto), Guile Wu (Noah's Ark Lab), Yuan Ren (Huawei), Bingbing Liu (Huawei Noah's Ark Lab, Canada), Yang Liu (Huawei Technologies Canada), and Jinjun Shan (York University)</i>	
Learning Vision-and-Language Navigation from YouTube Videos	8283
<i>Kunyang Lin (South China University of Technology, Peking University), Peihao Chen (South China University of Technology), Diwei Huang (South China University of Technology), Thomas H. Li (Peking University Shenzhen Graduate School), Mingkui Tan (South China University of Technology; Key Laboratory of Big Data and Intelligent Robot, Ministry of Education), and Chuang Gan (UMass Amherst, MIT-IBM Watson AI Lab)</i>	
TrajPAC: Towards Robustness Verification of Pedestrian Trajectory Prediction Models	8293
<i>Liang Zhang (Institute of Software, Chinese Academy of Sciences), Nathaniel Xu (Institute of Software Chinese Academy of Sciences), Pengfei Yang (Institute of Software, Chinese Academy of Sciences), Gaojie Jin (University of Liverpool), Cheng-Chao Huang (Nanjing Institute of Software Technology, ISCAS), and Lijun Zhang (Institute of Software, Chinese Academy of Sciences)</i>	

VAD: Vectorized Scene Representation for Efficient Autonomous Driving	8306
<i>Bo Jiang (Huazhong University of Science and Technology), Shaoyu Chen (Huazhong University of Science and Technology), Qing Xu (Horizon Robotics), Bencheng Liao (Huazhong University of Science and Technology), Jiajie Chen (Horizon Robotics), Helong Zhou (Horizon Robotics), Qian Zhang (Horizon Robotics), Wenyu Liu (Huazhong University of Science and Technology), Chang Huang (Horizon Robotics), and Xinggong Wang (Huazhong University of Science and Technology)</i>	
Traj-MAE: Masked Autoencoders for Trajectory Prediction	8317
<i>Hao Chen (The Chinese University of Hong Kong), Jiaze Wang (The Chinese University of Hong Kong), Kun Shao (Huawei Noah's Ark Lab), Furui Liu (Zhejiang Lab), Jianye Hao (Tianjin University), Chenyong Guan (Huazhong University of Science & Technology), Guangyong Chen (The Chinese University of Hong Kong), and Pheng-Ann Heng (The Chinese University of Hong Kong)</i>	
Sparse Point Guided 3D Lane Detection	8329
<i>Chengtang Yao (Beijing Institute of Technology), Lidong Yu (Deeproute), Yuwei Wu (Beijing Institute of Technology, (BIT), and Yunde Jia (Beijing Institute of Technology)</i>	
A Simple Vision Transformer for Weakly Semi-supervised 3D Object Detection	8339
<i>Dingyuan Zhang (Huazhong University of Science and Technology), Dingkan Liang (Huazhong University of Science and Technology), Zhikang Zou (Huazhong university of science and technology), Jingyu Li (Huazhong University of Science and Technology), Xiaoqing Ye (Baidu Inc.), Zhe Liu (Huazhong University of Science and Technology), Xiao Tan (Baidu Inc.), and Xiang Bai (Huazhong University of Science and Technology)</i>	
Learn TAROT with MENTOR: A Meta-Learned Self-Supervised Approach for Trajectory Prediction....	8350
<i>Mozghan Pourkeshavarz (Huawei), Changhe Chen (Huawei Technologies Canada Co., Ltd), and Amir Rasouli (Huawei Technologies Canada)</i>	
FocalFormer3D: Focusing on Hard Instance for 3D Object Detection	8360
<i>Yilun Chen (Chinese University of Hong Kong), Zhiding Yu (NVIDIA), Yukang Chen (The Chinese University of Hong Kong), Shiyi Lan (NVIDIA CORPORATION), Anima Anandkumar (Caltech), Jiaya Jia (Chinese University of Hong Kong), and Jose M. Alvarez (NVIDIA)</i>	
Scene as Occupancy	8372
<i>Wenwen Tong (SenseTime), Chonghao Sima (Purdue University), Tai Wang (The Chinese University of Hong Kong), Li Chen (Shanghai AI Laboratory), Silei Wu (sensetime), Hanming Deng (SenseTime), Yi Gu (Sun Yat-Sen University), Lewei Lu (SenseTime Research), Ping Luo (The University of Hong Kong), Dahua Lin (The Chinese University of Hong Kong), and Hongyang Li (Shanghai AI Lab / Shanghai Jiao Tong University)</i>	
Real-Time Neural Rasterization for Large Scenes	8382
<i>Jeffrey Yunfan Liu (University of Waterloo), Yun Chen (Waabi / University of Toronto), Ze Yang (University of Toronto), Jingkan Wang (University of Toronto), Sivabalan Manivasagam (University of Toronto), and Raquel Urtasun (Waabi)</i>	

A Game of Bundle Adjustment - Learning Efficient Convergence	8394
<i>Amir Belder (Technion), Refael Vivanti (Meta ltd.), and Ayellet Tal (Technion)</i>	
Efficient Transformer-based 3D Object Detection with Dynamic Token Halting	8404
<i>Mao Ye (The University of Texas at Austin), Gregory P. Meyer (Cruise), Yuning Chai (Alphabet), and Qiang Liu (UT Austin)</i>	
RegFormer: An Efficient Projection-Aware Transformer Network for Large-Scale Point Cloud Registration	8417
<i>Jiuming Liu (Shanghai Jiao Tong University), Guangming Wang (University of Cambridge), Zhe Liu (University of Cambridge), Chaokang Jiang (China University of Mining and Technology), Marc Pollefeys (ETH Zurich / Microsoft), and Hesheng Wang (SJTU)</i>	
CASSPR: Cross Attention Single Scan Place Recognition	8427
<i>Yan Xia (Technical University of Munich), Mariia Gladkova (Technical University of Munich), Rui Wang (Microsoft), Qianyun Li (Technical University of Munich), Uwe Stilla (Technical University of Munich), Joao F. Henriques (University of Oxford), and Daniel Cremers (TU Munich)</i>	
Recursive Video Lane Detection	8439
<i>Dongkwon Jin (Korea university), Dahyun Kim (Korea University), and Chang-Su Kim (Korea university)</i>	
Parametric Depth Based Feature Representation Learning for Object Detection and Segmentation in Bird’s-Eye View	8449
<i>Jiayu Yang (The Australian National University), Enze Xie (The University of Hong Kong), Miaomiao Liu (The Australian National University), and Jose M. Alvarez (NVIDIA)</i>	
SHIFT3D: Synthesizing Hard Inputs For Tricking 3D Detectors	8459
<i>Hongge Chen (Cruise), Zhao Chen (Cruise LLC), Gregory P. Meyer (Cruise), Dennis Park (Cruise LLC), Carl Vondrick (Columbia University), Ashish Shrivastava (Apple), and Yuning Chai (Alphabet)</i>	
Bootstrap Motion Forecasting With Self-Consistent Constraints	8470
<i>Maosheng Ye (HKUST), Jiamiao Xu (Deeproute.AI), Xunnong Xu (Deeproute.ai), Tengfei Wang (HKUST), Tongyi Cao (University of Massachusetts Amherst), and Qifeng Chen (HKUST)</i>	
Towards Viewpoint Robustness in Bird’s Eye View Segmentation	8481
<i>Tzofi Klinghoffer (Massachusetts Institute of Technology), Jonah Philion (University of Toronto, NVIDIA), Wenzheng Chen (University of Toronto), Or Litany (NVIDIA), Zan Gojcic (NVIDIA), Jungseock Joo (NVIDIA, University of California Los Angeles), Ramesh Raskar (Massachusetts Institute of Technology), Sanja Fidler (University of Toronto, NVIDIA), and Jose M. Alvarez (NVIDIA)</i>	
R-Pred: Two-Stage Motion Prediction Via Tube-Query Attention-Based Trajectory Refinement	8491
<i>Sehwan Choi (Hanyang University), Junggho Kim (Hanyang University), Junyong Yun (Hanyang University), and Jun Won Choi (Hanyang University)</i>	

INT2: Interactive Trajectory Prediction at Intersections	8502
<i>Zhijie Yan (Beihang University), Pengfei Li (Tsinghua University), Zheng Fu (Tsinghua University), Shaocong Xu (Xiamen University), Yongliang Shi (Tsinghua University), Xiaoxue Chen (Tsinghua University), Yuhang Zheng (Beihang University), Yang Li (Tsinghua University), Tianyu Liu (The Hong Kong University of Science and Technology), Chuxuan Li (Tsinghua University), Nairui Luo (Baidu Inc.), Xu Gao (Baidu Inc), Yilun Chen (Tsinghua University), Zuoxu Wang (Beihang University), Yifeng Shi (Baidu Inc), Pengfei Huang (Institute for AI Industry Research, (AIR), Zhengxiao Han (Beijing University of Chemical Technology), Jirui Yuan (Tsinghua University), Jiangtao Gong (Tsinghua University), Guyue Zhou (Tsinghua University), Hang Zhao (Tsinghua University), and Hao Zhao (Tsinghua University)</i>	
MatrixVT: Efficient Multi-Camera to BEV Transformation for 3D Perception	8514
<i>Hongyu Zhou (Megvii), Zheng Ge (MEGVII Technology), Zeming Li (Bytedance Inc), and Xiangyu Zhang (Megvii Technology)</i>	
Unsupervised Self-Driving Attention Prediction via Uncertainty Mining and Knowledge Embedding	8524
<i>Pengfei Zhu (Beijing University of Posts and Telecommunications), Mengshi Qi (Beijing University of Posts and Telecommunications), Xia Li (Peking University), Weijian Li (Amazon), and Huadong Ma (Beijing University of Posts and Telecommunications)</i>	
SVQNet: Sparse Voxel-Adjacent Query Network for 4D Spatio-Temporal LiDAR Semantic Segmentation	8535
<i>Xuechao Chen (Tsinghua University), Shuangjie Xu (HKUST), Xiaoyi Zou (deeproute.ai), Tongyi Cao (University of Massachusetts Amherst), Dit-Yan Yeung (HKUST), and Lu Fang (Tsinghua University)</i>	
MotionLM: Multi-Agent Motion Forecasting as Language Modeling	8545
<i>Ari Seff (Waymo), Brian Cera (Waymo), Dian Chen (UT Austin), Mason Ng (Waymo), Aurick Zhou (UC Berkeley), Nigamaa Nayakanti (Waymo), Khaled S. Refaat (Waymo), Rami Al-Rfou (Waymo Research), and Benjamin Sapp (Waymo)</i>	
Improving Online Lane Graph Extraction by Object-Lane Clustering	8557
<i>Yigit Baran Can (ETH Zurich), Alexander Liniger (ETH Zurich), Danda Pani Paudel (ETH Zürich, INSAIT Sofia), and Luc Van Gool (ETH Zurich)</i>	
Unsupervised 3D Perception with 2D Vision-Language Distillation for Autonomous Driving	8568
<i>Mahyar Najibi (Waymo), Jingwei Ji (Waymo), Yin Zhou (Waymo), Charles R. Qi (Waymo LLC), Xinchen Yan (Waymo), Scott Ettinger (Waymo), and Dragomir Anguelov (Waymo)</i>	
Self-Supervised Monocular Depth Estimation by Direction-aware Cumulative Convolution Network	8579
<i>Wencheng Han (University of Macau), Junbo Yin (Beijing Institute of Technology), and Jianbing Shen (University of Macau)</i>	
Ordered Atomic Activity for Fine-grained Interactive Traffic Scenario Understanding	8590
<i>Nakul Agarwal (Honda Research Institute USA) and Yi-Ting Chen (National Yang Ming Chiao Tung University)</i>	

DistillBEV: Boosting Multi-Camera 3D Object Detection with Cross-Modal Knowledge Distillation	8603
<i>Zeyu Wang (University of California, Santa Cruz), Dingwen Li (QCraft Inc.), Chenxu Luo (Johns Hopkins University), Cihang Xie (University of California, Santa Cruz), and Xiaodong Yang (NVIDIA Research)</i>	
Video Task Decathlon: Unifying Image and Video Tasks in Autonomous Driving	8613
<i>Thomas E. Huang (ETH Zürich), Yifan Liu (University of Adelaide), Luc Van Gool (ETH Zurich), and Fisher Yu (ETH Zurich)</i>	
MV-Map: Offboard HD-Map Generation with Multi-view Consistency	8624
<i>ZiYang Xie (FUDAN University), Ziqi Pang (UIUC), and Yu-Xiong Wang (University of Illinois at Urbana-Champaign)</i>	
Towards Universal LiDAR-Based 3D Object Detection by Multi-Domain Knowledge Transfer	8635
<i>Guile Wu (Noah's Ark Lab), Tongtong Cao (Huawei Technologies Canada), Bingbing Liu (Huawei Noah's Ark Lab, Canada), Xingxin Chen (Huawei Technologies Canada), and Yuan Ren (Huawei)</i>	
Forecast-MAE: Self-supervised Pre-training for Motion Forecasting with Masked Autoencoders...	8645
<i>Jie Cheng (HKUST), Xiaodong Mei (The Hong Kong University of Science and Technology), and Ming Liu (HKUST)</i>	
UniFusion: Unified Multi-View Fusion Transformer for Spatial-Temporal Representation in Bird's-Eye-View	8656
<i>Zeun Qin (Zhejiang University), Jingyu Chen (DJI), Chao Chen (DJI), Xiaozhi Chen (DJI), and Xi Li (Zhejiang University)</i>	
BEVPlace: Learning LiDAR-based Place Recognition using Bird's Eye View Images	8666
<i>Lun Luo (Zhejiang University, HAOMO.AI Technology Co., Ltd.), Shuhang Zheng (Zhejiang University), Yixuan Li (Zhejiang University), Yongzhi Fan (Zhejiang University), Beinan Yu (Zhejiang University), Si-Yuan Cao (Zhejiang University), Junwei Li (Zhejiang University), and Hui-Liang Shen (Zhejiang University)</i>	
CORE: Cooperative Reconstruction for Multi-Agent Perception	8676
<i>Binglu Wang (Northwestern Polytechnical University), Lei Zhang (Northwestern Polytechnical University), Zhaozhong Wang (NWPU), Yongqiang Zhao (Northwestern Polytechnical University), and Tianfei Zhou (ETH Zurich)</i>	
MetaBEV: Solving Sensor Failures for 3D Detection and Map Segmentation	8687
<i>Chongjian GE (The University of Hong Kong, (HKU), Junsong Chen (Dalian University of Technology), Enze Xie (The University of Hong Kong), Zhongdao Wang (Tsinghua University), Lanqing Hong (Huawei Noah's Ark Lab), Huchuan Lu (Dalian University of Technology), Zhenguo Li (Huawei Noah's Ark Lab), and Ping Luo (The University of Hong Kong)</i>	

3D from a Single Image and Shape-From-X

Aggregating Feature Point Cloud for Depth Completion	8698
<i>Zhu Yu (Zhejiang University), Zehua Sheng (Zhejiang University), Zili Zhou (Zhejiang University), Lun Luo (Zhejiang University, HAOMO.AI Technology Co., Ltd.), Si-Yuan Cao (Zhejiang University), Hong Gu (vivo), Huaqi Zhang (vivo Mobile Communication Co., Ltd.), and Hui-Liang Shen (Zhejiang University)</i>	

Coordinate Transformer: Achieving Single-stage Multi-person Mesh Recovery from Videos	8710
<i>Haoyuan Li (Sun Yet-sen University), Haoye Dong (Carnegie Mellon University), Hanchao Jia (Samsung Research China – Beijing, (SRC-B), Dong Huang (Carnegie Mellon University), Michael C. Kampffmeyer (UiT The Arctic University of Norway), Liang Lin (Sun Yat-sen University), and Xiaodan Liang (Sun Yat-sen University)</i>	
MAMo: Leveraging Memory and Attention for Monocular Video Depth Estimation	8720
<i>Rajeev Yasarla (Qualcomm AI Research), Hong Cai (Qualcomm AI Research), Jisoo Jeong (Qualcomm AI Research), Yunxiao Shi (Qualcomm AI Research), Risheek Garrepalli (Qualcomm AI Research), and Fatih Porikli (Australian National University)</i>	
SlaBins: Fisheye Depth Estimation using Slanted Bins on Road Environments	8731
<i>Jongsung Lee (UNIST), Gyeongsu Cho (UNIST), Jeongin Park (UNIST), Kyongjun Kim (UNIST), Seongoh Lee (UNIST), Jung-Hee Kim (42dot), Seong-Gyun Jeong (42dot.ai), and Kyungdon Joo (UNIST)</i>	
Creative Birds: Self-Supervised Single-View 3D Style Transfer	8741
<i>Renke Wang (Nanjing University of Science and Technology), Guimin Que (Nanjing University of Science and Technology), Shuo Chen (RIKEN), Xiang Li (Nankai University), Jun Li (Nanjing University of Science and Technology), and Jian Yang (Nanjing University of Science and Technology)</i>	
Dynamic PlenOctree for Adaptive Sampling Refinement in Explicit NeRF	8751
<i>Haotian Bai (Hong Kong University of Science and Technology, (Guangzhou), Yiqi Lin (Individual Researcher), Yize Chen (Hong Kong University of Science and Technology, (Guangzhou), and Lin Wang (HKUST)</i>	
CORE: Co-planarity Regularized Monocular Geometry Estimation with Weak Supervision	8762
<i>Yuguang Li (Samsung R&D Institute China XIAN), Kai Wang (samsung), Hui Li (Samsung Research and Development Institute China Xi'an), Seon-Min Rhee (Samsung Electronics), Seungju Han (Samsung Advanced Institute of Technology), Jihye Kim (Samsung Advanced Institute of Technology), Min Yang (Samsung R&D Institute China Xi'an, (SRCX), Ran Yang (Samsung R&D Institute China Xi'an, (SRCX), and Feng Zhu (Samsung)</i>	
Relightify: Relightable 3D Faces from a Single Image via Diffusion Models	8772
<i>Foivos Paraperas Papantoniou (Imperial College London), Alexandros Lattas (Imperial College London), Stylianos Moschoglou (Imperial College London), and Stefanos Zafeiriou (Imperial College London)</i>	
GLA-GCN: Global-local Adaptive Graph Convolutional Network for 3D Human Pose Estimation from Monocular Video	8784
<i>Bruce X.B. Yu (The Hong Kong Polytechnic University), Zhi Zhang (Hong Kong Polytechnic University), Yongxu Liu (The Hong Kong Polytechnic University), Sheng-hua Zhong (Shenzhen University), Yan Liu (The Hong Kong Polytechnic University), and Chang Wen Chen (The Hong Kong Polytechnic University)</i>	
Calibrating Panoramic Depth Estimation for Practical Localization and Mapping	8796
<i>Junho Kim (Seoul National University), Eun Sun Lee (Seoul National University), and Young Min Kim (Seoul National University)</i>	

SimNP: Learning Self-Similarity Priors Between Neural Points	8807
<i>Christopher Wewer (Max Planck Institute for Informatics), Eddy Ilg (Saarland University), Bernt Schiele (MPI Informatics), and Jan Eric Lenssen (MPI Informatics)</i>	
AGG-Net: Attention Guided Gated-Convolutional Network for Depth Image Completion	8819
<i>Dongyue Chen (Northeastern University), Tingxuan Huang (Northeastern University), Zhimin Song (Northeastern University(China), Shizhuo Deng (Northeastern University), and Tong Jia (Northeastern University)</i>	
Viewset Diffusion: (0-)Image-Conditioned 3D Generative Models from 2D Data	8829
<i>Stanislaw Szymanowicz (University of Oxford), Christian Rupprecht (University of Oxford), and Andrea Vedaldi (Oxford University)</i>	
CVSformer: Cross-View Synthesis Transformer for Semantic Scene Completion	8840
<i>Haotian Dong (Tianjin University), Enhui Ma (Tianjin University), Lubo Wang (Tianjin University), Miaohui Wang (Shenzhen University), Wuyuan Xie (Shenzhen University), Qing Guo (A*STAR), Ping Li (The Hong Kong Polytechnic University), Lingyu Liang (South China University of Technology), Kairui Yang (alibaba damo academy), and Di Lin (Tianjin University)</i>	
U-RED: Unsupervised 3D Shape Retrieval and Deformation for Partial Point Clouds	8850
<i>Yan Di (Technical University of Munich), Chenyangguang Zhang (Tsinghua University), Ruida Zhang (Tsinghua University), Fabian Manhardt (Google), Yongzhi Su (DFKI), Jason Rambach (DFKI), Didier Stricker (DFKI), Xiangyang Ji (Tsinghua University), and Federico Tombari (Google, TU Munich)</i>	
Single Depth-image 3D Reflection Symmetry and Shape Prediction	8862
<i>Zhaoxuan Zhang (Dalian University of Technology), Bo Dong (Princeton University), Tong Li (Dalian University of Technology), Felix Heide (Princeton / Algolux), Pieter Peers (College of William & Mary), Baocai Yin (Beijing University of Technology), and Xin Yang (Dalian University of Technology)</i>	
Self-supervised Monocular Depth Estimation: Let's Talk About The Weather	8873
<i>Kieran Saunders (Aston University), George Vogiatzis (Loughborough University), and Luis J. Manso (Aston University)</i>	
Mesh2Tex: Generating Mesh Textures from Image Queries	8884
<i>Aleksei Bokhovkin (TUM), Shubham Tulsiani (Carnegie Mellon University), and Angela Dai (Technical University of Munich)</i>	
Sketch and Text Guided Diffusion Model for Colored Point Cloud Generation	8895
<i>Zijie Wu (Hunan University), Yaonan Wang (Hunan University), Mingtao Feng (Xidian University), He Xie (Hunan University), and Ajmal Mian (University of Western Australia)</i>	
Learning a Room with the Occ-SDF Hybrid: Signed Distance Function Mingled with Occupancy Aids Scene Representation	8906
<i>Xiaoyang Lyu (The University of Hong Kong), Peng Dai (The University of Hong Kong), Zizhang Li (Zhejiang University), Dongyu Yan (Harbin Institute of Technology), Yi Lin (DJI), Yifan Peng (The University of Hong Kong), and Xiaojuan Qi (The University of Hong Kong)</i>	

Robust Geometry-Preserving Depth Estimation Using Differentiable Rendering	8917
<i>Chi Zhang (Nanyang Technological University), Wei Yin (DJI), Gang Yu (Tencent), Zhibin Wang (Tencent), Tao Chen (Fudan University), Bin Fu (Tencent), Joey Tianyi Zhou (A*STAR Centre for Frontier AI Research, (CFAR), and Chunhua Shen (Zhejiang University)</i>	
FeatureNeRF: Learning Generalizable NeRFs by Distilling Foundation Models	8928
<i>Jianglong Ye (UCSD), Naiyan Wang (TuSimple), and Xiaolong Wang (UCSD)</i>	
One-shot Implicit Animatable Avatars with Model-based Priors	8940
<i>Yangyi Huang (Zhejiang University), Hongwei Yi (Max Planck Institute for Intelligent Systems), Weiyang Liu (University of Cambridge), Haofan Wang (Carnegie Mellon University), Boxi Wu (Zhejiang University), Wenxiao Wang (Zhejiang University), Binbin Lin (Zhejiang University), Debing Zhang (Xiaohongshu), and Deng Cai (ZJU)</i>	
VeRi3D: Generative Vertex-based Radiance Fields for 3D Controllable Human Image Synthesis ..	8952
<i>Xinya Chen (Zhejiang University), Jiaxin Huang (Zhejiang University), Yanrui Bin (HUST), Lu Yu (Zhejiang University), and Yiyi Liao (Zhejiang University)</i>	
Diffuse3D: Wide-Angle 3D Photography via Bilateral Diffusion	8964
<i>Yutao Jiang (South China University of Technology), Yang Zhou (South China University of Technology), Yuan Liang (South China University of Technology), Wenxi Liu (Fuzhou University), Jianbo Jiao (University of Birmingham; University of Oxford), Yuhui Quan (South China University of Technology), and Shengfeng He (Singapore Management University)</i>	
AutoSynth: Learning to Generate 3D Training Data for Object Point Cloud Registration	8975
<i>Zheng Dang (EPFL) and Mathieu Salzmann (EPFL)</i>	
Body Knowledge and Uncertainty Modeling for Monocular 3D Human Body Reconstruction	8986
<i>Yufei Zhang (Rensselaer Polytechnic Institute), Hanjing Wang (Rensselaer Polytechnic Institute), Jeffrey O. Kephart (IBM Research, USA), and Qiang Ji (Rensselaer Polytechnic Institute)</i>	
Accurate 3D Face Reconstruction with Facial Component Tokens	8999
<i>Tianke Zhang (Tsinghua University), Xuangeng Chu (IDEA), Yunfei Liu (International Digital Economy Academy), Lijian Lin (Tencent), Zhendong Yang (Graduate school at ShenZhen, Tsinghua university), Zhengzhuo Xu (Tsinghua University), Chengkun Cao (Peking University), Fei Yu (Vistring), Changyin Zhou (Vistring), Chun Yuan (Graduate school at ShenZhen, Tsinghua university), and Yu Li (International Digital Economy Academy)</i>	
Metric3D: Towards Zero-shot Metric 3D Prediction from A Single Image	9009
<i>Wei Yin (DJI), Chi Zhang (Nanyang Technological University), Hao Chen (Zhejiang University), Zhipeng Cai (Intel Labs), Gang Yu (Tencent), Kaixuan Wang (Self Research), Xiaozhi Chen (DJI), and Chunhua Shen (University of Adelaide, Australia)</i>	
Reconstructing Interacting Hands with Interaction Prior from Monocular Images	9020
<i>Binghui Zuo (Southeast University), Zimeng Zhao (Southeast University), Wenqian Sun (Southeast University), Wei Xie (Southeast University), Zhou Xue (Bytedance), and Yangang Wang (Southeast University)</i>	

SparseNeRF: Distilling Depth Ranking for Few-shot Novel View Synthesis	9031
<i>Guangcong Wang (Nanyang Technological University), Zhaoxi Chen (Nanyang Technological University), Chen Change Loy (Nanyang Technological University), and Ziwei Liu (Nanyang Technological University)</i>	
Beyond the Limitation of Monocular 3D Detector via Knowledge Distillation	9043
<i>Yiran Yang (Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100190, China), Dongshuo Yin (UCAS), Xuee Rong (Aerospace Information Research Institute Chinese Academy of Sciences), Xian Sun (Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100190, China), Wenhui Diao (Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100190, China), and Xinming Li (Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100190, China)</i>	
HiFace: High-Fidelity 3D Face Reconstruction by Learning Static and Dynamic Details	9053
<i>Zenghao Chai (National University of Singapore), Tianke Zhang (Tsinghua University), Tianyu He (Microsoft Research Asia), Xu Tan (Microsoft Research Asia), Tadas Baltrusaitis (Microsoft), HsiangTao Wu (Microsoft), Runnan Li (Microsoft), Sheng Zhao (Microsoft), Chun Yuan (Graduate school at ShenZhen, Tsinghua university), and Jiang Bian (Microsoft Research)</i>	
Animal3D: A Comprehensive Dataset of 3D Animal Pose and Shape	9065
<i>Jiacong Xu (Johns Hopkins University), Yi Zhang (Johns Hopkins University), Jiawei Peng (Johns Hopkins University), Wufei Ma (Johns Hopkins University), Artur Jesslen (University of Freiburg), Pengliang Ji (Beihang University), Qixin Hu (Huazhong University of Science and Technology), Jiehua Zhang (University of Oulu), Qihao Liu (Johns Hopkins University), Jiahao Wang (Johns Hopkins University), Wei Ji (University of Alberta), Chen Wang (Tsinghua University), Xiaoding Yuan (Johns Hopkins University), Prakhar Kaushik (Johns Hopkins University), Guofeng Zhang (UCLA), Jie Liu (City University of Hong Kong), Yushan Xie (East China Normal University), Yawen Cui (University of Oulu), Alan Yuille (Johns Hopkins University), and Adam Kortylewski (Max Planck Institute for Informatics)</i>	
JOTR: 3D Joint Contrastive Learning with Transformers for Occluded Human Mesh Recovery	9076
<i>Jiahao Li (Zhejiang University), Zongxin Yang (Zhejiang University), Xiaohan Wang (Zhejiang University), Jianxin Ma (Alibaba Group), Chang Zhou (Alibaba Group), and Yi Yang (Zhejiang University)</i>	
D-IF: Uncertainty-aware Human Digitization via Implicit Distribution Field	9088
<i>Xueting Yang (Psyche.ai), Yihao Luo (Department of Bioengineering, Imperial College London), Yuliang Xiu (Max Planck Institute for Intelligent Systems), Wang Wei (BUPT), Hao Xu (HKUST), and Zhaoxin Fan (Renmin University of China)</i>	
3D Distillation: Improving Self-Supervised Monocular Depth Estimation on Reflective Surfaces	9099
<i>Xuepeng Shi (Imperial College London), Georgi Dikov (Qualcomm), Gerhard Reitmayr (Qualcomm), Tae-Kyun Kim (KAIST/Imperial College London), and Mohsen Ghafoorian (Qualcomm)</i>	

DeformToon3D: Deformable Neural Radiance Fields for 3D Toonification	9110
<i>Junzhe Zhang (Nanyang Technological University), Yushi Lan (Nanyang Technological University), Shuai Yang (Nanyang Technological University), Fangzhou Hong (Nanyang Technological University), Quan Wang (SenseTime), Chai Kiat Yeo (Nanyang Technological University), Ziwei Liu (Nanyang Technological University), and Chen Change Loy (Nanyang Technological University)</i>	
MonoDETR: Depth-guided Transformer for Monocular 3D Object Detection	9121
<i>Renrui Zhang (Shanghai AI Lab), Han Qiu (Nanyang Technological University), Tai Wang (The Chinese University of Hong Kong), Ziyu Guo (Peking University), Ziteng Cui (The University of Tokyo), Yu Qiao (Shanghai AI Laboratory), Hongsheng Li (The Chinese University of Hong Kong), and Peng Gao (Chinese university of hong kong)</i>	
ReLeaPS : Reinforcement Learning-based Illumination Planning for Generalized Photometric Stereo	9133
<i>Jun Hoong Chan (Peking University), Bohan Yu (Peking University), Heng Guo (Beijing University of Posts and Telecommunications; Osaka University), Jieji Ren (Shanghai Jiao Tong University), Zongqing Lu (Peking University), and Boxin Shi (Peking University)</i>	
Convex Decomposition of Indoor Scenes	9142
<i>Vaibhav Vavilala (University of Illinois at Urbana-Champaign) and David Forsyth (University of Illinois Urbana-Champaign)</i>	
NeO 360: Neural Fields for Sparse View Synthesis of Outdoor Scenes	9153
<i>Muhammad Zubair Irshad (Georgia Institute of Technology), Sergey Zakharov (Toyota Research Institute), Katherine Liu (Toyota Research Institute), Vitor Guizilini (Toyota Research Institute), Thomas Kollar (Toyota Research Institute), Adrien Gaidon (Toyota Research Institute), Zsolt Kira (Georgia Institute of Technology), and Rares Ambrus (Toyota Research Institute)</i>	
UrbanGIRAFFE: Representing Urban Scenes as Compositional Generative Neural Feature Fields	9165
<i>Yuanbo Yang (Zhejiang University), Yifei Yang (Zhejiang University), Hanlei Guo (Zhejiang University), Rong Xiong (Zhejiang University), Yue Wang (Zhejiang University), and Yiyi Liao (Zhejiang University)</i>	
Efficient Converted Spiking Neural Network for 3D and 2D Classification	9177
<i>Yuxiang Lan (Xiamen University), Yachao Zhang (Tsinghua University), Xu Ma (Northeastern University), Yanyun Qu (XMU), and Yun Fu (Northeastern University)</i>	
Distribution-Aligned Diffusion for Human Mesh Recovery	9187
<i>Lin Geng Foo (Singapore University of Technology and Design), Jia Gong (Singapore University of Technology and Design), Hossein Rahmani (Lancaster University), and Jun Liu (Singapore University of Technology and Design)</i>	
Towards Zero-Shot Scale-Aware Monocular Depth Estimation	9199
<i>Vitor Guizilini (Toyota Research Institute), Igor Vasiljevic (Toyota Research Institute), Dian Chen (Toyota Research Institute), Rares Ambrus Rareş Ambruş (Toyota Research Institute), and Adrien Gaidon (Toyota Research Institute)</i>	

Learning Depth Estimation for Transparent and Mirror Surfaces	9210
<i>Alex Costanzino (University of Bologna), Pierluigi Zama Ramirez (University of Bologna), Matteo Poggi (University of Bologna), Fabio Tosi (University of Bologna), Stefano Mattocchia (University of Bologna), and Luigi Di Stefano (University of Bologna)</i>	
Uni-3D: A Universal Model for Panoptic 3D Scene Reconstruction	9222
<i>Xiang Zhang (University of California San Diego), Zeyuan Chen (UC San Diego), Fangyin Wei (Princeton University), and Zhuowen Tu (UC San Diego)</i>	
3D VR Sketch Guided 3D Shape Prototyping and Exploration	9233
<i>Ling Luo (University of Surrey), Pinaki Nath Chowdhury (University of Surrey), Tao Xiang (University of Surrey), Yi-Zhe Song (University of Surrey), and Yulia Gryaditskaya (University of Surrey, CVSSP, Surrey Institute for People-Centred AI)</i>	
Transparent Shape from a Single View Polarization Image	9243
<i>Mingqi Shao (Graduate school at ShenZhen, Tsinghua university), Chongkun Xia (Tsinghua University), Zhendong Yang (Graduate school at ShenZhenTsinghua university), Junnan Huang (Tsinghua University), and Xueqian Wang (Tsinghua University)</i>	
Get3DHuman: Lifting StyleGAN-Human into a 3D Generative Model Using Pixel-Aligned Reconstruction Priors	9253
<i>Zhangyang Xiong (The Chinese University of Hong Kong, Shenzhen), Di Kang (Tencent), Derong Jin (The Chinese University of Hong Kong, Shenzhen), Weikai Chen (Tencent America), Linchao Bao (Tencent AI Lab), Shuguang Cui (The Chinese University of Hong Kong, Shenzhen), and Xiaoguang Han (Shenzhen Research Institute of Big Data, the Chinese University of Hong Kong, (Shenzhen))</i>	
Zero-1-to-3: Zero-shot One Image to 3D Object	9264
<i>Ruoshi Liu (Columbia University), Rundi Wu (Columbia University), Basile Van Hoorick (Columbia University), Pavel Tokmakov (Toyota Research Institute), Sergey Zakharov (Toyota Research Institute), and Carl Vondrick (Columbia University)</i>	
FrozenRecon: Pose-free 3D Scene Reconstruction with Frozen Depth Models	9276
<i>Guangkai Xu (University of Science and Technology of China), Wei Yin (DJI), Hao Chen (Zhejiang University), Chunhua Shen (University of Adelaide, Australia), Kai Cheng (University of Science and Technology of China), and Feng Zhao (University of Science and Technology of China)</i>	
LIST: Learning Implicitly from Spatial Transformers for Single-View 3D Reconstruction	9287
<i>Mohammad Samiul Arshad (University of Texas at Arlington) and William J. Beksı (University of Texas at Arlington)</i>	
3DMiner: Discovering Shapes from Large-Scale Unannotated Image Datasets	9297
<i>Ta-Ying Cheng (University of Oxford), Matheus Gadelha (University of Massachusetts Amherst), Sören Pirk (Adobe), Thibault Groueix (Adobe Research), Radomír Měch (Adobe Systems Incorporated), Andrew Markham (University of Oxford), and Niki Trigoni (University of Oxford)</i>	

Nonrigid Object Contact Estimation With Regional Unwrapping Transformer	9308
<i>Wei Xie (Southeast University), Zimeng Zhao (Southeast University), Shiyang Li (Southeast University), Binghui Zuo (Southeast University), and Yangang Wang (Southeast University)</i>	
SHERF: Generalizable Human NeRF from a Single Image	9318
<i>Shoukang Hu (Nanyang Technological University), Fangzhou Hong (Nanyang Technological University), Liang Pan (Nanyang Technological University), Haiyi Mei (SenseTime Group Limited), Lei Yang (SenseTime Group Limited), and Ziwei Liu (Nanyang Technological University)</i>	
Full-Body Articulated Human-Object Interaction	9331
<i>Nan Jiang (Peking University), Tengyu Liu (Beijing Institute for General Artificial Intelligence), Zhexuan Cao (Tsinghua University), Jieming Cui (Beijing Institute for General Artificial Intelligence), Zhiyuan Zhang (Tsinghua University), Yixin Chen (UCLA), He Wang (Peking University), Yixin Zhu (Peking University), and Siyuan Huang (Beijing Institute for General Artificial Intelligence)</i>	
PlaneRecTR: Unified Query Learning for 3D Plane Recovery from a Single View	9343
<i>Jingjia Shi (National University of Defense Technology), Shuailong Zhi (National University of Defense Technology), and Kai Xu (National University of Defense Technology)</i>	
SceneRF: Self-Supervised Monocular 3D Scene Reconstruction with Radiance Fields	9353
<i>Anh-Quan Cao (Inria) and Raoul de Charette (Inria)</i>	
3D-Aware Neural Body Fitting for Occlusion Robust 3D Human Pose Estimation	9365
<i>Yi Zhang (Johns Hopkins University), Pengliang Ji (Beihang University), Angtian Wang (Johns Hopkins University), Jieru Mei (Johns Hopkins University), Adam Kortylewski (Max Planck Institute for Informatics), and Alan Yuille (Johns Hopkins University)</i>	
Two-in-One Depth: Bridging the Gap Between Monocular and Binocular Self-Supervised Depth Estimation	9377
<i>Zhengming Zhou (NLPR-IA-CAS) and Qiulei Dong (NLPR-IA-CAS)</i>	
LRRU: Long-short Range Recurrent Updating Networks for Depth Completion	9388
<i>Yufei Wang (Northwestern Polytechnical University), Bo Li (Northwestern Polytechnical University), Ge Zhang (Northwestern Polytechnical University), Qi Liu (Northwestern Polytechnical University), Tao Gao (Chang'an university), and Yuchao Dai (Northwestern Polytechnical University)</i>	
OccFormer: Dual-path Transformer for Vision-based 3D Semantic Occupancy Prediction	9399
<i>Yunpeng Zhang (PhiGent Robotics), Zheng Zhu (Tsinghua University), and Dalong Du (PhiGent Technology Co., Ltd)</i>	
CHORD: Category-level Hand-held Object Reconstruction via Shape Deformation	9410
<i>Kailin Li (Shanghai Jiao Tong University), Lixin Yang (Shanghai Jiao Tong University), Haoyu Zhen (Shanghai Jiao Tong University), Zenan Lin (South China University of Technology), Xinyu Zhan (Shanghai Jiao Tong University), Licheng Zhong (Shanghai Jiao Tong University), Jian Xu (Nreal), Kejian Wu (Nreal), and Cewu Lu (Shanghai Jiao Tong University)</i>	

NDC-Scene: Boost Monocular 3D Semantic Scene Completion in Normalized Device Coordinates Space	9421
<i>Jiawei Yao (University of Washington), Chuming Li (University of Sydney), Keqiang Sun (Chinese University of Hong Kong), Yingjie Cai (CUHK), Hao Li (Chinese University of Hong Kong), Wanli Ouyang (The University of Sydney), and Hongsheng Li (The Chinese University of Hong Kong)</i>	
Neural Video Depth Stabilizer	9432
<i>Yiran Wang (Huazhong University of Science and Technology), Min Shi (Huazhong University of Science and Technology), Jiaqi Li (Huazhong University of Science and Technology), Zihao Huang (Huazhong University of Science and Technology), Zhiguo Cao (Huazhong University of Science and Technology), Jianming Zhang (Adobe Research), Ke Xian (Nanyang Technological University), and Guosheng Lin (Nanyang Technological University)</i>	
DiLiGenT-Pi: Photometric Stereo for Planar Surfaces with Rich Details – Benchmark Dataset and Beyond	9443
<i>Feishi Wang (Peking University), Jieji Ren (Shanghai Jiao Tong University), Heng Guo (Osaka University), Mingjun Ren (Shanghai Jiao Tong University), and Boxin Shi (Peking University)</i>	

Motion Estimation, Matching and Tracking

TMR: Text-to-Motion Retrieval Using Contrastive 3D Human Motion Synthesis	9454
<i>Mathis Petrovich (Ecole des Ponts), Michael J. Black (Max Planck Institute for Intelligent Systems), and Gül Varol (Ecole des Ponts ParisTech)</i>	
Sequential Texts Driven Cohesive Motions Synthesis with Natural Transitions	9464
<i>Shuai Li (Beihang University), Sisi Zhuang (Beihang University), Wenfeng Song (Beijing Information Science and Technology University), Xinyu Zhang (Beijing Information Science and Technology University), Hejia Chen (Beihang University), and Aimin Hao (Beihang University)</i>	
Auxiliary Tasks Benefit 3D Skeleton-based Human Motion Prediction	9475
<i>Chenxin Xu (Shanghai Jiao Tong University), Robby T. Tan (National University of Singapore), Yuhong Tan (Shanghai Jiaotong University), Siheng Chen (Shanghai Jiao Tong University, Shanghai AI Laboratory), Xinchao Wang (National University of Singapore), and Yan-Feng Wang (Cooperative medianet innovation center of Shanghai Jiao Tong University)</i>	
Explicit Motion Disentangling for Efficient Optical Flow Estimation	9487
<i>Changxing Deng (University of Macau), Ao Luo (Megvii), Haibin Huang (Kuaishou Technology), Shaodan Ma (University of Macau), Jiangyu Liu (megvii inc), and Shuaicheng Liu (UESTC Megvii)</i>	

TrackFlow: Multi-Object tracking with Normalizing Flows	9497
<i>Gianluca Mancusi (University of Modena and Reggio Emilia), Aniello Panariello (University of Modena and Reggio Emilia), Angelo Porrello (University of Modena and Reggio Emilia), Matteo Fabbri (University of Modena and Reggio Emilia), Simone Calderara (University of Modena and Reggio Emilia), and Rita Cucchiara (Università di Modena e Reggio Emilia)</i>	
HumanMAC: Masked Motion Completion for Human Motion Prediction	9510
<i>Ling-Hao Chen (Tsinghua University), JiaWei Zhang (Xidian University), Yewen Li (Nanyang Technological University), Yiren Pang (Xidian University), Xiaobo Xia (The University of Sydney), and Tongliang Liu (The University of Sydney)</i>	
Geometrized Transformer for Self-Supervised Homography Estimation	9522
<i>Jiazhen Liu (Renmin University of China) and Xirong Li (Renmin University of China)</i>	
SemARFlow: Injecting Semantics into Unsupervised Optical Flow Estimation for Autonomous Driving	9532
<i>Shuai Yuan (Duke University), Shuzhi Yu (Duke University), Hannah Kim (Amazon), and Carlo Tomasi (Duke University)</i>	
NeSS-ST: Detecting Good and Stable Keypoints with a Neural Stability Score and the Shi-Tomasi detector	9544
<i>Konstantin Pakulev (Skolkovo Institute of Science and Technology), Alexander Vakhitov (SLAMCore), and Gonzalo Ferrer (Skolkovo Institute of Science and Technology)</i>	
Robust Object Modeling for Visual Tracking	9555
<i>Yidong Cai (Nanjing University), Jie Liu (Nanjing University), Jie Tang (Nanjing University), and Gangshan Wu (Nanjing University)</i>	
Social Diffusion: Long-term Multiple Human Motion Anticipation	9567
<i>Julian Tanke (University of Bonn), Linguang Zhang (Facebook Reality Labs), Amy Zhao (Facebook Reality Labs), Chengcheng Tang (Facebook Reality Labs), Yujun Cai (Nanyang Technological University), Lezi Wang (Rutgers University), Po-Chen Wu (Meta Inc.), Juergen Gall (University of Bonn), and Cem Keskin (Facebook)</i>	
Exploring Lightweight Hierarchical Vision Transformers for Efficient Visual Tracking	9578
<i>Ben Kang (Dalian University of Technology), Xin Chen (Dalian University of Technology), Dong Wang (Dalian University of Technology), Houwen Peng (Microsoft Research), and Huchuan Lu (Dalian University of Technology)</i>	
HMD-NeMo: Online 3D Avatar Motion Generation From Sparse Observations	9588
<i>Sadegh Aliakbarian (Microsoft), Fatemeh Saleh (Microsoft), David Collier (Microsoft), Pashmina Cameron (Microsoft), and Darren Cosker (Microsoft)</i>	
Learning Fine-Grained Features for Pixel-Wise Video Correspondences	9598
<i>Rui Li (University of Science and Technology of China), Shenglong Zhou (University of Science and Technology of China), and Dong Liu (University of Science and Technology of China)</i>	

GAFlow: Incorporating Gaussian Attention into Optical Flow	9608
<i>Ao Luo (Megvii), Fan Yang (G42), Xin Li (AIQ), Lang Nie (Beijing Jiaotong University), Chunyu Lin (Beijing Jiaotong University), Haoqiang Fan (Megvii Incface++), and Shuaicheng Liu (UESTC, Megvii)</i>	
Occ ² Net: Robust Image Matching Based on 3D Occupancy Estimation for Occluded Regions	9618
<i>Miao Fan (megvii), Mingrui Chen (MEGVII), Chen Hu (Megvii Technology), and Shuchang Zhou (MEGVII Technology)</i>	
Locomotion-Action-Manipulation: Synthesizing Human-Scene Interactions in Complex 3D Environments	9629
<i>Jiye Lee (Seoul National University) and Hanbyul Joo (Seoul National University)</i>	
Trajectory Unified Transformer for Pedestrian Trajectory Prediction	9641
<i>Liushuai Shi (Xi'an Jiaotong University), Le Wang (Xi'an Jiaotong University), Sanping Zhou (Xi'an Jiaotong University), and Gang Hua (Wormpex AI Research)</i>	
TMA: Temporal Motion Aggregation for Event-based Optical Flow	9651
<i>Haotian Liu (Tongji University), Guang Chen (Tongji University), Sanqing Qu (Tongji University), Yanping Zhang (Tongji University), Zhijun Li (University of Science and Technology of China), Alois Knoll (Robotics and Embedded Systems), and Changjun Jiang (Tongji University)</i>	
Taming Contrast Maximization for Learning Sequential, Low-latency, Event-based Optical Flow	9661
<i>Federico Paredes-Vallés (Delft University of Technology), Kirk Y. W. Scheper (Sony), Christophe De Wagter (Delft University of Technology), and Guido C. H. E. de Croon (TU Delft)</i>	
GlueStick: Robust Image Matching by Sticking Points and Lines Together	9672
<i>Rémi Pautrat (ETH Zurich), Iago Suárez (New Horizon Technologies), Yifan Yu (ETH Zurich), Marc Pollefeys (ETH Zurich / Microsoft), and Viktor Larsson (Lund University)</i>	
DARTH: Holistic Test-time Adaptation for Multiple Object Tracking	9683
<i>Mattia Segu (ETH Zurich), Bernt Schiele (MPI Informatics), and Fisher Yu (ETH Zurich)</i>	
S-TREK: Sequential Translation and Rotation Equivariant Keypoints for Local Feature Extraction	9694
<i>Emanuele Santellani (Graz University of Technology), Christian Sormann (Graz University of Technology), Mattia Rossi (SONY Europe B.V.), Andreas Kuhn (SONY Europe), and Friedrich Fraundorfer (Graz University of Technology)</i>	
Integrating Boxes and Masks: A Multi-Object Framework for Unified Visual Tracking and Segmentation	9704
<i>Yuanyou Xu (Zhejiang University), Zongxin Yang (Zhejiang University), and Yi Yang (Zhejiang University)</i>	

Robust Frame-to-Frame Camera Rotation Estimation in Crowded Scenes	9718
<i>Fabien Delattre (University of Massachusetts Amherst), David Dirnfeld (University of Massachusetts Amherst), Phat Nguyen (University of Massachusetts Amherst), Stephen Scarano (University of Massachusetts, Amherst), Michael J. Jones (MERL), Pedro Miraldo (MERL Mitsubishi Electric Research Laboratories), and Erik Learned-Miller (University of Massachusetts, Amherst)</i>	
Sparse Instance Conditioned Multimodal Trajectory Prediction	9729
<i>Yonghao Dong (Xi'an Jiaotong University), Le Wang (Xi'an Jiaotong University), Sanping Zhou (Xi'an Jiaotong University), and Gang Hua (Wormpex AI Research)</i>	
PoseDiffusion: Solving Pose Estimation via Diffusion-aided Bundle Adjustment	9739
<i>Jianyuan Wang (University of Oxford), Christian Rupprecht (University of Oxford), and David Novotny (Facebook AI Research)</i>	
3DMOTFormer: Graph Transformer for Online 3D Multi-Object Tracking	9750
<i>Shuxiao Ding (Mercedes-Benz AG), Eike Rehder (Robert Bosch GmbH), Lukas Schneider (Mercedes Benz AG), Marius Cordts (Mercedes-Benz AG), and Juergen Gall (University of Bonn)</i>	
Fast Inference and Update of Probabilistic Density Estimation on Trajectory Prediction	9761
<i>Takahiro Maeda (Toyota Technological Institute) and Norimichi Ukita (TTI-J)</i>	
Supervised Homography Learning with Realistic Dataset Generation	9772
<i>Hai Jiang (Sichuan University), Haipeng Li (UESTC, Megvii), Songchen Han (Sichuan University), Haoqiang Fan (Megvii Inface++), Bing Zeng (University of Electronic Science and Technology of China), and Shuaicheng Liu (UESTC, Megvii)</i>	
Joint-Relation Transformer for Multi-Person Motion Prediction	9782
<i>Qingyao Xu (Shanghai Jiao Tong University), Weibo Mao (SJTU), Jingze Gong (Shanghai Jiao Tong University), Chenxin Xu (Shanghai Jiao Tong University), Siheng Chen (Shanghai Jiao Tong University, Shanghai AI Laboratory), Weidi Xie (Shanghai Jiao Tong University), Ya Zhang (Cooperative Medianet Innovation Center, Shanghai Jiao Tong University), and Yan-Feng Wang (Cooperative medianet innovation center of Shanghai Jiao Tong University)</i>	
Event-based Temporally Dense Optical Flow Estimation with Sequential Learning	9793
<i>Wachirawit Ponghiran (Purdue University), Chamika Mihiranga Liyanagedera (Purdue university), and Kaushik Roy (Purdue University)</i>	
3D Motion Magnification: Visualizing Subtle Motions from Time-Varying Radiance Fields	9803
<i>Brandon Y. Feng (University of Maryland, College Park), Hadi Alzayer (University of Maryland College Park), Michael Rubinstein (Google), William T. Freeman (MIT), and Jia-Bin Huang (University of Maryland College Park)</i>	
Learning Optical Flow from Event Camera with Rendered Dataset	9813
<i>Xinglong Luo (University of Electronic Science and Technology of China), Kunming Luo (HKUST), Ao Luo (Megvii), Zhengning Wang (University of Electronic Science and Technology of China), Ping Tan (Hong Kong University of Science and Technology), and Shuaicheng Liu (UESTC, Megvii)</i>	

Persistent-Transient Duality: A Multi-Mechanism Approach for Modeling Human-Object Interaction	9824
<i>Hung Tran (Deakin University), Vuong Le (Amazon.com), Svetha Venkatesh (Deakin University), and Truyen Tran (Deakin University)</i>	
Deep Homography Mixture for Single Image Rolling Shutter Correction	9834
<i>Weilong Yan (Department of Electrical and Computer Engineering), Robby T. Tan (National University of Singapore), Bing Zeng (University of Electronic Science and Technology of China), and Shuaicheng Liu (UESTC, Megvii)</i>	
Fast Neural Scene Flow	9844
<i>Xueqian Li (The University of Adelaide), Jianqiao Zheng (University of Adelaide), Francesco Ferroni (Argo AI), Jhony Kaesemodel Pontes (Argo AI), and Simon Lucey (University of Adelaide)</i>	
RLSAC: Reinforcement Learning Enhanced Sample Consensus for End-to-End Robust Estimation	9857
<i>Chang Nie (SJTU), Guangming Wang (University of Cambridge), Zhe Liu (University of Cambridge), Luca Cavalli (ETH Zurich), Marc Pollefeys (ETH Zurich / Microsoft), and Hesheng Wang (SJTU)</i>	
MeMOTR: Long-Term Memory-Augmented Transformer for Multi-Object Tracking	9867
<i>Ruopeng Gao (Nanjing University) and Limin Wang (Nanjing University)</i>	
MBPTrack: Improving 3D Point Cloud Tracking with Memory Networks and Box Priors	9877
<i>Tian-Xing Xu (Tsinghua University), Yuan-Chen Guo (Tsinghua University), Yu-Kun Lai (Cardiff University), and Song-Hai Zhang (Tsinghua University)</i>	
SportsMOT: A Large Multi-Object Tracking Dataset in Multiple Sports Scenes	9887
<i>Yutao Cui (Nanjing University), Chenkai Zeng (Nanjing University), Xiaoyu Zhao (Nanjing University), Yichun Yang (Nanjing University), Gangshan Wu (Nanjing University), and Limin Wang (Nanjing University)</i>	
Heterogeneous Diversity Driven Active Learning for Multi-Object Tracking	9898
<i>Rui Li (BJTU), Baopeng Zhang (BJTU), Jun Liu (Singapore University of Technology and Design), Wei Liu (BJTU), Jian Zhao (Institute of North Electronic Equipment), and Zhu Teng (Beijing Jiaotong University)</i>	
TM2D: Bimodality Driven 3D Dance Generation via Music-Text Integration	9908
<i>Kehong Gong (national university of singapore), Dongze Lian (National University of Singapore), Heng Chang (Tsinghua University), Chuan Guo (University of Alberta), Zihang Jiang (National University of Singapore), Xinxin Zuo (University of Alberta), Michael Bi Mi (Huawei International Pte Ltd), and Xinchao Wang (National University of Singapore)</i>	
Synchronize Feature Extracting and Matching: A Single Branch Framework for 3D Object Tracking	9919
<i>Teli Ma (Shanghai Artificial Intelligence Laboratory), Mengmeng Wang (Zhejiang University), Jimin Xiao (Xi'an Jiaotong-Liverpool University), Huifeng Wu (Hangzhou Dianzi University), and Yong Liu (Zhejiang University)</i>	
Collaborative Tracking Learning for Frame-Rate-Insensitive Multi-Object Tracking	9930
<i>Yiheng Liu (Bytedance Inc.), Junta Wu (ByteDance), and Yi Fu (Bytedance Inc.)</i>	

CiteTracker: Correlating Image and Text for Visual Tracking	9940
<i>Xin Li (Peng Cheng Laboratory), Yuqing Huang (Harbin Institute of Technology, Shenzhen), Zhenyu He (Harbin Institute of Technology, Shenzhen), Yaowei Wang (PengCheng Laboratory), Huchuan Lu (Dalian University of Technology), and Ming-Hsuan Yang (University of California at Merced)</i>	
SINC: Spatial Composition of 3D Human Motions for Simultaneous Action Generation	9950
<i>Nikos Athanasiou (Max Planck Institute for Intelligent Systems), Mathis Petrovich (Ecole des Ponts), Michael J. Black (Max Planck Institute for Intelligent Systems), and Gul Varol (Ecole des Ponts ParisTech)</i>	
Uncertainty-aware Unsupervised Multi-Object Tracking	9962
<i>Kai Liu (Zhejiang University), Sheng Jin (Nanyang Technological University), Zhihang Fu (Alibaba Group), Ze Chen (Zhejiang University), Rongxin Jiang (Zhejiang University), and Jieping Ye (Alibaba DAMO Academy)</i>	
PVT++: A Simple End-to-End Latency-Aware Visual Tracking Framework	9972
<i>Bowen Li (Carnegie Mellon University), Ziyuan Huang (National University of Singapore), Junjie Ye (Tongji University), Yiming Li (New York University), Sebastian Scherer (Carnegie Mellon University), Hang Zhao (Tsinghua University), and Changhong Fu (Tongji University)</i>	
EigenTrajectory: Low-Rank Descriptors for Multi-Modal Trajectory Forecasting	9983
<i>Inhwan Bae (Gwangju Institute of Science and Technology), Jean Oh (Carnegie Mellon University), and Hae-Gon Jeon (GIST)</i>	
RPEFlow: Multimodal Fusion of RGB-PointCloud-Event for Joint Optical Flow and Scene Flow Estimation	9996
<i>Zhexiong Wan (Northwestern Polytechnical University), Yuxin Mao (Northwestern Polytechnical University), Jing Zhang (Australian National University), and Yuchao Dai (Northwestern Polytechnical University)</i>	
Multi-Scale Bidirectional Recurrent Network with Hybrid Correlation for Point Cloud Based Scene Flow Estimation	10007
<i>Wencan Cheng (Sungkyunkwan University) and Jong Hwan Ko (Sungkyunkwan University)</i>	
ReST: A Reconfigurable Spatial-Temporal Graph Model for Multi-Camera Multi-Object Tracking.....	10017
<i>Cheng-Che Cheng (National Tsing Hua University), Min-Xuan Qiu (National Tsing Hua University), Chen-Kuo Chiang (National Chung Cheng University), and Shang-Hong Lai (National Tsing Hua University)</i>	
TAPIR: Tracking Any Point with Per-Frame Initialization and Temporal Refinement	10027
<i>Carl Doersch (DeepMind), Yi Yang (DeepMind), Mel Vecerik (University College London, Deepmind), Dilara Gokay (DeepMind), Ankush Gupta (DeepMind), Yusuf Aytar (DeepMind), Joao Carreira (DeepMind), and Andrew Zisserman (University of Oxford)</i>	

IHNet: Iterative Hierarchical Network Guided by High-Resolution Estimated Information for Scene Flow Estimation	10039
<i>Yun Wang (Huazhong University of Science and Technology), Cheng Chi (Huazhong University of Science and Technology), Min Lin (Huazhong University of Science and Technology), and Xin Yang (Huazhong University of Science and Technology)</i>	
Can Language Models Learn to Listen?	10049
<i>Evonne Ng (UC Berkeley), Sanjay Subramanian (UC Berkeley), Dan Klein (UC Berkeley), Angjoo Kanazawa (University of California Berkeley), Trevor Darrell (UC Berkeley), and Shiry Ginosar (UC Berkeley)</i>	
XVO: Generalized Visual Odometry via Cross-Modal Self-Training	10060
<i>Lei Lai (Boston University), Zhongkai Shangguan (Boston University), Jimuyang Zhang (Boston University), and Eshed Ohn-Bar (Boston University)</i>	
Distracting Downpour: Adversarial Weather Attacks for Motion Estimation	10072
<i>Jenny Schmalfluss (University of Stuttgart), Lukas Mehl (University of Stuttgart), and Andrés Bruhn (University of Stuttgart)</i>	
Foreground-Background Distribution Modeling Transformer for Visual Object Tracking	10083
<i>Dawei Yang (University of Science and Technology of China), Jianfeng He (University of Science and Technology of China), Yinchao Ma (University of Science and Technology of China), Qianjin Yu (University of Science and Technology of China), and Tianzhu Zhang (University of Science and Technology of China)</i>	

Action and Event Understanding

Weakly-Supervised Action Segmentation and Unseen Error Detection in Anomalous Instructional Videos	10094
<i>Reza Ghoddoosian (University of Texas at Arlington), Isht Dwivedi (Honda Research Institute USA), Nakul Agarwal (Honda Research Institute USA), and Behzad Dariush (Honda Research Institute US)</i>	
Diffusion Action Segmentation	10105
<i>Daochang Liu (University of Sydney), Qiyue Li (Peking University), Anh-Dung Dinh (USYD), Tingting Jiang (Peking University), Mubarak Shah (University of Central Florida), and Chang Xu (University of Sydney)</i>	
Audio-Visual Glance Network for Efficient Video Recognition	10116
<i>Muhammad Adi Nugroho (KAIST), Sangmin Woo (KAIST), Sumin Lee (KAIST), and Changick Kim (KAIST)</i>	
Learning from Noisy Pseudo Labels for Semi-Supervised Temporal Action Localization	10126
<i>Kun Xia (Xi'an Jiaotong University), Le Wang (Xi'an Jiaotong University), Sanping Zhou (Xi'an Jiaotong University), Gang Hua (Wormpex AI Research), and Wei Tang (University of Illinois at Chicago)</i>	
Video Action Recognition with Attentive Semantic Units	10136
<i>Yifei Chen (Huawei), Dapeng Chen (Huawei), Ruijin Liu (Xi'an Jiaotong University), Hao Li (Xiamen University), and Wei Peng (Huawei Technologies)</i>	

Masked Motion Predictors are Strong 3D Action Representation Learners	10147
<i>Yunyao Mao (University of Science and Technology of China), Jiajun Deng (The University of Sydney), Wengang Zhou (University of Science and Technology of China), Yao Fang (MUCFC), Wanli Ouyang (The University of Sydney), and Houqiang Li (University of Science and Technology of China)</i>	
Boosting Positive Segments for Weakly-Supervised Audio-Visual Video Parsing	10158
<i>Kranthi Kumar Rachavarapu (Indian Institute of Technology Madras) and Rajagopalan A. N. N. (Indian Institute of Technology Madras)</i>	
Weakly-Supervised Action Localization by Hierarchically-Structured Latent Attention Modeling	10169
<i>Guiqin Wang (Xi'an Jiaotong University), Peng Zhao (Xi'an Jiaotong University), Cong Zhao (Xi'an Jiaotong University), Shusen Yang (Xi'an Jiaotong University), Jie Cheng (Huawei Technology Co., Ltd.), Luziwei Leng (Huawei Technologies Co., Ltd), Jianxing Liao (HUAWEI TECHNOLOGIES CO.LTD), and Qinghai Guo (Huawei Technologies Co., Ltd.)</i>	
Few-Shot Common Action Localization via Cross-Attentional Fusion of Context and Temporal Dynamics	10180
<i>Juntae Lee (Qualcomm AI Research), Mihir Jain (Meta), and Sungrack Yun (Qualcomm AI Research)</i>	
Interaction-aware Joint Attention Estimation Using People Attributes	10190
<i>Chihiro Nakatani (TTI-J), Hiroaki Kawashima (University of Hyogo), and Norimichi Ukita (TTI-J)</i>	
FineDance: A Fine-grained Choreography Dataset for 3D Full Body Dance Generation	10200
<i>Ronghui Li (Tsinghua University), Junfan Zhao (Tsinghua University), Yachao Zhang (Tsinghua University), Mingyang Su (Tsinghua University), Zeping Ren (Tsinghua University), Han Zhang (Xidian University), Yansong Tang (Tsinghua University), and Xiu Li (Tsinghua University)</i>	
SOAR: Scene-debiasing Open-set Action Recognition	10210
<i>Yuanhao Zhai (State University of New York at Buffalo), Ziyi Liu (Xi'an Jiaotong University), Zhenyu Wu (Texas A&M University), Yi Wu (Wormpex AI Research), Chunluan Zhou (Ant Group), David Doermann (University at Buffalo), Junsong Yuan (State University of New York at Buffalo, USA), and Gang Hua (Wormpex AI Research)</i>	
Leveraging Spatio-Temporal Dependency for Skeleton-Based Action Recognition	10221
<i>Jungho Lee (Yonsei University), Minhyeok Lee (Yonsei University), Suhwan Cho (Yonsei University), Sungmin Woo (Yonsei University), Sungjun Jang (Yonsei University), and Sangyoun Lee (Yonsei University)</i>	
Cross-Modal Learning with 3D Deformable Attention for Action Recognition	10231
<i>Sangwon Kim (Keimyung University), Dasom Ahn (Keimyung University), and Byoung Chul Ko (Keimyung University)</i>	
Generative Action Description Prompts for Skeleton-based Action Recognition	10242
<i>Wangmeng Xiang (The Hong Kong Polytechnic University), Chao Li (Alibaba), Yuxuan Zhou (University of Mannheim), Biao Wang (Alibaba), and Lei Zhang (Hong Kong Polytechnic University, Hong Kong, China)</i>	
Self-Feedback DETR for Temporal Action Detection	10252
<i>Jihwan Kim (Sungkyunkwan University), Miso Lee (Sungkyunkwan University), and Jae-Pil Heo (Sungkyunkwan University)</i>	

Skip-Plan: Procedure Planning in Instructional Videos via Condensed Action Space Learning	10263
<i>Zhiheng Li (Tsinghua University), Wenjia Geng (Shenzhen International Graduate School, Tsinghua University), Muheng Li (Tsinghua University), Lei Chen (University of Science and Technology Beijing), Yansong Tang (Tsinghua University), Jiwen Lu (Tsinghua University), and Jie Zhou (Tsinghua University)</i>	
The Unreasonable Effectiveness of Large Language-Vision Models for Source-Free Video Domain Adaptation	10273
<i>Giacomo Zara (University of Trento), Alessandro Conti (University of Trento), Subhankar Roy (Telecom Paris), Stéphane Lathuilière (Telecom-Paris), Paolo Rota (University of Trento), and Elisa Ricci (FBK - Technologies of Vision)</i>	
Multimodal Motion Conditioned Diffusion Model for Skeleton-based Video Anomaly Detection	10284
<i>Alessandro Flaborea (Sapienza University of Rome), Luca Collorone (Universita' Degli Studi Di Roma La Sapienza), Guido Maria D'Amely di Melendugno (Sapienza University), Stefano D'Arrigo (Sapienza University of Rome), Bardh Prenkaj (Sapienza University of Rome), and Fabio Galasso (Sapienza University)</i>	
Video Anomaly Detection via Sequentially Learning Multiple Pretext Tasks	10296
<i>Chenrui Shi (Beijing Institute of Technology), Che Sun (Beijing Institute of Technology), Yuwei Wu (Beijing Institute of Technology, (BIT)), and Yunde Jia (Beijing Institute of Technology)</i>	
MiniROAD: Minimal RNN Framework for Online Action Detection	10307
<i>Joungbin An (Yonsei University), Hyolim Kang (Yonsei University), Su Ho Han (Yonsei University), Ming-Hsuan Yang (University of California at Merced), and Seon Joo Kim (Yonsei University)</i>	
How Much Temporal Long-Term Context is Needed for Action Segmentation?	10317
<i>Emad Bahrami (University of Bonn), Gianpiero Francesca (Toyota-Europe), and Juergen Gall (University of Bonn)</i>	
DiffTAD: Temporal Action Detection with Proposal Denoising Diffusion	10328
<i>Sauradip Nag (University of Surrey, UK; iFlyTek-Surrey Joint Research Center on Artificial Intelligence, UK), Xiatian Zhu (University of Surrey, UK; Surrey Institute for People-Centred Artificial Intelligence, UK), Jiankang Deng (Imperial College London, UK), Yi-Zhe Song (University of Surrey, UK; iFlyTek-Surrey Joint Research Center on Artificial Intelligence, UK), and Tao Xiang (University of Surrey, UK; iFlyTek-Surrey Joint Research Center on Artificial Intelligence, UK)</i>	
STEPS: Self-Supervised Key Step Extraction and Localization from Unlabeled Procedural Videos	10341
<i>Anshul Shah (Johns Hopkins University), Benjamin Lundell (Microsoft), Harpreet Sawhney (Microsoft), and Rama Chellappa (Johns Hopkins University)</i>	
Efficient Video Action Detection with Token Dropout and Context Refinement	10354
<i>Lei Chen (Nanjing University), Zhan Tong (Ant Research), Yibing Song (Fudan University), Gangshan Wu (Nanjing University), and Limin Wang (Nanjing University)</i>	

FSAR: Federated Skeleton-based Action Recognition with Adaptive Topology Structure and Knowledge Distillation	10366
<i>Jingwen Guo (Peking University Shenzhen Graduate School), Hong Liu (Peking University Shenzhen Graduate School), Shitong Sun (Queen Mary University of London), Tianyu Guo (Peking University Shenzhen Graduate School), Min Zhang (Harbin Institute of Technology Shenzhen Graduate School), and Chenyang Si (Singapore Nanyang Technological University)</i>	
Exploring Predicate Visual Context in Detecting of Human–Object Interactions	10377
<i>Frederic Z Zhang (Australian National University), Yuhui Yuan (Microsoft Research), Dylan Campbell (Australian National University), Zhuoyao Zhong (South China University of Technology), and Stephen Gould (Australian National University, Australia)</i>	
E2E-LOAD: End-to-End Long-form Online Action Detection	10388
<i>Shuqiang Cao (Shandong University), Weixin Luo (Meituan), Bairui Wang (Meituan), Wei Zhang (Shandong University), and Lin Ma (Meituan)</i>	
Revisiting Foreground and Background Separation in Weakly-supervised Temporal Action Localization: A Clustering-based Approach	10399
<i>Qinying Liu (University of Science and Technology of China), Zilei Wang (University of Science and Technology of China), Shenghai Rong (University of Science and Technology of China), Junjie Li (University of Science and Technology of China), and Yixin Zhang (University of Science and Technology of China)</i>	
Hierarchically Decomposed Graph Convolutional Networks for Skeleton-Based Action Recognition	10410
<i>Jungho Lee (Yonsei University), Minhyeok Lee (Yonsei University), Dogyoon Lee (Yonsei University), and Sangyoun Lee (Yonsei University)</i>	

Computational Imaging

Tiled Multiplane Images for Practical 3D Photography	10420
<i>Numair Khan (Meta), Lei Xiao (Reality Labs Research, Meta), and Douglas Lanman (Facebook Reality Labs)</i>	
Eulerian Single-Photon Vision	10431
<i>Shantanu Gupta (University of Wisconsin-Madison) and Mohit Gupta (University of Wisconsin-Madison, USA)</i>	
ProPainter: Improving Propagation and Transformer for Video Inpainting	10443
<i>Shangchen Zhou (Nanyang Technological University), Chongyi Li (Nanyang Technological University), Kelvin C.K. Chan (Google), and Chen Change Loy (Nanyang Technological University)</i>	
Global Perception Based Autoregressive Neural Processes	10453
<i>Jinyang Tai (Shanghai University)</i>	

DOLCE: A Model-Based Probabilistic Diffusion Framework for Limited-Angle CT Reconstruction.....	10464
<i>Jiaming Liu (Washington University in St. Louis), Rushil Anirudh (Lawrence Livermore National Laboratory), Jayaraman J. Thiagarajan (Lawrence Livermore National Laboratory), Stewart He (Lawrence Livermore National Laboratory), K. Aditya Mohan (Lawrence Livermore National Laboratory), Ulugbek S. Kamilov (Washington University in St. Louis), and Hyojin Kim (LLNL)</i>	
GlowGAN: Unsupervised Learning of HDR Images from LDR Images in the Wild	10475
<i>Chao Wang (Max Planck Institute for Informatics), Ana Serrano (Universidad de Zaragoza), Xingang Pan (Max Planck Institute for Informatics), Bin Chen (Max-Planck-Institut für Informatik), Karol Myszkowski (MPI-Informatik), Hans-Peter Seidel (Max Planck Institute for Informatics), Christian Theobalt (MPI Informatik), and Thomas Leimkühler (MPI Informatik)</i>	
Score-Based Diffusion Models as Principled Priors for Inverse Imaging	10486
<i>Berthy T. Feng (California Institute of Technology), Jamie Smith (Google), Michael Rubinstein (Google), Huiwen Chang (Google Research), Katherine L. Bouman (Caltech), and William T. Freeman (MIT)</i>	
NLOS-NeuS: Non-line-of-sight Neural Implicit Surface	10498
<i>Yuki Fujimura (Nara Institute of Science and Technology), Takahiro Kushida (Ritsumeikan University), Takuya Funatomi (Nara Institute of Science and Technology), and Yasuhiro Mukaigawa (NAIST)</i>	
MEFLUT: Unsupervised 1D Lookup Tables for Multi-exposure Image Fusion	10508
<i>Ting Jiang (MEGVII), Chuan Wang (Face++, (Megvii), Xinpeng Li (Megvii), Ru Li (Harbin Institute of Technology), Haoqiang Fan (Megvii Incface++), and Shuaicheng Liu (UESTC, Megvii)</i>	
Temporal-Coded Spiking Neural Networks with Dynamic Firing Threshold: Learning with Event-Driven Backpropagation	10518
<i>Wenjie Wei (University of Electronic Science and Technology of China), Malu Zhang (University of Electronic Science and Technology of China), Hong Qu (University of Electronic Science and Technology of China), Ammar Belatreche (Northumbria University), Jian Zhang (Tsinghua University), and Hong Chen (Tsinghua University)</i>	
Enhancing Non-line-of-sight Imaging via Learnable Inverse Kernel and Attention Mechanisms .	10529
<i>Yanhua Yu (ShanghaiTech University), Siyuan Shen (ShanghaiTech University), Zi Wang (ShanghaiTech University), Binbin Huang (ShanghaiTech University), Yuehan Wang (ShanghaiTech University), Xingyue Peng (ShanghaiTech University), Suan Xia (ShanghaiTech University), Ping Liu (ShanghaiTech University), Ruiqian Li (ShanghaiTech University), and Shiyong Li (ShanghaiTech University)</i>	
Aperture Diffraction for Compact Snapshot Spectral Imaging	10540
<i>Tao Lv (Nanjing University), Hao Ye (Nanjing University), Quan Yuan (Nanjing University), Zhan Shi (Nanjing University), Yibo Wang (Nanjing University), Shuming Wang (Nanjing University), and Xun Cao (Nanjing University)</i>	
Content-Aware Local GAN for Photo-Realistic Super-Resolution	10551
<i>JoonKyu Park (Seoul National Univ.), Sanghyun Son (Seoul National University), and Kyoung Mu Lee (Seoul National University)</i>	

RED-PSM: Regularization by Denoising of Partially Separable Models for Dynamic Imaging	10561
<i>Berk Iskender (University of Illinois at Urbana-Champaign), Marc L. Klasky (Los Alamos National Laboratory), and Yoram Bresler (UIUC)</i>	
Self-Supervised Burst Super-Resolution	10571
<i>Goutam Bhat (ETH Zurich), Michaël Gharbi (Adobe Research), Jiawen Chen (Adobe, Inc.), Luc Van Gool (ETH Zurich), and Zhihao Xia (Adobe)</i>	
Coherent Event Guided Low-Light Video Enhancement	10581
<i>Jinxu Liang (Peking University), Yixin Yang (Peking University), Boyu Li (Peking University), Peiqi Duan (Peking University), Yong Xu (South China University of Technology), and Boxin Shi (Peking University)</i>	
Panoramas from Photons	10592
<i>Sacha Jungerman (University of Wisconsin-Madison), Atul Ingle (Portland State University), and Mohit Gupta (University of Wisconsin-Madison, USA)</i>	
Designing Phase Masks for Under-Display Cameras	10603
<i>Anqi Yang (Carnegie Mellon University), Eunhee Kang (Samsung Electronics), Hyong-Euk Lee (Samsung Advanced Institute of Technology), and Aswin C. Sankaranarayanan (Carnegie Mellon University)</i>	
Deep Optics for Video Snapshot Compressive Imaging	10612
<i>Ping Wang (Zhejiang University), Lishun Wang (UCAS), and Xin Yuan (Westlake University)</i>	
TiDy-PSFs: Computational Imaging with Time-Averaged Dynamic Point-Spread-Functions	10623
<i>Sachin Shah (University of Maryland), Sakshum Kulshrestha (University of Maryland), and Christopher A. Metzler (University of Maryland College Park)</i>	
Generalized Lightness Adaptation with Channel Selective Normalization	10634
<i>Mingde Yao (University of Science and Technology of China), Jie Huang (University of Science and Technology of China), Xin Jin (Eastern Institute of Technology), Ruikang Xu (University of Science and Technology of China), Shenglong Zhou (University of Science and Technology of China), Man Zhou (Nanyang Technological University), and Zhiwei Xiong (University of Science and Technology of China)</i>	
Towards Nonlinear-Motion-Aware and Occlusion-Robust Rolling Shutter Correction	10646
<i>Delin Qu (Fudan University), Yizhen Lao (Hunan University), Zhigang Wang (shanghai AI lab), Dong Wang (Shanghai AI Laboratory), Bin Zhao (Northwestern Polytechnical University), and Xuelong Li (Northwestern Polytechnical University)</i>	
FCCNs: Fully Complex-valued Convolutional Networks using Complex-valued Color Model and Loss Function	10655
<i>Saurabh Yadav (IIIT Delhi) and Koteswar Rao Jerripothula (IIIT Delhi)</i>	
Event Camera Data Pre-training	10665
<i>Yan Yang (BDSI, ANU), Liyuan Pan (BITSZ & School of CSAT, BIT), and Liu Liu (Cyberverse Dept., Huawei)</i>	
Improving 3D Imaging with Pre-Trained Perpendicular 2D Diffusion Models	10676
<i>Suhyeon Lee (KAIST), Hyungjin Chung (KAIST), Minyoung Park (JLK Inc.), Jonghyuk Park (JLK Inc.), Wi-Sun Ryu (JLK Inc.), and Jong Chul Ye (Kim Jaechul Graduate School of AI, KAIST, Korea)</i>	

Multiscale Structure Guided Diffusion for Image Deblurring	10687
<i>Mengwei Ren (New York University), Mauricio Delbracio (Google Research), Hossein Talebi (Google), Guido Gerig (NYU), and Peyman Milanfar (Google)</i>	
Generalizing Event-Based Motion Deblurring in Real-World Scenarios	10700
<i>Xiang Zhang (Wuhan University), Lei Yu (Wuhan University), Wen Yang (Wuhan University), Jianzhuang Liu (Huawei Noah's Ark Lab), and Gui-Song Xia (Wuhan University)</i>	
On the Robustness of Normalizing Flows for Inverse Problems in Imaging	10711
<i>Seongmin Hong (Seoul National University), Inbum Park (Seoul National University), and Se Young Chun (Seoul National University)</i>	
Learned Compressive Representations for Single-Photon 3D Imaging	10722
<i>Felipe Gutierrez-Barragan (University of Wisconsin-Madison), Fangzhou Mu (University of Wisconsin-Madison), Andrei Ardelean (EPFL), Atul Ingle (Portland State University), Claudio Bruschini (EPFL), Edoardo Charbon (EPFL), Yin Li (University of Wisconsin-Madison), Mohit Gupta (University of Wisconsin-Madison, USA), and Andreas Velten (University of Wisconsin - Madison)</i>	
Recovering a Molecule's 3D Dynamics from Liquid-phase Electron Microscopy Movies	10733
<i>Enze Ye (Tsinghua University), Yuhang Wang (DP Technology, Ltd.), Hong Zhang (Peking University), Yiqin Gao (Peking University), Huan Wang (Peking University), and He Sun (Peking University)</i>	
NIR-assisted Video Enhancement via Unpaired 24-hour Data	10744
<i>Muyao Niu (The University of Tokyo), Zhihang Zhong (The University of Tokyo), and Yinqiang Zheng (The University of Tokyo)</i>	
SpinCam: High-Speed Imaging via a Rotating Point-Spread Function	10755
<i>Dorian Chan (Carnegie Mellon University), Mark Sheinin (Carnegie Mellon University), and Matthew O'Toole (Carnegie Mellon University)</i>	
RecRecNet: Rectangling Rectified Wide-Angle Images by Thin-Plate Spline Model and DoF-based Curriculum Learning	10766
<i>Kang Liao (Beijing Jiaotong University), Lang Nie (Beijing Jiaotong University), Chunyu Lin (Beijing Jiaotong University), Zishuo Zheng (Beijing Jiaotong University), and Yao Zhao (Beijing Jiaotong University)</i>	
Affective Image Filter: Reflecting Emotions from Text to Images	10776
<i>Shuchen Weng (Peking University), Peixuan Zhang (Beijing University of Posts and Telecommunications), Zheng Chang (Beijing University of Posts and Telecommunications), Xinlong Wang (Beijing Academy of Artificial Intelligence), Si Li (Beijing University of Posts and Telecommunications), and Boxin Shi (Peking University)</i>	
Towards General Low-Light Raw Noise Synthesis and Modeling	10786
<i>Feng Zhang (Huazhong University of Science and Technology), Bin Xu (Dji), Zhiqiang Li (Dji innovation), Xinran Liu (Huazhong University of Science and Technology), Qingbo Lu (DJI), Changxin Gao (Huazhong University of Science and Technology), and Nong Sang (Huazhong University of Science and Technology)</i>	

Unsupervised Video Deraining with An Event Camera	10797
<i>Jin Wang (University of Science and Technology of China), Wenming Weng (University of Science and Technology of China), Yueyi Zhang (University of Science and Technology of China), and Zhiwei Xiong (University of Science and Technology of China)</i>	
LoLep: Single-View View Synthesis with Locally-Learned Planes and Self-Attention Occlusion Inference	10807
<i>Cong Wang (Tsinghua University), Yu-Ping Wang (Beijing Institute of Technology), and Dinesh Manocha (University of Maryland at College Park)</i>	

Embodied Vision: Active Agents; Simulation

Skill Transformer: A Monolithic Policy for Mobile Manipulation	10818
<i>Xiaoyu Huang (Georgia Tech), Dhruv Batra (Meta AI, (FAIR, Georgia Tech), Akshara Rai (Meta AI, (FAIR), and Andrew Szot (Georgia Tech)</i>	
ENTL: Embodied Navigation Trajectory Learner	10829
<i>Klemen Kotar (Stanford CS), Aaron Walsman (University of Washington), and Roozbeh Mottaghi (FAIR @ Meta)</i>	
DREAMWALKER: Mental Planning for Continuous Vision-Language Navigation	10839
<i>Hanqing Wang (Beijing Institute of Technology), Wei Liang (Beijing Institute of Technology), Luc Van Gool (ETH Zurich), and Wenguan Wang (Zhejiang University)</i>	
Scene Graph Contrastive Learning for Embodied Navigation	10850
<i>Kunal Pratap Singh (Allen Institute for AI), Jordi Salvador (Allen Institute for AI), Luca Weihs (Allen Institute for Artificial Intelligence), and Aniruddha Kembhavi (Allen Institute for Artificial Intelligence)</i>	
Perpetual Humanoid Control for Real-time Simulated Avatars	10861
<i>Zhengyi Luo (Carnegie Mellon University), Jinkun Cao (Carnegie Mellon University), Alexander Winkler (Meta), Kris Kitani (Carnegie Mellon University), and Weipeng Xu (Reality Labs Research)</i>	
Grounding 3D Object Affordance from 2D Interactions in Images	10871
<i>Yuhang Yang (University of Science and Technology of China), Wei Zhai (University of Science and Technology of China), Hongchen Luo (University of Science and Technology of China), Yang Cao (University of Science and Technology of China), Jiebo Luo (U. Rochester), and Zheng-Jun Zha (University of Science and Technology of China)</i>	
Navigating to Objects Specified by Images	10882
<i>Jacob Krantz (Oregon State University), Theophile Gervet (Carnegie Mellon University), Karmesh Yadav (Facebook AI Research), Austin Wang (Facebook AI Research), Chris Paxton (Meta AI), Roozbeh Mottaghi (FAIR @ Meta), Dhruv Batra (Georgia Tech & Facebook AI Research), Jitendra Malik (University of California at Berkeley), Stefan Lee (Oregon State University), and Devendra Singh Chaplot (Facebook AI Research)</i>	
PEANUT: Predicting and Navigating to Unseen Targets	10892
<i>Albert J. Zhai (University of Illinois at Urbana-Champaign) and Shenlong Wang (UIUC)</i>	

Context-Aware Planning and Environment-Aware Memory for Instruction Following Embodied Agents	10902
<i>Byeonghwi Kim (Yonsei University), Jinyeon Kim (Yonsei University), Yuyeong Kim (GIST), Cheolhong Min (Yonsei University), and Jonghyun Choi (Yonsei University)</i>	
Learning Foresightful Dense Visual Affordance for Deformable Object Manipulation	10913
<i>Ruihai Wu (Peking University), Chuanruo Ning (Peking University), and Hao Dong (Peking University)</i>	
Exploiting Proximity-Aware Tasks for Embodied Social Navigation	10923
<i>Enrico Cancelli (University of Padua), Tommaso Campari (University of Padova), Luciano Serafini (Fondazione Bruno Kessler), Angel X. Chang (SFU), and Lamberto Ballan (University of Padova)</i>	
Bird's-Eye-View Scene Graph for Vision-Language Navigation	10934
<i>Rui Liu (Zhejiang University), Xiaohan Wang (Zhejiang University), Wenguan Wang (Zhejiang University), and Yi Yang (Zhejiang University)</i>	
Active Neural Mapping	10947
<i>Zike Yan (Peking University), Haoxiang Yang (Peking University), and Hongbin Zha (Peking University, China)</i>	
Omnidirectional Information Gathering for Knowledge Transfer-Based Audio-Visual Navigation.....	10959
<i>Jinyu Chen (Beihang University), Wenguan Wang (Zhejiang University), Si Liu (Beihang University), Hongsheng Li (The Chinese University of Hong Kong), and Yi Yang (Zhejiang University)</i>	
Multi-Object Navigation with Dynamically Learned Neural Implicit Representations	10970
<i>Pierre Marza (INSA Lyon), Laetitia Matignon (UCBL), Olivier Simonin (INSA Lyon), and Christian Wolf (Naver Labs Europe)</i>	

Recognition: Retrieval

Unsupervised Feature Representation Learning for Domain-generalized Cross-domain Image Retrieval	10982
<i>Conghui Hu (National University of Singapore), Can Zhang (National University of Singapore), and Gim Hee Lee (National University of Singapore)</i>	
DEDRIFT: Robust Similarity Search under Content Drift	10992
<i>Dmitry Baranchuk (MSU / Yandex), Matthijs Douze (Meta AI Research), Yash Upadhyay (University of Minnesota, Twin Cities), and I. Zeki Yalniz (Meta)</i>	
Global Features are All You Need for Image Retrieval and Reranking	11002
<i>Shihao Shao (Peking University), Kaifeng Chen (Google), Arjun Karpur (Google), Qinghua Cui (peking university), André Araujo (Google), and Bingyi Cao (Google)</i>	
HSE: Hybrid Species Embedding for Deep Metric Learning	11013
<i>Bailin Yang (Zhejiang Gongshang University), Haoqiang Sun (Zhejiang Gongshang University), Frederick W. B. Li (University of Durham), Zheng Chen (Zhejiang Gongshang University), Jianlu Cai (zhejiang Gongshang University), and Chao Song (Zhejiang Gongshang University)</i>	

Discrepant and Multi-Instance Proxies for Unsupervised Person Re-Identification	11024
<i>Chang Zou (Xi'an Jiaotong University), Zeqi Chen (Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an, China.), Zhichao Cui (Chang'an University), Yuehu Liu (Xi'an Jiaotong University), and Chi Zhang (Xi'an Jiaotong University)</i>	
Towards Grand Unified Representation Learning for Unsupervised Visible-Infrared Person Re-Identification	11035
<i>Bin Yang (Wuhan University), Jun Chen (Wuhan University), and Mang Ye (Wuhan University)</i>	
EigenPlaces: Training Viewpoint Robust Models for Visual Place Recognition	11046
<i>Gabriele Berton (Politecnico di Torino), Gabriele Trivigno (Polytechnic of Turin), Barbara Caputo (Politecnico di Torino), and Carlo Masone (Politecnico di Torino)</i>	
Simple Baselines for Interactive Video Retrieval with Questions and Answers	11057
<i>Kaiqu Liang (Princeton University) and Samuel Albanie (University of Cambridge)</i>	
Fan-Beam Binarization Difference Projection (FB-BDP): A Novel Local Object Descriptor for Fine-Grained Leaf Image Retrieval	11068
<i>Xin Chen (Griffith university), Bin Wang (Griffith University), and Yongsheng Gao (Griffith University)</i>	
Conditional Cross Attention Network for Multi-Space Embedding without Entanglement in Only a SINGLE Network	11078
<i>Chull Hwan Song (Dealicious Inc.), Taebaek Hwang (Dealicious Inc.), Jooyoung Yoon (Dealicious Inc.), Shunghyun Choi (Dealicious Inc.), and Yeong Hyeon Gu (Sejong University)</i>	
Learning Concordant Attention via Target-aware Alignment for Visible-Infrared Person Re-identification	11088
<i>Jianbing Wu (Peking University), Hong Liu (Peking University), Yuxin Su (Peking University), Wei Shi (Peking University), and Hao Tang (Computer Vision Lab, ETH Zürich, Switzerland)</i>	
Person Re-Identification without Identification via Event anonymization	11098
<i>Shafiq Ahmad (Istituto Italiano di Tecnologia), Pietro Morerio (Istituto Italiano di Tecnologia), and Alessio Del Bue (Istituto Italiano di Tecnologia, IIT)</i>	
Divide&Classify: Fine-Grained Classification for City-Wide Visual Geo-Localization	11108
<i>Gabriele Trivigno (Polytechnic of Turin), Gabriele Berton (Politecnico di Torino), Juan Aragon (Politecnico di Torino), Barbara Caputo (Politecnico di Torino), and Carlo Masone (Politecnico di Torino)</i>	
Dark Side Augmentation: Generating Diverse Night Examples for Metric Learning	11119
<i>Albert Mohwald (Vision Recognition Group, Czech Technical University in Prague), Tomas Jenicek (Vision Recognition Group, Czech Technical University in Prague), and Ondrej Chum (Vision Recognition Group, Czech Technical University in Prague)</i>	

PIDRo: Parallel Isomeric Attention with Dynamic Routing for Text-Video Retrieval	11130
<i>Peiyan Guan (The University of Hong Kong), Renjing Pei (HuaWei), Bin Shao (Huawei Noah's Ark Lab), Jianzhuang Liu (Huawei Noah's Ark Lab), Weimian Li (Huawei), Jiayi Gu (Huawei Noah's Ark Lab), Hang Xu (Huawei Noah's Ark Lab), Songcen Xu (Huawei Noah's Ark Lab), Youliang Yan (Huawei Noah's Ark Lab), and Edmund Y. Lam (The University of Hong Kong)</i>	
Unified Pre-Training with Pseudo Texts for Text-To-Image Person Re-Identification	11140
<i>Zhiyin Shao (South China University of Technology), Xinyu Zhang (Baidu Inc.), Changxing Ding (South China University of Technology), Jian Wang (Baidu), and Jingdong Wang (Baidu)</i>	
Modality Unifying Network for Visible-Infrared Person Re-Identification	11151
<i>Hao Yu (Nanjing University of Information Science and Technology), Xu Cheng (Nanjing University of Information Science and Technology), Wei Peng (Stanford University), Weihao Liu (Soochow University), and Guoying Zhao (University of Oulu)</i>	
DeepChange: A Long-Term Person Re-Identification Benchmark with Clothes Change	11162
<i>Peng Xu (Tsinghua University) and Xiatian Zhu (University of Surrey)</i>	
LexLIP: Lexicon-Bottlenecked Language-Image Pre-Training for Large-Scale Image-Text Sparse Retrieval	11172
<i>Ziyang Luo (Hong Kong Baptist University), Pu Zhao (Microsoft Research), Can Xu (microsoft), Xiubo Geng (STCA NLP Group, Microsoft), Tao Shen (Microsoft), Chongyang Tao (Microsoft), Jing Ma (Hong Kong Baptist University), Qingwei Lin (Microsoft Research), and Daxin Jiang (Microsoft, Beijing, China)</i>	
Dual Pseudo-Labels Interactive Self-Training for Semi-Supervised Visible-Infrared Person Re-Identification	11184
<i>Jiangming Shi (Xiamen University), Yachao Zhang (Tsinghua University), Xiangbo Yin (Xiamen University), Yuan Xie (East China Normal University), Zhizhong Zhang (East China Normal University), Jianping Fan (Lenovo), Zhongchao Shi (lenovo company), and Yanyun Qu (XMU)</i>	
BT ² : Backward-compatible Training with Basis Transformation	11195
<i>Yifei Zhou (Cornell University), Zilu Li (Cornell University), Abhinav Shrivastava (University of Maryland), Hengshuang Zhao (The University of Hong Kong), Antonio Torralba (MIT), Taipeng Tian (Meta), and Ser-Nam Lim (Facebook AI)</i>	
Prototypical Mixing and Retrieval-Based Refinement for Label Noise-Resistant Image Retrieval	11205
<i>Xinlong Yang (PKU), Haixin Wang (Peking University), Jinan Sun (Peking University), Shikun Zhang (Peking University), Chong Chen (Peking University), Xian-Sheng Hua (Terminus Group), and Xiao Luo (UCLA)</i>	
Learning Spatial-context-aware Global Visual Feature Representation for Instance Image Retrieval	11216
<i>Zhongyan Zhang (University of Wollongong), Lei Wang (University of Wollongong, Australia), Luping Zhou (University of Sydney), and Piotr Koniusz (ANU College of Engineering and Computer Science)</i>	

Coarse-to-Fine: Learning Compact Discriminative Representation for Single-Stage Image Retrieval	11226
<i>Yunquan Zhu (Tencent YouTu Lab), Xinkai Gao (Tencent Youtu Lab), Bo Ke (Tencent Youtu Lab), Ruizhi Qiao (Tencent Youtu Lab), and Xing Sun (Tencent)</i>	
Visible-Infrared Person Re-Identification via Semantic Alignment and Affinity Inference	11236
<i>Xingye Fang (Beijing Institute of Technology), Yang Yang (Institute of Automation, Chinese Academy of Sciences), and Ying Fu (Beijing Institute of Technology)</i>	
Part-Aware Transformer for Generalizable Person Re-identification	11246
<i>Hao Ni (The University of Electronic Science and Technology of China), Yuke Li (The University of Electronic Science and Technology of China), Lianli Gao (The University of Electronic Science and Technology of China), Heng Tao Shen (University of Electronic Science and Technology of China, UESTC), and Jingkuan Song (UESTC)</i>	
Towards Universal Image Embeddings: A Large-Scale Dataset and Challenge for Generic Image Representations	11256
<i>Nikolaos-Antonios Ypsilantis (Czech Technical University in Prague), Kaifeng Chen (Google), Bingyi Cao (Google), Mário Lipovský (Google), Pelin Dogan-Schönberger (Google), Grzegorz Makosa (Google), Boris Bluntschli (Google), Mojtaba Seyedhosseini (Google), Ondrej Chum (Vision Recognition Group, Czech Technical University in Prague), and André Araujo (Google)</i>	
Dual Learning with Dynamic Knowledge Distillation for Partially Relevant Video Retrieval	11268
<i>Jianfeng Dong (Zhejiang Gongshang University), Minsong Zhang (zhejiang Gongshang University), Zheng Zhang (Zhejiang Gongshang University), Xianke Chen (Zhejiang Gongshang University), Daizong Liu (Peking University), Xiaoye Qu (Huawei Cloud), Xun Wang (Zhejiang Gongshang University), and Baolong Liu (Zhejiang Gongshang University)</i>	
Fine-grained Unsupervised Domain Adaptation for Gait Recognition	11279
<i>Kang Ma (Beijing Institute of Technology), Ying Fu (Beijing Institute of Technology), Dezhi Zheng (Beijing Institute of technology), Yunjie Peng (Beihang University), Chunshui Cao (Watrix Technology), and Yongzhen Huang (School of Artificial Intelligence, Beijing Normal University)</i>	
FashionNTM: Multi-turn Fashion Image Retrieval via Cascaded Memory	11289
<i>Anwesan Pal (UC San Diego), Sahil Wadhwa (Amazon), Ayush Jaiswal (Amazon.com Inc.), Xu Zhang (Amazon.com Inc.), Yue Wu (Amazon.com Inc.), Rakesh Chada (Amazon), Pradeep Natarajan (Amazon.com Inc.), and Henrik I. Christensen (UC San Diego)</i>	
CrossLoc3D: Aerial-Ground Cross-Source 3D Place Recognition	11301
<i>Tianrui Guan (University of Maryland, College Park), Aswath Muthuselvam (University of Maryland - College Park), Montana Hoover (University of Maryland College Park), Xijun Wang (University of Maryland, College Park), Jing Liang (UMD), Adarsh Jagan Sathyamoorthy (University of Maryland College Park), Damon Conover (DEVCOM Army Research Laboratory), and Dinesh Manocha (University of Maryland at College Park)</i>	

Transfer, Low-Shot, Continual, Long-Tail Learning 1

ImbSAM: A Closer Look at Sharpness-Aware Minimization in Class-Imbalanced Recognition	11311
<i>Yixuan Zhou (UESTC), Yi Qu (UESTC), Xing Xu (University of Electronic Science and Technology of China), and Hengtao Shen (University of Electronic Science and Technology of China, (UESTC))</i>	
LFS-GAN: Lifelong Few-Shot Image Generation	11322
<i>Juwon Seo (Kyung Hee University), Ji-Su Kang (Klleon), and Gyeong-Moon Park (Kyung Hee University)</i>	
Augmented Box Replay: Overcoming Foreground Shift for Incremental Object Detection	11333
<i>Yuyang Liu (State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences), Yang Cong (Chinese Academy of Sciences), Dipam Goswami (Computer Vision Center Barcelona), Xialei Liu (Nankai University), and Joost van de Weijer (Computer Vision Center)</i>	
Contrastive Model Adaptation for Cross-Condition Robustness in Semantic Segmentation	11344
<i>David Brüggemann (ETH Zurich), Christos Sakaridis (ETH Zurich), Tim Broedermann (ETH Zurich), and Luc Van Gool (ETH Zurich)</i>	
Towards Effective Instance Discrimination Contrastive Loss for Unsupervised Domain Adaptation	11354
<i>Yixin Zhang (University of Science and Technology of China), Zilei Wang (University of Science and Technology of China), Junjie Li (University of Science and Technology of China), Jiafan Zhuang (Shantou University), and Zihan Lin (University of Science and Technology of China)</i>	
Adversarial Bayesian Augmentation for Single-Source Domain Generalization	11366
<i>Sheng Cheng (Arizona State University), Tejas Gokhale (Arizona State University), and Yezhou Yang (Arizona State University)</i>	
Measuring Asymmetric Gradient Discrepancy in Parallel Continual Learning	11377
<i>Fan Lyu (College of Intelligence and Computing, Tianjin University), Qing Sun (College of Intelligence and Computing, Tianjin University), Fanhua Shang (Tianjin University), Liang Wan (College of Intelligence and Computing, Tianjin University), and Wei Feng (College of Intelligence and Computing, Tianjin University, China)</i>	
CSDA: Learning Category-Scale Joint Feature for Domain Adaptive Object Detection	11387
<i>Changlong Gao (xi'an jiaotong university), Chengxu Liu (Xi'an Jiaotong University), Yujiu Dun (Xi'an Jiaotong University), and Xueming Qian (Xi'an Jiaotong University)</i>	
Distilling from Similar Tasks for Transfer Learning on a Budget	11397
<i>Kenneth Borup (Aarhus University), Cheng Perng Phoo (Cornell University), and Bharath Hariharan (Cornell University)</i>	
Complementary Domain Adaptation and Generalization for Unsupervised Continual Domain Shift Learning	11408
<i>Wonguk Cho (Seoul National University), Jinha Park (Seoul National University), and Taesup Kim (Seoul National University)</i>	

Camera-Driven Representation Learning for Unsupervised Domain Adaptive Person Re-identification	11419
<i>Geon Lee (Yonsei University), Sanghoon Lee (Yonsei University), Dohyung Kim (YONSEI UNIVERSITY), Younghoon Shin (Hyundai motor company), Yongsang Yoon (Hyundai Motor Group Robotics LAB), and Bumsu Ham (Yonsei University)</i>	
Introducing Language Guidance in Prompt-based Continual Learning	11429
<i>Muhammad Gul Zain Ali Ali Khan (Deutsches Forschungszentrum für künstliche Intelligenz, (DFKI), Muhammad Ferjad Naeem (ETH Zürich), Luc Van Gool (ETH Zurich), Didier Stricker (DFKI), Federico Tombari (Google, TU Munich), and Muhammad Zeshan Afzal (Deutsches Forschungszentrum für Künstliche Intelligenz, (DFKI, GmbH)</i>	
Fast and Accurate Transferability Measurement by Evaluating Intra-class Feature Variance	11440
<i>Huiwen Xu (Seoul National University) and U Kang (Seoul National University)</i>	
A Unified Continual Learning Framework with General Parameter-Efficient Tuning	11449
<i>Qiankun Gao (Peking University Shenzhen Graduate School), Chen Zhao (KAUST), Yifan Sun (Baidu Research), Teng Xi (Baidu Inc.), Gang Zhang (Baidu Inc.), Bernard Ghanem (KAUST), and Jian Zhang (Peking University)</i>	
SFHarmony: Source Free Domain Adaptation for Distributed Neuroimaging Analysis	11460
<i>Nicola K Dinsdale (University of Oxford), Mark Jenkinson (University of Oxford), and Ana IL Namburete (University of Oxford)</i>	
Towards Realistic Evaluation of Industrial Continual Learning Scenarios with an Emphasis on Energy Consumption and Computational Footprint	11472
<i>Vivek Chavan (Fraunhofer IPK), Paul Koch (Fraunhofer Institute for Production Systems and Design Technology IPK), Marian Schlüter (Fraunhofer Institute for Production Systems and Design Technology IPK), and Clemens Briese (Fraunhofer Institute for Production Systems and Design Technology)</i>	
CDAC: Cross-domain Attention Consistency in Transformer for Domain Adaptive Semantic Segmentation	11485
<i>Kaihong Wang (Boston University), Donghyun Kim (Korea University), Rogerio Feris (MIT-IBM Watson AI Lab), and Margrit Betke (Boston University)</i>	
PC-Adapter: Topology-Aware Adapter for Efficient Domain Adaption on Point Clouds with Rectified Pseudo-label	11496
<i>Joonhyung Park (KAIST), Hyunjin Seo (KAIST), and Eunho Yang (KAIST;AITRICS)</i>	
DETA: Denoised Task Adaptation for Few-Shot Learning	11507
<i>Ji Zhang (University of Electronic Science and Technology of China), Lianli Gao (Shenzhen Institute for Advanced Study), Xu Luo (University of Electronic Science and Technology of China), Hengtao Shen (University of Electronic Science and Technology of China), and Jingkuan Song (Shenzhen Institute for Advanced Study)</i>	

Activate and Reject: Towards Safe Domain Generalization under Category Shift	11518
<i>Chaoqi Chen (The University of Hong Kong), Luyao Tang (XMU), Leitian Tao (Wuhan University), Hong-Yu Zhou (The University of Hong Kong), Yue Huang (Xiamen University), Xiaoguang Han (Shenzhen Research Institute of Big Data, the Chinese University of Hong Kong, (Shenzhen), and Yizhou Yu (The University of Hong Kong)</i>	
Generalizable Decision Boundaries: Dualistic Meta-Learning for Open Set Domain Generalization	11530
<i>Xiran Wang (Nanjing University), Jian Zhang (Nanjing University), Lei Qi (Southeast University), and Yinghuan Shi (Nanjing University)</i>	
Continual Zero-Shot Learning through Semantically Guided Generative Random Walks	11540
<i>Wenxuan Zhang (King Abdullah University of Science and Technology), Paul Janson (University of Moratuwa), Kai Yi (King Abdullah University of Science and Technology), Ivan Skorokhodov (KAUST), and Mohamed Elhoseiny (KAUST)</i>	
Zero-Shot Point Cloud Segmentation by Semantic-Visual Aware Synthesis	11552
<i>Yuwei Yang (Sichuan University), Munawar Hayat (Monash University), Zhao Jin (Sichuan University), Hongyuan Zhu (Astar), and Yinjie Lei (Sichuan University)</i>	
MDCS: More Diverse Experts with Consistency Self-distillation for Long-tailed Recognition	11563
<i>Qihao Zhao (Beijing University of Chemical Technology), Chen Jiang (Beijing University of Chemical Technology), Wei Hu (Beijing University of Chemical Technology), Fan Zhang (Beijing University of Chemical Technology), and Jun Liu (Singapore University of Technology and Design)</i>	
Building a Winning Team: Selecting Source Model Ensembles using a Submodular Transferability Estimation Approach	11575
<i>Vimal K B (IIT Hyderabad), Saketh Bachu (Indian Institute of Technology, Hyderabad), Tanmay Garg (Indian Institute of Technology, Hyderabad), Niveditha Lakshmi Narasimhan (KLA Corporation), Raghavan Konuru (KLA Software India Pvt Ltd), and Vineeth N Balasubramanian (Indian Institute of Technology, Hyderabad)</i>	
Confidence-based Visual Dispersion for Few-shot Unsupervised Domain Adaptation	11587
<i>Yizhe Xiong (Tsinghua University), Hui Chen (Tsinghua University), Zijia Lin (Alibaba), Sicheng Zhao (Tsinghua University), and Guiguang Ding (Tsinghua University, China)</i>	
BEV-DG: Cross-Modal Learning under Bird's-Eye View for Domain Generalization of 3D Semantic Segmentation	11598
<i>Miaoyu Li (Xiamen University), Yachao Zhang (Tsinghua University), Xu Ma (Northeastern University), Yanyun Qu (XMU), and Yun Fu (Northeastern University)</i>	
CDFSL-V: Cross-Domain Few-Shot Learning for Videos	11609
<i>Sarinda Samarasinghe (University of Central Florida), Mamshad Nayeem Rizve (University of Central Florida), Navid Kardan (University of Central Florida), and Mubarak Shah (University of Central Florida)</i>	

Energy-based Self-Training and Normalization for Unsupervised Domain Adaptation	11619
<i>Samitha Herath (Monash University), Basura Fernando (Agency for Science, Technology and Research, A*STAR, Singapore), Ehsan Abbasnejad (The University of Adelaide), Munawar Hayat (Monash University), Shahram Khadivi (eBay, Inc.), Mehrtash Harandi (Monash University), Hamid Rezatofighi (Monash University), and Gholamreza Haffari (Monash University, Australia)</i>	
Regularized Mask Tuning: Uncovering Hidden Knowledge in Pre-Trained Vision-Language Models...	11629
<i>Kecheng Zheng (Zhejiang University), Wei Wu (University of Science and Technology of China), Ruili Feng (University of Science and Technology of China), Kai Zhu (USTC), Jiawei Liu (University of Science and Technology of China), Deli Zhao (Alibaba Group), Zheng-Jun Zha (University of Science and Technology of China), Wei Chen (Zhejiang University), and Yujun Shen (Dept. of IE, CUHK)</i>	
NAPA-VQ: Neighborhood-Aware Prototype Augmentation with Vector Quantization for Continual Learning	11640
<i>Tamasha Malepathirana (University of Melbourne), Damith Senanayake (University of Melbourne), and Saman Halgamuge (University of Melbourne)</i>	
A Sentence Speaks a Thousand Images: Domain Generalization through Distilling CLIP with Language Guidance	11651
<i>Zeyi Huang (University of Wisconsin-Madison), Andy Zhou (University of Illinois at Urbana-Champaign), Zijian Ling (Imperial College London), Mu Cai (University of Wisconsin-Madison), Haohan Wang (University of Illinois Urbana-Champaign), and Yong Jae Lee (University of Wisconsin-Madison)</i>	
ViM: Vision Middleware for Unified Downstream Transferring	11662
<i>Yutong Feng (Alibaba Group), Biao Gong (Alibaba Group), Jianwen Jiang (Tsinghua University), Yiliang Lv (Alibaba Group), Yujun Shen (Dept. of IE, CUHK), Deli Zhao (Alibaba Group), and Jingren Zhou (Alibaba Group)</i>	
Learning to Learn: How to Continuously Teach Humans and Machines	11674
<i>Parantak Singh (Agency for Science, Technology and Research), You Li (University of Wisconsin-Madison), Ankur Sikarwar (I2R, A*STAR Singapore), Weixian Lei (National University of Singapore), Difei Gao (NUS), Morgan B. Talbot (Harvard), Ying Sun (Institute for Infocomm Research, A*STAR), Mike Zheng Shou (National University of Singapore), Gabriel Kreiman (Hms), and Mengmi Zhang (NTU, A*STAR)</i>	
A Good Student is Cooperative and Reliable: CNN-Transformer Collaborative Learning for Semantic Segmentation	11686
<i>Jinjing Zhu (HKUST), Yunhao Luo (Brown University), Xu Zheng (HKUST), Hao Wang (Alibaba Group), and Lin Wang (HKUST)</i>	
Online Class Incremental Learning on Stochastic Blurry Task Boundary via Mask and Visual Prompt Tuning	11697
<i>Jun-Yeong Moon (Kyung Hee University), Keon-Hee Park (Kyung Hee University), Jung Uk Kim (Kyung Hee University), and Gyeong-Moon Park (Kyung Hee University)</i>	

Heterogeneous Forgetting Compensation for Class-Incremental Learning	11708
<i>Jiahua Dong (Shenyang Institute of Automation, Chinese Academy of Sciences), Wenqi Liang (Shenyang Institute of Automation Chinese Academy of Sciences), Yang Cong (Chinese Academy of Sciences), and Gan Sun (Chinese Academy of Sciences)</i>	
Disposable Transfer Learning for Selective Source Task Unlearning	11718
<i>Seunghye Koh (KAIST), Hyounguk Shon (KAIST), Janghyeon Lee (LG AI Research), Hyeong Gwon Hong (KAIST), and Junmo Kim (KAIST)</i>	
Online Continual Learning on Hierarchical Label Expansion	11727
<i>Byung Hyun Lee (Seoul National University), Okchul Jung (Seoul National University), Jonghyun Choi (Yonsei University), and Se Young Chun (Seoul National University)</i>	
Black-Box Unsupervised Domain Adaptation with Bi-Directional Atkinson-Shiffrin Memory	11737
<i>Jingyi Zhang (Nanyang Technological University), Jiaying Huang (Nanyang Technological University), Xueying Jiang (Nanyang Technological University), and Shijian Lu (Nanyang Technological University)</i>	
Local and Global Logit Adjustments for Long-Tailed Learning	11749
<i>Yingfan Tao (Tsinghua University), Jingna Sun (bytedance), Hao Yang (ByteDance), Li Chen (Bytedance), Xu Wang (ByteDance), Wenming Yang (Tsinghua University), Daniel Du (ByteDance Inc.), and Min Zheng (ByteDance)</i>	
FS-DETR: Few-Shot DETection TRansformer with Prompting and without Re-Training	11759
<i>Adrian Bulat (Samsung AI Center, Cambridge), Ricardo Guerrero (Samsung AI Center Cambridge), Brais Martinez (Samsung AI Center), and Georgios Tzimiropoulos (Queen Mary University of London)</i>	
Tuning Pre-trained Model via Moment Probing	11769
<i>Mingze Gao (Tianjin University), Qilong Wang (Tianjin University), Zhenyi Lin (Tianjin University), Pengfei Zhu (tianjin university), Qinghua Hu (Tianjin University), and Jingbo Zhou (Baidu Research)</i>	

(O) Transfer, Low-Shot, and Continual Learning

Frequency Guidance Matters in Few-Shot Learning	11780
<i>Hao Cheng (Nanyang Technological University), Siyuan Yang (Nanyang Technological University), Joey Tianyi Zhou (A*STAR Centre for Frontier AI Research, (CFAR), Lanqing Guo (Nanyang Technological University), and Bihan Wen (Nanyang Technological University)</i>	
Sensitivity-Aware Visual Parameter-Efficient Fine-Tuning	11791
<i>Haoyu He (Monash University), Jianfei Cai (Monash University), Jing Zhang (The University of Sydney), Dacheng Tao (The University of Sydney), and Bohan Zhuang (Monash University)</i>	
On the Robustness of Open-World Test-Time Training: Self-Training with Dynamic Prototype Expansion	11802
<i>Yushu Li (South China University of Technology), Xun Xu (I2R, A-STAR), Yongyi Su (South China University of Technology), and Kui Jia (South China University of Technology)</i>	

Generating Instance-level Prompts for Rehearsal-free Continual Learning	11813
<i>Dahuin Jung (Seoul National University), Dongyoon Han (NAVER AI Lab), Jihwan Bang (Naver Corp.), and Hwanjun Song (AWS AI Labs)</i>	
Boosting Novel Category Discovery Over Domains with Soft Contrastive Learning and All in One Classifier	11824
<i>Zelin Zang (Zhejiang University & Westlake University), Lei Shang (Alibaba Group), Senqiao Yang (Harbin Institute of Technology), Fei Wang (alibaba group), Baigui Sun (Alibaba Group), Xuansong Xie (Alibaba), and Stan Z. Li (Westlake University)</i>	
A Soft Nearest-Neighbor Framework for Continual Semi-Supervised Learning	11834
<i>Zhiqi Kang (INRIA), Enrico Fini (University of Trento), Moin Nabi (SAP), Elisa Ricci (FBK - Technologies of Vision), and Karteek Alahari (Inria)</i>	
GraphEcho: Graph-Driven Unsupervised Domain Adaptation for Echocardiogram Video Segmentation	11844
<i>Jiewen Yang (The Hong Kong University of Science and Technology), Xinpeng Ding (The Hong Kong University of Science and Technology), Ziyang Zheng (Harbin Institute of Technology), Xiaowei Xu (Guangdong Provincial People's Hospital), and Xiaomeng Li (The Hong Kong University of Science and Technology)</i>	
ViperGPT: Visual Inference via Python Execution for Reasoning	11854
<i>Dádac Surís (Columbia University), Sachit Menon (Columbia University), and Carl Vondrick (Columbia University)</i>	
Improved Visual Fine-tuning with Natural Language Supervision	11865
<i>Junyang Wang (Beijing Jiaotong University), Yuanhong Xu (Alibaba Group), Juhua Hu (University of Washington), Ming Yan (Alibaba Group), Jitao Sang (Beijing Jiaotong University, China), and Qi Qian (Alibaba Group)</i>	
Preparing the Future for Continual Semantic Segmentation	11876
<i>Zihan Lin (University of Science and Technology of China), Zilei Wang (University of Science and Technology of China), and Yixin Zhang (University of Science and Technology of China)</i>	
MAP: Towards Balanced Generalization of IID and OOD through Model-Agnostic Adapters	11887
<i>Min Zhang (Zhejiang University), Junkun Yuan (Zhejiang University), Yue He (Tsinghua University), Wenbin Li (Nanjing University), Zhengyu Chen (Zhejiang University), and Kun Kuang (Zhejiang University)</i>	
Space-time Prompting for Video Class-incremental Learning	11898
<i>Yixuan Pei (Xi'an Jiaotong University), Zhiwu Qing (Huazhong University of Science and Technology), Shiwei Zhang (DAMO Academy, Alibaba Group), Xiang Wang (Huazhong University of Science and Technology), Yingya Zhang (Alibaba Group), Deli Zhao (Alibaba Group), and Xueming Qian (Xi'an Jiaotong University)</i>	

(O) Vision and Language

Chinese Text Recognition with A Pre-Trained CLIP-Like Model Through Image-IDS Aligning ...	11909
<i>Haiyang Yu (Fudan University), Xiaocong Wang (Fudan University), Bin Li (Fudan University), and Xianguang Xue (Fudan University)</i>	

OmniLabel: A Challenging Benchmark for Language-Based Object Detection	11919
<i>Samuel Schuster (NEC Laboratories America), Vijay Kumar B G (NEC Laboratories America), Yumin Suh (NEC Labs America), Konstantinos M. Dafnis (Rutgers University), Zhixing Zhang (Rutgers University), Shiyu Zhao (Rutgers University), and Dimitris Metaxas (Rutgers)</i>	
IntentQA: Context-aware Video Intent Reasoning	11929
<i>Jiapeng Li (Xi'an Jiaotong University), Ping Wei (Xi'an Jiaotong University), Wenjuan Han (Beijing Institute for General Artificial Intelligence), and Lifeng Fan (University of California, Los Angeles)</i>	
Sigmoid Loss for Language Image Pre-Training	11941
<i>Xiaohua Zhai (Google Brain), Basil Mustafa (Google), Alexander Kolesnikov (Google Brain), and Lucas Beyer (Google Brain)</i>	
What does CLIP know about a red circle? Visual prompt engineering for VLMs	11953
<i>Aleksandar Shtedritski (University of Oxford), Christian Rupprecht (University of Oxford), and Andrea Vedaldi (Oxford University)</i>	
Equivariant Similarity for Vision-Language Foundation Models	11964
<i>Tan Wang (Nanyang Technological University), Kevin Lin (Microsoft), Linjie Li (Microsoft), Chung-Ching Lin (Microsoft), Zhengyuan Yang (Microsoft), Hanwang Zhang (Nanyang Technological University), Zicheng Liu (Microsoft), and Lijuan Wang (Microsoft)</i>	
Scaling Data Generation in Vision-and-Language Navigation	11975
<i>Zun Wang (Australian National University), Jialu Li (University of North Carolina at Chapel Hill), Yicong Hong (Australian National University), Yi Wang (Shanghai AI Laboratory), Qi Wu (University of Adelaide), Mohit Bansal (University of North Carolina at Chapel Hill), Stephen Gould (Australian National University, Australia), Hao Tan (UNC, Chapel Hill), and Yu Qiao (Shanghai AI Laboratory)</i>	
Name Your Colour For the Task: Artificially Discover Colour Naming via Colour Quantisation Transformer	11987
<i>Shenghan Su (Shanghai Jiao Tong University), Lin Gu (RIKEN, AIP / The University of Tokyo), Yue Yang (Shanghai Jiao Tong University), Zenghui Zhang (Shanghai Jiao Tong University), and Tatsuya Harada (The University of Tokyo / RIKEN)</i>	
G2L: Semantically Aligned and Uniform Video Grounding via Geodesic and Game Theory	11998
<i>Hongxiang Li (Peking University), Meng Cao (Peking University), Xuxin Cheng (Peking University), Yaowei Li (Peking University), Zhihong Zhu (Peking University), and Yuexian Zou (Peking University)</i>	
Grounded Entity-Landmark Adaptive Pre-Training for Vision-and-Language Navigation	12009
<i>Yibo Cui (National Innovation Institute of Defense Technology, Academy of Military Sciences China), Liang Xie (National Innovation Institute of Defense and Technology), Yakun Zhang (National Innovation Institute of Defense Technology, Academy of Military Sciences), Meishan Zhang (Harbin Institute of Technology, (Shenzhen), Ye Yan (National Innovation Institute of Defense and Technology), and Erwei Yin (National Institute of Defense Technology Innovation, Academy of Military Sciences China, Beijing, 100081, China, Tianjin Artificial Intelligence Innovation Center, (TAIC))</i>	

Audio-Enhanced Text-to-Video Retrieval using Text-Conditioned Feature Alignment	12020
<i>Sarah Ibrahim (University of Amsterdam), Xiaohang Sun (Amazon Prime Video), Pichao Wang (Amazon Prime Video), Amanmeet Garg (Amazon Prime Video), Ashutosh Sanan (Amazon Prime Video), and Mohamed Omar (Amazon Prime Video)</i>	
Open-domain Visual Entity Recognition: Towards Recognizing Millions of Wikipedia Entities ...	12031
<i>Hexiang Hu (Google), Yi Luan (Google Brain), Yang Chen (Georgia Institute of Technology), Urvashi Khandelwal (Google), Mandar Joshi (Google), Kenton Lee (Google), Kristina Toutanova (Google), and Ming-Wei Chang (Google)</i>	

Low-Level and Physics-Based Vision

Hierarchical Contrastive Learning for Pattern-Generalizable Image Corruption Detection	12042
<i>Xin Feng (Harbin Institute of Technology, Shenzhen), Yifeng Xu (Harbin Institute of Technology, Shenzhen), Guangming Lu (Harbin Institute of Technology, Shenzhen), and Wenjie Pei (Harbin Institute of Technology, Shenzhen)</i>	
DDS2M: Self-Supervised Denoising Diffusion Spatio-Spectral Model for Hyperspectral Image Restoration	12052
<i>Yuchun Miao (Wuhan University), Lefei Zhang (Wuhan University), Liangpei Zhang (Wuhan University), and Dacheng Tao (The University of Sydney)</i>	
From Sky to the Ground: A Large-scale Benchmark and Simple Baseline Towards Real Rain Removal	12063
<i>Yun Guo (Huazhong University of Science and Technology), Xueyao Xiao (Huazhong University of Science and Technology), Yi Chang (Huazhong University of Science and Technology), Shumin Deng (Huazhong University of Science and Technology), and Luxin Yan (Huazhong University of Science and Technology)</i>	
VAPCNet: Viewpoint-Aware 3D Point Cloud Completion	12074
<i>Zhiheng Fu (The University of Western Australia), Longguang Wang (National University of Defense Technology), Lian Xu (The University of Western Australia), Zhiyong Wang (The University of Sydney), Hamid Laga (Murdoch University, Australia), Yulan Guo (Sun Yat-sen University), Farid Boussaid (University of Western Australia), and Mohammed Bennamoun (University of Western Australia)</i>	
AccFlow: Backward Accumulation for Long-Range Optical Flow	12085
<i>Guangyang Wu (UESTC), Xiaohong Liu (Shanghai Jiao Tong University), Kunming Luo (HKUST), Xi Liu (UESTC), Qingqing Zheng (Chinese Academic of Sciences), Shuaicheng Liu (UESTC, Megvii), Xinyang Jiang (Microsoft Research Asia), Guangtao Zhai (Shanghai Jiao Tong University), and Wenyi Wang (UESTC)</i>	
Improving Transformer-based Image Matching by Cascaded Capturing Spatially Informative Keypoints	12095
<i>Chenjie Cao (fudan.edu.cn) and Yanwei Fu (Fudan University)</i>	

Low-Light Image Enhancement with Multi-Stage Residue Quantization and Brightness-Aware Attention	12106
<i>Yunlong Liu (Xidian University), Tao Huang (Xidian University), Weisheng Dong (Xidian University), Fangfang Wu (Xidian University), Xin Li (University at Albany), and Guangming Shi (Xidian University)</i>	
Random Sub-Samples Generation for Self-Supervised Real Image Denoising	12116
<i>Yizhong Pan (Sichuan University), Xiao Liu (Sichuan University), Xiangyu Liao (Sichuan University), Yuanzhouhan Cao (Beijing Jiaotong University), and Chao Ren (Sichuan University)</i>	
RSFNet: A White-Box Image Retouching Approach using Region-Specific Color Filters	12126
<i>Wenqi Ouyang (Alibaba), Yi Dong (Nanyang Technological University), Xiaoyang Kang (Alibaba), Peiran Ren (Alibaba), Xin Xu (xin xu), and Xuansong Xie (Alibaba)</i>	
Physics-Driven Turbulence Image Restoration with Stochastic Refinement	12136
<i>Ajay Jaiswal (UT Austin), Xingguang Zhang (Purdue University), Stanley H. Chan (Purdue University, USA), and Zhangyang Wang (University of Texas at Austin)</i>	
SYENet: A Simple Yet Effective Network for Multiple Low-Level Vision Tasks with Real-Time Performance on Mobile Device	12148
<i>Weiran Gou (Sanechips), Ziyao Yi (Sanechips), Yan Xiang (Sanechips), Shaoqing Li (Sanechips Ltd, Co), Zibin Liu (sanechips), Dehui Kong (SaneChips), and Ke Xu (SaneChips)</i>	
Self-supervised Image Denoising with Downsampled Invariance Loss and Conditional Blind-Spot Network	12162
<i>Yeong Il Jang (Seoul National University), Keuntek Lee (Seoul National University), Gu Yong Park (Seoul National University), Seyun Kim (Gauss Labs), and Nam Ik Cho (Seoul National University)</i>	
Variational Degeneration to Structural Refinement: A Unified Framework for Superimposed Image Decomposition	12172
<i>Wenyu Li (Tianjin University), Yan Xu (Tianjin University), Yang Yang (Tianjin University), Haoran Ji (Tianjin University), and Yue Lang (Hebei University of Technology)</i>	
Reconstructed Convolution Module Based Look-Up Tables for Efficient Image Super-Resolution.....	12183
<i>Guandu Liu (Tsinghua university), Yukang Ding (Kuaishou), Mading Li (Kuaishou Technology), Ming Sun (Kuaishou Technology), Xing Wen (Kuaishou), and Bin Wang (Tsinghua University)</i>	
Self-supervised Pre-training for Mirror Detection	12193
<i>Jiaying Lin (City University of Hong Kong) and Rynson W.H. Lau (City University of Hong Kong)</i>	
Downscaled Representation Matters: Improving Image Rescaling with Collaborative Downscaled Images	12203
<i>Bingna Xu (South China University of Technology), Yong Guo (Max Planck Institute for Informatics), Luoqian Jiang (South China University of Technology), Mianjie Yu (South China University of Technology), and Jian Chen (South China University of Technology, China)</i>	

Self-supervised Monocular Underwater Depth Recovery, Image Restoration, and a Real-sea Video Dataset	12214
<i>Nisha Varghese (IIT Madras), Ashish Kumar (Indian Institute of Technology Madras), and A. N. A. N. Rajagopalan (Indian Institute of Technology Madras)</i>	
Rethinking Video Frame Interpolation from Shutter Mode Induced Degradation	12225
<i>Xiang Ji (The University of Tokyo), Zhixiang Wang (The University of Tokyo, National Institute of Informatics), Zhihang Zhong (The University of Tokyo), and Yinqiang Zheng (The University of Tokyo)</i>	
Single Image Deblurring with Row-dependent Blur Magnitude	12235
<i>Xiang Ji (The University of Tokyo), Zhixiang Wang (The University of Tokyo, National Institute of Informatics), Shin'ichi Satoh (National Institute of Informatics), and Yinqiang Zheng (The University of Tokyo)</i>	
Multi-view Self-supervised Disentanglement for General Image Denoising	12247
<i>Hao Chen (University of Birmingham), Chenyuan Qu (University of Birmingham), Yu Zhang (Nanjing University), Chen Chen (University of Central Florida), and Jianbo Jiao (University of Birmingham; University of Oxford)</i>	
Joint Demosaicing and Deghosting of Time-Varying Exposures for Single-Shot HDR Imaging	12258
<i>Jungwoo Kim (KAIST) and Min H. Kim (KAIST)</i>	
Diff-Retinex: Rethinking Low-light Image Enhancement with A Generative Diffusion Model	12268
<i>Xunpeng Yi (Wuhan University), Han Xu (Wuhan University), Hao Zhang (Wuhan University), Linfeng Tang (Wuhan University), and Jiayi Ma (Wuhan University)</i>	
Dual Aggregation Transformer for Image Super-Resolution	12278
<i>Zheng Chen (Shanghai Jiao Tong University), Yulun Zhang (ETH Zurich), Jinjin Gu (The University of Sydney), Linghe Kong (Shanghai Jiao Tong University), Xiaokang Yang (Shanghai Jiao Tong University of China), and Fisher Yu (ETH Zurich)</i>	
Video Object Segmentation-aware Video Frame Interpolation	12288
<i>Jun-Sang Yoo (Korea University), Hongjae Lee (Korea University), and Seung-Won Jung (Korea University)</i>	
RawHDR: High Dynamic Range Image Reconstruction from a Single Raw Image	12300
<i>Yunhao Zou (Beijing Institute of Technology), Chenggang Yan (Hangzhou Dianzi University), and Ying Fu (Beijing Institute of Technology)</i>	
Multi-Scale Residual Low-Pass Filter Network for Image Deblurring	12311
<i>Jiangxin Dong (Nanjing University of Science and Technology), Jinshan Pan (Nanjing University of Science and Technology), Zhongbao Yang (Nanjing University of Science and Technology), and Jinhui Tang (Nanjing University of Science and Technology)</i>	
Indoor Depth Recovery Based on Deep Unfolding with Non-Local Prior	12321
<i>Yuhui Dai (East China Normal University), Junkang Zhang (East China Normal University), Faming Fang (East China Normal University), and Guixu Zhang (East China Normal University)</i>	

Learning Correction Filter via Degradation-Adaptive Regression for Blind Single Image Super-Resolution	12331
<i>Hongyang Zhou (University of Science and Technology Beijing), Xiaobin Zhu (University of Science and Technology Beijing), Jianqing Zhu (College of Engineering, Huaqiao University), Zheng Han (University of Science and Technology Beijing), Shi-Xue Zhang (University of Science and Technology Beijing), Jingyan Qin (University of Science and Technology Beijing), and Xu-Cheng Yin (University of Science and Technology Beijing)</i>	
Learning Non-Local Spatial-Angular Correlation for Light Field Image Super-Resolution	12342
<i>Zhengyu Liang (National University of Defense Technology), Yingqian Wang (National University of Defense Technology), Longguang Wang (National University of Defense Technology), Jungang Yang (National University of Defense Technology), Shilin Zhou (National University of Defense Technology), and Yulan Guo (Sun Yat-sen University)</i>	
Both Diverse and Realism Matter: Physical Attribute and Style Alignment for Rainy Image Generation	12353
<i>Changfeng Yu (Huazhong University of Science and Technology), Shiming Chen (Huazhong University of Science and Technology), Yi Chang (Huazhong University of Science and Technology), Yibing Song (Fudan University), and Luxin Yan (Huazhong University of Science and Technology)</i>	
Learned Image Reasoning Prior Penetrates Deep Unfolding Network for Panchromatic and Multi-spectral Image Fusion	12364
<i>Man Zhou (Chinese Academy of Sciences), Jie Huang (University of Science and Technology of China), Naishan Zheng (University of Science and Technology of China), and Chongyi Li (Nanyang Technological University)</i>	
The Devil is in the Upsampling: Architectural Decisions Made Simpler for Denoising with Deep Image Prior	12374
<i>Yilin Liu (University of North Carolina-Chapel Hill), Jiang Li (The University of North Carolina at Chapel Hill), Yunkui Pang (University of North Carolina Chapel Hill), Dong Nie (UNC), and Pew-Thian Yap (UNC Chapel Hill)</i>	
SimFIR: A Simple Framework for Fisheye Image Rectification with Self-supervised Representation Learning	12384
<i>Hao Feng (University of Science and Technology of China), Wendi Wang (University of Science and Technology of China), Jiajun Deng (The University of Sydney), Wengang Zhou (University of Science and Technology of China), Li Li (University of Science and Technology of China), and Houqiang Li (University of Science and Technology of China)</i>	
Exploring Temporal Frequency Spectrum in Deep Video Deblurring	12394
<i>Qi Zhu (University of Science and Technology of China), Man Zhou (University of Science and Technology of China), Naishan Zheng (University of Science and Technology of China), Chongyi Li (Nanyang Technological University), Jie Huang (University of Science and Technology of China), and Feng Zhao (University of Science and Technology of China)</i>	

ExposureDiffusion: Learning to Expose for Low-light Image Enhancement	12404
<i>Yufei Wang (Nanyang Technological University), Yi Yu (Nanyang Technological University), Wenhan Yang (Peng Cheng Laboratory), Lanqing Guo (Nanyang Technological University), Lap-Pui Chau (The Hong Kong Polytechnic University), Alex C. Kot (Nanyang Technological University), and Bihan Wen (Nanyang Technological University)</i>	
High-Resolution Document Shadow Removal via A Large-Scale Real-World Dataset and A Frequency-Aware Shadow Erasing Net	12415
<i>Zinuo Li (Monash University), Xuhan Chen (University of Macau), Chi-Man Pun (University of Macau), and Xiaodong Cun (Tencent AI Lab)</i>	
Towards Saner Deep Image Registration	12425
<i>Bin Duan (Illinois Institute of Technology), Ming Zhong (Illinois Institute of Technology), and Yan Yan (Illinois Institute of Technology)</i>	
VideoFlow: Exploiting Temporal Cues for Multi-frame Optical Flow Estimation	12435
<i>Xiaoyu Shi (CUHK), Zhaoyang Huang (CUHK), Weikang Bian (The Chinese University of Hong Kong), Dasong Li (Chinese University of Hong Kong), Manyuan Zhang (Sensetime), Ka Chun Cheung (Nvidia), Simon See (NVIDIA AI Tech Centre), Hongwei Qin (Sensetime), Jifeng Dai (Tsinghua University), and Hongsheng Li (The Chinese University of Hong Kong)</i>	
Scene Matters: Model-based Deep Video Compression	12447
<i>Lv Tang (University of Chinese Academy of Sciences), Xinfeng Zhang (University of Chinese Academy of Sciences), Gai Zhang (University of Chinese Academy of Sciences), and Xiaoqi Ma (University of Chinese Academy of Sciences)</i>	
Non-Coaxial Event-Guided Motion Deblurring with Spatial Alignment	12458
<i>Hoonhee Cho (KAIST), Yuhwan Jeong (KAIST), Taewoo Kim (KAIST), and Kuk-Jin Yoon (KAIST)</i>	
Retinexformer: One-stage Retinex-based Transformer for Low-light Image Enhancement	12470
<i>Yuanhao Cai (Tsinghua University, Tsinghua Shenzhen International Graduate School), Hao Bian (Tsinghua Shenzhen International Graduate School, Tsinghua University), Jing Lin (Tsinghua University, Tsinghua Shenzhen International Graduate School), Haoqian Wang (Tsinghua Shenzhen International Graduate School, Tsinghua University), Radu Timofte (University of Wurzburg & ETH Zurich), and Yulun Zhang (ETH Zurich)</i>	
Feature Modulation Transformer: Cross-Refinement of Global Representation via High-Frequency Prior for Image Super-Resolution	12480
<i>Ao Li (University of Electronic Science and Technology of China), Le Zhang (University of Electronic Science and Technology of China), Yun Liu (A*STAR), and Ce Zhu (University of Electronic Science & Technology of China)</i>	
MVPSNet: Fast Generalizable Multi-view Photometric Stereo	12491
<i>Dongxu Zhao (UNC-Chapel Hill), Daniel Lichy (University of Maryland), Pierre-Nicolas Perrin (University of North Carolina at Chapel Hill), Jan-Michael Frahm (UNC-Chapel Hill), and Soumyadip Sengupta (University of North Carolina at Chapel Hill)</i>	

FSI: Frequency and Spatial Interactive Learning for Image Restoration in Under-Display Cameras	12503
<i>Chengxu Liu (Xi'an Jiaotong University), Xuan Wang (megvii), Shuai Li (Megvii), Yuzhi Wang (Megvii Technology), and Xueming Qian (Xi'an Jiaotong University)</i>	
Spherical Space Feature Decomposition for Guided Depth Map Super-Resolution	12513
<i>Zixiang Zhao (Xi'an Jiaotong University), Jianshe Zhang (Xi'an Jiaotong University), Xiang Gu (Xi'an Jiaotong University), Chengli Tan (Xi'an Jiaotong University), Shuang Xu (Northwestern Polytechnical University), Yulun Zhang (ETH Zurich), Radu Timofte (University of Wurzburg & ETH Zurich), and Luc Van Gool (ETH Zurich)</i>	
Empowering Low-Light Image Enhancer through Customized Learnable Priors	12525
<i>Naishan Zheng (University of Science and Technology of China), Man Zhou (University of Science and Technology of China), Yanmeng Dong (University of Science and Technology of China), Xiangyu Rui (Xi'an Jiaotong University), Jie Huang (University of Science and Technology of China), Chongyi Li (Nanyang Technological University), and Feng Zhao (University of Science and Technology of China)</i>	
Learning Image Harmonization in the Linear Color Space	12536
<i>Ke Xu (City University of Hong Kong), Gerhard Petrus Hancke (City University of Hong Kong), and Rynson W.H. Lau (City University of Hong Kong)</i>	
Under-Display Camera Image Restoration with Scattering Effect	12546
<i>Binbin Song (University of Macau), Xiangyu Chen (University of Macau), Shuning Xu (University of Macau), and Jiantao Zhou (University of Macau)</i>	
Iterative Soft Shrinkage Learning for Efficient Image Super-Resolution	12556
<i>Jiamian Wang (Rochester Institute of Technology), Huan Wang (Northeastern University), Yulun Zhang (ETH Zurich), Yun Fu (Northeastern University), and Zhiqiang Tao (Rochester Institute of Technology)</i>	
Single Image Defocus Deblurring via Implicit Neural Inverse Kernels	12566
<i>Yuhui Quan (South China University of Technology), Xin Yao (South China University of Technology), and Hui Ji (National University of Singapore)</i>	
Degradation-Resistant Unfolding Network for Heterogeneous Image Fusion	12577
<i>Chunming He (Tsinghua University), Kai Li (NEC LABORATORIES AMERICA, INC), Guoxia Xu (Nanjing University of Posts and Telecommunications), Yulun Zhang (ETH Zurich), Runze Hu (Beijing Institute of Technology), Zhenhua Guo (Tianyi Traffic Technology), and Xiu Li (Tsinghua University)</i>	
Graphics2RAW: Mapping Computer Graphics Images to Sensor RAW Images	12588
<i>Donghwan Seo (Samsung Electronics), Abhijith Punnappurath (Samsung AI Center Toronto), Luxi Zhao (Samsung AI Center Toronto), Abdelrahman Abdelhamed (Samsung), Sai Kiran Tedla (York University), Sanguk Park (Samsung Electronics Company), Jihwan Choe (Samsung), and Michael S. Brown (York University)</i>	

Lighting up NeRF via Unsupervised Decomposition and Enhancement	12598
<i>Haoyuan Wang (City University of Hong Kong), Xiaogang Xu (The Chinese University of Hong Kong), Ke Xu (City University of Hong Kong), and Rynson W.H. Lau (City University of Hong Kong)</i>	
Unsupervised Image Denoising in Real-World Scenarios via Self-Collaboration Parallel Generative Adversarial Branches	12608
<i>Xin Lin (Sichuan University), Chao Ren (Sichuan University), Xiao Liu (Sichuan University), Jie Huang (Sichuan University), and Yinjie Lei (Sichuan University)</i>	
Adverse Weather Removal with Codebook Priors	12619
<i>Tian Ye (The Hong Kong University of Science and Technology, (Guangzhou), Sixiang Chen (The Hong Kong University of Science and Technology, (Guangzhou), Jinbin Bai (Nanjing University), Jun Shi (Xinjiang University), Chenghao Xue (Jimei University), Jingxia Jiang (jimei university), Junjie Yin (JMU), Erkang Chen (Jimei University), and Yun Liu (Southwest University)</i>	
Deep Video Demoiréing via Compact Invertible Dyadic Decomposition	12631
<i>Yuhui Quan (South China University of Technology), Haoran Huang (South China University of Technology), Shengfeng He (Singapore Management University), and Ruotao Xu (South China University of Technology)</i>	
SILT: Shadow-Aware Iterative Label Tuning for Learning to Detect Shadows from Noisy Labels.	12641
<i>Han Yang (The Chinese University of Hong Kong), Tianyu Wang (The Chinese University of Hong Kong), Xiaowei Hu (Shanghai AI Laboratory), and Chi-Wing Fu (The Chinese University of Hong Kong)</i>	
Innovating Real Fisheye Image Correction with Dual Diffusion Architecture	12653
<i>Shangrong Yang (Beijing Jiaotong University), Chunyu Lin (Beijing Jiaotong University), Kang Liao (Beijing Jiaotong University), and Yao Zhao (Beijing Jiaotong University)</i>	
Adaptive Illumination Mapping for Shadow Detection in Raw Images	12663
<i>Jiayu Sun (Dalian University of Technology), Ke Xu (City University of Hong Kong), Youwei Pang (Dalian University of Technology), Lihe Zhang (Dalian University of Technology), Huchuan Lu (Dalian University of Technology), Gerhard Hancke (City University of Hong Kong), and Rynson Lau (City University of Hong Kong)</i>	
GEDepth: Ground Embedding for Monocular Depth Estimation	12673
<i>Xiaodong Yang (NVIDIA Research), Zhuang Ma (QCraft), Zhiyu Ji (Qcraft), and Zhe Ren (Qcraft)</i>	
Lightweight Image Super-Resolution with Superpixel Token Interaction	12682
<i>Aiping Zhang (Sun Yat-sen University), Wenqi Ren (Sun Yat-Sen University), Yi Liu (Baidu Inc.), and Xiaochun Cao (Sun Yat-sen University)</i>	
Unfolding Framework with Prior of Convolution-Transformer Mixture and Uncertainty Estimation for Video Snapshot Compressive Imaging	12692
<i>Siming Zheng (CNIC) and Xin Yuan (Westlake University)</i>	

Efficient Unified Demosaicing for Bayer and Non-Bayer Patterned Image Sensors	12704
<i>Haechang Lee (Seoul National University), Dongwon Park (Seoul National University), Wongi Jeong (Seoul National University), Kijeong Kim (SK Hynix), Hyunwoo Je (SK hynix), Dongil Ryu (SK Hynix), and Se Young Chun (Seoul National University)</i>	
LAN-HDR: Luminance-based Alignment Network for High Dynamic Range Video Reconstruction	12714
<i>Haesoo Chung (Seoul National University) and Nam Ik Cho (Seoul National University)</i>	
Fine-grained Visible Watermark Removal	12724
<i>Li Niu (Shanghai Jiao Tong University), Xing Zhao (Shanghai Jiao Tong University), Bo Zhang (Shanghai Jiao Tong University), and Liqing Zhang (Shanghai Jiao Tong University)</i>	
SRFormer: Permuted Self-Attention for Single Image Super-Resolution	12734
<i>Yupeng Zhou (Nankai University), Zhen Li (Nankai University), Chun-Le Guo (Nankai University), Song Bai (University of Oxford), Ming-Ming Cheng (Nankai University), and Qibin Hou (Nankai University)</i>	
DLGSANet: Lightweight Dynamic Local and Global Self-Attention Networks for Image Super-Resolution	12746
<i>Xiang Li (Nanjing University of Science and Technology), Jiangxin Dong (Nanjing University of Science and Technology), Jinhui Tang (Nanjing University of Science and Technology), and Jinshan Pan (Nanjing University of Science and Technology)</i>	
MB-TaylorFormer: Multi-Branch Efficient Transformer Expanded by Taylor Formula for Image Dehazing	12756
<i>Qiu Yuwei (Sun Yat-sen University), Kaihao Zhang (Australian National University), Chenxi Wang (Sun yat-sen university), Wenhan Luo (Sun Yat-sen University), Hongdong Li (Australian National University, Australia), and Zhi Jin (Sun Yat-sen University)</i>	
Multi-Frequency Representation Enhancement with Privilege Information for Video Super-Resolution	12768
<i>Fei Li (China Agricultural University), Linfeng Zhang (Tsinghua University), Zikun Liu (Samsung Research China – Beijing, (SRC-B)), Juan Lei (Samsung Research China – Beijing, (SRC-B)), and Zhenbo Li (China Agricultural University)</i>	
COMPASS: High-Efficiency Deep Image Compression with Arbitrary-scale Spatial Scalability ...	12780
<i>Jongmin Park (KAIST), Jooyoung Lee (ETRI), and Munchurl Kim (Korea Advanced Institute of Science and Technology)</i>	
Alignment-free HDR Deghosting with Semantics Consistent Transformer	12790
<i>Steven Tel (University of Burgundy, France), Zongwei Wu (Univ. Bourgogne Franche-Comte, France), Yulun Zhang (ETH Zurich), Barthélemy Heyrman (University of Burgundy, France), Cédric Demonceaux (Université de Bourgogne), Radu Timofte (University of Wurzburg & ETH Zurich), and Dominique Ginhac (Le2i - University of Burgundy, France)</i>	

From Chaos Comes Order: Ordering Event Representations for Object Recognition and Detection	12800
<i>Nikola Zubić (University of Zurich), Daniel Gehrig (University of Zurich and ETH Zurich), Mathias Gehrig (University of Zurich), and Davide Scaramuzza (University of Zurich, Switzerland)</i>	
Towards High-Quality Specular Highlight Removal by Leveraging Large-Scale Synthetic Data ..	12811
<i>Gang Fu (The Hong Kong Polytechnic University), Qing Zhang (Sun Yat-sen University), Lei Zhu (The Hong Kong University of Science and Technology, (Guangzhou), Chunxia Xiao (Wuhan University), and Ping Li (The Hong Kong Polytechnic University)</i>	
DynamicISP: Dynamically Controlled Image Signal Processor for Image Recognition	12820
<i>Masakazu Yoshimura (Sony Group Corporation), Junji Otsuka (Sony Group Corporation), Atsushi Irie (Sony Group Corporation), and Takeshi Ohashi (Sony Group Corporation)</i>	
Dancing in the Dark: A Benchmark towards General Low-light Video Enhancement	12831
<i>Huiyuan Fu (Beijing University of Posts and Telecommunications), Wenkai Zheng (Beijing University of Posts and Telecommunications), Xicong Wang (Beijing University of Posts and Telecommunications), Jiaxuan Wang (Beijing University of Posts and Telecommunications), Heng Zhang (Xiaomi), and Huadong Ma (Beijing University of Posts and Telecommunications)</i>	
Dec-Adapter: Exploring Efficient Decoder-Side Adapter for Bridging Screen Content and Natural Image Compression	12841
<i>Sheng Shen (Tianjin University), Huanjing Yue (Tianjin University), and Jingyu Yang (Tianjin University)</i>	
OmniZoomer: Learning to Move and Zoom in on Sphere at High-Resolution	12851
<i>Zidong Cao (HKUST), Hao Ai (The Hong Kong University of Science and Technology), Yan-Pei Cao (Tencent), Ying Shan (Tencent), Xiaohu Qie (Tencent), and Lin Wang (HKUST)</i>	
Pyramid Dual Domain Injection Network for Pan-sharpening	12862
<i>Xuanhua He (University of Science and Technology of China), Keyu Yan (University of Science and Technology of China), Rui Li (Institute of Intelligent Machines, Chinese Academy of Sciences China), Chengjun Xie (Institute of Intelligent Machines, Chinese Academy of Sciences China), Jie Zhang (Hefei Institute of Physical Sciences, Chinese Academy of Sciences), and Man Zhou (University of Science and Technology of China)</i>	
Implicit Neural Representation for Cooperative Low-light Image Enhancement	12872
<i>Shuzhou Yang (Peking University), Moxuan Ding (Independent Researcher), Yanmin Wu (Peking University), Zihan Li (University of Washington), and Jian Zhang (Peking University)</i>	
Physically-Plausible Illumination Distribution Estimation	12882
<i>Egor Ershov (IITP RAS), Vasily Tesalin (IITP), Ivan Ermakov (IITP RAS), and Michael S. Brown (York University)</i>	

Score Priors Guided Deep Variational Inference for Unsupervised Real-World Single Image Denoising	12891
<i>Jun Cheng (Huazhong University of Science and Technology), Tao Liu (Huazhong University of Science and Technology), and Shan Tan (Huazhong University of Science and Technology)</i>	
Semantic-Aware Dynamic Parameter for Video Inpainting Transformer	12903
<i>Eunhye Lee (Hanyang University), Jinsu Yoo (The Ohio State University), Yunjeong Yang (Hanyang University), Sungyong Baik (Hanyang University), and Tae Hyun Kim (Hanyang University)</i>	
Pixel Adaptive Deep Unfolding Transformer for Hyperspectral Image Reconstruction	12913
<i>Miaoyu Li (Beijing Institute of Technology), Ying Fu (Beijing Institute of Technology), Ji Liu (Baidu Research), and Yulun Zhang (ETH Zurich)</i>	
Improving Lens Flare Removal with General-Purpose Pipeline and Multiple Light Sources Recovery	12923
<i>Yuyan Zhou (Nanjing University of Aeronautics and Astronautics), Dong Liang (Nanjing University of Aeronautics and Astronautics), Songcan Chen (Nanjing University of Aeronautics and Astronautics), Sheng-Jun Huang (Nanjing University of Aeronautics and Astronautics), Shuo Yang (vivo Mobile Communication Co., Ltd), and Chongyi Li (Nanyang Technological University)</i>	
RFD-ECNet: Extreme Underwater Image Compression with Reference to Feature Dictionary	12934
<i>Mengyao Li (Shanghai university), Liqun Shen (Shanghai University), Peng Ye (fudan university), Guorui Feng (Shanghai University), and Zheyin Wang (Shanghai University)</i>	
Learning Continuous Exposure Value Representations for Single-Image HDR Reconstruction	12944
<i>Su-Kai Chen (National Yang Ming Chiao Tung University), Hung-Lin Yen (National Yang Ming Chiao Tung University), Yu-Lun Liu (National Yang Ming Chiao Tung University), Min-Hung Chen (NVIDIA), Hou-Ning Hu (Mediatek Inc.), Wen-Hsiao Peng (National Yang Ming Chiao Tung University), and Yen-Yu Lin (National Yang Ming Chiao Tung University)</i>	
Focal Network for Image Restoration	12955
<i>Yuning Cui (Technical University of Munich), Wenqi Ren (Sun Yat-Sen University), Xiaochun Cao (Sun Yat-sen University), and Alois Knoll (Robotics and Embedded Systems)</i>	
CIRI: Curricular Inactivation for Residue-aware One-shot Video Inpainting	12966
<i>Weiyang Zheng (South China University of Technology), Cheng Xu (South China University of Technology), Xuemiao Xu (South China University of Technology), Wenxi Liu (Fuzhou University), and Shengfeng He (Singapore Management University)</i>	
Beyond Image Borders: Learning Feature Extrapolation for Unbounded Image Composition	12977
<i>Xiaoyu Liu (Harbin Institute of Technology), Ming Liu (Harbin Institute of Technology, China), Junyi Li (Harbin Institute of Technology), Shuai Liu (Xiaomi), Xiaotao Wang (Xiaomi), Lei Lei (Xiaomi), and Wangmeng Zuo (Harbin Institute of Technology, China)</i>	

MetaF2N: Blind Image Super-Resolution by Learning Efficient Model Adaptation from Faces	12987
<i>Zhicun Yin (Harbin Institute of Technology, China), Ming Liu (Harbin Institute of Technology, China), Xiaoming Li (Nanyang Technological University), Hui Yang (Transsion), Longan Xiao (Shanghai Transsion Information Technology Co., Ltd), and Wangmeng Zuo (Harbin Institute of Technology, China)</i>	
Boundary-Aware Divide and Conquer: A Diffusion-Based Solution for Unsupervised Shadow Removal	12999
<i>Lanqing Guo (Nanyang Technological University), Chong Wang (Nanyang Technological University), Wenhan Yang (Peng Cheng Laboratory), Yufei Wang (Nanyang Technological University), and Bihan Wen (Nanyang Technological University)</i>	
Leveraging Inpainting for Single-Image Shadow Removal	13009
<i>Xiaoguang Li (University of South Carolina), Qing Guo (A*STAR), Rabab Abdelfattah (University of South Carolina; University of Southern Mississippi), Di Lin (Tianjin University), Wei Feng (College of Intelligence and Computing, Tianjin University, China), Ivor Tsang (A*STAR), and Song Wang (University of South Carolina)</i>	
Hybrid Spectral Denoising Transformer with Guided Attention	13019
<i>Zeqiang Lai (Beijing Institute of Technology), Chenggang Yan (Hangzhou Dianzi University), and Ying Fu (Beijing Institute of Technology)</i>	
Examining Autoexposure for Challenging Scenes	13030
<i>SaiKiran Tedla (York University), Beixuan Yang (York University), and Michael S. Brown (York University)</i>	
Self-supervised Learning to Bring Dual Reversed Rolling Shutter Images Alive	13040
<i>Wei Shang (Harbin Institute of Technology), Dongwei Ren (Harbin Institute of Technology), Chaoyu Feng (xiaomi), Wang Xiaotao (Xiaomi), Lei Lei (Xiaomi), and Wangmeng Zuo (Harbin Institute of Technology, China)</i>	
DiffIR: Efficient Diffusion Model for Image Restoration	13049
<i>Bin Xia (Tsinghua university), Yulun Zhang (ETH Zurich), Shiyin Wang (ByteDance), Yitong Wang (Bytedance), Xinglong Wu (Bytedance), Yapeng Tian (The University of Texas at Dallas), Wenming Yang (Tsinghua University), and Luc Van Gool (ETH Zurich)</i>	
Sparse Sampling Transformer with Uncertainty-Driven Ranking for Unified Removal of Raindrops and Rain Streaks	13060
<i>Sixiang Chen (The Hong Kong University of Science and Technology, (Guangzhou), Tian Ye (The Hong Kong University of Science and Technology, (Guangzhou), Jinbin Bai (Nanjing University), Erkang Chen (Jimei University), Jun Shi (Xinjiang University), and Lei Zhu (The Hong Kong University of Science and Technology, (Guangzhou)</i>	
LMR: A Large-Scale Multi-Reference Dataset for Reference-Based Super-Resolution	13072
<i>Lin Zhang (CASIA), Xin Li (Baidu), Dongliang He (ByteDance), Fu Li (Baidu), Errui Ding (Baidu Inc.), and Zhaoxiang Zhang (Chinese Academy of Sciences, China)</i>	

Low-Light Image Enhancement with Illumination-Aware Gamma Correction and Complete Image Modelling Network	13082
<i>Yinglong Wang (Huawei Noah's Ark Lab), Zhen Liu (Megvii Technology Inc.), Jianzhuang Liu (Huawei Noah's Ark Lab), Songcen Xu (Huawei Noah's Ark Lab), and Shuaicheng Liu (UESTC, Megvii)</i>	
Single Image Reflection Separation via Component Synergy	13092
<i>Qiming Hu (Tianjin University) and Xiaojie Guo (Tianjin University)</i>	
Learning Rain Location Prior for Nighttime Deraining	13102
<i>Fan Zhang (Beijing Institute of Technology), Shaodi You (University of Amsterdam), Yu Li (International Digital Economy Academy), and Ying Fu (Beijing Institute of Technology)</i>	
Exploring Positional Characteristics of Dual-Pixel Data for Camera Autofocus	13112
<i>Myungsub Choi (Samsung Electronics), Hana Lee (Samsung Advanced Institute of Technology), and Hyong-euk Lee (Samsung Advanced Institute of Technology)</i>	
Continuously Masked Transformer for Image Inpainting	13123
<i>Keunsoo Ko (Catholic University of Korea) and Chang-Su Kim (Korea university)</i>	
Learning Data-Driven Vector-Quantized Degradation Model for Animation Video Super-Resolution	13133
<i>Zixi Tuo (Xi'an Jiaotong University), Huan Yang (Microsoft Research), Jianlong Fu (Microsoft Research), Yujie Dun (Xi'an Jiaotong University), and Xueming Qian (Xi'an Jiaotong University)</i>	
Spatially-Adaptive Feature Modulation for Efficient Image Super-Resolution	13144
<i>Long Sun (Nanjing University of Science and Technology), Jiangxin Dong (Nanjing University of Science and Technology), Jinhui Tang (Nanjing University of Science and Technology), and Jinshan Pan (Nanjing University of Science and Technology)</i>	
Video Adverse-Weather-Component Suppression Network via Weather Messenger and Adversarial Backpropagation	13154
<i>Yijun Yang (The Hong Kong University of Science and Technology, (Guangzhou), Angelica I. Aviles-Rivero (University of Cambridge), Huazhu Fu (Institute of High Performance Computing, A*STAR), Ye Liu (College of Intelligence and Computing, Tianjin University), Weiming Wang (Hong Kong Metropolitan University), and Lei Zhu (The Hong Kong University of Science and Technology, (Guangzhou))</i>	
Snow Removal in Video: A New Dataset and A Novel Method	13165
<i>Haoyu Chen (The Hong Kong University of Science and Technology, (Guangzhou), Jingjing Ren (The Hong Kong University of Science and Technology, (Guangzhou), Jinjin Gu (The University of Sydney), Hongtao Wu (The Hong Kong University of Science and Technology, (Guangzhou), Xuequan Lu (La Trobe University), Haoming Cai (University of Maryland, College Park), and Lei Zhu (The Hong Kong University of Science and Technology, (Guangzhou))</i>	

Boosting Single Image Super-Resolution via Partial Channel Shifting	13177
<i>Xiaoming Zhang (School of Computing and Artificial Intelligence, Southwest Jiaotong University), Tianrui Li (School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu, 611756, China), and Xiaole Zhao (School of Computing and Artificial Intelligence, Southwest Jiaotong University)</i>	
Towards Real-World Burst Image Super-Resolution: Benchmark and Method	13187
<i>Pengxu Wei (Sun Yat-sen University), Yujing Sun (Peking University), Xingbei Guo (Sun Yat-sen University), Chang Liu (Tsinghua University), Guanbin Li (Sun Yat-sen University), Jie Chen (Peking University), Xiangyang Ji (Tsinghua University), and Liang Lin (Sun Yat-sen University)</i>	
On the Effectiveness of Spectral Discriminators for Perceptual Quality Improvement	13197
<i>Xin Luo (University of Science and Technology of China), Yunan Zhu (University of Science and Technology of China), Shunxin Xu (University of Science and Technology of China), and Dong Liu (University of Science and Technology of China)</i>	
E2NeRF: Event Enhanced Neural Radiance Fields from Blurry Images	13208
<i>Yunshan Qi (Beihang University), Lin Zhu (Beijing Institute of Technology), Yu Zhang (SenseTime and Tetras.AI), and Jia Li (Beihang University; Peng Cheng Laboratory)</i>	
Iterative Denoiser and Noise Estimator for Self-Supervised Image Denoising	13219
<i>Yunhao Zou (Beijing Institute of Technology), Chenggang Yan (Hangzhou Dianzi University), and Ying Fu (Beijing Institute of Technology)</i>	
Lighting Every Darkness in Two Pairs: A Calibration-Free Pipeline for RAW Denoising	13229
<i>Xin Jin (Nankai University), Jia-Wen Xiao (Nankai University), Ling-Hao Han (Nankai University), Chunle Guo (Nankai University), Ruixun Zhang (Peking University), Xialei Liu (Nankai University), and Chongyi Li (Nanyang Technological University)</i>	
Fingerprinting Deep Image Restoration Models	13239
<i>Yuhui Quan (South China University of Technology), Huan Teng (South China University of Technology), Ruotao Xu (South China University of Technology), Jun Huang (Alibaba Group), and Hui Ji (National University of Singapore)</i>	

Computer Vision Theory

Environment-Invariant Curriculum Relation Learning for Fine-Grained Scene Graph Generation.....	13250
<i>Yukuan Min (Xidian University), Aming Wu (Xidian University), and Cheng Deng (Xidian University)</i>	
DCPB: Deformable Convolution Based on the Poincaré Ball for Top-view Fisheye Cameras	13262
<i>Xuan Wei (Southeast University), Zhidan Ran (Southeast University), and Xiaobo Lu (Southeast University)</i>	
FemtoDet: An Object Detection Baseline for Energy Versus Performance Tradeoffs	13272
<i>Peng Tu (MicroBT Inc.), Xu Xie (MicroBT Inc.), Guo Ai (MicroBT Inc.), Yuexiang Li (Jarvis Lab, Tencent), Yawen Huang (Jarvis Lab, Tencent), and Yefeng Zheng (Jarvis Lab, Tencent)</i>	

Curvature-Aware Training for Coordinate Networks	13282
<i>Hemant Saratchandran (University of Adelaide/Australian Institute of Machine Learning), Shin-Fang Chng (The University of Adelaide), Sameera Ramasinghe (University of Adelaide), Lachlan MacDonald (Johns Hopkins University), and Simon Lucey (University of Adelaide)</i>	
Yes, we CANN: Constrained Approximate Nearest Neighbors for Local Feature-Based Visual Localization	13293
<i>Dror Aiger (Google), Andre Araujo (Google), and Simon Lynen (Google)</i>	
Unleashing the Potential of Spiking Neural Networks with Dynamic Confidence	13304
<i>Chen Li (Mr), Edward G Jones (University of Manchester), and Steve Furber (University of Manchester)</i>	
Minimal Solutions to Uncalibrated Two-view Geometry with Known Epipoles	13315
<i>Gaku Nakano (NEC Corporation)</i>	
FBLNet: FeedBack Loop Network for Driver Attention Prediction	13325
<i>Yilong Chen (Chongqing University), Zhixiong Nan (Chongqing University), and Tao Xiang (Chongqing University)</i>	
Deep Feature Deblurring Diffusion for Detecting Out-of-Distribution Objects	13335
<i>Aming Wu (Xidian University), Da Chen (University of Bath), and Cheng Deng (Xidian University)</i>	

Video Analysis and Understanding

Long-range Multimodal Pretraining for Movie Understanding	13346
<i>Dawit Mureja Argaw (KAIST), Joon-Young Lee (Adobe Research), Markus Woodson (Adobe), In So Kweon (KAIST), and Fabian Caba Heilbron (Adobe Research)</i>	
Cross-view Semantic Alignment for Livestreaming Product Recognition	13358
<i>Wenjie Yang (Kuaishou), Yiyi Chen (kuaishou), Yan Li (Kuaishou), Yanhua Cheng (Kuaishou), Xudong Liu (Kuaishou Inc.), Quan Chen (Kuaishou), and Han Li (Kuaishou)</i>	
HTML: Hybrid Temporal-scale Multimodal Learning Framework for Referring Video Object Segmentation	13368
<i>Mingfei Han (University of Technology Sydney), Yali Wang (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences), Zhihui Li (University of New South Wales), Lina Yao (University of New South Wales), Xiaojun Chang (University of Technology Sydney), and Yu Qiao (Shanghai AI Laboratory)</i>	
DyGait: Exploiting Dynamic Representations for High-performance Gait Recognition	13378
<i>Ming Wang (Beijing Jiaotong university), Xianda Guo (PhiGent Robotics), Beibei Lin (National University of Singapore), Tian Yang (University of Auckland), Zheng Zhu (Tsinghua University), Lincheng Li (NetEase Fuxi AI Lab), Shunli Zhang (Beijing Jiaotong University), and Xin Yu (University of Queensland)</i>	
Identity-Consistent Aggregation for Video Object Detection	13388
<i>Chaorui Deng (University of Adelaide), Da Chen (University of Bath), and Qi Wu (University of Adelaide)</i>	

Augmenting and Aligning Snippets for Few-Shot Video Domain Adaptation	13399
<i>Yuecong Xu (Institute for Infocomm Research, A*STAR, Singapore), Jianfei Yang (Nanyang Technological University), Yunjiao Zhou (Nanyang Technological University), Zhenghua Chen (Institute for Infocomm Research, A*STAR, Singapore), Min Wu (Institute for Infocomm Research, A*STAR, Singapore), and Xiaoli Li (Institute for Infocomm Research, A*STAR, Singapore/Nanyang Technological University)</i>	
Action Sensitivity Learning for Temporal Action Localization	13411
<i>Jiayi Shao (Zhejiang University), Xiaohan Wang (Zhejiang University), Ruijie Quan (University of Technology Sydney), Junjun Zheng (Alibaba Group), Jiang Yang (Alibaba Group), and Yi Yang (Zhejiang University)</i>	
SwinLSTM: Improving Spatiotemporal Prediction Accuracy using Swin Transformer and LSTM	13424
<i>Song Tang (Hainan University), Chuang Li (Hainan University), Pu Zhang (Hainan University), and RongNian Tang (HAINAN UNIVERSITY)</i>	
LVOS: A Benchmark for Long-term Video Object Segmentation	13434
<i>Lingyi Hong (Fudan University), Wenchao chen (Fudan university), Zhongying Liu (Fudan University), Wei Zhang (Fudan University), Pinxue Guo (Fudan University), Zhaoyu Chen (Fudan University), and Wenqiang Zhang (Fudan University)</i>	
MGMAE: Motion Guided Masking for Video Masked Autoencoding	13447
<i>Bingkun Huang (Nanjing University), Zhiyu Zhao (Nanjing University), Guozhen Zhang (Nanjing University), Yu Qiao (Shanghai AI Laboraotry), and Limin Wang (Nanjing University)</i>	
Markov Game Video Augmentation for Action Segmentation	13459
<i>Nicolas Aziere (Oregon State University) and Sinisa Todorovic (Oregon State U)</i>	
COOL-CHIC: Coordinate-based Low Complexity Hierarchical Image Codec	13469
<i>Théo Ladune (Orange), Pierrick Philippe (Orange, Rennes), Félix Henry (Orange), Gordon Clare (orange), and Thomas Leguay (Orange Innovation)</i>	
ReGen: A good Generative Zero-Shot Video Classifier Should be Rewarded	13477
<i>Adrian Bulat (Samsung AI Center, Cambridge), Enrique Sanchez (Samsung AI Centre, Cambridge), Brais Martinez (Samsung AI Center), and Georgios Tzimiropoulos (Queen Mary University of London)</i>	
Task Agnostic Restoration of Natural Video Dynamics	13488
<i>Muhammad Kashif Ali (Hanyang University), Dongjin Kim (Hanyang University), and Tae Hyun Kim (Hanyang Univeristy)</i>	
Normalizing Flows for Human Pose Anomaly Detection	13499
<i>Or Hirschorn (Tel Aviv University) and Shai Avidan (Tel Aviv University)</i>	
Movement Enhancement toward Multi-Scale Video Feature Representation for Temporal Action Detection	13509
<i>Zixuan Zhao (Shanghai Jiao Tong University), Dongqi Wang (Shanghai Jiao Tong University), and Xu Zhao (Shanghai Jiao Tong University)</i>	
Event-Guided Procedure Planning from Instructional Videos with Text Supervision	13519
<i>An-Lan Wang (Sun Yat-sen University), Kun-Yu Lin (Sun Yat-sen University), Jia-Run Du (Sun Yat-sen University), Jingke Meng (Sun Yat-Sen University), and Wei-Shi Zheng (Sun Yat-sen University, China)</i>	

SCANet: Scene Complexity Aware Network for Weakly-Supervised Video Moment Retrieval ...	13530
<i>Sunjae Yoon (KAIST), Gwanhyeong Koo (KAIST), Dahyun Kim (KAIST), and Chang D. Yoo (KAIST)</i>	
Spatio-temporal Prompting Network for Robust Video Feature Extraction	13541
<i>Guanxiong Sun (Queen's University Belfast), Chi Wang (Queen's University Belfast), Zhaoyu Zhang (Queen's University Belfast), Jiankang Deng (Imperial College London), Stefanos Zafeiriou (Imperial College London), and Yang Hua (Queen's University Belfast)</i>	
TeD-SPAD: Temporal Distinctiveness for Self-Supervised Privacy-Preservation for Video Anomaly Detection	13552
<i>Joseph Fioresi (University of Central Florida), Ishan Rajendrakumar Dave (University of Central Florida), and Mubarak Shah (University of Central Florida)</i>	
Non-Semantics Suppressed Mask Learning for Unsupervised Video Semantic Compression	13564
<i>Yuan Tian (Shanghai Jiao Tong University), Guo Lu (Shanghai Jiao Tong University), Guangtao Zhai (Shanghai Jiao Tong University), and Zhiyong Gao (Shanghai Jiao Tong University)</i>	
UnLoc: A Unified Framework for Video Localization Tasks	13577
<i>Shen Yan (Google), Xuehan Xiong (Google), Arsha Nagrani (Google), Anurag Arnab (Google), Zhonghao Wang (UIUC), Weina Ge (Google), David Ross (Google), and Cordelia Schmid (Google)</i>	
SkeleTR: Towards Skeleton-based Action Recognition in the Wild	13588
<i>Haodong Duan (The Chinese University of Hong Kong), Mingze Xu (Amazon), Bing Shuai (Amazon), Davide Modolo (Amazon), Zhuowen Tu (UC San Diego), Joseph Tighe (Amazon), and Alessandro Bergamo (Amazon)</i>	
AutoAD II: The Sequel – Who, When, and What in Movie Audio Description	13599
<i>Tengda Han (University of Oxford), Max Bain (University of Oxford), Arsha Nagrani (Google), Gul Varol (Ecole des Ponts ParisTech), Weidi Xie (Shanghai Jiao Tong University), and Andrew Zisserman (University of Oxford)</i>	
What Can a Cook in Italy Teach a Mechanic in India? Action Recognition Generalisation Over Scenarios and Locations	13610
<i>Chiara Plizzari (Politecnico di Torino), Toby Perrett (University of Bristol), Barbara Caputo (Politecnico di Torino), and Dima Damen (University of Bristol)</i>	
Localizing Moments in Long Video Via Multimodal Guidance	13621
<i>Wayner Barrios (Dartmouth), Mattia Soldan (KAUST), Alberto Mario Ceballos-Arroyo (Northeastern University), Fabian Caba Heilbron (Adobe Research), and Bernard Ghanem (KAUST)</i>	
LAC - Latent Action Composition for Skeleton-based Action Segmentation	13633
<i>Di Yang (INRIA), Yaohui Wang (INRIA), Antitza Dantcheva (INRIA), Quan Kong (Woven Planet Holdings), Lorenzo Garattoni (Toyota-Europe), Gianpiero Francesca (Toyota-Europe), and Francois Bremond (Inria Sophia Antipolis, France)</i>	
RIGID: Recurrent GAN Inversion and Editing of Real Face Videos	13645
<i>Yangyang Xu (The University of HongKong), Shengfeng He (Singapore Management University), Kwan-Yee K. Wong (The University of Hong Kong), and Ping Luo (The University of Hong Kong)</i>	

Uncertainty-aware State Space Transformer for Egocentric 3D Hand Trajectory Forecasting	13656
<i>Wentao Bao (Michigan State University), Lele Chen (University of Rochester), Libing Zeng (Texas A&M University), Zhong Li (InnoPeak Technology, Inc.), Yi Xu (OPPO US Research Center), Junsong Yuan (State University of New York at Buffalo, USA), and Yu Kong (Michigan State University)</i>	
What Can Simple Arithmetic Operations Do for Temporal Modeling?	13666
<i>Wenhao Wu (The University of Sydney), Yuxin Song (Baidu), Zhun Sun (Tohoku University), Jingdong Wang (Baidu), Chang Xu (University of Sydney), and Wanli Ouyang (The University of Sydney)</i>	
UATVR: Uncertainty-Adaptive Text-Video Retrieval	13677
<i>Bo Fang (UCAS), Wenhao Wu (The University of Sydney), Chang Liu (Tsinghua University), Yu Zhou (Institute of Information Engineering, CAS; Also with University of Chinese Academy of Sciences), Yuxin Song (Baidu), Weiping Wang (Institute of Information Engineering, CAS, China), Xiangbo Shu (Nanjing University of Science and Technology), Xiangyang Ji (Tsinghua University), and Jingdong Wang (Baidu)</i>	
D3G: Exploring Gaussian Prior for Temporal Sentence Grounding with Glance Annotation	13688
<i>Hanjun Li (Tencent Youtu Lab), Xiujun Shu (Tencent Youtu Lab), Sunan He (Hong Kong University of Science and Technology), Ruizhi Qiao (Tencent Youtu Lab), Wei Wen (Tencent Youtu Lab), Taian Guo (Tencent Youtu Lab), Bei Gan (Tencent), and Xing Sun (Tencent)</i>	
Unsupervised Open-Vocabulary Object Localization in Videos	13701
<i>Ke Fan (Fudan University), Zechen Bai (Amazon), Tianjun Xiao (Amazon), Dominik Zietlow (Amazon Web Services), Max Horn (Amazon Web Services), Zixu Zhao (Amazon), Carl-Johann Simon-Gabriel (ETH Zürich), Mike Zheng Shou (National University of Singapore), Francesco Locatello (Amazon), Bernt Schiele (MPI Informatics), Thomas Brox (University of Freiburg), Zheng Zhang (Amazon), Yanwei Fu (Fudan University), and Tong He (Amazon)</i>	
HiVLP: Hierarchical Interactive Video-Language Pre-Training	13710
<i>Bin Shao (Huawei Noah's Ark Lab), Jianzhuang Liu (Huawei Noah's Ark Lab), Renjing Pei (HuaWei), Songcen Xu (Huawei Noah's Ark Lab), Peng Dai (Huawei Technologies Canada), Juwei Lu (Huawei Noah's Ark Lab), Weimian Li (Huawei), and Youliang Yan (Huawei Noah's Ark Lab)</i>	
Scanning Only Once: An End-to-end Framework for Fast Temporal Grounding in Long Videos .	13721
<i>Yulin Pan (Alibaba Group), Xiangteng He (Peking University), Biao Gong (Alibaba Group), Yiliang Lv (Alibaba Group), Yujun Shen (Dept. of IE, CUHK), Yuxin Peng (Peking University), and Deli Zhao (Alibaba Group)</i>	
Video-FocalNets: Spatio-Temporal Focal Modulation for Video Action Recognition	13732
<i>Syed Talal Wasim (Mohamed Bin Zayed University of Artificial Intelligence), Muhammad Uzair Khattak (Mohamed Bin Zayed University of Artificial Intelligence, (MBZUAI), Muzammal Naseer (MBZUAI), Salman Khan (MBZUAI), Mubarak Shah (University of Central Florida), and Fahad Shahbaz Khan (MBZUAI)</i>	

Lip2Vec: Efficient and Robust Visual Speech Recognition via Latent-to-Latent Visual to Audio Representation Mapping	13744
<i>Yasser Abdelaziz Dahou Djilali (Dublin City UNIVERISTY), Sanath Narayan (Technology Innovation Institute), Haithem Boussaid (Philips), Ebtessam Almazrouei (Technology Innovation Institute, (TII), and Merouane Debbah (Technology Innovation Institute)</i>	
Video OWL-ViT: Temporally-consistent Open-world Localization in Video	13756
<i>Georg Heigold (Google), Matthias Minderer (Google Research), Alexey Gritsenko (Google Brain), Alex Bewley (Google), Daniel Keysers (Google), Mario Lucic (Google Brain), Fisher Yu (ETH Zurich), and Thomas Kipf (Google Brain)</i>	
Tubelet-Contrastive Self-Supervision for Video-Efficient Generalization	13766
<i>Fida Mohammad Thoker (University of Amsterdam), Hazel Doughty (Leiden University), and Cees G. M. Snoek (University of Amsterdam)</i>	
Memory-and-Anticipation Transformer for Online Action Understanding	13778
<i>Jiahao Wang (Nanjing University), Guo Chen (Nanjing University), Yifei Huang (The University of Tokyo), Limin Wang (Nanjing University), and Tong Lu (Nanjing University)</i>	
Video Action Segmentation via Contextually Refined Temporal Keypoints	13790
<i>Borui Jiang (Peking University), Yang Jin (Peking University), Zhentao Tan (Peking University), and Yadong Mu (Peking University)</i>	
Knowing Where to Focus: Event-aware Transformer for Video Grounding	13800
<i>Jinhyun Jang (Yonsei University), Jungin Park (Yonsei University), Jin Kim (Yonsei University), Hyeongjun Kwon (Yonsei University), and Kwanghoon Sohn (Yonsei Univ.)</i>	
MPI-Flow: Learning Realistic Optical Flow with Multiplane Images	13811
<i>Yingping Liang (Beijing Institute of Technology), Jiaming Liu (Baidu Inc.), Debing Zhang (Xiaohongshu), and Ying Fu (Beijing Institute of Technology)</i>	
Discovering Spatio-Temporal Rationales for Video Question Answering	13823
<i>Yicong Li (National University of Singapore), Junbin Xiao (National University of Singapore), Chun Feng (University of Science and Technology of China), Xiang Wang (National University of Singapore), and Tat-Seng Chua (National university of Singapore)</i>	
Scalable Video Object Segmentation with Simplified Framework	13833
<i>Qiangqiang Wu (City University of Hong Kong), Tianyu Yang (International Digital Economy Academy), Wei Wu (City University of Hong Kong), and Antoni B. Chan (City University of Hong Kong, Hong Kong)</i>	
Root Pose Decomposition Towards Generic Non-rigid 3D Reconstruction with Monocular Videos	13844
<i>Yikai Wang (Tsinghua University), Yinpeng Dong (Tsinghua University), Fuchun Sun (Tsinghua University), and Xiao Yang (Tsinghua University)</i>	
Helping Hands: An Object-Aware Ego-Centric Video Recognition Model	13855
<i>Chuhan Zhang (University of Oxford), Ankush Gupta (DeepMind), and Andrew Zisserman (University of Oxford)</i>	

Modeling the Relative Visual Tempo for Self-supervised Skeleton-based Action Recognition	13867
<i>Yisheng Zhu (Nanjing University of Posts and Telecommunications), Hu Han (Institute of Computing Technology, Chinese Academy of Sciences), Zhengtao Yu (Kunming University of Science and Technology), and Guangcan Liu (Southeast University)</i>	
Tube-Link: A Flexible Cross Tube Framework for Universal Video Segmentation	13877
<i>Xiangtai Li (Nanyang Technological University), Haobo Yuan (Nanyang Technological University), Wenwei Zhang (NTU), Guangliang Cheng (University of Liverpool), Jiangmiao Pang (Shanghai AI Laboratory), and Chen Change Loy (Nanyang Technological University)</i>	
Disentangling Spatial and Temporal Learning for Efficient Image-to-Video Transfer Learning....	13888
<i>Zhiwu Qing (Huazhong University of Science and Technology), Shiwei Zhang (DAMO Academy, Alibaba Group), Ziyuan Huang (National University of Singapore), Yingya Zhang (Alibaba Group), Changxin Gao (Huazhong University of Science and Technology), Deli Zhao (Alibaba Group), and Nong Sang (Huazhong University of Science and Technology)</i>	
Tem-Adapter: Adapting Image-Text Pretraining for Video Question Answer	13899
<i>Guangyi Chen (MBZUAI), Xiao Liu (Eindhoven University of Technology), Guangrun Wang (University of Oxford), Kun Zhang (Carnegie Mellon University), Philip H.S. Torr (University of Oxford), Xiao-Ping Zhang (Ryerson University), and Yansong Tang (Tsinghua University)</i>	

Object Pose Estimation and Tracking

MixCycle: Mixup Assisted Semi-Supervised 3D Single Object Tracking with Cycle Consistency ..	13910
<i>Qiao Wu (Northwestern Polytechnical University), Jiaqi Yang (Northwestern Polytechnical University), Kun Sun (China University of Geosciences(Wuhan), Chu'ai Zhang (Northwestern Polytechnical University), Yanning Zhang (Northwestern Polytechnical University), and Mathieu Salzmann (EPFL)</i>	
Deep Fusion Transformer Network with Weighted Vector-Wise Keypoints Voting for Robust 6D Object Pose Estimation	13921
<i>Jun Zhou (The Hong Kong Polytechnic University), Kai Chen (The Chinese University of Hong Kong), Linlin Xu (University of Waterloo), Qi Dou (The Chinese University of Hong Kong), and Jing Qin (The Hong Kong Polytechnic University)</i>	
IST-Net: Prior-Free Category-Level Pose Estimation with Implicit Space Transformation	13932
<i>Jianhui Liu (The University of Hong Kong), Yukang Chen (The Chinese University of Hong Kong), Xiaoqing Ye (Baidu Inc.), and Xiaojuan Qi (The University of Hong Kong)</i>	
Adaptive and Background-Aware Vision Transformer for Real-Time UAV Tracking	13943
<i>Shuiwang Li (Guilin University of Technology), Yangxiang Yang (Guilin university of technology), Dan Zeng (Southern University of Science and Technology), and Xucheng Wang (Guilin University of Technology)</i>	
VI-Net: Boosting Category-level 6D Object Pose Estimation via Learning Decoupled Rotations on the Spherical Representations	13955
<i>Jiehong Lin (South China University of Technology), Zewei Wei (South China University of Technology), Yabin Zhang (Hong Kong Polytechnic University), and Kui Jia (South China University of Technology)</i>	

Tracking by Natural Language Specification with Long Short-term Context Decoupling	13966
<i>Ding Ma (Harbin Institute of Technology) and Xiangqian Wu (Harbin Institute of Technology)</i>	
CheckerPose: Progressive Dense Keypoint Localization for Object Pose Estimation with Graph Neural Network	13976
<i>Ruyi Lian (Stony Brook University) and Haibin Ling (Stony Brook University)</i>	
Deep Active Contours for Real-time 6-DoF Object Tracking	13988
<i>Long Wang (SenseTime), Shen Yan (National University of Defense Technology), Jianan Zhen (SenseTime), Yu Liu (National University of Defense Technology), Maojun Zhang (National University of Defense Technology), Guofeng Zhang (Zhejiang University), and Xiaowei Zhou (Zhejiang University)</i>	
Learning Symmetry-Aware Geometry Correspondences for 6D Object Pose Estimation	13999
<i>Heng Zhao (Hikvision Research Institute), Shenxing Wei (Xi'an JiaoTong university), Dahu Shi (Hikvision Research Institute), Wenming Tan (Hikvision Research Institute), Zheyang Li (Hikvision Research Institute), Ye Ren (Hikvision Research Institute), Xing Wei (Xi'an Jiaotong University), Yi Yang (Zhejiang University), and Shiliang Pu (Hikvision Research Institute)</i>	
Query6DoF: Learning Sparse Queries as Implicit Shape Prior for Category-Level 6DoF Pose Estimation	14009
<i>Ruiqi Wang (Huazhong University of Science and Technology), Xinggong Wang (Huazhong University of Science and Technology), Te Li (Zhejiang Lab), Rong Yang (Zhejiang Lab), Minhong Wan (Zhejiang Lab), and Wenyu Liu (Huazhong University of Science and Technology)</i>	
SOCS: Semantically-Aware Object Coordinate Space for Category-Level 6D Object Pose Estimation under Large Shape Variations	14019
<i>Boyan Wan (National University of Defense Technology), Yifei Shi (National University of Defense Technology), and Kai Xu (National University of Defense Technology)</i>	
Pseudo Flow Consistency for Self-Supervised 6D Object Pose Estimation	14029
<i>Yang Hai (Xidian University), Rui Song (Xidian University), Jiaojiao Li (xidian university), David Ferstl (Magic Leap), and Yinlin Hu (MagicLeap)</i>	
Tracking by 3D Model Estimation of Unknown Objects in Videos	14040
<i>Denys Rozumnyi (ETH Zurich / CTU in Prague), Jiri Matas (Czech Technical University, Prague), Marc Pollefeys (ETH Zurich / Microsoft), Vittorio Ferrari (Google Research), and Martin R. Oswald (ETH Zurich)</i>	
Algebraically Rigorous Quaternion Framework for the Neural Network Pose Estimation Problem.....	14051
<i>Chen Lin (Flatiron Institute), Andrew J. Hanson (Indiana University), and Sonya M. Hanson (Flatiron Institute)</i>	
Linear-Covariance Loss for End-to-End Learning of 6D Pose Estimation	14061
<i>Fulin Liu (Beihang University), Yinlin Hu (MagicLeap), and Mathieu Salzmann (EPFL)</i>	

Vanishing Point Estimation in Uncalibrated Images with Prior Gravity Direction	14072
<i>Rémi Pautrat (ETH Zurich), Shaohui Liu (ETH Zurich), Petr Hruby (ETH Zürich), Marc Pollefeys (ETH Zurich / Microsoft), and Daniel Barath (ETH Zürich)</i>	

3D Shape Modeling and Processing

2D3D-MATR: 2D-3D Matching Transformer for Detection-Free Registration Between Images and Point Clouds	14082
<i>Minhao Li (National University of Defense Technology), Zheng Qin (Defense Innovation Institute, Academy of Military Sciences), Zhirui Gao (National University of Defense Technology), Renjiao Yi (National University of Defense Technology), Chenyang Zhu (National University of Defense Technology), Yulan Guo (Sun Yat-sen University), and Kai Xu (National University of Defense Technology)</i>	
Learning Versatile 3D Shape Generation with Improved Auto-regressive Models	14093
<i>Simian Luo (Tsinghua University), Xuelin Qian (Fudan University), Yanwei Fu (Fudan University), Yinda Zhang (Google), Ying Tai (Tencent YouTu), Zhenyu Zhang (Tencent), Chengjie Wang (Tencent, Shanghai Jiao Tong University), and Xiangyang Xue (Fudan University)</i>	
CaPhy: Capturing Physical Properties for Animatable Human Avatars	14104
<i>Zhaoqi Su (Tsinghua University), Liangxiao Hu (Harbin Institute of Technology), Siyou Lin (Tsinghua University), Hongwen Zhang (Tsinghua University), Shengping Zhang (Harbin Institute of Technology), Justus Thies (Max Planck Institute for Intelligent Systems), and Yebin Liu (Tsinghua University)</i>	
Instance-aware Dynamic Prompt Tuning for Pre-trained Point Cloud Models	14115
<i>Yaohua Zha (Tsinghua University), Jinpeng Wang (Tsinghua University), Tao Dai (Shenzhen University), Bin Chen (Harbin Institute of Technology, Shenzhen), Zhi Wang (Tsinghua University), and Shu-Tao Xia (Tsinghua University)</i>	
Structure-Aware Surface Reconstruction via Primitive Assembly	14125
<i>Jingen Jiang (UCAS), Mingyang Zhao (University of Chinese Academy and Sciences&Beijing Academy of Artificial Intelligence), Shiqing Xin (Shandong University), Yanchao Yang (Stanford University), Hanxiao Wang (Casia), Xiaohong Jia (Chinese Academy of Sciences), and Dong-Ming Yan (NLPR, CASIA)</i>	
BaRe-ESA: A Riemannian Framework for Unregistered Human Body Shapes	14135
<i>Emmanuel Hartman (Florida State University), Emery Pierson (Université de Lille), Martin Bauer (Florida State University), Nicolas Charon (Johns Hopkins University), and Mohamed Daoudi (IMT Nord Europe)</i>	
Speech4Mesh: Speech-Assisted Monocular 3D Facial Reconstruction for Speech-Driven 3D Facial Animation	14146
<i>Shan He (USTC), Haonan He (Iflytek), Shuo Yang (IFLYTEK Co., Ltd.), Xiaoyan Wu (Iflytek), Pengcheng Xia (Iflytek), Bing Yin (IFLYTEK Co., Ltd), Cong Liu (iFLYTEK Research), Lirong Dai (University of Science and Technology of China), and Chang Xu (University of Sydney)</i>	

Learning Point Cloud Completion without Complete Point Clouds: A Pose-Aware Approach	14157
<i>Jihun Kim (KAIST), Hyeokjun Kweon (KAIST), Yunseo Yang (KAIST), and Kuk-Jin Yoon (KAIST)</i>	
GeoUDF: Surface Reconstruction from 3D Point Clouds via Geometry-guided Distance Representation	14168
<i>Siyu Ren (City University of HongKong), Junhui Hou (City University of Hong Kong, Hong Kong), Xiaodong Chen (Tianjin University), Ying He (Nanyang Technological University), and Wenping Wang (Texas A&M University)</i>	
SurfsUP: Learning Fluid Simulation for Novel Surfaces	14179
<i>Arjun Mani (Columbia University), Ishaan Preetam Chandratreya (Columbia University), Elliot Creager (University of Toronto), Carl Vondrick (Columbia University), and Richard Zemel (University of Toronto)</i>	
DeFormer: Integrating Transformers with Deformable Models for 3D Shape Abstraction from a Single Image	14190
<i>Di Liu (Rutgers University), Xiang Yu (Amazon), Meng Ye (Rutgers University), Qilong Zhangli (Rutgers University), Zhuowei Li (Rutgers), Zhixing Zhang (Rutgers University), and Dimitris N. Metaxas (Rutgers)</i>	
Neural Deformable Models for 3D Bi-Ventricular Heart Shape Reconstruction and Modeling from 2D Sparse Cardiac Magnetic Resonance Imaging	14201
<i>Meng Ye (Rutgers University), Dong Yang (NVIDIA Corporation), Mikael Kanski (New York University School of Medicine), Leon Axel (NYUMC), and Dimitris Metaxas (Rutgers)</i>	
DiffFacto: Controllable Part-Based 3D Point Cloud Generation with Cross Diffusion	14211
<i>Kiyohiro Nakayama (Stanford University), Mikaela Angelina Uy (Stanford University), Jiahui Huang (Tsinghua University), Shi-Min Hu (Tsinghua University), Ke Li (Simon Fraser University), and Leonidas Guibas (Stanford University)</i>	
Self-supervised Learning of Implicit Shape Representation with Dense Correspondence for Deformable Objects	14222
<i>Baowen Zhang (Institute of Software Chinese Academy of Sciences), Jiahe Li (Institute of Software, Chinese Academy of Science), Xiaoming Deng (Institute of Software, Chinese Academy of Sciences), Yinda Zhang (Google), Cuixia Ma (Institute of Software Chinese Academy of Sciences, University of Chinese Academy of Sciences), and Hongan Wang (Institute of Software, Chinese Academy of Sciences)</i>	
Neural Implicit Surface Evolution	14233
<i>Tiago Novello (IMPA), Vinicius da Silva (Pontifical Catholic University of Rio de Janeiro, (PUC-Rio), Guilherme Schardong (University of Coimbra), Luiz Schirmer (Unisinos), Helio Lopes (Pontifícia Universidade Católica do Rio de Janeiro), and Luiz Velho (IMPA)</i>	

PointDC: Unsupervised Semantic Segmentation of 3D Point Clouds via Cross-Modal Distillation and Super-Voxel Clustering	14244
<i>Zisheng Chen (South China University of Technology), Hongbin Xu (South China University of Technology), Weitao Chen (personal), Zhipeng Zhou (Chinese Academy of Sciences), Haihong Xiao (South China University of Technology), Baigui Sun (Alibaba Group), Xuansong Xie (Alibaba Group), and Wenxiong kang (South China University of Technology)</i>	
HyperDiffusion: Generating Implicit Neural Fields with Weight-Space Diffusion	14254
<i>Ziya Erkoç (Technical University of Munich), Fangchang Ma (Apple Inc.), Qi Shan (Apple Inc.), Matthias Nießner (Technical University of Munich), and Angela Dai (Technical University of Munich)</i>	
Leveraging SE(3) Equivariance for Learning 3D Geometric Shape Assembly	14265
<i>Ruihai Wu (Peking University), Chenrui Tie (Peking University), Yushi Du (Peking University), Yan Zhao (Peking University), and Hao Dong (Peking University)</i>	
DPF-Net: Combining Explicit Shape Priors in Deformable Primitive Field for Unsupervised Structural Reconstruction of 3D Objects	14275
<i>Qingyao Shuai (University of Science and Technology of China), Chi Zhang (University of Science and Technology of China), Kaizhi Yang (University of Science and Technology of China), and Xuejin Chen (University of Science and Technology of China)</i>	
Sample-adaptive Augmentation for Point Cloud Recognition Against Real-world Corruptions ..	14284
<i>Jie Wang (Beijing Institute of Technology), Lihe Ding (Beijing Institute of Technology), Tingfa Xu (Beijing Institute of Technology), Shaocong Dong (Beijing Institute of Technology), Xinli Xu (Beijing Institute of Technology), Long Bai (The Chinese University of Hong Kong), and Jianan Li (Beijing Institute of Technology)</i>	
3DHacker: Spectrum-based Decision Boundary Generation for Hard-label 3D Point Cloud Attack.....	14294
<i>Yunbo Tao (School of Cyberspace Security Huazhong University of Science and Technology), Daizong Liu (Peking University), Pan Zhou (Huazhong University of Science and Technology), Yulai Xie (huazhong university of science and technology), Wei Du (hust.edu.cn), and Wei Hu (Peking University)</i>	
P2C: Self-Supervised Point Cloud Completion from Single Partial Clouds	14305
<i>Ruikai Cui (Australian National University), Shi Qiu (ANU), Saeed Anwar (The Australian National University), Jiawei Liu (Australian National University), Chaoyue Xing (Australian National University), Jing Zhang (Australian National University), and Nick Barnes (ANU)</i>	
Towards Multi-Layered 3D Garments Animation	14315
<i>Yidi Shao (Nanyang Technological University), Chen Change Loy (Nanyang Technological University), and Bo Dai (Shanghai AI Lab)</i>	
AvatarCraft: Transforming Text into Neural Human Avatars with Parameterized Shape and Pose Control	14325
<i>Ruixiang Jiang (The Hong Kong Polytechnic University), Can Wang (City University of Hong Kong), Jingbo Zhang (City University of Hong Kong), Menglei Chai (Google), Mingming He (Netflix), Dongdong Chen (Microsoft Cloud AI), and Jing Liao (City University of Hong Kong)</i>	

Blending-NeRF: Text-Driven Localized Editing in Neural Radiance Fields	14337
<i>Hyeonseop Song (LG Electronics), Seokhun Choi (LG Electronics), Hoseok Do (LG Electronics), Chul Lee (LG Electronics), and Taehyeong Kim (Seoul National University)</i>	
SIRA-PCR: Sim-to-Real Adaptation for 3D Point Cloud Registration	14348
<i>Suyi Chen (UESTC, Megvii), Hao Xu (CUHK), Ru Li (Harbin Institute of Technology), Guanghui Liu (UESTC), Chi-Wing Fu (The Chinese University of Hong Kong), and Shuaicheng Liu (UESTC, Megvii)</i>	
3D Semantic Subspace Traverser: Empowering 3D Generative Model with Shape Editing Capability	14360
<i>Ruowei Wang (Sichuan university), Yu Liu (Sichuan University), Pei Su (Sichuan University), Jianwei Zhang (Sichuan University), and Qijun Zhao (Sichuan University)</i>	
DMNet: Delaunay Meshing Network for 3D Shape Representation	14372
<i>Chen Zhang (Huazhong University of Science and Technology), Ganzhangqin Yuan (Huazhong University of Science and Technology), and Wenbing Tao (Huazhong University of Science and Technology)</i>	
Attention Discriminant Sampling for Point Clouds	14383
<i>Cheng-Yao Hong (Academia Sinica), Yu-Ying Chou (National Taiwan University), and Tyng-Luh Liu (Academia Sinica)</i>	
SALAD: Part-Level Latent Diffusion for 3D Shape Generation and Manipulation	14395
<i>Juil Koo (KAIST), Seungwoo Yoo (KAIST), Minh Hieu Nguyen (KAIST), and Minhyuk Sung (KAIST)</i>	
MAPConNet: Self-supervised 3D Pose Transfer with Mesh and Point Contrastive Learning	14406
<i>Jiaze Sun (Imperial College London), Zhixiang Chen (University of Sheffield), and Tae-Kyun Kim (KAIST/Imperial College London)</i>	
Invariant Training 2D-3D Joint Hard Samples for Few-Shot Point Cloud Recognition	14417
<i>Xuanyu Yi (Nanyang Technological University), Jiajun Deng (The University of Sydney), Qianru Sun (Singapore Management University), Xian-Sheng Hua (Terminus Group), Joo-Hwee Lim (Institute for Infocomm Research), and Hanwang Zhang (Nanyang Technological University)</i>	
EPiC: Ensemble of Partial Point Clouds for Robust Classification	14429
<i>Meir Yossef Levi (Technion) and Guy Gilboa (Technion)</i>	
Leveraging Intrinsic Properties for Non-Rigid Garment Alignment	14439
<i>Siyu Lin (Tsinghua University), Boyao Zhou (Tsinghua University), Zerong Zheng (Tsinghua University), Hongwen Zhang (Tsinghua University), and Yebin Liu (Tsinghua University)</i>	
Spatially and Spectrally Consistent Deep Functional Maps	14451
<i>Mingze Sun (tbsi), Shiwei Mao (Tsinghua University), Puhua Jiang (Tsinghua University), Maks Ovsjanikov (Ecole polytechnique), and Ruqi Huang (Tsinghua Shenzhen International Graduate School, (SIGS, and Tsinghua-Berkeley Shenzhen Institute, (TBSI, Tsinghua University)</i>	

SVDFormer: Complementing Point Cloud via Self-view Augmentation and Self-structure Dual-generator	14462
<i>Zhe Zhu (Nanjing University of Aeronautics and Astronautics), Honghua Chen (Nanyang Technological University), Xing He (Nanjing University of Aeronautics and Astronautics), Weiming Wang (Hong Kong Metropolitan University), Jing Qin (The Hong Kong Polytechnic University), and Mingqiang Wei (Nanjing University of Aeronautics and Astronautics)</i>	
Batch-based Model Registration for Fast 3D Sherd Reconstruction	14473
<i>Jiepeng Wang (The University of Hong Kong), Congyi Zhang (The University of Hong Kong), Peng Wang (The University of Hong Kong), Xin Li (Texas A&M University), Peter J. Cobb (HKU), Christian Theobalt (MPI Informatik), and Wenping Wang (Texas A&M University)</i>	
Implicit Autoencoder for Point-Cloud Self-Supervised Representation Learning	14484
<i>Siming Yan (The University of Texas at Austin), Zhenpei Yang (The University of Texas at Austin), Haoxiang Li (Wormpex AI Research), Chen Song (The University of Texas at Austin), Li Guan (Meta), Hao Kang (Wormpex AI Research), Gang Hua (Wormpex AI Research), and Qixing Huang (The University of Texas at Austin)</i>	
E3Sym: Leveraging E(3) Invariance for Unsupervised 3D Planar Reflective Symmetry Detection.	14497
<i>Ren-Wu Li (Institute of Computing Technology, Chinese Academy of Sciences), Ling-Xiao Zhang (Institute of Computing Technology, Chinese Academy of Sciences), Chunpeng Li (Institute of Computing Technology, Chinese Academy of Sciences), Yu-Kun Lai (Cardiff University), and Lin Gao (Institute of Computing Technology, Chinese Academy of Sciences)</i>	
Semantify: Simplifying the Control of 3D Morphable Models Using CLIP	14508
<i>Omer Galnik (Reichman University), Guy Gafni (Technical University of Munich), and Ariel Shamir (The Interdisciplinary Center)</i>	
VoroMesh: Learning Watertight Surface Meshes with Voronoi Diagrams	14519
<i>Nissim Maruani (Inria), Roman Klokov (INRIA), Maks Ovsjanikov (Ecole polytechnique), Pierre Alliez (INRIA Sophie-Antipolis, France), and Mathieu Desbrun (Inria/Ecole Polytechnique)</i>	
DG3D: Generating High Quality 3D Textured Shapes by Learning to Discriminate Multi-Modal Diffusion-Renderings	14529
<i>Qi Zuo (Alibaba DAMO Academy), Yafei Song (Alibaba Group), Jianfang Li (alibaba), Lin Liu (Alibaba Group), and Liefeng Bo (Alibaba DAMO Academy)</i>	
Unaligned 2D to 3D Translation with Conditional Vector-Quantized Code Diffusion using Transformers	14539
<i>Abril Corona-Figueroa (Durham University), Sam Bond-Taylor (Durham University), Neelanjan Bhowmik (Durham University), Yona A. Gaus (Durham University), Toby P. Breckon (Durham University), Hubert P. H. Shum (Durham University), and Chris G. Willcocks (Durham University)</i>	
Hyperbolic Chamfer Distance for Point Cloud Completion	14549
<i>Fangzhou Lin (Tohoku University), Yun Yue (Worcester Polytechnic Institute), Songlin Hou (Worcester Polytechnic Institute), Xuechu Yu (Worcester Polytechnic Institute), Yajun Xu (Hokkaido University), Kazunori D Yamada (Tohoku University), and Ziming Zhang (Worcester Polytechnic Institute)</i>	

SKED: Sketch-guided Text-based 3D Editing 14561
Aryan Mikaeili (Simon Fraser University), Or Perel (NVIDIA), Mehdi Safaee (Simon Fraser University), Danny Cohen-Or (Tel Aviv University), and Ali Mahdavi-Amiri (Simon Fraser University)

Adaptive Spiral Layers for Efficient 3D Representation Learning on Meshes 14574
Francesca Babiloni (Huawei Noah's Ark Lab), Matteo Maggioni (Huawei Noah's Ark Lab), Thomas Tanay (Huawei Noah's Ark Lab), Jiankang Deng (Imperial College London), Ales Leonardis (Huawei Noah's Ark Lab), and Stefanos Zafeiriou (Imperial College London)

Human Pose/Shape Estimation

EMDB: The Electromagnetic Database of Global 3D Human Pose and Shape in the Wild 14586
Manuel Kaufmann (ETH Zurich), Jie Song (ETH Zurich), Chen Guo (ETH Zurich), Kaiyue Shen (ETH Zurich), Tianjian Jiang (ETH Zurich), Chengcheng Tang (Facebook Reality Labs), Juan José Zárate (ETH Zürich), and Otmar Hilliges (ETH Zurich)

ReFit: Recurrent Fitting Network for 3D Human Recovery 14598
Yufu Wang (University of Pennsylvania) and Kostas Daniilidis (University of Pennsylvania)

Global Adaptation Meets Local Generalization: Unsupervised Domain Adaptation for 3D Human Pose Estimation 14609
Wenhao Chai (University of Washington), Zhongyu Jiang (University of Washington), Jenq-Neng Hwang (University of Washington), and Gaoang Wang (Zhejiang University)

Spectral Graphormer: Spectral Graph-Based Transformer for Egocentric Two-Hand Reconstruction using Multi-View Color Images 14620
Tze Ho Elden Elden Tse (University of Birmingham), Franziska Mueller (Google Inc.), Zhengyang Shen (Google), Danhang Tang (Google Inc.), Thabo Beeler (Google), Mingsong Dou (Google), Yinda Zhang (Google), Sasa Petrovic (Google), Hyung Jin Chang (University of Birmingham), Jonathan Taylor (Google Inc.), and Bardia Doosti (Google)

Realistic Full-Body Tracking from Sparse Observations via Joint-Level Modeling 14632
Xiaozheng Zheng (ByteDance), Zhuo Su (Bytedance), Chao Wen (Bytedance), Zhou Xue (Bytedance), and Xiaojie Jin (Bytedance Inc. USA)

Rethinking Pose Estimation in Crowds: Overcoming the Detection Information Bottleneck and Ambiguity 14643
Mu Zhou (EPFL), Lucas Stoffl (EPFL), Mackenzie Weygandt Mathis (EPFL), and Alexander Mathis (EPFL)

HDG-ODE: A Hierarchical Continuous-Time Model for Human Pose Forecasting 14654
Yucheng Xing (Stony Brook University) and Xin Wang (Stony Brook University)

AffordPose: A Large-Scale Dataset of Hand-Object Interactions with Affordance-Driven Hand Pose 14667
Juntao Jian (Dalian University of Technology), Xiuping Liu (Dalian University of Technology), Manyi Li (Shandong University), Ruizhen Hu (Shenzhen University), and Jian Liu (Tsinghua university)

PhaseMP: Robust 3D Pose Estimation via Phase-conditioned Human Motion Prior	14679
<i>Mingyi Shi (Hong Kong University), Sebastian Starke (Meta), Yuting Ye (Meta Reality Labs Research), Taku Komura (The University of Hong Kong), and Jungdam Won (Meta AI)</i>	
Synthesizing Diverse Human Motions in 3D Indoor Scenes	14692
<i>Kaifeng Zhao (ETH Zurich), Yan Zhang (ETH Zurich), Shaofei Wang (ETH Zurich), Thabo Beeler (Google), and Siyu Tang (ETH Zurich)</i>	
TEMPO: Efficient Multi-View Pose Estimation, Tracking, and Forecasting	14704
<i>Rohan Choudhury (Carnegie Mellon University), Kris M. Kitani (Carnegie Mellon University), and László A. Jeni (Carnegie Mellon University)</i>	
Diffusion-Based 3D Human Pose Estimation with Multi-Hypothesis Aggregation	14715
<i>Wenkang Shan (Peking University), Zhenhua Liu (Huawei Noah's Ark Lab), Xinfeng Zhang (University of Chinese Academy of Sciences), Zhao Wang (Peking University), Kai Han (Huawei Noah's Ark Lab), Shanshe Wang (Peking University), Siwei Ma (Peking University, China), and Wen Gao (PKU)</i>	
Towards Robust and Smooth 3D Multi-Person Pose Estimation from Monocular Videos in the Wild	14726
<i>Sungchan Park (Seoul National University), Eunyi You (Seoul national university), Inhoe Lee (Seoul National University), and Joonseok Lee (Google Research & Seoul National University)</i>	
Humans in 4D: Reconstructing and Tracking Humans with Transformers	14737
<i>Shubham Goel (University of California at Berkeley), Georgios Pavlakos (UC Berkeley), Jathushan Rajasegaran (UC Berkeley), Angjoo Kanazawa (University of California Berkeley), and Jitendra Malik (University of California at Berkeley)</i>	
NPC: Neural Point Characters from Video	14749
<i>Shih-Yang Su (University of British Columbia), Timur Bagautdinov (Facebook), and Helge Rhodin (UBC)</i>	
Priority-Centric Human Motion Generation in Discrete Latent Space	14760
<i>Hanyang Kong (National University of Singapore), Kehong Gong (national university of singapore), Dongze Lian (National University of Singapore), Michael Bi Mi (Huawei International Pte Ltd), and Xinchao Wang (National University of Singapore)</i>	
NCHO: Unsupervised Learning for Neural 3D Composition of Humans and Objects	14771
<i>Taeksoo Kim (Seoul National University), Shunsuke Saito (Facebook), and Hanbyul Joo (Seoul National University)</i>	
Cyclic Test-Time Adaptation on Monocular Video for 3D Human Mesh Reconstruction	14783
<i>Hyeongjin Nam (Seoul National University), Daniel Sungho Jung (Seoul National University), Yeonguk Oh (Seoul National University), and Kyoung Mu Lee (Seoul National University)</i>	
MHEntropy: Entropy Meets Multiple Hypotheses for Pose and Shape Recovery	14794
<i>Rongyu Chen (National University of Singapore), Linlin Yang (National University of Singapore), and Angela Yao (National University of Singapore)</i>	

Probabilistic Triangulation for Uncalibrated Multi-View 3D Human Pose Estimation	14804
<i>Boyuan Jiang (Institute of Computing Technology, Chinese Academy of Sciences), Lei Hu (Institute of Computing Technology, Chinese Academy of Sciences), and Shihong Xia (institute of computing technology of the Chinese academy of sciences)</i>	
DiffPose: SpatioTemporal Diffusion Model for Video-Based Human Pose Estimation	14815
<i>Runyang Feng (Jilin University), Yixing Gao (Jilin University), Tze Ho Elden Elden Tse (University of Birmingham), Xueqing Ma (Jilin University), and Hyung Jin Chang (University of Birmingham)</i>	
Reconstructing Groups of People with Hypergraph Relational Reasoning	14827
<i>Buzhen Huang (Southeast University), Jingyi Ju (Southeast University), Zhihao Li (Huawei Noah's Ark Lab), and Yangang Wang (Southeast University)</i>	
MixSynthFormer: A Transformer Encoder-like Structure with Mixed Synthetic Self-attention for Efficient Human Pose Estimation	14838
<i>Yuran Sun (The University of Hong Kong), Alan William Dougherty (The University of Hong Kong), Zhuoying Zhang (The University of Hong Kong), Yi King Choi (The University of Hong Kong), and Chuan Wu (The University of Hong Kong)</i>	
Dynamic Hyperbolic Attention Network for Fine Hand-object Reconstruction	14848
<i>Zhiying Leng (State Key Laboratory of Virtual Reality Technology and Systems, Beihang University), Shun-Cheng Wu (Technical University of Munich), Mahdi Saleh (Technical University of Munich), Antonio Montanaro (Politecnico di Torino), Hao Yu (Technical University of Munich), Yin Wang (State Key Laboratory of Virtual Reality Technology and Systems, Beihang University), Nassir Navab (TU Munich, Germany), Xiaohui Liang (State Key Laboratory of Virtual Reality Technology and Systems, Beihang University; Zhongguancun Laboratory, Beijing), and Federico Tombari (Google, TU Munich)</i>	
Human from Blur: Human Pose Tracking from Blurry Images	14859
<i>Yiming Zhao (ETH Zürich), Denys Rozumnyi (ETH Zurich / CTU in Prague), Jie Song (ETH Zurich), Otmar Hilliges (ETH Zurich), Marc Pollefeys (ETH Zurich / Microsoft), and Martin R. Oswald (ETH Zurich)</i>	
AG3D: Learning to Generate 3D Avatars from 2D Image Collections	14870
<i>Zijian Dong (ETH Zurich), Xu Chen (ETH Zürich), Jinlong Yang (Max Planck Institute for Intelligent Systems), Michael J. Black (Max Planck Institute for Intelligent Systems), Otmar Hilliges (ETH Zurich), and Andreas Geiger (University of Tuebingen)</i>	
InterDiff: Generating 3D Human-Object Interactions with Physics-Informed Diffusion	14882
<i>Sirui Xu (University of Illinois Urbana-Champaign), Zhengyuan Li (University of Illinois Urbana-Champaign), Yu-Xiong Wang (University of Illinois at Urbana-Champaign), and Liang-Yan Gui (University of Illinois Urbana-Champaign)</i>	
SEFD: Learning to Distill Complex Pose and Occlusion	14895
<i>ChangHee Yang (Sogang University), Kyeongbo Kong (Pusan National University), SungJun Min (Sogang University), Dongyoon Wee (Clova AI Video, NAVER Corp.), Ho-Deok Jang (NAVER Corp.), Geonho Cha (Clova AI Video, NAVER Corp.), and SukJu Kang (Sogang University)</i>	

3D Human Mesh Recovery with Sequentially Global Rotation Estimation	14907
<i>Dongkai Wang (Peking University) and Shiliang Zhang (Peking University)</i>	
Co-Evolution of Pose and Mesh for 3D Human Body Estimation from Video	14917
<i>Yingxuan You (Peking University), Hong Liu (Peking University), Ti Wang (Peking University), Wenhao Li (Peking University), Runwei Ding (Peking University), and Xia Li (ETH Zurich)</i>	
PHRIT: Parametric Hand Representation with Implicit Template	14928
<i>Zhisheng Huang (Wuhan University), Yujin Chen (Technical University of Munich), Di Kang (Tencent), Jinlu Zhang (Wuhan University), and Zhigang Tu (Wuhan University)</i>	
HopFIR: Hop-wise GraphFormer with Intragroup Joint Refinement for 3D Human Pose Estimation.....	14939
<i>Kai Zhai (Hefei University of Technology), Qiang Nie (Tencent Youtu Lab), Bo Ouyang (Hefei University of Technology), Xiang Li (Tsinghua University), and Shanlin Yang (Hefei University of Technology)</i>	
Prior-guided Source-free Domain Adaptation for Human Pose Estimation	14950
<i>Dripta S. Raychaudhuri (AWS AI Labs), Calvin-Khang Ta (University of California, Riverside), Arindam Dutta (University of California Riverside), Rohit Lal (University of California, Riverside), and Amit K. Roy-Chowdhury (University of California, Riverside)</i>	
Cloth2Body: Generating 3D Human Body Mesh from 2D Clothing	14961
<i>Lu Dai (The Hong Kong University of Science and Technology), Liqian Ma (ZMO AI), Shenhan Qian (Technical University of Munich), Hao Liu (HKUST), Ziveli Liu (Nanyang Technological University), and Hui Xiong (Hong Kong University of Science and Tech)</i>	
PoseFix: Correcting 3D Human Poses with Natural Language	14972
<i>Ginger Delmas (NAVER LABS EUROPE), Philippe Weinzaepfel (NAVER LABS Europe), Francesc Moreno-Noguer (IRI), and Grégory Rogez (NAVER LABS Europe)</i>	
Group Pose: A Simple Baseline for End-to-End Multi-Person Pose Estimation	14983
<i>Huan Liu (Beijing Jiaotong University), Qiang Chen (Baidu Inc), Zichang Tan (Baidu Research), Jiang-Jiang Liu (Baidu Inc.), Jian Wang (Baidu), Xiangbo Su (Baidu Inc.), Xiaolong Li (Beijing Jiaotong University, Institute of Information Science), Kun Yao (Baidu Inc.), Junyu Han (Baidu Inc.), Errui Ding (Baidu Inc.), Yao Zhao (Beijing Jiaotong University), and Jingdong Wang (Baidu)</i>	
Make-An-Animation: Large-Scale Text-conditional 3D Human Motion Generation	14993
<i>Samaneh Azadi (Meta AI), Akbar Shah (Meta Platforms Inc), Thomas Hayes (Meta), Devi Parikh (Georgia Tech & Facebook AI Research), and Sonal Gupta (Facebook)</i>	

NSF: Neural Surface Fields for Human Modeling from Monocular Depth	15003
<i>Yuxuan Xue (Tübingen AI Center; University of Tübingen), Bharat Lal Bhatnagar (Tübingen AI Center; University of Tübingen; Max Planck Institute for Informatics; Meta Reality Labs Research), Riccardo Marin (Tübingen AI Center; University of Tübingen), Nikolaos Sarafianos (Meta Reality Labs Research), Yuanlu Xu (Meta Reality Labs Research), Gerard Pons-Moll (Tübingen AI Center; University of Tübingen; Max Planck Institute for Informatics), and Tony Tung (Meta Reality Labs Research)</i>	
Hierarchical Generation of Human-Object Interactions with Diffusion Probabilistic Models	15015
<i>Huaijin Pi (Zhejiang University), Sida Peng (Zhejiang University), Minghui Yang (Ant Group), Xiaowei Zhou (Zhejiang University), and Hujun Bao (Zhejiang University)</i>	
Dynamic Mesh Recovery from Partial Point Cloud Sequence	15028
<i>Hojun Jang (Seoul National University), Minkwan Kim (Seoul National University), Jinseok Bae (Seoul National University), and Young Min Kim (Seoul National University)</i>	
MotionBERT: A Unified Perspective on Learning Human Motion Representations	15039
<i>Wentao Zhu (Peking University), Xiaoxuan Ma (Peking University), Zhaoyang Liu (SenseTime Research), Libin Liu (Peking University), Wayne Wu (Shanghai AI Laboratory), and Yizhou Wang (PKU)</i>	
Novel-View Synthesis and Pose Estimation for Hand-Object Interaction from Sparse Views	15054
<i>Wentian Qu (Institute of Software Chinese Academy of Sciences; University of Chinese Academy of Sciences), Zhaopeng Cui (Zhejiang University), Yinda Zhang (Google), Chenyu Meng (Institute of Software Chinese Academy of Sciences), Cuixia Ma (Institute of Software Chinese Academy of Sciences; University of Chinese Academy of Sciences), Xiaoming Deng (Institute of Software, Chinese Academy of Sciences), and Hongan Wang (Institute of Software, Chinese Academy of Sciences)</i>	
OCHID-Fi: Occlusion-Robust Hand Pose Estimation in 3D via RF-Vision	15066
<i>Shujie Zhang (Nanyang Technological University), Tianyue Zheng (Nanyang Technological University), Zhe Chen (Fudan University), Jingzhi Hu (Nanyang Technological University), Abdelwahed Khamis (CSIRO), Jiajun Liu (CSIRO), and Jun Luo (Nanyang Technological University)</i>	
Neural Interactive Keypoint Detection	15076
<i>Jie Yang (The Chinese University of HongKong, Shenzhen), Ailing Zeng (International Digital Economy Academy), Feng Li (Hong Kong University of Science and Technology), Shilong Liu (Tsinghua University), Ruimao Zhang (The Chinese University of Hong Kong, Shenzhen), and Lei Zhang (International Digital Economy Academy, (IDEA)</i>	
Plausible Uncertainties for Human Pose Regression	15087
<i>Lennart Bramlage (Reutlingen University), Michelle Karg (ADC Automotive Distance Control Systems GmbH, Continental), and Cristóbal Curio (Reutlingen University)</i>	

TORE: Token Reduction for Efficient Human Mesh Recovery with Transformer	15097
<i>Zhiyang Dou (The University of Hong Kong), Qingxuan Wu (University of Oxford), Cheng Lin (Tencent Games), Zeyu Cao (University of Cambridge), Qiangqiang Wu (City University of Hong Kong), Weilin Wan (The University of Hong Kong), Taku Komura (The University of Hong Kong), and Wenping Wang (Texas A&M University)</i>	
Weakly-supervised 3D Pose Transfer with Keypoints	15110
<i>Jinnan Chen (National University of Singapore), Chen Li (National University of Singapore), and Gim Hee Lee (National University of Singapore)</i>	

Vision and Language 2

SATR: Zero-Shot Semantic Segmentation of 3D Shapes	15120
<i>Ahmed Abdelreheem (KAUST), Ivan Skorokhodov (KAUST), Maks Ovsjanikov (LIX, Ecole Polytechnique), and Peter Wonka (KAUST)</i>	
CiT: Curation in Training for Effective Vision-Language Data	15134
<i>Hu Xu (Meta AI, FAIR), Saining Xie (Meta AI, FAIR), Po-Yao Huang (Meta AI, FAIR), Licheng Yu (Meta AI, FAIR), Russell Howes (Meta AI, FAIR), Gargi Ghosh (Meta AI, FAIR), Luke Zettlemoyer (Meta AI, FAIR), and Christoph Feichtenhofer (Meta AI, FAIR)</i>	
Self-regulating Prompts: Foundational Model Adaptation without Forgetting	15144
<i>Muhammad Uzair Khattak (Mohamed Bin Zayed University of Artificial Intelligence, MBZUAI), Syed Talal Wasim (Mohamed Bin Zayed University of Artificial Intelligence), Muzammal Naseer (MBZUAI), Salman Khan (MBZUAI), Ming-Hsuan Yang (University of California at Merced), and Fahad Shahbaz Khan (MBZUAI)</i>	
Learning to Ground Instructional Articles in Videos through Narrations	15155
<i>Effrosyni Mavroudi (Meta AI), Triantafyllos Afouras (Meta AI), and Lorenzo Torresani (Meta AI)</i>	
RefEgo: Referring Expression Comprehension Dataset from First-Person Perception of Ego4D ...	15168
<i>Shuhei Kurita (RIKEN), Naoki Katsura (University of Tsukuba), and Eri Onami (RIKEN)</i>	
Multi3DRefer: Grounding Text Description to Multiple 3D Objects	15179
<i>Yiming Zhang (Simon Fraser University), ZeMing Gong (SFU), and Angel X. Chang (SFU)</i>	
Bayesian Prompt Learning for Image-Language Model Generalization	15191
<i>Mohammad Mahdi Derakhshani (University of Amsterdam), Enrique Sanchez (Samsung AI Centre, Cambridge), Adrian Bulat (Samsung AI Center, Cambridge), Victor Guilherme Turrisi da Costa (University of Trento), Cees G. M. Snoek (University of Amsterdam), Georgios Tzimiropoulos (Queen Mary University of London), and Brais Martinez (Samsung AI Center)</i>	
Who Are You Referring To? Coreference Resolution In Image Narrations	15201
<i>Arushi Goel (University of Edinburgh), Basura Fernando (Agency for Science, Technology and Research, A*STAR, Singapore), Frank Keller (University of Edinburgh), and Hakan Bilen (University of Edinburgh)</i>	

Guiding Image Captioning Models Toward More Specific Captions	15213
<i>Simon Kornblith (Google Brain), Lala Li (Google), Zirui Wang (Apple AI/ML), and Thao Nguyen (University of Washington)</i>	
PreSTU: Pre-Training for Scene-Text Understanding	15224
<i>Jihyung Kil (The Ohio State University), Soravit Changpinyo (Google AI), Xi Chen (Google Research), Hexiang Hu (Google), Sebastian Goodman (Google), Wei-Lun Chao (The Ohio State University), and Radu Soricut (Google)</i>	
Exploring Group Video Captioning with Efficient Relational Approximation	15235
<i>Wang Lin (Zhejiang University), Tao Jin (Zhejiang University), Ye Wang (Zhejiang University), Wenwen Pan (Hangzhou Dianzi University), Linjun Li (Zhejiang University), Xize Cheng (Zhejiang University), and Zhou Zhao (Zhejiang University)</i>	
VLSlice: Interactive Vision-and-Language Slice Discovery	15245
<i>Eric Slyman (Oregon State University), Minsuk Kahng (Google), and Stefan Lee (Oregon State University)</i>	
Pretrained Language Models as Visual Planners for Human Assistance	15256
<i>Dhruvesh Patel (University of Massachusetts Amherst), Hamid Eghbalzadeh (Meta), Nitin Kamra (Reality Labs Research, Meta), Michael Louis Iuzzolino (Meta), Unnat Jain (Facebook AI Research), and Ruta Desai (Meta Reality Labs Research)</i>	
VQA Therapy: Exploring Answer Differences by Visually Grounding Answers	15269
<i>Chongyan Chen (University of Texas at Austin), Samreen Anjum (University of Colorado Boulder), and Danna Gurari (University of Colorado Boulder)</i>	
Towards High-Fidelity Text-Guided 3D Face Generation and Manipulation Using only Images .	15280
<i>Cuican Yu (Xi'an Jiaotong University), Guansong Lu (Huawei Noah's Ark Lab), Yihan Zeng (Shanghai Jiao Tong University), Jian Sun (Xi'an Jiaotong University), Xiaodan Liang (Sun Yat-sen University), Huibin Li (Xian JiaoTong University, China), Zongben Xu (XJTU), Songcen Xu (Huawei Noah's Ark Lab), Wei Zhang (Noah's Ark Lab, Huawei Technologies), and Hang Xu (Huawei Noah's Ark Lab)</i>	
Zero-Shot Composed Image Retrieval with Textual Inversion	15292
<i>Alberto Baldrati (Università degli Studi di Firenze), Lorenzo Agnolucci (University of Florence), Marco Bertini (University of Florence), and Alberto Del Bimbo (University of Florence)</i>	
PatchCT: Aligning Patch Set and Label Set with Conditional Transport for Multi-Label Image Classification	15302
<i>Miaoge Li (Xidian University), Dongsheng Wang (Xidian University), Xinyang Liu (Xidian University), Zequn Zeng (Xidian University), Ruiying Lu (Xidian University), Bo Chen (Xidian University), and Mingyuan Zhou (University of Texas at Austin)</i>	
Lip Reading for Low-resource Languages by Learning and Combining General Speech Knowledge and Language-specific Knowledge	15313
<i>Minsu Kim (KAIST), Jeong Hun Yeo (Korea Advanced Institute of Science and Technology), Jeongsoo Choi (KAIST), and Yong Man Ro (KAIST)</i>	

ViewRefer: Grasp the Multi-view Knowledge for 3D Visual Grounding	15326
<i>Zoey Guo (Peking University), Yiwen Tang (Shanghai AI Laboratory), Ray Zhang (Shanghai AI Lab), Dong Wang (Shanghai AI Laboratory), Zhigang Wang (shanghai AI lab), Bin Zhao (Northwestern Polytechnical University), and Xuelong Li (Northwestern Polytechnical University)</i>	
AerialVLN: Vision-and-Language Navigation for UAVs	15338
<i>Shubo Liu (Northwestern Polytechnical University), Hongsheng Zhang (Northwestern Polytechnical University), Yuankai Qi (The University of Adelaide), Peng Wang (Northwestern Polytechnical University), Yanning Zhang (Northwestern Polytechnical University), and Qi Wu (University of Adelaide)</i>	
Linear Spaces of Meanings: Compositional Structures in Vision-Language Models	15349
<i>Matthew Trager (AWS AI Labs), Pramuditha Perera (AWS AI Labs), Luca Zancato (AWS AI Labs), Alessandro Achille (AWS), Parminder Bhatia (Amazon), and Stefano Soatto (AWS Amazon ML)</i>	
HiTeA: Hierarchical Temporal-Aware Video-Language Pre-training	15359
<i>Qinghao Ye (Alibaba Group), Guohai Xu (Alibaba Group), Ming Yan (Alibaba Group), Haiyang Xu (Alibaba DAMO Academy), Qi Qian (Alibaba Group), Ji Zhang (Alibaba Inc.), and Fei Huang (Alibaba)</i>	
EgoTV: Egocentric Task Verification from Natural Language Task Descriptions	15371
<i>Rishi Hazra (Örebro University), Brian Chen (Columbia University), Akshara Rai (Facebook), Nitin Kamra (Reality Labs Research, Meta), and Ruta Desai (Meta Reality Labs Research)</i>	
SINC: Self-Supervised In-Context Learning for Vision-Language Tasks	15384
<i>Yi-Syuan Chen (National Yang Ming Chiao Tung University), Yun-Zhu Song (National Yang Ming Chiao Tung University), Cheng Yu Yeo (National Yang Ming Chiao Tung University), Bei Liu (Microsoft Research), Jianlong Fu (Microsoft Research), and Hong-Han Shuai (National Yang Ming Chiao Tung University)</i>	
VLN-PETL: Parameter-Efficient Transfer Learning for Vision-and-Language Navigation	15397
<i>Yanyuan Qiao (University of Adelaide), Zheng Yu (University of Adelaide), and Qi Wu (University of Adelaide)</i>	
Going Denser with Open-Vocabulary Part Segmentation	15407
<i>Peize Sun (The University of Hong Kong), Shoufa Chen (The University of Hong Kong), Chenchen Zhu (Meta), Fanyi Xiao (Meta), Ping Luo (The University of Hong Kong), Saining Xie (New York University), and Zhicheng Yan (Meta Reality Labs)</i>	
Temporal Collection and Distribution for Referring Video Object Segmentation	15420
<i>Jiajin Tang (ShanghaiTech University), Ge Zheng (ShanghaiTech University), and Sibe Yang (ShanghaiTech University)</i>	
Inverse Compositional Learning for Weakly-supervised Relation Grounding	15431
<i>Huan Li (Xi'an Jiaotong University), Ping Wei (Xi'an Jiaotong University), Zeyu Ma (Xi'an Jiaotong University), and Nanning Zheng (Xi'an Jiaotong University)</i>	

Why Is Prompt Tuning for Vision-Language Models Robust to Noisy Labels?	15442
<i>Cheng-En Wu (University of Wisconsin-Madison), Yu Tian (Rutgers), Haichao Yu (UIUC), Heng Wang (TikTok), Pedro Morgado (University of Wisconsin-Madison), Yu Hen Hu (University of Wisconsin-Madison), and Linjie Yang (ByteDance AI Lab)</i>	
CHAMPAGNE: Learning Real-world Conversation from Large-Scale Web Videos	15452
<i>Seungju Han (Seoul National University), Jack Hessel (Allen Institute for AI), Nouha Dziri (Allen Institute for AI), Yejin Choi (University of Washington), and Youngjae Yu (Yonsei University)</i>	
RCA-NOC: Relative Contrastive Alignment for Novel Object Captioning	15464
<i>Jiashuo Fan (Tsinghua University), Yaoyuan Liang (Tsinghua University), Leyao Liu (Tsinghua University), Shaolun Huang (Tsinghua-Berkeley Shenzhen Institute, Tsinghua University), and Lei Zhang (International Digital Economy Academy, IDEA)</i>	
DIME-FM : Distilling Multimodal and Efficient Foundation Models	15475
<i>Ximeng Sun (Boston University), Pengchuan Zhang (Microsoft Research AI), Peizhao Zhang (Meta), Hardik Shah (Meta Inc), Kate Saenko (Boston University), and Xide Xia (Meta)</i>	
Black Box Few-Shot Adaptation for Vision-Language Models	15488
<i>Yassine Ouali (Samsung), Adrian Bulat (Samsung AI Center, Cambridge), Brais Matinez (Samsung AI Center), and Georgios Tzimiropoulos (Queen Mary University of London)</i>	
Shatter and Gather: Learning Referring Image Segmentation with Text Supervision	15501
<i>Dongwon Kim (POSTECH), Namyup Kim (POSTECH), Cuiling Lan (Microsoft Research), and Suha Kwak (POSTECH)</i>	
Accurate and Fast Compressed Video Captioning	15512
<i>Yaojie Shen (Institute of Software, Chinese Academy of Sciences), Xin Gu (UCAS), Kai Xu (ByteDance Inc), Heng Fan (University of North Texas), Longyin Wen (Bytedance, Inc.), and Libo Zhang (Institute of Software Chinese Academy of Sciences)</i>	
Exploring Temporal Concurrency for Video-Language Representation Learning	15522
<i>Heng Zhang (Gaoling School of Artificial Intelligence, Renmin University of China), Daqing Liu (JD.com Inc), Zezhong Lv (Renmin University of China), Bing Su (Renmin University of China), and Dacheng Tao (The University of Sydney)</i>	
Verbs in Action: Improving Verb Understanding in Video-Language Models	15533
<i>Liliane Momeni (University of Oxford), Mathilde Caron (Google Research), Arsha Nagrani (Google), Andrew Zisserman (University of Oxford), and Cordelia Schmid (Google)</i>	
Sign Language Translation with Iterative Prototype	15546
<i>Huijie Yao (University of Science and Technology of China), Wengang Zhou (University of Science and Technology of China), Hao Feng (University of Science and Technology of China), Hezhen Hu (University of Science and Technology of China), Hao Zhou (University of Science and Technology of China), and Houqiang Li (University of Science and Technology of China)</i>	

Contrastive Feature Masking Open-Vocabulary Vision Transformer	15556
<i>Dahun Kim (Google), Anelia Angelova (Google), and Weicheng Kuo (Google)</i>	
Toward Unsupervised Realistic Visual Question Answering	15567
<i>Yuwei Zhang (University of California San Diego), Chih-Hui Ho (University of California San Diego), and Nuno Vasconcelos (University of California San Diego)</i>	
GridMM: Grid Memory Map for Vision-and-Language Navigation	15579
<i>Zihan Wang (Institute of Computing Technology, Chinese Academy of Sciences), Xiangyang Li (Institute of Computing Technology, Chinese Academy of Sciences), Jiahao Yang (Institute of Computing Technology, Chinese Academy of Sciences), Yeqi Liu (Institute of Computing Technology, Chinese Academy of Sciences), and Shuqiang Jiang (ICT, China Academy of Science)</i>	
Video Background Music Generation: Dataset, Method and Evaluation	15591
<i>Le Zhuo (Beihang University), Zhaokai Wang (Shanghai Jiao Tong University), Baisen Wang (Beihang University), Yue Liao (Beihang University), Chenxi Bao (University of Edinburgh), Stanley Peng (BUAA), Songhao Han (Beihang University), Aixi Zhang (Alibaba Group), Fei Fang (Alibaba Group), and Si Liu (Beihang University)</i>	
Prompt Switch: Efficient CLIP Adaptation for Text-Video Retrieval	15602
<i>Chaorui Deng (University of Adelaide), Qi Chen (University of Adelaide), Pengda Qin (Alibaba), Da Chen (University of Bath), and Qi Wu (University of Adelaide)</i>	
Prompt-aligned Gradient for Prompt Tuning	15613
<i>Beier Zhu (Nanyang Technological University), Yulei Niu (Columbia University), Yucheng Han (Nanyang Technological University), Yue Wu (Damo Academy, Alibaba Group), and Hanwang Zhang (Nanyang Technological University)</i>	
Knowledge-Aware Prompt Tuning for Generalizable Vision-Language Models	15624
<i>Baoshuo Kan (Qilu University of Technology, (Shandong Academy of Sciences), Teng Wang (The University of Hong Kong), Wenpeng Lu (Qilu University of Technology, (Shandong Academy of Sciences), Xiantong Zhen (United Imaging Healthcare), Weili Guan (Monash Univerisity), and Feng Zheng (SUSTech)</i>	
Order-Prompted Tag Sequence Generation for Video Tagging	15635
<i>Zongyang Ma (Institute of Automation, Chinese Academy of Sciences), Ziqi Zhang (CASIA), Yuxin Chen (Institute of Automation, Chinese Academy of Sciences), Zhongang Qi (Tencent), Yingmin Luo (Tencent), Zekun Li (National Computer Network Emergency Response Technical Team/Coordination Center of China), Chunfeng Yuan (NLPR), Bing Li (National Laboratory of Pattern Recognition, (NLPR), Xiaohu Qie (Tencent), Ying Shan (Tencent), and Weiming Hu (Institute of Automation Chinese Academy of Sciences)</i>	
What Does a Platypus Look Like? Generating Customized Prompts for Zero-Shot Image Classification	15645
<i>Sarah Pratt (University of Washington), Ian Covert (University of Washington), Rosanne Liu (ML Collective), and Ali Farhadi (University of Washington, Allen Institue for AI, Apple)</i>	

PromptStyler: Prompt-driven Style Generation for Source-free Domain Generalization	15656
<i>Junhyeong Cho (ADD), Gilhyun Nam (ADD), Sungyeon Kim (POSTECH), Hunmin Yang (KAIST), and Suha Kwak (POSTECH)</i>	
DiffDis: Empowering Generative Diffusion Model with Cross-Modal Discrimination Capability	15667
<i>Runhui Huang (Sun Yat-sen University), Jianhua Han (Huawei Noah's Ark Lab), Guansong Lu (Huawei Noah's Ark Lab), Xiaodan Liang (Sun Yat-sen University), Yihan Zeng (Shanghai Jiao Tong University), Wei Zhang (Noah's Ark Lab, Huawei Technologies), and Hang Xu (Huawei Noah's Ark Lab)</i>	
EdaDet: Open-Vocabulary Object Detection Using Early Dense Alignment	15678
<i>Cheng Shi (ShanghaiTech University) and Sibe Yang (ShanghaiTech University)</i>	
MixSpeech: Cross-Modality Self-Learning with Audio-Visual Stream Mixup for Visual Speech Translation and Recognition	15689
<i>Xize Cheng (Zhejiang University), Tao Jin (Zhejiang University), Rongjie Huang (Zhejiang University), Linjun Li (Zhejiang University), Wang Lin (Zhejiang University), Zehan Wang (Zhejiang University), Ye Wang (Zhejiang University), Huadai Liu (Zhejiang University), Aoxiong Yin (Zhejiang University), and Zhou Zhao (Zhejiang University)</i>	
Waffling Around for Performance: Visual Classification with Random Words and Broad Concepts	15700
<i>Karsten Roth (University of Tuebingen), Jae Myung Kim (University of Tuebingen), A. Sophia Koepke (University of Tübingen), Oriol Vinyals (DeepMind), Cordelia Schmid (Inria/Google), and Zeynep Akata (University of Tübingen)</i>	
March in Chat: Interactive Prompting for Remote Embodied Referring Expression	15712
<i>Yanyuan Qiao (University of Adelaide), Yuankai Qi (The University of Adelaide), Zheng Yu (University of Adelaide), Jing Liu (National Lab of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences), and Qi Wu (University of Adelaide)</i>	

(O) Self-, Semi-, and Unsupervised Learning

Kick Back & Relax: Learning to Reconstruct the World by Watching SlowTV	15722
<i>Jaime Spencer (University of Surrey), Chris Russell (University of Oxford), Simon Hadfield (University of Surrey), and Richard Bowden (University of Surrey)</i>	
Novel Scenes & Classes: Towards Adaptive Open-set Object Detection	15734
<i>Wuyang Li (City University of Hong Kong), Xiaoqing Guo (University of Oxford), and Yixuan Yuan (Chinese University of Hong Kong)</i>	
Improving Unsupervised Visual Program Inference with Code Rewriting Families	15745
<i>Aditya Ganeshan (Brown University), R. Kenny Jones (Brown University), and Daniel Ritchie (Brown University)</i>	
Denosing Diffusion Autoencoders are Unified Self-supervised Learners	15756
<i>Weilai Xiang (Beihang University), Hongyu Yang (Beihang University), Di Huang (Beihang University, China), and Yunhong Wang (State Key Laboratory of Virtual Reality Technology and System, Beihang University, Beijing 100191, China)</i>	

Self-Ordering Point Clouds	15767
<i>Pengwan Yang (University of Amsterdam), Cees G. M. Snoek (University of Amsterdam), and Yuki M. Asano (University of Amsterdam)</i>	
MOST: Multiple Object Localization with Self-Supervised Transformers for Object Discovery	15777
<i>Sai Saketh Rambhatla (Meta AI Research), Ishan Misra (Facebook AI Research), Rama Chellappa (Johns Hopkins University), and Abhinav Shrivastava (University of Maryland)</i>	
CHORUS : Learning Canonicalized 3D Human-Object Spatial Relations from Unbounded Synthesized Images	15789
<i>Sookwan Han (Seoul National University) and Hanbyul Joo (Seoul National University)</i>	
Identity-Seeking Self-Supervised Representation Learning for Generalizable Person Re-Identification	15801
<i>Zhaopeng Dou (Tsinghua University), Zhongdao Wang (Tsinghua University), Yali Li (Tsinghua University), and Shengjin Wang (Tsinghua University)</i>	
Anatomical Invariance Modeling and Semantic Alignment for Self-supervised Learning in 3D Medical Image Analysis	15813
<i>Yankai Jiang (College of Computer Science and Technology, Zhejiang University), Mingze Sun (tbsi), Heng Guo (Alibaba Group Damo Academy), Xiaoyu Bai (Alibaba Group DAMO Academy), Ke Yan (Alibaba DAMO Academy), Le Lu (Alibaba Group), and Minfeng Xu (Alibaba)</i>	
IOMatch: Simplifying Open-Set Semi-Supervised Learning with Joint Inliers and Outliers Utilization	15824
<i>Zekun Li (Nanjing University), Lei Qi (Southeast University), Yinghuan Shi (Nanjing University), and Yang Gao (Nanjing University)</i>	
Enhancing Sample Utilization through Sample Adaptive Augmentation in Semi-Supervised Learning	15834
<i>Guan Gui (Nanjing University), Zhen Zhao (University of Sydney), Lei Qi (Southeast University), Luping Zhou (University of Sydney), Lei Wang (University of Wollongong, Australia), and Yinghuan Shi (Nanjing University)</i>	
When Noisy Labels Meet Long Tail Dilemmas: A Representation Calibration Method	15844
<i>Manyi Zhang (Tsinghua University), Xuyang Zhao (Peking University), Jun Yao (Huawei Technologies), Chun Yuan (Graduate school at ShenZhen Tsinghua university), and Weiran Huang (Qing Yuan Research Institute, Shanghai Jiao Tong University)</i>	

(O) Generative AI 2

Cross-Ray Neural Radiance Fields for Novel-View Synthesis from Unconstrained Image Collections	15855
<i>Yifan Yang (South China University of Technology), Shuhai Zhang (South China University of Technology), Zixiong Huang (scut), Yubing Zhang (Guangzhou Shiyuan Electronics Co., Ltd(CVTE), and Mingkui Tan (South China University of Technology)</i>	

Effective Real Image Editing with Accelerated Iterative Diffusion Inversion	15866
<i>Zhihong Pan (OPPO), Riccardo Gherardi (Oppo), Xiufeng Xie (OPPO), and Stephen Huang (OPPO)</i>	
Simulating Fluids in Real-World Still Images	15876
<i>Siming Fan (SenseTime), Jingtian Piao (CUHK MMLab), Chen Qian (SenseTime), Hongsheng Li (The Chinese University of Hong Kong), and Kwan-Yee Lin (SenseTime Research)</i>	
FateZero: Fusing Attentions for Zero-shot Text-based Video Editing	15886
<i>Chenyang Qi (The Hong Kong University of Science and Technology), Xiaodong Cun (Tencent AI Lab), Yong Zhang (Tencent AI Lab), Chenyang Lei (CAIR, HKISI-CAS / Princeton University), Xintao Wang (Tencent), Ying Shan (Tencent), and Qifeng Chen (HKUST)</i>	
ELITE: Encoding Visual Concepts into Textual Embeddings for Customized Text-to-Image Generation	15897
<i>Yuxiang Wei (Harbin Institute of Technology), Yabo Zhang (Harbin Institute of Technology), Zhilong Ji (Tomorrow Advancing Life), Jinfeng Bai (Tomorrow Advance Life), Lei Zhang (Hong Kong Polytechnic University, Hong Kong, China), and Wangmeng Zuo (Harbin Institute of Technology, China)</i>	
Text2Video-Zero: Text-to-Image Diffusion Models are Zero-Shot Video Generators	15908
<i>Levon Khachatryan (Picsart AI Research), Andranik Movsisyan (Picsart), Vahram Tadevosyan (Picsart), Roberto Henschel (Leibniz University of Hannover), Zhangyang Wang (University of Texas at Austin), Shant Navasardyan (Picsart Inc.), and Humphrey Shi (U of Oregon; UIUC; PAIR)</i>	
Chupa: Carving 3D Clothed Humans from Skinned Shape Priors using 2D Diffusion Probabilistic Models	15919
<i>Byungjun Kim (Seoul National University), Patrick Kwon (Naver Webtoon), Kwangho Lee (NAVER WEBTOON AI), Myunggi Lee (Seoul National University), Sookwan Han (Seoul National University), Daesik Kim (Naver webtoon), and Hanbyul Joo (Seoul National University)</i>	
DiffPose: Multi-hypothesis Human Pose Estimation using Diffusion Models	15931
<i>Karl Holmquist (Linköping University) and Bastian Wandt (Linköping University)</i>	
HumanSD: A Native Skeleton-Guided Diffusion Model for Human Image Generation	15942
<i>Xuan Ju (The Chinese University of Hong Kong), Ailing Zeng (International Digital Economy Academy), Chenchen Zhao (The Chinese University of Hong Kong), Jianan Wang (International Digital Economy Academy, IDEA), Lei Zhang (International Digital Economy Academy, IDEA), and Qiang Xu (The Chinese University of Hong Kong)</i>	
Role-Aware Interaction Generation from Textual Description	15953
<i>Mikihiro Tanaka (LINE Corporation) and KentFujiwara KentFujiwara (LINE Corporation)</i>	
PhysDiff: Physics-Guided Human Motion Diffusion Model	15964
<i>Ye Yuan (NVIDIA Research), Jiaming Song (NVIDIA), Umar Iqbal (NVIDIA Research), Arash Vahdat (NVIDIA), and Jan Kautz (NVIDIA)</i>	

Forward Flow for Novel View Synthesis of Dynamic Scenes	15976
<i>Xiang Guo (Northwestern Polytechnical University), Jiadai Sun (Northwestern Polytechnical University), Yuchao Dai (Northwestern Polytechnical University), Guanying Chen (The Chinese University of Hong Kong, Shenzhen), Xiaoqing Ye (Baidu Inc.), Xiao Tan (Baidu Inc.), Errui Ding (Baidu Inc.), Yumeng Zhang (Tsinghua University), and Jingdong Wang (Baidu)</i>	

Self-, Semi-, Meta-, Unsupervised Learning

Noise2Info: Noisy Image to Information of Noise for Self-Supervised Image Denoising	15988
<i>Jiachuan Wang (HKUST), Shimin Di (The Hong Kong University of Science and Technology), Lei Chen (Hong Kong University of Science and Technology), and Charles Wang Wai Wai Ng (Hong Kong University of Science and Technology)</i>	
Box-based Refinement for Weakly Supervised and Unsupervised Localization Tasks	15998
<i>Eyal Gomel (None), Tal Shaharbany (Tel Aviv University), and Lior Wolf (Tel Aviv University, Israel)</i>	
Diverse Cotraining Makes Strong Semi-Supervised Segmentor	16009
<i>Yijiang Li (Johns Hopkins University), Xinjiang Wang (SenseTime Research), Lihe Yang (Nanjing University), Litong Feng (SenseTime Research), Wayne Zhang (SenseTime Research), and Ying Gao (south china university of technology)</i>	
SSB: Simple but Strong Baseline for Boosting Performance of Open-Set Semi-Supervised Learning	16022
<i>Yue Fan (Max Planck Institute for Informatics), Anna Kukleva (MPII), Dengxin Dai (MPI for Informatics), and Bernt Schiele (MPI Informatics)</i>	
Late Stopping: Avoiding Confidently Learning from Mislabeled Examples	16033
<i>Suqin Yuan (The University of Sydney), Lei Feng (Nanyang Technological University), and Tongliang Liu (The University of Sydney)</i>	
Ponder: Point Cloud Pre-training via Neural Rendering	16043
<i>Di Huang (Zhejiang University), Sida Peng (Zhejiang University), Tong He (Shanghai AI lab), Honghui Yang (Zhejiang University), Xiaowei Zhou (Zhejiang University), and Wanli Ouyang (The University of Sydney)</i>	
Semantics-Consistent Feature Search for Self-Supervised Visual Representation Learning	16053
<i>Kaiyou Song (MEGVII Technology), Shan Zhang (MEGVII Technology), Zimeng Luo (MEGVII Technology), Tong Wang (MEGVII Technology), and Jin Xie (MEGVII Technology)</i>	
Stable and Causal Inference for Discriminative Self-supervised Deep Visual Representations.....	16063
<i>Yuewei Yang (Duke University), Hai Li (Duke University), and Yiran Chen (Duke University)</i>	
Towards Semi-supervised Learning with Non-random Missing Labels	16075
<i>Yue Duan (Nanjing University), Zhen Zhao (University of Sydney), Lei Qi (Southeast University), Luping Zhou (University of Sydney), Lei Wang (University of Wollongong, Australia), and Yinghuan Shi (Nanjing University)</i>	

Hallucination Improves the Performance of Unsupervised Visual Representation Learning	16086
<i>Jing Wu (University of Illinois Urbana-Champaign), Jennifer Hobbs (Intel/Air), and Naira Hovakimyan (UIUC)</i>	
Audiovisual Masked Autoencoders	16098
<i>Mariana-Iuliana Georgescu (University of Bucharest), Eduardo Fonseca (Google Research), Radu Tudor Ionescu (University of Bucharest), Mario Lucic (Google Brain), Cordelia Schmid (Google), and Anurag Arnab (Google)</i>	
PADCLIP: Pseudo-labeling with Adaptive Debiasing in CLIP for Unsupervised Domain Adaptation	16109
<i>Zhengfeng Lai (University of California, Davis), Noranart Vesdapunt (Amazon), Ning Zhou (Amazon), Jun Wu (Amazon), Cong Phuoc Huynh (Amazon), Xuelu Li (Amazon), Kah Kuen Fu (Amazon), and Chen-Nee Chuah (University of California Davis)</i>	
Removing Anomalies as Noises for Industrial Defect Localization	16120
<i>Fanbin Lu (The Chinese University of Hong Kong), Xufeng Yao (The Chinese University of Hong Kong), Chi-Wing Fu (The Chinese University of Hong Kong), and Jiaya Jia (The Chinese University of Hong Kong)</i>	
SparseMAE: Sparse Training Meets Masked Autoencoders	16130
<i>Aojun Zhou (The Chinese University of Hong Kong), Yang Li (SenseTime Research), Zipeng Qin (The Chinese University of Hong Kong), Jianbo Liu (The Chinese University of Hong Kong), Junting Pan (The Chinese University of Hong Kong), Renrui Zhang (Shanghai AI Lab), Rui Zhao (SenseTime Group Limited), Peng Gao (Chinese university of hong kong), and Hongsheng Li (The Chinese University of Hong Kong)</i>	
Shrinking Class Space for Enhanced Certainty in Semi-Supervised Learning	16141
<i>Lihe Yang (Nanjing University), Zhen Zhao (University of Sydney), Lei Qi (Southeast University), Yu Qiao (Shanghai AI Laboratory), Yinghuan Shi (Nanjing University), and Hengshuang Zhao (The University of Hong Kong)</i>	
Logic-induced Diagnostic Reasoning for Semi-supervised Semantic Segmentation	16151
<i>Chen Liang (Zhejiang University), Wenguan Wang (Zhejiang University), Jiaxu Miao (Zhejiang University), and Yi Yang (Zhejiang University)</i>	
GasMono: Geometry-Aided Self-Supervised Monocular Depth Estimation for Indoor Scenes	16163
<i>Chaoqiang Zhao (East China University of Science and Technology), Matteo Poggi (University of Bologna), Fabio Tosi (University of Bologna), Lei Zhou (East china university of science and technology), Qiyu Sun (East China University of Science and Technology), Yang Tang (East China University of Science and Technology), and Stefano Mattoccia (University of Bologna)</i>	
Is Imitation All You Need? Generalized Decision-Making with Dual-Phase Training	16175
<i>Yao Wei (Zhejiang University), Yanchao Sun (UMD), Ruijie Zheng (University of Maryland, College Park), Sai Vemprala (Microsoft), Rogerio Bonatti (Microsoft), Shuhang Chen (Microsoft), Ratnesh Madaan (Personal), Zhongjie Ba (Zhejiang University), Ashish Kapoor (Microsoft), and Shuang Ma (Microsoft)</i>	

Benchmarking Low-Shot Robustness to Natural Distribution Shifts	16186
<i>Aaditya Singh (Georgia Institute of Technology), Kartik Sarangmath (Georgia Institute of Technology), Prithvijit Chattopadhyay (Georgia Institute of Technology), and Judy Hoffman (Georgia Institute of Technology)</i>	
All4One: Symbiotic Neighbour Contrastive Learning via Self-Attention and Redundancy Reduction	16197
<i>Imanol G. Estepa (Universitat de Barcelona), Ignacio Sarasua (NVIDIA AI Technology Center), Bhalaji Nagarajan (University of Barcelona), and Petia Radeva (University of Barcelona)</i>	
Weakly Supervised Learning of Semantic Correspondence through Cascaded Online Correspondence Refinement	16208
<i>Yiwen Huang (Fudan University), Yixuan Sun (Fudan University), Chenghang Lai (Fudan university), Qing Xu (UniDT), Xiaomei Wang (UniDT), Xuli Shen (Fudan University), and Weifeng Ge (Fudan University)</i>	
Tracking without Label: Unsupervised Multiple Object Tracking via Contrastive Similarity Learning	16218
<i>Sha Meng (Northwestern Polytechnical University), Dian Shao (Northwestern Polytechnical Univesity), Jiacheng Guo (Northwestern Polytechnical University), and Shan Gao (nwpu)</i>	
Active Self-Supervised Learning: A Few Low-Cost Relationships Are All You Need	16228
<i>Vivien Cabannes (Meta AI), Leon Bottou (Meta AI), Yann Lecun (Meta AI), and Randall Balestriero (Meta AI)</i>	
Diffusion Models as Masked Autoencoders	16238
<i>Chen Wei (FAIR, Meta AI; Johns Hopkins University), Karttikeya Mangalam (FAIR, Meta AI), Po-Yao Huang (FAIR, Meta AI), Yanghao Li (FAIR, Meta AI), Haoqi Fan (FAIR, Meta AI), Hu Xu (FAIR, Meta AI), Huiyu Wang (FAIR, Meta AI), Cihang Xie (UC Santa Cruz), Alan Yuille (Johns Hopkins Unioersity), and Christoph Feichtenhofer (FAIR, Meta AI)</i>	
Enhanced Meta Label Correction for Coping with Label Corruption	16249
<i>Mitchell Keren Taraday (Technion) and Chaim Baskin (Technion)</i>	
Randomized Quantization: A Generic Augmentation for Data Agnostic Self-supervised Learning	16259
<i>Huimin Wu (The Hong Kong University of Science and Technology), Chenyang Lei (CAIR, HKISI-CAS / Princeton University), Xiao Sun (Shanghai Artificial Intelligence Laboratory), Peng-Shuai Wang (Peking University), Qifeng Chen (HKUST), Kwang-Ting Cheng (Hong Kong University of Science and Technology), Stephen Lin (Microsoft Research), and Zhirong Wu (Microsoft Research)</i>	
Prototypes-oriented Transductive Few-shot Learning with Conditional Transport	16271
<i>Long Tian (Xidian University), Jingyi Feng (Xidian University), Xiaoqiang Chai (Xidian University), Wenchao Chen (Xi'dian University), Liming Wang (Xidian University), Xiyang Liu (Xidian University), and Bo Chen (Xidian University)</i>	

Contrastive Learning Relies More on Spatial Inductive Bias Than Supervised Learning: An Empirical Study	16281
<i>Yuanyi Zhong (University of Illinois at Urbana-Champaign), Haoran Tang (University of Pennsylvania), Jun-Kun Chen (University of Illinois at Urbana-Champaign), and Yu-Xiong Wang (University of Illinois at Urbana-Champaign)</i>	
Pseudo-label Alignment for Semi-supervised Instance Segmentation	16291
<i>Jie Hu (Xiamen University), Chen Chen (Xiamen University), Liujuan Cao (Xiamen University), Shengchuan Zhang (Xiamen University), Annan Shu (Contemporary Amperex Technology Co. Limited), Guannan Jiang (CATL), and Rongrong Ji (Xiamen University, China)</i>	
CFCG: Semi-Supervised Semantic Segmentation via Cross-Fusion and Contour Guidance Supervision	16302
<i>Shuo Li (Institute of Computing Technology, Chinese Academy of Sciences), Yue He (Baidu), Weiming Zhang (baidu), Wei Zhang (Baidu Inc), Xiao Tan (Baidu Inc.), Junyu Han (Baidu Inc.), Errui Ding (Baidu Inc.), and Jingdong Wang (Baidu)</i>	
Pixel-Wise Contrastive Distillation	16313
<i>Junqiang Huang (Shopee) and Zichao Guo (Shopee)</i>	
Rethinking Safe Semi-supervised Learning: Transferring the Open-set Problem to A Close-set One	16324
<i>Qiankun Ma (Sichuan University), Jiyao Gao (Sichuan University), Bo Zhan (Sichuan University), Yunpeng Guo (), Jiliu Zhou (Sichuan University), and Yan Wang (Sichuan University)</i>	
Towards Open-Set Test-Time Adaptation Utilizing the Wisdom of Crowds in Entropy Minimization	16334
<i>Jungsoo Lee (Graduate School of Artificial Intelligence, KAIST), Debasmit Das (Qualcomm AI Research), Jaegul Choo (Korea Advanced Institute of Science and Technology), and Sungha Choi (Qualcomm AI Research)</i>	
Gradient-based Sampling for Class Imbalanced Semi-supervised Object Detection	16344
<i>Jiaming Li (Sun Yat-sen University), Xiangru Lin (HKU), Wei Zhang (Baidu Inc), Xiao Tan (Baidu Inc.), Yingying Li (Baidu), Junyu Han (Baidu Inc.), Errui Ding (Baidu Inc.), Jingdong Wang (Baidu), and Guanbin Li (Sun Yat-sen University)</i>	
Remembering Normality: Memory-guided Knowledge Distillation for Unsupervised Anomaly Detection	16355
<i>Zhihao Gu (Shanghai Jiao Tong University), Liang Liu (Tencent), Xu Chen (Tencent), Ran Yi (Shanghai Jiao Tong University), Jiangning Zhang (Zhejiang University), Yabiao Wang (Tencent), Chengjie Wang (Tencent, Shanghai Jiao Tong University), Annan Shu (Contemporary Amperex Technology Co. Limited), Guannan Jiang (CATL), and Lizhuang Ma (Shanghai Jiao Tong University)</i>	
Semi-Supervised Learning via Weight-Aware Distillation under Class Distribution Mismatch ...	16364
<i>Pan Du (Renmin University of China), Suyun Zhao (Renmin University of China), Zisen Sheng (Renmin university), Cuiping Li (Renmin University of China), and Hong Chen (Renmin University, China)</i>	

Label Shift Adapter for Test-Time Adaptation under Covariate and Label Shifts	16375
<i>Sunghyun Park (KAIST), Seunghan Yang (Qualcomm AI Research), Jaegul Choo (Korea Advanced Institute of Science and Technology), and Sungrack Yun (Qualcomm AI Research)</i>	
SimMatchV2: Semi-Supervised Learning with Graph Consistency	16386
<i>Mingkai Zheng (The University of Sydney), Shan You (SenseTime), Lang Huang (The University of Tokyo), Chen Luo (State grid anhui electric power research institute), Fei Wang (University of Science and Technology of China), Chen Qian (SenseTime), and Chang Xu (University of Sydney)</i>	
Unsupervised Accuracy Estimation of Deep Visual Models using Domain-Adaptive Adversarial Perturbation without Source Samples	16397
<i>JoonHo Lee (Samsung SDS Technology Research), Jae Oh Woo (Samsung SDS America), Hankyu Moon (Samsung SDS Research America), and Kwonho Lee (Samsung SDS)</i>	
Learning by Sorting: Self-supervised Learning with Group Ordering Constraints	16407
<i>Nina Shvetsova (Goethe University Frankfurt), Felix Petersen (Stanford University), Anna Kukleva (MPII), Bernt Schiele (MPI Informatics), and Hilde Kuehne (University of Bonn)</i>	
L-DAWA: Layer-wise Divergence Aware Weight Aggregation in Federated Self-Supervised Visual Representation Learning	16418
<i>Yasar Abbas Ur Rehman (TCL Corporate Research(Hong Kong), Yan Gao (University of Cambridge), Pedro Porto Buarque de Gusmão (University of Cambridge), Mina Alibeigi (Zenseact), Jiajun Shen (TCL Research), and Nicholas D. Lane (University of Cambridge and Samsung AI)</i>	
Class-relation Knowledge Distillation for Novel Class Discovery	16428
<i>Peiyan Gu (ShanghaiTech University), Chuyu Zhang (PLUS Lab, ShanghaiTech University), Ruijie Xu (shanghaitech university), and Xuming He (ShanghaiTech University)</i>	
Representation Uncertainty in Self-Supervised Learning as Variational Inference	16438
<i>Hiroki Nakamura (Panasonic Holdings Corporation), Masashi Okada (Panasonic Holdings Corporation), and Tadahiro Taniguchi (Ritsumeikan University)</i>	
Point-TTA: Test-Time Adaptation for Point Cloud Registration Using Multitask Meta-Auxiliary Learning	16448
<i>Ahmed Hatem (University of Manitoba), Yiming Qian (Amazon), and Yang Wang (Concordia University)</i>	
Adaptive Similarity Bootstrapping for Self-Distillation Based Representation Learning	16459
<i>Tim Lebailly (KU Leuven), Thomas Stegmüller (EPFL), Behzad Bozorgtabar (EPFL), Jean-Philippe Thiran (École Polytechnique Fédérale de Lausanne), and Tinne Tuytelaars (KU Leuven)</i>	
Point Contrastive Prediction with Semantic Clustering for Self-Supervised Learning on Point Cloud Videos	16469
<i>Xiaoxiao Sheng (Shanghai Jiao Tong University), Zhiqiang Shen (Shanghai Jiao Tong University), Gang Xiao (Shanghai Jiao Tong University), Longguang Wang (National University of Defense Technology), Yulan Guo (Sun Yat-sen University), and Hehe Fan (ZJU)</i>	

MHCN: A Hyperbolic Neural Network Model for Multi-view Hierarchical Clustering	16479
<i>Fangfei Lin (University of Electronic Science and Technology of China), Bing Bai (Tsinghua University), Yiwen Guo (Independent Researcher), Hao Chen (UC Davis), Yazhou Ren (University of Electronic Science and Technology of China), and Zenglin Xu (Harbin Institute of Technology, Shenzhen)</i>	
Time Does Tell: Self-Supervised Time-Tuning of Dense Image Representations	16490
<i>Mohammadreza Salehi (University of Amsterdam), Efstratios Gavves (University of Amsterdam), Cees G.M. Snoek (University of Amsterdam), and Yuki M. Asano (University of Amsterdam)</i>	
To Adapt or Not to Adapt? Real-Time Adaptation for Semantic Segmentation	16502
<i>Marc Botet Colomer (KTH Royal Institute of Technology), Pier Luigi Dovesi (Univrses), Theodoros Panagiotakopoulos (King), Joao Frederico Carvalho (Univrses AB), Linus Härenstam-Nielsen (TUM), Hossein Azizpour (KTH, (Royal Institute of Technology), Hedvig Kjellström (KTH Royal Institute of Technology), Daniel Cremers (TU Munich), and Matteo Poggi (University of Bologna)</i>	
Simple and Effective Out-of-Distribution Detection via Cosine-based Softmax Loss	16514
<i>SoonCheol Noh (Sungkyunkwan University), DongEon Jeong (Sungkyunkwan University), and Jee-Hyong Lee (Sungkyunkwan University)</i>	
MixBag: Bag-Level Data Augmentation for Learning from Label Proportions	16524
<i>Takanori Asanomi (Kyushu University), Shinnosuke Matsuo (Kyushu University), Daiki Suehiro (Yokohama City University), and Ryoma Bise (Kyushu University)</i>	
Masked Spatio-Temporal Structure Prediction for Self-supervised Learning on Point Cloud Videos	16534
<i>Zhiqiang Shen (Shanghai Jiao Tong University), Xiaoxiao Sheng (Shanghai Jiao Tong University), Hehe Fan (ZJU), Longguang Wang (National University of Defense Technology), Yulan Guo (Sun Yat-sen University), Qiong Liu (CloudWalk), Hao Wen (CloudWalk), and Xi Zhou (CloudWalk Technology)</i>	
Parametric Classification for Generalized Category Discovery: A Baseline Study	16544
<i>Xin Wen (The University of Hong Kong), Bingchen Zhao (University of Edinburgh), and Xiaojuan Qi (The University of Hong Kong)</i>	
Object-Centric Multiple Object Tracking	16555
<i>Zixu Zhao (Amazon), Jiaze Wang (The Chinese University of Hong Kong), Max Horn (Amazon Web Services), Yizhuo Ding (Fudan University), Tong He (Amazon), Zechen Bai (Amazon), Dominik Zietlow (Amazon Web Services), Carl-Johann Simon-Gabriel (ETH Zürich), Bing Shuai (Amazon), Zhuowen Tu (UC San Diego), Thomas Brox (University of Freiburg), Bernt Schiele (MPI Informatics), Yanwei Fu (Fudan University), Francesco Locatello (Amazon), Zheng Zhang (Amazon), and Tianjun Xiao (Amazon)</i>	
Locating Noise is Halfway Denoising for Semi-Supervised Segmentation	16566
<i>Yan Fang (Beijing Jiaotong University), Feng Zhu (University of Technology Sydney), Bowen Cheng (UIUC), Luoqi Liu (meitu), Yao Zhao (Beijing Jiaotong University), and Yunchao Wei (UTS)</i>	

Learning Semi-supervised Gaussian Mixture Models for Generalized Category Discovery	16577
<i>Bingchen Zhao (University of Edinburgh), Xin Wen (The University of Hong Kong), and Kai Han (The University of Hong Kong)</i>	
LoCUS: Learning Multiscale 3D-consistent Features from Posed Images	16588
<i>Dominik A. Kloepper (University of Oxford), Dylan Campbell (Australian National University), and João F. Henriques (University of Oxford)</i>	
Stable Cluster Discrimination for Deep Clustering	16599
<i>Qi Qian (Alibaba Group)</i>	
Cross-modal Scalable Hierarchical Clustering in Hyperbolic space	16609
<i>Teng Long (University of Amsterdam) and Nanne van Noord (University of Amsterdam)</i>	
Collaborative Propagation on Multiple Instance Graphs for 3D Instance Segmentation with Single-point Supervision	16619
<i>Shichao Dong (Nanyang Technological University), Ruibo Li (Nanyang Technological University), Jiacheng Wei (Nanyang Technological University), Fayao Liu (Institute for Infocomm Research, A*STAR), and Guosheng Lin (Nanyang Technological University)</i>	
Semantics Meets Temporal Correspondence: Self-supervised Object-centric Learning in Videos..	16629
<i>Rui Qian (The Chinese University of Hong Kong), Shuangrui Ding (Shanghai Jiao Tong University), Xian Liu (The Chinese University of Hong Kong), and Dahua Lin (The Chinese University of Hong Kong)</i>	
Proxy Anchor-based Unsupervised Learning for Continuous Generalized Category Discovery ..	16642
<i>Hyungmin Kim (KAIST), Sungho Suh (DFKI), Daehwan Kim (Samsung Electro-mechanics), Daun Jeong (Samsung electro-mechanics), Hansang Cho (Samsung Electro-Mechanics), and Junmo Kim (KAIST)</i>	
DreamTeacher: Pretraining Image Backbones with Deep Generative Models	16652
<i>Daiqing Li (NVIDIA), Huan Ling (University of Toronto, NVIDIA), Amlan Kar (University of Toronto, NVIDIA), David Acuna (University of Toronto, NVIDIA), Seung Wook Kim (University of Toronto, NVIDIA), Karsten Kreis (NVIDIA), Antonio Torralba (MIT), and Sanja Fidler (University of Toronto, NVIDIA)</i>	
MATE: Masked Autoencoders are Online 3D Test-Time Learners	16663
<i>M. Jehanzeb Mirza (Technical University of Graz), Inkyu Shin (KAIST), Wei Lin (Graz University of Technology), Andreas Schriebl (Graz University of Technology), Kunyang Sun (Southeast University), Jaesung Choe (KAIST), Mateusz Kozinski (ICG TUGRAZ), Horst Possegger (Graz University of Technology), In So Kweon (KAIST), Kuk-jin Yoon (KAIST), and Horst Bischof (Graz University of Technology)</i>	
PADDLES: Phase-Amplitude Spectrum Disentangled Early Stopping for Learning with Noisy Labels	16673
<i>Huaxi Huang (CSIRO), Hui Kang (USYD), Sheng Liu (NYU), Olivier Salvado (CSIRO), Thierry Rakotoarivelo (Data61-CSIRO), Dadong Wang (CSIRO Data61), and Tongliang Liu (The University of Sydney)</i>	
Calibrating Uncertainty for Semi-Supervised Crowd Counting	16685
<i>Chen Li (Stony Brook University), Xiaoling Hu (Stony Brook University), Shahira Abousamra (Stony brook university), and Chao Chen (Stony Brook University)</i>	

Test Time Adaptation for Blind Image Quality Assessment	16696
<i>Subhadeep Roy (Indian Institute of Science), Shankhanil Mitra (Indian Institute of Science), Soma Biswas (Indian Institute of Science, Bangalore), and Rajiv Soundararajan (Indian Institute of Science)</i>	
Deep Multiview Clustering by Contrasting Cluster Assignments	16706
<i>Jie Chen (College of Computer Science, Sichuan University), Hua Mao (Northumbria University), Wai Lok Woo (Northumbria University), and Xi Peng (College of Computer Science, Sichuan University)</i>	

Photogrammetry and Remote Sensing

Re:PolyWorld - A Graph Neural Network for Polygonal Scene Parsing	16716
<i>Stefano Zorzi (Graz University of Technology) and Friedrich Fraundorfer (Graz University of Technology)</i>	
SatlasPretrain: A Large-Scale Dataset for Remote Sensing Image Understanding	16726
<i>Fayen Bastani (Allen Institute for AI), Piper Wolters (Allen Institute for AI), Ritwik Gupta (University of California, Berkeley), Joe Ferdinando (Allen Institute for Artificial Intelligence), and Aniruddha Kembhavi (Allen Institute for Artificial Intelligence)</i>	
Large-Scale Land Cover Mapping with Fine-Grained Classes via Class-Aware Semi-Supervised Semantic Segmentation	16737
<i>Runmin Dong (Tsinghua University), Lichao Mou (DLR&TUM), Mengxuan Chen (Tsinghua University), Weijia Li (Sun Yat-sen University), Xin-Yi Tong (Technical University of Munich), Shuai Yuan (City University of Hong Kong), Lixian Zhang (Tsinghua University), Juepeng Zheng (Sun Yat-Sen University), Xiao Xiang Zhu (Technical University of Munich, Germany), and Haohuan Fu (missing)</i>	
Large Selective Kernel Network for Remote Sensing Object Detection	16748
<i>Yuxuan Li (Nankai University), Qibin Hou (Nankai University), Zhaohui Zheng (Nankai University), Ming-Ming Cheng (Nankai University), Jian Yang (Nankai University), and Xiang Li (Nankai University)</i>	
Towards Geospatial Foundation Models via Continual Pretraining	16760
<i>Matías Mendieta (University of Central Florida), Boran Han (AWS), Xingjian Shi (HKUST), Yi Zhu (Amazon), and Chen Chen (University of Central Florida)</i>	
Regularized Primitive Graph Learning for Unified Vector Mapping	16771
<i>Lei Wang (Huawei), Min Dai (Huawei Technologies Co.), Jianan He (Rieman Lab, Huawei Technologies.), and Jingwei Huang (Huawei)</i>	
Class Prior-Free Positive-Unlabeled Learning with Taylor Variational Loss for Hyperspectral Remote Sensing Imagery	16781
<i>Hengwei Zhao (Wuhan University), Xinyu Wang (Wuhan University), Jingtao Li (Whuhan university), and Yanfei Zhong (Wuhan University)</i>	
MapFormer: Boosting Change Detection by Using Pre-change Information	16791
<i>Maximilian Bernhard (Ludwig-Maximilians-Universität München), Niklas Strauß (LMU Munich), and Matthias Schubert (Ludwig-Maximilians-Universität München)</i>	

Sample4Geo: Hard Negative Sampling For Cross-View Geo-Localisation	16801
<i>Fabian Deuser (University of the Bundeswehr Munich), Konrad Habel (University of the Bundeswehr Munich), and Norbert Oswald (University of the Bundeswehr Munich)</i>	
PanFlowNet: A Flow-Based Deep Network for Pan-Sharpening	16811
<i>Gang Yang (University of Science and Technology of China), Xiangyong Cao (Xi'an Jiaotong University), Wenzhe Xiao (Xi'an Jiaotong University), Man Zhou (University of Science and Technology of China), Aiping Liu (University of Science and Technology of China), Xun Chen (University of Science and Technology of China), and Deyu Meng (Xi'an Jiaotong University)</i>	
Seeing Beyond the Patch: Scale-Adaptive Semantic Segmentation of High-resolution Remote Sensing Imagery based on Reinforcement Learning	16822
<i>Yinhe Liu (Wuhan University), Sunan Shi (Wuhan University), Junjue Wang (Wuhan University), and Yanfei Zhong (Wuhan University)</i>	

Efficient and Scalable Vision

AdaNIC: Towards Practical Neural Image Compression via Dynamic Transform Routing	16833
<i>Lufang Tao (Tencent AI Lab), Wei Gao (SECE, Shenzhen Graduate School, Peking University), Ge Li (SECE, Shenzhen Graduate School, Peking University), and Chenhao Zhang (PKUSZ)</i>	
Rethinking Vision Transformers for MobileNet Size and Speed	16843
<i>Yanyu Li (Northeastern University), Ju Hu (Snap Inc.), Yang Wen (Snap Inc.), Georgios Evangelidis (Snap Inc), Kamyar Salahi (UC Berkeley), Yanzhi Wang (Northeastern University), Sergey Tulyakov (Snap Inc), and Jian Ren (Snap Inc.)</i>	
DELFlow: Dense Efficient Learning of Scene Flow for Large-Scale Point Clouds	16855
<i>Chensheng Peng (Shanghai Jiao Tong University), Guangming Wang (University of Cambridge), Xian Wan Lo (Shanghai Jiao Tong University), Xinrui Wu (Shanghai Jiao Tong University), Chenfeng Xu (UC Berkeley), Masayoshi Tomizuka (MSC Lab), Wei Zhan (University of California, Berkeley), and Hesheng Wang (SJTU)</i>	
Eventful Transformers: Leveraging Temporal Redundancy in Vision Transformers	16865
<i>Matthew Dutson (University of Wisconsin-Madison), Yin Li (University of Wisconsin-Madison), and Mohit Gupta (University of Wisconsin-Madison, USA)</i>	
Inherent Redundancy in Spiking Neural Networks	16878
<i>Man Yao (Xi'an Jiaotong University), Jiakui Hu (Peking University), Guangshe Zhao (Xi'an Jiaotong University), Yaoyuan Wang (Huawei Technologies Co., Ltd.), Ziyang Zhang (HUAWEI TECHNOLOGIES CO.LTD), Bo Xu (Institute of Automation, Chinese Academy of Sciences), and Guoqi Li (Tsinghua University)</i>	
Achievement-Based Training Progress Balancing for Multi-Task Learning	16889
<i>Hayoung Yun (Samsung Research) and Hanjoo Cho (Samsung Research)</i>	

Prune Spatio-temporal Tokens by Semantic-aware Temporal Accumulation	16899
<i>Shuangrui Ding (Shanghai Jiao Tong University), Peisen Zhao (Huawei), Xiaopeng Zhang (Huawei Cloud EI), Rui Qian (The Chinese University of Hong Kong), Hongkai Xiong (Shanghai Jiao Tong University), and Qi Tian (Huawei Cloud BU)</i>	
Differentiable Transportation Pruning	16911
<i>Yunqiang Li (TU Delft), Jan C. van Gemert (Delft University of Technology), Torsten Hoeffler (ETH Zurich), Bert Moons (Synopsis), Evangelos Eleftheriou (Axelera AI), and Bram-Ernst Verhoef (Imec)</i>	
XiNet: Efficient Neural Networks for tinyML	16922
<i>Alberto Ancilotto (Fondazione Bruno Kessler), Francesco Paissan (Fondazione Bruno Kessler), and Elisabetta Farella (Fondazione Bruno Kessler)</i>	
Jumping through Local Minima: Quantization in the Loss Landscape of Vision Transformers	16932
<i>Natalia Frumkin (The University of Texas at Austin), Dibakar Gope (Arm Research), and Diana Marculescu (The University of Texas at Austin)</i>	
A2Q: Accumulator-Aware Quantization with Guaranteed Overflow Avoidance	16943
<i>Ian Colbert (AMD), Alessandro Pappalardo (Advanced Micro Devices), and Jakoba Petri-Koenig (AMD)</i>	
Workie-Talkie: Accelerating Federated Learning by Overlapping Computing and Communications via Contrastive Regularization	16953
<i>Rui Chen (University of Houston), Qiyu Wan (University of Houston), Pavana Prakash (University of Houston), Lan Zhang (Michigan Technological University), Xu Yuan (University of Delaware), Yanmin Gong (UTSA), Xin Fu (University of Houston), and Miao Pan (University of Houston)</i>	
DenseShift: Towards Accurate and Efficient Low-Bit Power-of-Two Quantization	16964
<i>Xinlin Li (Huawei Noah's Ark Lab), Bang Liu (University of Montreal), Rui Heng Yang (Huawei Noah's Ark Lab), Vanessa Courville (Huawei Noah's Ark Lab), Chao Xing (Huawei Noah's Ark Lab), and Vahid Partovi Nia (Huawei Noah's Ark Lab)</i>	
PRANC: Pseudo Random Networks for Compacting Deep Models	16975
<i>Parsa Nooralinejad (University of California-Davis), Ali Abbasi (Vanderbilt University), Soroush Abbasi Koochpayegani (University of California, Davis), Kossar Pourahmadi Meibodi (University of California, Davis), Rana Muhammad Shahroz Khan (Vanderbilt University), Soheil Kolouri (Vanderbilt University), and Hamed Pirsivavash (University of California Davis)</i>	
Reinforce Data, Multiply Impact: Improved Model Accuracy and Robustness with Dataset Reinforcement	16986
<i>Fartash Faghri (Apple), Hadi Pouransari (Apple), Sachin Mehta (University of Washington), Mehrdad Farajtabar (Apple), Ali Farhadi (University of Washington, Allen Institute for AI, Apple), Mohammad Rastegari (University of Washington), and Oncel Tuzel (Apple)</i>	
A Fast Unified System for 3D Object Detection and Tracking	16998
<i>Thomas Heitzinger (Technische Universität Wien) and Martin Kampel (Vienna University of Technology, Computer Vision Lab)</i>	

Estimator Meets Equilibrium Perspective: A Rectified Straight Through Estimator for Binary Neural Networks Training	17009
<i>Xiao-Ming Wu (Sun Yat-Sen University), Dian Zheng (Sun Yat-Sen University), Zuhao Liu (Sun Yat-sen University), and Wei-Shi Zheng (Sun Yat-sen University, China)</i>	
I-ViT: Integer-only Quantization for Efficient Vision Transformer Inference	17019
<i>Zhikai Li (Institute of Automation, Chinese Academy of Sciences; School of Artificial Intelligence, University of Chinese Academy of Sciences) and Qingyi Gu (Institute of Automation, Chinese Academy of Sciences)</i>	
EMQ: Evolving Training-free Proxies for Automated Mixed Precision Quantization	17030
<i>Peijie Dong (Data Science and Analytics, The Hong Kong University of Science and Technology, (Guangzhou), Lujun Li (The Hong Kong University of Science and Technology), Zimian Wei (School of Computer Science, National University of Defense Technology), Xin Niu (NUDT), Zhiliang Tian (National University of Defense Technology), and Hengyue Pan (National University of Defense Technology)</i>	
Local or Global: Selective Knowledge Assimilation for Federated Learning with Limited Labels	17041
<i>Yae Jee Cho (Carnegie Mellon University), Gauri Joshi (Carnegie Mellon University), and Dimitrios Dimitriadis (Microsoft Research)</i>	
DataDAM: Efficient Dataset Distillation with Attention Matching	17051
<i>Ahmad Sajedi (University of Toronto), Samir Khaki (University of Toronto), Ehsan Amjadian (University of Waterloo), Lucy Z. Liu (RBC), Yuri A. Lawryshyn (University of Toronto), and Konstantinos N. Plataniotis (UofT)</i>	
SAFE: Machine Unlearning With Shard Graphs	17062
<i>Yonatan Dukler (AWS AI Labs), Benjamin Bowman (UCLA), Alessandro Achille (AWS), Aditya Golatkar (University of California, Los Angeles), Ashwin Swaminathan (Qualcomm, USA), and Stefano Soatto (AWS Amazon ML)</i>	
ResQ: Residual Quantization for Video Perception	17073
<i>Davide Abati (Qualcomm AI Research), Haitam Ben Yahia (Qualcomm AI Research), Markus Nagel (Qualcomm), and Amirhossein Habibian (Qualcomm AI Research)</i>	
Efficient Computation Sharing for Multi-Task Visual Scene Understanding	17084
<i>Sara Shoouri (University of Michigan, Ann Arbor), Mingyu Yang (University of Michigan), Zichen Fan (University of Michigan), and Hun-Seok Kim (Nil)</i>	
Essential Matrix Estimation using Convex Relaxations in Orthogonal Space	17096
<i>Arman Karimian (Boston University) and Roberto Tron (Boston University)</i>	

TripLe: Revisiting Pretrained Model Reuse and Progressive Learning for Efficient Vision Transformer Scaling and Searching	17107
<i>Cheng Fu (University of California, San Diego), Hanxian Huang (UC San Diego), Zixuan Jiang (The University of Texas at Austin), Yun Ni (Google), Lifeng Nai (Google), Gang Wu (Google), Liqun Cheng (Google), Yanqi Zhou (Google), Sheng Li (Google), Andrew Li (Google), and Jishen Zhao (UCSD)</i>	
DiffRate : Differentiable Compression Rate for Efficient Vision Transformers	17118
<i>Mengzhao Chen (Xiamen University), Wenqi Shao (The Chinese University of HongKong), Peng Xu (The University of Hong Kong), Mingbao Lin (Xiamen University, China), Kaipeng Zhang (Shanghai AI Laboratory), Fei Chao (Xiamen University), Rongrong Ji (Xiamen University, China), Yu Qiao (Shanghai AI Laboraotry), and Ping Luo (The University of Hong Kong)</i>	
Bridging Cross-task Protocol Inconsistency for Distillation in Dense Object Detection	17129
<i>Longrong Yang (Zhejiang University), Xianpan Zhou (Zhejiang University), Xuewei Li (Zhejiang University), Liang Qiao (Zhejiang University & Hikvision Research Institute), Zheyang Li (Hikvision Research Institute), Zitwei Yang (Hikvision Research Institute), Gaoang Wang (Zhejiang University), and Xi Li (Zhejiang University)</i>	
From Knowledge Distillation to Self-Knowledge Distillation: A Unified Approach with Normalized Loss and Customized Soft Labels	17139
<i>Zhendong Yang (Graduate school at ShenZhenTsinghua university), Ailing Zeng (International Digital Economy Academy), Zhe Li (Institute of Automation, Chinese Academy of Sciences), Tianke Zhang (Tsinghua University), Chun Yuan (Graduate school at ShenZhenTsinghua university), and Yu Li (International Digital Economy Academy)</i>	
Efficient 3D Semantic Segmentation with Superpoint Transformer	17149
<i>Damien Robert (IGN), Hugo Raguet (INSA Centre Val-de-Loire Université de Tours), and Loic Landrieu (IGN)</i>	
Dataset Quantization	17159
<i>Daquan Zhou (National University of Singapore), Kai Wang (National University of Singapore), Jianyang Gu (College of Control Science and Engineering, Zhejiang University), Xiangyu Peng (National University of Singapore), Dongze Lian (National University of Singapore), Yifan Zhang (National University of Singapore), Yang You (National University of Singapore), and Jiashi Feng (ByteDance)</i>	
Revisiting the Parameter Efficiency of Adapters from the Perspective of Precision Redundancy	17171
<i>Shibo Jie (Peking University), Haoqing Wang (Peking University), and Zhi-Hong Deng (Peking University)</i>	
RepQ-ViT: Scale Reparameterization for Post-Training Quantization of Vision Transformers	17181
<i>Zhikai Li (Institute of Automation, Chinese Academy of Sciences, School of Artificial Intelligence, University of Chinese Academy of Sciences), Junrui Xiao (Institute of Automation, Chinese Academy of Sciences, School of Artificial Intelligence, University of Chinese Academy of Sciences), Lianwei Yang (Institute of Automation, Chinese Academy of Sciences), and Qingyi Gu (Institute of Automation, Chinese Academy of Sciences)</i>	

Semantically Structured Image Compression via Irregular Group-Based Decoupling	17191
<i>Ruoyu Feng (University of Science and Technology of China), Yixin Gao (University of Science and Technology of China), Xin Jin (Eastern Institute for Advanced Study), Runsen Feng (University of Science and Technology of China), and Zhibo Chen (University of Science and Technology of China)</i>	
SeiT: Storage-Efficient Vision Training with Tokens Using 1% of Pixel Storage	17202
<i>Song Park (NAVER AI Lab), Sanghyuk Chun (NAVER AI Lab), Byeongho Heo (NAVER AI Lab), Wonjae Kim (NAVER AI Lab), and Sangdoon Yun (NAVER AI Lab)</i>	
SMMix: Self-Motivated Image Mixing for Vision Transformers	17214
<i>Mengzhao Chen (Xiamen University), Mingbao Lin (Xiamen University), Zhihang Lin (Xiamen University), Yuxin Zhang (Xiamen University), Fei Chao (Xiamen University), and Rongrong Ji (Xiamen University)</i>	
Multi-Label Knowledge Distillation	17225
<i>Penghui Yang (Nanjing University of Aeronautics and Astronautics), Ming-Kun Xie (Nanjing University of Aeronautics and Astronautics), Chen-Chen Zong (Nanjing University of Aeronautics and Astronautics), Lei Feng (Nanyang Technological University), Gang Niu (RIKEN), Masashi Sugiyama (RIKEN/The University of Tokyo), and Sheng-Jun Huang (Nanjing University of Aeronautics and Astronautics)</i>	
UGC: Unified GAN Compression for Efficient Image-to-Image Translation	17235
<i>Yuxi Ren (ByteDance Inc), Jie Wu (ByteDance Inc), Peng Zhang (Bytedance), Manlin Zhang (Sun Yat-sen University), Xuefeng Xiao (ByteDance Inc), Qian He (Bytedance), Rui Wang (ByteDance), Min Zheng (ByteDance), and Xin Pan (ByteDance)</i>	
MotionDeltaCNN: Sparse CNN Inference of Frame Differences in Moving Camera Videos with Spherical Buffers and Padded Convolutions	17246
<i>Mathias Parger (Graz University of Technology), Chengcheng Tang (Facebook Reality Labs), Thomas Neff (Graz University of Technology), Christopher D. Twigg (Meta), Cem Keskin (Facebook), Robert Wang (Facebook Reality Labs), and Markus Steinberger (Graz University of Technology)</i>	
EfficientViT: Lightweight Multi-Scale Attention for High-Resolution Dense Prediction	17256
<i>Han Cai (Massachusetts Institute of Technology), Junyan Li (Zhejiang University), Muyan Hu (Tsinghua University), Chuang Gan (MIT-IBM Watson AI Lab), and Song Han (MIT)</i>	
DREAM: Efficient Dataset Distillation by Representative Matching	17268
<i>Yanqing Liu (Zhejiang University), Jianyang Gu (College of Control Science and Engineering, Zhejiang University), Kai Wang (National University of Singapore), Zheng Zhu (Tsinghua University), Wei Jiang (Department of Control Science and Engineering, Zhejiang University), and Yang You (National University of Singapore)</i>	
INSTA-BNN: Binary Neural Network with INSTAnce-aware Threshold	17279
<i>Changhun Lee (POSTECH), Hyungjun Kim (SqueezeBits Inc.), Eunhyeok Park (POSTECH), and Jae-Joon Kim (Seoul National University)</i>	

Deep Incubation: Training Large Models by Divide-and-Conquering	17289
<i>Zanlin Ni (Tsinghua University), Yulin Wang (Tsinghua University), Jiangwei Yu (Tsinghua University), Haojun Jiang (Tsinghua University), Yue Cao (Beijing Academy of Artificial Intelligence), and Gao Huang (Tsinghua)</i>	
AdaMV-MoE: Adaptive Multi-Task Vision Mixture-of-Experts	17300
<i>Tianlong Chen (University of Texas at Austin), Xuxi Chen (University of Texas at Austin), Xianzhi Du (Google Brain), Abdullah Rashwan (University of Waterloo), Fan Yang (Google), Huizhong Chen (Google), Zhangyang Wang (University of Texas at Austin), and Yeqing Li (Google)</i>	
Overcoming Forgetting Catastrophe in Quantization-Aware Training	17312
<i>Ting-An Chen (National Taiwan University), De-Nian Yang (Academia Sinica), and Ming-Syan Chen (National Taiwan University)</i>	
Window-Based Early-Exit Cascades for Uncertainty Estimation: When Deep Ensembles are More Efficient than Single Models	17322
<i>Guoxuan Xia (Imperial College London) and Christos-Savvas Bouganis (Imperial College London)</i>	
ORC: Network Group-based Knowledge Distillation using Online Role Change	17335
<i>Junyong Choi (Hyundai Motor Company), Hyeon Cho (Ajou University), Seokhwa Cheung (Ajou University), and Wonjun Hwang (Ajou University)</i>	
RMP-Loss: Regularizing Membrane Potential Distribution for Spiking Neural Networks	17345
<i>Yufei Guo (Intelligent Science Technology Academy of CASIC), Xiaode Liu (Intelligent Science & Technology Academy of CASIC), Yuanpei Chen (X LAB, The Second Academy of CASIC, Beijing), Liwen Zhang (X Lab, the Second Academy of CASIC, Beijing), Weihang Peng (buaa), Yuhan Zhang (Intelligent Science & Technology Academy of CASIC), Xuhui Huang (Intelligent Science and Technology Academy of CASIC), and Zhe Ma (Xlab, the Second Academy of CASIC, Beijing)</i>	
Structural Alignment for Network Pruning through Partial Regularization	17356
<i>Shangqian Gao (University of Pittsburgh), Zeyu Zhang (University of Arizona), Yanfu Zhang (University of Pittsburgh), Feihu Huang (Nanjing University of Aeronautics and Astronautics), and Heng Huang (University of Maryland at College Park)</i>	
Automated Knowledge Distillation via Monte Carlo Tree Search	17367
<i>Lujun Li (The Hong Kong University of Science and Technology), Peijie Dong (Data Science and Analytics, The Hong Kong University of Science and Technology, (Guangzhou), Zimian Wei (School of Computer Science, National University of Defense Technology), and Yang Ya (City University of Hong Kong)</i>	
SwiftFormer: Efficient Additive Attention for Transformer-based Real-time Mobile Vision Applications	17379
<i>Abdelrahman Shaker (MBZUAI), Muhammad Maaz (MBZUAI), Hanoona Rasheed (MBZUAI), Salman Khan (MBZUAI), Ming-Hsuan Yang (University of California at Merced), and Fahad Shahbaz Khan (MBZUAI)</i>	
Causal-DFQ: Causality Guided Data-Free Network Quantization	17391
<i>Yuzhang Shang (Illinois Institute of Technology), Bingxin Xu (Illinois Institute of Technology), Gaowen Liu (Cisco Research), Ramana Rao Kompella (Cisco), and Yan Yan (Illinois Institute of Technology)</i>	

Efficient Joint Optimization of Layer-Adaptive Weight Pruning in Deep Neural Networks	17401
<i>Kaixin Xu (institute of infocomm research), Zhe Wang (Institute for Infocomm Research, Singapore), Xue Geng (I2R, A*STAR), Min Wu (Institute for Infocomm Research, A*STAR, Singapore), Xiaoli Li (Institute for Infocomm Research, A*STAR, Singapore/Nanyang Technological University), and Weisi Lin (Nanyang Technological University, Singapore)</i>	
Automatic Network Pruning via Hilbert-Schmidt Independence Criterion Lasso under Information Bottleneck Principle	17412
<i>Song Guo (Xiamen University), Lei Zhang (Xiamen University), Xiawu Zheng (Peng Cheng Laboratory), Yan Wang (Samsara), Yuchao Li (Alibaba Group), Fei Chao (Xiamen University), Chenglin Wu (fuzhi.ai), Shengchuan Zhang (Xiamen University), and Rongrong Ji (Xiamen University, China)</i>	
Distribution Shift Matters for Knowledge Distillation with Webly Collected Images	17424
<i>Jialiang Tang (Nanjing University of Science and Technology), Shuo Chen (RIKEN), Gang Niu (RIKEN), Masashi Sugiyama (RIKEN/The University of Tokyo), and Chen Gong (Nanjing University of Science and Technology)</i>	
FastRecon: Few-shot Industrial Anomaly Detection via Fast Feature Reconstruction	17435
<i>Zheng Fang (XJTLU), Xiaoyang Wang (Univerisity of Liverpool), Haocheng Li (Dinnar Automation Technology), Jiejie Liu (Xi'an Jiaotong-liverpool University), Qiugui Hu (Dinnar Automation Technology), and Jimin Xiao (Xi'an Jiaotong-Liverpool University)</i>	
E ² VPT: An Effective and Efficient Approach for Visual Prompt Tuning	17445
<i>Cheng Han (Rochester Institute of Technology), Qifan Wang (Meta AI), Yiming Cui (University of Florida), Zhiwen Cao (Purdue University), Wenguan Wang (Zhejiang University), Siyuan Qi (UCLA), and Dongfang Liu (Rochester Institute of Technology)</i>	
Bridging Vision and Language Encoders: Parameter-Efficient Tuning for Referring Image Segmentation	17457
<i>Zunnan Xu (Sun Yat-sen University), Zhihong Chen (Shenzhen Research Institute of Big Data, The Chinese University of Hong Kong, Shenzhen), Yong Zhang (Tencent AI Lab), Yibing Song (Fudan University), Xiang Wan (Shenzhen Research Institute of Big Data, the Chinese University of Hong Kong, Shenzhen), and Guanbin Li (Sun Yat-sen University)</i>	
SHACIRA: Scalable HAsH-grid Compression for Implicit Neural Representations	17467
<i>Sharath Girish (University of Maryland, College Park), Abhinav Shrivastava (University of Maryland), and Kamal Gupta (University of Maryland)</i>	
Efficient Deep Space Filling Curve	17479
<i>Wanli Chen (CUHK), Xufeng Yao (Chinese University of HongKong), Xinyun Zhang (The Chinese University of Hong Kong), and Bei Yu (CUHK)</i>	
Q-Diffusion: Quantizing Diffusion Models	17489
<i>Xiuyu Li (UC Berkeley), Yijiang Liu (Nanjing University), Long Lian (UC Berkeley), Huanrui Yang (UC Berkeley), Zhen Dong (UC Berkeley), Daniel Kang (UIUC), Shanghang Zhang (Peking University), and Kurt Keutzer (EECS, UC Berkeley)</i>	

Lossy and Lossless (L2) Post-training Model Size Compression 17500
Yumeng Shi (Beihang University), Shihao Bai (sensetime), Xiuying Wei (Beihang University), Ruihao Gong (SenseTime), and Jianlei Yang (Beihang University)

Robustifying Token Attention for Vision Transformers 17511
Yong Guo (Max Planck Institute for Informatics), David Stutz (Max Planck Institute for Informatics), and Bernt Schiele (MPI Informatics)

3D from Multi-view and Sensors 2

Strivec: Sparse Tri-Vector Radiance Fields 17523
Quankai Gao (University of Southern California), Qiangeng Xu (USC), Hao Su (UCSD), Ulrich Neumann (USC), and Zexiang Xu (Adobe Research)

LDP-Feat: Image Features with Local Differential Privacy 17534
Francesco Pittaluga (NEC Laboratories America) and Bingbing Zhuang (NEC Labs)

SparseFusion: Fusing Multi-Modal Sparse Representations for Multi-Sensor 3D Object Detection 17545
Yichen Xie (University of California, Berkeley), Chenfeng Xu (UC Berkeley), Marie-Julie Rakotosaona (Google), Patrick Rim (California Institute of Technology), Federico Tombari (Google, TU Munich), Kurt Keutzer (EECS, UC Berkeley), Masayoshi Tomizuka (MSC Lab), and Wei Zhan (University of California, Berkeley)

Strata-NeRF : Neural Radiance Fields for Stratified Scenes 17557
Ankit Dhiman (Indian Institute of Science, Bangalore), R Srinath (IISc), Harsh Rangwani (Indian Institute of Science), Rishubh Parihar (Indian Institute of Science, Bangalore), Lokesh R Boregowda (Samsung Research Institute Bangalore), Srinath Sridhar (Brown University), and R Venkatesh Babu (Indian Institute of Science)

CRN: Camera Radar Net for Accurate, Robust, Efficient 3D Perception 17569
Youngseok Kim (Korea Advanced Institute of Science and Technology), Juyeon Shin (KAIST), Sanmin Kim (Korea Advanced Institute of Science and Technology), In-Jae Lee (Korea Advanced Institute of Science and Technology), Jun Won Choi (Hanyang University), and Dongsuk Kum (Korea Advanced Institute of Science and Technology)

LightGlue: Local Feature Matching at Light Speed 17581
Philipp Lindenberger (ETH Zurich), Paul-Edouard Sarlin (ETH Zurich), and Marc Pollefeys (ETH Zurich / Microsoft)

ExBluRF: Efficient Radiance Fields for Extreme Motion Blurred Images 17593
Dongwoo Lee (Seoul National University), Jeongtaek Oh (Seoul National University), Jaesung Rim (POSTECH), Sunghyun Cho (POSTECH), and Kyoung Mu Lee (Seoul National University)

Generalized Differentiable RANSAC 17603
Tong Wei (Czech Technical University in Prague), Yash Patel (Czech Technical University in Prague), Alexander Shekhovtsov (Czech Technical University in Prague, Czech Republic), Jiri Matas (Czech Technical University, Prague), and Daniel Barath (ETH Zürich)

Constraining Depth Map Geometry for Multi-View Stereo: A Dual-Depth Approach with Saddle-shaped Depth Cells	17615
<i>Xinyi Ye (Huazhong University of Science and Technology), Weiyue Zhao (Huazhong University of Science and Technology), Tianqi Liu (Huazhong University of Science and Technology), Zihao Huang (Huazhong University of Science and Technology), Zhiguo Cao (Huazhong Univ. of Sci.&Tech.), and Xin Li (University at Albany)</i>	
Total-Recon: Deformable Scene Reconstruction for Embodied View Synthesis	17625
<i>Chonghyuk Song (Carnegie Mellon University), Gengshan Yang (Carnegie Mellon University), Kangle Deng (Carnegie Mellon University), Jun-Yan Zhu (Carnegie Mellon University), and Deva Ramanan (CMU)</i>	
Seal-3D: Interactive Pixel-Level Editing for Neural Radiance Fields	17637
<i>Xiangyu Wang (Zhejiang University), Jingsen Zhu (Zhejiang University), Qi Ye (Zhejiang Univeristy), Yuchi Huo (Zhejiang Lab; Zhejiang University), Yunlong Ran (Zhejiang University), Zhihua Zhong (Zhejiang University), and Jiming Chen (Zhejiang University)</i>	
PointMBF: A Multi-scale Bidirectional Fusion Network for Unsupervised RGB-D Point Cloud Registration	17648
<i>Mingzhi Yuan (Fudan University), Kexue Fu (Digital Medical Research Center, School of Basic Medical Science, Fudan University), Zhihao Li (Fudan), Yucong Meng (Fudan university), and Manning Wang (Fudan University)</i>	
PARF: Primitive-Aware Radiance Fusion for Indoor Scene Novel View Synthesis	17660
<i>Haiyang Ying (Tsinghua University), Baowei Jiang (QiYuan Lab), Jinzhi Zhang (Tsinghua University), Di Xu (Huawei Cloud), Tao Yu (Tsinghua University), Qionghai Dai (Tsinghua University), and Lu Fang (Tsinghua University)</i>	
Rethinking Point Cloud Registration as Masking and Reconstruction	17671
<i>Guangyan Chen (Beijing Institute of Technology), Meiling Wang (Beijing Institute of Technology), Li Yuan (Peking University), Yi Yang (Beijing Institute of Technology), and Yufeng Yue (Beijing Institute of Technology)</i>	
Ada3D : Exploiting the Spatial Redundancy with Adaptive Inference for Efficient 3D Object Detection	17682
<i>Tianchen Zhao (Tsinghua University; Novauto), Xuefei Ning (Tsinghua University), Ke Hong (Tsinghua University), Zhongyuan Qiu (Novauto), Pu Lu (Tsinghua University), Yali Zhao (Novauto), Linfeng Zhang (Tsinghua University), Lipu Zhou (Meituan), Guohao Dai (Shanghai Jiao Tong University), Huazhong Yang (Tsinghua University), and Yu Wang (Tsinghua University)</i>	
Delicate Textured Mesh Recovery from NeRF via Adaptive Surface Refinement	17693
<i>Jiaxiang Tang (Peking University), Hang Zhou (Baidu Inc.), Xiaokang Chen (Peking University), Tianshu Hu (Baidu), Errui Ding (Baidu Inc.), Jingdong Wang (Baidu), and Gang Zeng (Peking University)</i>	
CVRecon: Rethinking 3D Geometric Feature Learning For Neural Reconstruction	17704
<i>Ziyue Feng (Clemson University), Liang Yang (Apple Inc), Pengsheng Guo (Carnegie Mellon University), and Bing Li (Clemson University)</i>	

RICO: Regularizing the Unobservable for Indoor Compositional Reconstruction	17715
<i>Zizhang Li (Zhejiang University), Xiaoyang Lyu (The University of Hong Kong), Yuanyuan Ding (Zhejiang University), Mengmeng Wang (Zhejiang University), Yiyi Liao (Zhejiang University), and Yong Liu (Zhejiang University)</i>	
Multiscale Representation for Real-Time Anti-Aliasing Neural Rendering	17726
<i>Dongting Hu (The University of Melbourne), Zhenkai Zhang (University of Melbourne), Tingbo Hou (Google Research), Tongliang Liu (The University of Sydney), Huan Fu (University of Sydney), and Mingming Gong (University of Melbourne)</i>	
ELFNet: Evidential Local-global Fusion for Stereo Matching	17738
<i>Jieming Lou (Institute for Infocomm Research, A*STAR; National University of Singapore), Weide Liu (Institute for Infocomm Research, A*STAR), Zhuo Chen (Institute for Infocomm Research, A*STAR), Fayao Liu (Institute for Infocomm Research, A*STAR), and Jun Cheng (Institute for Infocomm Research, A*STAR)</i>	
GaPro: Box-Supervised 3D Point Cloud Instance Segmentation Using Gaussian Processes as Pseudo Labelers	17748
<i>Tuan Duc Ngo (VinAI Research), Binh-Son Hua (Trinity College Dublin), and Khoi Nguyen (VinAI Research)</i>	
Multi-body Depth and Camera Pose Estimation from Multiple Views	17758
<i>Andrea Porfiri Dal Cin (Politecnico di Milano), Giacomo Boracchi (Politecnico di Milano), and Luca Magri (Politecnico di Milano)</i>	
Reference-guided Controllable Inpainting of Neural Radiance Fields	17769
<i>Ashkan Mirzaei (University of Toronto), Tristan Aumentado-Armstrong (University of Toronto), Marcus A. Brubaker (York University), Jonathan Kelly (University of Toronto), Alex Levinshtein (Samsung AI Centre Toronto), Konstantinos G. Derpanis (York University), and Igor Gilitschenski (University of Toronto)</i>	
Retro-FPN: Retrospective Feature Pyramid Network for Point Cloud Semantic Segmentation	17780
<i>Peng Xiang (Tsinghua University), Xin Wen (JD.com), Yu-Shen Liu (Tsinghua University), Hui Zhang (Tsinghua University), Yi Fang (New York University), and Zhizhong Han (Wayne State University)</i>	
GeoMIM: Towards Better 3D Knowledge Transfer via Masked Image Modeling for Multi-view 3D Understanding	17793
<i>Jihao Liu (The Chinese University of Hong Kong), Tai Wang (The Chinese University of Hong Kong), Boxiao Liu (Institute of Computing Technology, Chinese Academy of Sciences), Qihang Zhang (The Chinese University of Hong Kong), Yu Liu (SenseTime Group LTD), and Hongsheng Li (The Chinese University of Hong Kong)</i>	
OpenOccupancy: A Large Scale Benchmark for Surrounding Semantic Occupancy Perception ...	17804
<i>Xiaofeng Wang (Institute of Automation, Chinese Academy of Sciences, School of Artificial Intelligence, University of Chinese Academy of Sciences), Zheng Zhu (Tsinghua University), Wenbo Xu (PhiGent Robotics), Yunpeng Zhang (PhiGent Robotics), Yi Wei (Tsinghua University), Xu Chi (Phigent Robotics), Yun Ye (XForwardAI Technology Co., Ltd, Beijing, China), Dalong Du (PhiGent Technology Co., Ltd), Jiwen Lu (Tsinghua University), and Xingang Wang (Institute of Automation, CAS)</i>	

Surface Normal Clustering for Implicit Representation of Manhattan Scenes	17814
<i>Nikola Popović (ETH Zürich), Danda Pani Paudel (ETH Zürich, INSAIT Sofia), and Luc Van Gool (ETH Zurich)</i>	
Spacetime Surface Regularization for Neural Dynamic Scene Reconstruction	17825
<i>Jaesung Choe (KAIST), Christopher Choy (Nvidia), Jaesik Park (POSTECH), In So Kweon (KAIST), and Animashree Anandkumar (Caltech)</i>	
LDL: Line Distance Functions for Panoramic Localization	17836
<i>Junho Kim (Seoul National University), Changwoon Choi (Seoul National University), Hojun Jang (Seoul National University), and Young Min Kim (Seoul National University)</i>	
Learning Neural Implicit Surfaces with Object-Aware Radiance Fields	17847
<i>Yiheng Zhang (JD Explore Academy), Zhaofan Qiu (JD.com), Yingwei Pan (USTC), Ting Yao (JD AI Research), and Tao Mei (AI Research of JD.com)</i>	
MonoNeRF: Learning a Generalizable Dynamic Radiance Field from Monocular Videos	17857
<i>Fengrui Tian (Xi'an Jiaotong Univeristy), Shaoyi Du (Xi'an Jiaotong Unviersity), and Yueqi Duan (Tsinghua University)</i>	
Neural Radiance Field with LiDAR maps	17868
<i>MingFang Chang (Carnegie Mellon University), Akash Sharma (Carnegie Mellon University), Michael Kaess (Carnegie Mellon University), and Simon Lucey (University of Adelaide)</i>	
Deformable Model-Driven Neural Rendering for High-Fidelity 3D Reconstruction of Human Heads Under Low-View Settings	17878
<i>Baixin Xu (Nanyang Technological University), Jiarui Zhang (Peking University), Kwan-Yee Lin (SenseTime Research), Chen Qian (SenseTime), and Ying He (Nanyang Technological University)</i>	
DeLiRa: Self-Supervised Depth, Light, and Radiance Fields	17889
<i>Vitor Guizilini (Toyota Research Institute), Igor Vasiljevic (Toyota Research Institute), Jiading Fang (Toyota Technological Institute at Chicago), Rares Ambrus (Toyota Research Institute), Sergey Zakharov (Toyota Research Institute), Vincent Sitzmann (MIT), and Adrien Gaidon (Toyota Research Institute)</i>	
ATT3D: Amortized Text-to-3D Object Synthesis	17900
<i>Jonathan Lorraine (University of Toronto), Kevin Xie (University of Toronto), Xiaohui Zeng (University of Toronto), Chen-Hsuan Lin (NVIDIA Research), Towaki Takikawa (University of Toronto), Nicholas Sharp (NVIDIA), Tsung-Yi Lin (Nvidia), Ming-Yu Liu (NVIDIA), Sanja Fidler (University of Toronto, NVIDIA), and James Lucas (University of Toronto and Vector Institute)</i>	
ScatterNeRF: Seeing Through Fog with Physically-Based Inverse Neural Rendering	17911
<i>Andrea Ramazzina (Mercedes-Benz AG), Mario Bijelic (Princeton University), Stefanie Walz (Mercedes-Benz AG), Alessandro Sanvito (Mercedes-Benz AG), Dominik Scheuble (Mercedes-Benz AG), and Felix Heide (Princeton / Algolux)</i>	

CroCo v2: Improved Cross-view Completion Pre-training for Stereo Matching and Optical Flow	17923
<i>Philippe Weinzaepfel (NAVER LABS Europe), Thomas Lucas (Naver), Vincent Leroy (Naver Labs Europe), Yohann Cabon (Naver Labs Europe), Vaibhav Arora (Université Paris Sud), Romain Brégier (Naver Labs Europe), Gabriela Csurka (Naver Labs Europe), Leonid Antsfeld (Naver Labs Europe), Boris Chidlovskii (Naver Labs Europe), and Jerome Revaud (Naver Labs Europe)</i>	
Guiding Local Feature Matching with Surface Curvature	17935
<i>Shuzhe Wang (Aalto University), Juho Kannala (Aalto University, Finland), Marc Pollefeys (ETH Zurich / Microsoft), and Daniel Barath (ETH Zürich)</i>	
NaviNeRF: NeRF-based 3D Representation Disentanglement by Latent Semantic Navigation	17946
<i>Baao Xie (EIT Institute for Advanced Study), Bohan Li (Shanghai Jiao Tong University), Zequn Zhang (Northwest Normal University), Junting Dong (Zhejiang University), Xin Jin (Eastern Institute for Advanced Study), Jingyu Yang (Tianjin University), and Wenjun Zeng (EIT Institute for Advanced Study)</i>	
Efficient LiDAR Point Cloud Oversegmentation Network	17957
<i>Le Hui (Nanjing University of Science and Technology), Linghua Tang (Nanjing University of Science and Technology), Yuchao Dai (Northwestern Polytechnical University), Jin Xie (Nanjing University of Science and Technology), and Jian Yang (Nanjing University of Science and Technology)</i>	
Iterative Superquadric Recomposition of 3D Objects from Multiple Views	17967
<i>Stephan Alaniz (University of Tübingen), Massimiliano Mancini (University of Trento), and Zeynep Akata (University of Tübingen)</i>	
S3IM: Stochastic Structural SIMilarity and Its Unreasonable Effectiveness for Neural Fields	17978
<i>Zeke Xie (Baidu Research), Xindi Yang (Baidu Research), Yujie Yang (Baidu Research), Qi Sun (Baidu Research), Yixiang Jiang (Baidu Research Insitution), Haoran Wang (Baidu), Yunfeng Cai (Baidu Research), and Mingming Sun (Baidu Research)</i>	
LiveHand: Real-time and Photorealistic Neural Hand Rendering	17989
<i>Akshay Mundra (MPI Informatics), Mallikarjun B R (Max Planck Institute for Informatics), Jiayi Wang (Max Planck Institut Informatik), Marc Habermann (Max Planck Institute for Informatics), Christian Theobalt (MPI Informatik), and Mohamed Elgharib (Max Planck Institute for Informatics)</i>	
Neural-PBIR Reconstruction of Shape, Material, and Illumination	18000
<i>Cheng Sun (National Tsing Hua University), Guangyan Cai (University of California, Irvine), Zhengqin Li (Meta), Kai Yan (University of California, Irvine), Cheng Zhang (Meta Reality Labs Research), Carl Marshall (Reality Labs Research at Meta), Jia-Bin Huang (University of Maryland College Park), Shuang Zhao (University of California, Irvine), and Zhao Dong (Meta Reality Labs Research)</i>	

Predict to Detect: Prediction-guided 3D Object Detection using Sequential Images	18011
<i>Sanmin Kim (Korea Advanced Institute of Science and Technology), Youngseok Kim (Korea Advanced Institute of Science and Technology), In-Jae Lee (Korea Advanced Institute of Science and Technology), and Dongsuk Kum (Korea Advanced Institute of Science and Technology)</i>	
ObjectFusion: Multi-modal 3D Object Detection with Object-Centric Fusion	18021
<i>Qi Cai (University of Science and Technology of China), Yingwei Pan (USTC), Ting Yao (JD AI Research), Chong-Wah Ngo (Singapore Management University), and Tao Mei (AI Research of JD.com)</i>	
Domain Generalization of 3D Semantic Segmentation in Autonomous Driving	18031
<i>Jules Sanchez (Mines Paris - PSL), Jean-Emmanuel Deschaud (Mines Paris - PSL), and François Goulette (MINES ParisTech)</i>	
When Epipolar Constraint Meets Non-Local Operators in Multi-View Stereo	18042
<i>Tianqi Liu (Huazhong University of Science and Technology), Xinyi Ye (Huazhong University of Science and Technology), Weiyue Zhao (Huazhong University of Science and Technology), Zhiyu Pan (Huazhong Univ. of Sci.&Tech.), Min Shi (Huazhong University of Science and Technology), and Zhiguo Cao (Huazhong Univ. of Sci.&Tech.)</i>	
Hierarchical Point-based Active Learning for Semi-supervised Point Cloud Semantic Segmentation	18052
<i>Zongyi Xu (Chongqing University of Posts and Telecommunications), Bo Yuan (Chongqing University of Posts and Telecommunications), Shanshan Zhao (JD.COM), Qianni Zhang (Queen Mary University of London), and Xinbo Gao (Chongqing University of Posts and Telecommunications)</i>	
UniT3D: A Unified Transformer for 3D Dense Captioning and Visual Grounding	18063
<i>Dave Zhenyu Chen (Technical University of Munich), Ronghang Hu (Facebook), Xinlei Chen (FAIR), Matthias Nießner (Technical University of Munich), and Angel X. Chang (SFU)</i>	
Nerfbusters: Removing Ghostly Artifacts from Casually Captured NeRFs	18074
<i>Frederik Warburg (The Technical University of Denmark), Ethan Weber (UC Berkeley), Matthew Tancik (UC Berkeley), Aleksander Holynski (UC Berkeley / Google Research), and Angjoo Kanazawa (University of California Berkeley)</i>	
Clutter Detection and Removal in 3D Scenes with View-Consistent Inpainting	18085
<i>Fangyin Wei (Princeton University), Thomas Funkhouser (Google Research & Princeton University), and Szymon Rusinkiewicz (Princeton University)</i>	
PG-RCNN: Semantic Surface Point Generation for 3D Object Detection	18096
<i>Inyong Koo (KAIST), Inyoung Lee (KAIST), Se-Ho Kim (KAIST), Hee-Seon Kim (KAIST), Woo-jin Jeon (KAIST), and Changick Kim (KAIST)</i>	

Distributed Bundle Adjustment with Block-Based Sparse Matrix Compression for Super Large Scale Datasets	18106
<i>Maoteng Zheng (China University of Geosciences, (Wuhan), Nengcheng Chen (China University of GeoSciences), Junfeng Zhu (Mirauge3D Technology Inc.), Xiaoru Zeng (Mirauge3D Technology Inc.), Huanbin Qiu (Guangzhou Jiantong Surveying, Mapping And Geoinfo Co., LTD), Yuyao Jiang (China University of Geosciences(Wuhan), Xingyue Lu (China University of Geosciences, (Wuhan), and Hao Qu (Mirauge3D Technology Inc.)</i>	
Adaptive Reordering Sampler with Neurally Guided MAGSAC	18117
<i>Tong Wei (Czech Technical University in Prague), Jiri Matas (Czech Technical University, Prague), and Daniel Barath (ETH Zürich)</i>	
Privacy Preserving Localization via Coordinate Permutations	18128
<i>Linfei Pan (ETHZ), Johannes L. Schönberger (Microsoft), Viktor Larsson (Lund University), and Marc Pollefeys (ETH Zurich / Microsoft)</i>	
DG-Recon: Depth-Guided Neural 3D Scene Reconstruction	18138
<i>Jihong Ju (Qualcomm), Ching-Wei Tseng (Qualcomm), Oleksandr Bailo (Qualcomm), Georgi Dikov (Qualcomm), and Mohsen Ghafoorian (Qualcomm)</i>	
WaveNeRF: Wavelet-based Generalizable Neural Radiance Fields	18149
<i>Muyu Xu (Nanyang Technological University), Fangneng Zhan (Max Planck Institute for Informatics), Jiahui Zhang (Nanyang Technological University), Yingchen Yu (Nanyang Technological University), Xiaoqin Zhang (Wenzhou University), Christian Theobalt (MPI Informatik), Ling Shao (UCAS-Terminus AI Lab, UCAS), and Shijian Lu (Nanyang Technological University)</i>	
TransIFF: An Instance-Level Feature Fusion Framework for Vehicle-Infrastructure Cooperative 3D Detection with Transformers	18159
<i>Ziming Chen (Beihang University), Yifeng Shi (Baidu Inc), and Jinrang Jia (Baidu Inc.)</i>	
Density-invariant Features for Distant Point Cloud Registration	18169
<i>Quan Liu (Shanghai Jiao Tong University), Hongzi Zhu (Shanghai Jiao Tong University), Yunsong Zhou (Shanghai Jiao Tong University), Hongyang Li (Shanghai AI Lab / Shanghai Jiao Tong University), Shan Chang (Donghua University), and Minyi Guo (Shanghai Jiaotong University)</i>	
UMIFormer: Mining the Correlations between Similar Tokens for Multi-View 3D Reconstruction.....	18180
<i>Zhenwei Zhu (Macau University of Science and Technology), Liying Yang (Macau University of Science and Technology), Ning Li (Macau University of Science and Technology), Chaohao Jiang (Macau University of Science and Technology), and Yanyan Liang (Macau University of Science and Technology, Macao SAR, China)</i>	
Neural LiDAR Fields for Novel View Synthesis	18190
<i>Shengyu Huang (ETH Zürich), Zan Gojcic (NVIDIA), Zian Wang (University of Toronto), Francis Williams (NVIDIA), Yoni Kasten (NVIDIA Research), Sanja Fidler (University of Toronto, NVIDIA), Konrad Schindler (ETH Zurich), and Or Litany (NVIDIA)</i>	

Learning Unified Decompositional and Compositional NeRF for Editable Novel View Synthesis	18201
<i>Yuxin Wang (Hong Kong University of Science and Technology), Wayne Wu (Shanghai AI Laboratory), and Dan Xu (The Hong Kong University of Science and Technology)</i>	
Long-Range Grouping Transformer for Multi-View 3D Reconstruction	18211
<i>Liyang Yang (Macau University of Science and Technology), Zhenwei Zhu (Macau University of Science and Technology), Xuxin Lin (Macau University of Science and Technology), Jian Nong (Macau University of Science and Technology), and Yanyan Liang (Macau University of Science and Technology, Macao SAR, China)</i>	
Cross Modal Transformer: Towards Fast and Robust 3D Object Detection	18222
<i>Junjie Yan (Fudan University), Yingfei Liu (Megvii Technology), Jianjian Sun (Tianjin University), Fan Jia (Tianjin University), Shuailin Li (Megvii Technology), Tiancai Wang (Megvii Technology), and Xiangyu Zhang (Megvii Technology)</i>	
KECOR: Kernel Coding Rate Maximization for Active 3D Object Detection	18233
<i>Yadan Luo (University of Queensland), Zhuoxiao Chen (University of Queensland), Zhen Fang (Australian Artificial Intelligence Institute, University of Technology Sydney), Zheng Zhang (Harbin Institute of Technology, Shenzhen), Mahsa Baktashmotlagh (University of Queensland), and Zi Huang (University of Queensland)</i>	
C2F2NeUS: Cascade Cost Frustum Fusion for High Fidelity and Generalizable Neural Surface Reconstruction	18245
<i>Luoyuan Xu (Huazhong University of Science and Technology), Tao Guan (Huazhong University of Science and Technology), Yuesong Wang (Huazhong University of Science and Technology), Wenkai Liu (Huazhong University of Science and Technology), Zhaojie Zeng (Huazhong University of Science and Technology), Junle Wang (Tencent), and Wei Yang (Huazhong University of Science and Technology)</i>	
End-to-end 3D Tracking with Decoupled Queries	18256
<i>Yanwei Li (The Chinese University of Hong Kong), Zhiding Yu (NVIDIA), Jonah Philion (University of Toronto, NVIDIA), Anima Anandkumar (Caltech), Sanja Fidler (University of Toronto, NVIDIA), Jiaya Jia (Chinese University of Hong Kong), and Jose Alvarez (NVIDIA)</i>	
LU-NeRF: Scene and Pose Estimation by Synchronizing Local Unposed NeRFs	18266
<i>Zezhou Cheng (University of Massachusetts, Amherst), Carlos Esteves (Google Research), Varun Jampani (Google), Abhishek Kar (Google), Subhransu Maji (University of Massachusetts, Amherst), and Ameesh Makadia (Google Research)</i>	
GridPull: Towards Scalability in Learning Implicit Representations from 3D Point Clouds	18276
<i>Chao Chen (Tsinghua University), Yu-Shen Liu (Tsinghua University), and Zhizhong Han (Wayne State University)</i>	
Robust e-NeRF: NeRF from Sparse & Noisy Events under Non-Uniform Motion	18289
<i>Weng Fei Low (National University of Singapore) and Gim Hee Lee (National University of Singapore)</i>	

Parameterized Cost Volume for Stereo Matching	18301
<i>Jiayi Zeng (Beijing Institute of Technology), Chengtang Yao (Beijing Institute of Technology), Lidong Yu (Deeproute), Yuwei Wu (Beijing Institute of Technology, (BIT), and Yunde Jia (Beijing Institute of Technology)</i>	
Coordinate Quantized Neural Implicit Representations for Multi-view Reconstruction	18312
<i>Sijia Jiang (Wayne State University), Jing Hua (Wayne State University), and Zhizhong Han (Wayne State University)</i>	
Pixel-Aligned Recurrent Queries for Multi-View 3D Object Detection	18324
<i>Yiming Xie (Northeastern University), Huaizu Jiang (Northeastern University), Georgia Gkioxari (Caltech), and Julian Straub (MIT)</i>	
Optimizing the Placement of Roadside LiDARs for Autonomous Driving	18335
<i>Wentao Jiang (Beihang University), Hao Xiang (University of California, Los Angeles), Xinyu Cai (Shanghai AI Laboratory), Runsheng Xu (University of California, Los Angeles), Jiaqi Ma (University of California, Los Angeles), Yikang Li (Shanghai AI Lab), Gim Hee Lee (National University of Singapore), and Si Liu (Beihang University)</i>	
ActorsNeRF: Animatable Few-shot Human Rendering with Generalizable NeRFs	18345
<i>Jiteng Mu (University of California, San Diego), Shen Sang (UC San Diego; ByteDance Inc.), Nuno Vasconcelos (missing), and Xiaolong Wang (UCSD)</i>	
NeRFrac: Neural Radiance Fields through Refractive Surface	18356
<i>Yifan Zhan (the University of Tokyo), Shohei Nobuhara (Kyoto University), Ko Nishino (Kyoto University), and Yinqiang Zheng (The University of Tokyo)</i>	
CPCM: Contextual Point Cloud Modeling for Weakly-supervised Point Cloud Semantic Segmentation	18367
<i>Lizhao Liu (South China University of Technology), Zhuangwei Zhuang (South China University of Technology), Shangxin Huang (South China University of Technology), Xunlong Xiao (South China University of Technology), Tianhang Xiang (South China University of Technology), Cen Chen (Institute for Infocomm Research, (I2R), Jingdong Wang (Baidu), and Mingkui Tan (South China University of Technology)</i>	
FineRecon: Depth-aware Feed-forward Network for Detailed 3D Reconstruction	18377
<i>Noah Stier (University of California, Santa Barbara), Anurag Ranjan (Apple), Alex Colburn (Apple Inc.), Yajie Yan (Apple), Liang Yang (Apple Inc), Fangchang Ma (Apple Inc.), and Baptiste Angles (Apple)</i>	
Point-SLAM: Dense Neural Point Cloud-based SLAM	18387
<i>Erik Sandström (ETH Zürich), Yue Li (ETH Zürich), Luc Van Gool (ETH Zurich), and Martin R. Oswald (ETH Zurich)</i>	
You Never Get a Second Chance To Make a Good First Impression: Seeding Active Learning for 3D Semantic Segmentation	18399
<i>Nermin Samet (Ecole des Ponts ParisTech), Oriane Siméoni (valeo.ai), Gilles Puy (valeo.ai), Georgy Ponimatkin (Ecole des Ponts ParisTech), Renaud Marlet (Ecole des Ponts ParisTech / Valeo), and Vincent Lepetit (Ecole des Ponts ParisTech)</i>	

Tetra-NeRF: Representing Neural Radiance Fields Using Tetrahedra	18412
<i>Jonas Kulhanek (Czech Technical University in Prague) and Torsten Sattler (Czech Technical University in Prague)</i>	
Active Stereo Without Pattern Projector	18424
<i>Luca Bartolomei (Università di Bologna), Matteo Poggi (University of Bologna), Fabio Tosi (University of Bologna), Andrea Conti (University of Bologna), and Stefano Mattoccia (University of Bologna)</i>	
HOSNeRF: Dynamic Human-Object-Scene Neural Radiance Fields from a Single Video	18437
<i>Jia-Wei Liu (National University of Singapore), Yan-Pei Cao (Tencent), Tianyuan Yang (National University of Singapore), Zhongcong Xu (National University of Singapore), Jussi Keppo (National University of Singapore), Ying Shan (Tencent), Xiaohu Qie (Tencent), and Mike Zheng Shou (National University of Singapore)</i>	
PlankAssembly: Robust 3D Reconstruction from Three Orthographic Views with Learnt Shape Programs	18449
<i>Wentao Hu (Sun Yat-Sen University; Guangdong Key Laboratory of Big Data Analysis and Processing), Jia Zheng (Manycore Tech Inc.), Zixin Zhang (University of Electronic Science and Technology of China), Xiaojun Yuan (University of Electronic Science and Technology of China), Jian Yin (Sun Yat-Sen University; Guangdong Key Laboratory of Big Data Analysis and Processing), and Zihan Zhou (Manycore Tech Inc.)</i>	
Efficient View Synthesis with Neural Radiance Distribution Field	18460
<i>Yushuang Wu (CUHKSZ), Xiao Li (Microsoft Research Asia), Jinglu Wang (Microsoft Research Asia), Xiaoguang Han (Shenzhen Research Institute of Big Data, the Chinese University of Hong Kong, (Shenzhen), Shuguang Cui (The Chinese University of Hong Kong, Shenzhen), and Yan Lu (Microsoft Research Asia)</i>	
Query Refinement Transformer for 3D Instance Segmentation	18470
<i>Jiahao Lu (University of Science and Technology of China), Jiacheng Deng (University of Science and Technology of China), Chuxin Wang (University of Science and Technology of China), Jianfeng He (University of Science and Technology of China), and Tianzhu Zhang (University of Science and Technology of China)</i>	
TrajectoryFormer: 3D Object Tracking Transformer with Predictive Trajectory Hypotheses	18481
<i>Xuesong Chen (CUHK), Shaoshuai Shi (MPI Informatics), Chao Zhang (Samsung Telecommunication Research Institute), Benjin Zhu (CUHK), Qiang Wang (Samsung Research China, Beijing), Ka Chun Cheung (Nvidia), Simon See (NVIDIA AI Tech Centre), and Hongsheng Li (The Chinese University of Hong Kong)</i>	
NerfAcc: Efficient Sampling Accelerates NeRFs	18491
<i>Ruilong Li (UC Berkeley), Hang Gao (UC Berkeley), Matthew Tancik (UC Berkeley), and Angjoo Kanazawa (University of California Berkeley)</i>	
NeTO: Neural Reconstruction of Transparent Objects with Self-Occlusion Aware Refraction-Tracing	18501
<i>Zongcheng Li (WHU), Xiaoxiao Long (The University of Hong Kong), Yusen Wang (WHU), Tuo Cao (Wuhan University), Wenping Wang (Texas A&M University), Fei Luo (Wuhan University), and Chunxia Xiao (Wuhan University)</i>	

Text2Tex: Text-driven Texture Synthesis via Diffusion Models	18512
<i>Dave Zhenyu Chen (Technical University of Munich), Yawar Siddiqui (Technical University of Munich), Hsin-Ying Lee (Snap Inc), Sergey Tulyakov (Snap Inc), and Matthias Nießner (Technical University of Munich)</i>	
Learning Long-Range Information with Dual-Scale Transformers for Indoor Scene Completion .	18523
<i>Ziqi Wang (Wuhan University), Fei Luo (Wuhan University), Xiaoxiao Long (The University of Hong Kong), Wenxiao Zhang (Singapore University of Technology & Design), and Chunxia Xiao (Wuhan University)</i>	
SparseBEV: High-Performance Sparse 3D Object Detection from Multi-Camera Videos	18534
<i>Haisong Liu (Nanjing University), Yao Teng (Nanjing University), Tao Lu (Nanjing University), Haiguang Wang (Nanjing University), and Limin Wang (Nanjing University)</i>	
NeRF-MS: Neural Radiance Fields with Multi-Sequence	18545
<i>Peihao Li (Tsinghua University), Shaohui Wang (Tsinghua University), Chen Yang (Shanghai Jiao Tong University), Bingbing Liu (Huawei Noah's Ark Lab, Canada), Weichao Qiu (Huawei Noah's Ark Lab), and Haoqian Wang (Tsinghua Shenzhen International Graduate School, Tsinghua University)</i>	

Transfer, Low-Shot, Continual, Long-Tail Learning 2

Label-Guided Knowledge Distillation for Continual Semantic Segmentation on 2D Images and 3D Point Clouds	18555
<i>Ze Yang (Nanyang Technological University), Ruibo Li (Nanyang Technological University), Ewan Ling (Hyundai Motor Group Innovation Center in Singapore), Chi Zhang (Nanyang Technological University), Yiming Wang (Nanyang Technological University), Dezhaoh Huang (Hyundai Motor Group Innovation Center in Singapore, (HMGICS), Keng Teck Ma (Hyundai Motor Group Innovation Center in Singapore), Minhoe Hur (AIRS Company, Hyundai Motor Group), and Guosheng Lin (Nanyang Technological University)</i>	
ETran: Energy-Based Transferability Estimation	18567
<i>Mohsen Gholami (University of British Columbia), Mohammad Akbari (Huawei Technologies Canada Co. Ltd.), Xinglu Wang (Simon Fraser University), Behnam Kamranian (Huawei Technologies Canada Co. Ltd.), and Yong Zhang (Huawei Technologies Canada Co., Ltd.)</i>	
PØDA: Prompt-driven Zero-shot Domain Adaptation	18577
<i>Mohammad Fahes (inria), Tuan-Hung Vu (Valeo.ai), Andrei Bursuc (valeo.ai), Patrick Pérez (Valeo.ai), and Raoul de Charette (Inria)</i>	
Local Context-Aware Active Domain Adaptation	18588
<i>Tao Sun (Stony Brook University), Cheng Lu (Xiaopeng), and Haibin Ling (Stony Brook University)</i>	
MRN: Multiplexed Routing Network for Incremental Multilingual Text Recognition	18598
<i>Tianlun Zheng (Fudan University), Zhineng Chen (School of Computer Science, Fudan University), BingChen Huang (Fudan University), Wei Zhang (JD AI Research), and Yu-Gang Jiang (Fudan University)</i>	

Few-Shot Dataset Distillation via Translative Pre-Training	18608
<i>Songhua Liu (National University of Singapore) and Xinchao Wang (National University of Singapore)</i>	
Wasserstein Expansible Variational Autoencoder for Discriminative and Generative Continual Learning	18619
<i>Fei Ye (University of York) and Adrian G. Bors (University of York)</i>	
Tangent Model Composition for Ensembling and Continual Fine-tuning	18630
<i>Tian Yu Liu (UCLA) and Stefano Soatto (UCLA)</i>	
Look at the Neighbor: Distortion-aware Unsupervised Domain Adaptation for Panoramic Semantic Segmentation	18641
<i>Xu Zheng (HKUST), Tianbo Pan (HKUST(GZ), Yunhao Luo (Brown University), and Lin Wang (HKUST)</i>	
Homeomorphism Alignment for Unsupervised Domain Adaptation	18653
<i>Lihua Zhou (University of Electronic Science and Technology of China), Mao Ye (University of Electronic Science and Technology of China), Xiatian Zhu (University of Surrey), Siying Xiao (University of Electronic Science and Technology of China), Xu-Qian Fan (Jinan University), and Ferrante Neri (University of Surrey)</i>	
Knowledge Restore and Transfer for Multi-Label Class-Incremental Learning	18665
<i>Songlin Dong (Xi'an Jiaotong University), Haoyu Luo (Xi'an Jiaotong University), Yuhang He (Xi'an Jiaotong University), Xing Wei (Xi'an Jiaotong University), Jie Cheng (Huawei Technology Co., Ltd.), and Yihong Gong (Xi'an Jiaotong University)</i>	
Unsupervised Domain Adaptation for Training Event-Based Networks Using Contrastive Learning and Uncorrelated Conditioning	18675
<i>Dayuan Jian (University of Southern California) and Mohammad Rostami (University of Southern California)</i>	
A Simple Recipe to Meta-Learn Forward and Backward Transfer	18686
<i>Edoardo Cetin (King's College London), Antonio Carta (Università di Pisa), and Oya Celiktutan (King's College London)</i>	
Dynamic Residual Classifier for Class Incremental Learning	18697
<i>Xiuwei Chen (Sun Yat-sen University) and Xiaobin Chang (Sun Yat-Sen University)</i>	
Concept-wise Fine-tuning Matters in Preventing Negative Transfer	18707
<i>Yunqiao Yang (City University of Hong Kong), Long-Kai Huang (Tencent AI Lab), and Ying Wei (City University of Hong Kong)</i>	
Online Prototype Learning for Online Continual Learning	18718
<i>Yujie Wei (Fudan University), Jiaxin Ye (Fudan University), Zhizhong Huang (Fudan University), Junping Zhang (Fudan University), and Hongming Shan (Fudan University)</i>	
Bidirectional Alignment for Domain Adaptive Detection with Transformers	18729
<i>Liqiang He (Oregon State University), Wei Wang (Amazon Inc), Albert Chen (Amazon), Min Sun (NTHU), Cheng-Hao Kuo (Amazon), and Sinisa Todorovic (Oregon State U)</i>	

Borrowing Knowledge From Pre-trained Language Model: A New Data-efficient Visual Learning Paradigm	18740
<i>Wenxuan Ma (Beijing Institute of Technology), Shuang Li (Beijing Institute of Technology), JinMing Zhang (Beijing Institute of Technology), Chi Harold Liu (Beijing Institute of Technology), Jingxuan Kang (University of Liverpool), Yulin Wang (Tsinghua University), and Gao Huang (Tsinghua)</i>	
CLR: Channel-wise Lightweight Reprogramming for Continual Learning	18752
<i>Yunhao Ge (University of Southern California), Yuecheng Li (University of Southern California), Shuo Ni (University of Southern California), Jiaping Zhao (Google Inc.), Ming-Hsuan Yang (University of California at Merced), and Laurent Itti (University of Southern California)</i>	
Multi-Modal Continual Test-Time Adaptation for 3D Semantic Segmentation	18763
<i>Haozhi Cao (Nanyang Technological University), Yuecong Xu (Institute for Infocomm Research, A*STAR, Singapore), Jianfei Yang (Nanyang Technological University), Pengyu Yin (Nanyang Technological University), Shenghai Yuan (Nanyang Technological University), and Lihua Xie (Nanyang Technological University)</i>	
First Session Adaptation: A Strong Replay-Free Baseline for Class-Incremental Learning	18774
<i>Aristeidis Panos (University of Cambridge), Yuriko Kobe (Cambridge University), Daniel Olmeda Reino (Toyota Motor Europe), Rahaf Aljundi (Toyota Motor Europe), and Richard E. Turner (University of Cambridge)</i>	
Domain Adaptive Few-Shot Open-Set Learning	18785
<i>Debabrata Pal (Indian Institute of Technology, Bombay), Deeptej More (Manipal Institute of Technology), Sai Bhargav (Indian Institute of Technology, Bombay), Dipesh Tamboli (Purdue University), Vaneet Aggarwal (Purdue University), and Biplab Banerjee (Indian Institute of Technology, Bombay)</i>	
Rethinking the Role of Pre-Trained Networks in Source-Free Domain Adaptation	18795
<i>Wenyu Zhang (Institute for Infocomm Research), Li Shen (Institute for Infocomm Research), and Chuan-Sheng Foo (Institute for Infocomm Research, A*STAR)</i>	
Rapid Adaptation in Online Continual Learning: Are We Evaluating It Right?	18806
<i>Hasan Abed Al Kader Al Kader Hammoud (King Abdullah University of Science and Technology), Ameya Prabhu (University of Oxford), Ser-Nam Lim (Facebook AI), Philip H.S. Torr (University of Oxford), Adel Bibi (University of Oxford), and Bernard Ghanem (KAUST)</i>	
Multi-grained Temporal Prototype Learning for Few-shot Video Object Segmentation	18816
<i>Nian Liu (MBZUAI), Kepan Nan (Northwestern Polytechnical University), Wangbo Zhao (National University of Singapore), Yuanwei Liu (Northwestern Polytechnical University), Xiwen Yao (Northwestern Polytechnical University), Salman Khan (MBZUAI), Hisham Cholakkal (MBZUAI), Rao Muhammad Anwer (MBZUAI/AALTO), Junwei Han (Northwestern Polytechnical University), and Fahad Shahbaz Khan (MBZUAI)</i>	
A Low-Shot Object Counting Network With Iterative Prototype Adaptation	18826
<i>Nikola Đukić (Ghent University), Alan Lukežič (University of Ljubljana), Vitjan Zavrtanik (University of Ljubljana), and Matej Kristan (University of Ljubljana)</i>	

Towards Better Robustness against Common Corruptions for Unsupervised Domain Adaptation	18836
<i>Zhiqiang Gao (Xi'an Jiaotong-Liverpool University), Kaizhu Huang (Duke Kunshan University), Rui Zhang (Xi'an Jiaotong-Liverpool University), Dawei Liu (De Montfort University), and Jieming Ma (Xi'an Jiaotong-liverpool University)</i>	
Alleviating Catastrophic Forgetting of Incremental Object Detection via Within-Class and Between-Class Knowledge Distillation	18848
<i>Mengxue Kang (Intelligent Science & Technology Academy), Jinpeng Zhang (Computer Science and Technology College), Jinming Zhang (Xinjiang University), Xiashuang Wang (X lab), Yang Chen (Brainnetome Center, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences;University of Chinese Academy of Sciences;), Zhe Ma (Xlab, the Second Academy of CASIC, Beijing), and Xuhui Huang (Intelligent Science and Technology Academy of CASIC)</i>	
Class-Aware Patch Embedding Adaptation for Few-Shot Image Classification	18859
<i>Fusheng Hao (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences), Fengxiang He (The University of Edinburgh), Liu Liu (The University of Sydney), Fuxiang Wu (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences), Dacheng Tao (The University of Sydney), and Jun Cheng (Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences)</i>	
Order-preserving Consistency Regularization for Domain Adaptation and Generalization	18870
<i>Mengmeng Jing (University of Electronic Science and Technology of China), Xiantong Zhen (United Imaging Healthcare), Jingjing Li (University of Electronic Science and Technology of China), and Cees G. M. Snoek (University of Amsterdam)</i>	
Domain-Specificity Inducing Transformers for Source-Free Domain Adaptation	18882
<i>Sunandini Sanyal (Indian Institute of Science), Ashish Ramayee Asokan (Indian Institute of Science), Suvaansh Bhambri (Indian Institute of Science), Akshay Kulkarni (University of California San Diego), Jogendra Nath Kundu (Indian Institute of Science), and R Venkatesh Babu (Indian Institute of Science)</i>	
Diffusion Model as Representation Learner	18892
<i>Xingyi Yang (National University of Singapore) and Xinchao Wang (National University of Singapore)</i>	
σ -Adaptive Decoupled Prototype for Few-Shot Object Detection	18904
<i>Jinhao Du (Baidu), Shan Zhang (Australian National University), Qiang Chen (Baidu Inc), Haifeng Le (Beijing Union University), Yanpeng Sun (Nanjing University of Science And Technology), Yao Ni (Australian National University), Jian Wang (Baidu), Bin He (baidu), and Jingdong Wang (Baidu)</i>	
Growing a Brain with Sparsity-Inducing Generation for Continual Learning	18915
<i>Hyundong Jin (Chung-Ang University), Gyeong-hyeon Kim (Chung-Ang University), Chanho Ahn (Samsung Advanced Institute of Technology), and Eunwoo Kim (Chung-Ang University)</i>	
DomainAdaptor: A Novel Approach to Test-time Adaptation	18925
<i>Jian Zhang (Nanjing University), Lei Qi (Southeast University), Yinghuan Shi (Nanjing University), and Yang Gao (Nanjing University)</i>	

Reconciling Object-Level and Global-Level Objectives for Long-Tail Detection	18936
<i>Shaoyu Zhang (Institute of Automation, Chinese Academy of Sciences), Chen Chen (The Chinese academy of science), and Silong Peng (The Chinese academy of science)</i>	
Domain Generalization via Balancing Training Difficulty and Model Capability	18947
<i>Xueying Jiang (Nanyang Technological University), Jiaying Huang (Nanyang Technological University), Sheng Jin (Nanyang Technological University), and Shijian Lu (Nanyang Technological University)</i>	
Understanding Hessian Alignment for Domain Generalization	18958
<i>Sobhan Hemati (Noah’s ark Lab), Guojun Zhang (Huawei Noah’s Ark Lab), Amir Estiri (Huawei), and Xi Chen (Huawei Noah’s Ark Lab)</i>	
Vision Transformer Adapters for Generalizable Multitask Learning	18969
<i>Deblina Bhattacharjee (EPFL), Sabine Süsstrunk (EPFL), and Mathieu Salzmann (EPFL)</i>	
Focus on Your Target: A Dual Teacher-Student Framework for Domain-Adaptive Semantic Segmentation	18981
<i>Xinyue Huo (University of Science and Technology of China), Lingxi Xie (Huawei Inc.), Wengang Zhou (University of Science and Technology of China), Houqiang Li (University of Science and Technology of China), and Qi Tian (Huawei Cloud BU)</i>	
Masked Retraining Teacher-Student Framework for Domain Adaptive Object Detection	18993
<i>Zijing Zhao (Peking University), Sitong Wei (Peking University), Qingchao Chen (Peking University), Dehui Li (Tencent intelligent mobility), Yifan Yang (Tencent Intelligent Mobility), Yuxin Peng (Peking University), and Yang Liu (Peking University)</i>	
DandelionNet: Domain Composition with Instance Adaptive Classification for Domain Generalization	19004
<i>Lanqing Hu (Institute of Computing Technology, Chinese Academy of Sciences), Meina Kan (Institute of Computing Technology, Chinese Academy of Sciences), Shiguang Shan (Institute of Computing Technology, Chinese Academy of Sciences), and Xilin Chen (Institute of Computing Technology, Chinese Academy of Sciences)</i>	
CAFA: Class-Aware Feature Alignment for Test-Time Adaptation	19014
<i>Sanghun Jung (University of Washington), Jungsoo Lee (Graduate School of Artificial Intelligence, KAIST), Nanhee Kim (Enssel Inc.), Amirreza Shaban (University of Washington), Byron Boots (University of Washington), and Jaegul Choo (Korea Advanced Institute of Science and Technology)</i>	
Image-Free Classifier Injection for Zero-Shot Classification	19026
<i>Anders Christensen (Technical University of Denmark), Massimiliano Mancini (University of Trento), A. Sophia Koepke (University of Tübingen), Ole Winther (DTU and KU), and Zeynep Akata (University of Tübingen)</i>	
CBA: Improving Online Continual Learning via Continual Bias Adaptor	19036
<i>Quanzhang Wang (Xi’an Jiaotong Univeristy), Renzhen Wang (Xi’an Jiaotong University), Yichen Wu (City University of Hong Kong), Xixi Jia (Xidian University), and Deyu Meng (Xi’an Jiaotong University)</i>	

Masked Autoencoders are Efficient Class Incremental Learners	19047
<i>Jiang-Tian Zhai (Nankai University), Xialei Liu (Nankai University), Andrew D. Bagdanov (University of Florence, Italy), Ke Li (Tencent), and Ming-Ming Cheng (Nankai University)</i>	
DomainDrop: Suppressing Domain-Sensitive Channels for Domain Generalization	19057
<i>Jintao Guo (Nanjing University), Lei Qi (Southeast University), and Yinghuan Shi (Nanjing University)</i>	
Preventing Zero-Shot Transfer Degradation in Continual Learning of Vision-Language Models .	19068
<i>Zangwei Zheng (National University of Singapore), Mingyuan Ma (UC Berkeley), Kai Wang (National University of Singapore), Ziheng Qin (National University of Singapore), Xiangyu Yue (The Chinese University of Hong Kong), and Yang You (National University of Singapore)</i>	
Incremental Generalized Category Discovery	19080
<i>Bingchen Zhao (University of Edinburgh) and Oisin Mac Aodha (University of Edinburgh)</i>	
SLCA: Slow Learner with Classifier Alignment for Continual Learning on a Pre-trained Model...	19091
<i>Gengwei Zhang (University of Technology Sydney), Liyuan Wang (Tsinghua University), Guoliang Kang (Beihang University), Ling Chen (University of Technology, Sydney, Australia), and Yunchao Wei (UTS)</i>	
Efficient Model Personalization in Federated Learning via Client-Specific Prompt Generation	19102
<i>Fu-En Yang (National Taiwan University), Chien-Yi Wang (NVIDIA), and Yu-Chiang Frank Wang (National Taiwan University)</i>	
iDAG: Invariant DAG Searching for Domain Generalization	19112
<i>Zenan Huang (Zhejiang University), Haobo Wang (Zhejiang University), Junbo Zhao (Zhejiang University), and Nenggan Zheng (Zhejiang University)</i>	
SSDA: Secure Source-Free Domain Adaptation	19123
<i>Sabbir Ahmed (Binghamton University, (SUNY)), Abdullah Al Arafat (North Carolina State University), Mamshad Nayeem Rizve (University of Central Florida), Rahim Hossain (STATE UNIVERSITY OF NEW YORK AT BINGHAMTON), Zhishan Guo (North Carolina State University), and Adnan Siraj Rakin (Binghamton University, (SUNY))</i>	
Learning Pseudo-Relations for Cross-domain Semantic Segmentation	19134
<i>Dong Zhao (Xidian University), Shuang Wang (Xidian University), Qi Zang (Xidian University), Dou Quan (School of Artificial Intelligence, Xidian University, China), Xiutiao Ye (Xidian University), Rui Yang (XIDIAN UNIVERSITY), and Licheng Jiao (Xidian University)</i>	
Self-Organizing Pathway Expansion for Non-Exemplar Class-Incremental Learning	19147
<i>Kai Zhu (USTC), Kecheng Zheng (Zhejiang University), Ruili Feng (University of Science and Technology of China), Deli Zhao (Alibaba Group), Yang Cao (University of Science and Technology of China), and Zheng-Jun Zha (University of Science and Technology of China)</i>	

Improved Knowledge Transfer for Semi-Supervised Domain Adaptation via Trico Training Strategy	19157
<i>Ba Hung Ngo (Chonnam National University), Yeon Jeong Chae (Dongguk University), Jung Eun Kwon (Dongguk university), Jae Hyeon Park (Dongguk University), and Sung In Cho (Dongguk University)</i>	
Few-shot Continual Infomax Learning	19167
<i>Ziqi Gu (Nanjing University of Science and Technology), Chunyan Xu (Nanjing University of Science and Technology), Jian Yang (Nanjing University of Science and Technology), and Zhen Cui (Nanjing University of Science and Technology)</i>	
EDAPS: Enhanced Domain-Adaptive Panoptic Segmentation	19177
<i>Suman Saha (ETH Zurich), Lukas Hoyer (ETH Zurich), Anton Obukhov (ETH Zurich), Dengxin Dai (MPI for Informatics), and Luc Van Gool (ETH Zurich)</i>	
Label-Efficient Online Continual Object Detection in Streaming Video	19189
<i>Jay Zhangjie Wu (National University of Singapore), David Junhao Zhang (National University of Singapore), Wynne Hsu (National University of Singapore), Mengmi Zhang (NTU, A*STAR), and Mike Zheng Shou (National University of Singapore)</i>	
Prototypical Kernel Learning and Open-set Foreground Perception for Generalized Few-shot Semantic Segmentation	19199
<i>Kai Huang (Alibaba Group), Feigege Wang (Alibaba), Ye Xi (alibaba), and Yutao Gao (alibaba)</i>	
MSI: Maximize Support-Set Information for Few-Shot Segmentation	19209
<i>Seonghyeon Moon (Rutgers University), Samuel S. Sohn (Rutgers University), Honglu Zhou (NEC Laboratories America, Inc.), Sejong Yoon (The College of New Jersey), Vladimir Paolovic (Rutgers University), Muhammad Haris Khan (Muhammad Bin Zayed University of Artificial Intelligence), and Mubbasir Kapadia (Rutgers)</i>	
AREA: Adaptive Reweighting via Effective Area for Long-Tailed Classification	19220
<i>Xiaohua Chen (Institute of Information Engineering, Chinese Academy of Sciences), Yucan Zhou (Institute of Information Engineering, Chinese Academy of Sciences), Dayan Wu (Institute of Information Engineering, Chinese Academy of Sciences), Chule Yang (Defense Innovation InstituteDII), Bo Li (Institute of Information Engineering, Chinese Academy of Sciences), Qinghua Hu (Tianjin University), and Weiping Wang (Institute of Information Engineering, CAS, China)</i>	
PASTA: Proportional Amplitude Spectrum Training Augmentation for Syn-to-Real Domain Generalization	19231
<i>Prithvijit Chattopadhyay (Georgia Institute of Technology), Kartik Sarangmath (Georgia Institute of Technology), Vivek Vijaykumar (Georgia Institute of Technology), and Judy Hoffman (Georgia Institute of Technology)</i>	
Personalized Semantics Excitation for Federated Image Classification	19244
<i>Haifeng Xia (Tulane University), Kai Li (NEC LABORATORIES AMERICA, INC), and Zhengming Ding (Tulane University)</i>	

Few-Shot Video Classification via Representation Fusion and Promotion Learning	19254
<i>Haifeng Xia (Tulane University), Kai Li (NEC LABORATORIES AMERICA, INC), Martin Renqiang Min (NEC Labs America-Princeton), and Zhengming Ding (Tulane University)</i>	
Segmenting Known Objects and Unseen Unknowns without Prior Knowledge	19264
<i>Stefano Gasperini (Technical University of Munich), Alvaro Marcos-Ramiro (BMW Group), Michael Schmidt (BMW), Nassir Navab (TU Munich, Germany), Benjamin Busam (Technical University of Munich), and Federico Tombari (Google, TU Munich)</i>	

Machine Learning (Other than Deep Learning)

Adaptive Calibrator Ensemble: Navigating Test Set Difficulty in Out-of-Distribution Scenarios	19276
<i>Yuli Zou (The Hong Kong Polytechnic University), Weijian Deng (Australian National University), and Liang Zheng (Australian National University)</i>	
Anchor Structure Regularization Induced Multi-view Subspace Clustering via Enhanced Tensor Rank Minimization	19286
<i>Jintian Ji (School of Computer and Information Technology, Beijing Jiaotong University) and Songhe Feng (School of Computer and Information Technology, Beijing Jiaotong University)</i>	
Meta OOD Learning For Continuously Adaptive OOD Detection	19296
<i>Xinheng Wu (University of Technology Sydney), Jie Lu (University of Technology Sydney), Zhen Fang (Australian Artificial Intelligence Institute, University of Technology Sydney), and Guangquan Zhang (University of Technology Sydney)</i>	
Learning with Diversity: Self-Expanded Equalization for Better Generalized Deep Metric Learning	19308
<i>Jiexi Yan (Xidian University), Zhihui Yin (Xidian University), Erkun Yang (Xidian University), Yanhua Yang (Xidian University), and Heng Huang (University of Maryland at College Park)</i>	
Bold but Cautious: Unlocking the Potential of Personalized Federated Learning through Cautiously Aggressive Collaboration	19318
<i>Xinghao Wu (Beihang University), Xuefeng Liu (Beihang university), Jianwei Niu (beihang university), Guogang Zhu (Beihang University), and Shaojie Tang (University of Texas at Dallas)</i>	
Federated Learning Over Images: Vertical Decompositions and Pre-Trained Backbones Are Difficult to Beat	19328
<i>Erdong Hu (Rice University), Yuxin Tang (Rice University), Anastasios Kyrillidis (Rice University), and Chris Jermaine (Rice University)</i>	
Towards Inadequately Pre-trained Models in Transfer Learning	19340
<i>Andong Deng (University of Central Florida), Xingjian Li (Baidu Research), Di Hu (Renmin University of China), Tianyang Wang (University of Alabama at Birmingham), Haoyi Xiong (Baidu Research), and Cheng-Zhong Xu (University of Macau)</i>	

Reducing Training Time in Cross-Silo Federated Learning Using Multigraph Topology	19352
<i>Tuong Do (AIOZ), Binh X. Nguyen (AIOZ), Vuong Pham (AIOZ), Toan Tran (VinAI Research), Erman Tjiputra (AIOZ), Quang D. Tran (AIOZ), and Anh Nguyen (University of Liverpool)</i>	
Membrane Potential Batch Normalization for Spiking Neural Networks	19363
<i>Yufei Guo (Intelligent Science Technology Academy of CASIC), Yuhan Zhang (Intelligent Science & Technology Academy of CASIC), Yuanpei Chen (X LAB, The Second Academy of CASIC, Beijing), Weihang Peng (buaa), Xiaode Liu (Intelligent Science & Technology Academy of CASIC), Liwen Zhang (X Lab, the Second Academy of CASIC, Beijing), Xuhui Huang (Intelligent Science and Technology Academy of CASIC), and Zhe Ma (Xlab, the Second Academy of CASIC, Beijing)</i>	
Revisit PCA-based Technique for Out-of-Distribution Detection	19374
<i>Xiaoyuan Guan (Sun yat-sen University), Zhouwu Liu (Sun Yat-sen University), Wei-Shi Zheng (Sun Yat-sen University, China), Yuren Zhou (Sun Yat-sen University), and Ruixuan Wang (Sun Yat-sen University)</i>	
Cross-view Topology Based Consistent and Complementary Information for Deep Multi-view Clustering	19383
<i>Zhibin Dong (nudt), Siwei Wang (Intelligent Game and Decision Lab), Jiaqi Jin (National University of Defence Technology), Xinwang Liu (National University of Defense Technology), and En Zhu (National University of Defense Technology)</i>	

Document Analysis and Understanding

A Benchmark for Chinese-English Scene Text Image Super-Resolution	19395
<i>Jianqi Ma (the Hong Kong Polytechnic University), Zhefong Liang (Guangdong Oppo Mobile Telecommunications Corp., Ltd), Wangmeng Xiang (The Hong Kong Polytechnic University), Xi Yang (The Hong Kong Polytechnic University), and Lei Zhang (Hong Kong Polytechnic University, Hong Kong, China)</i>	
Vision Grid Transformer for Document Layout Analysis	19405
<i>Cheng Da (Alibaba DAMO Academy), Chuwei Luo (Alibaba Group), Qi Zheng (Alibaba Group), and Cong Yao (Alibaba DAMO Academy)</i>	
Self-Supervised Character-to-Character Distillation for Text Recognition	19416
<i>Tongkun Guan (Shanghai Jiao Tong University), Wei Shen (SJTU), Xue Yang (Shanghai Jiao Tong University), Qi Feng (Shanghai Jiao Tong University), Zekun Jiang (Shanghai Jiao Tong University), and Xiaokang Yang (Shanghai Jiao Tong University of China)</i>	
ICL-D3IE: In-Context Learning with Diverse Demonstrations Updating for Document Information Extraction	19428
<i>Jiabang He (University of Electronic Science and Technology of China), Lei Wang (Singapore Management University), Yi Hu (University of Electronic Science and Technology of China), Ning Liu (Beijing Forestry University), Hui Liu (Beijing Rongda Technology Co., Ltd.), Xing Xu (University of Electronic Science and Technology of China), and Heng Tao Shen (University of Electronic Science and Technology of China, (UESTC)</i>	

ESTextSpotter: Towards Better Scene Text Spotting with Explicit Synergy in Transformer	19438
<i>Mingxin Huang (South China University of Technology), Jiaxin Zhang (ByteDance), Dezhi Peng (South China University of Technology), Hao Lu (Huazhong University of Science and Technology), Can Huang (ByteDance AI Lab), Yuliang Liu (Huazhong University of Science and Technology), Xiang Bai (Huazhong University of Science and Technology), and Lianwen Jin (South China University of Technology)</i>	
Few Shot Font Generation Via Transferring Similarity Guided Global Style and Quantization	
Local Style	19449
<i>Wei Pan (Wuhan University of Technology), Anna Zhu (Wuhan University of Technology), Xinyu Zhou (Wuhan University of Technology), Brian Kenji Iwana (Kyushu University), and Shilin Li (WUT)</i>	
Attention Where It Matters: Rethinking Visual Document Understanding with Selective Region Concentration	19460
<i>Haoyu Cao (Tencent), Changcun Bao (Tencent), Chaohu Liu (University of Science and Technology of China), Huang Chen (Tencent), Kun Yin (tencent), Hao Liu (Tencent Youtu Lab), Yinsong Liu (Tencent), Deqiang Jiang (Tencent), and Xing Sun (Tencent)</i>	
Document Understanding Dataset and Evaluation (DUDE)	19471
<i>Jordy Van Landeghem (Contract.fit), Rubèn Tito (Computer Vision Center, (CVC), Łukasz Borchmann (Snowflake), Michał Pietruszka (Jagiellonian University), Pawel Joziak (Warsaw University of Technology), Rafal Powalski (instabase), Dawid Jurkiewicz (Snowflake), Mickael Coustaty (L3i laboratory), Bertrand Anckaert (Smart and Easy NV), Ernest Valveny (Universitat Autònoma de Barcelona), Matthew Blaschko (KU Leuven), Sien Moens (KU Leuven), and Tomasz Stanislawek (Snowflake)</i>	
LISTER: Neighbor Decoding for Length-Insensitive Scene Text Recognition	19484
<i>Changxu Cheng (Alibaba DAMO Academy), Peng Wang (Alibaba DAMO Academy), Cheng Da (Alibaba DAMO Academy), Qi Zheng (Alibaba Group), and Cong Yao (Alibaba DAMO Academy)</i>	
MolGrapher: Graph-based Visual Recognition of Chemical Structures	19495
<i>Lucas Morin (ETH Zurich, IBM Research), Martin Danelljan (ETH Zurich), Maria Isabel Agea (IBM Research), Ahmed Nassar (IBM Research), Valery Weber (ibm), Ingmar Meijer (IBM Research -- Zurich), Peter Staar (IBM Research), and Fisher Yu (ETH Zurich)</i>	
SCOB: Universal Text Understanding via Character-wise Supervised Contrastive Learning with Online Text Rendering for Bridging Domain Gap	19505
<i>Daehee Kim (Naver Cloud Corp.), Yoonsik Kim (Text Vision, NAVER Cloud Corp.), DongHyun Kim (NAVER Cloud), Yumin Lim (Seoul National University), Geewook Kim (NAVER Corporation), and Taeho Kil (Naver Cloud Corp.)</i>	
Foreground and Text-lines Aware Document Image Rectification	19517
<i>Heng Li (Harbin Institute of Technology, Shenzhen), Xiangping Wu (Harbin Institute of Technology, Shenzhen), Qingcai Chen (Harbin Institute of Technology, Shenzhen), and Qianjin Xiang (Harbin Institute of Technology, Shenzhen)</i>	

DocTr: Document Transformer for Structured Information Extraction in Documents	19527
<i>Haofu Liao (Amazon), Aruni RoyChowdhury (University of Massachusetts, Amherst), Weijian Li (Amazon), Ankan Bansal (Amazon), Yuting Zhang (Amazon), Zhuowen Tu (UC San Diego), Ravi Kumar Satzoda (Amazon Web Services), R. R. Manmatha (Amazon), and Vijay Mahadevan (Amazon Inc.)</i>	

Biometrics

GPgait: Generalized Pose-based Gait Recognition	19538
<i>Yang Fu (Beijing Normal University), Shibe Meng (Beijing Normal University), Saihui Hou (Beijing Normal University), Xuecai Hu (Beijing Normal University), and Yongzhen Huang (School of Artificial Intelligence, Beijing Normal University)</i>	
RPG-Palm: Realistic Pseudo-data Generation for Palmprint Recognition	19548
<i>Lei Shen (Tencent), Jianlong Jin (Hefei University of Technology), Ruixin Zhang (Tencent), Huaen Li (Hefei University of Technology), Kai Zhao (UCLA), Yingyi Zhang (Tencent), Jingyun Zhang (Tencent), Shouhong Ding (Tencent), Yang Zhao (Hefei University of Technology), and Wei Jia (Hefei University of Technology)</i>	
Learning Clothing and Pose Invariant 3D Shape Representation for Long-Term Person Re-Identification	19560
<i>Feng Liu (Michigan State University), Minchul Kim (Michigan State University), ZiAng Gu (Michigan State University), Anil Jain (Michigan State University), and Xiaoming Liu (Michigan State University)</i>	
Physics-Augmented Autoencoder for 3D Skeleton-Based Gait Recognition	19570
<i>Hongji Guo (Rensselaer Polytechnic Institute) and Qiang Ji (Rensselaer Polytechnic Institute)</i>	
Hierarchical Spatio-Temporal Representation Learning for Gait Recognition	19582
<i>Lei Wang (Hebei Agricultural University), Bo Liu (Hebei Agricultural University), Fangfang Liang (Hebei Agricultural University), and Bincheng Wang (Hebei Agricultural University)</i>	
IDiff-Face: Synthetic-based Face Recognition through Fizzy Identity-Conditioned Diffusion Model	19593
<i>Fadi Boutros (Fraunhofer IGD), Jonas Henry Grebe (Fraunhofer IGD), Arjan Kuijper (Fraunhofer Institute for Computer Graphics Research IGD and Mathematical and Applied Visual Computing group, TU Darmstadt), and Naser Damer (Fraunhofer Institute for Computer Graphics Research IGD and TU Darmstadt)</i>	
Template Inversion Attack against Face Recognition Systems using 3D Face Reconstruction	19605
<i>Hatef Otroschi Shahreza (Idiap Research Institute) and Sebastien Marcel (Idiap Research Institute)</i>	
Privacy-Preserving Face Recognition Using Random Frequency Components	19616
<i>Yuxi Mi (Fudan University), Yuge Huang (Tencent YouTu), Jiazhen Ji (Tencent), Minyi Zhao (Fudan University), Jiayang Wu (Tencent), Xingkun Xu (Tencent), Shouhong Ding (Tencent), and Shuigeng Zhou (Fudan University)</i>	

FLIP: Cross-domain Face Anti-spoofing with Language Guidance	19628
<i>Koushik Srivatsan (MBZUAI), Muzammal Naseer (MBZUAI), and Karthik Nandakumar (Mohamed Bin Zayed University of Artificial Intelligence)</i>	

(O) 3D from Multi-view and Sensors 2

Zip-NeRF: Anti-Aliased Grid-Based Neural Radiance Fields	19640
<i>Jonathan T. Barron (Google Research), Ben Mildenhall (Google Research), Dor Verbin (Google Research), Pratul P. Srinivasan (Google Research), and Peter Hedman (Google Research)</i>	
Mixed Neural Voxels for Fast Multi-view Video Synthesis	19649
<i>Feng Wang (Tsinghua University), Sinan Tan (Tsinghua University), Xinghang Li (Tsinghua University), Zeyue Tian (Hong Kong University of Science and Technology), Yafei Song (Alibaba Group), and Huaping Liu (Tsinghua University)</i>	
Diffusion-Guided Reconstruction of Everyday Hand-Object Interaction Clips	19660
<i>Yufei Ye (Carnegie Mellon University), Poorvi Hebbbar (Carnegie Mellon University), Abhinav Gupta (CMU/FAIR), and Shubham Tulsiani (Carnegie Mellon University)</i>	
LERF: Language Embedded Radiance Fields	19672
<i>Justin Kerr (Berkeley), Chung Min Kim (UC Berkeley), Ken Goldberg (UC Berkeley), Angjoo Kanazawa (University of California Berkeley), and Matthew Tancik (UC Berkeley)</i>	
Instruct-NeRF2NeRF: Editing 3D Scenes with Instructions	19683
<i>Ayaan Haque (UC Berkeley), Matthew Tancik (UC Berkeley), Alexei A. Efros (UC Berkeley), Aleksander Holynski (UC Berkeley / Google Research), and Angjoo Kanazawa (University of California Berkeley)</i>	
PIAC: Revisiting Absolute Pose From a Single Affine Correspondence	19694
<i>Jonathan Ventura (California Polytechnic State University, San Luis Obispo), Zuzana Kukelova (Czech Technical University in Prague), Torsten Sattler (Czech Technical University in Prague), and Dániel Baráth (ETH Zürich)</i>	
Neural Haircut: Prior-Guided Strand-Based Hair Reconstruction	19705
<i>Vanessa Sklyarova (Max Planck Institute for Intelligent Systems), Jenya Chelishev (Samsung AI Center, Moscow), Andreea Dogaru (Friedrich-Alexander-Universität Erlangen-Nürnberg), Igor Medvedev (Samsung AI Center), Victor Lempitsky (Cinemersive Labs), and Egor Zakharov (Samsung AI Center)</i>	
Tri-MipRF: Tri-Mip Representation for Efficient Anti-Aliasing Neural Radiance Fields	19717
<i>Wenbo Hu (Pico MR, Bytedance Ltd.), Yuling Wang (Tsinghua University), Lin Ma (Bytedance), Bangbang Yang (Bytedance Inc), Lin Gao (Institute of Computing Technology, Chinese Academy of Sciences), Xiao Liu (Bytedance Inc.), and Yuewen Ma (Pico)</i>	
LiDAR-UDA: Self-ensembling Through Time for Unsupervised LiDAR Domain Adaptation	19727
<i>Amirreza Shaban (University of Washington), JoonHo Lee (University of Washington), Sanghun Jung (University of Washington), Xiangyun Meng (University of Washington), and Byron Boots (University of Washington)</i>	

Tracking Everything Everywhere All at Once	19738
<i>Qianqian Wang (Cornell), Yen-Yu Chang (Cornell University), Ruojin Cai (Cornell University), Zhengqi Li (Google Inc.), Bharath Hariharan (Cornell University), Aleksander Holynski (UC Berkeley / Google Research), and Noah Snavely (Cornell University and Google AI)</i>	
Ego-Humans: An Ego-Centric 3D Multi-Human Benchmark	19750
<i>Rawal Khirodkar (Carnegie Mellon University), Aayush Bansal (Carnegie Mellon University), Lingni Ma (Facebook Reality Labs), Richard Newcombe (Facebook), Minh Vo (Facebook Reality Labs), and Kris Kitani (Carnegie Mellon University)</i>	
Once Detected, Never Lost: Surpassing Human Performance in Offline LiDAR based 3D Object Detection	19763
<i>Lue Fan (Institute of Automation, Chinese Academy of Sciences), Yuxue Yang (Institute of Automation, Chinese Academy of Sciences), Yiming Mao (TuSimple), Feng Wang (TuSimple), Yuntao Chen (CAIR, HKISI, CAS), Naiyan Wang (TuSimple), and Zhaoxiang Zhang (Chinese Academy of Sciences, China)</i>	

(O) Machine Learning and Dataset

DiffusionDet: Diffusion Model for Object Detection	19773
<i>Shoufa Chen (The University of Hong Kong), Peize Sun (The University of Hong Kong), Yibing Song (Fudan University), and Ping Luo (The University of Hong Kong)</i>	
V3Det: Vast Vocabulary Visual Detection Dataset	19787
<i>Jiaqi Wang (Shanghai AI Laboratory), Pan Zhang (Shanghai AI Laboratory), Tao Chu (South China University of Technology), Yuhang Cao (CUHK), Yujie Zhou (Renmin University of China), Tong Wu (The Chinese University of Hong Kong), Bin Wang (Shanghai AI Laboratory), Conghui He (Shanghai AI Lab), and Dahua Lin (The Chinese University of Hong Kong)</i>	
PointOdyssey: A Large-Scale Synthetic Dataset for Long-Term Point Tracking	19798
<i>Yang Zheng (Stanford University), Adam W. Harley (Stanford University), Bokui Shen (Stanford), Gordon Wetzstein (Stanford University), and Leonidas J. Guibas (Stanford University)</i>	
Label-Free Event-based Object Recognition via Joint Learning with Image Reconstruction from Events	19809
<i>Hoonhee Cho (KAIST), Hyeonseong Kim (KAIST), Yujeong Chae (KAIST), and Kuk-Jin Yoon (KAIST)</i>	
Vision HGNN: An Image is More than a Graph of Nodes	19821
<i>Yan Han (UT Austin), Peihao Wang (University of Texas at Austin), Souvik Kundu (Intel Labs), Ying Ding (The University of Texas at Austin), and Zhangyang Wang (University of Texas at Austin)</i>	
Revisiting Vision Transformer from the View of Path Ensemble	19832
<i>Shuning Chang (National University of Singapore), Pichao Wang (Alibaba Group), Hao Luo (Alibaba group), Fan Wang (Alibaba Group), and Mike Zheng Shou (National University of Singapore)</i>	

All in Tokens: Unifying Output Space of Visual Tasks via Soft Token	19843
<i>Jia Ning (Huazhong University of Science and Technology), Chen Li (Xi'an Jiaotong University), Zheng Zhang (MSRA), Chunyu Wang (Microsoft Research asia), Zigang Geng (University of Science and Technology of China), Qi Dai (Microsoft Research), Kun He (Huazhong University of Science and Technology), and Han Hu (Microsoft Research Asia)</i>	
Mitigating and Evaluating Static Bias of Action Representations in the Background and the Foreground	19854
<i>Haoxin Li (Nanyang Technological University), Yuan Liu (Guangzhou University), Hanwang Zhang (Nanyang Technological University), and Boyang Li (Nanyang Technological University)</i>	
Deep Multitask Learning with Progressive Parameter Sharing	19867
<i>Haosen Shi (The Chinese University of Hong Kong), Shen Ren (Continental Automotive Singapore), Tianwei Zhang (Nanyang Technological University), and Sinno Jialin Pan (The Chinese University of Hong Kong)</i>	
Implicit Temporal Modeling with Learnable Alignment for Video Recognition	19879
<i>Shuyuan Tu (Fudan University), Qi Dai (Microsoft Research), Zuxuan Wu (Fudan University), Zhi-Qi Cheng (Carnegie Mellon University), Han Hu (Microsoft Research Asia), and Yu-Gang Jiang (Fudan University)</i>	
Unmasked Teacher: Towards Training-Efficient Video Foundation Models	19891
<i>Kunchang Li (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences), Yali Wang (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences), Yizhuo Li (The University of Hong Kong), Yi Wang (Shanghai AI Laboratory), Yinan He (Shanghai AI Laboratory), Limin Wang (Nanjing University), and Yu Qiao (Shanghai AI Laboratory)</i>	
Large-Scale Person Detection and Localization Using Overhead Fisheye Cameras	19904
<i>Lu Yang (Beijing University of Posts and Telecommunications), Liulei Li (University of Technology Sydney), Xueshi Xin (Beijing University of Posts and Telecommunications), Yifan Sun (Baidu Research), Qing Song (Beijing University of Posts and Telecommunications), and Wenguan Wang (Zhejiang University)</i>	

Datasets and Evaluation

A step towards understanding why classification helps regression	19915
<i>Silvia L. Pintea (TUDelft), Yancong Lin (Delft University of Technology), Jouke Dijkstra (Leiden University Medical Center), and Jan C. van Gemert (Delft University of Technology)</i>	

DNA-Rendering: A Diverse Neural Actor Repository for High-Fidelity Human-Centric Rendering.....	19925
<i>Wei Cheng (HKUST), Ruixiang Chen (SenseTime), Siming Fan (SenseTime), Wanqi Yin (SenseTime Research), Keyu Chen (Shanghai Artificial Intelligence Lab), Zhongang Cai (SenseTime International Pte Ltd), Jingbo Wang (The Chinese University of HongKong), Yang Gao (SenseTime), Zhengming Yu (SenseTime), Zhengyu Lin (SenseTime), Daxuan Ren (Nanyang Technological University), Lei Yang (SenseTime Group Limited), Ziwei Liu (Nanyang Technological University), Chen Change Loy (Nanyang Technological University), Chen Qian (SenseTime), Wayne Wu (Shanghai AI Laboratory), Dahua Lin (The Chinese University of Hong Kong), Bo Dai (Shanghai AI Lab), and Kwan-Yee Lin (SenseTime Research)</i>	
Robo3D: Towards Robust and Reliable 3D Perception against Corruptions	19937
<i>Lingdong Kong (National University of Singapore), Youquan Liu (Hochschule Bremerhaven), Xin Li (East China Normal University), Runnan Chen (The University of Hong Kong), Wenwei Zhang (NTU), Jiawei Ren (Nanyang Technological University), Liang Pan (Nanyang Technological University), Kai Chen (Shanghai AI Laboratory), and Ziwei Liu (Nanyang Technological University)</i>	
Efficient Discovery and Effective Evaluation of Visual Perceptual Similarity: A Benchmark and Beyond	19950
<i>Oren Barkan (The Open University), Tal Reiss (The Hebrew University of Jerusalem), Jonathan Weill (Tel Aviv University), Ori Katz (Technion), Roy Hirsch (Tel-Aviv University), Itzik Malkiel (Tel Aviv University), and Noam Koenigstein (Microsoft)</i>	
DetermiNet: A Large-Scale Diagnostic Dataset for Complex Visually-Grounded Referencing using Determiners	19962
<i>Clarence Lee (Singapore University of Technology and Design), M Ganesh Kumar (National University of Singapore, (NUS), and Cheston Tan (Institute for Infocomm Research, Singapore)</i>	
Beyond Object Recognition: A New Benchmark towards Object Concept Learning	19972
<i>Yong-Lu Li (Shanghai Jiao Tong University), Yue Xu (Shanghai Jiao Tong University), Xinyu Xu (Shanghai Jiao Tong University), Xiaohan Mao (Shanghai Jiao Tong University), Yuan Yao (University of Rochester), Siqi Liu (Shanghai Jiao Tong University), and Cewu Lu (Shanghai Jiao Tong University)</i>	
HRS-Bench: Holistic, Reliable and Scalable Benchmark for Text-to-Image Models	19984
<i>Eslam Mohamed Bakr (KAUST), Pengzhan Sun (National University of Singapore), Xiaoqian Shen (King Abdullah University of Science and Technology), Faizan Farooq Khan (KAUST), Li Erran Li (Columbia University), and Mohamed Elhoseiny (KAUST)</i>	
SegRCDB: Semantic Segmentation via Formula-Driven Supervised Learning	19997
<i>Risa Shinoda (Kyoto University, National Institute of Advanced Industrial Science and Technology, (AIST), Ryo Hayamizu (National Institute of Advanced Industrial Science and Technology, (AIST), Kodai Nakashima (CyberAgent, Univ. of Tsukuba, AIST), Nakamasa Inoue (Tokyo Institute of Technology), Rio Yokota (Tokyo Institute of Technology), and Hirokatsu Kataoka (National Institute of Advanced Industrial Science and Technology, (AIST)</i>	

LoTE-Animal: A Long Time-span Dataset for Endangered Animal Behavior Understanding	20007
<i>Dan Liu (Northwest A&F University), Jin Hou (Beijing Normal University), Shaoli Huang (Tencent AI-Lab), Jing Liu (Northwest A&F University), Yuxin He (North Sichuan Medical College), Bochuan Zheng (China West Normal University), Jifeng Ning (Northwest A&F University), and Jingdong Zhang (China West Normal University)</i>	
Building3D: A Urban-Scale Dataset and Benchmarks for Learning Roof Structures from Point Clouds	20019
<i>Ruisheng Wang (University of Calgary, Canada), Shangfeng Huang (University of Calgary), and Hongxin Yang (University of Calgary)</i>	
Lecture Presentations Multimodal Dataset: Towards Understanding Multimodality in Educational Videos	20030
<i>Dong Won Lee (Massachusetts Institute of Technology), Chaitanya Ahuja (Meta AI), Paul Pu Liang (Carnegie Mellon University), Sanika Natu (Tecton), and Louis-Philippe Morency (Carnegie Mellon University)</i>	
Probabilistic Precision and Recall Towards Reliable Evaluation of Generative Models	20042
<i>Dogyun Park (Korea University) and Suhyun Kim (Korea Institute of Science and Technology)</i>	
EgoObjects: A Large-Scale Egocentric Dataset for Fine-Grained Object Understanding	20053
<i>Chenchen Zhu (Meta), Fanyi Xiao (Meta), Andres Alvarado (Meta), Yasmine Babaei (Meta), Jiabo Hu (Meta), Hichem El-Mohri (Meta), Sean Chang Culatana (Facebook), Roshan Sumbaly (Facebook), and Zhicheng Yan (Meta Reality Labs)</i>	
CAME: Contrastive Automated Model Evaluation	20064
<i>Ru Peng (Zhejiang University), Qiuyang Duan (Zhejiang University), Haobo Wang (Zhejiang University), Jiachen Ma (Zhejiang University), Yanbo Jiang (Zhejiang University), Yongjun Tu (Guangdong OPPO Mobile Telecommunications Corp., Ltd.), Xiu Jiang (Guangdong OPPO Mobile Telecommunications Corp., Ltd.), and Junbo Zhao (Zhejiang University)</i>	
Aria Digital Twin: A New Benchmark Dataset for Egocentric 3D Machine Perception	20076
<i>Xiaqing Pan (Meta Platforms, Inc.), Nicholas Charron (Meta), Yongqian Yang (Meta Platform), Scott Peters (Meta), Thomas Whelan (Meta), Chen Kong (Meta), Omkar Parkhi (Facebook), Richard Newcombe (Facebook), and Yuheng Yuheng (n/a, Meta)</i>	
Exploring Video Quality Assessment on User Generated Contents from Aesthetic and Technical Perspectives	20087
<i>Haoning Wu (Nanyang Technological University), Erli Zhang (Nanyang Technological University), Liang Liao (Nanyang Technological University), Chaofeng Chen (Nanyang Technological University), Jingwen Hou (Nanyang Technological University), Annan Wang (Nanyang Technological University), Wenxiu Sun (SenseTime Research and Tetras.AI), Qiong Yan (SenseTime Group Limited), and Weisi Lin (Nanyang Technological University, Singapore)</i>	

Going Beyond Nouns With Vision & Language Models Using Synthetic Data	20098
<i>Paola Cascante-Bonilla (Rice University), Khaled Shehada (Massachusetts Institute of Technology), James Seale Smith (Georgia Institute of Technology), Sivan Doveh (IBM-Research), Donghyun Kim (MIT-IBM Watson AI Lab), Rameswar Panda (MIT-IBM Watson AI Lab), Gul Varol (Ecole des Ponts ParisTech), Aude Oliva (MIT), Vicente Ordonez (Rice University), Rogerio Feris (MIT-IBM Watson AI Lab, IBM Research), and Leonid Karlinsky (IBM-Research)</i>	
H3WB: Human3.6M 3D WholeBody Dataset and Benchmark	20109
<i>Yue Zhu (ENPC), Nermin Samet (Ecole des Ponts ParisTech), and David Picard (ENPC)</i>	
Zenseact Open Dataset: A Large-Scale and Diverse Multimodal Dataset for Autonomous Driving.....	20121
<i>Mina Alibeigi (Zenseact), William Ljungbergh (Linköping University), Adam Tonderski (Lund University), Georg Hess (Chalmers University of Technology), Adam Lilja (Zenseact), Carl Lindström (Chalmers University of Technology), Daria Motorniuk (Zenseact), Junsheng Fu (Zenseact), Jenny Widahl (Zenseact AB), and Christoffer Petersson (Zenseact)</i>	
CAD-Estate: Large-scale CAD Model Annotation in RGB Videos	20132
<i>Kevis-Kokitsi Maninis (Google Research), Stefan Popov (Google Research), Matthias Nießner (Technical University of Munich), and Vittorio Ferrari (Google Research)</i>	
Neglected Free Lunch - Learning Image Classifiers Using Annotation Byproducts	20143
<i>Dongyoon Han (NAVER AI Lab), Junsuk Choe (Sogang University), Seonghyeok Chun (Dante Company), John Joon Young Young Chung (SpaceCraft), Minsuk Chang (Google), Sangdoon Yun (NAVER AI LAB), Jean Y. Song (DGIST), and Seong Joon Oh (Tübingen AI Center)</i>	
Chaotic World: A Large and Challenging Benchmark for Human Behavior Understanding in Chaotic Events	20156
<i>Kian Eng Ong (Singapore University of Technology and Design), Xun Long Ng (Singapore University of Technology and Design), Yanchao Li (Singapore University of Technology and Design), Wenjie Ai (Singapore University of Technology and Design), Kuangyi Zhao (missing), Si Yong (SUTD), and Jun Liu (Singapore University of Technology and Design)</i>	
MOSE: A New Dataset for Video Object Segmentation in Complex Scenes	20167
<i>Henghui Ding (Nanyang Technological University), Chang Liu (Nanyang Technological University), Shuting He (Zhejiang University), Xudong Jiang (Nanyang Technological University), Philip H.S. Torr (University of Oxford), and Song Bai (University of Oxford)</i>	
Spurious Features Everywhere - Large-Scale Detection of Harmful Spurious Features in ImageNet	20178
<i>Yannic Neuhaus (University of Tübingen), Maximilian Augustin (University of Tuebingen), Valentyn Boreiko (University of Tübingen), and Matthias Hein (University of Tübingen)</i>	

Chop & Learn: Recognizing and Generating Object-State Compositions	20190
<i>Nirat Saini (University of Maryland), Hanyu Wang (University of Maryland - College Park), Archana Swaminathan (University of Maryland), Vinoj Jayasundara (University of Maryland), Bo He (University of Maryland), Kamal Gupta (University of Maryland), and Abhinav Shrivastava (University of Maryland)</i>	
Building Bridge Across the Time: Disruption and Restoration of Murals In the Wild	20202
<i>Huiyang Shao (Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Science), Qianqian Xu (Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences), Peisong Wen (Institute of Computing Technology, Chinese Academy of Sciences, University of Chinese Academy of Sciences), Peifeng Gao (School of Computer Science and Technology, University of Chinese Academy of Sciences), Zhiyong Yang (University of Chinese Academy of Sciences, UCAS), and Qingming Huang (University of Chinese Academy of Sciences)</i>	
HoloAssist: an Egocentric Human Interaction Dataset for Interactive AI Assistants in the Real World	20213
<i>Xin Wang (Microsoft Research), Taemin Kwon (ETH Zurich), Mahdi Rad (Microsoft), Bowen Pan (Microsoft), Ishani Chakraborty (Microsoft AI & R), Sean Andrist (Microsoft Research), Dan Bohus (Microsoft Research), Ashley Feniello (Microsoft), Bugra Tekin (Microsoft), Felipe Vieira Frujeri (Microsoft Research), Neel Joshi (MICROSOFT RESEARCH), and Marc Pollefeys (ETH Zurich / Microsoft)</i>	
SynBody: Synthetic Dataset with Layered Human Models for 3D Human Perception and Modeling	20225
<i>Zhitao Yang (SenseTime Group Limited), Zhongang Cai (SenseTime International Pte Ltd), Haiyi Mei (SenseTime Group Limited), Shuai Liu (Shanghai AI Laboratory), Zhaoxi Chen (Nanyang Technological University), Weiye Xiao (Southeast University), Yukun Wei (SenseTime Group Limited), Zhongfei Qing (SenseTime Group Limited), Chen Wei (Nanyang Technological University), Bo Dai (Shanghai AI Lab), Wayne Wu (Shanghai AI Laboratory), Chen Qian (SenseTime), Dahua Lin (The Chinese University of Hong Kong), Ziwei Liu (Nanyang Technological University), and Lei Yang (SenseTime Group Limited)</i>	
OxfordTVG-HIC: Can Machine Make Humorous Captions from Images?	20236
<i>Runjia Li (University of Oxford), Shuyang Sun (University of Oxford), Mohamed Elhoseiny (KAUST), and Philip Torr (University of Oxford)</i>	
LaRS: A Diverse Panoptic Maritime Obstacle Detection Dataset and Benchmark	20247
<i>Lojze Žust (University of Ljubljana), Janez Perš (University of Ljubljana), and Matej Kristan (University of Ljubljana)</i>	
Joint Metrics Matter: A Better Standard for Trajectory Forecasting	20258
<i>Erica Weng (Carnegie Mellon University), Hana Hoshino (Carnegie Mellon University), Deva Ramanan (Carnegie Mellon University), and Kris Kitani (Carnegie Mellon University)</i>	

LPPF: A Portrait Dataset for Face Generators Across Large Poses	20270
<i>Yiqian Wu (Zhejiang University), Jing Zhang (Zhejiang University), Hongbo Fu (City University of Hong Kong), and Xiaogang Jin (Zhejiang University)</i>	
Replay: Multi-modal Multi-view Acted Videos for Casual Holography	20281
<i>Roman Shapovalov (Facebook), Yanir Kleiman (Meta AI), Ignacio Rocco (Facebook AI Research), David Novotny (Facebook AI Research), Andrea Vedaldi (Meta and University of Oxford), Changan Chen (University of Texas at Austin), Filippos Kokkinos (Meta), Ben Graham (Facebook Research), and Natalia Neverova (Facebook AI Research)</i>	
Human-centric Scene Understanding for 3D Large-scale Scenarios	20292
<i>Yiteng Xu (ShanghaiTech University), Peishan Cong (ShanghaiTech), Yichen Yao (ShanghaiTech University), Runnan Chen (The University of Hong Kong), Yuenan Hou (Shanghai AI Lab), Xinge Zhu (The Chinese University of Hong Kong), Xuming He (ShanghaiTech University), Jingyi Yu (Shanghai Tech University), and Yuexin Ma (ShanghaiTech University)</i>	
Pre-training Vision Transformers with Very Limited Synthesized Images	20303
<i>Ryo Nakamura (Fukuoka University, National Institute of Advanced Industrial Science and Technology, (AIST), Hirokatsu Kataoka (National Institute of Advanced Industrial Science and Technology, (AIST), Sora Takashima (Tokyo Institute of Technology / National Institute of Advanced Industrial Science and Technology, (AIST), Edgar Josafat Martinez Noriega (AIST), Rio Yokota (Tokyo Institute of Technology), and Nakamasa Inoue (Tokyo Institute of Technology)</i>	
FACET: Fairness in Computer Vision Evaluation Benchmark	20313
<i>Laura Gustafson (Facebook AI Research), Chloe Rolland (Meta AI), Nikhila Ravi (Facebook AI Research), Quentin Duval (Facebook), Aaron Adcock (Facebook AI Research), Cheng-Yang Fu (Facebook.com), Melissa Hall (Facebook), and Candace Ross (Massachusetts Institute of Technology)</i>	
EmoSet: A Large-scale Visual Emotion Dataset with Rich Attributes	20326
<i>Jingyuan Yang (Shenzhen Univeristy), Qirui Huang (Shenzhen University), Tingting Ding (Shenzhen University), Dani Lischinski (The Hebrew University of Jerusalem), Danny Cohen-Or (Tel Aviv University), and Hui Huang (Shenzhen University)</i>	
RenderIH: A Large-Scale Synthetic Dataset for 3D Interacting Hand Pose Estimation	20338
<i>Lijun Li (Shanghai Artificial Intelligence Laboratory), Linrui Tian (alibaba), Xindi Zhang (Queen Mary University of London), Qi Wang (Alibaba group), Bang Zhang (Damo Academy Alibaba), Liefeng Bo (Alibaba DAMO Academy), Mengyuan Liu (Peking University, Shenzhen Graduate School), and Chen Chen (University of Central Florida)</i>	
TIFA: Accurate and Interpretable Text-to-Image Faithfulness Evaluation with Question Answering	20349
<i>Yushi Hu (University of Washington), Benlin Liu (University of Washington), Jungo Kasai (University of Washington), Yizhong Wang (University of Washington), Mari Ostendorf (University of Washington), Ranjay Krishna (University of Washington), and Noah A. Smith (University of Washington and Allen Institute for AI)</i>	

Exploring the Sim2Real Gap Using Digital Twins	20361
<i>Sruthi Sudhakar (Columbia University), Jon Hanzelka (Microsoft Research), Josh Bobillot (Microsoft), Tanmay Randhavane (Microsoft Research), Neel Joshi (MICROSOFT RESEARCH), and Vibhav Vineet (Microsoft Research)</i>	
ClothesNet: An Information-Rich 3D Garment Model Repository with Simulated Clothes Environment	20371
<i>Bingyang Zhou (Harbin Institute of Technology Shenzhen), Haoyu Zhou (Beihang University), Tianhai Liang (Harbin Institute of Technology, Shenzhen), Qiaojun Yu (Shanghai Jiaotong University), Siheng Zhao (National University of Singapore), Yuwei Zeng (National University of Singapore), Jun Lv (Shanghai Jiao Tong University), Siyuan Luo (Xi'an Jiaotong University), Qiancai Wang (Harbin Institute of technology, Shenzhen), Xinyuan Yu (National University of Singapore), Haonan Chen (Nanjing University), Cewu Lu (Shanghai Jiao Tong University), and Lin Shao (National University of Singapore)</i>	
Video State-Changing Object Segmentation	20382
<i>Jiangwei Yu (Tsinghua University), Xiang Li (University of Illinois at Urbana-Champaign), Xinran Zhao (Stanford University), Hongming Zhang (Tencent AI Lab, Seattle), and Yu-Xiong Wang (University of Illinois at Urbana-Champaign)</i>	
PlanarTrack: A Large-scale Challenging Benchmark for Planar Object Tracking	20392
<i>Xinran Liu (Institute of Software, Chinese Academy of Science), Xiaoqiong Liu (University of North Texas), Ziruo Yi (University of North Texas), Xin Zhou (UESTC), Thanh Le (University of North Texas), Libo Zhang (Institute of Software Chinese Academy of Sciences), Yan Huang (University of North Texas), Qing Yang (University of North Texas), and Heng Fan (University of North Texas)</i>	
AIDE: A Vision-Driven Multi-View, Multi-Modal, Multi-Tasking Dataset for Assistive Driving Perception	20402
<i>Dingkang Yang (Fudan University; Institute of Meta-Medical), Shuai Huang (Fudan University), Zhi Xu (Fudan University), Zhenpeng Li (Fudan University), Shunli Wang (Fudan University), Mingcheng Li (Fudan University), Yuzheng Wang (Fudan University), Yang Liu (Fudan University), Kun Yang (Fudan University), Zhaoyu Chen (Fudan University), Yan Wang (Fudan University), Jing Liu (Fudan University), Peixuan Zhang (Boli Technology Co., Ltd., Changchun, China), Peng Zhai (Fudan University), and Lihua Zhang (Fudan University; Institute of Meta-Medical; Engineering Research Center of AI and Robotics, Ministry of Education, Shanghai, China; AI and Unmanned Systems Engineering Research Center of Jilin Province, Changchun, China; Boli Technology Co., Ltd., Changchun)</i>	
Harvard Glaucoma Detection and Progression: A Multimodal Multitask Dataset and Generalization-Reinforced Semi-Supervised Learning	20414
<i>Yan Luo (Harvard University), Min Shi (Harvard University), Yu Tian (Harvard University), Tobias Elze (Schepens Eye Research Inst., Harvard Med School), and Mengyu Wang (Harvard University)</i>	

ARNOLD: A Benchmark for Language-Grounded Task Learning with Continuous States in Realistic 3D Scenes	20426
<i>Ran Gong (UCLA), Jiangyong Huang (Peking University), Yizhou Zhao (UCLA), Haoran Geng (Peking University), Xiaofeng Gao (UCLA), Qingyang Wu (Columbia University), Wensi Ai (UCLA), Ziheng Zhou (University of California, Los Angeles), Demetri Terzopoulos (University of California, Los Angeles), Song-Chun Zhu (UCLA), Baoxiong Jia (UCLA), and Siyuan Huang (Beijing Institute for General Artificial Intelligence)</i>	
FishNet: A Large-scale Dataset and Benchmark for Fish Recognition, Detection, and Functional Trait Prediction	20439
<i>Faizan Farooq Khan (KAUST), Xiang Li (King Abdullah University of Science and Technology), Andrew J. Temple (King Abdullah University of Science and Technology), and Mohamed Elhoseiny (KAUST)</i>	
Towards Content-based Pixel Retrieval in Revisited Oxford and Paris	20450
<i>Guoyuan An (KAIST), Woo Jae Kim (KAIST), Saelyne Yang (KAIST), Rong Li (Zhejiang Lab), Yuchi Huo (Zhejiang University), and Sung-Eui Yoon (KAIST)</i>	
A Large-scale Study of Spatiotemporal Representation Learning with a New Benchmark on Action Recognition	20462
<i>Andong Deng (University of Central Florida), Taojiannan Yang (University of Central Florida), and Chen Chen (University of Central Florida)</i>	
SQAD: Automatic Smartphone Camera Quality Assessment and Benchmarking	20475
<i>Zilin Fang (National University of Singapore), Andrey Ignatov (ETH Zurich), Eduard Zamfir (University of Wurzburg), and Radu Timofte (University of Wurzburg & ETH Zurich)</i>	
Revisiting Scene Text Recognition: A Data Perspective	20486
<i>Qing Jiang (South China University of Technology), Jiapeng Wang (South China University of Technology), Dezhi Peng (South China University of Technology), Chongyu Liu (South China University of Technology), and Lianwen Jin (South China University of Technology)</i>	
Will Large-scale Generative Models Corrupt Future Datasets?	20498
<i>Ryuichiro Hataya (RIKEN), Han Bao (Kyoto University), and Hiromi Arai (RIKEN AIP)</i>	
360VOT: A New Benchmark Dataset for Omnidirectional Visual Object Tracking	20509
<i>Huajian Huang (The Hong Kong University of Science and Technology), Yinzhe Xu (The Hong Kong University of Science and Technology), Yingshu Chen (The Hong Kong University of Science and Technology), and Sai-Kit Yeung (Hong Kong University of Science and Technology)</i>	

Faces and Gestures

DeePoint: Visual Pointing Recognition and Direction Estimation	20520
<i>Shu Nakamura (Kyoto University), Yasutomo Kawanishi (RIKEN), Shohei Nobuhara (Kyoto University), and Ko Nishino (Kyoto University)</i>	

Contactless Pulse Estimation Leveraging Pseudo Labels and Self-Supervision	20531
<i>Zhihua Li (Binghamton University) and Lijun Yin (State University of New York at Binghamton)</i>	
Most Important Person-Guided Dual-Branch Cross-Patch Attention for Group Affect Recognition	20541
<i>Hongxia Xie (National Yang Ming Chiao Tung University), Ming-Xian Lee (National Yang Ming Chiao Tung University), Tzu-Jui Chen (National Yang Ming Chiao Tung University), Hung-Jen Chen (National Yang Ming Chiao Tung University), Hou-I Liu (National Yang Ming Chiao Tung University), Hong-Han Shuai (National Yang Ming Chiao Tung University), and Wen-Huang Cheng (National Taiwan University)</i>	
ContactGen: Generative Contact Modeling for Grasp Generation	20552
<i>Shaowei Liu (UIUC), Yang Zhou (Adobe Research), Jimei Yang (Adobe), Saurabh Gupta (UIUC), and Shenlong Wang (UIUC)</i>	
Imitator: Personalized Speech-driven 3D Facial Animation	20564
<i>Balamurugan Thambiraja (Max Planck Institute for Intelligent Systems), Ikhsanul Habibie (Max Planck Institute for Informatics), Sadegh Aliakbarian (Microsoft), Darren Cosker (Microsoft), Christian Theobalt (MPI Informatik), and Justus Thies (Max Planck Institute for Intelligent Systems)</i>	
DVGaze: Dual-View Gaze Estimation	20575
<i>Yihua Cheng (University of Birmingham) and Feng Lu (Beihang University)</i>	
TransFace: Calibrating Transformer Training for Face Recognition from a Data-Centric Perspective	20585
<i>Jun Dan (Zhejiang University), Yang Liu (Alibaba Group), Haoyu Xie (Northeastern University), Jiankang Deng (Imperial College London), Haoran Xie (Tsinghua university), Xuansong Xie (Alibaba), and Baigui Sun (Alibaba Group)</i>	
Towards Unsupervised Domain Generalization for Face Anti-Spoofing	20597
<i>Yuchen Liu (Shanghai Jiao Tong university), Yabo Chen (Shanghai Jiao Tong University), Mengran Gou (Qualcomm), Chun-Ting Huang (Qualcomm), Yaoming Wang (Shanghai Jiao Tong University), Wenrui Dai (Shanghai Jiao Tong University), and Hongkai Xiong (Shanghai Jiao Tong University)</i>	
Reinforced Disentanglement for Face Swapping without Skip Connection	20608
<i>Xiaohang Ren (Xiaobing.AI), Xingyu Chen (Xiaobing.AI), Pengfei Yao (Xiaobing.AI), Heung-Yeung Shum (Hong Kong University of Science and Technology), and Baoyuan Wang (Xiaobing.AI)</i>	
CoSign: Exploring Co-occurrence Signals in Skeleton-based Continuous Sign Language Recognition	20619
<i>Peiqi Jiao (Institute of Computing Technology, Chinese Academy of Sciences), Yuecong Min (Institute of Computing Technology, Chinese Academy of Sciences), Yanan Li (Xiaomi), Xiaotao Wang (Xiaomi), Lei Lei (Xiaomi), and Xilin Chen (Institute of Computing Technology, Chinese Academy of Sciences)</i>	

EmoTalk: Speech-Driven Emotional Disentanglement for 3D Face Animation	20630
<i>Ziqiao Peng (Renmin University of China), Haoyu Wu (Renmin University of China), Zhenbo Song (Nanjing University of Science and Technology), Hao Xu (HKUST), Xiangyu Zhu (Chinese Academy of Sciences), Jun He (Renmin University of China), Hongyan Liu (Tsinghua University, China), and Zhaoxin Fan (Renmin University of China)</i>	
LA-Net: Landmark-Aware Learning for Reliable Facial Expression Recognition under Label Noise	20641
<i>Zhiyu Wu (Peking University) and Jinshi Cui (Peking University, China)</i>	
ASM: Adaptive Skinning Model for High-Quality 3D Face Modeling	20651
<i>Kai Yang (Tencent), Hong Shang (Tencent AI Lab), Tianyang Shi (Tencent), Xinghan Chen (Tencent medical center), Jingkai Zhou (South China University of Technology), Zhongqian Sun (Tencent AI Lab), and Wei Yang (Tencent AI Lab)</i>	
Troubleshooting Ethnic Quality Bias with Curriculum Domain Adaptation for Face Image Quality Assessment	20661
<i>Fu-Zhao Ou (City University of Hong Kong), Baoliang Chen (City University of Hongkong), Chongyi Li (Nanyang Technological University), Shiqi Wang (City University of Hong Kong), and Sam Kwong (City Univeristy of Hong Kong)</i>	
UniFace: Unified Cross-Entropy Loss for Deep Face Recognition	20673
<i>Jiancan Zhou (Aqara), Xi Jia (University of Birmingham), Qiufu Li (shenzhen university), Linlin Shen (Shenzhen University), and Jinming Duan (University of Birmingham)</i>	
Human Part-wise 3D Motion Context Learning for Sign Language Recognition	20683
<i>Taeryung Lee (Seoul National University), Yeonguk Oh (Seoul National University), and Kyoung Mu Lee (Seoul National University)</i>	
Weakly-Supervised Text-Driven Contrastive Learning for Facial Behavior Understanding	20694
<i>Xiang Zhang (State University of New York at Binghamton), Taoyue Wang (State Univeristy of New York at Binghamton), Xiaotian Li (Binghamton University), Huiyuan Yang (Rice University), and Lijun Yin (State University of New York at Binghamton)</i>	
HaMuCo: Hand Pose Estimation via Multiview Collaborative Self-Supervised Learning	20706
<i>Xiaozheng Zheng (ByteDance), Chao Wen (Bytedance), Zhou Xue (Bytedance), Pengfei Ren (Beijing University of Posts and Telecommunications), and Jingyu Wang (Beijing University of Posts and Telecommunications)</i>	
ReactionNet: Learning High-Order Facial Behavior from Universal Stimulus-Reaction by Dyadic Relation Reasoning	20717
<i>Xiaotian Li (Binghamton University), Taoyue Wang (State Univeristy of New York at Binghamton), Geran Zhao (Binghamton University), Xiang Zhang (State University of New York at Binghamton), Xi Kang (Binghamton University), and Lijun Yin (State University of New York at Binghamton)</i>	

CLIP-Cluster: CLIP-Guided Attribute Hallucination for Face Clustering	20729
<i>Shuai Shen (Tsinghua University), Wanhua Li (Harvard University), Xiaobing Wang (Samsung Research China-Beijing), Dafeng Zhang (Samsung Research China – Beijing, (SRCB), Zhezhu Jin (Samsung Research Institute China – Beijing, (SRC-B), Jie Zhou (Tsinghua University), and Jiwen Lu (Tsinghua University)</i>	
Learning Human Dynamics in Autonomous Driving Scenarios	20739
<i>Jingbo Wang (The Chinese University of HongKong), Ye Yuan (NVIDIA Research), Zhengyi Luo (Carnegie Mellon University), Kevin Xie (University of Toronto), Dahua Lin (The Chinese University of Hong Kong), Umar Iqbal (NVIDIA Research), Sanja Fidler (University of Toronto, NVIDIA), and Sameh Khamis (Nvidia)</i>	
LivelySpeaker: Towards Semantic-Aware Co-Speech Gesture Generation	20750
<i>Yihao Zhi (ShanghaiTech University), Xiaodong Cun (Tencent AI Lab), Xuelin Chen (Tencent AI Lab), Xi Shen (École des Ponts ParisTech), Wen Guo (INRIA), Shaoli Huang (Tencent AI-Lab), and Shenghua Gao (ShanghaiTech University)</i>	
Controllable Guide-Space for Generalizable Face Forgery Detection	20761
<i>Ying Guo (Meituan), Cheng Zhen (Meituan), and Pengfei Yan (Meituan)</i>	
Unpaired Multi-domain Attribute Translation of 3D Facial Shapes with a Square and Symmetric Geometric Map	20771
<i>Zhenfeng Fan (Institute of computing technology, Chinese Academy of Sciences), Zhiheng Zhang (Institute of Computing Technology, Chinese Academy of Sciences), Shuang Yang (Institute of Computing Technology, Chinese Academy of Sciences), Chongyang Zhong (Institute of Computing Technology, Chinese Academy of Sciences), Min Cao (Soochow University), and Shihong Xia (institute of computing technology of the Chinese academy of sciences)</i>	
Emotional Listener Portrait: Neural Listener Head Generation with Emotion	20782
<i>Luchuan Song (Universtiy of Rochester), Guojun Yin (University of Science and Technology of China), Zhenchao Jin (The University of Hong Kong), Xiaoyi Dong (University of Science and Technology of China), and Chenliang Xu (University of Rochester)</i>	
Steered Diffusion: A Generalized Framework for Plug-and-Play Conditional Image Synthesis	20793
<i>Nithin Gopalakrishnan Nair (Johns Hopkins University), Anoop Cherian (MERL), Suhas Lohit (Mitsubishi Electric Research Laboratories), Ye Wang (Mitsubishi Electric Research Laboratories), Toshiaki Koike-Akino (Mitsubishi Electric Research Laboratories), Vishal M. Patel (Johns Hopkins University), and Tim K. Marks (Mitsubishi Electric Research Laboratories, USA)</i>	
Invariant Feature Regularization for Fair Face Recognition	20804
<i>Jiali Ma (Panasonic R&D Center Singapore), Zhongqi Yue (Nanyang Technological University), Kagaya Tomoyuki (Panasonic Connect Corporation), Suzuki Tomoki (Panasonic Connect), Karlekar Jayashree (Panasonic), Sugiri Pranata (Panasonic R&D Center Singapore), and Hanwang Zhang (Nanyang Technological University)</i>	

Gloss-Free Sign Language Translation: Improving from Visual-Language Pretraining	20814
<i>Benjia Zhou (Macau University of Science and Technology), Zhigang Chen (CASIA), Albert Clapés (Aalborg University), Jun Wan (NLPR, CASIA), Yanyan Liang (Macau University of Science and Technology, Macao SAR, China), Sergio Escalera (Computer Vision Center, (UAB & University of Barcelona), Zhen Lei (NLPR, CASIA, China), and Du Zhang (Macau University of Science and Technology)</i>	
Contrastive Pseudo Learning for Open-World DeepFake Attribution	20825
<i>Zhimin Sun (Shanghai Jiao Tong University), Shen Chen (Tencent YouTu Lab), Taiping Yao (Tencent YouTu), Bangjie Yin (Tencent), Ran Yi (Shanghai Jiao Tong University), Shouhong Ding (Tencent), and Lizhuang Ma (Shanghai Jiao Tong University)</i>	
Continual Learning for Personalized Co-speech Gesture Generation	20836
<i>Chaitanya Ahuja (Meta AI), Pratik Joshi (Carnegie Mellon University), Ryo Ishii (NTT), and Louis-Philippe Morency (Carnegie Mellon University)</i>	
HandR2N2: Iterative 3D Hand Pose Estimation Using a Residual Recurrent Neural Network	20847
<i>Wencan Cheng (Sungkyunkwan University) and Jong Hwan Ko (Sungkyunkwan University)</i>	
SPACE: Speech-driven Portrait Animation with Controllable Expression	20857
<i>Siddharth Gururani (NVIDIA), Arun Mallya (NVIDIA), Ting-Chun Wang (NVIDIA), Rafael Valle (NVIDIA), and Ming-Yu Liu (NVIDIA)</i>	
How to Boost Face Recognition with StyleGAN?	20867
<i>Artem Sevastopolskiy (Technical University of Munich), Yury Malkov (Twitter), Nikita Durasov (École polytechnique fédérale de Lausanne), Luisa Verdoliva (University Federico II of Naples), and Matthias Nießner (Technical University of Munich)</i>	
ChildPlay: A New Benchmark for Understanding Children's Gaze Behaviour	20878
<i>Samy Tafasca (Idiap Research Institute), Anshul Gupta (Idiap and EPFL), and Jean-Marc Odobez (IDIAP/EPFL, SWITZERLAND)</i>	
Robust One-Shot Face Video Re-enactment using Hybrid Latent Spaces of StyleGAN2	20890
<i>Trevine Oorloff (University of Maryland - College Park) and Yaser Yacoob (UMD)</i>	
Data-Free Class-Incremental Hand Gesture Recognition	20901
<i>Shubhra Aich (Carnegie Mellon University), Jesus Ruiz-Santaquiteria (UCLM), Zhenyu Lu (Carnegie Mellon University), Prachi Garg (International Institute of Information Technology, (IIIT), K J Joseph (Adobe Research), Alvaro Fernandez Garcia (Carnegie Mellon University), Vineeth N Balasubramanian (Indian Institute of Technology, Hyderabad), Kenrick Kin (Meta Reality Labs), Chengde Wan (Meta Reality Labs), Necati Cihan Camgoz (University of Surrey), Shugao Ma (Meta Reality Labs), and Fernando De la Torre (Carnegie Mellon)</i>	
Learning Robust Representations with Information Bottleneck and Memory Network for RGB-D-based Gesture Recognition	20911
<i>Yunan Li (Xidian University), Huizhou Chen (Xidian University), Guanwen Feng (Xidian University), and Qiguang Miao (Xidian University)</i>	

Knowledge-Spreader: Learning Semi-Supervised Facial Action Dynamics by Consistifying Knowledge Granularity	20922
<i>Xiaotian Li (Binghamton University), Xiang Zhang (State University of New York at Binghamton), Taoyue Wang (State University of New York at Binghamton), and Lijun Yin (State University of New York at Binghamton)</i>	
Face Clustering via Graph Convolutional Networks with Confidence Edges	20933
<i>Yang Wu (Institute of Automation, Chinese Academy of Sciences), Zhiwei Ge (JD.com), Yuhao Luo (JD.com), Lin Liu (JD.com), and Sulong Xu (JD.com)</i>	
StyleGANEX: StyleGAN-Based Manipulation Beyond Cropped Aligned Faces	20943
<i>Shuai Yang (Nanyang Technological University), Liming Jiang (Nanyang Technological University), Ziwei Liu (Nanyang Technological University), and Chen Change Loy (Nanyang Technological University)</i>	
SeeABLE: Soft Discrepancies and Bounded Contrastive Learning for Exposing Deepfakes	20954
<i>Nicolas Larue (ETIS), Ngoc-Son Vu (ENSEA), Vitomir Struc (University of Ljubljana), Peter Peer (University of Ljubljana), and Vassilis Christophides (ETIS)</i>	
Adaptive Nonlinear Latent Transformation for Conditional Face Editing	20965
<i>Zhizhong Huang (Fudan University), Siteng Ma (Fudan University), Junping Zhang (Fudan University), and Hongming Shan (Fudan University)</i>	
Semi-supervised Speech-driven 3D Facial Animation via Cross-modal Encoding	20975
<i>Peiji Yang (Tencent Technology, (Shenzhen Co.Ltd), Huawei Wei (Tencent), Yicheng Zhong (Tencent), and Zhisheng Wang (Tencent)</i>	
ICD-Face: Intra-class Compactness Distillation for Face Recognition	20985
<i>Zhipeng Yu (University of Chinese Academy of Sciences), Jiaheng Liu (Beihang University), Haoyu Qin (SenseTime), Yichao Wu (SenseTime Group Limited), Kun Hu (senseTime), Jiayi Tian (Beihang University), and Ding Liang (SenseTime Group Limited)</i>	
C2ST: Cross-Modal Contextualized Sequence Transduction for Continuous Sign Language Recognition	20996
<i>Huaiwen Zhang (Inner Mongolia University), Zihang Guo (Inner Mongolia University), Yang Yang (Inner Mongolia University), Xin Liu (Inner Mongolia University), and De Hu (Inner Mongolia University)</i>	

Medical and Biological Vision; Cell Microscopy

CO-PILOT: Dynamic Top-Down Point Cloud with Conditional Neighborhood Aggregation for Multi-Gigapixel Histopathology Image Representation	21006
<i>Ramin Nakhli (The University of British Columbia), Allen Zhang (University of British Columbia), Ali Mirabadi (University of British Columbia), Katherine Rich (AIM Lab), Maryam Asadi (UBC), Blake Gilks (Vancouver General Hospital), Hossein Farahani (UBC), and Ali Bashashati (The University of British Columbia)</i>	

SKiT: a Fast Key Information Video Transformer for Online Surgical Phase Recognition	21017
<i>Yang Liu (King's College London), Jiayu Huo (King's College London), Jingjing Peng (Kings college london), Rachel Sparks (King's College London), Prokar Dasgupta (King's College London), Alejandro Granados (King's College London), and Sebastien Ourselin (King's College London)</i>	
XNet: Wavelet-Based Low and High Frequency Fusion Networks for Fully- and Semi-Supervised Semantic Segmentation of Biomedical Images	21028
<i>Yanfeng Zhou (Institute of Automation, Chinese Academy of Sciences), Jiaxing Huang (CASIA), Chenlong Wang (BioMap, (Beijing), Le Song (Georgia Institute of Technology & Ant Financial), and Ge Yang (National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences)</i>	
Probabilistic Modeling of Inter- and Intra-observer Variability in Medical Image Segmentation	21040
<i>Arne Schmidt (Universidad de Granada), Pablo Morales-Álvarez Morales-Álvarez (Universidad de Granada), and Rafael Molina (University of Granada, Spain)</i>	
Learning Cross-Representation Affinity Consistency for Sparsely Supervised Biomedical Instance Segmentation	21050
<i>Xiaoyu Liu (University of Science and Technology of China), Wei Huang (University of Science and Technology of China), Zhiwei Xiong (University of Science and Technology of China), Shenglong Zhou (University of Science and Technology of China), Yueyi Zhang (University of Science and Technology of China), Xuejin Chen (University of Science and Technology of China), Zheng-Jun Zha (University of Science and Technology of China), and Feng Wu (University of Science and Technology of China)</i>	
Dual Meta-Learning with Longitudinally Consistent Regularization for One-Shot Brain Tissue Segmentation Across the Human Lifespan	21061
<i>Yongheng Sun (Xi'an Jiaotong University), Fan Wang (Xi'an Jiaotong University), Jun Shu (Xi'an Jiaotong University), Haifeng Wang (Xi'an Jiaotong University), Li Wang (UNC), Deyu Meng (Xi'an Jiaotong University), and Chunfeng Lian (Xi'an Jiaotong University)</i>	
BlindHarmony: "Blind" Harmonization for MR Images via Flow Model	21072
<i>Hwihun Jeong (Seoul national university), Heejoon Byun (Seoul National University), Dong Un Kang (Seoul National University), and Jongho Lee (Seoul National University)</i>	
Continual Segment: Towards a Single, Unified and Non-forgetting Continual Segmentation Model of 143 Whole-body Organs in CT Scans	21083
<i>Zhanghexuan Ji (University at Buffalo, SUNY), Dazhou Guo (Alibaba DAMO Academy USA), Puyang Wang (Damo Academy, Alibaba Group), Ke Yan (Alibaba DAMO Academy), Le Lu (Alibaba Group), Minfeng Xu (Alibaba), Qifeng Wang (Sichuan Cancer Hospital and Institution, Sichuan Cancer Center, Radiation Oncology Key Laboratory of Sichuan Province), Jia Ge (Zhejiang University), Mingchen Gao (University at Buffalo, SUNY), Xianghua Ye (Zhejiang University), and Dakai Jin (Alibaba USA Inc.)</i>	

CLIP-Driven Universal Model for Organ Segmentation and Tumor Detection	21095
<i>Jie Liu (City University of Hong Kong), Yixiao Zhang (Johns Hopkins University), Jie-Neng Chen (Johns Hopkins University), Junfei Xiao (Johns Hopkins University), Yongyi Lu (Guangdong University of Technology), Bennett A. Landman (Vanderbilt University), Yixuan Yuan (Chinese University of Hong Kong), Alan Yuille (Johns Hopkins University), Yucheng Tang (Vanderbilt University), and Zongwei Zhou (Johns Hopkins University)</i>	
LIMITR: Leveraging Local Information for Medical Image-Text Representation	21108
<i>Gefen Dawidowicz (Technion), Elad Hirsch (Technion), and Ayellet Tal (Technion)</i>	
Taxonomy Adaptive Cross-Domain Adaptation in Medical Imaging via Optimization Trajectory Distillation	21117
<i>Jianan Fan (University of Sydney), Dongnan Liu (University of Sydney), Hang Chang (Lawrence Berkeley National Laboratory), Heng Huang (University of Maryland at College Park), Mei Chen (Microsoft), and Weidong Cai (University of Sydney)</i>	
CuNeRF: Cube-Based Neural Radiance Field for Zero-Shot Medical Image Arbitrary-Scale Super Resolution	21128
<i>Zixuan Chen (Sun Yat-sen University), Lingxiao Yang (Sun Yat-sen University), Jian-Huang Lai (Sun Yat-sen University), and Xiaohua Xie (Sun Yat-sen University)</i>	
Learning to Distill Global Representation for Sparse-View CT	21139
<i>Zilong Li (Fudan University), Chenglong Ma (Fudan University), Jie Chen (Fudan University), Junping Zhang (Fudan University), and Hongming Shan (Fudan University)</i>	
Preserving Tumor Volumes for Unsupervised Medical Image Registration	21151
<i>Qihua Dong (City University of Hong Kong), Hao Du (City University of Hong Kong), Ying Song (Chinese Academy of Medical Sciences and Peking Union Medical College), Yan Xu (Beihang University), and Jing Liao (City University of Hong Kong)</i>	
uSplit: Image Decomposition for Fluorescence Microscopy	21162
<i>Ashesh Ashesh (Fondazione Human Technopole), Alexander Krull (University of Birmingham), Moises Di Sante (University of Pavia), Francesco Silvio Pasqualini (University of Pavia), and Florian Jug (Human Technopole/ MPI-CBG)</i>	
Rethinking Multi-Contrast MRI Super-Resolution: Rectangle-Window Cross-Attention Transformer and Arbitrary-Scale Upsampling	21173
<i>Guangyuan Li (Zhejiang University), Lei Zhao (Zhejiang University), Jiakai Sun (Zhejiang University), Zehua Lan (Zhejiang University), Zhanjie Zhang (Zhejiang University), Jiafu Chen (Zhejiang University), Zhijie Lin (Zhejiang University of Science and Technology), Huaizhong Lin (Zhejiang University), and Wei Xing (Zhejiang University)</i>	
Multimodal Optimal Transport-based Co-Attention Transformer with Global Structure Consistency for Survival Prediction	21184
<i>Yingxue Xu (The Hong Kong university of Science and Technology) and Hao Chen (The Hong Kong University of Science and Technology)</i>	

4D Myocardium Reconstruction with Decoupled Motion and Shape Model	21195
<i>Xiaohan Yuan (School of Automation, Southeast University), Cong Liu (Southeast University), and Yangang Wang (Southeast University)</i>	
Unsupervised Learning of Object-Centric Embeddings for Cell Instance Segmentation in Microscopy Images	21206
<i>Steffen Wolf (MRC Laboratory of Molecular Biology), Manan Lalit (MPI-CBG), Katie McDole (MRC Laboratory of Molecular Biology), and Jan Funke (HHMI Janelia Research Campus)</i>	
LightDepth: Single-View Depth Self-Supervision from Illumination Decline	21216
<i>Javier Rodriguez-Puigvert (Universidad de Zaragoza), Victor M. Batlle (I3A. Universidad de Zaragoza), J.M.M. J.M.M. Montiel (I3A. Universidad de Zaragoza), Ruben Martinez-Cantin (University of Zaragoza), Pascal Fua (EPFL, Switzerland), Juan D. Tardós (I3A. Universidad de Zaragoza), and Javier Civera (Universidad de Zaragoza)</i>	
BoMD: Bag of Multi-label Descriptors for Noisy Chest X-ray Classification	21227
<i>Yuanhong Chen (University of Adelaide), Fengbei Liu (University of Adelaide), Hu Wang (the University of Adelaide), Chong Wang (Australian Institute for Machine Learning, The University of Adelaide), Yuyuan Liu (University of Adelaide), Yu Tian (Harvard University), and Gustavo Carneiro (University of Surrey)</i>	
Decomposition-Based Variational Network for Multi-Contrast MRI Super-Resolution and Reconstruction	21239
<i>Pengcheng Lei (East China Normal University), Faming Fang (East China Normal University), Guixu Zhang (East China Normal University), and Tiejong Zeng (The Chinese University of Hong Kong)</i>	
TopoSeg: Topology-Aware Nuclear Instance Segmentation	21250
<i>Hongliang He (School of Electronic and Computer Engineering, Peking University), Jun Wang (School of Electronic and Computer Engineering, Peking University), Pengxu Wei (Sun Yat-sen University), Fan Xu (Peng Cheng Laboratory), Xiangyang Ji (Tsinghua University), Chang Liu (Tsinghua University), and Jie Chen (Peking University)</i>	
Scratch Each Other's Back: Incomplete Multi-Modal Brain Tumor Segmentation via Category Aware Group Self-Support Learning	21260
<i>Yansheng Qiu (Wuhan University), Delin Chen (Wuhan University), Hongdou Yao (Wuhan University), Yongchao Xu (School of Computer Science, Wuhan University), and Zheng Wang (Wuhan University)</i>	

CancerUniT: Towards a Single Unified Model for Effective Detection, Segmentation, and Diagnosis of Eight Major Cancers Using a Large Collection of CT Scans	21270
<i>Jieneng Chen (Johns Hopkins University), Yingda Xia (Alibaba Group), Jiawen Yao (DAMO Academy, Alibaba Group), Ke Yan (Alibaba DAMO Academy), Jianpeng Zhang (Alibaba DAMO Academy), Le Lu (Alibaba Group), Fakai Wang (University of Maryland, College Park), Bo Zhou (Yale University), Mingyan Qiu (Fudan university), Qihang Yu (Johns Hopkins University), Mingze Yuan (Peking University), Wei Fang (DAMO Academy, Alibaba Group), Yuxing Tang (Alibaba USA Inc.), Minfeng Xu (Alibaba), Jian Zhou (Sun Yat-sen University Cancer Center), Yuqian Zhao (Sichuan Cancer Center), Qifeng Wang (Sichuan Cancer Hospital and Institution, Sichuan Cancer Center, Radiation Oncology Key Laboratory of Sichuan Province), Xianghua Ye (Zhejiang University), Xiaoli Yin (China Medical University), Yu Shi (Shengjing Hospital), Xin Chen (Guangzhou First People's Hospital), Jimgren Zhou (missing), Alan Yuille (Johns Hopkins University), Zaiyi Liu (Department of Radiology, Guangdong General Hospital, Guangdong Academy of Medical Science), and Ling Zhang (Alibaba USA Inc.)</i>	
Gram-based Attentive Neural Ordinary Differential Equations Network for Video Nystagmography Classification	21282
<i>Xihe Qiu (Shanghai University of Engineering Science), Shaojie Shi (Shanghai University of Engineering Science), Xiaoyu Tan (INF Technology, (shanghai), Chao Qu (Ant Financial Services Group), Zhijun Fang (Shanghai university of engineering science), Hailing Wang (Shanghai University of Engineering Science), Yongbin Gao (Shanghai university of engineering science), Peixia Wu (Fudan University), and Huawei Li (Fudan University)</i>	
ConSlide: Asynchronous Hierarchical Interaction Transformer with Breakup-Reorganize Rehearsal for Continual Whole Slide Image Analysis	21292
<i>Yanyan Huang (Zhejiang University), Weiqin Zhao (Department of Statistics and Actuarial Science, The University of Hong Kong), Shujun Wang (PolyU), Yu Fu (Zhejiang University), Yuming Jiang (Stanford University), and Lequan Yu (The University of Hong Kong)</i>	
PRIOR: Prototype Representation Joint Learning from Medical Images and Reports	21304
<i>Pujin Cheng (Southern University of Science and Technology), Li Lin (Southern University of Science and Technology, The University of Hong Kong), Junyan Lyu (Southern University of Science and Technology), Yijin Huang (Southern University of Science and Technology), Wenhan Luo (Sun Yat-sen University), and Xiaoying Tang (Southern University of Science and Technology)</i>	
MedKLIP: Medical Knowledge Enhanced Language-Image Pre-Training for X-ray Diagnosis	21315
<i>Chaoyi Wu (Shanghai Jiao Tong University), Xiaoman Zhang (Shanghai Jiao Tong University), Ya Zhang (Cooperative Medianet Innovation Center, Shanghai Jiao Tong University), Yanfeng Wang (Cooperative medianet innovation center of Shanghai Jiao Tong University), and Weidi Xie (Shanghai Jiao Tong University)</i>	

Affine-Consistent Transformer for Multi-Class Cell Nuclei Detection	21327
<i>Junjia Huang (Sun Yat-Sen University), Haofeng Li (Shenzhen Research Institute of Big Data, The Chinese University of Hong Kong, (Shenzhen), Xiang Wan (Shenzhen Research Institute of Big Data, the Chinese University of Hong Kong, (Shenzhen), and Guambin Li (Sun Yat-sen University)</i>	
A Skeletonization Algorithm for Gradient-Based Optimization	21337
<i>Martin J. Menten (Imperial College London), Johannes C. Paetzold (TUM), Veronika A. Zimmer (Technical University Munich), Suprosanna Shit (TUM), Ivan Ezhov (TUM), Robbie Holland (Imperial College London), Monika Probst (Technical University Munich), Julia A. Schnabel (Technical University of Munich / Helmholtz Center Munich), and Daniel Rueckert (Technische Universität München)</i>	
Improving Representation Learning for Histopathologic Images with Cluster Constraints	21347
<i>Weiyi Wu (Dartmouth College), Chongyang Gao (Northwestern University), Joseph DiPalma (Dartmouth College), Soroush Vosoughi (Dartmouth College), and Saeed Hassanpour (Dartmouth College)</i>	
Enhancing Modality-Agnostic Representations via Meta-Learning for Brain Tumor Segmentation.....	21358
<i>Aishik Konwer (Stony Brook University), Xiaoling Hu (Stony Brook University), Joseph Bae (Stony Brook University), Xuan Xu (Stony Brook University), Chao Chen (Stony Brook University), and Prateek Prasanna (Stony Brook University)</i>	
CauSSL: Causality-inspired Semi-supervised Learning for Medical Image Segmentation	21369
<i>Juzheng Miao (The Chinese University of Hong Kong), Cheng Chen (Harvard Medical School, Massachusetts General Hospital), Furui Liu (Zhejiang Lab), Hao Wei (The Chinese University of Hong Kong), and Pheng-Ann Heng (The Chinese University of Hong Kong)</i>	
UniverSeg: Universal Medical Image Segmentation	21381
<i>Victor Ion Butoi (MIT), Jose Javier Gonzalez Ortiz (MIT), Tianyu Ma (Cornell University), Mert R. Sabuncu (Cornell), John Guttag (MIT), and Adrian V. Dalca (MIT)</i>	
MRM: Masked Relation Modeling for Medical Image Pre-Training with Genetics	21395
<i>Qiushi Yang (City University of Hong Kong), Wuyang Li (City University of Hong Kong), Baopu Li (Independent Researcher), and Yixuan Yuan (Chinese University of Hong Kong)</i>	
Boosting Whole Slide Image Classification from the Perspectives of Distribution, Correlation and Magnification	21406
<i>Linhao Qu (Digital Medical Research Center, School of Basic Medical Science, Fudan University), Zhiwei Yang (Fudan University), Minghong Duan (Fudan University), Yingfan Ma (Fudan University), Shuo Wang (Digital Medical Reserach Center, Schoool of Basic Medical Sciences, Fudan University), Manning Wang (Fudan University), and Zhijian Song (Fudan University)</i>	

Adaptive Template Transformer for Mitochondria Segmentation in Electron Microscopy Images	21417
<i>Yuwen Pan (University of Science and Technology of China), Naisong Luo (University of Science and Technology of China), Rui Sun (University of Science and Technology of China), Meng Meng (University of Science and Technology of China), Tianzhu Zhang (University of Science and Technology of China), Zhiwei Xiong (University of Science and Technology of China), and Yongdong Zhang (University of Science and Technology of China)</i>	
Cross-Modal Translation and Alignment for Survival Analysis	21428
<i>Fengtao Zhou (The Hong Kong University of Science and Technology) and Hao Chen (The Hong Kong University of Science and Technology)</i>	
LNPL-MIL: Learning from Noisy Pseudo Labels for Promoting Multiple Instance Learning in Whole Slide Image	21438
<i>Zhuchen Shao (Tsinghua Shenzhen International Graduate School, Tsinghua University), Yifeng Wang (Harbin Institute of Technology, Shenzhen), Yang Chen (Tsinghua Shenzhen International Graduate School, Tsinghua University), Hao Bian (Tsinghua Shenzhen International Graduate School, Tsinghua University), Shaohui Liu (Harbin Institute of Technology), Haoqian Wang (Tsinghua Shenzhen International Graduate School, Tsinghua University), and Yongbing Zhang (Harbin Institute of Technology, (Shenzhen))</i>	

Scene Analysis and Understanding

Generalized Few-Shot Point Cloud Segmentation via Geometric Words	21449
<i>Yating Xu (National University of Singapore), Conghui Hu (National University of Singapore), Na Zhao (SUTD), and Gim Hee Lee (National University of Singapore)</i>	
Boosting 3-DoF Ground-to-Satellite Camera Localization Accuracy via Geometry-Guided Cross-View Transformer	21459
<i>Yujiao Shi (ANU), Fei Wu (Australian National University), Akhil Perincherry (Ford Motor Company), Ankit Vora (Ford Motor Company), and Hongdong Li (Australian National University, Australia)</i>	
EP2P-Loc: End-to-End 3D Point to 2D Pixel Localization for Large-Scale Visual Localization.....	21470
<i>Minjung Kim (Seoul National University), Junseo Koo (Seoul National University), and Gunhee Kim (Seoul National University)</i>	
Multi-task View Synthesis with Neural Radiance Fields	21481
<i>Shuhong Zheng (University of Illinois at Urbana-Champaign), Zhipeng Bao (Carnegie Mellon University), Martial Hebert (Carnegie Mellon School of Computer Science), and Yu-Xiong Wang (University of Illinois at Urbana-Champaign)</i>	
Multi-Task Learning with Knowledge Distillation for Dense Prediction	21493
<i>Yangyang Xu (Wuhan University), Yibo Yang (Peking University), and Lefei Zhang (Wuhan University)</i>	

Visually-Prompted Language Model for Fine-Grained Scene Graph Generation in an Open World	21503
<i>Qifan Yu (Zhejiang University), Juncheng Li (Zhejiang University), Yu Wu (Wuhan University), Siliang Tang (Zhejiang University), Wei Ji (National University of Singapore), and Yueting Zhuang (Zhejiang University)</i>	
CMDA: Cross-Modality Domain Adaptation for Nighttime Semantic Segmentation	21515
<i>Ruihao Xia (East China University of Science and Technology), Chaoqiang Zhao (East China University of Science and Technology), Meng Zheng (Rensselaer Polytechnic Institute), Ziyang Wu (United Imaging Intelligence), Qiyu Sun (East China University of Science and Technology), and Yang Tang (East China University of Science and Technology)</i>	
VQA-GNN: Reasoning with Multimodal Knowledge via Graph Neural Networks for Visual Question Answering	21525
<i>Yanan Wang (KDDI Research), Michihiro Yasunaga (Stanford University), Hongyu Ren (Stanford University), Shinya Wada (KDDI Research, Inc.), and Jure Leskovec (Stanford University)</i>	
Disentangle then Parse: Night-time Semantic Segmentation with Illumination Disentanglement.	21536
<i>Zhixiang Wei (University of Science and Technology of China), Lin Chen (University of Science and Technology of China), Tao Tu (University of Science and Technology of China), Pengyang Ling (University of Science and Technology of China), Huaian Chen (University of Science and Technology of China), and Yi Jin (University of Science and Technology of China)</i>	
Visual Traffic Knowledge Graph Generation from Scene Images	21547
<i>Yunfei Guo (Institute of Automation, Chinese Academy of Sciences), Fei Yin (Institute of Automation of Chinese Academy of Sciences), Xiao-hui Li (National Laboratory of Pattern Recognition, Institute of Automation of Chinese Academy of Sciences, University of Chinese Academy of Sciences), Xudon Yan (Tencent), Tao Xue (tencent), Shuqi Mei (Tencent), and Cheng-Lin Liu (Institute of Automation of Chinese Academy of Sciences)</i>	
Agglomerative Transformer for Human-Object Interaction Detection	21557
<i>Danyang Tu (Shanghai Jiao Tong University), Wei Sun (Shanghai Jiao Tong University), Guangtao Zhai (Shanghai Jiao Tong University), and Wei Shen (SJTU)</i>	
3D Neural Embedding Likelihood: Probabilistic Inverse Graphics for Robust 6D Pose Estimation	21568
<i>Guangyao Zhou (DeepMind), Nishad Gothoskar (MIT), Lirui Wang (MIT), Joshua B. Tenenbaum (MIT), Dan Gutfreund (IBM), Miguel Lázaro-Gredilla (DeepMind), Dileep George (Vicarious), and Vikash K. Mansinghka (Massachusetts Institute of Technology)</i>	
HiLo: Exploiting High Low Frequency Relations for Unbiased Panoptic Scene Graph Generation.....	21580
<i>Zijian Zhou (King's College London), Miaojing Shi (King's College London), and Holger Caesar (Motional)</i>	

RLIPv2: Fast Scaling of Relational Language-Image Pre-Training	21592
<i>Hangjie Yuan (Zhejiang University), Shiwei Zhang (DAMO Academy, Alibaba Group), Xiang Wang (Huazhong University of Science and Technology), Samuel Albanie (University of Cambridge), Yining Pan (Singapore University of Technology and Design), Tao Feng (Sichuan University), Jianwen Jiang (Tsinghua University), Dong Ni (Zhejiang University), Yingya Zhang (Alibaba Group), and Deli Zhao (Alibaba Group)</i>	
UniSeg: A Unified Multi-Modal LiDAR Segmentation Network and the OpenPCSeg Codebase ..	21605
<i>Youquan Liu (Hochschule Bremerhaven), Runnan Chen (The University of Hong Kong), Xin Li (East China Normal University), Lingdong Kong (National University of Singapore), Yuchen Yang (Fudan University), Zhaoyang Xia (Shanghai AI Laboratory), Yeqi Bai (Nanyang Technological University), Xinge Zhu (The Chinese University of Hong Kong), Yuexin Ma (ShanghaiTech University), Yikang Li (Shanghai AI Lab), Yu Qiao (Shanghai AI Laboratory), and Yuenan Hou (Shanghai AI Lab)</i>	
See More and Know More: Zero-shot Point Cloud Segmentation via Multi-modal Visual Data ...	21617
<i>Yuhang Lu (ShanghaiTech University), Qi Jiang (ShanghaiTech University), Runnan Chen (The University of Hong Kong), Yuenan Hou (Shanghai AI Lab), Xinge Zhu (The Chinese University of Hong Kong), and Yuexin Ma (ShanghaiTech University)</i>	
Compositional Feature Augmentation for Unbiased Scene Graph Generation	21628
<i>Lin Li (Zhejiang University), Guikun Chen (Zhejiang University), Jun Xiao (Zhejiang University), Yi Yang (Zhejiang University), Chunping Wang (FinVolution), and Long Chen (The Hong Kong University of Science and Technology)</i>	
Multi-weather Image Restoration via Domain Translation	21639
<i>Prashant W. Patil (Deakin University), Sunil Gupta (Deakin University, Australia), Santu Rana (Deakin University, Australia), Svetha Venkatesh (Deakin University), and Subrahmanyam Murala (Trinity College Dublin)</i>	
CLIPTEER: Looking at the Bigger Picture in Scene Text Recognition	21649
<i>Aviad Aberdam (Amazon), David Bensaïd (Technion, Israel Institute of Technology), Alona Golts (Technion), Roy Ganz (Technion), Oren Nuriel (Amazon), Royce Tichauer (Amazon), Shai Mazor (Amazon), and Ron Litman (Amazon)</i>	
Towards Models that Can See and Read	21661
<i>Roy Ganz (Technion), Oren Nuriel (Amazon), Aviad Aberdam (Amazon), Yair Kittenplon (Tel-Aviv University), Shai Mazor (Amazon), and Ron Litman (Amazon)</i>	
SurroundOcc: Multi-camera 3D Occupancy Prediction for Autonomous Driving	21672
<i>Yi Wei (Tsinghua University), Linqing Zhao (Tianjin University), Wenzhao Zheng (Tsinghua University), Zheng Zhu (Tsinghua University), Jie Zhou (Tsinghua University), and Jiwen Lu (Tsinghua University)</i>	

DDP: Diffusion Model for Dense Visual Prediction	21684
<i>Yuanfeng Ji (HKU), Zhe Chen (Nanjing University), Enze Xie (The University of Hong Kong), Lanqing Hong (Huawei Noah's Ark Lab), Xihui Liu (UC Berkeley), Zhaoqiang Liu (Huawei), Tong Lu (Nanjing University), Zhenguo Li (Huawei Noah's Ark Lab), and Ping Luo (The University of Hong Kong)</i>	
Understanding 3D Object Interaction from a Single Image	21696
<i>Shengyi Qian (University of Michigan) and David F. Fouhey (New York University)</i>	
ObjectSDF++: Improved Object-Compositional Neural Implicit Surfaces	21707
<i>Qianyi Wu (Monash University), Kaisiyuan Wang (University of Sydney), Kejie Li (University of Oxford), Jianmin Zheng (Nanyang Technological University), and Jianfei Cai (Monash University)</i>	
Improving Equivariance in State-of-the-Art Supervised Depth and Normal Predictors	21718
<i>Yuanyi Zhong (University of Illinois at Urbana-Champaign), Anand Bhattad (UIUC), Yu-Xiong Wang (University of Illinois at Urbana-Champaign), and David Forsyth (University of Illinois at Urbana-Champaign)</i>	
CrossMatch: Source-Free Domain Adaptive Semantic Segmentation via Cross-Modal Consistency Training	21729
<i>Yifang Yin (Institute for Infocomm Research, A*STAR), Wenmiao Hu (National University of Singapore), Zhenguang Liu (Zhejiang University), Guanfeng Wang (Grab), Shili Xiang (Institute for Infocomm Research, A*STAR), and Roger Zimmermann (NUS)</i>	
Semantic Attention Flow Fields for Monocular Dynamic Scene Decomposition	21740
<i>Yiqing Liang (Brown University), Eliot Laidlaw (Brown University), Alexander Meyerowitz (Brown University), Srinath Sridhar (Brown University), and James Tompkin (Brown University)</i>	
Holistic Geometric Feature Learning for Structured Reconstruction	21750
<i>Ziqiong Lu (Wuhan University), Linxi Huan (Wuhan University), Qiyuan Ma (Wuhan University), and Xianwei Zheng (The State Key Lab. LIESMARS, Wuhan University)</i>	
Scalable Multi-Temporal Remote Sensing Change Data Generation via Simulating Stochastic Change Process	21761
<i>Zhuo Zheng (Wuhan University), Shiqi Tian (Wuhan University), Ailong Ma (Wuhan University), Liangpei Zhang (Wuhan University), and Yanfei Zhong (Wuhan University)</i>	
TaskExpert: Dynamically Assembling Multi-Task Representations with Memorial Mixture-of-Experts	21771
<i>Hanrong Ye (The Hong Kong University of Science and Technology) and Dan Xu (The Hong Kong University of Science and Technology)</i>	
Thinking Image Color Aesthetics Assessment: Models, Datasets and Benchmarks	21781
<i>Shuai He (Beijing University of Posts and Telecommunications), Anlong Ming (Beijing University of Posts and Telecommunications), Yaqi Li (Beijing University of Posts and Telecommunications), Jinyuan Sun (Beijing University of Posts and Telecommunications), ShunTian Zheng (Beijing University of Posts and Telecommunications), and Huadong Ma (Beijing University of Posts and Telecommunications)</i>	

STEERER: Resolving Scale Variations for Counting and Localization via Selective Inheritance Learning	21791
<i>Tao Han (Shanghai AI Lab), Lei Bai (Shanghai AI Laboratory), Lingbo Liu (Hong Kong Polytechnic University), and Wanli Ouyang (The University of Sydney)</i>	
Object-aware Gaze Target Detection	21803
<i>Francesco Tonini (University of Trento), Nicola Dall'Asen (University of Trento), Cigdem Beyan (Università degli Studi di Trento), and Elisa Ricci (FBK - Technologies of Vision)</i>	
Weakly Supervised Referring Image Segmentation with Intra-Chunk and Inter-Chunk Consistency	21813
<i>Jungbeom Lee (Seoul National University), Sungjin Lee (Amazon Alexa), Jinseok Nam (Amazon), Seunghak Yu (MIT), Jaeyoung Do (Microsoft Research), and Tara Taghavi (Amazon Inc)</i>	
Vision Relation Transformer for Unbiased Scene Graph Generation	21825
<i>Gopika Sudhakaran (Technical University of Darmstadt), Devendra Singh Dhami (TU Darmstadt), Kristian Kersting (TU Darmstadt), and Stefan Roth (TU Darmstadt)</i>	
DDIT: Semantic Scene Completion via Deformable Deep Implicit Templates	21837
<i>Haoang Li (Technical University of Munich), Jinhui Dong (The Chinese University of Hong Kong), Binghui Wen (Technical University of Munich), Ming Gao (Technical University of Munich), Tianyu Huang (The Chinese University of Hong Kong), Yun-Hui Liu (CUHK), and Daniel Cremers (TU Munich)</i>	
DQS3D: Densely-matched Quantization-aware Semi-supervised 3D Detection	21848
<i>Huan-ang Gao (Tsinghua University), Beiwen Tian (Tsinghua University), Pengfei Li (Tsinghua University), Hao Zhao (Tsinghua University), and Guyue Zhou (Tsinghua University)</i>	
Shape Anchor Guided Holistic Indoor Scene Understanding	21859
<i>Mingyue Dong (Wuhan University), Linxi Huan (Wuhan University), Hanjiang Xiong (Wuhan University), Shuhan Shen (Institute of Automation, Chinese Academy of Sciences), and Xianwei Zheng (The State Key Lab. LIESMARS, Wuhan University)</i>	
SGAligner: 3D Scene Alignment with Scene Graphs	21870
<i>Sayan Deb Sarkar (ETH Zurich), Ondrej Miksik (Microsoft), Marc Pollefeys (ETH Zurich / Microsoft), Daniel Barath (ETH Zürich), and Iro Armeni (ETH Zurich)</i>	
Betrayed by Captions: Joint Caption Grounding and Generation for Open Vocabulary Instance Segmentation	21881
<i>Jianzong Wu (Peking University), Xiangtai Li (Nanyang Technological University), Henghui Ding (Nanyang Technological University), Xia Li (ETH Zurich), Guangliang Cheng (University of Liverpool), Yunhai Tong (Peking University), and Chen Change Loy (Nanyang Technological University)</i>	

Multimodal Learning

- SLAN: Self-Locator Aided Network for Vision-Language Understanding 21892
Jiang-Tian Zhai (Nankai University), Qi Zhang (Tencent), Tong Wu (Tencent Youtu Lab), Xing-Yu Chen (Youtu Lab), Jiang-Jiang Liu (Baidu Inc.), and Ming-Ming Cheng (Nankai University)
- Task-Oriented Multi-Modal Mutual Learning for Vision-Language Models 21902
Sifan Long (Jilin University), Zhen Zhao (University of Sydney), Junkun Yuan (Zhejiang University), Zichang Tan (Baidu Research), Jiangjiang Liu (Baidu Inc.), Luping Zhou (University of Sydney), Shengsheng Wang (Jilin University), and Jingdong Wang (Baidu)
- TinyCLIP: CLIP Distillation via Affinity Mimicking and Weight Inheritance 21913
Kan Wu (Sun Yat-sen University), Houwen Peng (Microsoft Research), Zhenghong Zhou (Huazhong university of Science and Technology), Bin Xiao (Microsoft), Mengchen Liu (Microsoft), Lu Yuan (Microsoft), Hong Xuan (Microsoft), Michael Valenzuela (Microsoft), Xi Xi (Stoehen Chen, (Microsoft), Xinggang Wang (Huazhong University of Science and Technology), Hongyang Chao (Sun Yat-sen University), and Han Hu (Microsoft Research Asia)
- In-Style: Bridging Text and Uncurated Videos with Style Transfer for Text-Video Retrieval 21924
Nina Shvetsova (Goethe University Frankfurt), Anna Kukleva (MPII), Bernt Schiele (MPI Informatics), and Hilde Kuehne (University of Bonn)
- Preserving Modality Structure Improves Multi-Modal Learning 21936
Sirnam Swetha (University of Central Florida), Mamshad Nayeem Rizve (University of Central Florida), Nina Shvetsova (Goethe University Frankfurt Germany; University of Bonn Germany), Hilde Kuehne (Goethe University Frankfurt Germany; University of Bonn Germany; MIT-IBM Watson AI Lab), and Mubarak Shah (University of Central Florida)
- Distribution-Aware Prompt Tuning for Vision-Language Models 21947
Eulrang Cho (Korea University), Jooyeon Kim (Korea University), and Hyunwoo J Kim (Korea University)
- SupFusion: Supervised LiDAR-Camera Fusion for 3D Object Detection 21957
Yiran Qin (The Chinese University of Hong Kong, Shenzhen), Chaoqun Wang (The Chinese University of Hong Kong, Shenzhen), Zijian Kang (Nio Inc), Ningning Ma (NIO), Zhen Li (The Chinese University of Hong Kong, Shenzhen), and Ruimao Zhang (The Chinese University of Hong Kong, Shenzhen)
- Distribution-Consistent Modal Recovering for Incomplete Multimodal Learning 21968
Yuanzhi Wang (Nanjing University of Science and Technology), Zhen Cui (Nanjing University of Science and Technology), and Yong Li (Nanjing University of Science and Technology)
- Fg-T2M: Fine-Grained Text-Driven Human Motion Generation via Diffusion Model 21978
Yin Wang (State Key Laboratory of Virtual Reality Technology and Systems, Beihang University), Zhiying Leng (State Key Laboratory of Virtual Reality Technology and Systems, Beihang University), Frederick W. B. Li (University of Durham), Shun-Cheng Wu (Technical University of Munich), and Xiaohui Liang (State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Zhongguancun Laboratory, Beijing)

Cross-Modal Orthogonal High-Rank Augmentation for RGB-Event Transformer-Trackers	21988
<i>Zhiyu Zhu (City University of Hong Kong), Junhui Hou (City University of Hong Kong, Hong Kong), and Dapeng Oliver Wu (City University of Hong Kong)</i>	
eP-ALM: Efficient Perceptual Augmentation of Language Models	21999
<i>Mustafa Shukor (Sorbonne University), Corentin Dancette (Sorbonne Universite), and Matthieu Cord (Sorbonne University)</i>	
Generating Visual Scenes from Touch	22013
<i>Fengyu Yang (University of Michigan), Jiacheng Zhang (University of Michigan), and Andrew Owens (University of Michigan)</i>	
Multimodal High-order Relation Transformer for Scene Boundary Detection	22024
<i>Xi Wei (University of Science and Technology of China), Zhangxiang Shi (University of Science and Technology of China), Tianzhu Zhang (University of Science and Technology of China), Xiaoyuan Yu (Huawei Cloud), and Lei Xiao (Huawei Cloud)</i>	
Muscles in Action	22034
<i>Mia Chiquier (Columbia University) and Carl Vondrick (Columbia University)</i>	
Self-Evolved Dynamic Expansion Model for Task-Free Continual Learning	22045
<i>Fei Ye (University of York) and Adrian G. Bors (University of York)</i>	
Multi-Event Video-Text Retrieval	22056
<i>Gengyuan Zhang (Ludwig-Maximilians-Universität München), Jisen Ren (LMU Munich), Jindong Gu (University of Oxford), and Volker Tresp (Siemens AG and Ludwig Maximilian University of Munich)</i>	
Referring Image Segmentation Using Text Supervision	22067
<i>Fang Liu (City University of Hong Kong), Yuhao Liu (City University of Hong Kong), Yuqiu Kong (Dalian University of Technology), Ke Xu (City University of Hong Kong), Lihe Zhang (Dalian University of Technology), Baocai Yin (Beijing University of Technology), Gerhard Hancke (City University of Hong Kong), and Rynson Lau (City University of Hong Kong)</i>	
Audio-Visual Deception Detection: DOLOS Dataset and Parameter-Efficient Crossmodal Learning	22078
<i>Xiaobao Guo (Nanyang Technological University), Nithish Muthuchamy Selvaraj (Nanyang Technological University), Zitong Yu (Great Bay University), Adams Wai-Kin Kong (Nanyang Technological University), Bingquan Shen (DSO), and Alex Kot (Nanyang Technological University)</i>	
EMMN: Emotional Motion Memory Network for Audio-driven Emotional Talking Face Generation ...	22089
<i>Shuai Tan (Shanghai Jiao Tong University), Bin Ji (Shanghai Jiao Tong University), and Ye Pan (Shanghai Jiaotong University)</i>	
CLIP2Point: Transfer CLIP to Point Cloud Classification with Image-Depth Pre-Training	22100
<i>Tianyu Huang (Harbin Institute of Technology), Bowen Dong (Harbin Institute of Technology), Yunhan Yang (The University of Hong Kong), Xiaoshui Huang (Shanghai AI Laboratory), Rynson W.H. Lau (City University of Hong Kong), Wanli Ouyang (The University of Sydney), and Wangmeng Zuo (Harbin Institute of Technology, China)</i>	

Speech2Lip: High-fidelity Speech to Lip Generation by Learning from a Short Video	22111
<i>Xiuzhe Wu (The University of Hong Kong), Pengfei Hu (Tsinghua University), Yang Wu (Tencent AI Lab), Xiaoyang Lyu (The University of Hong Kong), Yan-Pei Cao (Tencent), Ying Shan (Tencent), Wenming Yang (Tsinghua University), Zhongqian Sun (Tencent AI Lab), and Xiaojuan Qi (The University of Hong Kong)</i>	
GrowCLIP: Data-Aware Automatic Model Growing for Large-scale Contrastive Language-Image Pre-Training	22121
<i>Xinchi Deng (Sun Yat-sen University), Han Shi (The Hong Kong University of Science and Technology), Runhui Huang (Sun Yat-sen University), Changlin Li (University of Technology Sydney), Hang Xu (Huawei Noah's Ark Lab), Jianhua Han (Huawei Noah's Ark Lab), James Kwok (The Hong Kong University of Science and Technology), Shen Zhao (Sun Yat-Sen University), Wei Zhang (Noah's Ark Lab, Huawei Technologies), and Xiaodan Liang (Sun Yat-sen University)</i>	
A Retrospect to Multi-prompt Learning across Vision and Language	22133
<i>Ziliang Chen (Jinan University), Xin Huang (Sun Yat-sen University), Quanlong Guan (Jinan University), Liang Lin (Sun Yat-sen University), and Weiqi Luo (Jinan University)</i>	
ChartReader: A Unified Framework for Chart Derendering and Comprehension without Heuristic Rules	22145
<i>Zhi-Qi Cheng (Carnegie Mellon University), Qi Dai (Microsoft Research), and Alexander G. Hauptmann (Carnegie Mellon University)</i>	
Boosting Multi-modal Model Performance with Adaptive Gradient Modulation	22157
<i>Hong Li (School of Information Science and Technology, ShanghaiTech University), Xingyu Li (Shanghai Center for Brain Science and Brain-Inspired Technology), Pengbo Hu (University of Science and Technology of China), Yinuo Lei (School of Information Science and Technology, ShanghaiTech University), Chunxiao Li (Shanghai Jiao Tong University), and Yi Zhou (University of Science and Technology of China)</i>	
ViLLA: Fine-Grained Vision-Language Representation Learning from Real-World Data	22168
<i>Maya Varma (Stanford University), Jean-Benoit Delbrouck (Stanford University), Sarah Hooper (Stanford University), Akshay Chaudhari (Stanford University), and Curtis Langlotz (Stanford University)</i>	
Robust Referring Video Object Segmentation with Cyclic Structural Consensus	22179
<i>Xiang Li (Carnegie Mellon University), Jinglu Wang (Microsoft Research Asia), Xiaohao Xu (University of Michigan, Ann Arbor), Xiao Li (Microsoft Research Asia), Bhiksha Raj (Carnegie Mellon University), and Yan Lu (Microsoft Research Asia)</i>	
Fantasia3D: Disentangling Geometry and Appearance for High-quality Text-to-3D Content Creation	22189
<i>Rui Chen (South China University of Technology), Yongwei Chen (South China University of Technology), Ningxin Jiao (South China University of Technology), and Kui Jia (South China University of Technology)</i>	

CTP:Towards Vision-Language Continual Pretraining via Compatible Momentum Contrast and Topology Preservation 22200

Hongguang Zhu (Beijing Jiaotong university), Yunchao Wei (UTS), Xiaodan Liang (Sun Yat-sen University), Chunjie Zhang (Beijing Jiaotong University), and Yao Zhao (Beijing Jiaotong University)

Narrator: Towards Natural Control of Human-Scene Interaction Generation via Relationship Reasoning 22211

Haibiao Xuan (Tianjin University), Xiongzhen Li (Tianjin university), Jinsong Zhang (Tianjin University), Hongwen Zhang (Tsinghua University), Yebin Liu (Tsinghua University), and Kun Li (Tianjin University)

Human-in-the-Loop Computer Vision

Knowledge-Aware Federated Active Learning with Non-IID Data 22222

Yu-Tong Cao (University of Sydney), Ye Shi (ShanghaiTech University), Baosheng Yu (The University of Sydney), Jingya Wang (ShanghaiTech University), and Dacheng Tao (The University of Sydney)

SimpleClick: Interactive Image Segmentation with Simple Vision Transformers 22233

Qin Liu (UNC), Zhenlin Xu (UNC Chapel Hill), Gedas Bertasius (UNC Chapel Hill), and Marc Niethammer (UNC)

InterFormer: Real-time Interactive Image Segmentation 22244

You Huang (Xiamen University), Hao Yang (Xiamen University), Ke Sun (Xiamen University), Shengchuan Zhang (Xiamen University), Liujuan Cao (Xiamen University), Guannan Jiang (CATL), and Rongrong Ji (Xiamen University, China)

Interactive Class-Agnostic Object Counting 22255

Yifeng Huang (Stony Brook University), Viresh Ranjan (Stony Brook University), and Minh Hoai (Stony Brook University)

Agile Modeling: From Concept to Classifier in Minutes 22266

Otilia Stretcu (Google), Edward Vendrow (Stanford University), Kenji Hata (Google), Krishnamurthy Viswanathan (Google Research), Vittorio Ferrari (Google Research), Sasan Tavakkol (Google), Wenlei Zhou (Google), Aditya Avinash (Google Inc), Emming Luo (Google), Neil Gordon Alldrin (Google), MohammadHossein Bateni (Google Research), Gabriel Berger (Google), Andrew Bunner (Google), Chun-Ta Lu (Google Research), Javier Rey (Google), Giulia DeSalvo (Google Research), Ranjay Krishna (University of Washington), and Ariel Fuxman (Google)

TiDAL: Learning Training Dynamics for Active Learning 22278

Seong Min Kye (KAIST), Kwanghee Choi (Carnegie Mellon University), Hyeonmin Byun (Hyperconnect), and Buru Chang (Hyperconnect)

Image and Video Forensics

Pre-Training-Free Image Manipulation Localization through Non-Mutually Exclusive Contrastive Learning	22289
<i>Jizhe Zhou (Sichuan University), Xiaochen Ma (Sichuan University), Xia Du (Xiamen University of Technology), Ahmed Y. Alhammadi (Mohamed Bin Zayed University for Humanities), and Wentao Feng (Sichuan University)</i>	
VADER: Video Alignment Differencing and Retrieval	22300
<i>Alexander Black (University of Surrey), Simon Jenni (Adobe Research), Tu Bui (University of Surrey), Md. Mehrab Md. Mehrab Tanjim (University of California, San Diego), Stefano Petrangeli (Adobe), Ritwik Sinha (Adobe Research), Viswanathan Swaminathan (missing), and John Collomosse (Adobe Research)</i>	
PIRNet: Privacy-Preserving Image Restoration Network via Wavelet Lifting	22311
<i>Xin Deng (Beihang university), Chao Gao (Beihang University), and Mai Xu (BUAA)</i>	
Quality-Agnostic Deepfake Detection with Intra-model Collaborative Learning	22321
<i>Binh M. Le (Sungkyunkwan University) and Simon S. Woo (Sungkyunkwan University, (SKKU)</i>	
Towards Generic Image Manipulation Detection with Weakly-Supervised Self-Consistency Learning	22333
<i>Yuanhao Zhai (State University of New York at Buffalo), Tianyu Luan (State University of New York at Buffalo), David Doermann (University at Buffalo), and Junsong Yuan (State University of New York at Buffalo, USA)</i>	
CopyRNeRF: Protecting the CopyRight of Neural Radiance Fields	22344
<i>Ziyuan Luo (Hong Kong Baptist University), Qing Guo (A*STAR), Ka Chun Cheung (Nvidia), Simon See (NVIDIA AI Tech Centre), and Renjie Wan (Hong Kong Baptist University)</i>	
UCF: Uncovering Common Features for Generalizable Deepfake Detection	22355
<i>Zhiyuan Yan (The Chinese University of Hong Kong, Shenzhen, Shenzhen Research Institute of Big Data), Yong Zhang (Tencent AI Lab), Yanbo Fan (Tencent AI Lab), and Baoyuan Wu (The Chinese University of Hong Kong, Shenzhen, Shenzhen Research Institute of Big Data)</i>	
SAFL-Net: Semantic-Agnostic Feature Learning Network with Auxiliary Plugins for Image Manipulation Detection	22367
<i>Zhihao Sun (Institute of Computing Technology, Chinese Academy of Sciences, University of Chinese Academy of Sciences), Haoran Jiang (University of Chinese Academy of Sciences), Danding Wang (Institute of Computing Technology, Chinese Academy of Sciences; Research Institute of Intelligent Computing, Zhejiang Lab;), Xirong Li (Renmin University of China), and Juan Cao (Institute of Computing Technology, Chinese Academy of Sciences; University of Chinese Academy of Sciences)</i>	

DRAW: Defending Camera-shooted RAW Against Image Manipulation	22377
<i>Xiaoxiao Hu (School of Computer Science, Fudan University), Qichao Ying (School of Computer Science, Fudan University), Zhenxing Qian (School of Computer Science, Fudan University), Sheng Li (Fudan University), and Xinpeng Zhang (School of Computer Science, Fudan University)</i>	
DIRE for Diffusion-Generated Image Detection	22388
<i>Zhendong Wang (University of Science and Technology of China), Jianmin Bao (Microsoft Research Asia), Wengang Zhou (University of Science and Technology of China), Weilun Wang (University of Science and Technology of China), Hezhen Hu (University of Science and Technology of China), Hong Chen (MUCFC), and Houqiang Li (University of Science and Technology of China)</i>	
Uncertainty-guided Learning for Improving Image Manipulation Detection	22399
<i>Kaixiang Ji (Ant Group), Feng Chen (The University of Adelaide), Xin Guo (Ant Group), Yadong Xu (Tsinghua University), Jian Wang (Ant Group), and Jingdong Chen (Ant Group)</i>	
The Stable Signature: Rooting Watermarks in Latent Diffusion Models	22409
<i>Pierre Fernandez (Meta AI), Guillaume Couairon (Facebook AI Research), Hervé Jégou (Facebook AI Research), Matthijs Douze (Meta AI Research), and Teddy Furon (Inria)</i>	

Geometric Deep Learning

Get the Best of Both Worlds: Improving Accuracy and Transferability by Grassmann Class Representation	22421
<i>Haoqi Wang (SenseTime), Zhizhong Li (Sensetime), and Wayne Zhang (SenseTime Research)</i>	
4D Panoptic Segmentation as Invariant and Equivariant Field Prediction	22431
<i>Minghan Zhu (University of Michigan, Ann Arbor), Shizhong Han (Qualcomm AI Research), Hong Cai (Qualcomm AI Research), Shubhankar Borse (Qualcomm AI Research), Maani Ghaffari (Univ. of Michigan), and Fatih Porikli (Australian National University)</i>	
SiLK: Simple Learned Keypoints	22442
<i>Pierre Gleize (Meta), Weiyao Wang (Facebook AI Research), and Matt Feiszli (Facebook AI Research)</i>	
SC3K: Self-supervised and Coherent 3D Keypoints Estimation from Rotated, Noisy, and Decimated Point Cloud Data	22452
<i>Mohammad Zohaib (Italian Institute of Technology, (IIT) and Alessio Del Bue (Istituto Italiano di Tecnologia, (IIT)</i>	
Geometric Viewpoint Learning with Hyper-Rays and Harmonics Encoding	22463
<i>Zhixiang Min (Stevens Institute of Technology), Juan Carlos Dibene Simental (Stevens Institute of Technology), and Enrique Dunn (Stevens Institute of Technology)</i>	

Surface Extraction from Neural Unsigned Distance Fields	22474
<i>Congyi Zhang (The University of Hong Kong), Guying Lin (The University of Hong Kong), Lei Yang (The University of Hong Kong), Xin Li (Texas A&M University), Taku Komura (The University of Hong Kong), Scott Schaefer (Texas A&M University), John Keyser (Texas A&M University), and Wenping Wang (Texas A&M University)</i>	
Learning Adaptive Neighborhoods for Graph Neural Networks	22484
<i>Avishkar Saha (University of Surrey), Oscar Mendez (University of Surrey), Chris Russell (University of Oxford), and Richard Bowden (University of Surrey)</i>	
Why do networks have inhibitory/negative connections?	22494
<i>Qingyang Wang (Johns Hopkins University), Mike A. Powell (Johns Hopkins University), Ali Geisa (Johns Hopkins University), Eric Bridgeford (Johns Hopkins University), Carey E. Priebe (Johns Hopkins University), and Joshua T. Vogelstein (Johns Hopkins University)</i>	

Image and Video Synthesis 2

MasaCtrl: Tuning-Free Mutual Self-Attention Control for Consistent Image Synthesis and Editing	22503
<i>Mingdeng Cao (The University of Tokyo), Xintao Wang (Tencent), Zhongqiang Qi (Tencent), Ying Shan (Tencent), Xiaohu Qie (Tencent), and Yinqiang Zheng (The University of Tokyo)</i>	
Personalized Image Generation for Color Vision Deficiency Population	22514
<i>Shuyi Jiang (University of Sydney), Daochang Liu (University of Sydney), Dingquan Li (Peng Cheng Laboratory), and Chang Xu (University of Sydney)</i>	
ReNeRF: Relightable Neural Radiance Fields with Nearfield Lighting	22524
<i>Yingyan Xu (ETH Zurich), Gaspard Zoss (DisneyResearch; Studios), Prashanth Chandran (ETH Zurich), Markus Gross (ETH Zurich), Derek Bradley (DisneyResearch; Studios), and Paulo Gotardo (Google)</i>	
MagicFusion: Boosting Text-to-Image Generation Performance by Fusing Diffusion Models	22535
<i>Jing Zhao (Institute for Quantum Information & State Key Laboratory of High Performance Computing, College of Computer, National University of Defense Technology), Heliang Zheng (USTC), Chaoyue Wang (JD.com), Long Lan (NUDT), and Wenjing Yang (National University of Defense Technology)</i>	
PODIA-3D: Domain Adaptation of 3D Generative Model Across Large Domain Gap Using Pose-Preserved Text-to-Image Diffusion	22546
<i>Gwanghyun Kim (Department of Electrical and Computer Engineering, Seoul National University), Ji Ha Jang (Seoul National University), and Se Young Chun (Seoul National University)</i>	
Pluralistic Aging Diffusion Autoencoder	22556
<i>Peipei Li (Beijing University of Posts and Telecommunications), Rui Wang (Beijing University of Posts and Telecommunications), Huaibo Huang (Institute of Automation, Chinese Academy of Sciences), Ran He (Institute of Automation, Chinese Academy of Sciences), and Zhaofeng He (Beijing University of Posts and Telecommunications)</i>	

DPM-OT: A New Diffusion Probabilistic Model Based on Optimal Transport	22567
<i>Zezen Li (Dalian University of Technology), Shenghao Li (Dalian University of Technology), Zhanpeng Wang (Dalian University of Technology), Na Lei (Dalian University of Technology), Zhongxuan Luo (Dalian University of Technology), and David Xianfeng Gu (Stony Brook University)</i>	
Efficient Emotional Adaptation for Audio-Driven Talking-Head Generation	22577
<i>Yuan Gan (Zhejiang University), Zongxin Yang (Zhejiang University), Xihang Yue (Zhejiang University), Lingyun Sun (Zhejiang University), and Yi Yang (Zhejiang University)</i>	
DiFaReli: Diffusion Face Relighting	22589
<i>Puntawat Ponglertnapakorn (Vidyasirimedhi Institute of Science and Technology), Nontawat Tritrong (Vidyasirimedhi Institute of Science and Technology), and Supasorn Suwajanakorn (Vistec)</i>	
TALL: Thumbnail Layout for Deepfake Video Detection	22601
<i>Yuting Xu (Institute of Automation, Chinese Academy of Sciences), Jian Liang (CASIA), Gengyun Jia (Nanjing University of Posts and Telecommunications), Ziming Yang (Institute of Information Engineering, Chinese Academy of Sciences), Yanhao Zhang (OPPO Research Institute), and Ran He (Institute of Automation, Chinese Academy of Sciences)</i>	
LAW-Diffusion: Complex Scene Generation by Diffusion with Layouts	22612
<i>Binbin Yang (Sun Yat-sen University), Yi Luo (Sun Yat-sen University), Ziliang Chen (Jinan University), Guangrun Wang (University of Oxford), Xiaodan Liang (Sun Yat-sen University), and Liang Lin (Sun Yat-sen University)</i>	
DreamPose: Fashion Video Synthesis with Stable Diffusion	22623
<i>Johanna Karras (University of Washington), Aleksander Holynski (UC Berkeley / Google Research), Ting-Chun Wang (NVIDIA), and Ira Kemelmacher-Shlizerman (University of Washington)</i>	
Ablating Concepts in Text-to-Image Diffusion Models	22634
<i>Nupur Kumari (Carnegie Mellon University), Bingliang Zhang (Tsinghua University), Sheng-Yu Wang (Carnegie Mellon University), Eli Shechtman (Adobe Research, US), Richard Zhang (Adobe), and Jun-Yan Zhu (Carnegie Mellon University)</i>	
DReg-NeRF: Deep Registration for Neural Radiance Fields	22646
<i>Yu Chen (National University of Singapore) and Gim Hee Lee (National University of Singapore)</i>	
The Euclidean Space is Evil: Hyperbolic Attribute Editing for Few-shot Image Generation	22657
<i>Lingxiao Li (Columbia University), Yi Zhang (University of Oxford), and Shuhui Wang (VIPL, ICT, Chinese academic of science)</i>	
Discriminative Class Tokens for Text-to-Image Diffusion Models	22668
<i>Idan Schwartz (Technion), Vésteinn Snæbjarnarson (University of Copenhagen), Hila Chefer (Tel Aviv University), Serge Belongie (University of Copenhagen), Lior Wolf (Tel Aviv University, Israel), and Sagie Benaïm (University of Copenhagen)</i>	

General Image-to-Image Translation with One-Shot Image Guidance	22679
<i>Bin Cheng (NetEase Games AI Lab), Zuhao Liu (NetEase Games AI Lab), Yunbo Peng (NetEase Games AI Lab), and Yue Lin (NetEase Games AI Lab)</i>	
Text2Performer: Text-Driven Human Video Generation	22690
<i>Yuming Jiang (Nanyang Technological University), Shuai Yang (Nanyang Technological University), Tong Liang Koh (Nanyang Technological University), Wayne Wu (Shanghai AI Laboratory), Chen Change Loy (Nanyang Technological University), and Ziwei Liu (Nanyang Technological University)</i>	
AesPA-Net: Aesthetic Pattern-Aware Style Transfer Networks	22701
<i>Kibeom Hong (Yonsei University; SwatchOn; NAVER WEBTOON AI), Seogkyu Jeon (Yonsei University), Junsoo Lee (NAVER WEBTOON AI), Namhyuk Ahn (NAVER WEBTOON AI), Kunhee Kim (KAIST AI; NAVER WEBTOON AI), Pilhyeon Lee (Yonsei University), Daesik Kim (NAVER WEBTOON AI), Youngjung Uh (Yonsei University), and Hyeran Byun (Yonsei University)</i>	
Controllable Person Image Synthesis with Pose-Constrained Latent Diffusion	22711
<i>Xiao Han (University of Surrey; iFlyTek-Surrey Joint Research Centre on AI), Xiatian Zhu (University of Surrey; Surrey Institute for People-Centred AI), Jiankang Deng (Imperial College London), Yi-Zhe Song (University of Surrey; iFlyTek-Surrey Joint Research Centre on AI), and Tao Xiang (University of Surrey; iFlyTek-Surrey Joint Research Centre on AI)</i>	
PATMAT: Person Aware Tuning of Mask-Aware Transformer for Face Inpainting	22721
<i>Saman Motamed (Carnegie Mellon University, Pittsburgh, PA; Sofia University, Bulgaria), Jianjin Xu (Carnegie Mellon University, Pittsburgh, PA), Chen Henry Wu (Carnegie Mellon University, Pittsburgh, PA), Christian Häne (Independent Researcher), Jean-Charles Bazin (Independent Researcher), and Fernando De la Torre (Carnegie Mellon University, Pittsburgh, PA)</i>	
Virtual Try-On with Pose-Garment Keypoints Guided Inpainting	22731
<i>Zhi Li (Nanyang Technological University), Pengfei Wei (AI Lab, Bytedance), Xiang Yin (ByteDance), Zejun Ma (Bytedance), and Alex C. Kot (Nanyang Technological University)</i>	
Online Clustered Codebook	22741
<i>Chuanxia Zheng (University of Oxford) and Andrea Vedaldi (Oxford University)</i>	
InfiniCity: Infinite-Scale City Synthesis	22751
<i>Chieh Hubert Lin (University of California at Merced), Hsin-Ying Lee (Snap Inc), Willi Menapace (University of Trento), Menglei Chai (Google), Aliaksandr Siarohin (SNAP), Ming-Hsuan Yang (University of California at Merced), and Sergey Tulyakov (Snap Inc)</i>	
Make-It-3D: High-fidelity 3D Creation from A Single Image with Diffusion Prior	22762
<i>Junshu Tang (Shanghai Jiao Tong University), Tengfei Wang (HKUST), Bo Zhang (Microsoft Research Asia), Ting Zhang (MSRA), Ran Yi (Shanghai Jiao Tong University), Lizhuang Ma (Shanghai Jiao Tong University), and Dong Chen (Microsoft Research Asia)</i>	

SAMPLING: Scene-adaptive Hierarchical Multiplane Images Representation for Novel View Synthesis from a Single Image	22773
<i>Xiaoyu Zhou (Peking University), Zhiwei Lin (Peking University), Xiaojun Shan (Peking University), Yongtao Wang (Peking University), Deqing Sun (Google), and Ming-Hsuan Yang (University of California at Merced)</i>	
StyleLipSync: Style-based Personalized Lip-sync Video Generation	22784
<i>Taekyung Ki (AITRICS) and Dongchan Min (KAIST)</i>	
StyleInV: A Temporal Style Modulated Inversion Network for Unconditional Video Generation	22794
<i>Yuhan Wang (Nanyang Technological University), Liming Jiang (Nanyang Technological University), and Chen Change Loy (Nanyang Technological University)</i>	
3D-Aware Generative Model for Improved Side-View Image Synthesis	22805
<i>Kyungmin Jo (Korea Advanced Institute of Science and Technology), Wonjoon Jin (POSTECH), Jaegul Choo (Korea Advanced Institute of Science and Technology), Hyunjoon Lee (Kakao Brain), and Sunghyun Cho (POSTECH)</i>	
Zero-Shot Contrastive Loss for Text-Guided Diffusion Image Style Transfer	22816
<i>Serin Yang (KAIST), Hyunmin Hwang (KAIST), and Jong Chul Ye (Kim Jaechul Graduate School of AI, KAIST, Korea)</i>	
FlipNeRF: Flipped Reflection Rays for Few-shot Novel View Synthesis	22826
<i>Seunghyeon Seo (Seoul National University), Yeonjin Chang (Seoul National University), and Nojun Kwak (Seoul National University)</i>	
Inverse Problem Regularization with Hierarchical Variational Autoencoders	22837
<i>Jean Prost (Institut de Mathématiques de Bordeaux, IMB), Antoine Houdard (Ubisoft), Andrés Almansa (CNRS & Université Paris Cité, MAP5), and Nicolas Papadakis (CNRS/IMB)</i>	
3D-aware Blending with Generative NeRFs	22849
<i>Hyunsu Kim (NAVER AI Lab), Gayoung Lee (NAVER AI Lab), Yunjey Choi (NAVER AI Lab), Jin-Hwa Kim (NAVER AI Lab), and Jun-Yan Zhu (Carnegie Mellon University)</i>	
NeMF: Inverse Volume Rendering with Neural Microflake Field	22862
<i>Youjia Zhang (Huazhong University of Science and Technology), Teng Xu (Huazhong University of Science and Technology), Junqing Yu (Huazhong University of Science & Technology), Yuteng Ye (Huazhong University of Science & Technology), Yanqing Jing (Tencent), Junle Wang (Tencent), Jingyi Yu (Shanghai Tech University), and Wei Yang (Huazhong University of Science and Technology)</i>	
Preserve Your Own Correlation: A Noise Prior for Video Diffusion Models	22873
<i>Songwei Ge (University of Maryland, College Park), Seungjun Nah (NVIDIA), Guilin Liu (NVIDIA), Tyler Poon (NVIDIA), Andrew Tao (NVIDIA), Bryan Catanzaro (NVIDIA), David Jacobs (University of Maryland, USA), Jia-Bin Huang (University of Maryland College Park), Ming-Yu Liu (NVIDIA), and Yogesh Balaji (Nvidia)</i>	
iVS-Net: Learning Human View Synthesis from Internet Videos	22885
<i>Junting Dong (Zhejiang University), Qi Fang (NetEase), Tianshuo Yang (Zhejiang University), Qing Shuai (Zhejiang University), Chengyu Qiao (Zhejiang University), and Sida Peng (Zhejiang University)</i>	

EGC: Image Generation and Classification via a Diffusion Energy-Based Model	22895
<i>Qiushan Guo (The University of Hong Kong), Chuofan Ma (The University of Hong Kong), Yi Jiang (Bytedance), Zehuan Yuan (Bytedance.Inc), Yizhou Yu (The University of Hong Kong), and Ping Luo (The University of Hong Kong)</i>	
Automatic Animation of Hair Blowing in Still Portrait Photos	22906
<i>Wenpeng Xiao (South China University of Technology), Wentao Liu (ByteDance), Yitong Wang (Bytedance), Bernard Ghanem (KAUST), and Bing Li (KAUST)</i>	
HoloFusion: Towards Photo-realistic 3D Generative Modeling	22919
<i>Animesh Karnear (University College London), Niloy J. Mitra (university college london), Andrea Vedaldi (Meta and University of Oxford), and David Novotny (Facebook AI Research)</i>	
Foreground Object Search by Distilling Composite Image Feature	22929
<i>Bo Zhang (Shanghai Jiao Tong University), Jiacheng Sui (Xian Jiao Tong University), and Li Niu (Shanghai Jiao Tong University)</i>	
OrthoPlanes: A Novel Representation for Better 3D-Awareness of GANs	22939
<i>Honglin He (Shanghai AI Lab), Zhuoqian Yang (EPFL), Shikai Li (SenseTime Research), Bo Dai (Shanghai AI Lab), and Wayne Wu (Shanghai AI Laboratory)</i>	
3DHumanGAN: 3D-Aware Human Image Generation with 3D Pose Mapping	22951
<i>Zhuoqian Yang (EPFL), Shikai Li (SenseTime Research), Wayne Wu (Shanghai AI Laboratory), and Bo Dai (Shanghai AI Lab)</i>	
MODA: Mapping-Once Audio-driven Portrait Animation with Dual Attentions	22963
<i>Yunfei Liu (International Digital Economy Academy), Lijian Lin (Tencent), Fei Yu (Vistring), Changyin Zhou (Vistring), and Yu Li (International Digital Economy Academy)</i>	
Minimum Latency Deep Online Video Stabilization	22973
<i>Zhuofan Zhang (University of Electronic Science and Technology of China), Zhen Liu (Megvii Technology Inc.), Ping Tan (Hong Kong University of Science and Technology), Bing Zeng (University of Electronic Science and Technology of China), and Shuaicheng Liu (UESTC, Megvii)</i>	
StableVideo: Text-driven Consistency-aware Diffusion Video Editing	22983
<i>Wenhao Chai (University of Washington), Xun Guo (Microsoft Research Asia), Gaoang Wang (Zhejiang University), and Yan Lu (Microsoft Research Asia)</i>	
Localizing Object-Level Shape Variations with Text-to-Image Diffusion Models	22994
<i>Or Patashnik (Tel Aviv University), Daniel Garibi (Tel Aviv University), Idan Azuri (Meta), Hadar Averbuch-Elor (Tel Aviv University), and Daniel Cohen-Or (Tel Aviv University)</i>	
Implicit Identity Representation Conditioned Memory Compensation Network for Talking Head video Generation	23005
<i>Fa-Ting Hong (The Hong Kong University of Science and Technology) and Dan Xu (The Hong Kong University of Science and Technology)</i>	

ESSAformer: Efficient Transformer for Hyperspectral Image Super-resolution	23016
<i>Mingjin Zhang (Xidian University), Chi Zhang (Xidian University), Qiming Zhang (The University of Sydney), Jie Guo (Xidian University), Xinbo Gao (Chongqing University of Posts and Telecommunications), and Jing Zhang (The University of Sydney)</i>	
GlueGen: Plug and Play Multi-modal Encoders for X-to-image Generation	23028
<i>Can Qin (Northeastern University), Ning Yu (Salesforce Research), Chen Xing (Salesforce Research), Shu Zhang (Salesforce Research), Zeyuan Chen (Salesforce Research), Stefano Ermon (Stanford University), Yun Fu (Northeastern University), Caiming Xiong (Salesforce Research), and Ran Xu (Salesforce Research)</i>	
UHDNeRF: Ultra-High-Definition Neural Radiance Fields	23040
<i>Quewei Li (National Key Lab for Novel Software Technology, Nanjing University), Feichao Li (Nanjing University), Jie Guo (Nanjing University), and Yanwen Guo (Nanjing University)</i>	
All-to-Key Attention for Arbitrary Style Transfer	23052
<i>Mingrui Zhu (Xidian University), Xiao He (Xidian University), Nannan Wang (Xidian University), Xiaoyu Wang (Hong Kong University of Science and Technology, (HKUST)), and Xinbo Gao (Chongqing University of Posts and Telecommunications)</i>	
Diverse Inpainting and Editing with GAN Inversion	23063
<i>Ahmet Burak Yildirim (Bilkent University), Hamza Pehlivan (Bilkent University), Bahri Batuhan Bilecen (Aselsan Research), and Aysegul Dundar (Bilkent University)</i>	
MoTIF: Learning Motion Trajectories with Local Implicit Neural Functions for Continuous Space-Time Video Super-Resolution	23074
<i>Yi-Hsin Chen (National Yang Ming Chiao Tung University), Si-Cun Chen (NYCU), Yi-Hsin Chen (National Yang Ming Chiao Tung University), Yen-Yu Lin (National Yang Ming Chiao Tung University), and Wen-Hsiao Peng (National Yang Ming Chiao Tung University)</i>	
RANA: Relightable Articulated Neural Avatars	23085
<i>Umar Iqbal (NVIDIA Research), Akin Caliskan (Center for Vision Speech and Signal Processing - University of Surrey), Koki Nagano (NVIDIA), Sameh Khamis (Nvidia), Pavlo Molchanov (NVIDIA), and Jan Kautz (NVIDIA)</i>	
DiffCloth: Diffusion Based Garment Synthesis and Manipulation via Structural Cross-modal Semantic Alignment	23097
<i>Xujie Zhang (Sun Yat-sen University), Binbin Yang (Sun Yat-sen University), Michael C. Kampffmeyer (UiT The Arctic University of Norway), Wenqing Zhang (Sun Yat-sen University), Shiyue Zhang (Sun Yat-sen University), Guansong Lu (Huawei Noah's Ark Lab), Liang Lin (Sun Yat-sen University), Hang Xu (Huawei Noah's Ark Lab), and Xiaodan Liang (Sun Yat-sen University)</i>	
Masked Diffusion Transformer is a Strong Image Synthesizer	23107
<i>Shanghua Gao (Nankai University), Pan Zhou (Sea AI Lab), Ming-Ming Cheng (Nankai University), and Shuicheng Yan (National University of Singapore, Department of Electrical and Computer Engineering)</i>	

FreeDoM: Training-Free Energy-Guided Conditional Diffusion Model	23117
<i>Jiwen Yu (Peking University), Yinhuai Wang (Peking University Shenzhen Graduate School), Chen Zhao (KAUST), Bernard Ghanem (KAUST), and Jian Zhang (Peking University)</i>	
CLNeRF: Continual Learning Meets NeRF	23128
<i>Zhipeng Cai (Intel Labs) and Matthias Müller (Intel Labs)</i>	
Rethinking Fast Fourier Convolution in Image Inpainting	23138
<i>Tianyi Chu (Zhejiang University), Jiafu Chen (Zhejiang University), Jiakai Sun (Zhejiang University), Shuobin Lian (Zhejiang University), Zhizhong Wang (Zhejiang University), Zhiwen Zuo (Zhejiang Gongshang University), Lei Zhao (Zhejiang University), Wei Xing (Zhejiang University), and Dongming Lu (Zhejiang University)</i>	
Pix2Video: Video Editing using Image Diffusion	23149
<i>Duygu Ceylan (Adobe Research), Chun-Hao P. Huang (Adobe), and Niloy J. Mitra (university college london)</i>	
Multi-view Spectral Polarization Propagation for Video Glass Segmentation	23161
<i>Yu Qiao (Dalian University of Technology), Bo Dong (Princeton University), Ao Jin (Dalian University of Technology), Yu Fu (Dalian University of Technology), Seung-Hwan Baek (POSTECH), Felix Heide (Princeton / Algolux), Pieter Peers (College of William & Mary), Xiaopeng Wei (Dalian University of Technology), and Xin Yang (Dalian University of Technology)</i>	
WALDO: Future Video Synthesis Using Object Layer Decomposition and Parametric Flow Prediction	23172
<i>Guillaume Le Moing (Inria), Jean Ponce (Inria), and Cordelia Schmid (Inria/Google)</i>	
Ray Conditioning: Trading Photo-consistency for Photo-realism in Multi-view Image Generation	23185
<i>Eric Ming Chen (Cornell University), Sidhanth Holalkere (Cornell University), Ruyun Yan (Cornell University), Kai Zhang (Adobe Research), and Abe Davis (Cornell University)</i>	
Text-Conditioned Sampling Framework for Text-to-Image Generation with Masked Generative Models	23195
<i>Jaewoong Lee (KAIST), Sangwon Jang (Yonsei University), Jaehyeong Jo (KAIST), Jaehong Yoon (KAIST), Yunji Kim (NAVER AI Lab), Jin-Hwa Kim (NAVER AI Lab), Jung-Woo Ha (NAVER Cloud AI Lab), and Sung Ju Hwang (KAIST, AITRICS)</i>	
Efficient Video Prediction via Sparsely Conditioned Flow Matching	23206
<i>Aram Davtyan (University of Bern), Sepehr Sameni (University of Bern), and Paolo Favaro (University of Bern)</i>	

Vision Applications and Systems

Democratising 2D Sketch to 3D Shape Retrieval Through Pivoting	23218
<i>Pinaki Nath Chowdhury (University of Surrey), Ayan Kumar Bhunia (University of Surrey), Aneeshan Sain (University of Surrey), Subhadeep Koley (University of Surrey), Tao Xiang (University of Surrey), and Yi-Zhe Song (University of Surrey)</i>	

Towards Instance-adaptive Inference for Federated Learning	23230
<i>Chun-Mei Feng (Institute of High Performance Computing, A*STAR), Kai Yu (Agency for Science, Technology and Research, (A*STAR), Nian Liu (MBZUAI), Xinxing Xu (Institute of High Performance Computing, A*Star), Salman Khan (MBZUAI), and Wangmeng Zuo (Harbin Institute of Technology, China)</i>	
TransTIC: Transferring Transformer-based Image Compression from Human Perception to Machine Perception	23240
<i>Yi-Hsin Chen (National Yang Ming Chiao Tung University), Ying-Chieh Weng (National Yang Ming Chiao Tung University), Chia-Hao Kao (National Yang Ming Chiao Tung University), Cheng Chien (National Yang Ming Chiao Tung University), Wei-Chen Chiu (National Chiao Tung University), and Wen-Hsiao Peng (National Yang Ming Chiao Tung University)</i>	
Counting Crowds in Bad Weather	23251
<i>Zhi-Kai Huang (National Taiwan University), Wei-Ting Chen (National Taiwan University), Yuan-Chun Chiang (National Taiwan University), Sy-Yen Kuo (National Taiwan University), and Ming-Hsuan Yang (University of California at Merced)</i>	
NeRF-Det: Learning Geometry-Aware Volumetric Representation for Multi-View 3D Object Detection	23263
<i>Chenfeng Xu (UC Berkeley), Bichen Wu (Facebook Research), Ji Hou (Meta Reality Labs), Sam Tsai (Facebook), Ruilong Li (UC Berkeley), Jialiang Wang (Facebook Inc.), Wei Zhan (University of California, Berkeley), Zijian He (Facebook), Peter Vajda (Facebook), Kurt Keutzer (EECS, UC Berkeley), and Masayoshi Tomizuka (MSC Lab)</i>	
MEGA: Multimodal Alignment Aggregation and Distillation For Cinematic Video Segmentation	23274
<i>Najmeh Sadoughi (Amazon), Xinyu Li (Amazon), Avijit Vajpayee (Amazon), David Fan (Amazon), Bing Shuai (Amazon), Hector Santos-Villalobos (Amazon Inc), Vimal Bhat (Amazon), and Rohith MV (Amazon Lab126)</i>	
Bring Clipart to Life	23284
<i>Nanxuan Zhao (Adobe Research), Shengqi Dang (Tongji University), Hexun Lin (Imperial College London), Yang Shi (Tongji University), and Nan Cao (Tongji University)</i>	
UpCycling: Semi-supervised 3D Object Detection without Sharing Raw-level Unlabeled Scenes .	23294
<i>Sunwook Hwang (Seoul National University), Youngseok Kim (Seoul National University), Seongwon Kim (SK Telecom), Saewoong Bahk (Seoul National University), and Hyung-Sin Kim (Seoul National University)</i>	
Graph Matching with Bi-level Noisy Correspondence	23305
<i>Yijie Lin (Sichuan University), Mouxing Yang (Sichuan University), Jun Yu (Hangzhou Dianzi University), Peng Hu (Sichuan University), Changqing Zhang (Tianjin University), and Xi Peng (Sichuan University)</i>	

Anomaly Detection using Score-based Perturbation Resilience	23315
<i>Woosang Shin (Korea Institute of Industrial Technology), Jonghyeon Lee (Korea Institute of Industrial Technology), Taehan Lee (Korea Institute of Industrial Technology), Sangmoon Lee (Kyungpook National University), and Jong Pil Yun (Korea Institute of Industrial Technology)</i>	
Spatio-Temporal Domain Awareness for Multi-Agent Collaborative Perception	23326
<i>Kun Yang (Fudan University), Dingkang Yang (Fudan University), Jingyu Zhang (Fudan University), Mingcheng Li (Fudan University), Yang Liu (Fudan University), Jing Liu (Fudan University), Hanqi Wang (Fudan University), Peng Sun (Duke Kunshan University), and Liang Song (Fudan University)</i>	
Multimodal Garment Designer: Human-Centric Latent Diffusion Models for Fashion Image Editing	23336
<i>Alberto Baldrati (Università degli Studi di Firenze), Davide Morelli (UNIMORE), Giuseppe Cartella (University of Modena and Reggio Emilia), Marcella Cornia (University of Modena and Reggio Emilia), Marco Bertini (University of Florence), and Rita Cucchiara (Università di Modena e Reggio Emilia)</i>	
Towards Unifying Medical Vision-and-Language Pre-Training via Soft Prompts	23346
<i>Zhihong Chen (Shenzhen Research Institute of Big Data, The Chinese University of Hong Kong, Shenzhen), Shizhe Diao (The Hong Kong University of Science and Technology), Benyou Wang (The Chinese University of Hong Kong, Shenzhen), Guanbin Li (Sun Yat-sen University), and Xiang Wan (Shenzhen Research Institute of Big Data, the Chinese University of Hong Kong, Shenzhen)</i>	
MAS: Towards Resource-Efficient Federated Multiple-Task Learning	23357
<i>Weiming Zhuang (Nanyang Technological University), Yonggang Wen (Nanyang Technological University), Lingjuan Lyu (Sony AI), and Shuai Zhang (SenseTime Ltd)</i>	
Hierarchical Visual Categories Modeling: A Joint Representation Learning and Density Estimation Framework for Out-of-Distribution Detection	23368
<i>Jinglun Li (Fudan University), Xinyu Zhou (Fudan University), Pinxue Guo (Fudan University), Yixuan Sun (Fudan University), Yiwen Huang (Fudan University), Weifeng Ge (Fudan University), and Wenqiang Zhang (Fudan University)</i>	
Improving Generalization in Visual Reinforcement Learning via Conflict-aware Gradient Agreement Augmentation	23379
<i>Siao Liu (Fudan University), Zhaoyu Chen (Fudan University), Yang Liu (Fudan University), Yuzheng Wang (Fudan University), Dingkang Yang (Fudan University), Zhile Zhao (Fudan University), Ziqing Zhou (Fudan University), Xie Yi (Fudan University), Wei Li (Academy for Engineering and Technology, Fudan University, Shanghai, China), Wenqiang Zhang (Fudan University), and Zhongxue Gan (Fudan University)</i>	
Tiny Updater: Towards Efficient Neural Network-Driven Software Updating	23390
<i>Linfeng Zhang (Tsinghua University) and Kaisheng Ma (Tsinghua University)</i>	

Multiple Planar Object Tracking	23403
<i>Zhicheng Zhang (Nankai University), Shengzhe Liu (Nankai University), and Jufeng Yang (Nankai University)</i>	
OmnimatteRF: Robust Omnimatte with 3D Background Modeling	23414
<i>Geng Lin (University of Maryland, College Park), Chen Gao (Meta), Jia-Bin Huang (University of Maryland College Park), Changil Kim (Facebook), Yipeng Wang (Meta), Matthias Zwicker (University of Maryland), and Ayush Saraf (Meta Platforms)</i>	
Ordinal Label Distribution Learning	23424
<i>Changsong Wen (Nankai University), Xin Zhang (Nankai University), Xingxu Yao (Didi Chuxing), and Jufeng Yang (Nankai University)</i>	
Re-mine, Learn and Reason: Exploring the Cross-modal Semantic Correlations for Language-guided HOI detection	23435
<i>Yichao Cao (Southeast University), Qingfei Tang (Nanjing Enbo Technology Co., LTD), Feng Yang (southeast university), Xiu Su (University of Sydney), Shan You (SenseTime), Xiaobo Lu (Southeast University), and Chang Xu (University of Sydney)</i>	
MUVA: A New Large-Scale Benchmark for Multi-View Amodal Instance Segmentation in the Shopping Scenario	23447
<i>Zhixuan Li (Peking University), Weining Ye (Peking University), Juan Terven (AiFi Inc), Zachary Bennett (AiFi), Ying Zheng (Stanford), Tingting Jiang (Peking University), and Tiejun Huang (Peking University)</i>	
Editable Image Geometric Abstraction via Neural Primitive Assembly	23457
<i>Ye Chen (Shanghai Jiao Tong University), Bingbing Ni (Shanghai Jiao Tong University), Xuanhong Chen (Shanghai Jiao Tong University), and Zhangli Hu (Shanghai Jiao Tong University)</i>	
One-Shot Recognition of Any Material Anywhere Using Contrastive Learning with Physics-Based Rendering	23467
<i>Manuel S. Drehwald (University of Toronto), Sagi Eppel (Vector Institute), Jolina Li (The Matter Lab), Han Hao (University of Toronto), and Alan Aspuru-Guzik (University of Toronto)</i>	
Fast Full-frame Video Stabilization with Iterative Optimization	23477
<i>Weiyue Zhao (Huazhong University of Science and Technology), Xin Li (University at Albany), Zhan Peng (Huazhong University of Science and Technology), Xianrui Luo (Huazhong University of Science and Technology), Xinyi Ye (Huazhong University of Science and Technology), Hao Lu (Huazhong University of Science and Technology), and Zhiguo Cao (Huazhong Univ. of Sci.&Tech.)</i>	
Two Birds, One Stone: A Unified Framework for Joint Learning of Image and Video Style Transfers	23488
<i>Bohai Gu (Institute of Software Chinese Academy of Sciences), Heng Fan (University of North Texas), and Libo Zhang (Institute of Software Chinese Academy of Sciences)</i>	
Multi-Modal Gated Mixture of Local-to-Global Experts for Dynamic Image Fusion	23498
<i>Bing Cao (Tianjin University), Yiming Sun (Tianjin University), Pengfei Zhu (Tianjin University), and Qinghua Hu (Tianjin University)</i>	

SAFE: Sensitivity-Aware Features for Out-of-Distribution Object Detection	23508
<i>Samuel Wilson (Queensland University of Technology), Tobias Fischer (Queensland University of Technology), Feras Dayoub (The University of Adelaide), Dimity Miller (Queensland University of Technology), and Niko Sünderhauf (Queensland University of Technology)</i>	
GeT: Generative Target Structure Debiasing for Domain Adaptation	23520
<i>Can Zhang (National University of Singapore) and Gim Hee Lee (National University of Singapore)</i>	
HairCLIPv2: Unifying Hair Editing via Proxy Feature Blending	23532
<i>Tianyi Wei (University of Science and Technology of China), Dongdong Chen (Microsoft Cloud AI), Wenbo Zhou (University of Science and Technology of China), Jing Liao (City University of Hong Kong), Weiming Zhang (University of Science and Technology of China), Gang Hua (Wormpex AI Research), and Nenghai Yu (University of Science and Technology of China)</i>	
Deformer: Dynamic Fusion Transformer for Robust Hand Pose Estimation	23543
<i>Qichen Fu (Carnegie Mellon University), Xingyu Liu (Carnegie Mellon University), Ran Xu (Salesforce Research), Juan Carlos Niebles (Salesforce & Stanford University), and Kris M. Kitani (Carnegie Mellon University)</i>	
Improving Continuous Sign Language Recognition with Cross-Lingual Signs	23555
<i>Fangyun Wei (Microsoft Research Asia) and Yutong Chen (Tsinghua University)</i>	
A Parse-Then-Place Approach for Generating Graphic Layouts from Textual Descriptions	23565
<i>Jiawei Lin (Xi'an Jiao Tong University), Jiaqi Guo (Microsoft), Shizhao Sun (Microsoft Research), Weijiang Xu (MSRA), Ting Liu (Xi'an Jiaotong University), Jian-Guang Lou (Microsoft), and Dongmei Zhang (Microsoft Research Asia)</i>	
DISeR: Designing Imaging Systems with Reinforcement Learning	23575
<i>Tzofi Klinghoffer (Massachusetts Institute of Technology), Kushagra Tiwary (MIT), Nikhil Behari (MIT Media Lab), Bhavya Agrawalla (MIT), and Ramesh Raskar (Massachusetts Institute of Technology)</i>	
Segmentation of Tubular Structures Using Iterative Training with Tailored Samples	23586
<i>Wei Liao (Independent Researcher)</i>	
Time-to-Contact Map by Joint Estimation of Up-to-Scale Inverse Depth and Global Motion using a Single Event Camera	23596
<i>Urbano Miguel Nunes (Sorbonne University), Laurent Udo Perrinet (Institut de Neurosciences de la Timone, (UMR 7289, Aix Marseille Université, CNRS), and Sio-Hoi Ieng (Sorbonne University)</i>	

Author Index