2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)

New Orleans, Louisiana, USA
18-24 June 2022

Pages 1-548

IEEE Catalog Number: CFP22003-POD
ISBN: 978-1-6654-6947-0

Copyright © 2022 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:	CFP22003-POD
ISBN (Print-On-Demand):	978-1-6654-6947-0
ISBN (Online):	978-1-6654-6946-3
ISSN:	1063-6919

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

2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)

CVPR 2022

Table of Contents

Message from the General and Program Chairs ..... **ccxxiii**
CVPR 2022 Organizing Committee ..... **ccxxiv**
CVPR 2022 Area Chairs ..... **ccxxv**
CVPR 2022 Outstanding Reviewers ..... **ccxxvi**
Donors ..... **ccxxvii**

Poster 1.1: Machine Learning

Clipped Hyperbolic Classifiers Are Super-Hyperbolic Classifiers 1
Yunhui Guo (UC Berkeley / ICSI), Xudong Wang (UC Berkeley / ICSI), Yubei Chen (UC Berkeley), and Stella X. Yu (UC Berkeley / ICSI)

CO-SNE: Dimensionality Reduction and Visualization for Hyperbolic Data 11
Yunhui Guo (UC Berkeley / ICSI), Haoran Guo (UC Berkeley), and Stella X. Yu (UC Berkeley / ICSI)

Efficient Deep Embedded Subspace Clustering 21
Jinyu Cai (Fuzhou University), Jicong Fan (The Chinese University of Hong Kong (Shenzhen)), Wenzhong Guo (Fuzhou University), Shiping Wang (Fuzhou University), Yunhe Zhang (Fuzhou University), and Zhao Zhang (Hefei University of Technology)

Noise Is Also Useful: Negative Correlation-Steered Latent Contrastive Learning 31
Jiexi Yan (Xidian University), Lei Luo (University of Pittsburgh), Chenghao Xu (Xidian University), Cheng Deng (Xidian University), and Heng Huang (University of Pittsburgh)

Active Learning for Open-Set Annotation 41
Kun-Peng Ning (Nanjing University of Aeronautics and Astronautics), Xun Zhao (Tencent Company), Yu Li (International Digital Economy Academy), and Sheng-Jun Huang (Nanjing University of Aeronautics and Astronautics)

Understanding and Increasing Efficiency of Frank-Wolfe Adversarial Training 50
Theodoros Tsiligkaridis (MIT Lincoln Laboratory) and Jay Roberts (MIT Lincoln Laboratory)

Robust Optimization As Data Augmentation for Large-Scale Graphs 60
Kezhi Kong (University of Maryland, College Park), Guohao Li (King Abdullah University of Science and Technology (KAUST)), Mucong Ding (University of Maryland), Zuxuan Wu (UMD), Chen Zhu (University of Maryland, College Park), Bernard Ghanem (KAUST), Gavin Taylor (US Naval Academy), and Tom Goldstein (University of Maryland, College Park)

A Re-Balancing Strategy for Class-Imbalanced Classification Based on Instance Difficulty 70
Sihao Yu (University of Chinese Academy of Sciences), Jiafeng Guo (Institute of Computing Technology, Chinese Academy of Sciences), Ruqing Zhang (The Institute of Computing Technology of the Chinese Academy of Sciences), Yixing Fan (CAS Key Lab of Network Data Science and Technology, Institute of Computing Technology, Chinese Academy of Sciences), Zizhen Wang (institute of computing technology), and Xueqi Cheng (Institute of Computing Technology, Chinese Academy of Sciences)

The Devil Is in the Margin: Margin-Based Label Smoothing for Network Calibration 80
Bingyuan Liu (ETS Montreal), Ismail Ben Ayed (ETS Montreal), Adrian Galdran (Pompeu Fabra University), and Jose Dolz (ETS Montreal)

Towards Better Plasticity-Stability Trade-Off in Incremental Learning: A Simple Linear Connector 89
Guoliang Lin (Sun Yat-sen University), Hanlu Chu (South China Normal University), and Hanjiang Lai (Sun Yat-Sen university)

GCR: Gradient Coreset Based Replay Buffer Selection for Continual Learning 99
Rishabh Tiwari (Indian Institute of Technology (ISM) - Dhanbad), Krishnateja Killamsetty (University of Texas at Dallas), Rishabh Iyer (University of Texas at Dallas), and Pradeep Shenoy (Google Research India)

Learning Bayesian Sparse Networks With Full Experience Replay for Continual Learning 109
Qingsen Yan (The University of Adelaide), Dong Gong (The University of New South Wales), Yuhang Liu (The University of Adelaide), Anton van den Hengel (University of Adelaide), and Javen Qinfeng Shi (University of Adelaide)

A Variational Bayesian Method for Similarity Learning in Non-Rigid Image Registration 119
Daniel Grzech (Imperial College London), Mohammad Farid Azampour (Imperial College London), Ben Glocker (Imperial College London), Julia Schnabel (King's College London), Nassir Navab (TU Munich, Germany), Bernhard Kainz (Imperial College London, FAU Erlangen-Nürnberg), and Loïc Le Folgoc (Imperial College London)

Learning To Learn by Jointly Optimizing Neural Architecture and Weights 129
Yadong Ding (Zhejiang University), Yu Wu (University of Technology Sydney), Chengyue Huang (Zhejiang University), Siliang Tang (Zhejiang University), Yi Yang (UTS), Longhui Wei (University of Science and Technology of China), Yueting Zhuang (Zhejiang University), and Qi Tian (Huawei Cloud & AI)

Learning To Prompt for Continual Learning 139
Zifeng Wang (Northeastern University), Zizhao Zhang (Google), Chen-Yu Lee (Google), Han Zhang (Google), Ruoxi Sun (Google), Xiaoqi Ren (Google), Guolong Su (Google), Vincent Perot (Google AI), Jennifer Dy (Northeastern), and Tomas Pfister (Google)

Meta-Attention for ViT-Backed Continual Learning 150
Mengqi Xue (Zhejiang University), Haofei Zhang (Zhejiang University), Jie Song (Zhejiang University), and Mingli Song (Zhejiang University)

Multi-Frame Self-Supervised Depth With Transformers 160
Vitor Guizilini (Toyota Research Institute), Rareș Ambruș (Toyota Research Institute), Dian Chen (Toyota Research Institute), Sergey Zakharov (Toyota Research Institute), and Adrien Gaidon (Toyota Research Institute)

Continual Learning With Lifelong Vision Transformer 171
Zhen Wang (The University of Sydney), Liu Liu (The University of Sydney), Yiqun Duan (University of Technology Sydney), Yajing Kong (The University of Sydney), and Dacheng Tao (JD.com)

Rethinking Bayesian Deep Learning Methods for Semi-Supervised Volumetric Medical Image Segmentation 182
Jianfeng Wang (University of Oxford) and Thomas Lukasiewicz (University of Oxford)

Revisiting Random Channel Pruning for Neural Network Compression 191
Yawei Li (ETH Zurich), Kamil Adamczewski (Max Planck Institute for Intelligent Systems), Wen Li (University of Electronic Science and Technology of China), Shuhang Gu (ETH Zurich, Switzerland), Radu Timofte (ETH Zurich), and Luc Van Gool (ETH Zurich)

Deep Safe Multi-View Clustering: Reducing the Risk of Clustering Performance Degradation Caused by View Increase 202
Huayi Tang (University of Electronic Science and Technology of China) and Yong Liu (Renmin University of China)

Hypergraph-Induced Semantic Tuplet Loss for Deep Metric Learning 212
Jongin Lim (Seoul National University), Sangdoo Yun (NAVER AI LAB), Seulki Park (Seoul National University), and Jin Young Choi (Seoul National University)

Towards Robust and Reproducible Active Learning Using Neural Networks 223
Prateek Munjal (G42 Healthcare), Nasir Hayat (NYUAD), Munawar Hayat (Monash University), Jamshid Sourati (missing), and Shadab Khan (Inception Institute of Artificial Intelligence)

Non-Iterative Recovery From Nonlinear Observations Using Generative Models 233
Jiulong Liu (Academy of Mathematics and System Sciences) and Zhaoqiang Liu (National University of Singapore)

Gaussian Process Modeling of Approximate Inference Errors for Variational Autoencoders 244
Minyoung Kim (Samsung AI Center Cambridge)

Robust Combination of Distributed Gradients Under Adversarial Perturbations 254
Kwang In Kim (UNIST)

Do Learned Representations Respect Causal Relationships? 264
Lan Wang (Michigan State University) and Vishnu Naresh Boddeti (Michigan State University)

How Much More Data Do I Need? Estimating Requirements for Downstream Tasks 275
Rafid Mahmood (NVIDIA), James Lucas (University of Toronto and Vector Institute), David Acuna (University of Toronto, NVIDIA), Daiqing Li (NVIDIA), Jonah Philion (University of Toronto, NVIDIA), Jose M. Alvarez (NVIDIA), Zhiding Yu (NVIDIA), Sanja Fidler (University of Toronto, NVIDIA), and Marc T. Law (NVIDIA)

Pushing the Envelope of Gradient Boosting Forests via Globally-Optimized Oblique Trees 285
Magzhan Gabidolla (University of California, Merced) and Miguel Á. Carreira-Perpiñán (UC Merced)

Contrastive Test-Time Adaptation 295
Dian Chen (UC Berkeley), Dequan Wang (UC Berkeley), Trevor Darrell (UC Berkeley), and Sayna Ebrahimi (Google)

AutoSDF: Shape Priors for 3D Completion, Reconstruction and Generation 306
Paritosh Mittal (Carnegie Mellon University), Yen-Chi Cheng (Carnegie Mellon University), Maneesh Singh (Verisk Analytics), and Shubham Tulsiani (Carnegie Mellon University)

Selective-Supervised Contrastive Learning With Noisy Labels 316
Shikun Li (Chinese Academy of Sciences), Xiaobo Xia (The University of Sydney), Shiming Ge (Chinese Academy of Sciences), and Tongliang Liu (The University of Sydney)

RecDis-SNN: Rectifying Membrane Potential Distribution for Directly Training Spiking Neural Networks 326
Yufei Guo (The Second Academy of China Aerospace Science and Industry Corporation), Xinyi Tong (The Second Academy of China Aerospace Science and Industry Corporation), Yuanpei Chen (X LAB, The Second Academy of CASIC, Beijing), Liwen Zhang (X Lab, the Second Academy of CASIC, Beijing), Xiaode Liu (X Lab, The Second Academy of China Aerospace Science and Industry Corporation), Zhe Ma (Xlab, the Second Academy of CASIC, Beijing), and Xuhui Huang (X Lab, The Second Academy of CASIC)

Hierarchical Nearest Neighbor Graph Embedding for Efficient Dimensionality Reduction 336
Saquib Sarfraz (Karlsruhe Institute of Technology), Marios Koulakis (KIT), Constantin Seibold (Karlsruhe Institute of Technology), and Rainer Stiefelhagen (Karlsruhe Institute of Technology)

Poster 1.1: Statistical Methods

Scalable Penalized Regression for Noise Detection in Learning With Noisy Labels 346
Yikai Wang (Fudan University), Xinwei Sun (MSRA), and Yanwei Fu (Fudan University)

Nested Hyperbolic Spaces for Dimensionality Reduction and Hyperbolic NN Design 356
Xiran Fan (University of Florida), Chun-Hao Yang (National Taiwan University), and Baba C. Vemuri (University of Florida)

Learning Structured Gaussians To Approximate Deep Ensembles 366
Ivor J. A. Simpson (University of Sussex), Sara Vicente (Niantic), and Neill D. F. Campbell (University of Bath)

Out-of-Distribution Generalization With Causal Invariant Transformations 375
Ruoyu Wang (Chinese Academy of Science), Mingyang Yi (Chinese Academy of Science), Zhitang Chen (Huawei Noah's Ark Lab), and Shengyu Zhu (Huawei Noah's Ark Lab)

Split Hierarchical Variational Compression 386
Tom Ryder (Huawei, Noah's Ark Lab), Chen Zhang (Huawei Noah's Ark Lab), Ning Kang (Huawei Noah's Ark Lab), and Shifeng Zhang (Huawei Noah's Ark Lab)

Implicit Feature Decoupling With Depthwise Quantization 396
Iordanis Fostiropoulos (University of Southern California) and Barry Boehm (USC)

Understanding Uncertainty Maps in Vision With Statistical Testing 406
Jurijs Nazarovs (University of Wisconsin - Madison), Zhichun Huang (CMU), Songwong Tasneeyapant (University of Wisconsin-Madison), Rudrasis Chakraborty (Butlr), and Vikas Singh (University of Wisconsin Madison)

Poster 1.1: Optimization Methods

A Hybrid Quantum-Classical Algorithm for Robust Fitting 417
Anh-Dzung Doan (The University of Adelaide), Michele Sasdelli (The University of Adelaide), David Suter (Edith Cowan University), and Tat-Jun Chin (The University of Adelaide)

A Scalable Combinatorial Solver for Elastic Geometrically Consistent 3D Shape Matching 428
Paul Roetzer (TUM), Paul Swoboda (MPI fuer Informatik, Saarbruecken), Daniel Cremers (TU Munich), and Florian Bernard (University of Bonn)

FastDOG: Fast Discrete Optimization on GPU 439
Ahmed Abbas (MPI-INF) and Paul Swoboda (MPI fuer Informatik, Saarbruecken)

Data-Free Network Compression via Parametric Non-Uniform Mixed Precision Quantization 450
Vladimir Chikin (Huawei Noah's Ark Lab) and Mikhail Antiukh (HSE University)

AdaSTE: An Adaptive Straight-Through Estimator To Train Binary Neural Networks 460
Huu Le (NA), Rasmus Kjær Høier (Chalmers University of Technology), Che-Tsung Lin (Chalmers University of Techonology), and Christopher Zach (Chalmers University)

Training Quantised Neural Networks With STE Variants: The Additive Noise Annealing Algorithm 470
Matteo Spallanzani (ETH Zurich), Gian Paolo Leonardi (Universita di Trento), and Luca Benini (ETHZ, University of Bologna)

AME: Attention and Memory Enhancement in Hyper-Parameter Optimization 480
Nuo Xu (NLPR, Institute of Automation, Chinese Academy of Sciences, University of Chinese Academy of Sciences), Jianlong Chang (Huawei Cloud), Xing Nie (Institute of Automation, Chinese Academy of Sciences, University of Chinese Academy of Sciences), Chunlei Huo (National Lab. Patter Recognition, Institute of Automation, Chinese Academy of Sciences), Shiming Xiang (Chinese Academy of Sciences, China), and Chunhong Pan (Institute of Automation, Chinese Academy of Sciences)

Efficient Maximal Coding Rate Reduction by Variational Forms 490
Christina Baek (Carnegie Mellon University), Ziyang Wu (Cornell University), Kwan Ho Ryan Chan (University of California, Berkeley), Tianjiao Ding (Johns Hopkins University), Yi Ma (UC Berkeley), and Benjamin D. Haeffele (Johns Hopkins University)

A Unified Framework for Implicit Sinkhorn Differentiation 499
Marvin Eisenberger (TU Munich), Aysim Toker (TUM), Laura Leal-Taixé (TUM), Florian Bernard (University of Bonn), and Daniel Cremers (TU Munich)

Computing Wasserstein-p Distance Between Images With Linear Cost 509
Yidong Chen (Computer Network Information Center, Chinese Academy of Sciences; University of Chinese Academy of Sciences), Chen Li (Computer Network Information Center, Chinese Academy of Sciences), and Zhonghua Lu (Computer Network Information Center, CAS)

An Iterative Quantum Approach for Transformation Estimation From Point Sets 519
Natacha Kuete Meli (University of Luebeck), Florian Mannel (University of Lübeck), and Jan Lellmann (University of Luebeck)

Poster 1.1: Deep Learning Architectures & Techniques

BoosterNet: Improving Domain Generalization of Deep Neural Nets Using Culpability-Ranked Features 528
Nourhan Bayasi (University of British Columbia), Ghassan Hamarneh (Simon Fraser University), and Rafeef Garbi (University of British Columbia)

Pooling Revisited: Your Receptive Field Is Suboptimal 539
Dong-Hwan Jang (Seoul National University), Sanghyeok Chu (Seoul National University), Joonhyuk Kim (Seoul National University), and Bohyung Han (Seoul National University)

Why Discard if You Can Recycle?: A Recycling Max Pooling Module for 3D Point Cloud Analysis 549
Jiajing Chen (Syracuse University), Burak Kakillioglu (Motorola Solutions), Huantao Ren (Syracuse University), and Senem Velipasalar (Syracuse University)

Online Convolutional Re-Parameterization 558
Mu Hu (Zhejiang University), Junyi Feng (Alibaba group), Jiashen Hua (Damo Academy, Alibaba Group), Baisheng Lai (Alibaba Group), Jianqiang Huang (Alibaba Group), Xiaojin Gong (Zhejiang University), and Xian-Sheng Hua (Damo Academy, Alibaba Group)

RepMLPNet: Hierarchical Vision MLP With Re-Parameterized Locality 568
Xiaohan Ding (Tsinghua University), Honghao Chen (Institute of Automation, Chinese Academy of Sciences), Xiangyu Zhang (Megvii Technology), Jungong Han (Aberystwyth University), and Guiguang Ding (Tsinghua University, China)

DyRep: Bootstrapping Training With Dynamic Re-Parameterization 578
Tao Huang (SenseTime Research), Shan You (SenseTime), Bohan Zhang (UMass Amherst), Yuxuan Du (University of Sydney), Fei Wang (University of Science and Technology of China), Chen Qian (SenseTime), and Chang Xu (University of Sydney)

Quarantine: Sparsity Can Uncover the Trojan Attack Trigger for Free 588
Tianlong Chen (Unversity of Texas at Austin), Zhenyu Zhang (University of Science and Technology of China), Yihua Zhang (Michgan State University), Shiyu Chang (UCSB), Sijia Liu (Michigan State University), and Zhangyang Wang (University of Texas at Austin)

Condensing CNNs With Partial Differential Equations 600
Anil Kag (Boston University) and Venkatesh Saligrama (Boston University)

Deep Equilibrium Optical Flow Estimation 610
Shaojie Bai (Carnegie Mellon University), Zhengyang Geng (Peking University), Yash Savani (Carnegie Mellon University), and J. Zico Kolter (Carnegie Mellon University)

Frame Averaging for Equivariant Shape Space Learning 621
Matan Atzmon (Weizmann Institute of Science), Koki Nagano (NVIDIA), Sanja Fidler (University of Toronto, NVIDIA), Sameh Khamis (Nvidia), and Yaron Lipman (Weizmann Institute of Science/ Facebook AI Research)

Dual-Generator Face Reenactment 632
Gee-Sern Hsu (National Taiwan University of Science and Technology), Chun-Hung Tsai (National Taiwan University of Science and Technology), and Hung-Yi Wu (National Taiwan University of Science and Technology)

Convolution of Convolution: Let Kernels Spatially Collaborate 641
Rongzhen Zhao (Lynxi Technologies, Co., Ltd.), Jian Li (Lynxi Technologies, Co., Ltd.), and Zhenzhi Wu (Lynxi Technologies, Co., Ltd.)

SASIC: Stereo Image Compression With Latent Shifts and Stereo Attention 651
Matthias Wödlinger (TU Wien), Jan Kotera (Computer Vision Lab, TU Wien), Jan Xu (Deep Render Ltd), and Robert Sablatnig (TU Wien)

RADU: Ray-Aligned Depth Update Convolutions for ToF Data Denoising 661
Michael Schelling (Ulm University - Institute of Media Informatics), Pedro Hermosilla (Ulm University), and Timo Ropinski (Ulm University)

Co-Domain Symmetry for Complex-Valued Deep Learning 671
Utkarsh Singhal (UC Berkeley), Yifei Xing (UC Berkeley), and Stella X. Yu (UC Berkeley / ICSI)

Paramixer: Parameterizing Mixing Links in Sparse Factors Works Better Than Dot-Product Self-Attention 681
Tong Yu (NTNU), Ruslan Khalitov (NTNU), Lei Cheng (Norwegian University of Science and Technology), and Zhirong Yang (Norwegian University of Science and Technology)

Compressing Models With Few Samples: Mimicking Then Replacing 691
Huanyu Wang (Nanjing University), Junjie Liu (Meituan), Xin Ma (Meituan), Yang Yong (Xi'an Jiaotong University), Zhenhua Chai (Meituan), and Jianxin Wu (Nanjing University)

Total Variation Optimization Layers for Computer Vision 701
Raymond A. Yeh (TTIC), Yuan-Ting Hu (UIUC), Zhongzheng Ren (UIUC), and Alexander G. Schwing (UIUC)

AIM: An Auto-Augmenter for Images and Meshes 712
Vinit Veerendraveer Singh (University Of Delaware) and Chandra Kambhamettu (University of Delaware)

Recurrent Variational Network: A Deep Learning Inverse Problem Solver Applied to the Task of Accelerated MRI Reconstruction 722
George Yiasemis (The Netherlands Cancer Institute), Jan-Jakob Sonke (The Netherlands Cancer Institute), Clarisa Sánchez (University of Amsterdam), and Jonas Teuwen (NKI)

Deep Orientation-Aware Functional Maps: Tackling Symmetry Issues in Shape Matching 732
Nicolas Donati (Ecole Polytechnique), Etienne Corman (CNRS), and Maks Ovsjanikov (Ecole polytechnique)

Weakly-Supervised Metric Learning With Cross-Module Communications for the Classification of Anterior Chamber Angle Images 742
Jingqi Huang (Guangdong Polytechnic Normal University), Yue Ning (Stevens Institute of Technology), Dong Nie (UNC), Linan Guan (Guangdong Polytechnic Normal University), and Xiping Jia (Guangdong Polytechnic Normal University)

Delving Into the Estimation Shift of Batch Normalization in a Network 753
Lei Huang (Beihang University), Yi Zhou (Southeast University), Tian Wang (Beihang University), Jie Luo (Beihang University), and Xianglong Liu (BUAA)

Generalizing Interactive Backpropagating Refinement for Dense Prediction Networks 763
Fanqing Lin (Brigham Young University), Brian Price (Adobe), and Tony Martinez (Brigham Young University)

Brain-Inspired Multilayer Perceptron With Spiking Neurons 773
Wenshuo Li (Huawei Noah's Ark Lab), Hanting Chen (Peking University), Jianyuan Guo (Huawei), Ziyang Zhang (HUAWEI TECHNOLOGIES CO.LTD), and Yunhe Wang (Huawei Technologies)

Smooth Maximum Unit: Smooth Activation Function for Deep Networks Using Smoothing Maximum Technique 784
Koushik Biswas (Indraprastha Institute of Information Technology, New Delhi, India), Sandeep Kumar (Shaheed Bhagat Singh College, University of Delhi, Delhi), Shilpak Banerjee (Indian Institute of Technology Tirupati), and Ashish Kumar Pandey (Indraprastha Institute of Information Technology, New Delhi, India)

Revisiting Weakly Supervised Pre-Training of Visual Perception Models 794
Mannat Singh (Facebook AI Research), Laura Gustafson (Facebook AI Research), Aaron Adcock (Facebook AI Research), Vinicius de Freitas Reis (Nabla Bio), Bugra Gedik (Facebook, Inc), Raj Prateek Kosaraju (FACEBOOK INC), Dhruv Mahajan (Facebook), Ross Girshick (FAIR), Piotr Dollár (FAIR), and Laurens van der Maaten (Facebook)

On the Integration of Self-Attention and Convolution 805
Xuran Pan (Tsinghua University), Chunjiang Ge (Tsinghua University), Rui Lu (Tsinghua University), Shiji Song (Department of Automation, Tsinghua University), Guanfu Chen (Huawei), Zeyi Huang (Huawei), and Gao Huang (Tsinghua)

Hire-MLP: Vision MLP via Hierarchical Rearrangement 816
Jianyuan Guo (Noah's Ark Lab, Huawei Technologies), Yehui Tang (Peking University), Kai Han (Noah's Ark Lab, Huawei Technologies), Xinghao Chen (Huawei Noah's Ark Lab), Han Wu (University of Sydney), Chao Xu (Peking University), Chang Xu (University of Sydney), and Yunhe Wang (Huawei Technologies)

Stable Long-Term Recurrent Video Super-Resolution 827
Benjamin Naoto Chiche (Université Paris-Saclay / Safran Electronics & Defense), Arnaud Woiselle (Safran Electronics & Defense), Joana Frontera-Pons (CEA & IPSA), and Jean-Luc Starck (Commissariat à l'énergie atomique (CEA))

Single-Domain Generalized Object Detection in Urban Scene via Cyclic-Disentangled Self-Distillation 837
Aming Wu (Xidian University) and Cheng Deng (Xidian University)

Progressive End-to-End Object Detection in Crowded Scenes 847
Anlin Zheng (Megvii Technology), Yuang Zhang (Shanghai Jiao Tong University), Xiangyu Zhang (Megvii Technology), Xiaojuan Qi (The University of Hong Kong), and Jian Sun (Megvii Technology)

Zero-Shot Text-Guided Object Generation With Dream Fields 857
Ajay Jain (UC Berkeley), Ben Mildenhall (Google Research), Jonathan T. Barron (Google Research), Pieter Abbeel (UC Berkeley), and Ben Poole (Google Brain)

Poster 1.1: Recognition: Detection, Categorization, Retrieval

ISNet: Shape Matters for Infrared Small Target Detection 867
Mingjin Zhang (Xidian University), Rui Zhang (XDU), Yuxiang Yang (Hangzhou Dianzi University), Haichen Bai (Xidian University), Jing Zhang (The University of Sydney), and Jie Guo (Xidian University)

Pseudo-Stereo for Monocular 3D Object Detection in Autonomous Driving 877
Yi-Nan Chen (Zhejiang University), Hang Dai (Mohamed bin Zayed University of Artificial Intelligence), and Yong Ding (Zhejiang University)

CLRNet: Cross Layer Refinement Network for Lane Detection 888
Tu Zheng (FABU), Yifei Huang (zhejiang university), Yang Liu (Zhejiang University), Wenjian Tang (Hangzhou Feibu Technology Co., Ltd.), Zheng Yang (FABU), Deng Cai (ZJU), and Xiaofei He (Zhejiang University)

CAT-Det: Contrastively Augmented Transformer for Multi-Modal 3D Object Detection 898
Yanan Zhang (Beihang University), Jiaxin Chen (Beihang University), and Di Huang (Beihang University, China)

Modality-Agnostic Learning for Radar-Lidar Fusion in Vehicle Detection 908
Yu-Jhe Li (Carnegie Mellon University), Jinhyung Park (Carnegie Mellon University), Matthew O'Toole (Carnegie Mellon University), and Kris Kitani (Carnegie Mellon University)

Group Contextualization for Video Recognition 918
Yanbin Hao (University of Science and Technology of China), Hao Zhang (Singapore Management University), Chong-Wah Ngo (Singapore Management University), and Xiangnan He (University of Science and Technology of China)

Learning Transferable Human-Object Interaction Detector With Natural Language Supervision 929
Suchen Wang (Nanyang Technological University), Yueqi Duan (Tsinghua University), Henghui Ding (ETH Zurich), Yap-Peng Tan (Nanyang Technological University, Singapore), Kim-Hui Yap (Nanyang Technological University), and Junsong Yuan (State University of New York at Buffalo, USA)

Accelerating DETR Convergence via Semantic-Aligned Matching 939
Gongjie Zhang (Nanyang Technological University), Zhipeng Luo (Nanyang Technological University), Yingchen Yu (Nanyang Technological University), Kaiwen Cui (Nanyang Technological University), and Shijian Lu (Nanyang Technological University)

Efficient Video Instance Segmentation via Tracklet Query and Proposal 949
Jialian Wu (State University of New York at Buffalo), Sudhir Yarram (University at Buffalo), Hui Liang (Amazon), Tian Lan (Amazon Inc), Junsong Yuan (State University of New York at Buffalo, USA), Jayan Eledath (Amazon), and Gérard Medioni (USC)

Class Re-Activation Maps for Weakly-Supervised Semantic Segmentation 959
Zhaozheng Chen (Singapore Management University), Tan Wang (Nanyang Technological University), Xiongwei Wu (Singapore Management U), Xian-Sheng Hua (Damo Academy, Alibaba Group), Hanwang Zhang (Nanyang Technological University), and Qianru Sun (Singapore Management University)

Democracy Does Matter: Comprehensive Feature Mining for Co-Salient Object Detection 969
Siyue Yu (Xi'an Jiaotong-Liverpool University), Jimin Xiao (Xi'an Jiaotong-Liverpool University), Bingfeng Zhang (Xi'an Jiaotong-Liverpool University), and Eng Gee Lim (Xian jiaotong-liverpool University)

C2AM: Contrastive Learning of Class-Agnostic Activation Map for Weakly Supervised Object Localization and Semantic Segmentation 979
Jinheng Xie (Shenzhen University), Jianfeng Xiang (ShenZhen University), Junliang Chen (Shenzhen University), Xianxu Hou (Shenzhen University), Xiaodong Zhao (Shenzhen University), and Linlin Shen (Shenzhen University)

Sketching Without Worrying: Noise-Tolerant Sketch-Based Image Retrieval 989
Ayan Kumar Bhunia (University of Surrey), Subhadeep Koley (West Bengal University of Technology), Abdullah Faiz Ur Rahman Khilji (National Institute of Technology Silchar), Aneeshan Sain (University of Surrey), Pinaki Nath Chowdhury (University of Surrey), Tao Xiang (University of Surrey), and Yi-Zhe Song (University of Surrey)

AutoLoss-Zero: Searching Loss Functions From Scratch for Generic Tasks 999
Hao Li (Chinese University of Hong Kong), Tianwen Fu (Chinese University of Hong Kong), Jifeng Dai (SenseTime), Hongsheng Li (The Chinese University of Hong Kong), Gao Huang (Tsinghua), and Xizhou Zhu (SenseTime)

Consistency Learning via Decoding Path Augmentation for Transformers in Human Object Interaction Detection 1009
Jihwan Park (Korea university), SeungJun Lee (Korea University), Hwan Heo (Korea University), Hyeong Kyu Choi (Korea University), and Hyunwoo J. Kim (Korea University)

A Proposal-Based Paradigm for Self-Supervised Sound Source Localization in Videos 1019
Hanyu Xuan (Nanjing University of Science and Technology), Zhiliang Wu (Nanjing University of Science and Technology), Jian Yang (Nanjing University of Science and Technology), Yan Yan (Nanjing University of Science and Technology), and Xavier Alameda-Pineda (INRIA)

SimAN: Exploring Self-Supervised Representation Learning of Scene Text via Similarity-Aware Normalization 1029
Canjie Luo (South China University of Technology), Lianwen Jin (South China University of Technology), and Jingdong Chen (Ant Group)

Towards End-to-End Unified Scene Text Detection and Layout Analysis 1039
Shangbang Long (Google Research), Siyang Qin (Google Research), Dmitry Panteleev (Google Research), Alessandro Bissacco (Google Inc.), Yasuhisa Fujii (Google Inc.), and Michalis Raptis (Google Inc.)

Clothes-Changing Person Re-Identification With RGB Modality Only 1050
Xinqian Gu (University of Chinese Academy of Sciences), Hong Chang (Chinese Academy of Sciences), Bingpeng Ma (University of Chinese Academy of Sciences), Shutao Bai (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)

MonoJSG: Joint Semantic and Geometric Cost Volume for Monocular 3D Object Detection 1060
Qing Lian (Hong Kong University of Science and Technology), Peiliang Li (DJI), and Xiaozhi Chen (DJI)

Homography Loss for Monocular 3D Object Detection 1070
Jiaqi Gu (Zhejiang University), Bojian Wu (Alibaba Group), Lubin Fan (Alibaba Group), Jianqiang Huang (Alibaba Group), Shen Cao (Alibaba Group), Zhiyu Xiang (Zhejiang University), and Xian-Sheng Hua (Damo Academy, Alibaba Group)

TransFusion: Robust LiDAR-Camera Fusion for 3D Object Detection With Transformers 1080
Xuyang Bai (HKUST), Zeyu Hu (Hong Kong University of Science and Technology), Xinge Zhu (The Chinese University of Hong Kong), Qingqiu Huang (Huawei), Yilun Chen (HUAWEI Corporation), Hongbo Fu (City University of Hong Kong), and Chiew-Lan Tai (Hong Kong University of Science & Technology)

TWIST: Two-Way Inter-Label Self-Training for Semi-Supervised 3D Instance Segmentation 1090
Ruihang Chu (The Chinese University of Hong Kong), Xiaoqing Ye (baidu), Zhengzhe Liu (The Chinese University of Hong Kong), Xiao Tan (Baidu Inc.), Xiaojuan Qi (The University of Hong Kong), Chi-Wing Fu (The Chinese University of Hong Kong), and Jiaya Jia (Chinese University of Hong Kong)

RBGNet: Ray-Based Grouping for 3D Object Detection 1100
Haiyang Wang (Peking University), Shaoshuai Shi (MPI Informatics), Ze Yang (University of Toronto), Rongyao Fang (Chinese University of Hong Kong), Qi Qian (Alibaba Group), Hongsheng Li (The Chinese University of Hong Kong), Bernt Schiele (MPI Informatics), and Liwei Wang (Peking University)

Voxel Field Fusion for 3D Object Detection 1110
Yanwei Li (The Chinese University of Hong Kong), Xiaojuan Qi (The University of Hong Kong), Yukang Chen (The Chinese University of Hong Kong), Liwei Wang (CUHK), Zeming Li (Megvii (Face++) Inc.), Jian Sun (Megvii Technology), and Jiaya Jia (Chinese University of Hong Kong)

Learning To Detect Mobile Objects From LiDAR Scans Without Labels 1120
Yurong You (Cornell University), Katie Luo (Cornell University), Cheng Perng Phoo (Cornell University), Wei-Lun Chao (The Ohio State University), Wen Sun (Cornell University), Bharath Hariharan (Cornell University), Mark Campbell (Cornell University), and Kilian Q. Weinberger (Cornell University)

OccAM's Laser: Occlusion-Based Attribution Maps for 3D Object Detectors on LiDAR Data 1131
David Schinagl (Graz University of Technology), Georg Krispel (Graz University of Technology), Horst Possegger (Graz University of Technology), Peter M. Roth (TU Munich), and Horst Bischof (Graz University of Technology)

Confidence Propagation Cluster: Unleash Full Potential of Object Detectors 1141
Yichun Shen (Nvidia), Wanli Jiang (Nvidia), Zhen Xu (Nvidia), Rundong Li (NVIDIA), Junghyun Kwon (Nvidia), and Siyi Li (NVIDIA)

TransGeo: Transformer Is All You Need for Cross-View Image Geo-Localization 1152
Sijie Zhu (University of Central Florida), Mubarak Shah (University of Central Florida), and Chen Chen (University of Central Florida)

A Voxel Graph CNN for Object Classification With Event Cameras 1162
Yongjian Deng (Beijing University of Technology), Hao Chen (Southeast University), Hai Liu (Central China Norma University), and Youfu Li (City University of Hong Kong)

OSKDet: Orientation-Sensitive Keypoint Localization for Rotated Object Detection 1172
Dongchen Lu (Tsinghua University), Dongmei Li (Tsinghua University), Yali Li (Tsinghua University), and Shengjin Wang (Tsinghua University)

Canonical Voting: Towards Robust Oriented Bounding Box Detection in 3D Scenes 1183
Yang You (Shanghai Jiao Tong University), Zelin Ye (Shanghai Jiao Tong University), Yujing Lou (Shanghai Jiao Tong University), Chengkun Li (Shanghai Jiao Tong University), Yong-Lu Li (Shanghai Jiao Tong University), Lizhuang Ma (Shanghai Jiao Tong University), Weiming Wang (shanghai jiaotong university), and Cewu Lu (Shanghai Jiao Tong University)

Poster 1.1: Segmentation, Grouping and Shape Analysis

Category Contrast for Unsupervised Domain Adaptation in Visual Tasks 1193
Jiaxing Huang (Nanyang Technological University), Dayan Guan (Nanyang Technological University), Aoran Xiao (Nanyang Technological University), Shijian Lu (Nanyang Technological University), and Ling Shao (NCAI/IIAI/MBZUAI)

Amodal Segmentation Through Out-of-Task and Out-of-Distribution Generalization With a Bayesian Model 1205
Yihong Sun (Johns Hopkins University), Adam Kortylewski (Max Planck Institute for Informatics), and Alan Yuille (Johns Hopkins University)

GANSeg: Learning To Segment by Unsupervised Hierarchical Image Generation 1215
Xingzhe He (University of British Columbia), Bastian Wandt (University of British Columbia), and Helge Rhodin (UBC)

Segment-Fusion: Hierarchical Context Fusion for Robust 3D Semantic Segmentation 1226
Anirud Thyagharajan (Intel Corporation), Benjamin Ummenhofer (Intel Labs), Prashant Laddha (Intel), Om Ji Omer (Intel Corporation), and Sreenivas Subramoney (Intel)

Deep Hierarchical Semantic Segmentation 1236
Liulei Li (Beijing Institute of Technology), Tianfei Zhou (ETH Zurich), Wenguan Wang (Eidgenössische Technische Hochschule Zürich), Jianwu Li (Beijing Institute of Technology), and Yi Yang (UTS)

Semantic Segmentation by Early Region Proxy 1248
Yifan Zhang (Shanghai Jiao Tong University), Bo Pang (Shanghai Jiao Tong University), and Cewu Lu (Shanghai Jiao Tong University)

Panoptic, Instance and Semantic Relations: A Relational Context Encoder To Enhance Panoptic Segmentation 1259
Shubhankar Borse (Qualcomm AI Research), Hyojin Park (Qualcomm AI Research), Hong Cai (Qualcomm AI Research), Debasmit Das (Qualcomm AI Research), Risheek Garrepalli (Qualcomm AI Research), and Fatih Porikli (Qualcomm AI Research)

Panoptic SegFormer: Delving Deeper Into Panoptic Segmentation With Transformers 1270
Zhiqi Li (Nanjing University), Wenhai Wang (Nanjing University), Enze Xie (The University of Hong Kong), Zhiding Yu (NVIDIA), Anima Anandkumar (Caltech), Jose M. Alvarez (NVIDIA), Ping Luo (The University of Hong Kong), and Tong Lu (Nanjing University)

Masked-Attention Mask Transformer for Universal Image Segmentation 1280
Bowen Cheng (UIUC), Ishan Misra (Facebook AI Research), Alexander G. Schwing (UIUC), Alexander Kirillov (Facebook AI Reserach), and Rohit Girdhar (Facebook AI Research)

FocalClick: Towards Practical Interactive Image Segmentation 1290
Xi Chen (Alibaba Group), Zhiyan Zhao (alibaba), Yilei Zhang (Alibaba), Manni Duan (alibaba-inc.com), Donglian Qi (Zhejiang University), and Hengshuang Zhao (University of Oxford)

High Quality Segmentation for Ultra High-Resolution Images 1300
Tiancheng Shen (The Chinese University of Hong Kong), Yuechen Zhang (The Chinese University of Hong Kong), Lu Qi (The Chinese University of Hong Kong), Jason Kuen (Adobe Research), Xingyu Xie (Peking University), Jianlong Wu (Shandong University), Zhe Lin (Adobe Research), and Jiaya Jia (Chinese University of Hong Kong)

Wnet: Audio-Guided Video Object Segmentation via Wavelet-Based Cross-Modal Denoising Networks 1310
Wenwen Pan (Zhejiang University), Haonan Shi (Zhejiang University), Zhou Zhao (Zhejiang University), Jieming Zhu (Huawei Noah's Ark Lab), Xiuqiang He (Huawei Noah's Ark Lab), Zhigeng Pan (Hangzhou Normal University), Lianli Gao (The University of Electronic Science and Technology of China), Jun Yu (HDU), Fei Wu (Zhejiang University), and Qi Tian (Huawei Cloud & AI)

Recurrent Dynamic Embedding for Video Object Segmentation 1322
Mingxing Li (University of Science and Technology of China), Li Hu (Alibaba), Zhiwei Xiong (University of Science and Technology of China), Bang Zhang (Damo Academy Alibaba), Pan Pan (Alibaba Group), and Dong Liu (University of Science and Technology of China)

Accelerating Video Object Segmentation With Compressed Video 1332
Kai Xu (National University of Singapore) and Angela Yao (National University of Singapore)

Per-Clip Video Object Segmentation 1342
Kwanyong Park (KAIST), Sanghyun Woo (KAIST), Seoung Wug Oh (Adobe Research), In So Kweon (KAIST), and Joon-Young Lee (Adobe Research)

SWEM: Towards Real-Time Video Object Segmentation With Sequential Weighted Expectation-Maximization 1352
Zhihui Lin (Tsinghua University), Tianyu Yang (Tencent AI Lab), Maomao Li (Tencent AIlab), Ziyu Wang (Tsinghua University), Chun Yuan (Graduate school at ShenZhenTsinghua university), Wenhao Jiang (Tencent), and Wei Liu (Tencent)

Neural Recognition of Dashed Curves With Gestalt Law of Continuity 1363
Hanyuan Liu (The Chinese University of Hong Kong), Chengze Li (The Chinese University of Hong Kong; Style2Paints Research; Caritas Institute of Higher Education), Xueting Liu (Caritas Institute of Higher Education), and Tien-Tsin Wong (The Chinese University of Hong Kong)

CVNet: Contour Vibration Network for Building Extraction 1373
Ziqiang Xu (Nanjing University of Science and Technology), Chunyan Xu (Nanjing University of Science and Technology), Zhen Cui (Nanjing University of Science and Technology), Xiangwei Zheng (Shandong Normal University), and Jian Yang (Nanjing University of Science and Technology)

A Keypoint-Based Global Association Network for Lane Detection 1382
Jinsheng Wang (Peking University), Yinchao Ma (University of Science and Technology of China), Shaofei Huang (Sensetime Technology), Tianrui Hui (Institute of Information Engineering, Chinese Academy of Sciences), Fei Wang (University of Science and Technology of China), Chen Qian (SenseTime), and Tianzhu Zhang (University of Science and Technology of China)

EDTER: Edge Detection With Transformer 1392
Mengyang Pu (Beijing Jiaotong University), Yaping Huang (Beijing Jiaotong University), Yuming Liu (Beijing Jiaotong University), Qingji Guan (Beijing Jiaotong University), and Haibin Ling (Stony Brook University)

Fixing Malfunctional Objects With Learned Physical Simulation and Functional Prediction 1403
Yining Hong (University of California, Los Angeles), Kaichun Mo (Stanford), Li Yi (Tsinghua University), Leonidas J. Guibas (Stanford University), Antonio Torralba (MIT), Joshua B. Tenenbaum (MIT), and Chuang Gan (MIT-IBM Watson AI Lab)

Coherent Point Drift Revisited for Non-Rigid Shape Matching and Registration 1414
Aoxiang Fan (Wuhan University), Jiayi Ma (Wuhan University), Xin Tian (Wuhan University), Xiaoguang Mei (Wuhan University), and Wei Liu (Tencent)

CodedVTR: Codebook-Based Sparse Voxel Transformer With Geometric Guidance 1425
Tianchen Zhao (Tsinghua University), Niansong Zhang (Tsinghua University), Xuefei Ning (Tsinghua University), He Wang (Peking University), Li Yi (Tsinghua University), and Yu Wang (Tsinghua University)

FLOAT: Factorized Learning of Object Attributes for Improved Multi-Object Multi-Part Scene Parsing 1435
Rishubh Singh (Google Research), Pranav Gupta (International Institute of Information Technology, Hyderabad), Pradeep Shenoy (Google Research India), and Ravikiran Sarvadevabhatla (IIIT Hyderabad)

Rotationally Equivariant 3D Object Detection 1446
Hong-Xing Yu (Stanford University), Jiajun Wu (Stanford University), and Li Yi (Tsinghua University)

AUV-Net: Learning Aligned UV Maps for Texture Transfer and Synthesis 1455
Zhiqin Chen (Simon Fraser University), Kangxue Yin (NVIDIA), and Sanja Fidler (University of Toronto, NVIDIA)

Poster 1.1: 3D From Single Images

Learning To Estimate Robust 3D Human Mesh From In-the-Wild Crowded Scenes 1465
Hongsuk Choi (Seoul National University), Gyeongsik Moon (Seoul National University), JoonKyu Park (Seoul National Univ.), and Kyoung Mu Lee (Seoul National University)

Human Mesh Recovery From Multiple Shots 1475
Georgios Pavlakos (UC Berkeley), Jitendra Malik (University of California at Berkeley), and Angjoo Kanazawa (University of California Berkeley)

HandOccNet: Occlusion-Robust 3D Hand Mesh Estimation Network 1486
JoonKyu Park (Seoul National Univ.), Yeonguk Oh (Seoul National University), Gyeongsik Moon (Seoul National University), Hongsuk Choi (Seoul National University), and Kyoung Mu Lee (Seoul National University)

Photorealistic Monocular 3D Reconstruction of Humans Wearing Clothing 1496
Thiemo Alldieck (Google), Mihai Zanfir (Google), and Cristian Sminchisescu (Google)

Disentangled3D: Learning a 3D Generative Model With Disentangled Geometry and Appearance From Monocular Images 1506
Ayush Tewari (MIT), Mallikarjun B R (Max Planck Institute for Informatics), Xingang Pan (Max Planck Institute for Informatics), Ohad Fried (IDC Herzliya), Maneesh Agrawala (Stanford University), and Christian Theobalt (MPI Informatik)

NeuralHDHair: Automatic High-Fidelity Hair Modeling From a Single Image Using Implicit Neural Representations 1516
Keyu Wu (Zhejiang University), Yifan Ye (Zhejiang University), Lingchen Yang (ETH, Zurich), Hongbo Fu (City University of Hong Kong), Kun Zhou (Zhejiang University), and Youyi Zheng (Zhejiang University)

Topologically-Aware Deformation Fields for Single-View 3D Reconstruction 1526
Shivam Duggal (Carnegie Mellon University) and Deepak Pathak (Carnegie Mellon University)

Generating Diverse 3D Reconstructions From a Single Occluded Face Image 1537
Rahul Dey (Michigan State University) and Vishnu Naresh Boddeti (Michigan State University)

LOLNerf: Learn From One Look 1548
Daniel Rebain (Google Inc.), Mark Matthews (Google Inc.), Kwang Moo Yi (University of British Columbia), Dmitry Lagun (Google Brain and University of Toronto), and Andrea Tagliasacchi (Google)

Learning Local Displacements for Point Cloud Completion 1558
Yida Wang (Technische Universität München), David Joseph Tan (Google), Nassir Navab (TU Munich, Germany), and Federico Tombari (Google, TU Munich)

Exploiting Pseudo Labels in a Self-Supervised Learning Framework for Improved Monocular Depth Estimation 1568
Andra Petrovai (Technical University of Cluj-Napoca) and Sergiu Nedevschi (Technical University of Cluj-Napoca)

Dimension Embeddings for Monocular 3D Object Detection 1579
Yunpeng Zhang (Tsinghua University), Wenzhao Zheng (Tsinghua University), Zheng Zhu (Tsinghua University), Guan Huang (Institute of Automation, Chinese Academy of Sciences), Dalong Du (PhiGent Technology Co., Ltd), Jie Zhou (Tsinghua University), and Jiwen Lu (Tsinghua University)

Understanding 3D Object Articulation in Internet Videos 1589
Shengyi Qian (University of Michigan), Linyi Jin (University of Michigan), Chris Rockwell (University of Michigan), Siyi Chen (University of Michigan), and David F. Fouhey (University of Michigan)

P3Depth: Monocular Depth Estimation With a Piecewise Planarity Prior 1600
Vaishakh Patil (ETH Zurich), Christos Sakaridis (ETH Zurich), Alexander Liniger (ETH Zurich), and Luc Van Gool (ETH Zurich)

Neural Face Identification in a 2D Wireframe Projection of a Manifold Object 1612
Kehan Wang (University of California, Berkeley), Jia Zheng (Manycore Tech Inc.), and Zihan Zhou (Manycore Tech Inc.)

PanopticDepth: A Unified Framework for Depth-Aware Panoptic Segmentation 1622
Naiyu Gao (Institue of Automation, Chinese Academy of Sciences; University of Chinese Academy of Sciences), Fei He (Institue of Automation, Chinese Academy of Sciences; University of Chinese Academy of Sciences), Jian Jia (Institute of Automation, Chinese Academy of Sciences), Yanhu Shan (horizon robotics), Haoyang Zhang (Horizon Robotics), Xin Zhao (Institute of Automation, Chinese Academy of Sciences), and Kaiqi Huang (Institute of Automation, Chinese Academy of Sciences)

Stability-Driven Contact Reconstruction From Monocular Color Images 1633
Zimeng Zhao (Southeast University), Binghui Zuo (Southeast University), Wei Xie (Southeast University), and Yangang Wang (Southeast University)

LGT-Net: Indoor Panoramic Room Layout Estimation With Geometry-Aware Transformer Network .. 1644
Zhigang Jiang (East China Normal University), Zhongzheng Xiang (Shanghai Yiwo Information Technology Co., Ltd.), Jinhua Xu (East China Normal University), and Ming Zhao (Shanghai Yiwo Information Technology Limited)

Collaborative Learning for Hand and Object Reconstruction With Attention-Guided Graph Convolution 1654
Tze Ho Elden Elden Tse (University of Birmingham), Kwang In Kim (UNIST), Ales Ales Leonardis (University of Birmingham), and Hyung Jin Chang (University of Birmingham)

RM-Depth: Unsupervised Learning of Recurrent Monocular Depth in Dynamic Scenes 1665
Tak-Wai Hui (H-1 Research)

Exploring Geometric Consistency for Monocular 3D Object Detection 1675
Qing Lian (Hong Kong University of Science and Technology), Botao Ye (Institute of Computing Technology, Chinese Academy of Sciences), Ruijia Xu (Hong Kong University of Science and Technology), Weilong Yao (Shanghai Xiantu Intelligent Technology Co., Ltd.), and Tong Zhang (Hong Kong University of Science and Technology)

Learning 3D Object Shape and Layout Without 3D Supervision 1685
Georgia Gkioxari (Facebook AI Resaerch), Nikhila Ravi (Facebook AI Research), and Justin Johnson (University of Michigan)

Single-Stage 3D Geometry-Preserving Depth Estimation Model Training on Dataset Mixtures With Uncalibrated Stereo Data 1695
Nikolay Patakin (Samsung AI Center Moscow), Anna Vorontsova (Samsung AI Center), Mikhail Artemyev (Samsung), and Anton Konushin (Samsung AI Center Moscow)

Occluded Human Mesh Recovery 1705
Rawal Khirodkar (Carnegie Mellon University), Shashank Tripathi (Max Planck Institute for Intelligent Systems), and Kris Kitani (Carnegie Mellon University)

LAKe-Net: Topology-Aware Point Cloud Completion by Localizing Aligned Keypoints 1716
Junshu Tang (Shanghai Jiao Tong University), Zhijun Gong (Shanghai Jiao Tong University), Ran Yi (Shanghai Jiao Tong University), Yuan Xie (East China Normal University), and Lizhuang Ma (Shanghai Jiao Tong University)

OcclusionFusion: Occlusion-Aware Motion Estimation for Real-Time Dynamic 3D Reconstruction..... 1726
Wenbin Lin (Tsinghua University), Chengwei Zheng (Tsinghua University), Jun-Hai Yong (Tsinghua University), and Feng Xu (Tsinghua University)

Depth Estimation by Combining Binocular Stereo and Monocular Structured-Light 1736
Yuhua Xu (Orbbec Research), Xiaoli Yang (Orbbec), Yushan Yu (SCUT), Wei Jia (Heifei University of Technology), Zhaobi Chu (Hefei University of Technology), and Yulan Guo (Sun Yat-sen University)

Learning From Pixel-Level Noisy Label: A New Perspective for Light Field Saliency Detection 1746
Mingtao Feng (Xidian Unversity), Kendong Liu (Xidian University), Liang Zhang (School of Computer Science and Technology, Xidian University, China), Hongshan Yu (Hunan University), Yaonan Wang (Hunan University), and Ajmal Mian (University of Western Australia)

Poster 1.1: Photogrammetry and Remote Sensing

HyperTransformer: A Textural and Spectral Feature Fusion Transformer for Pansharpening 1757
Wele Gedara Chaminda Bandara (Johns Hopkins University) and Vishal M. Patel (Johns Hopkins University)

Revisiting Near/Remote Sensing With Geospatial Attention 1768
Scott Workman (DZYNE Technologies), M. Usman Rafique (Kitware Inc.), Hunter Blanton (University of Kentucky), and Nathan Jacobs (University of Kentucky)

Memory-Augmented Deep Conditional Unfolding Network for Pan-Sharpening 1778
Gang Yang (University of Science and Technology of China), Man Zhou (University of Science and Technology of China), Keyu Yan (University of Science and Technology of China), Aiping Liu (University of Science and Technology of China), Xueyang Fu (University of Science and Technology of China), and Fan Wang (University of Science and Technology of China)

Mutual Information-Driven Pan-Sharpening 1788
Man Zhou (University of Science and Technology of China), Keyu Yan (University of Science and Technology of China), Jie Huang (University of Science and Technology of China), Zihe Yang (University of Science and Technology of China), Xueyang Fu (University of Science and Technology of China), and Feng Zhao (University of Science and Technology of China)

Sparse and Complete Latent Organization for Geospatial Semantic Segmentation 1799
Fengyu Yang (University of Michigan) and Chenyang Ma (University of Michigan)

The Probabilistic Normal Epipolar Constraint for Frame-to-Frame Rotation Optimization Under Uncertain Feature Positions 1809
Dominik Muhle (Technical University of Munich), Lukas Koestler (Technical University of Munich), Nikolaus Demmel (TU Munich), Florian Bernard (University of Bonn), and Daniel Cremers (TU Munich)

Oriented RepPoints for Aerial Object Detection 1819
Wentong Li (Zhejiang University), Yijie Chen (Zhejiang University), Kaixuan Hu (University of Electronic Science and Technology of China), and Jianke Zhu (Zhejiang University)

Using 3D Topological Connectivity for Ghost Particle Reduction in Flow Reconstruction 1829
Christina Tsalicoglou (ETH Zürich) and Thomas Rösgen (ETH Zurich)

PolyWorld: Polygonal Building Extraction With Graph Neural Networks in Satellite Images 1838
Stefano Zorzi (Graz University of Technology), Shabab Bazrafkan (Blackshark.ai), Stefan Habenschuss (Blackshark.ai), and Friedrich Fraundorfer (Graz University of Technology)

Self-Supervised Super-Resolution for Multi-Exposure Push-Frame Satellites 1848
Ngoc Long Nguyen (ENS Paris-Saclay), Jérémy Anger (ENS Paris-Saclay), Axel Davy (ENS Paris-Saclay), Pablo Arias (ENS Paris-Saclay), and Gabriele Facciolo (ENS Paris - Saclay)

Poster 1.1: Low-Level Vision

MISF: Multi-Level Interactive Siamese Filtering for High-Fidelity Image Inpainting 1859
Xiaoguang Li (University of South Carolina), Qing Guo (Nanyang Technological University), Di Lin (Tianjin University), Ping Li (The Hong Kong Polytechnic University), Wei Feng (College of Intelligence and Computing, Tianjin University, China), and Song Wang (University of South Carolina)

Iterative Deep Homography Estimation 1869
Si-Yuan Cao (Zhejiang University), Jianxin Hu (Zhejiang University), Zehua Sheng (Zhejiang University), and Hui-Liang Shen (Zhejiang University)

GCFSR: A Generative and Controllable Face Super Resolution Method Without Facial and GAN Priors 1879
Jingwen He (Bytedance Inc), Wu Shi (Shenzhen Institutes of Advanced Technology), Kai Chen (Bytedance Inc), Lean Fu (ByteDance), and Chao Dong (SIAT)

Deep Color Consistent Network for Low-Light Image Enhancement 1889
Zhao Zhang (Hefei University of Technology), Huan Zheng (Hefei University of Technology), Richang Hong (Hefei University of Technology), Mingliang Xu (Zhengzhou University), Shuicheng Yan (Sea AI Labs), and Meng Wang (Hefei University of Technology)

LAR-SR: A Local Autoregressive Model for Image Super-Resolution 1899
Baisong Guo (Shanghai Jiao Tong University), Xiaoyun Zhang (Shanghai Jiao Tong University), Haoning Wu (Shanghai Jiao Tong University), Yu Wang (Shanghai Jiao Tong University), Ya Zhang (Cooperative medianet innovation center of Shanghai Jiao Tong University), and Yan-Feng Wang (Cooperative Medianet Innovation Center, Shang hai Jiao Tong University)

Multi-Scale Memory-Based Video Deblurring 1909
Bo Ji (National University of Singapore) and Angela Yao (National University of Singapore)

Local Texture Estimator for Implicit Representation Function 1919
Jaewon Lee (DGIST) and Kyong Hwan Jin (DGIST)

Chitransformer: Towards Reliable Stereo From Cues 1929
Qing Su (Georgia State University) and Shihao Ji (Georgia State University)

BNUDC: A Two-Branched Deep Neural Network for Restoring Images From Under-Display Cameras 1940
Jaihyun Koh (Samsung Display), Jangho Lee (Seoul National University), and Sungroh Yoon (Seoul National University)

ISNAS-DIP: Image-Specific Neural Architecture Search for Deep Image Prior 1950
Metin Ersin Arican (Boğaziçi University), Ozgur Kara (Bogazici University), Gustav Bredell (ETH Zürich), and Ender Konukoglu (ETH Zurich)

IFRNet: Intermediate Feature Refine Network for Efficient Frame Interpolation 1959
Lingtong Kong (Shanghai Jiao Tong University), Boyuan Jiang (Tencent Youtu Lab), Donghao Luo (Tencent), Wenqing Chu (Tencent), Xiaoming Huang (Tencent Youtu), Ying Tai (Tencent YouTu), Chengjie Wang (Tencent), and Jie Yang (Shanghai Jiao Tong University)

Learning Graph Regularisation for Guided Super-Resolution 1969
Riccardo de Lutio (ETHZ), Alexander Becker (ETH Zurich), Stefano D'Aronco (ETHZ), Stefania Russo (ETH Zurich), Jan D. Wegner (ETH Zurich), and Konrad Schindler (ETH Zurich)

Self-Supervised Deep Image Restoration via Adaptive Stochastic Gradient Langevin Dynamics ... 1979
Weixi Wang (National University of Singapore), Ji Li (National University of Singapore), and Hui Ji (National University of Singapore)

Self-Supervised Arbitrary-Scale Point Clouds Upsampling via Implicit Neural Representation..... 1989
Wenbo Zhao (Peng Cheng Laboratory), Xianming Liu (Harbin Institute of Technology), Zhiwei Zhong (Harbin Institute of Technology), Junjun Jiang (Harbin Institute of Technology), Wei Gao (SECE, Shenzhen Graduate School, Peking University), Ge Li (SECE, Shenzhen Graduate School, Peking University), and Xiangyang Ji (Tsinghua University)

Noise Distribution Adaptive Self-Supervised Image Denoising Using Tweedie Distribution and Score Matching 1998
Kwanyoung Kim (Korea Advanced Institute of Science and Technology), Taesung Kwon (KAIST), and Jong Chul Ye (Kim Jaechul Graduate School of AI, KAIST, Korea)

Unpaired Deep Image Deraining Using Dual Contrastive Learning 2007
Xiang Chen (Shenyang Aerospace University), Jinshan Pan (Nanjing University of Science and Technology), Kui Jiang (National Engineering Research Center for Multimedia Software, School of Computer Science, Wuhan University), Yufeng Li (Shenyang Aerospace University), Yufeng Huang (Shenyang Aerospace University), Caihua Kong (Shenyang Aerospace University), Longgang Dai (Shenyang Aerospace University), and Zhentao Fan (Shenyang Aerospace University)

Blind2Unblind: Self-Supervised Image Denoising With Visible Blind Spots 2017
Zejin Wang (Institute of Automation, Chinese Academy of Sciences), Jiazheng Liu (Institute of AutomationChinese Academy of Sciences), Guoqing Li (Institute of Automation, Chinese Academy of Sciences), and Hua Han (Institute of Automation, Chinese Academy of Sciences)

Self-Augmented Unpaired Image Dehazing via Density and Depth Decomposition 2027
Yang Yang (Tianjin University), Chaoyue Wang (JD.com), Risheng Liu (Dalian University of Technology), Lin Zhang (Tongji University, China), Xiaojie Guo (Tianjin University), and Dacheng Tao (JD.com)

VideoINR: Learning Video Implicit Neural Representation for Continuous Space-Time Super-Resolution 2037
Zeyuan Chen (University of Science and Technology of China), Yinbo Chen (UC San Diego), Jingwen Liu (University of California, San Diego), Xingqian Xu (UIUC), Vidit Goel (Indian Institute of Technology, Kharagpur), Zhangyang Wang (University of Texas at Austin), Humphrey Shi (U of Oregon/UIUC), and Xiaolong Wang (UCSD)

Fast Algorithm for Low-Rank Tensor Completion in Delay-Embedded Space 2048
Ryuki Yamamoto (Nagoya Institute of Technology), Hidekata Hontani (Nagoya Institute of Technology), Akira Imakura (University of Tsukuba), and Tatsuya Yokota (Nagoya Institute of Technology)

Exploring and Evaluating Image Restoration Potential in Dynamic Scenes 2057
Cheng Zhang (Northwestern Polytechnical University), Shaolin Su (Northwestern Polytechnical University), Yu Zhu (Northwestern Polytechnical University), Qingsen Yan (The University of Adelaide), Jinqiu Sun (Northwestern Polytechnical University), and Yanning Zhang (Northwestern Polytechnical University)

GIQE: Generic Image Quality Enhancement via Nth Order Iterative Degradation 2067
Pranjay Shyam (Korea Advanced Institute of Science and technology), Kyung-Soo Kim (KAIST), and Kuk-Jin Yoon (KAIST)

Does Text Attract Attention on E-Commerce Images: A Novel Saliency Prediction Dataset and Method 2078
Lai Jiang (BUAA), Yifei Li (Beihang university), Shengxi Li (Imperial College London), Mai Xu (BUAA), Se Lei (Beihang university), Yichen Guo (Beihang university), and Bo Huang (ALIBABA GROUP)

IDR: Self-Supervised Image Denoising via Iterative Data Refinement 2088
Yi Zhang (CUHK), Dasong Li (Chinese University of Hong Kong), Ka Lung Law (Sensetime), Xiaogang Wang (Chinese University of Hong Kong, Hong Kong), Hongwei Qin (Sensetime), and Hongsheng Li (The Chinese University of Hong Kong)

ABPN: Adaptive Blend Pyramid Network for Real-Time Local Retouching of Ultra High-Resolution Photo 2098
Biwen Lei (Alibaba), Xiefan Guo (Individual), Hongyu Yang (Individual), Miaomiao Cui (Alibaba-inc), Xuansong Xie (Alibaba), and Di Huang (Individual)

Texture-Based Error Analysis for Image Super-Resolution 2108
Salma Abdel Magid (Harvard University), Zudi Lin (Harvard University), Donglai Wei (Boston College), Yulun Zhang (ETH Zurich), Jinjin Gu (The University of Sydney), and Hanspeter Pfister (Harvard University)

Blind Image Super-Resolution With Elaborate Degradation Modeling on Noise and Kernel 2118
Zongsheng Yue (The University of Hong Kong), Qian Zhao (Xi'an Jiaotong University), Jianwen Xie (Baidu Research), Lei Zhang (Hong Kong Polytechnic University, Hong Kong, China), Deyu Meng (Xi'an Jiaotong University), and Kwan-Yee K. Wong (The University of Hong Kong)

KNN Local Attention for Image Restoration 2129
Hunsang Lee (Yonsei Univ.), Hyesong Choi (Ewha w. university), Kwanghoon Sohn (Yonsei Univ.), and Dongbo Min (Ewha Womans University)

Can You Spot the Chameleon? Adversarially Camouflaging Images From Co-Salient Object Detection 2140
Ruijun Gao (College of Intelligence and Computing, Tianjin University, China), Qing Guo (Nanyang Technological University), Felix Juefei-Xu (Alibaba Group, USA), Hongkai Yu (Cleveland State University), Huazhu Fu (IHPC, ASTAR), Wei Feng (College of Intelligence and Computing, Tianjin University, China), Yang Liu (Nanyang Technology University, Singapore), and Song Wang (University of South Carolina)

Zoom in and Out: A Mixed-Scale Triplet Network for Camouflaged Object Detection 2150
Youwei Pang (Dalian University of Technology), Xiaoqi Zhao (Dalian University of Technology), Tian-Zhu Xiang (Inception Institute of Artificial Intelligence), Lihe Zhang (Dalian University of Technology), and Huchuan Lu (Dalian University of Technology)

Poster 1.1: Behavior Analysis

Self-Supervised Keypoint Discovery in Behavioral Videos 2161
Jennifer J. Sun (Caltech), Serim Ryou (California Institute of Technology), Roni H. Goldshmid (California Institute of Technology), Brandon Weissbourd (Caltech), John O. Dabiri (California Institute of Technology), David J. Anderson (Caltech), Ann Kennedy (Northwestern University), Yisong Yue (Caltech), and Pietro Perona (California Institute of Technology)

Learning To Align Sequential Actions in the Wild 2171
Weizhe Liu (EPFL), Bugra Tekin (Microsoft), Huseyin Coskun (Technical University of Munich), Vibhav Vineet (Microsoft Research), Pascal Fua (EPFL, Switzerland), and Marc Pollefeys (ETH Zurich / Microsoft)

Dynamic 3D Gaze From Afar: Deep Gaze Estimation From Temporal Eye-Head-Body Coordination ... 2182
Soma Nonaka (Kyoto University), Shohei Nobuhara (Kyoto University), and Ko Nishino (Kyoto University)

End-to-End Human-Gaze-Target Detection With Transformers 2192
Danyang Tu (Shanghai Jiao Tong University), Xiongkuo Min (Shanghai Jiao Tong University), Huiyu Duan (Shanghai Jiao Tong University), Guodong Guo (Baidu), Guangtao Zhai (Shanghai Jiao Tong University), and Wei Shen (Shanghai Jiao Tong University)

Automatic Synthesis of Diverse Weak Supervision Sources for Behavior Analysis 2201
Albert Tseng (Caltech), Jennifer J. Sun (Caltech), and Yisong Yue (Caltech)

MUSE-VAE: Multi-Scale VAE for Environment-Aware Long Term Trajectory Prediction 2211
Mihee Lee (Rutgers University), Samuel S. Sohn (Rutgers University), Seonghyeon Moon (Rutgers University), Sejong Yoon (The College of New Jersey), Mubbasir Kapadia (Rutgers University), and Vladimir Pavlovic (Rutgers University)

Graph-Based Spatial Transformer With Memory Replay for Multi-Future Pedestrian Trajectory Prediction 2221
Lihuan Li (University of New South Wales), Maurice Pagnucco (UNSW), and Yang Song (University of New South Wales)

End-to-End Trajectory Distribution Prediction Based on Occupancy Grid Maps 2232
Ke Guo (The University of Hong Kong), Wenxi Liu (Fuzhou University), and Jia Pan (The University of Hong Kong)

Learning Affordance Grounding From Exocentric Images 2242
Hongchen Luo (University of Science and Technology of China), Wei Zhai (University of Science and Technology of China), Jing Zhang (The University of Sydney), Yang Cao (University of Science and Technology of China), and Dacheng Tao (JD.com)

Poster 1.1: Vision Applications & Systems

3D Scene Painting via Semantic Image Synthesis 2252
Jaebong Jeong (POSTECH), Janghun Jo (POSTECH), Sunghyun Cho (POSTECH), and Jaesik Park (POSTECH)

Learning Invisible Markers for Hidden Codes in Offline-to-Online Photography 2263
Jun Jia (Shanghai Jiao Tong University), Zhongpai Gao (Shanghai Jiao Tong University), Dandan Zhu (Shanghai Jiao Tong University), Xiongkuo Min (Shanghai Jiao Tong University), Guangtao Zhai (Shanghai Jiao Tong University), and Xiaokang Yang (Shanghai Jiao Tong University of China)

ETHSeg: An Amodel Instance Segmentation Network and a Real-World Dataset for X-Ray Waste Inspection 2273
Lingteng Qiu (Shenzhen Research Institute of Big Data, the Chinese University of Hong Kong (Shenzhen)), Zhangyang Xiong (The Chinese University of Hong Kong, Shenzhen), Xuhao Wang (The Chinese University of Hong Kong, Shenzhen), Kenkun Liu (Chinese University of Hong Kong (Shenzhen)), Yihan Li (CUHK (SZ)), Guanying Chen (The Chinese University of Hong Kong, Shenzhen), Xiaoguang Han (Shenzhen Research Institute of Big Data, the Chinese University of Hong Kong (Shenzhen)), and Shuguang Cui (The Chinese University of Hong Kong, Shenzhen)

Doodle It Yourself: Class Incremental Learning by Drawing a Few Sketches 2283
Ayan Kumar Bhunia (University of Surrey), Viswanatha Reddy Gajjala (University of Wisconsin-Madison), Subhadeep Koley (West Bengal University of Technology), Rohit Kundu (Jadavpur University), Aneeshan Sain (University of Surrey), Tao Xiang (University of Surrey), and Yi-Zhe Song (University of Surrey)

Image Disentanglement Autoencoder for Steganography Without Embedding 2293
Xiyao Liu (Central South University), Ziping Ma (Peking University), Junxing Ma (Central South University), Jian Zhang (Central South University), Gerald Schaefer (Loughborough University), and Hui Fang (Loughborough University)

Adaptive Hierarchical Representation Learning for Long-Tailed Object Detection 2303
Banghuai Li (Momenta)

Semiconductor Defect Detection by Hybrid Classical-Quantum Deep Learning 2313
Yuan-Fu Yang (National Tsing Hua University) and Min Sun (NTHU)

Density-Preserving Deep Point Cloud Compression 2323
Yun He (Fudan University), Xinlin Ren (Fudan University), Danhang Tang (Google Inc.), Yinda Zhang (Google), Xiangyang Xue (Fudan University), and Yanwei Fu (Fudan University)

Graph-Context Attention Networks for Size-Varied Deep Graph Matching 2333
Zheheng Jiang (Lancaster University), Hossein Rahmani (Lancaster University), Plamen Angelov (Lancaster University), Sue Black (Lancaster University), and Bryan M. Williams (Lancaster University)

TransWeather: Transformer-Based Restoration of Images Degraded by Adverse Weather Conditions 2343
Jeya Maria Jose Valanarasu (Johns Hopkins University), Rajeev Yasarla (Johns Hopkins University, Whiting School of Engineering), and Vishal M. Patel (Johns Hopkins University)

ObjectFormer for Image Manipulation Detection and Localization 2354
Junke Wang (Fudan University), Zuxuan Wu (UMD), Jingjing Chen (Fudan University), Xintong Han (Huya Inc), Abhinav Shrivastava (University of Maryland), Ser-Nam Lim (Meta AI), and Yu-Gang Jiang (Fudan University)

Sequential Voting With Relational Box Fields for Active Object Detection 2364
Qichen Fu (Carnegie Mellon University), Xingyu Liu (Carnegie Mellon University), and Kris Kitani (Carnegie Mellon University)

Efficient Classification of Very Large Images With Tiny Objects 2374
Fanjie Kong (Duke University) and Ricardo Henao (Duke University)

Partially Does It: Towards Scene-Level FG-SBIR With Partial Input 2385
Pinaki Nath Chowdhury (University of Surrey), Ayan Kumar Bhunia (University of Surrey), Viswanatha Reddy Gajjala (University of Wisconsin-Madison), Aneeshan Sain (University of Surrey), Tao Xiang (University of Surrey), and Yi-Zhe Song (University of Surrey)

Long-Term Visual Map Sparsification With Heterogeneous GNN 2396
Ming-Fang Chang (Carnegie Mellon University), Yipu Zhao (Facebook Inc), Rajvi Shah (IIIT Hyderabad), Jakob J. Engel (TU Munich (now Oculus), ,), Michael Kaess (Carnegie Mellon University), and Simon Lucey (CMU)

Connecting the Complementary-View Videos: Joint Camera Identification and Subject Association 2406
Ruize Han (College of Intelligence and Computing, Tianjin University), Yiyang Gan (College of Intelligence and Computing, Tianjin University), Jiacheng Li (College of Intelligence and Computing, Tianjin University), Feifan Wang (College of Intelligence and Computing, Tianjin University), Wei Feng (College of Intelligence and Computing, Tianjin University, China), and Song Wang (University of South Carolina)

DiffusionCLIP: Text-Guided Diffusion Models for Robust Image Manipulation 2416
Gwanghyun Kim (Department of Bio and Brain Engineering, KAIST, Korea), Taesung Kwon (KAIST), and Jong Chul Ye (Kim Jaechul Graduate School of AI, KAIST, Korea)

Aesthetic Text Logo Synthesis via Content-Aware Layout Inferring 2426
Yizhi Wang (Peking University), Guo Pu (Peking University), Wenhan Luo (Tencent), Yexin Wang (Tencent), Pengfei Xiong (Tencent), Hongwen Kang (Tencent), and Zhouhui Lian (Peking University)

Rethinking Image Cropping: Exploring Diverse Compositions From Global Views 2436
Gengyun Jia (University of Chinese Academy of Sciences), Huaibo Huang (Institute of Automation, Chinese Academy of Sciences), Chaoyou Fu (Institute of Automation, Chinese Academy of Sciences), and Ran He (Institute of Automation, Chinese Academy of Sciences)

Defensive Patches for Robust Recognition in the Physical World 2446
Jiakai Wang (Beihang University), Zixin Yin (Beihang University), Pengfei Hu (Beihang University), Aishan Liu (Beihang University), Renshuai Tao (Beihang University), Haotong Qin (Beihang University), Xianglong Liu (BUAA), and Dacheng Tao (JD.com)

Semi-Supervised Video Paragraph Grounding With Contrastive Encoder 2456
Xun Jiang (UESTC), Xing Xu (University of Electronic Science and Technology of China), Jingran Zhang (University of Electronic Science and Technology of China (UESTC)), Fumin Shen (UESTC), Zuo Cao (MEITUAN), and Heng Tao Shen (University of Electronic Science and Technology of China (UESTC))

Large-Scale Pre-Training for Person Re-Identification With Noisy Labels 2466
Dengpan Fu (University of Science and Technology of China), Dongdong Chen (Microsoft Cloud AI), Hao Yang (Microsoft Research Asia), Jianmin Bao (Microsoft Research Asia), Lu Yuan (Microsoft), Lei Zhang (International Digital Economy Academy (IDEA), ,), Houqiang Li (University of Science and Technology of China), Fang Wen (Microsoft Research Asia), and Dong Chen (Microsoft Research Asia)

Meta Distribution Alignment for Generalizable Person Re-Identification 2477
Hao Ni (The University of Electronic Science and Technology of China), Jingkuan Song (UESTC), Xiaopeng Luo (University of Electronic Science and Technology of China), Feng Zheng (SUSTech), Wen Li (University of Electronic Science and Technology of China), and Heng Tao Shen (University of Electronic Science and Technology of China (UESTC), ,)

FvOR: Robust Joint Shape and Pose Optimization for Few-View Object Reconstruction 2487
Zhenpei Yang (The University of Texas at Austin), Zhile Ren (Apple Inc.), Miguel Angel Bautista (Apple), Zaiwei Zhang (Amazon), Qi Shan (Apple Inc.), and Qixing Huang (The University of Texas at Austin)

It's About Time: Analog Clock Reading in the Wild 2498
Charig Yang (University of Oxford), Weidi Xie (Shanghai Jiao Tong University), and Andrew Zisserman (University of Oxford)

Consistency Driven Sequential Transformers Attention Model for Partially Observable Scenes..... 2508
Samrudhdhi B. Rangrej (McGill University), Chetan L. Srinidhi (University of Toronto), and James J. Clark (McGill University)

SmartAdapt: Multi-Branch Object Detection Framework for Videos on Mobiles 2518
Ran Xu (Purdue University), Fangzhou Mu (University of Wisconsin-Madison), Jayoung Lee (Purdue University), Preeti Mukherjee (Purdue University), Somali Chaterji (Purdue University), Saurabh Bagchi (Purdue University), and Yin Li (University of Wisconsin-Madison)

Generating 3D Bio-Printable Patches Using Wound Segmentation and Reconstruction To Treat Diabetic Foot Ulcers 2529
Han Joo Chae (ROKIT Healthcare), Seunghwan Lee (ROKIT Healthcare), Hyewon Son (ROKIT Healthcare), Seungyeob Han (ROKIT Healthcare), and Taebin Lim (ROKIT Healthcare)

Investigating the Impact of Multi-LiDAR Placement on Object Detection for Autonomous Driving 2540
Hanjiang Hu (Carnegie Mellon University), Zuxin Liu (Carnegie Mellon University), Sharad Chitlangia (Amazon), Akhil Agnihotri (University of Southern California), and Ding Zhao (Carnegie Mellon University)

Oral 1.2.1: Segmentation, Grouping and Shape Analysis

CMT-DeepLab: Clustering Mask Transformers for Panoptic Segmentation 2550
Qihang Yu (Johns Hopkins University), Huiyu Wang (JHU), Dahun Kim (KAIST), Siyuan Qiao (Google), Maxwell Collins (Google Inc.), Yukun Zhu (Google Inc.), Hartwig Adam (Google), Alan Yuille (Johns Hopkins University), and Liang-Chieh Chen (Google Inc.)

Unsupervised Hierarchical Semantic Segmentation With Multiview Cosegmentation and Clustering Transformers 2561
Tsung-Wei Ke (UC Berkeley), Jyh-Jing Hwang (Waymo), Yunhui Guo (UC Berkeley / ICSI), Xudong Wang (UC Berkeley / ICSI), and Stella X. Yu (UC Berkeley / ICSI)

Rethinking Semantic Segmentation: A Prototype View 2572
Tianfei Zhou (ETH Zurich), Wenguan Wang (Eidgenössische Technische Hochschule Zürich), Ender Konukoglu (ETH Zurich), and Luc Van Gool (ETH Zurich)

Semantic-Aware Domain Generalized Segmentation 2584
Duo Peng (Sichuan University), Yinjie Lei (Sichuan University), Munawar Hayat (Monash University), Yulan Guo (Sun Yat-sen University), and Wen Li (University of Electronic Science and Technology of China)

Adaptive Early-Learning Correction for Segmentation From Noisy Annotations 2596
Sheng Liu (NYU), Kangning Liu (New York University), Weicheng Zhu (New York University Center for Data Science), Yiqiu Shen (New York University), and Carlos Fernandez-Granda (NYU)

Pointly-Supervised Instance Segmentation 2607
Bowen Cheng (UIUC), Omkar Parkhi (Facebook), and Alexander Kirillov (Facebook AI Reserach)

Joint Forecasting of Panoptic Segmentations With Difference Attention 2617
Colin Graber (UIUC), Cyril Jazra (University of Illinois at Urbana-Champaign), Wenjie Luo (Waymo LLC), Liangyan Gui (University of Illinois Urbana-Champaign), and Alexander G. Schwing (UIUC)

FocusCut: Diving Into a Focus View in Interactive Segmentation 2627
Zheng Lin (Nankai University), Zheng-Peng Duan (Nankai University), Zhao Zhang (Nankai University), Chun-Le Guo (Nankai University), and Ming-Ming Cheng (Nankai University)

Human Instance Matting via Mutual Guidance and Multi-Instance Refinement 2637
Yanan Sun (HKUST), Chi-Keung Tang (Hong Kong University of Science and Technology), and Yu-Wing Tai (Kuaishou Technology / HKUST)

Deformable Sprites for Unsupervised Video Decomposition 2647
Vickie Ye (UC Berkeley), Zhengqi Li (Google Inc.), Richard Tucker (Google), Angjoo Kanazawa (University of California Berkeley), and Noah Snavely (Google)

Eigencontours: Novel Contour Descriptors Based on Low-Rank Approximation 2657
Wonhui Park (Korea university), Dongkwon Jin (Korea university), and Chang-Su Kim (Korea university)

Robust and Accurate Superquadric Recovery: A Probabilistic Approach 2666
Weixiao Liu (Johns Hopkins University), Yuwei Wu (National University of Singapore), Sipu Ruan (National University of Singapore), and Gregory S. Chirikjian (National University of Singapore)

Medial Spectral Coordinates for 3D Shape Analysis 2676
Morteza Rezanejad (University of Toronto), Mohammad Khodadad (Sharif University of Technology), Hamidreza Mahyar (McMaster University), Herve Lombaert (ETS Montreal / Inria), Michael Gruninger (University of Toronto), Dirk Walther (University of Toronto), and Kaleem Siddiqi (McGill University)

Scribble-Supervised LiDAR Semantic Segmentation 2687
Ozan Unal (ETH Zurich), Dengxin Dai (ETH Zurich), and Luc Van Gool (ETH Zurich)

SoftGroup for 3D Instance Segmentation on Point Clouds 2698
Thang Vu (KAIST), Kookhoi Kim (KAIST), Tung M. Luu (Korea Advanced Institute of Science And Technology), Thanh Nguyen (KAIST), and Chang D. Yoo (KAIST)

Oral 1.2.2: 3D From Single Images

Accurate 3D Body Shape Regression Using Metric and Semantic Attributes 2708
Vasileios Choutas (ETH Zurich), Lea Müller (Max Planck Institute for Intelligent Systems), Chun-Hao P. Huang (Max Planck Institute for Intelligent Systems), Siyu Tang (ETH Zurich), Dimitrios Tzionas (Max Planck Institute for Intelligent Systems), and Michael J. Black (Max Planck Institute for Intelligent Systems)

JIFF: Jointly-Aligned Implicit Face Function for High Quality Single View Clothed Human Reconstruction 2719
Yukang Cao (the University of Hong Kong), Guanying Chen (The Chinese University of Hong Kong, Shenzhen), Kai Han (The University of Hong Kong), Wenqi Yang (The University of Hong Kong), and Kwan-Yee K. Wong (The University of Hong Kong)

Tracking People by Predicting 3D Appearance, Location and Pose 2730
Jathushan Rajasegaran (UC Berkeley), Georgios Pavlakos (UC Berkeley), Angjoo Kanazawa (University of California Berkeley), and Jitendra Malik (University of California at Berkeley)

ArtiBoost: Boosting Articulated 3D Hand-Object Pose Estimation via Online Exploration and Synthesis 2740
Lixin Yang (Shanghai Jiao Tong University), Kailin Li (Shanghai Jiao Tong University), Xinyu Zhan (Shanghai Jiao Tong University), Jun Lv (Shanghai Jiao Tong University), Wenqiang Xu (Shanghai Jiao Tong University), Jiefeng Li (Shanghai Jiao Tong University), and Cewu Lu (Shanghai Jiao Tong University)

Interacting Attention Graph for Single Image Two-Hand Reconstruction 2751
Mengcheng Li (Tsinghua University), Liang An (Tsinghua University), Hongwen Zhang (Tsinghua University), Lianpeng Wu (Hisense/Juhaokan), Feng Chen (Tsinghua University), Tao Yu (Tsinghua University), and Yebin Liu (Tsinghua University)

3D Human Tongue Reconstruction From Single "In-the-Wild" Images 2761
Stylianos Ploumpis (Huawei Technologies Co. Ltd), Stylianos Moschoglou (Huawei Technologies Co. Ltd), Vasileios Triantafyllou (Huawei Technologies Co. Ltd), and Stefanos Zafeiriou (Huawei Technologies Co. Ltd)

EPro-PnP: Generalized End-to-End Probabilistic Perspective-N-Points for Monocular Object Pose Estimation 2771
Hansheng Chen (Tongji University), Pichao Wang (Alibaba Group), Fan Wang (Alibaba Group), Wei Tian (Tongji University), Lu Xiong (Tongji University), and Hao Li (Alibaba Group)

Diversity Matters: Fully Exploiting Depth Clues for Reliable Monocular 3D Object Detection..... 2781
Zhuoling Li (Tsinghua University), Zhan Qu (Noah's Ark Lab, Huawei Technologies Co., Ltd.), Yang Zhou (Noah's Ark Lab, Huawei Technologies Co., Ltd.), Jianzhuang Liu (Huawei Noah's Ark Lab), Haoqian Wang (Tsinghua Shenzhen International Graduate School, Tsinghua University), and Lihui Jiang (Huawei Technologies Co., Ltd)

OmniFusion: 360 Monocular Depth Estimation via Geometry-Aware Fusion 2791
Yuyan Li (University of Missouri), Yuliang Guo (BOSCH Research North America), Zhixin Yan (Bosch Research China), Xinyu Huang (BOSCH Research North America), Ye Duan (University of Missouri - Columbia), and Liu Ren (BOSCH Research North America)

Gated2Gated: Self-Supervised Depth Estimation From Gated Images 2801
Amanpreet Walia (Algolux), Stefanie Walz (Mercedes-Benz AG), Mario Bijelic (Princeton University), Fahim Mannan (Algolux), Frank Julca-Aguilar (Algolux), Michael Langer (McGill University), Werner Ritter (Mercedes-Benz AG), and Felix Heide (Princeton University)

IRISformer: Dense Vision Transformers for Single-Image Inverse Rendering in Indoor Scenes 2812
Rui Zhu (University of California San Diego), Zhengqin Li (Meta), Janarbek Matai (Qualcomm AI Research), Fatih Porikli (Qualcomm AI Research), and Manmohan Chandraker (UC San Diego)

Egocentric Scene Understanding via Multimodal Spatial Rectifier 2822
Tien Do (University of Minnesota), Khiem Vuong (University of Minnesota), and Hyun Soo Park (The University of Minnesota)

Multi-View Depth Estimation by Fusing Single-View Depth Probability With Multi-View Geometry 2832
Gwangbin Bae (University of Cambridge), Ignas Budvytis (Department of Engineering, University of Cambridge), and Roberto Cipolla (University of Cambridge)

The Implicit Values of a Good Hand Shake: Handheld Multi-Frame Neural Depth Refinement 2842
Ilya Chugunov (Princeton University), Yuxuan Zhang (Princeton University), Zhihao Xia (Adobe), Xuaner Zhang (Adobe), Jiawen Chen (Adobe, Inc.), and Felix Heide (Princeton University)

BANMo: Building Animatable 3D Neural Models From Many Casual Videos 2853
Gengshan Yang (Carnegie Mellon University), Minh Vo (Facebook Reality Labs), Natalia Neverova (Facebook AI Research), Deva Ramanan (Carnegie Mellon University), Andrea Vedaldi (University of Oxford / Facebook AI Research), and Hanbyul Joo (Facebook AI Research)

Oral 1.2.3: Video Analysis & Understanding

Self-Supervised Video Transformer 2864
Kanchana Ranasinghe (Stony Brook University), Muzammal Naseer (Australian National University (ANU), ,), Salman Khan (MBZUAI/ANU), Fahad Shahbaz Khan (MBZUAI), and Michael S. Ryoo (Stony Brook/Google)

Temporally Efficient Vision Transformer for Video Instance Segmentation 2875
Shusheng Yang (Huazhong University of Science and Technology), Xinggang Wang (Huazhong University of Science and Technology), Yu Li (International Digital Economy Academy), Yuxin Fang (Huazhong University of Science and Technology), Jiemin Fang (Huazhong University of Science and Technology), Wenyu Liu (Huazhong University of Science and Technology), Xun Zhao (Tencent Company), and Ying Shan (Tencent)

VISOLO: Grid-Based Space-Time Aggregation for Efficient Online Video Instance Segmentation.. 2886
Su Ho Han (Yonsei University), Sukjun Hwang (Yonsei University), Seoung Wug Oh (Adobe Research), Yeonchool Park (LG Electronics), Hyunwoo Kim (LG AI Research), Min-Jung Kim (KAIST), and Seon Joo Kim (Yonsei University)

Temporal Alignment Networks for Long-Term Video 2896
Tengda Han (University of Oxford), Weidi Xie (Shanghai Jiao Tong University), and Andrew Zisserman (University of Oxford)

Revisiting the "Video" in Video-Language Understanding 2907
Shyamal Buch (Stanford University), Cristóbal Eyzaguirre (Stanford University), Adrien Gaidon (Toyota Research Institute), Jiajun Wu (Stanford University), Li Fei-Fei (Stanford University), and Juan Carlos Niebles (Salesforce & Stanford University)

Invariant Grounding for Video Question Answering 2918
Yicong Li (National University of Singapore), Xiang Wang (National University of Singapore), Junbin Xiao (Sea AI Lab), Wei Ji (National University of Singapore), and Tat-Seng Chua (National university of Singapore)

P3IV: Probabilistic Procedure Planning From Instructional Videos With Weak Supervision 2928
He Zhao (York University), Isma Hadji (Samsung AI Center - Toronto), Nikita Dvornik (Samsung), Konstantinos G. Derpanis (York University), Richard P. Wildes (York University), and Allan D. Jepson (Samsung Toronto AIC)

FineDiving: A Fine-Grained Dataset for Procedure-Aware Action Quality Assessment 2939
Jinglin Xu (Tsinghua University), Yongming Rao (Tsinghua University), Xumin Yu (Tsinghua University), Guangyi Chen (MBZUAI), Jie Zhou (Tsinghua University), and Jiwen Lu (Tsinghua University)

Cross-Model Pseudo-Labeling for Semi-Supervised Action Recognition 2949
Yinghao Xu (Chinese University of Hong Kong), Fangyun Wei (Microsoft Research Asia), Xiao Sun (Microsoft Research Asia), Ceyuan Yang (Chinese University of Hong Kong), Yujun Shen (Dept. of IE, CUHK), Bo Dai (Shanghai AI Lab), Bolei Zhou (UCLA), and Stephen Lin (Microsoft Research)

Revisiting Skeleton-Based Action Recognition 2959
Haodong Duan (The Chinese University of Hong Kong), Yue Zhao (University of Texas at Austin), Kai Chen (SenseTime Research), Dahua Lin (The Chinese University of Hong Kong), and Bo Dai (Shanghai AI Lab)

OpenTAL: Towards Open Set Temporal Action Localization 2969
Wentao Bao (Rochester Institute of Technology), Qi Yu (Rochester Institute of Technology), and Yu Kong (Rochester Institute of Technology)

Dual-AI: Dual-Path Actor Interaction Learning for Group Activity Recognition 2980
Mingfei Han (University of Technology Sydney; DATA61, CSIRO), David Junhao Zhang (National University of Singapore), Yali Wang (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences), Rui Yan (Nanjing University of Science and Technology), Lina Yao (University of New South Wales), Xiaojun Chang (University of Technology Sydney), and Yu Qiao (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences)

TransRank: Self-Supervised Video Representation Learning via Ranking-Based Transformation Recognition 2990
Haodong Duan (The Chinese University of Hong Kong), Nanxuan Zhao (Chinese University of Hong Kong), Kai Chen (SenseTime Research), and Dahua Lin (The Chinese University of Hong Kong)

Revealing Occlusions With 4D Neural Fields 3001
Basile Van Hoorick (Columbia University), Purva Tendulkar (Columbia University), Dídac Surís (Columbia University), Dennis Park (Toyota Research Institute), Simon Stent (Toyota Research Institute), and Carl Vondrick (Columbia University)

HODOR: High-Level Object Descriptors for Object Re-Segmentation in Video Learned From Static Images 3012
Ali Athar (RWTH Aachen), Jonathon Luiten (RWTH Aachen University), Alexander Hermans (RWTH Aachen University, Computer Vision Group), Deva Ramanan (Carnegie Mellon University), and Bastian Leibe (RWTH Aachen University)

Poster 1.2: Video Analysis & Understanding

Compositional Temporal Grounding With Structured Variational Cross-Graph Correspondence Learning 3022
Juncheng Li (Zhejiang University), Junlin Xie (Zhejiang University), Long Qian (Zhejiang University), Linchao Zhu (University of Technology, Sydney), Siliang Tang (Zhejiang University), Fei Wu (Zhejiang University, China), Yi Yang (UTS), Yueting Zhuang (Zhejiang University), and Xin Eric Wang (University of California, Santa Cruz)

UMT: Unified Multi-Modal Transformers for Joint Video Moment Retrieval and Highlight Detection 3032
Ye Liu (The Hong Kong Polytechnic University), Siyuan Li (Tencent), Yang Wu (ARC LabTencent PCG), Chang-Wen Chen (The Hong Kong Polytechnic University), Ying Shan (Tencent), and Xiaohu Qie (Tencent)

Future Transformer for Long-Term Action Anticipation 3042
Dayoung Gong (POSTECH), Joonseok Lee (POSTECH), Manjin Kim (POSTECH), Seong Jong Ha (NCSOFT), and Minsu Cho (POSTECH)

MLP-3D: A MLP-Like 3D Architecture With Grouped Time Mixing 3052
Zhaofan Qiu (JD.com), Ting Yao (JD AI Research), Chong-Wah Ngo (Singapore Management University), and Tao Mei (AI Research of JD.com)

Learning Pixel-Level Distinctions for Video Highlight Detection 3063
Fanyue Wei (University of Electronic Science and Technology of China), Biao Wang (Alibaba Group), Tiezheng Ge (Alibaba Group), Yuning Jiang (Alibaba Group), Wen Li (University of Electronic Science and Technology of China), and Lixin Duan (University of Electronic Science and Technology of China)

DR.VIC: Decomposition and Reasoning for Video Individual Counting 3073
Tao Han (Northwestern Polytechnical University, Center for OPTical IMagery Analysis and Learning), Lei Bai (University of Sydney), Junyu Gao (Northwestern Polytechnical University, Center for OPTical IMagery Analysis and Learning), Qi Wang (Northwestern Polytechnical University), and Wanli Ouyang (The University of Sydney)

Slot-VPS: Object-Centric Representation Learning for Video Panoptic Segmentation 3083
Yi Zhou (Samsung Advanced Institute of Technology, China), Hui Zhang (Beijing Samsung Telecommunication R&D Center), Hana Lee (Samsung Advanced Institute of Technology), Shuyang Sun (University of Oxford), Pingjun Li (Beijing Samsung Telecommunication R&D Center), Yangguang Zhu (University of Chinese Academy of Sciences, Beijing 100049, China), ByungIn Yoo (Samsung Advanced Institute of Technology), Xiaojuan Qi (The University of Hong Kong), and Jae-Joon Han (Samsung)

Explore Spatio-Temporal Aggregation for Insubstantial Object Detection: Benchmark Dataset and Baseline 3094
Kailai Zhou (Nanjing University), Yibo Wang (Nanjing University), Tao Lv (Nanjing University), Yunqian Li (Nanjing University), Linsen Chen (Nanjing University), Qiu Shen (Nanjing University), and Xun Cao (Nanjing University)

Video Shadow Detection via Spatio-Temporal Interpolation Consistency Training 3106
Xiao Lu (Hunan Normmal University), Yihong Cao (Hunan University), Sheng Liu (Hunan Normal University), Chengjiang Long (Meta Reality Labs), Zipei Chen (WuHan University), Xuanyu Zhou (Hunan Normal University), Yimin Yang (Lakehead University / Vector Institute), and Chunxia Xiao (Wuhan University)

Coarse-To-Fine Feature Mining for Video Semantic Segmentation 3116
Guolei Sun (ETH Zurich), Yun Liu (ETH Zurich), Henghui Ding (ETH Zurich), Thomas Probst (ETH Zurich), and Luc Van Gool (ETH Zurich)

Tencent-MVSE: A Large-Scale Benchmark Dataset for Multi-Modal Video Similarity Evaluation .. 3128
Zhaoyang Zeng (Tencent), Yongsheng Luo (Tencent), Zhenhua Liu (Tencent), Fengyun Rao (Tencent), Dian Li (Tencent.com), Weidong Guo (Tencent), and Zhen Wen (Tencent Technology (Shenzhen) Co., Ltd, ,)

Object-Region Video Transformers 3138
Roei Herzig (Tel Aviv University), Elad Ben-Avraham (Tel Aviv University), Karttikeya Mangalam (UC Berkeley), Amir Bar (Tel Aviv University), Gal Chechik (NVIDIA), Anna Rohrbach (UC Berkeley), Trevor Darrell (UC Berkeley), and Amir Globerson (Tel Aviv University, Google)

Colar: Effective and Efficient Online Action Detection by Consulting Exemplars 3150
Le Yang (Northwestern Polytechnical University), Junwei Han (NWPU, China), and Dingwen Zhang (NWPU)

SimVP: Simpler Yet Better Video Prediction 3160
Zhangyang Gao (westlake university), Cheng Tan (Westlake University), Lirong Wu (Westlake University), and Stan Z. Li (Westlake University)

Imposing Consistency for Optical Flow Estimation 3171
Jisoo Jeong (Seoul National University), Jamie Menjay Lin (Google Research), Fatih Porikli (Qualcomm AI Research), and Nojun Kwak (Seoul National University)

Stand-Alone Inter-Frame Attention in Video Models 3182
Fuchen Long (JD.com), Zhaofan Qiu (JD.com), Yingwei Pan (JD AI Research), Ting Yao (JD AI Research), Jiebo Luo (U. Rochester), and Tao Mei (AI Research of JD.com)

Video Swin Transformer 3192
Ze Liu (USTC), Jia Ning (Huazhong University of Science and Technology), Yue Cao (Microsoft Research), Yixuan Wei (Tsinghua University), Zheng Zhang (MSRA), Stephen Lin (Microsoft Research), and Han Hu (Microsoft Research Asia)

Bayesian Nonparametric Submodular Video Partition for Robust Anomaly Detection 3202
Hitesh Sapkota (Rochester Institute of Technology) and Qi Yu (Rochester Institute of Technology)

Self-Supervised Predictive Learning: A Negative-Free Method for Sound Source Localization in Visual Scenes 3212
Zengjie Song (Institute of Automation, Chinese Academy of Sciences), Yuxi Wang (CASIA), Junsong Fan (Chinese Academy of Sciences, China), Tieniu Tan (NLPR, China), and Zhaoxiang Zhang (Chinese Academy of Sciences, China)

Likert Scoring With Grade Decoupling for Long-Term Action Assessment 3222
Angchi Xu (Sun Yat-sen University), Ling-An Zeng (Sun Yat-sen University), and Wei-Shi Zheng (Sun Yat-sen University, China)

Complex Video Action Reasoning via Learnable Markov Logic Network 3232
Yang Jin (Peking University), Linchao Zhu (University of Technology, Sydney), and Yadong Mu (Peking University)

Learning From Temporal Gradient for Semi-Supervised Action Recognition 3242
Junfei Xiao (Johns Hopkins University), Longlong Jing (Waymo LLC), Lin Zhang (Carnegie Mellon University), Ju He (Johns Hopkins University), Qi She (Bytedance AI Lab), Zongwei Zhou (Johns Hopkins University), Alan Yuille (Johns Hopkins University), and Yingwei Li (Johns Hopkins University)

Semi-Supervised Video Semantic Segmentation With Inter-Frame Feature Reconstruction 3253
Jiafan Zhuang (University of Science and Technology of China), Zilei Wang (University of Science and Technology of China), and Yuan Gao (University of Science and Technology of Chlna)

Weakly Supervised Temporal Action Localization via Representative Snippet Knowledge Propagation 3262
Linjiang Huang (The Chinese University of Hong Kong), Liang Wang (NLPR, China), and Hongsheng Li (The Chinese University of Hong Kong)

Joint Hand Motion and Interaction Hotspots Prediction From Egocentric Videos 3272
Shaowei Liu (UCSD), Subarna Tripathi (Intel Labs), Somdeb Majumdar (Intel Labs), and Xiaolong Wang (UCSD)

Human Hands As Probes for Interactive Object Understanding 3283
Mohit Goyal (University of Illinois at Urbana Champaign), Sahil Modi (University of Illinois at Urbana-Champaign), Rishabh Goyal (University of Illinois Urbana-Champaign), and Saurabh Gupta (UIUC)

LD-ConGR: A Large RGB-D Video Dataset for Long-Distance Continuous Gesture Recognition ... 3294
Dan Liu (Institute of Software Chinese Academy of Sciences), Libo Zhang (Institute of Software Chinese Academy of Sciences), and Yanjun Wu (Institute of Software Chinese Academy of Sciences)

Object-Aware Video-Language Pre-Training for Retrieval 3303
Jinpeng Wang (National University of Singapore), Yixiao Ge (Tencent), Guanyu Cai (Tongji University), Rui Yan (Nanjing University of Science and Technology), Xudong Lin (Columbia University), Ying Shan (Tencent), Xiaohu Qie (Tencent), and Mike Zheng Shou (National University of Singapore)

Fast and Unsupervised Action Boundary Detection for Action Segmentation 3313
Zexing Du (Northwestern Polytechnical University), Xue Wang (Northwestern Polytechnical University), Guoqing Zhou (Northwestern Polytechnical University), and Qing Wang (Northwestern Polytechnical University)

Multiview Transformers for Video Recognition 3323
Shen Yan (Michigan State University), Xuehan Xiong (Google), Anurag Arnab (Google), Zhichao Lu (Google), Mi Zhang (Michigan State University), Chen Sun (Brown University), and Cordelia Schmid (Google)

Semi-Weakly-Supervised Learning of Complex Actions From Instructional Task Videos 3334
Yuhan Shen (Northeastern University) and Ehsan Elhamifar (Northeastern University)

Progressive Attention on Multi-Level Dense Difference Maps for Generic Event Boundary Detection 3345
Jiaqi Tang (Nanjing University), Zhaoyang Liu (SenseTime Research), Chen Qian (SenseTime), Wayne Wu (SenseTime Research), and Limin Wang (Nanjing University)

Comparing Correspondences: Video Prediction With Correspondence-Wise Losses 3355
Daniel Geng (University of Michigan), Max Hamilton (University of Michigan, Ann Arbor), and Andrew Owens (University of Michigan)

Poster 1.2: Image & Video Synthesis and Generation

Sound-Guided Semantic Image Manipulation 3367
Seung Hyun Lee (Korea University), Wonseok Roh (Korea University), Wonmin Byeon (NVIDIA Research), Sang Ho Yoon (KAIST), Chanyoung Kim (Korea University), Jinkyu Kim (Korea University), and Sangpil Kim (Korea University)

Expressive Talking Head Generation With Granular Audio-Visual Control 3377
Borong Liang (Baidu-Vis), Yan Pan (The Chinese University of Hong Kong, Shenzhen), Zhizhi Guo (Department of Computer Vision Technology (VIS), Baidu Inc.), Hang Zhou (The Chinese University of Hong Kong), Zhibin Hong (Baidu Inc.), Xiaoguang Han (Shenzhen Research Institute of Big Data, the Chinese University of Hong Kong (Shenzhen)), Junyu Han (Baidu Inc.), Jingtuo Liu (Baidu Inc.), Errui Ding (Baidu Inc.), and Jingdong Wang (Baidu)

Depth-Aware Generative Adversarial Network for Talking Head Video Generation 3387
Fa-Ting Hong (The Hong Kong University of Science and Technology), Longhao Zhang (Beijing University of Posts and Telecommunications), Li Shen (Alibaba), and Dan Xu (The Hong Kong University of Science and Technology)

Learning Motion-Dependent Appearance for High-Fidelity Rendering of Dynamic Humans From a Single Camera 3397
Jae Shin Yoon (U of Minnesota), Duygu Ceylan (Adobe Research), Tuanfeng Y. Wang (adobe research), Jingwan Lu (Adobe Research), Jimei Yang (Adobe), Zhixin Shu (Adobe Research), and Hyun Soo Park (The University of Minnesota)

Audio-Driven Neural Gesture Reenactment With Video Motion Graphs 3408
Yang Zhou (Adobe Research), Jimei Yang (Adobe), Dingzeyu Li (Adobe Research), Jun Saito (Adobe), Deepali Aneja (Adobe), and Evangelos Kalogerakis (UMass Amherst)

Portrait Eyeglasses and Shadow Removal by Leveraging 3D Synthetic Data 3419
Junfeng Lyu (Tsinghua University), Zhibo Wang (Tsinghua University), and Feng Xu (Tsinghua University)

Weakly Supervised High-Fidelity Clothing Model Generation 3430
Ruili Feng (University of Science and Technology of China), Cheng Ma (Zhejiang University), Chengji Shen (Zhejiang University), Xin Gao (Alibaba Group), Zhenjiang Liu (Alibaba Group), Xiaobo Li (Alibaba), Kairi Ou (Alibaba Group), Deli Zhao (Ant Research), and Zheng-Jun Zha (University of Science and Technology of China)

TemporalUV: Capturing Loose Clothing With Temporally Coherent UV Coordinates 3440
You Xie (TUM), Huiqi Mao (National University of Singapore), Angela Yao (National University of Singapore), and Nils Thuerey (Technical University of Munich)

Full-Range Virtual Try-On With Recurrent Tri-Level Transform 3450
Han Yang (ETH Zurich), Xinrui Yu (Harbin Institute of Technology (Shenzhen)), and Ziwei Liu (Nanyang Technological University)

Style-Based Global Appearance Flow for Virtual Try-On 3460
Sen He (University of Surrey), Yi-Zhe Song (University of Surrey), and Tao Xiang (University of Surrey)

Dressing in the Wild by Watching Dance Videos 3470
Xin Dong (bytedance), Fuwei Zhao (Sun Yat-Sen University), Zhenyu Xie (Sun Yat-sen University), Xijin Zhang (ByteDance Inc.), Daniel K. Du (ByteDance Inc.), Min Zheng (ByteDance), Xiang Long (Bytedance), Xiaodan Liang (Sun Yat-sen University), and Jianchao Yang (Bytedance Inc.)

A Brand New Dance Partner: Music-Conditioned Pluralistic Dancing Controlled by Multiple Dance Genres 3480
Jinwoo Kim (Yonsei University), Heeseok Oh (Hansung University), Seongjean Kim (Yonsei University), Hoseok Tong (Yonsei University), and Sanghoon Lee (Yonsei University, Korea)

Unpaired Cartoon Image Synthesis via Gated Cycle Mapping 3491
Yifang Men (Alibaba), Yuan Yao (Alibaba Group), Miaomiao Cui (Alibaba-inc), Zhouhui Lian (Peking University), Xuansong Xie (Alibaba), and Xian-Sheng Hua (Damo Academy, Alibaba Group)

DLFormer: Discrete Latent Transformer for Video Inpainting 3501
Jingjing Ren (South China University of Technology), Qingqing Zheng (Chinese Academic of Sciences), Yuanyuan Zhao (tencent), Xuemiao Xu (South China University of Technology), and Chen Li (Tencent)

ST-MFNet: A Spatio-Temporal Multi-Flow Network for Frame Interpolation 3511
Duolikun Danier (University of Bristol), Fan Zhang (University of Bristol), and David Bull (University of bristol)

Video Frame Interpolation With Transformer 3522
Liying Lu (The Chinese University of Hong Kong), Ruizheng Wu (The Chinese University of Hong KONG), Huaijia Lin (The Chinese University of Hong Kong), Jiangbo Lu (SmartMore Corporation), and Jiaya Jia (Chinese University of Hong Kong)

Long-Term Video Frame Interpolation via Feature Propagation 3533
Dawit Mureja Argaw (KAIST) and In So Kweon (KAIST)

Many-to-Many Splatting for Efficient Video Frame Interpolation 3543
Ping Hu (Boston University), Simon Niklaus (Adobe Research), Stan Sclaroff (Boston University), and Kate Saenko (Boston University)

Look Outside the Room: Synthesizing a Consistent Long-Term 3D Scene Video From a Single Image 3553
Xuanchi Ren (HKUST) and Xiaolong Wang (UCSD)

Spatial-Temporal Space Hand-in-Hand: Spatial-Temporal Video Super-Resolution via Cycle-Projected Mutual Learning 3564
Mengshun Hu (Wuhan University), Kui Jiang (Wuhan University), Liang Liao (National Institute of Informatics), Jing Xiao (Wuhan University), Junjun Jiang (Harbin Institute of Technology), and Zheng Wang (Wuhan University)

Playable Environments: Video Manipulation in Space and Time 3574
Willi Menapace (University of Trento), Stéphane Lathuilière (Telecom-Paris), Aliaksandr Siarohin (Snapchat), Christian Theobalt (MPI Informatik), Sergey Tulyakov (Snap Inc), Vladislav Golyanik (MPI for Informatics), and Elisa Ricci (University of Trento)

Event-Based Video Reconstruction via Potential-Assisted Spiking Neural Network 3584
Lin Zhu (Peking University), Xiao Wang (Peng Cheng Laboratory), Yi Chang (PengCheng Laboratory), Jianing Li (Peking University), Tiejun Huang (Peking University), and Yonghong Tian (Peking University)

Modular Action Concept Grounding in Semantic Video Prediction 3595
Wei Yu (University of Toronto), Wenxin Chen (University of Toronto), Songheng Yin (University of Toronto), Steve Easterbrook (University of Toronto), and Animesh Garg (University of Toronto, Vector Institute, Nvidia)

Show Me What and Tell Me How: Video Synthesis via Multimodal Conditioning 3605
Ligong Han (Rutgers University), Jian Ren (Snap Inc.), Hsin-Ying Lee (Snap Inc), Francesco Barbieri (Snap Inc.), Kyle Olszewski (Snap Inc.), Shervin Minaee (Snap), Dimitris Metaxas (Rutgers), and Sergey Tulyakov (Snap Inc)

StyleGAN-V: A Continuous Video Generator With the Price, Image Quality and Perks of StyleGAN2 3616
Ivan Skorokhodov (KAUST), Sergey Tulyakov (Snap Inc), and Mohamed Elhoseiny (KAUST)

Structure-Aware Motion Transfer With Deformable Anchor Model 3627
Jiale Tao (University of Electronic Science and Technology of China), Biao Wang (Alibaba Group), Borun Xu (University of Electronic Science and Technology of China), Tiezheng Ge (Alibaba Group), Yuning Jiang (Alibaba Group), Wen Li (University of Electronic Science and Technology of China), and Lixin Duan (University of Electronic Science and Technology of China)

Image Animation With Perturbed Masks 3637
Yoav Shalev (Tel Aviv University) and Lior Wolf (Tel Aviv University, Israel)

Thin-Plate Spline Motion Model for Image Animation 3647
Jian Zhao (Tsinghua University) and Hui Zhang (Tsinghua University)

Controllable Animation of Fluid Elements in Still Images 3657
Aniruddha Mahapatra (Adobe) and Kuldeep Kulkarni (Adobe Research)

Watch It Move: Unsupervised Discovery of 3D Joints for Re-Posing of Articulated Objects 3667
Atsuhiro Noguchi (The University of Tokyo), Umar Iqbal (NVIDIA Research), Jonathan Tremblay (NVIDIA), Tatsuya Harada (The University of Tokyo / RIKEN), and Orazio Gallo (NVIDIA Research)

Geometric Structure Preserving Warp for Natural Image Stitching 3678
Peng Du (Northwest A&F University), Jifeng Ning (Northwest A&F University), Jiguang Cui (Northwest A&F University), Shaoli Huang (Tencent AI-Lab), Xinchao Wang (National University of Singapore), and Jiaxin Wang (Northwest A&F University)

Few-Shot Incremental Learning for Label-to-Image Translation 3687
Pei Chen (Zhejiang University), Yangkang Zhang (Zhejiang University), Zejian Li (Zhejiang University), and Lingyun Sun (Zhejiang University)

Exemplar-Based Pattern Synthesis With Implicit Periodic Field Network 3698
Haiwei Chen (University of Southern California), Jiayi Liu (University of Southern California), Weikai Chen (Tencent America), Shichen Liu (University of Southern California), and Yajie Zhao (Institution for Creative Technologies, University of Southern California)

SIMBAR: Single Image-Based Scene Relighting for Effective Data Augmentation for Automated Driving Vision Tasks 3708
Xianling Zhang (Ford Motor Company), Nathan Tseng (University of Michigan), Ameerah Syed (Ford Motor Company), Rohan Bhasin (Ford Motor Company), and Nikita Jaipuria (Ford Motor Company)

SoftCollage: A Differentiable Probabilistic Tree Generator for Image Collage 3719
Jiahao Yu (Tsinghua University), Li Chen (Tsinghua University), Mingrui Zhang (Tsinghua University), and Mading Li (Kuaishou Technology)

PILC: Practical Image Lossless Compression With an End-to-End GPU Oriented Neural Framework 3729
Ning Kang (Huawei Noah's Ark Lab), Shanzhao Qiu (Tsinghua University), Shifeng Zhang (Huawei Noah's Ark Lab), Zhenguo Li (Huawei Noah's Ark Lab), and Shu-Tao Xia (Tsinghua University)

Kubric: A Scalable Dataset Generator 3739
Klaus Greff (Google), Francois Belletti (Google), Lucas Beyer (Google Brain), Carl Doersch (DeepMind), Yilun Du (MIT), Daniel Duckworth (Google), David J. Fleet (University of Toronto), Dan Gnanapragasam (Google Inc), Florian Golemo (Mila, ElementAI), Charles Herrmann (Google), Thomas Kipf (Google Brain), Abhijit Kundu (Google), Dmitry Lagun (Google), Issam Laradji (ServiceNow), Hsueh-Ti Derek Liu (University of Toronto), Henning Meyer (Google), Yishu Miao (Haiper Ltd), Derek Nowrouzezahrai (McGill University), Cengiz Oztireli (University of Cambridge, Google), Etienne Pot (Google), Noha Radwan (Google), Daniel Rebain (Google Inc.), Sara Sabour (Google), Mehdi S. M. Sajjadi (Google Brain), Matan Sela (Google), Vincent Sitzmann (Massachusetts Institute of Technology), Austin Stone (Google), Deqing Sun (Google), Suhani Vora (Google), Ziyu Wang (Haiper Ltd), Tianhao Wu (University of Cambridge), Kwang Moo Yi (University of British Columbia), Fangcheng Zhong (University of Cambridge), and Andrea Tagliasacchi (Google Brain and University of Toronto)

Poster 1.2: 3D From Single Images

360MonoDepth: High-Resolution 360° Monocular Depth Estimation 3752
Manuel Rey-Area (University of Bath), Mingze Yuan (University of Bath), and Christian Richardt (University of Bath)

Pre-Train, Self-Train, Distill: A Simple Recipe for Supersizing 3D Reconstruction 3763
Kalyan Vasudev Alwala (Facebook), Abhinav Gupta (CMU/FAIR), and Shubham Tulsiani (Carnegie Mellon University)

DGECN: A Depth-Guided Edge Convolutional Network for End-to-End 6D Pose Estimation 3773
Tuo Cao (Wuhan University), Fei Luo (Wuhan University), Yanping Fu (Anhui University), Wenxiao Zhang (Wuhan University), Shengjie Zheng (Wuhan university), and Chunxia Xiao (Wuhan University)

MonoGround: Detecting Monocular 3D Objects From the Ground 3783
Zequn Qin (Zhejiang University) and Xi Li (Zhejiang University)

3D Shape Reconstruction From 2D Images With Disentangled Attribute Flow 3793
Xin Wen (Tsinghua University and JD.com), Junsheng Zhou (Tsinghua University), Yu-Shen Liu (Tsinghua University), Hua Su (Kuaishou Technology), Zhen Dong (Wuhan University), and Zhizhong Han (Wayne State University)

Toward Practical Monocular Indoor Depth Estimation 3804
Cho-Ying Wu (University of Southern California), Jialiang Wang (Facebook Inc.), Michael Hall (Facebook), Ulrich Neumann (USC), and Shuochen Su (Facebook Inc.)

Focal Length and Object Pose Estimation via Render and Compare 3815
Georgy Ponimatkin (Ecole des Ponts ParisTech), Yann Labbé (Inria), Bryan Russell (Adobe Research), Mathieu Aubry (École des ponts ParisTech), and Josef Sivic (Czech Technical University)

CLIP-NeRF: Text-and-Image Driven Manipulation of Neural Radiance Fields 3825
Can Wang (City University of Hong Kong), Menglei Chai (Snap Inc.), Mingming He (USC ICT), Dongdong Chen (Microsoft Cloud AI), and Jing Liao (City University of Hong Kong)

Registering Explicit to Implicit: Towards High-Fidelity Garment Mesh Reconstruction From Single Images 3835
Heming Zhu (Shenzhen Research Institute of Big Data, The Chinese University of Hong Kong (Shenzhen)), Lingteng Qiu (Shenzhen Research Institute of Big Data, the Chinese University of Hong Kong (Shenzhen)), Yuda Qiu (Shenzhen Research Institute of Big Data, The Chinese University of Hong Kong (Shenzhen)), and Xiaoguang Han (Shenzhen Research Institute of Big Data, the Chinese University of Hong Kong (Shenzhen))

Layered Depth Refinement With Mask Guidance 3845
Soo Ye Kim (KAIST), Jianming Zhang (Adobe Research), Simon Niklaus (Adobe Research), Yifei Fan (Adobe), Simon Chen (Adobe Research), Zhe Lin (Adobe Research), and Munchurl Kim (Korea Advanced Institute of Science and Technology)

HEAT: Holistic Edge Attention Transformer for Structured Reconstruction 3856
Jiacheng Chen (Simon Fraser University), Yiming Qian (University of Manitoba), and Yasutaka Furukawa (Simon Fraser University)

BARC: Learning To Regress 3D Dog Shape From Images by Exploiting Breed Information 3866
Nadine Rüegg (ETH Zürich, Max-Planck Institute for Intelligent Systems), Silvia Zuffi (IMATI-CNR), Konrad Schindler (ETH Zurich), and Michael J. Black (Max Planck Institute for Intelligent Systems)

Time3D: End-to-End Joint Monocular 3D Object Detection and Tracking for Autonomous Driving..... 3875
Peixuan Li (Autonomous Driving Center SAIC MOTOR) and Jieyu Jin (Autonomous Driving Center SAIC MOTOR)

What's in Your Hands? 3D Reconstruction of Generic Objects in Hands 3885
Yufei Ye (Carnegie Mellon University), Abhinav Gupta (CMU/FAIR), and Shubham Tulsiani (Carnegie Mellon University)

3D Moments From Near-Duplicate Photos 3896
Qianqian Wang (Cornell), Zhengqi Li (Google Inc.), David Salesin (Google Research), Noah Snavely (Cornell University and Google AI), Brian Curless (University of Washington), and Janne Kontkanen (Google)

Neural Window Fully-Connected CRFs for Monocular Depth Estimation 3906
Weihao Yuan (Alibaba A.I. Labs), Xiaodong Gu (Alibaba A.I. Labs), Zuozhuo Dai (Alibaba A.I. Labs), Siyu Zhu (Alibaba A.I. Labs), and Ping Tan (Simon Fraser University)

PUMP: Pyramidal and Uniqueness Matching Priors for Unsupervised Learning of Local Descriptors 3916
Jérome Revaud (Naver Labs Europe), Vincent Leroy (Naver Labs Europe), Philippe Weinzaepfel (NAVER LABS Europe), and Boris Chidlovskii (Naver Labs Europe)

CroMo: Cross-Modal Learning for Monocular Depth Estimation 3927
Yannick Verdié (Huawei Noah's Ark Lab), Jifei Song (Huawei Noah's Ark Lab), Barnabé Mas (Ecole Polytechnique), Benjamin Busam (Technical University of Munich), Ales Ales Leonardis (Huawei Noah's Ark Lab), and Steven McDonagh (Huawei Noah's Ark Lab)

f-SfT: Shape-From-Template With a Physics-Based Deformation Model 3938
Navami Kairanda (Max Planck Institute for Informatics), Edith Tretschk (Max-Planck-Institut für Informatik), Mohamed Elgharib (Max Planck Institute for Informatics), Christian Theobalt (MPI Informatik), and Vladislav Golyanik (MPI for Informatics)

Human-Aware Object Placement for Visual Environment Reconstruction 3949
Hongwei Yi (Max Planck Institute for Intelligent Systems), Chun-Hao P. Huang (Max Planck Institute for Intelligent Systems), Dimitrios Tzionas (Max Planck Institute for Intelligent Systems), Muhammed Kocabas (Max Planck Institute for Intelligent Systems / ETH Zurich), Mohamed Hassan (Max Planck Institute for Intelligent Systems), Siyu Tang (ETH Zurich), Justus Thies (Max Planck Institute for Intelligent Systems), and Michael J. Black (Max Planck Institute for Intelligent Systems)

AutoRF: Learning 3D Object Radiance Fields From Single View Observations 3961
Norman Müller (Technical University of Munich), Andrea Simonelli (University of Trento), Lorenzo Porzi (Facebook), Samuel Rota Bulò (Facebook), Matthias Nießner (Technical University of Munich), and Peter Kontschieder (Facebook)

Pix2NeRF: Unsupervised Conditional p-GAN for Single Image to Neural Radiance Fields Translation 3971
Shengqu Cai (ETH Zurich), Anton Obukhov (ETH Zurich), Dengxin Dai (ETH Zurich), and Luc Van Gool (ETH Zurich)

MonoScene: Monocular 3D Semantic Scene Completion 3981
Anh-Quan Cao (Inria) and Raoul de Charette (Inria)

GenDR: A Generalized Differentiable Renderer 3992
Felix Petersen (University of Konstanz), Bastian Goldluecke (University of Konstanz), Christian Borgelt (University of Salzburg), and Oliver Deussen (University of Konstanz)

MonoDTR: Monocular 3D Object Detection With Depth-Aware Transformer 4002
Kuan-Chih Huang (National Taiwan University), Tsung-Han Wu (National Taiwan University), Hung-Ting Su (National Taiwan University), and Winston H. Hsu (National Taiwan University)

ROCA: Robust CAD Model Retrieval and Alignment From a Single Image 4012
Can Gümeli (Technical University of Munich), Angela Dai (Technical University of Munich), and Matthias Nießner (Technical University of Munich)

Poster 1.2: Face & Gestures

HP-Capsule: Unsupervised Face Part Discovery by Hierarchical Parsing Capsule Network 4022
Chang Yu (Institute of Automation, Chinese Academy of Sciences), Xiangyu Zhu (Chinese Academy of Sciences), Xiaomei Zhang (National Lab of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences), Zidu Wang (Institute of Automation, Chinese Academy of Sciences), Zhaoxiang Zhang (Chinese Academy of Sciences, China), and Zhen Lei (NLPR, CASIA, China)

Killing Two Birds With One Stone: Efficient and Robust Training of Face Recognition CNNs by Partial FC 4032
Xiang An (DeepGlint), Jiankang Deng (Imperial College London), Jia Guo (InsightFace.ai), Ziyong Feng (DeepGlint), XuHan Zhu (University of Chinese Academy of Sciences), Jing Yang (University of Nottingham), and Tongliang Liu (The University of Sydney)

Sparse Local Patch Transformer for Robust Face Alignment and Landmarks Inherent Relation Learning 4042
Jiahao Xia (School of Electrical and Data Engineering, FEIT, University of Technology Sydney), Weiwei Qu (Southern University of Science and Technology), Wenjian Huang (Southern University of Science and Technology), Jianguo Zhang (Southern University of Science and Technology), Xi Wang (CalmCar), and Min Xu (Univeristy of Technology Sydney)

Enhancing Face Recognition With Self-Supervised 3D Reconstruction 4052
Mingjie He (Institute of Computing Technology, Chinese Academy of Sciences; University of Chinese Academy of Sciences), Jie Zhang (ICT, CAS), Shiguang Shan (Institute of Computing Technology, Chinese Academy of Sciences), and Xilin Chen (Institute of Computing Technology, Chinese Academy of Sciences)

Learning To Learn Across Diverse Data Biases in Deep Face Recognition 4062
Chang Liu (Northeastern University), Xiang Yu (NEC Labs), Yi-Hsuan Tsai (Phiar Technologies), Masoud Faraki (NEC Labs America), Ramin Moslemi (NEC Laboratories America, Inc.), Manmohan Chandraker (UC San Diego), and Yun Fu (Northeastern University)

An Efficient Training Approach for Very Large Scale Face Recognition 4073
Kai Wang (National University of Singapore), Shuo Wang (Alibaba Group), Panpan Zhang (National University of Singapore), Zhipeng Zhou (Alibaba Group), Zheng Zhu (Tsinghua University), Xiaobo Wang (CBSR&NLPR, Institute of Automation, Chinese Academy of Sciences), Xiaojiang Peng (Shenzhen Technology University), Baigui Sun (Alibaba Group), Hao Li (Alibaba Group), and Yang You (National University of Singapore)

MogFace: Towards a Deeper Appreciation on Face Detection 4083
Yang Liu (Alibaba Group), Fei Wang (alibaba group), Jiankang Deng (Imperial College London), Zhipeng Zhou (Alibaba Group), Baigui Sun (Alibaba Group), and Hao Li (Alibaba Group)

Exploring Frequency Adversarial Attacks for Face Forgery Detection 4093
Shuai Jia (Shanghai Jiao Tong University), Chao Ma (Shanghai Jiao Tong University), Taiping Yao (Tencent YouTu), Bangjie Yin (Tencent), Shouhong Ding (Tencent), and Xiaokang Yang (Shanghai Jiao Tong University of China)

End-to-End Reconstruction-Classification Learning for Face Forgery Detection 4103
Junyi Cao (Shanghai Jiao Tong University), Chao Ma (Shanghai Jiao Tong University), Taiping Yao (Tencent YouTu), Shen Chen (Tencent YouTu Lab), Shouhong Ding (Tencent), and Xiaokang Yang (Shanghai Jiao Tong University of China)

Domain Generalization via Shuffled Style Assembly for Face Anti-Spoofing 4113
Zhuo Wang (Beijing University of Posts and Telecommunications), Zezheng Wang (Kuaishou), Zitong Yu (CMVS, University of Oulu), Weihong Deng (Beijing University of Posts and Telecommunications), Jiahong Li (Kuaishou), Tingting Gao (kuaishou), and Zhongyuan Wang (Kuaishou)

Privacy-Preserving Online AutoML for Domain-Specific Face Detection 4124
Chenqian Yan (Xiamen University), Yuge Zhang (Microsoft), Quanlu Zhang (Microsoft), Yaming Yang (Microsoft), Xinyang Jiang (Microsoft Research Asia), Yuqing Yang (Microsoft), and Baoyuan Wang (Xiaobing.AI)

Simulated Adversarial Testing of Face Recognition Models 4135
Nataniel Ruiz (Boston University), Adam Kortylewski (Max Planck Institute for Informatics), Weichao Qiu (Johns Hopkins University), Cihang Xie (University of California, Santa Cruz), Sarah Adel Bargal (Boston University), Alan Yuille (Johns Hopkins University), and Stan Sclaroff (Boston University)

Decoupled Multi-Task Learning With Cyclical Self-Regulation for Face Parsing 4146
Qingping Zheng (Northwestern Polytechnical University), Jiankang Deng (Imperial College London), Zheng Zhu (Tsinghua University), Ying Li (Northwestern Polytechnical University), and Stefanos Zafeiriou (Imperial College London)

Towards Semi-Supervised Deep Facial Expression Recognition With an Adaptive Confidence Margin 4156
Hangyu Li (Xidian University), Nannan Wang (Xidian University), Xi Yang (Xidian University), Xiaoyu Wang (The Chinese University of Hong Kong (Shenzhen), ,), and Xinbo Gao (Chongqing University of Posts and Telecommunications)

Towards Accurate Facial Landmark Detection via Cascaded Transformers 4166
Hui Li (Samsung Research and Development Institute China Xi'an), Zidong Guo (Samsung Research and Development Institute China Xi'an), Seon-Min Rhee (Samsung Electronics), Seungju Han (Samsung Advanced Institute of Technology), and Jae-Joon Han (Samsung)

PhysFormer: Facial Video-Based Physiological Measurement With Temporal Difference Transformer 4176
Zitong Yu (CMVS, University of Oulu), Yuming Shen (University of Oxford), Jingang Shi (Xi'an Jiaotong University), Hengshuang Zhao (University of Oxford), Philip H.S. Torr (University of Oxford), and Guoying Zhao (University of Oulu)

GazeOnce: Real-Time Multi-Person Gaze Estimation 4187
Mingfang Zhang (Pengcheng Laboratory), Yunfei Liu (Beihang University), and Feng Lu (Beihang University)

Generalizing Gaze Estimation With Rotation Consistency 4197
Yiwei Bao (Beihang University), Yunfei Liu (Beihang University), Haofei Wang (Peng Cheng Laboratory), and Feng Lu (Beihang University)

Face Relighting With Geometrically Consistent Shadows 4207
Andrew Hou (Michigan State University), Michel Sarkis (Qualcomm Technologies Inc.), Ning Bi (Qualcomm), Yiying Tong (Michigan State University), and Xiaoming Liu (Michigan State University)

HairMapper: Removing Hair From Portraits Using GANs 4217
Yiqian Wu (Zhejiang University), Yong-Liang Yang (University of Bath), and Xiaogang Jin (Zhejiang University)

Learning To Restore 3D Face From In-the-Wild Degraded Images 4227
Zhenyu Zhang (Tencent), Yanhao Ge (Tencent YouTu), Ying Tai (Tencent YouTu), Xiaoming Huang (Tencent Youtu), Chengjie Wang (Tencent), Hao Tang (ETH Zurich), Dongjin Huang (Shanghai University), and Zhifeng Xie (Shanghai University)

Poster 1.2: Segmentation, Grouping and Shape Analysis

Semi-Supervised Semantic Segmentation Using Unreliable Pseudo-Labels 4238
Yuchao Wang (Shanghai Jiao tong University), Haochen Wang (Shanghai Jiao Tong University), Yujun Shen (Dept. of IE, CUHK), Jingjing Fei (SenseTime Research), Wei Li (SenseTime Research), Guoqiang Jin (SenseTime Research), Liwei Wu (SenseTime Research), Rui Zhao (SenseTime Group Limited), and Xinyi Le (Shanghai Jiao Tong University)

Perturbed and Strict Mean Teachers for Semi-Supervised Semantic Segmentation 4248
Yuyuan Liu (University of Adelaide), Yu Tian (Australian Institute for Machine Learning, University of Adelaide), Yuanhong Chen (Australian Institute for Machine Learning, University of Adelaide), Fengbei Liu (University of Adelaide), Vasileios Belagiannis (Otto von Guericke University Magdeburg), and Gustavo Carneiro (University of Adelaide)

ST++: Make Self-Training Work Better for Semi-Supervised Semantic Segmentation 4258
Lihe Yang (Nanjing University), Wei Zhuo (Tencent), Lei Qi (Southeast University), Yinghuan Shi (Nanjing University), and Yang Gao (Nanjing University)

Beyond Semantic to Instance Segmentation: Weakly-Supervised Instance Segmentation via Semantic Knowledge Transfer and Self-Refinement 4268
Beomyoung Kim (NAVER), YoungJoon Yoo (Clova AI Research, NAVER Corp.), Chae Eun Rhee (Inha University), and Junmo Kim (KAIST)

Self-Supervised Image-Specific Prototype Exploration for Weakly Supervised Semantic Segmentation 4278
Qi 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)

Regional Semantic Contrast and Aggregation for Weakly Supervised Semantic Segmentation 4289
Tianfei Zhou (ETH Zurich), Meijie Zhang (Beijing Institute of Technology), Fang Zhao (Tencent AI Lab), and Jianwu Li (Beijing Institute of Technology)

Multi-Class Token Transformer for Weakly Supervised Semantic Segmentation 4300
Lian Xu (The University of Western Australia), Wanli Ouyang (The University of Sydney), Mohammed Bennamoun (University of Western Australia), Farid Boussaid (University of Western Australia), and Dan Xu (The Hong Kong University of Science and Technology)

Weakly Supervised Semantic Segmentation by Pixel-to-Prototype Contrast 4310
Ye Du (Beihang University), Zehua Fu (Hangzhou Innovation Institute, 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)

Threshold Matters in WSSS: Manipulating the Activation for the Robust and Accurate Segmentation Model Against Thresholds 4320
Minhyun Lee (Yonsei University), Dongseob Kim (Yonsei University), and Hyunjung Shim (Yonsei University)

Novel Class Discovery in Semantic Segmentation 4330
Yuyang Zhao (National University of Singapore), Zhun Zhong (University of Trento), Nicu Sebe (University of Trento), and Gim Hee Lee (National University of Singapore)

Pin the Memory: Learning To Generalize Semantic Segmentation 4340
Jin Kim (Yonsei University), Jiyoung Lee (NAVER AI LAB), Jungin Park (Yonsei University), Dongbo Min (Ewha Womans University), and Kwanghoon Sohn (Yonsei Univ.)

ISDNet: Integrating Shallow and Deep Networks for Efficient Ultra-High Resolution Segmentation 4351
Shaohua Guo (Shanghai Jiao Tong University), Liang Liu (Tencent), Zhenye Gan (Tencent), Yabiao Wang (Tencent), Wuhao Zhang (Tencent), Chengjie Wang (Tencent), Guannan Jiang (CATL), Wei Zhang (CATL), Ran Yi (Shanghai Jiao Tong University), Lizhuang Ma (Shanghai Jiao Tong University), and Ke Xu (City University of Hong Kong)

Incremental Learning in Semantic Segmentation From Image Labels 4361
Fabio Cermelli (Politecnico di Torino), Dario Fontanel (Politecnico di Torino), Antonio Tavera (Politecnico di Torino), Marco Ciccone (Politecnico di Torino), and Barbara Caputo (Politecnico di Torino)

Instance Segmentation With Mask-Supervised Polygonal Boundary Transformers 4372
Justin Lazarow (UC San Diego), Weijian Xu (University of California, San Diego), and Zhuowen Tu (UC San Diego)

SharpContour: A Contour-Based Boundary Refinement Approach for Efficient and Accurate Instance Segmentation 4382
Chenming Zhu (CUHKSZ), Xuanye Zhang (Shenzhen Research Institute of Big Data, the Chinese University of Hong Kong (Shenzhen)), Yanran Li (Bournemouth University), Liangdong Qiu (Shenzhen Research Institute of Big Data, the Chinese University of Hong Kong (Shenzhen)), Kai Han (The University of Hong Kong), and Xiaoguang Han (Shenzhen Research Institute of Big Data, the Chinese University of Hong Kong (Shenzhen))

Sparse Object-Level Supervision for Instance Segmentation With Pixel Embeddings 4392
Adrian Wolny (European Molecular Biology Laboratory), Qin Yu (European Molecular Biology Laboratory), Constantin Pape (University of Heidelberg), and Anna Kreshuk (European Molecular Biology Laboratory)

Mask Transfiner for High-Quality Instance Segmentation 4402
Lei Ke (HKUST), Martin Danelljan (ETH Zurich), Xia Li (ETH Zurich), Yu-Wing Tai (Kuaishou Technology / HKUST), Chi-Keung Tang (Hong Kong University of Science and Technology), and Fisher Yu (ETH Zurich)

Open-World Instance Segmentation: Exploiting Pseudo Ground Truth From Learned Pairwise Affinity 4412
Weiyao Wang (Facebook Research), Matt Feiszli (Facebook Research), Heng Wang (Facebook AI), Jitendra Malik (University of California at Berkeley), and Du Tran (Facebook AI)

Sparse Instance Activation for Real-Time Instance Segmentation 4423
Tianheng Cheng (Huazhong University of Science and Technology), Xinggang Wang (Huazhong University of Science and Technology), Shaoyu Chen (Huazhong University of Science and Technology), Wenqiang Zhang (Huazhong University of Science and Technology), Qian Zhang (Horizon Robotics), Chang Huang (Horizon Robotics), Zhaoxiang Zhang (Chinese Academy of Sciences, China), and Wenyu Liu (Huazhong University of Science and Technology)

E2EC: An End-to-End Contour-Based Method for High-Quality High-Speed Instance Segmentation.... 4433
Tao Zhang (Wuhan University), Shiqing Wei (Wuhan University), and Shunping Ji (Wuhan University)

Hyperbolic Image Segmentation 4443
Mina Ghadimi Atigh (University of Amsterdam), Julian Schoep (Promaton), Erman Acar (Vrije Universiteit Amsterdam), Nanne van Noord (University of Amsterdam), and Pascal Mettes (University of Amsterdam)

SeeThroughNet: Resurrection of Auxiliary Loss by Preserving Class Probability Information 4453
Dasol Han (Samsung Electronics), Jaewook Yoo (Samsung), and Dokwan Oh (Samsung Electronics)

CDGNet: Class Distribution Guided Network for Human Parsing 4463
Kunliang Liu (Ajou university; Tiangong university), Ouk Choi (Incheon National University), Jianming Wang (Tiangong University), and Wonjun Hwang (Ajou University)

CLIMS: Cross Language Image Matching for Weakly Supervised Semantic Segmentation 4473
Jinheng Xie (Shenzhen University), Xianxu Hou (Shenzhen University), Kai Ye (Shenzhen University), and Linlin Shen (Shenzhen University)

Sparse Non-Local CRF 4483
Olga Veksler (University of Waterloo) and Yuri Boykov (University of Waterloo)

Detecting Camouflaged Object in Frequency Domain 4494
Yijie Zhong (Nanjing University), Bo Li (Nanjing University), Lv Tang (University of Chinese Academy of Sciences), Senyun Kuang (Southwest Jiaotong University), Shuang Wu (Tencent), and Shouhong Ding (Tencent)

Progressive Minimal Path Method With Embedded CNN 4504
Wei Liao (Independent Researcher)

Poster 1.2: Document Analysis & Understanding

Open-Set Text Recognition via Character-Context Decoupling 4513
Chang Liu (University of Science and Technology Beijing), Chun Yang (University of Science and Technology Beijing), and Xu-Cheng Yin (University of Science and Technology Beijing)

Neural Collaborative Graph Machines for Table Structure Recognition 4523
Hao Liu (Tencent), Xin Li (Tencent), Bing Liu (Tencent), Deqiang Jiang (Tencent), Yinsong Liu (Tencent), and Bo Ren (Tencent)

Revisiting Document Image Dewarping by Grid Regularization 4533
Xiangwei Jiang (Wuhan University), Rujiao Long (Alibaba Group), Nan Xue (Wuhan University), Zhibo Yang (Alibaba Group), Cong Yao (Alibaba DAMO Academy), and Gui-Song Xia (Wuhan University)

Syntax-Aware Network for Handwritten Mathematical Expression Recognition 4543
Ye Yuan (Tomorrow Advancing Life), Xiao Liu (Tencent), Wondimu Dikubab (Huazhong University of Science and Technology), Hui Liu (TAL), Zhilong Ji (Tomorrow Advancing Life), Zhongqin Wu (Tomorrow Advancing Life), and Xiang Bai (Huazhong University of Science and Technology)

Few Could Be Better Than All: Feature Sampling and Grouping for Scene Text Detection 4553
Jingqun Tang (Netease), Wenqing Zhang (Huazhong University of Science and Technology), Hongye Liu (China JiLiang University), MingKun Yang (Huazhong University of Science and Technology), Bo Jiang (Netease Inc.), Guanglong Hu (Netease), and Xiang Bai (Huazhong University of Science and Technology)

Fourier Document Restoration for Robust Document Dewarping and Recognition 4563
Chuhui Xue (Nanyang Technological University), Zichen Tian (Nanyang Technological University), Fangneng Zhan (Nanyang Technological University), Shijian Lu (Nanyang Technological University), and Song Bai (University of Oxford)

XYLayoutLM: Towards Layout-Aware Multimodal Networks for Visually-Rich Document Understanding 4573
Zhangxuan Gu (Shanghai Jiao Tong University), Changhua Meng (Ant Group), Ke Wang (Ant Group Co., Ltd.), Jun Lan (antgroup), Weiqiang Wang (Ant Group), Ming Gu (Ant Group), and Liqing Zhang (Shanghai Jiao Tong University)

SwinTextSpotter: Scene Text Spotting via Better Synergy Between Text Detection and Text Recognition 4583
Mingxin Huang (South China University of Technology), Yuliang Liu (South China University of Technology), Zhenghao Peng (Chinese University of Hong Kong), Chongyu Liu (South China University of Tenology), Dahua Lin (The Chinese University of Hong Kong), Shenggao Zhu (Huawei), Nicholas Yuan (Huawei Cloud), Kai Ding (IntSig Information Co., Ltd), and Lianwen Jin (South China University of Technology)

Towards Weakly-Supervised Text Spotting Using a Multi-Task Transformer 4594
Yair Kittenplon (Tel-Aviv University), Inbal Lavi (Amazon), Sharon Fogel (Tel Aviv University), Yarin Bar (Amazon), R. Manmatha (Amazon), and Pietro Perona (Amazon Web Services (AWS), ,)

TableFormer: Table Structure Understanding With Transformers 4604
Ahmed Nassar (IBM Research), Nikolaos Livathinos (IBM Research), Maksym Lysak (IBM Research), and Peter Staar (IBM Research)

Knowledge Mining With Scene Text for Fine-Grained Recognition 4614
Hao Wang (Huazhong University of Science and Technology), Junchao Liao (Huazhong University of Science and Technology), Tianheng Cheng (Huazhong University of Science and Technology), Zewen Gao (Huazhong University of Science and Technology), Hao Liu (Tencent), Bo Ren (Tencent), Xiang Bai (Huazhong University of Science and Technology), and Wenyu Liu (Huazhong University of Science and Technology)

PubTables-1M: Towards Comprehensive Table Extraction From Unstructured Documents 4624
Brandon Smock (Microsoft), Rohith Pesala (Microsoft Research), and Robin Abraham (Microsoft)

Poster 1.2: Recognition: Detection, Categorization, Retrieval

Focal and Global Knowledge Distillation for Detectors 4633
Zhendong Yang (Graduate school at ShenZhen, Tsinghua university), Zhe Li (Bytedance Inc.), Xiaohu Jiang (Graduate School at Shenzhen, Tsinghua University), Yuan Gong (Graduate School at Shenzhen, Tsinghua University), Zehuan Yuan (Bytedance.Inc), Danpei Zhao (Beihang University), and Chun Yuan (Graduate school at ShenZhenTsinghua university)

Speed Up Object Detection on Gigapixel-Level Images With Patch Arrangement 4643
Jiahao Fan (Shanghai Jiao Tong University), Huabin Liu (Shanghai Jiao Tong University), Wenjie Yang (ShangHaiJiaoTong University), John See (Heriot-Watt University Malaysia), Aixin Zhang (Shanghai Jiao Tong University), and Weiyao Lin (Shanghai Jiao Tong university)

Training Object Detectors From Scratch: An Empirical Study in the Era of Vision Transformer 4652
Weixiang Hong (National University of Singapore), Jiangwei Lao (Ant Group), Wang Ren (Ant Group), Jian Wang (Ant Group), Jingdong Chen (Ant Group), and Wei Chu (Ant Group)

Learning With Neighbor Consistency for Noisy Labels 4662
Ahmet Iscen (Google), Jack Valmadre (University of Adelaide), Anurag Arnab (Google), and Cordelia Schmid (Google)

Meta Convolutional Neural Networks for Single Domain Generalization 4672
Chaoqun Wan (Alibaba Group), Xu Shen (Alibaba Group), Yonggang Zhang (University of Science and Technology of China), Zhiheng Yin (University of Michigan), Xinmei Tian (USTC), Feng Gao (Zhejiang Lab), Jianqiang Huang (Alibaba Group), and Xian-Sheng Hua (Alibaba Group)

Dual Cross-Attention Learning for Fine-Grained Visual Categorization and Object Re-Identification 4682
Haowei Zhu (Xilinx), Wenjing Ke (Xilinx), Dong Li (Xilinx), Ji Liu (Xilinx Beijing), Lu Tian (Xilinx, Inc.), and Yi Shan (Xilinx)

Geometry-Aware Guided Loss for Deep Crack Recognition 4693
Zhuangzhuang Chen (shenzhen university), Jin Zhang (Shenzhen University), Zhuonan Lai (Shenzhen University), Jie Chen (Shenzhen University), Zun Liu (Shenzhen University), and Jianqiang Li (Shenzhen University)

Segment, Magnify and Reiterate: Detecting Camouflaged Objects the Hard Way 4703
Qi Jia (Dalian University of Technology), Shuilian Yao (Dalian University Of Technology), Yu Liu (Dalian University of Technology), Xin Fan (Dalian University of Technology), Risheng Liu (Dalian University of Technology), and Zhongxuan Luo (Dalian University of Technology)

Dynamic Sparse R-CNN 4713
Qinghang Hong (Xilinx), Fengming Liu (Xilinx), Dong Li (Xilinx), Ji Liu (Xilinx Beijing), Lu Tian (Xilinx, Inc.), and Yi Shan (Xilinx)

Deep Hybrid Models for Out-of-Distribution Detection 4723
Senqi Cao (Binghamton University) and Zhongfei Zhang (Binghamton University)

AutoLoss-GMS: Searching Generalized Margin-Based Softmax Loss Function for Person Re-Identification 4734
Hongyang Gu (Tsinghua University; Xi'an Institute of High-Tech), Jianmin Li (Tsinghua University), Guangyuan Fu (Xi'an Institute of High-Tech), Chifong Wong (Tsinghua University), Xinghao Chen (Huawei Noah's Ark Lab), and Jun Zhu (Tsinghua University)

Feature Erasing and Diffusion Network for Occluded Person Re-Identification 4744
Zhikang Wang (Monash University), Feng Zhu (University of Science and Technology of China), Shixiang Tang (The University of Sydney), Rui Zhao (SenseTime Group Limited), Lihuo He (Xidian University), and Jiangning Song (Monash University)

Multi-Label Classification With Partial Annotations Using Class-Aware Selective Loss 4754
Emanuel Ben-Baruch (Alibaba), Tal Ridnik (Alibaba), Itamar Friedman (Alibaba), Avi Ben-Cohen (Alibaba group), Nadav Zamir (Alibaba), Asaf Noy (Alibaba), and Lihi Zelnik-Manor (Technion)

BoxeR: Box-Attention for 2D and 3D Transformers 4763
Duy-Kien Nguyen (University of Amsterdam), Jihong Ju (TomTom Global Content B.V.), Olaf Booij (TomTom), Martin R. Oswald (ETH Zurich), and Cees G. M. Snoek (University of Amsterdam)

Multi-Label Iterated Learning for Image Classification With Label Ambiguity 4773
Sai Rajeswar (University of Montreal), Pau Rodríguez (ServiceNow), Soumye Singhal (Mila, University of Montreal), David Vazquez (Element AI), and Aaron Courville (MILA, Université de Montréal)

Vision Transformer With Deformable Attention 4784
Zhuofan Xia (Tsinghua University), Xuran Pan (Tsinghua University), Shiji Song (Department of Automation, Tsinghua University), Li Erran Li (Amazon / Columbia University), and Gao Huang (Tsinghua)

MViTv2: Improved Multiscale Vision Transformers for Classification and Detection 4794
Yanghao Li (Facebook AI Research), Chao-Yuan Wu (Facebook), Haoqi Fan (Facebook AI Research), Karttikeya Mangalam (UC Berkeley), Bo Xiong (Facebook AI Research), Jitendra Malik (University of California at Berkeley), and Christoph Feichtenhofer (Facebook AI Research)

Dense Learning Based Semi-Supervised Object Detection 4805
Binghui Chen (Alibaba), Pengyu Li (Alibaba Group), Xiang Chen (Alibaba), Biao Wang (Alibaba), Lei Zhang (Hong Kong Polytechnic University, Hong Kong, China), and Xian-Sheng Hua (Alibaba Group)

R(Det)2: Randomized Decision Routing for Object Detection 4815
Yali Li (Tsinghua University) and Shengjin Wang (Tsinghua University)

GlideNet: Global, Local and Intrinsic Based Dense Embedding NETwork for Multi-Category Attributes Prediction 4825
Kareem Metwaly (The Pennsylvania State University), Aerin Kim (Scale AI), Elliot Branson (Scale AI), and Vishal Monga (The Pennsylvania State University)

Self-Supervised Equivariant Learning for Oriented Keypoint Detection 4837
Jongmin Lee (POSTECH), Byungjin Kim (Postech), and Minsu Cho (POSTECH)

Label Relation Graphs Enhanced Hierarchical Residual Network for Hierarchical Multi-Granularity Classification 4848
Jingzhou Chen (Zhejiang University), Peng Wang (Zhejiang University), Jian Liu (Ant Group), and Yuntao Qian (Zhejiang University, China)

Object Localization Under Single Coarse Point Supervision 4858
Xuehui Yu (University of Chinese Academy of Sciences), Pengfei Chen (University of Chinese Academy of Sciences), Di Wu (University of Chinese Academy of Sciences), Najmul Hassan (University of Oregon), Guorong Li (University of Chinese Academy of Sciences), Junchi Yan (Shanghai Jiao Tong University), Humphrey Shi (U of Oregon/UIUC), Qixiang Ye (University of Chinese Academy of Sciences, China), and Zhenjun Han (University of Chinese Academy of Sciences)

Rethinking Visual Geo-Localization for Large-Scale Applications 4868
Gabriele Berton (Politecnico di Torino), Carlo Masone (Consorzio Interuniversitario Nazionale per l'Informatica), and Barbara Caputo (Politecnico di Torino)

Whose Hands Are These? Hand Detection and Hand-Body Association in the Wild 4879
Supreeth Narasimhaswamy (Stony Brook University), Thanh Nguyen (VinAI Research), Mingzhen Huang (University at Buffalo), and Minh Hoai (Stony Brook University)

Cloning Outfits From Real-World Images to 3D Characters for Generalizable Person Re-Identification 4890
Yanan Wang (Inception Institute of Artificial Intelligence), Xuezhi Liang (Inception Institute of Artificial Intelligence), and Shengcai Liao (Inception Institute of Artificial Intelligence)

Towards Unsupervised Domain Generalization 4900
Xingxuan Zhang (Tsinghua University), Linjun Zhou (Tsinghua University), Renzhe Xu (Tsinghua University), Peng Cui (Tsinghua University), Zheyan Shen (Tsinghua University), and Haoxin Liu (HaoxinLiu)

ViM: Out-of-Distribution With Virtual-Logit Matching 4911
Haoqi Wang (SenseTime), Zhizhong Li (Sensetime), Litong Feng (Sensetime Research), and Wayne Zhang (SenseTime Research)

Vision Transformer Slimming: Multi-Dimension Searching in Continuous Optimization Space 4921
Arnav Chavan (Indian Institute of Technology - Dhanbad), Zhiqiang Shen (Carnegie Mellon University), Zhuang Liu (UC Berkeley), Zechun Liu (Carnegie Mellon University), Kwang-Ting Cheng (Hong Kong University of Science and Technology), and Eric P. Xing (MBZUAI, CMU, and Petuum Inc.)

Nonuniform-to-Uniform Quantization: Towards Accurate Quantization via Generalized Straight-Through Estimation 4932
Zechun Liu (Carnegie Mellon University), Kwang-Ting Cheng (Hong Kong University of Science and Technology), Dong Huang (Carnegie Mellon University), Eric P. Xing (MBZUAI, CMU, and Petuum Inc.), and Zhiqiang Shen (Carnegie Mellon University)

Poster 1.2: Vision & Language

Align and Prompt: Video-and-Language Pre-Training With Entity Prompts 4943
Dongxu Li (Salesforce), Junnan Li (Salesforce), Hongdong Li (Australian National University, Australia), Juan Carlos Niebles (Salesforce & Stanford University), and Steven C.H. Hoi (Salesforce)

Language-Bridged Spatial-Temporal Interaction for Referring Video Object Segmentation 4954
Zihan Ding (Beihang University), Tianrui Hui (Institute of Information Engineering, Chinese Academy of Sciences), Junshi Huang (Meituan), Xiaoming Wei (Meituan), Jizhong Han (Institute of Information Engineering, Chinese Academy of Sciences), and Si Liu (Beihang University)

Language As Queries for Referring Video Object Segmentation 4964
Jiannan Wu (The University of Hong Kong), Yi Jiang (Bytedance), Peize Sun (The University of Hong Kong), Zehuan Yuan (Bytedance.Inc), and Ping Luo (The University of Hong Kong)

End-to-End Referring Video Object Segmentation With Multimodal Transformers 4975
Adam Botach (Technion), Evgenii Zheltonozhskii (Technion), and Chaim Baskin (Technion)

Multi-Level Representation Learning With Semantic Alignment for Referring Video Object Segmentation 4986
Dongming Wu (Beijing Institute of Technology), Xingping Dong (Inception Institute of Artificial Intelligence), Ling Shao (NCAI/IIAI/MBZUAI), and Jianbing Shen (Inception Institute of Artificial Intelligence)

X-Pool: Cross-Modal Language-Video Attention for Text-Video Retrieval 4996
Satya Krishna Gorti (Layer6 AI), Noël Vouitsis (Layer6 AI), Junwei Ma (Layer6 AI), Keyvan Golestan (Layer 6 AI), Maksims Volkovs (Layer6 AI), Animesh Garg (University of Toronto, Vector Institute, Nvidia), and Guangwei Yu (Layer6 AI)

Video-Text Representation Learning via Differentiable Weak Temporal Alignment 5006
Dohwan Ko (Korea University), Joonmyung Choi (Korea University), Juyeon Ko (Korea University), Shinyeong Noh (KOREA UNIVERSITY), Kyoung-Woon On (Kakao Brain), Eun-Sol Kim (Hanyang University), and Hyunwoo J. Kim (Korea University)

MAD: A Scalable Dataset for Language Grounding in Videos From Movie Audio Descriptions 5016
Mattia Soldan (KAUST), Alejandro Pardo (KAUST), Juan León Alcázar (KAUST), Fabian Caba (Adobe Research), Chen Zhao (KAUST), Silvio Giancola (KAUST), and Bernard Ghanem (KAUST)

Advancing High-Resolution Video-Language Representation With Large-Scale Video Transcriptions 5026
Hongwei Xue (Microsoft Research Asia), Tiankai Hang (Microsoft Research), Yanhong Zeng (Microsoft Research), Yuchong Sun (Microsoft), Bei Liu (Microsoft Research), Huan Yang (Microsoft Research), Jianlong Fu (Microsoft Research), and Baining Guo (MSR Asia)

Measuring Compositional Consistency for Video Question Answering 5036
Mona Gandhi (Veermata Jijabai Technological Institute), Mustafa Omer Gul (Stanford University), Eva Prakash (Stanford University), Madeleine Grunde-McLaughlin (University of Washington), Ranjay Krishna (University of Washington), and Maneesh Agrawala (Stanford University)

SimVQA: Exploring Simulated Environments for Visual Question Answering 5046
Paola Cascante-Bonilla (Rice University), Hui Wu (MIT-IBM Watson AI Lab, IBM Research), Letao Wang (University of Virginia), Rogerio S. Feris (MIT-IBM Watson AI Lab, IBM Research), and Vicente Ordonez (Rice University)

Transform-Retrieve-Generate: Natural Language-Centric Outside-Knowledge Visual Question Answering 5057
Feng Gao (UCLA), Qing Ping (Amazon Alexa AI), Govind Thattai (Amazon), Aishwarya Reganti (Amazon Alexa AI), Ying Nian Wu (University of California, Los Angeles), and Prem Natarajan (Amazon.com Inc.)

SwapMix: Diagnosing and Regularizing the Over-Reliance on Visual Context in Visual Question Answering 5068
Vipul Gupta (Pennsylvania State University), Zhuowan Li (Johns Hopkins University), Adam Kortylewski (Max Planck Institute for Informatics), Chenyu Zhang (Johns Hopkins University), Yingwei Li (Johns Hopkins University), and Alan Yuille (Johns Hopkins University)

MuKEA: Multimodal Knowledge Extraction and Accumulation for Knowledge-Based Visual Question Answering 5079
Yang Ding (Institute of Information Engineering, Chinese Academy of Sciences. School of Cyber Security, University of Chinese Academy of Sciences.), Jing Yu (Institute of Information Engineering, Chinese Academy of Sciences), Bang Liu (University of Montreal), Yue Hu (Institute of Information Engineering, Chinese Academy of Sciences), Mingxin Cui (Institute of Information Engineering, CAS), and Qi Wu (University of Adelaide)

Maintaining Reasoning Consistency in Compositional Visual Question Answering 5089
Chenchen Jing (Beijing Institute of Technology), Yunde Jia (Beijing Institute of Technology), Yuwei Wu (Beijing Institute of Technology (BIT), China), Xinyu Liu (Beijing Institute of Technology), and Qi Wu (University of Adelaide)

MLSLT: Towards Multilingual Sign Language Translation 5099
Aoxiong Yin (Zhejiang University), Zhou Zhao (Zhejiang University), Weike Jin (Zhejiang University), Meng Zhang (Huawei Noah's Ark Lab), Xingshan Zeng (Huawei Noah's Ark Lab), and Xiaofei He (Zhejiang University)

A Simple Multi-Modality Transfer Learning Baseline for Sign Language Translation 5110
Yutong Chen (Tsinghua University), Fangyun Wei (Microsoft Research Asia), Xiao Sun (Microsoft Research Asia), Zhirong Wu (Microsoft Research), and Stephen Lin (Microsoft Research)

C2SLR: Consistency-Enhanced Continuous Sign Language Recognition 5121
Ronglai Zuo (The Hong Kong University of Science and Technology) and Brian Mak (The Hong Kong University of Science and Technology)

Signing at Scale: Learning to Co-Articulate Signs for Large-Scale Photo-Realistic Sign Language Production 5131
Ben Saunders (University of Surrey), Necati Cihan Camgoz (University of Surrey), and Richard Bowden (University of Surrey)

Generating Diverse and Natural 3D Human Motions From Text 5142
Chuan Guo (University of Alberta), Shihao Zou (University of Alberta), Xinxin Zuo (University of Alberta), Sen Wang (University of Alberta), Wei Ji (University of Alberta), Xingyu Li (University of Alberta), and Li Cheng (ECE dept., University of Alberta)

Sub-Word Level Lip Reading With Visual Attention 5152
K R Prajwal (VGG, Oxford), Triantafyllos Afouras (University of Oxford), and Andrew Zisserman (University of Oxford)

Habitat-Web: Learning Embodied Object-Search Strategies From Human Demonstrations at Scale..... 5163
Ram Ramrakhya (Georgia Institute of Technology), Eric Undersander (Facebook), Dhruv Batra (Georgia Tech & Facebook AI Research), and Abhishek Das (Facebook AI Research)

ViSTA: Vision and Scene Text Aggregation for Cross-Modal Retrieval 5174
Mengjun Cheng (Peking University), Yipeng Sun (Baidu Inc.), Longchao Wang (baidu), Xiongwei Zhu (baidu), Kun Yao (Baidu Inc.), Jie Chen (Peking University), Guoli Song (Peng Cheng Laboratory), Junyu Han (Baidu Inc.), Jingtuo Liu (baidu), Errui Ding (Baidu Inc.), and Jingdong Wang (Baidu)

Cross Modal Retrieval With Querybank Normalisation 5184
Simion-Vlad Bogolin (Institute of Mathematics of the Romanian Academy), Ioana Croitoru (Institute of Mathematics of the Romanian Academy), Hailin Jin (Adobe Research), Yang Liu (Peking University), and Samuel Albanie (University of Cambridge)

Prompt Distribution Learning 5196
Yuning Lu (USTC), Jianzhuang Liu (Huawei Noah's Ark Lab), Yonggang Zhang (University of Science and Technology of China), Yajing Liu (USTC), and Xinmei Tian (USTC)

VALHALLA: Visual Hallucination for Machine Translation 5206
Yi Li (UC San Diego), Rameswar Panda (MIT-IBM Watson AI Lab), Yoon Kim (MIT), Chun-Fu Richard Chen (JPMorgan Chase Bank, N.A.), Rogerio S. Feris (MIT-IBM Watson AI Lab, IBM Research), David Cox (MIT-IBM Watson AI Lab), and Nuno Vasconcelos (UC San Diego)

VL-Adapter: Parameter-Efficient Transfer Learning for Vision-and-Language Tasks 5217
Yi-Lin Sung (UNC Chapel Hill), Jaemin Cho (UNC Chapel Hill), and Mohit Bansal (University of North Carolina at Chapel Hill)

Winoground: Probing Vision and Language Models for Visio-Linguistic Compositionality 5228
Tristan Thrush (Hugging Face), Ryan Jiang (facebook), Max Bartolo (University College London), Amanpreet Singh (Facebook), Adina Williams (Facebook AI Research), Douwe Kiela (Facebook AI Research), and Candace Ross (Massachusetts Institute of Technology)

Oral 2.1.1: Recognition: Detection, Categorization, Retrieval

MixFormer: Mixing Features Across Windows and Dimensions 5239
Qiang Chen (Baidu Inc), Qiman Wu (Baidu Inc.), Jian Wang (Baidu), Qinghao Hu (Institute of Automation, Chinese Academy of Sciences), Tao Hu (baidu), Errui Ding (Baidu Inc.), Jian Cheng (Chinese Academy of Sciences, China), and Jingdong Wang (Baidu)

Recurrent Glimpse-Based Decoder for Detection With Transformer 5250
Zhe Chen (The University of Sydney), Jing Zhang (The University of Sydney), and Dacheng Tao (JD.com)

Mobile-Former: Bridging MobileNet and Transformer 5260
Yinpeng Chen (Microsoft), Xiyang Dai (Microsoft), Dongdong Chen (Microsoft Cloud AI), Mengchen Liu (Microsoft), Xiaoyi Dong (University of Science and Technology of China), Lu Yuan (Microsoft), and Zicheng Liu (Microsoft)

Unsupervised Domain Generalization by Learning a Bridge Across Domains 5270
Sivan Harary (IBM-Research), Eli Schwartz (IBM-Research), Assaf Arbelle (IBM Research AI), Peter Staar (IBM Research), Shady Abu-Hussein (Tel Aviv University), Elad Amrani (IBM / Technion), Roei Herzig (Tel Aviv University), Amit Alfassy (IBM-Research), Raja Giryes (University of Frankfurt), Hilde Kuehne (Massachusetts Institute of Technology), Dina Katabi (Tel Aviv University), Kate Saenko (Boston University), Rogerio S. Feris (MIT-IBM Watson AI Lab, IBM Research), and Leonid Karlinsky (IBM-Research)

SIGMA: Semantic-Complete Graph Matching for Domain Adaptive Object Detection 5281
Wuyang Li (City University of Hong Kong), Xinyu Liu (City University of Hong Kong), and Yixuan Yuan (City University of Hong Kong)

Target-Relevant Knowledge Preservation for Multi-Source Domain Adaptive Object Detection 5291
Jiaxi Wu (Beihang University), Jiaxin Chen (Beihang University), Mengzhe He (Shenzhen Institutes of Advanced Technology Chinese Academy of Science), Yiru Wang (SenseTime Group Limited), Bo Li (SenseTime Group Limited), Bingqi Ma (Sensetime Group Limited), Weihao Gan (SenseTime Group Limited), Wei Wu (SenseTime Group Limited), Yali Wang (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences), and Di Huang (Beihang University, China)

PNP: Robust Learning From Noisy Labels by Probabilistic Noise Prediction 5301
Zeren Sun (Nanjing University of Science and Technology), Fumin Shen (UESTC), Dan Huang (BIT), Qiong Wang (Nanjing University of Science and Technology), Xiangbo Shu (Nanjing University of Science and Technology), Yazhou Yao (Nanjing University of Science and Technology), and Jinhui Tang (Nanjing University of Science and Technology)

Few-Shot Object Detection With Fully Cross-Transformer 5311
Guangxing Han (Columbia University), Jiawei Ma (Columbia University), Shiyuan Huang (Columbia University), Long Chen (Columbia University), and Shih-Fu Chang (Columbia University)

Task Discrepancy Maximization for Fine-Grained Few-Shot Classification 5321
SuBeen Lee (Sungkyunkwan University), WonJun Moon (Sungkyunkwan University), and Jae-Pil Heo (Sungkyunkwan University)

Leveraging Self-Supervision for Cross-Domain Crowd Counting 5331
Weizhe Liu (EPFL), Nikita Durasov (École polytechnique fédérale de Lausanne), and Pascal Fua (EPFL, Switzerland)

What To Look at and Where: Semantic and Spatial Refined Transformer for Detecting Human-Object Interactions 5343
A S M Iftekhar (University of California, Santa Barbara), Hao Chen (Amazon), Kaustav Kundu (Amazon), Xinyu Li (ByteDance), Joseph Tighe (Amazon), and Davide Modolo (Amazon)

AdaMixer: A Fast-Converging Query-Based Object Detector 5354
Ziteng Gao (Nanjing University), Limin Wang (Nanjing University), Bing Han (antgroup), and Sheng Guo (Ant Group)

Correlation Verification for Image Retrieval 5364
Seongwon Lee (Yonsei university), Hongje Seong (Yonsei University), Suhyeon Lee (Yonsei University), and Euntai Kim (Yonsei University)

Real-Time Object Detection for Streaming Perception 5375
Jinrong Yang (Huazhong University of Science and Technology), Songtao Liu (MEGVII), Zeming Li (Megvii (Face++) Inc, ,), Xiaoping Li (Huazhong University of Science and Technology), and Jian Sun (Megvii Technology)

Deep Visual Geo-Localization Benchmark 5386
Gabriele Berton (Politecnico di Torino), Riccardo Mereu (Politecnico di Torino), Gabriele Trivigno (Polytechnic of Turin), Carlo Masone (Consorzio Interuniversitario Nazionale per l'Informatica), Gabriela Csurka (Naver Labs Europe), Torsten Sattler (Czech Technical University in Prague), and Barbara Caputo (Politecnico di Torino)

RendNet: Unified 2D/3D Recognizer With Latent Space Rendering 5398
Ruoxi Shi (Shanghai Jiao Tong University), Xinyang Jiang (Microsoft Research Asia), Caihua Shan (microsoft), Yansen Wang (Microsoft), and Dongsheng Li (Microsoft Research Asia)

Sparse Fuse Dense: Towards High Quality 3D Detection With Depth Completion 5408
Xiaopei Wu (ZhejiangUniversity), Liang Peng (ZJU), Honghui Yang (Zhejiang University), Liang Xie (Zhejiang University), Chenxi Huang (Zhejiang University), Chengqi Deng (Zhejiang University), Haifeng Liu (ZJU), and Deng Cai (ZJU)

Focal Sparse Convolutional Networks for 3D Object Detection 5418
Yukang Chen (The Chinese University of Hong Kong), Yanwei Li (The Chinese University of Hong Kong), Xiangyu Zhang (Megvii Technology), Jian Sun (Megvii Technology), and Jiaya Jia (Chinese University of Hong Kong)

Oral 2.1.2: 3D From Multi-View & Sensors

Point-NeRF: Point-Based Neural Radiance Fields 5428
Qiangeng Xu (USC), Zexiang Xu (Adobe Research), Julien Philip (Adobe), Sai Bi (Adobe Research), Zhixin Shu (Adobe Research), Kalyan Sunkavalli (Adobe Research), and Ulrich Neumann (USC)

NeRFusion: Fusing Radiance Fields for Large-Scale Scene Reconstruction 5439
Xiaoshuai Zhang (UC San Diego), Sai Bi (Adobe Research), Kalyan Sunkavalli (Adobe Research), Hao Su (University of California San Diego), and Zexiang Xu (Adobe Research)

Direct Voxel Grid Optimization: Super-Fast Convergence for Radiance Fields Reconstruction 5449
Cheng Sun (National Tsing Hua University), Min Sun (NTHU), and Hwann-Tzong Chen (National Tsing Hua University)

Mip-NeRF 360: Unbounded Anti-Aliased Neural Radiance Fields 5460
Jonathan T. Barron (Google Research), Ben Mildenhall (Google Research), Dor Verbin (Harvard University), Pratul P. Srinivasan (Google Research), and Peter Hedman (Google Research)

RegNeRF: Regularizing Neural Radiance Fields for View Synthesis From Sparse Inputs 5470
Michael Niemeyer (Max Planck Institute for Intelligent Systems, Tübingen and University of Tübingen), Jonathan T. Barron (Google Research), Ben Mildenhall (Google Research), Mehdi S. M. Sajjadi (Google Brain), Andreas Geiger (University of Tuebingen), and Noha Radwan (Google)

Ref-NeRF: Structured View-Dependent Appearance for Neural Radiance Fields 5481
Dor Verbin (Harvard University), Peter Hedman (Google Research), Ben Mildenhall (Google Research), Todd Zickler (Harvard University), Jonathan T. Barron (Google Research), and Pratul P. Srinivasan (Google Research)

Plenoxels: Radiance Fields Without Neural Networks 5491
Sara Fridovich-Keil (UC Berkeley), Alex Yu (UC Berkeley), Matthew Tancik (UC Berkeley), Qinhong Chen (University of California, Berkeley), Benjamin Recht (UC Berkeley), and Angjoo Kanazawa (University of California Berkeley)

Neural 3D Scene Reconstruction With the Manhattan-World Assumption 5501
Haoyu Guo (Zhejiang University), Sida Peng (Zhejiang University), Haotong Lin (Zhejiang University), Qianqian Wang (Cornell), Guofeng Zhang (Zhejiang University), Hujun Bao (Zhejiang University), and Xiaowei Zhou (Zhejiang University)

Neural 3D Video Synthesis From Multi-View Video 5511
Tianye Li (University of Southern California), Mira Slavcheva (Reality Labs Research), Michael Zollhöfer (Facebook Reality Labs), Simon Green (Facebook Reality Labs), Christoph Lassner (Meta Reality Labs Research), Changil Kim (Facebook), Tanner Schmidt (Facebook Reality Labs), Steven Lovegrove (Facebook Reality Labs), Michael Goesele (Facebook Reality Labs), Richard Newcombe (Facebook), and Zhaoyang Lv (Facebook)

Learning To Solve Hard Minimal Problems 5522
Petr Hruby (ETH Zürich), Timothy Duff (University of Washington), Anton Leykin (Georgia Institute of Technology), and Tomas Pajdla (Czech Technical University in Prague)

Learning a Structured Latent Space for Unsupervised Point Cloud Completion 5533
Yingjie Cai (CUHK), Kwan-Yee Lin (SenseTime Research), Chao Zhang (Samsung Telecommunication Research Institute), Qiang Wang (Samsung Research China, Beijing), Xiaogang Wang (Chinese University of Hong Kong, Hong Kong), and Hongsheng Li (The Chinese University of Hong Kong)

Lepard: Learning Partial Point Cloud Matching in Rigid and Deformable Scenes 5544
Yang Li (The University of Tokyo) and Tatsuya Harada (The University of Tokyo / RIKEN)

IRON: Inverse Rendering by Optimizing Neural SDFs and Materials From Photometric Images .. 5555
Kai Zhang (Cornell University), Fujun Luan (Adobe Research), Zhengqi Li (Cornell University), and Noah Snavely (Cornell University and Google AI)

Learning Multi-View Aggregation in the Wild for Large-Scale 3D Semantic Segmentation 5565
Damien Robert (IGN), Bruno Vallet (IGN), and Loic Landrieu (IGN)

HyperDet3D: Learning a Scene-Conditioned 3D Object Detector 5575
Yu Zheng (Tsinghua University), Yueqi Duan (Tsinghua University), Jiwen Lu (Tsinghua University), Jie Zhou (Tsinghua University), and Qi Tian (Huawei Cloud & AI)

KeyTr: Keypoint Transporter for 3D Reconstruction of Deformable Objects in Videos 5585
David Novotny (Facebook AI Research), Ignacio Rocco (Facebook), Samarth Sinha (Facebook AI Research), Alexandre Carlier (ETH Zurich), Gael Kerchenbaum (Facebook AI Research), Roman Shapovalov (Facebook AI Research), Nikita Smetanin (Facebook AI Research), Natalia Neverova (Facebook AI Research), Benjamin Graham (Facebook Research), and Andrea Vedaldi (University of Oxford / Facebook AI Research)

SelfRecon: Self Reconstruction Your Digital Avatar From Monocular Video 5595
Boyi Jiang (University of Science and Technology of China), Yang Hong (University of Science and Technology of China), Hujun Bao (Zhejiang University), and Juyong Zhang (University of Science and Technology of China)

Ditto: Building Digital Twins of Articulated Objects From Interaction 5606
Zhenyu Jiang (University of Texas at Austin), Cheng-Chun Hsu (Academia Sinica), and Yuke Zhu (University of Texas - Austin)

Oral 2.1.3: Low-Level Vision

Bijective Mapping Network for Shadow Removal 5617
Yurui Zhu (University of Science and Technology of China), Jie Huang (University of Science and Technology of China), Xueyang Fu (University of Science and Technology of China), Feng Zhao (University of Science and Technology of China), Qibin Sun (University of Science and Technology of China), and Zheng-Jun Zha (University of Science and Technology of China)

Toward Fast, Flexible, and Robust Low-Light Image Enhancement 5627
Long Ma (Dalian University of Technology), Tengyu Ma (Dalian University of Technology), Risheng Liu (Dalian University of Technology), Xin Fan (Dalian University of Technology), and Zhongxuan Luo (Dalian University of Technology)

Robust Equivariant Imaging: A Fully Unsupervised Framework for Learning To Image From Noisy and Partial Measurements 5637
Dongdong Chen (University of Edinburgh), Julián Tachella (CNRS & ENS de Lyon), and Mike E. Davies (University of Edinburgh)

Details or Artifacts: A Locally Discriminative Learning Approach to Realistic Image Super-Resolution 5647
Jie Liang (The Hong Kong Polytechnic University), Hui Zeng (OPPO), and Lei Zhang (Hong Kong Polytechnic University, Hong Kong, China)

Dual Adversarial Adaptation for Cross-Device Real-World Image Super-Resolution 5657
Xiaoqian Xu (Sun Yat-sen University), Pengxu Wei (Sun Yat-sen University), Weikai Chen (Tencent America), Yang Liu (Sun Yat-sen University), Mingzhi Mao (Sun Yat-sen University), Liang Lin (Sun Yat-sen University), and Guanbin Li (Sun Yat-sen University)

SphereSR: 360° Image Super-Resolution With Arbitrary Projection via Continuous Spherical Image Representation 5667
Youngho Yoon (KAIST), Inchul Chung (KAIST), Lin Wang (HKUST), and Kuk-Jin Yoon (KAIST)

Learning Trajectory-Aware Transformer for Video Super-Resolution 5677
Chengxu Liu (Xi'an Jiaotong University), Huan Yang (Microsoft Research), Jianlong Fu (Microsoft Research), and Xueming Qian (Xi'an Jiaotong University)

Discrete Cosine Transform Network for Guided Depth Map Super-Resolution 5687
Zixiang Zhao (Xi'an Jiaotong University), Jiangshe Zhang (Xi'an Jiaotong University), Shuang Xu (Xi'an Jiaotong University), Zudi Lin (Harvard University), and Hanspeter Pfister (Harvard University)

Faithful Extreme Rescaling via Generative Prior Reciprocated Invertible Representations 5698
Zhixuan Zhong (South China University of Technology), Liangyu Chai (South China University of Technology), Yang Zhou (South China University of Technology), Bailin Deng (Cardiff University), Jia Pan (The University of Hong Kong), and Shengfeng He (South China University of Technology)

ELIC: Efficient Learned Image Compression With Unevenly Grouped Space-Channel Contextual Adaptive Coding 5708
Dailan He (SenseTime), Ziming Yang (SenseTime), Weikun Peng (SenseTime), Rui Ma (Beihang University), Hongwei Qin (Sensetime), and Yan Wang (Tsinghua University)

Restormer: Efficient Transformer for High-Resolution Image Restoration 5718
Syed Waqas Zamir (IIAI), Aditya Arora (IIAI), Salman Khan (MBZUAI/ANU), Munawar Hayat (Monash University), Fahad Shahbaz Khan (MBZUAI), and Ming-Hsuan Yang (University of California at Merced)

Deep Rectangling for Image Stitching: A Learning Baseline 5730
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)

Parametric Scattering Networks 5739
Shanel Gauthier (University of Montreal, Mila), Benjamin Thérien (Waterloo University), Laurent Alsène-Racicot (UdeM/Mila), Muawiz Chaudhary (MILA), Irina Rish (Mila/UdeM), Eugene Belilovsky (Mila), Michael Eickenberg (Flatiron Institute), and Guy Wolf (Université de Montréal)

Burst Image Restoration and Enhancement 5749
Akshay Dudhane (Mohamed bin Zayed University of Artificial Intelligence), Syed Waqas Zamir (IIAI), Salman Khan (MBZUAI/ANU), Fahad Shahbaz Khan (MBZUAI), and Ming-Hsuan Yang (University of California at Merced)

MAXIM: Multi-Axis MLP for Image Processing 5759
Zhengzhong Tu (University of Texas at Austin), Hossein Talebi (Google), Han Zhang (Google), Feng Yang (Google Research), Peyman Milanfar (Google), Alan Bovik (University of Texas at Austin), and Yinxiao Li (Google)

Event-Aided Direct Sparse Odometry 5771
Javier Hidalgo-Carrió (University of Zurich), Guillermo Gallego (TU Berlin), and Davide Scaramuzza (University of Zurich & ETH Zurich, Switzerland)

CamLiFlow: Bidirectional Camera-LiDAR Fusion for Joint Optical Flow and Scene Flow Estimation 5781
Haisong Liu (Nanjing University), Tao Lu (Nanjing University), Yihui Xu (Nanjing University), Jia Liu (Nanjing University), Wenjie Li (Nanjing University), and Lijun Chen (Nanjing University)

Target-Aware Dual Adversarial Learning and a Multi-Scenario Multi-Modality Benchmark To Fuse Infrared and Visible for Object Detection 5792
Jinyuan Liu (Dalian University of Technology), Xin Fan (Dalian University of Technology), Zhanbo Huang (Dalian University of Technology), Guanyao Wu (Dalian University of Technology), Risheng Liu (Dalian University of Technology), Wei Zhong (Dalian University of Technology), and Zhongxuan Luo (Dalian University of Technology)

Poster 2.1: Low-Level Vision

Image Dehazing Transformer With Transmission-Aware 3D Position Embedding 5802
Chun-Le Guo (Nankai University), Qixin Yan (Tianjin University), Saeed Anwar (The Australian National University), Runmin Cong (Beijing Jiaotong University), Wenqi Ren (Institute of Information Engineering, Chinese Academy of Sciences), and Chongyi Li (Nanyang Technological University)

Unsupervised Deraining: Where Contrastive Learning Meets Self-Similarity 5811
Yuntong Ye (Huazhong University of Science and Technology), Changfeng Yu (Huazhong University of Science and Technology), Yi Chang (PengCheng Laboratory), Lin Zhu (Peking University), Xi-Le Zhao (School of Mathematical Sciences, University of Electronic Science and Technology of China), Luxin Yan (Huazhong University of Science and Technology), and Yonghong Tian (Peking University)

Towards Multi-Domain Single Image Dehazing via Test-Time Training 5821
Huan Liu (McMaster University), Zijun Wu (McMaster University), Liangyan Li (McMaster university), Sadaf Salehkalaibar (McMaster university), Jun Chen (McMaster University), and Keyan Wang (State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an, China)

Physically Disentangled Intra- and Inter-Domain Adaptation for Varicolored Haze Removal 5831
Yi Li (Huazhong University of Science and Technology), Yi Chang (PengCheng Laboratory), Yan Gao (Huazhong University of Science and Technology), Changfeng Yu (Huazhong University of Science and Technology), and Luxin Yan (Huazhong University of Science and Technology)

Incorporating Semi-Supervised and Positive-Unlabeled Learning for Boosting Full Reference Image Quality Assessment 5841
Yue Cao (Harbin Institute of Technology), Zhaolin Wan (Dalian Maritime University), Dongwei Ren (Harbin Institute of Technology), Zifei Yan (Harbin Institute of Technology), and Wangmeng Zuo (Harbin Institute of Technology, China)

Practical Learned Lossless JPEG Recompression With Multi-Level Cross-Channel Entropy Model in the DCT Domain 5852
Lina Guo (Sensetime), Xinjie Shi (SenseTime), Dailan He (SenseTime), Yuanyuan Wang (Tsinghua University), Rui Ma (Sensetime), Hongwei Qin (Beihang University), and Yan Wang (Sensetime)

Neural Compression-Based Feature Learning for Video Restoration 5862
Cong Huang (USTC), Jiahao Li (Microsoft Research), Bin Li (Microsoft Research Asia), Dong Liu (University of Science and Technology of China), and Yan Lu (Microsoft Research Asia)

Bi-Directional Object-Context Prioritization Learning for Saliency Ranking 5872
Xin Tian (Dalian University of Technology; City University of Hong Kong), Ke Xu (City University of Hong Kong), Xin Yang (Dalian University of Technology), Lin Du (AI Application Research Center, Huawei Technologies Co., Ltd), Baocai Yin (Dalian University of Technology), and Rynson W.H. Lau (City University of Hong Kong)

Pixel Screening Based Intermediate Correction for Blind Deblurring 5882
Meina Zhang (Institute of applied physics and computational mathematics), Yingying Fang (Imperial College London), Guoxi Ni (Institute of applied physics and computational mathematics, Beijing), and Tieyong Zeng (The Chinese University of Hong Kong)

URetinex-Net: Retinex-Based Deep Unfolding Network for Low-Light Image Enhancement 5891
Wenhui Wu (Shenzhen University), Jian Weng (ShenZhen University), Pingping Zhang (City University of Hong Kong), Xu Wang (Shenzhen University), Wenhan Yang (NTU), and Jianmin Jiang (Shenzhen University)

A Text Attention Network for Spatial Deformation Robust Scene Text Image Super-Resolution ... 5901
Jianqi Ma (the Hong Kong Polytechnic University), Zhetong Liang (Alibaba DAMO Academy), and Lei Zhang (Hong Kong Polytechnic University, Hong Kong, China)

Coarse-To-Fine Deep Video Coding With Hyperprior-Guided Mode Prediction 5911
Zhihao Hu (Beihang University), Guo Lu (Beijing Institute of Technology), Jinyang Guo (The University of Sydney), Shan Liu (Tencent America), Wei Jiang (Tencent America), and Dong Xu (University of Sydney)

Task Decoupled Framework for Reference-Based Super-Resolution 5921
Yixuan Huang (Shanghai Jiao Tong University), Xiaoyun Zhang (Shanghai Jiao Tong University), Yu Fu (Shanghai Jiao Tong University), Siheng Chen (Shanghai Jiao Tong University), Ya Zhang (Cooperative Medianet Innovation Center, Shang hai Jiao Tong University), Yan-Feng Wang (Cooperative medianet innovation center of Shanghai Jiao Tong University), and Dazhi He (Shanghai Jiao Tong University)

Learning Semantic Associations for Mirror Detection 5931
Huankang Guan (City University of Hong Kong), Jiaying Lin (City University of Hong Kong), and Rynson W.H. Lau (City University of Hong Kong)

SketchEdit: Mask-Free Local Image Manipulation With Partial Sketches 5941
Yu Zeng (Johns Hopkins University), Zhe Lin (Adobe Research), and Vishal M. Patel (Johns Hopkins University)

Investigating Tradeoffs in Real-World Video Super-Resolution 5952
Kelvin C.K. Chan (Nanyang Technological University), Shangchen Zhou (Nanyang Technological University), Xiangyu Xu (Sea AI Lab), and Chen Change Loy (Nanyang Technological University)

BasicVSR++: Improving Video Super-Resolution With Enhanced Propagation and Alignment 5962
Kelvin C.K. Chan (Nanyang Technological University), Shangchen Zhou (Nanyang Technological University), Xiangyu Xu (Sea AI Lab), and Chen Change Loy (Nanyang Technological University)

Inertia-Guided Flow Completion and Style Fusion for Video Inpainting 5972
Kaidong Zhang (University of Science and Technology of China), Jingjing Fu (Microsoft), and Dong Liu (University of Science and Technology of China)

Joint Global and Local Hierarchical Priors for Learned Image Compression 5982
Jun-Hyuk Kim (Yonsei University), Byeongho Heo (NAVER AI LAB), and Jong-Seok Lee (Yonsei University, Korea)

Reflash Dropout in Image Super-Resolution 5992
Xiangtao Kong (SIAT), Xina Liu (Shenzhen Institutes of Advanced TechnologyChinese Academy of Sciences), Jinjin Gu (The University of Sydney), Yu Qiao (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences), and Chao Dong (SIAT)

Towards Robust Rain Removal Against Adversarial Attacks: A Comprehensive Benchmark Analysis and Beyond 6003
Yi Yu (Nanyang Technological University), Wenhan Yang (NTU), Yap-Peng Tan (Nanyang Technological University, Singapore), and Alex C. Kot (Nanyang Technological University)

Dreaming To Prune Image Deraining Networks 6013
Weiqi Zou (University of Science and Technology of China), Yang Wang (University of Science and Technology of China), Xueyang Fu (University of Science and Technology of China), and Yang Cao (University of Science and Technology of China)

LC-FDNet: Learned Lossless Image Compression With Frequency Decomposition Network 6023
Hochang Rhee (Seoul National University), Yeong Il Jang (Seoul National University), Seyun Kim (Gauss Labs), and Nam Ik Cho (Seoul National University)

Exposure Normalization and Compensation for Multiple-Exposure Correction 6033
Jie Huang (University of Science and Technology of China), Yajing Liu (USTC), Xueyang Fu (University of Science and Technology of China), Man Zhou (University of Science and Technology of China), Yang Wang (University of Science and Technology of China), Feng Zhao (University of Science and Technology of China), and Zhiwei Xiong (University of Science and Technology of China)

Revisiting Temporal Alignment for Video Restoration 6043
Kun Zhou (SmartMore Corporation), Wenbo Li (The Chinese University of Hong Kong), Liying Lu (The Chinese University of Hong Kong), Xiaoguang Han (Shenzhen Research Institute of Big Data, the Chinese University of Hong Kong (Shenzhen)), and Jiangbo Lu (SmartMore Corporation)

Learning the Degradation Distribution for Blind Image Super-Resolution 6053
Zhengxiong Luo (Institute of AutomationChinese Academy of Sciences), Yan Huang (Institute of Automation, Chinese Academy of Sciences), Shang Li (CASIA), Liang Wang (NLPR, China), and Tieniu Tan (NLPR, China)

LSVC: A Learning-Based Stereo Video Compression Framework 6063
Zhenghao Chen (University of Sydney), Guo Lu (Beijing Institute of Technology), Zhihao Hu (Beihang University), Shan Liu (Tencent America), Wei Jiang (Tencent America), and Dong Xu (University of Sydney)

Learning Based Multi-Modality Image and Video Compression 6073
Guo Lu (Beijing Institute of Technology), Tianxiong Zhong (Beijing Institute of Technology), Jing Geng (Beijing institute of technology), Qiang Hu (ShanghaiTech University), and Dong Xu (University of Sydney)

Transformer Based Line Segment Classifier With Image Context for Real-Time Vanishing Point Detection in Manhattan World 6083
Xin Tong (Peking University), Xianghua Ying (Peking University), Yongjie Shi (Peking University), Ruibin Wang (Peking University), and Jinfa Yang (Peking University)

Deep Vanishing Point Detection: Geometric Priors Make Dataset Variations Vanish 6093
Yancong Lin (Delft University of Technology), Ruben Wiersma (Delft University of Technology), Silvia L. Pintea (TU Delft), Klaus Hildebrandt (TU Delft), Elmar Eisemann (Delft University of Technology), and Jan C. van Gemert (Delft University of Technology)

Stereo Depth From Events Cameras: Concentrate and Focus on the Future 6104
Yeongwoo Nam (GIST), Mohammad Mostafavi (GIST, Lunit), Kuk-Jin Yoon (KAIST), and Jonghyun Choi (Yonsei University)

Poster 2.1: 3D From Multi-View & Sensors

Volumetric Bundle Adjustment for Online Photorealistic Scene Capture 6114
Ronald Clark (Imperial College London)

Neural Volumetric Object Selection 6123
Zhongzheng Ren (UIUC), Aseem Agarwala (Google), Bryan Russell (Adobe Research), Alexander G. Schwing (UIUC), and Oliver Wang (Adobe Systems Inc)

HVH: Learning a Hybrid Neural Volumetric Representation for Dynamic Hair Performance Capture 6133
Ziyan Wang (Carnegie Mellon University), Giljoo Nam (Facebook Inc.), Tuur Stuyck (Meta Reality Labs Research), Stephen Lombardi (Facebook), Michael Zollhöfer (Facebook Reality Labs), Jessica Hodgins (Carnegie Mellon University), and Christoph Lassner (Meta Reality Labs Research)

NeuralHOFusion: Neural Volumetric Rendering Under Human-Object Interactions 6145
Yuheng Jiang (ShanghaiTech University), Suyi Jiang (ShanghaiTech University), Guoxing Sun (ShanghaiTech University), Zhuo Su (Tencent), Kaiwen Guo (Google), Minye Wu (KU Leuven), Jingyi Yu (Shanghai Tech University), and Lan Xu (ShanghaiTech)

BNV-Fusion: Dense 3D Reconstruction Using Bi-Level Neural Volume Fusion 6156
Kejie Li (University of Oxford), Yansong Tang (Tsinghua University), Victor Adrian Prisacariu (University of Oxford), and Philip H.S. Torr (University of Oxford)

Input-Level Inductive Biases for 3D Reconstruction 6166
Wang Yifan (Stanford University), Carl Doersch (DeepMind), Relja Arandjelović (DeepMind), João João Carreira (DeepMind), and Andrew Zisserman (University of Oxford)

Multi-View Mesh Reconstruction With Neural Deferred Shading 6177
Markus Worchel (Fraunhofer Heinrich Hertz Institute), Rodrigo Diaz (Fraunhofer Heinrich-Hertz-Institute), Weiwen Hu (Fraunhofer Heinrich-Hertz-Institute), Oliver Schreer (Fraunhofer Heinrich-Hertz-Institute), Ingo Feldmann (Fraunhofer HHI), and Peter Eisert (Fraunhofer Heinrich Hertz Institute)

StyleMesh: Style Transfer for Indoor 3D Scene Reconstructions 6188
Lukas Höllein (Technical University of Munich), Justin Johnson (University of Michigan), and Matthias Nießner (Technical University of Munich)

RGB-Depth Fusion GAN for Indoor Depth Completion 6199
Haowen Wang (Beijing University of Posts and Telecommunications), Mingyuan Wang (Beijing University of Posts and Telecommunications), Zhengping Che (Midea Group), Zhiyuan Xu (Midea Group), Xiuquan Qiao (Beijing University of Posts and Telecommunications), Mengshi Qi (Beijing University of Posts and Telecommunications), Feifei Feng (Midea Grooup), and Jian Tang (Midea Group)

PlanarRecon: Real-Time 3D Plane Detection and Reconstruction From Posed Monocular Videos .. 6209
Yiming Xie (Northeastern University), Matheus Gadelha (University of Massachusetts Amherst), Fengting Yang (Pennsylvania State University), Xiaowei Zhou (Zhejiang University), and Huaizu Jiang (Northeastern University)

Scene Representation Transformer: Geometry-Free Novel View Synthesis Through Set-Latent Scene Representations 6219
Mehdi S. M. Sajjadi (Google Brain), Henning Meyer (Google), Etienne Pot (Google), Urs Bergmann (Google Research), Klaus Greff (Google), Noha Radwan (Google), Suhani Vora (Google), Mario Lučić (Google Brain), Daniel Duckworth (Google), Alexey Dosovitskiy (Inceptive), Jakob Uszkoreit (Google), Thomas Funkhouser (Google Research), and Andrea Tagliasacchi (Google Brain and University of Toronto)

ShapeFormer: Transformer-Based Shape Completion via Sparse Representation 6229
Xingguang Yan (Shenzhen University), Liqiang Lin (Shenzhen University), Niloy J. Mitra (University College London), Dani Lischinski (The Hebrew University of Jerusalem), Daniel Cohen-Or (Tel Aviv University), and Hui Huang (Shenzhen University)

GuideFormer: Transformers for Image Guided Depth Completion 6240
Kyeongha Rho (Agency for Defense Development), Jinsung Ha (LUXROBO), and Youngjung Kim (ADD)

Improving Neural Implicit Surfaces Geometry With Patch Warping 6250
François Darmon (École des ponts ParisTech / Thales LAS France), Bénédicte Bascle (Thales LAS), Jean-Clément Devaux (Thales LAS), Pascal Monasse (Ecole des Ponts ParisTech), and Mathieu Aubry (École des ponts ParisTech)

Critical Regularizations for Neural Surface Reconstruction in the Wild 6260
Jingyang Zhang (The Hong Kong University of Science and Technology), Yao Yao (Apple Inc.), Shiwei Li (Apple Inc.), Tian Fang (Apple), David McKinnon (Apple), Yanghai Tsin (Apple Inc), and Long Quan (Apple)

Gradient-SDF: A Semi-Implicit Surface Representation for 3D Reconstruction 6270
Christiane Sommer (Technical University of Munich), Lu Sang (Technical University of Munich), David Schubert (Technical University of Munich), and Daniel Cremers (TU Munich)

Neural RGB-D Surface Reconstruction 6280
Dejan Azinović (Technical University of Munich), Ricardo Martin-Brualla (Google), Dan B Goldman (Google, Inc.), Matthias Nießner (Technical University of Munich), and Justus Thies (Max Planck Institute for Intelligent Systems)

POCO: Point Convolution for Surface Reconstruction 6292
Alexandre Boulch (valeo.ai) and Renaud Marlet (Ecole des Ponts ParisTech)

Reconstructing Surfaces for Sparse Point Clouds With On-Surface Priors 6305
Baorui Ma (Tsinghua university), Yu-Shen Liu (Tsinghua University), and Zhizhong Han (Wayne State University)

Surface Reconstruction From Point Clouds by Learning Predictive Context Priors 6316
Baorui Ma (Tsinghua university), Yu-Shen Liu (Tsinghua University), Matthias Zwicker (University of Maryland), and Zhizhong Han (Wayne State University)

IDEA-Net: Dynamic 3D Point Cloud Interpolation via Deep Embedding Alignment 6328
Yiming Zeng (City University of Hong Kong), Yue Qian (City University of Hong Kong, Hong Kong), Qijian Zhang (City University of Hong Kong), Junhui Hou (City University of Hong Kong, Hong Kong), Yixuan Yuan (City University of Hong Kong), and Ying He (Nanyang Technological University)

Deterministic Point Cloud Registration via Novel Transformation Decomposition 6338
Wen Chen (The Chinese University of Hong Kong), Haoang Li (The Chinese University of Hong Kong), Qiang Nie (Tencent Youtu Lab), and Yun-Hui Liu (CUHK)

Global-Aware Registration of Less-Overlap RGB-D Scans 6347
Che Sun (Beijing Institute of Technology), Yunde Jia (Beijing Institute of Technology), Yi Guo (Beijing Institute of Technology), and Yuwei Wu (Beijing Institute of Technology (BIT), China, ,)

Finding Good Configurations of Planar Primitives in Unorganized Point Clouds 6357
Mulin Yu (INRIA) and Florent Lafarge (INRIA)

Self-Supervised Global-Local Structure Modeling for Point Cloud Domain Adaptation With Reliable Voted Pseudo Labels 6367
*Hehe Fan (NUS), Xiaojun Chang (University of Technology Sydney), Wanyue Zhang (A*STAR), Yi Cheng (I2R, ASTAR), Ying Sun (Institute for Infocomm Research, A*STAR), and Mohan Kankanhalli (National University of Singapore,)*

AziNorm: Exploiting the Radial Symmetry of Point Cloud for Azimuth-Normalized 3D Perception 6377
Shaoyu Chen (Huazhong University of Science and Technology), Xinggang Wang (Huazhong University of Science and Technology), Tianheng Cheng (Huazhong University of Science and Technology), Wenqiang Zhang (Huazhong University of Science and Technology), Qian Zhang (Horizon Robotics), Chang Huang (Horizon Robotics), and Wenyu Liu (Huazhong University of Science and Technology)

WarpingGAN: Warping Multiple Uniform Priors for Adversarial 3D Point Cloud Generation 6387
Yingzhi Tang (City University of Hong Kong), Yue Qian (City University of Hong Kong, Hong Kong), Qijian Zhang (City University of Hong Kong), Yiming Zeng (City University of Hong Kong), Junhui Hou (City University of Hong Kong, Hong Kong), and Xuefei Zhe (Tencent AI lab)

Poster 2.1: Motion & Tracking

Forward Propagation, Backward Regression, and Pose Association for Hand Tracking in the Wild 6396
Mingzhen Huang (University at Buffalo), Supreeth Narasimhaswamy (Stony Brook University), Saif Vazir (Tulip Interfaces), Haibin Ling (Stony Brook University), and Minh Hoai (Stony Brook University)

Neural MoCon: Neural Motion Control for Physically Plausible Human Motion Capture 6407
Buzhen Huang (Southeast University), Liang Pan (Southeast University), Yuan Yang (Southeast University), Jingyi Ju (Southeast University), and Yangang Wang (Southeast University)

MotionAug: Augmentation With Physical Correction for Human Motion Prediction 6417
Takahiro Maeda (Toyota Technological Institute) and Norimichi Ukita (TTI-J)

Progressively Generating Better Initial Guesses Towards Next Stages for High-Quality Human Motion Prediction 6427
Tiezheng Ma (SCUT), Yongwei Nie (South China University of Technology), Chengjiang Long (Meta Reality Labs), Qing Zhang (Sun Yat-sen University), and Guiqing Li (South China University of Technology)

Spatio-Temporal Gating-Adjacency GCN for Human Motion Prediction 6437
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), Yongjing Ye (ICT), and Shihong Xia (institute of computing technology of the Chinese academy of sciences)

Motron: Multimodal Probabilistic Human Motion Forecasting 6447
Tim Salzmann (Technical University of Munich), Marco Pavone (Stanford University), and Markus Ryll (Technical University of Munich)

Human Trajectory Prediction With Momentary Observation 6457
Jianhua Sun (Shanghai Jiao Tong University), Yuxuan Li (Shanghai Jiao Tong University), Liang Chai (Shanghai Jiao Tong University), Hao-Shu Fang (SJTU), Yong-Lu Li (Shanghai Jiao Tong University), and Cewu Lu (Shanghai Jiao Tong University)

Non-Probability Sampling Network for Stochastic Human Trajectory Prediction 6467
Inhwan Bae (Gwangju Institute of Science and Technology), Jin-Hwi Park (GIST), and Hae-Gon Jeon (GIST)

Remember Intentions: Retrospective-Memory-Based Trajectory Prediction 6478
Chenxin Xu (Shanghai Jiao Tong University), Weibo Mao (SJTU), Wenjun Zhang (Shanghai Jiao Tong University), and Siheng Chen (Shanghai Jiao Tong University)

GroupNet: Multiscale Hypergraph Neural Networks for Trajectory Prediction With Relational Reasoning 6488
Chenxin Xu (Shanghai Jiao Tong University), Maosen Li (Cooperative Medianet Innovation Center, Shanghai Jiao Tong University), Zhenyang Ni (Shanghai Jiao Tong University), Ya Zhang (Cooperative Medianet Innovation Center, Shang hai Jiao Tong University), and Siheng Chen (Shanghai Jiao Tong University)

Learning Pixel Trajectories With Multiscale Contrastive Random Walks 6498
Zhangxing Bian (University of Michigan), Allan Jabri (UC Berkeley), Alexei A. Efros (UC Berkeley), and Andrew Owens (University of Michigan)

Adaptive Trajectory Prediction via Transferable GNN 6510
Yi Xu (Northeastern University), Lichen Wang (Northeastern University), Yizhou Wang (Northeastern University), and Yun Fu (Northeastern University)

Neural Prior for Trajectory Estimation 6522
Chaoyang Wang (Carnegie Mellon University), Xueqian Li (Carnegie Mellon University), Jhony Kaesemodel Pontes (Argo AI), and Simon Lucey (University of Adelaide)

M2I: From Factored Marginal Trajectory Prediction to Interactive Prediction 6533
Qiao Sun (Tsinghua University), Xin Huang (MIT), Junru Gu (Tsinghua University), Brian C. Williams (Massachusetts Institute of Technology), and Hang Zhao (Tsinghua University)

How Many Observations Are Enough? Knowledge Distillation for Trajectory Forecasting 6543
Alessio Monti (AImageLab - University of Modena and Reggio Emilia), Angelo Porrello (University of Modena and Reggio Emilia), Simone Calderara (University of Modena and Reggio Emilia, Italy), Pasquale Coscia (University of Padova), Lamberto Ballan (University of Padova), and Rita Cucchiara (Università di Modena e Reggio Emilia)

ATPFL: Automatic Trajectory Prediction Model Design Under Federated Learning Framework ... 6553
Chunnan Wang (HIT), Xiang Chen (Harbin Institute of Technology), Junzhe Wang (Harbin Institute of Technology), and Hongzhi Wang (Harbin Institute of Technology)

Whose Track Is It Anyway? Improving Robustness to Tracking Errors With Affinity-Based Trajectory Prediction 6563
Xinshuo Weng (Carnegie Mellon University), Boris Ivanovic (Stanford University), Kris Kitani (Carnegie Mellon University), and Marco Pavone (Stanford University)

Convolutions for Spatial Interaction Modeling 6573
Zhaoen Su (Aurora), Chao Wang (Aurora Innovation), David Bradley (Waymo), Carlos Vallespi-Gonzalez (Uber), Carl Wellington (Aurora), and Nemanja Djuric (Aurora Innovation)

Style-ERD: Responsive and Coherent Online Motion Style Transfer 6583
Tianxin Tao (University of British Columbia), Xiaohang Zhan (The Chinese University of Hong Kong), Zhongquan Chen (University of California, Davis), and Michiel van de Panne (University of British Columbia)

Neural Inertial Localization 6594
Sachini Herath (Simon Fraser University), David Caruso (Meta), Chen Liu (Meta), Yufan Chen (Facebook Reality Labs), and Yasutaka Furukawa (Simon Fraser University)

RIO: Rotation-Equivariance Supervised Learning of Robust Inertial Odometry 6604
Xiya Cao (Huawei Technologies Co. Ltd), Caifa Zhou (Huawei Technologies), Dandan Zeng (Huawei), and Yongliang Wang (Huawei Technologies Co., Ltd.)

CaDeX: Learning Canonical Deformation Coordinate Space for Dynamic Surface Representation via Neural Homeomorphism 6614
Jiahui Lei (University of Pennsylvania) and Kostas Daniilidis (University of Pennsylvania)

Poster 2.1: Pose Estimation & Tracking

ElePose: Unsupervised 3D Human Pose Estimation by Predicting Camera Elevation and Learning Normalizing Flows on 2D Poses 6625
Bastian Wandt (University of British Columbia), James J. Little (University of British Columbia, Canada), and Helge Rhodin (UBC)

Projective Manifold Gradient Layer for Deep Rotation Regression 6636
Jiayi Chen (Peking University), Yingda Yin (Peking University), Tolga Birdal (Stanford University), Baoquan Chen (Peking University), Leonidas J. Guibas (Stanford University), and He Wang (Peking University)

Multimodal Colored Point Cloud to Image Alignment 6646
Noam Rotstein (Technion), Amit Bracha (Technion), and Ron Kimmel (Technion)

Multi-Instance Point Cloud Registration by Efficient Correspondence Clustering 6657
Weixuan Tang (shanghai jiao tong university) and Danping Zou (Shang Jiao Tong University)

REGTR: End-to-End Point Cloud Correspondences With Transformers 6667
Zi Jian Yew (National University of Singapore) and Gim Hee Lee (National University of Singapore)

Text2Pos: Text-to-Point-Cloud Cross-Modal Localization 6677
Manuel Kolmet (Technical University of Munich), Qunjie Zhou (Technical University of Munich), Aljoša Ošep (TUM Munich), and Laura Leal-Taixé (TUM)

BCOT: A Markerless High-Precision 3D Object Tracking Benchmark 6687
Jiachen Li (Shandong University), Bin Wang (Shandong University), Shiqiang Zhu (Zhejiang Lab), Xin Cao (Shandong University), Fan Zhong (Shandong University), Wenxuan Chen (Zhejiang Lab), Te Li (Zhejiang Lab), Jason Gu (Dalhousie University), and Xueying Qin (Shandong University)

SAR-Net: Shape Alignment and Recovery Network for Category-Level 6D Object Pose and Size Estimation 6697
Haitao Lin (Fudan University), Zichang Liu (Fudan University), Chilam Cheang (Fudan University), Yanwei Fu (Fudan University), Guodong Guo (Baidu), and Xiangyang Xue (Fudan University)

ES6D: A Computation Efficient and Symmetry-Aware 6D Pose Regression Framework 6708
Ningkai Mo (Shenzhen Institutes of Advanced TechnologyChinese Academy of Sciences), Wanshui Gan (The university of tokyo), Naoto Yokoya (The University of Tokyo), and Shifeng Chen (SIAT)

Coupled Iterative Refinement for 6D Multi-Object Pose Estimation 6718
Lahav Lipson (Princeton University), Zachary Teed (Princeton University), Ankit Goyal (Princeton University), and Jia Deng (Princeton University)

ZebraPose: Coarse To Fine Surface Encoding for 6DoF Object Pose Estimation 6728
Yongzhi Su (DFKI), Mahdi Saleh (Technical University of Munich), Torben Fetzer (TU Kaiserslautern), Jason Rambach (DFKI), Nassir Navab (TU Munich, Germany), Benjamin Busam (Technical University of Munich), Didier Stricker (DFKI), and Federico Tombari (Google, TU Munich)

SurfEmb: Dense and Continuous Correspondence Distributions for Object Pose Estimation With Learnt Surface Embeddings 6739
Rasmus Laurvig Haugaard (University of Southern Denmark) and Anders Glent Buch (University of Southern Denmark)

MetaPose: Fast 3D Pose From Multiple Views Without 3D Supervision 6749
Ben Usman (Boston University), Andrea Tagliasacchi (Google Brain and University of Toronto), Kate Saenko (Boston University), and Avneesh Sud (Google)

Templates for 3D Object Pose Estimation Revisited: Generalization to New Objects and Robustness to Occlusions 6761
Van Nguyen Nguyen (Ecole des Ponts ParisTech), Yinlin Hu (EPFL), Yang Xiao (École des ponts ParisTech), Mathieu Salzmann (EPFL), and Vincent Lepetit (Ecole des Ponts ParisTech)

GPV-Pose: Category-Level Object Pose Estimation via Geometry-Guided Point-Wise Voting 6771
Yan Di (Technical University of Munich), Ruida Zhang (Tsinghua University), Zhiqiang Lou (Tsinghua University), Fabian Manhardt (Google), Xiangyang Ji (Tsinghua University), Nassir Navab (TU Munich, Germany), and Federico Tombari (Google, TU Munich)

HSC4D: Human-Centered 4D Scene Capture in Large-Scale Indoor-Outdoor Space Using Wearable IMUs and LiDAR 6782
Yudi Dai (Xiamen University), Yitai Lin (xiamen university), Chenglu Wen (Xiamen University), Siqi Shen (Xiamen University), Lan Xu (ShanghaiTech), Jingyi Yu (Shanghai Tech University), Yuexin Ma (ShanghaiTech University), and Cheng Wang (Xiamen University)

OVE6D: Object Viewpoint Encoding for Depth-Based 6D Object Pose Estimation 6793
Dingding Cai (Tampere University), Janne Heikkilä (University of Oulu, Finland), and Esa Rahtu (Tampere University)

FS6D: Few-Shot 6D Pose Estimation of Novel Objects 6804
Yisheng He (Hong Kong University of Science and Technology), Yao Wang (Megvii), Haoqiang Fan (Megvii Incface++), Jian Sun (Megvii Technology), and Qifeng Chen (HKUST)

OnePose: One-Shot Object Pose Estimation Without CAD Models 6815
Jiaming Sun (Zhejiang University), Zihao Wang (Zhejiang University), Siyu Zhang (SenseTime), Xingyi He (Zhejiang University), Hongcheng Zhao (TUM), Guofeng Zhang (Zhejiang University), and Xiaowei Zhou (Zhejiang University)

OSOP: A Multi-Stage One Shot Object Pose Estimation Framework 6825
Ivan Shugurov (TU Munich / Siemens Corporate Technology), Fu Li (National University of Defense Technology; Technical University of Munich), Benjamin Busam (Technical University of Munich), and Slobodan Ilic (TUM)

DiffPoseNet: Direct Differentiable Camera Pose Estimation 6835
Chethan M. Parameshwara (University of Maryland, College Park), Gokul Hari (University of Maryland, College Park), Cornelia Fermüller (University of Maryland, College Park), Nitin J. Sanket (University of Maryland, College Park), and Yiannis Aloimonos (University of Maryland, College Park)

Iterative Corresponding Geometry: Fusing Region and Depth for Highly Efficient 3D Tracking of Textureless Objects 6845
Manuel Stoiber (German Aerospace Center (DLR)), Martin Sundermeyer (German Aerospace Center (DLR)), and Rudolph Triebel (German Aerospace Center (DLR))

CPPF: Towards Robust Category-Level 9D Pose Estimation in the Wild 6856
Yang You (Shanghai Jiao Tong University), Ruoxi Shi (Shanghai Jiao Tong University), Weiming Wang (shanghai jiaotong university), and Cewu Lu (Shanghai Jiao Tong University)

Leveraging Equivariant Features for Absolute Pose Regression 6866
Mohamed Adel Musallam (SnT), Vincent Gaudillière (SnT), Miguel Ortiz del Castillo (SnT), Kassem Al Ismaeil (SnT), and Djamila Aouada (SnT)

Poster 2.1: Transfer / Low-Shot / Long-Tail Learning

The Majority Can Help the Minority: Context-Rich Minority Oversampling for Long-Tailed Classification 6877
Seulki Park (Seoul National University), Youngkyu Hong (NAVER AI LAB), Byeongho Heo (NAVER AI LAB), Sangdoo Yun (NAVER AI LAB), and Jin Young Choi (Seoul National University)

Long-Tailed Recognition via Weight Balancing 6887
Shaden Alshammari (Massachusetts Institute of Technology), Yu-Xiong Wang (University of Illinois at Urbana-Champaign), Deva Ramanan (Carnegie Mellon University), and Shu Kong (Carnegie Mellon University)

Balanced Contrastive Learning for Long-Tailed Visual Recognition 6898
Jianggang Zhu (Fudan University), Zheng Wang (Fudan University), Jingjing Chen (Fudan University), Yi-Ping Phoebe Chen (La Trobe University), and Yu-Gang Jiang (Fudan University)

Targeted Supervised Contrastive Learning for Long-Tailed Recognition 6908
Tianhong Li (MIT), Peng Cao (Massachusetts Institute of Technology), Yuan Yuan (MIT), Lijie Fan (MIT), Yuzhe Yang (MIT), Rogerio S. Feris (MIT-IBM Watson AI Lab, IBM Research), Piotr Indyk (MIT), and Dina Katabi (Massachusetts Institute of Technology)

Long-Tailed Visual Recognition via Gaussian Clouded Logit Adjustment 6919
Mengke Li (Hong Kong Baptist University), Yiu-ming Cheung (Hong Kong Baptist University), and Yang Lu (Xiamen University)

Long-Tail Recognition via Compositional Knowledge Transfer 6929
Sarah Parisot (Huawei Noah's Ark Lab), Pedro M. Esperança (Huawei Noah's Ark Lab), Steven McDonagh (Huawei Noah's Ark Lab), Tamas J. Madarasz (Mediatek Research), Yongxin Yang (University of Edinburgh), and Zhenguo Li (Huawei Noah's Ark Lab)

Nested Collaborative Learning for Long-Tailed Visual Recognition 6939
Jun Li (NLPR, CASIA, China), Zichang Tan (Baidu Research), Jun Wan (NLPR, CASIA), Zhen Lei (NLPR, CASIA, China), and Guodong Guo (Baidu)

Retrieval Augmented Classification for Long-Tail Visual Recognition 6949
Alexander Long (Amazon), Wei Yin (University of Adelaide), Thalaiyasingam Ajanthan (ANU), Vu Nguyen (Amazon), Pulak Purkait (Amazon), Ravi Garg (Amazon), Alan Blair (University of New South Wales), Chunhua Shen (University of Adelaide, Australia), and Anton van den Hengel (University of Adelaide)

Trustworthy Long-Tailed Classification 6960
Bolian Li (Tianjin University), Zongbo Han (Tianjin University), Haining Li (Xidian University), Huazhu Fu (IHPC, ASTAR), and Changqing Zhang (Tianjin university)

C2AM Loss: Chasing a Better Decision Boundary for Long-Tail Object Detection 6970
Tong Wang (CASIA), Yousong Zhu (Institute of Automation, Chinese Academy of Sciences), Yingying Chen (CASIA), Chaoyang Zhao (National Laboratory of Pattern Recognition, CASIA), Bin Yu (Institute of Automation, Chinese Academy of Sciences), Jinqiao Wang (Institute of Automation, Chinese Academy of Sciences), and Ming Tang (Institute of Automation, Chinese Academy of Sciences)

Equalized Focal Loss for Dense Long-Tailed Object Detection 6980
Bo Li (Tongji University), Yongqiang Yao (sensetime), Jingru Tan (Tongji University), Gang Zhang (Tsinghua University), Fengwei Yu (Sensetime Research), Jianwei Lu (Tongji University), and Ye Luo (Tongji University)

Relieving Long-Tailed Instance Segmentation via Pairwise Class Balance 6990
Yin-Yin He (Nanjing University), Peizhen Zhang (Megvii Technology), Xiu-Shen Wei (Nanjing University of Science and Technology), Xiangyu Zhang (Megvii Technology), and Jian Sun (Megvii Technology)

iFS-RCNN: An Incremental Few-Shot Instance Segmenter 7000
Khoi Nguyen (VinAI Research) and Sinisa Todorovic (Oregon State U)

Open-Vocabulary Instance Segmentation via Robust Cross-Modal Pseudo-Labeling 7010
Dat Huynh (Northeastern University), Jason Kuen (Adobe Research), Zhe Lin (Adobe Research), Jiuxiang Gu (Adobe Research), and Ehsan Elhamifar (Northeastern University)

SimT: Handling Open-Set Noise for Domain Adaptive Semantic Segmentation 7022
Xiaoqing Guo (City University of Hong Kong), Jie Liu (City University of Hong Kong), Tongliang Liu (The University of Sydney), and Yixuan Yuan (City University of Hong Kong)

Undoing the Damage of Label Shift for Cross-Domain Semantic Segmentation 7032
Yahao Liu (University of Electronic Science and Technology of China), Jinhong Deng (University of Electronic Science and Technology of China), Jiale Tao (University of Electronic Science and Technology of China), Tong Chu (University of Electronic Science and Technology of China), Lixin Duan (University of Electronic Science and Technology of China), and Wen Li (University of Electronic Science and Technology of China)

Representation Compensation Networks for Continual Semantic Segmentation 7043
Chang-Bin Zhang (Nankai University), Jia-Wen Xiao (Nankai University), Xialei Liu (Nankai University), Ying-Cong Chen (Hong Kong University of Science and Technology), and Ming-Ming Cheng (Nankai University)

Remember the Difference: Cross-Domain Few-Shot Semantic Segmentation via Meta-Memory Transfer 7055
Wenjian Wang (Beijing University Of Technology), Lijuan Duan (Beijing University of Technology), Yuxi Wang (CASIA), Qing En (Carleton University), Junsong Fan (Chinese Academy of Sciences, China), and Zhaoxiang Zhang (Chinese Academy of Sciences, China)

Domain-Agnostic Prior for Transfer Semantic Segmentation 7065
Xinyue Huo (University of Science and Technology of China), Lingxi Xie (Huawei Inc.), Hengtong Hu (Hefei University of Technology), Wengang Zhou (University of Science and Technology of China), Houqiang Li (University of Science and Technology of China), and Qi Tian (Huawei Cloud & AI)

Image Segmentation Using Text and Image Prompts 7076
Timo Lüddecke (University of Göttingen) and Alexander Ecker (University of Göttingen)

PCL: Proxy-Based Contrastive Learning for Domain Generalization 7087
Xufeng Yao (Chinese University of HongKong), Yang Bai (CUHK), Xinyun Zhang (The Chinese University of Hong Kong), Yuechen Zhang (The Chinese University of Hong Kong), Qi Sun (The Chinese University of Hong Kong), Ran Chen (The Chinese University of Hong Kong), Ruiyu Li (Smartmore), and Bei Yu (CUHK)

Localized Adversarial Domain Generalization 7098
Wei Zhu (University of Rochester), Le Lu (Alibaba DAMO Academy USA), Jing Xiao (Ping An Insurance (Group) Company of China, ,), Mei Han (PAII Inc.), Jiebo Luo (U. Rochester), and Adam P. Harrison (PAII Inc.)

Compound Domain Generalization via Meta-Knowledge Encoding 7109
Chaoqi Chen (The University of Hong Kong), Jiongcheng Li (Xiamen University), Xiaoguang Han (Shenzhen Research Institute of Big Data, the Chinese University of Hong Kong (Shenzhen), ,), Xiaoqing Liu (DeepWise AI Lab), and Yizhou Yu (The University of Hong Kong)

Style Neophile: Constantly Seeking Novel Styles for Domain Generalization 7120
Juwon Kang (POSTECH), Sohyun Lee (POSTECH), Namyup Kim (POSTECH), and Suha Kwak (POSTECH)

Slimmable Domain Adaptation 7131
Rang Meng (Hikvision Research Institute), Weijie Chen (Zhejiang University), Shicai Yang (Hikvision Research Institute), Jie Song (Zhejiang University), Luojun Lin (Fuzhou University), Di Xie (Hikvision Research Institute), Shiliang Pu (Hikvision Research Institute), Xinchao Wang (National University of Singapore), Mingli Song (Zhengjiang University), and Yueting Zhuang (Zhejiang University)

Exploring Domain-Invariant Parameters for Source Free Domain Adaptation 7141
Fan Wang (Shandong University), Zhongyi Han (Shandong University), Yongshun Gong (Shandong University), and Yilong Yin (Shandong University)

Cross-Domain Few-Shot Learning With Task-Specific Adapters 7151
Wei-Hong Li (University of Edinburgh), Xialei Liu (Nankai University), and Hakan Bilen (University of Edinburgh)

Task-Adaptive Negative Envision for Few-Shot Open-Set Recognition 7161
Shiyuan Huang (Columbia University), Jiawei Ma (Columbia University), Guangxing Han (Columbia University), and Shih-Fu Chang (Columbia University)

Reusing the Task-Specific Classifier as a Discriminator: Discriminator-Free Adversarial Domain Adaptation 7171
Lin Chen (University of Science and Technology of China), Huaian Chen (University of Science and Technology of China), Zhixiang Wei (University of Science and Technology of China), Xin Jin (University of Science and Technology of China), Xiao Tan (University of Science and Technology of China), Yi Jin (University of Science and Technology of China), and Enhong Chen (University of Science and Technology of China)

Safe Self-Refinement for Transformer-Based Domain Adaptation 7181
Tao Sun (Stony Brook University), Cheng Lu (Xiaopeng), Tianshuo Zhang (Meta Platforms Inc.), and Haibin Ling (Stony Brook University)

Continual Test-Time Domain Adaptation 7191
Qin Wang (ETH Zurich), Olga Fink (ETH Zurich), Luc Van Gool (ETH Zurich), and Dengxin Dai (ETH Zurich)

Source-Free Domain Adaptation via Distribution Estimation 7202
Ning Ding (Peking University), Yixing Xu (Huawei Technologies), Yehui Tang (Peking University), Chao Xu (Peking University), Yunhe Wang (Huawei Technologies), and Dacheng Tao (JD.com)

Domain Adaptation on Point Clouds via Geometry-Aware Implicits 7213
Yuefan Shen (Zhejiang University), Yanchao Yang (Stanford University), Mi Yan (Peking University), He Wang (Peking University), Youyi Zheng (Zhejiang University), and Leonidas J. Guibas (Stanford University)

Deformation and Correspondence Aware Unsupervised Synthetic-to-Real Scene Flow Estimation for Point Clouds 7223
Zhao Jin (Sichuan University), Yinjie Lei (Sichuan University), Naveed Akhtar (The University of Western Australia), Haifeng Li (Central South University), and Munawar Hayat (Monash University)

Hyperspherical Consistency Regularization 7234
Cheng Tan (Westlake University), Zhangyang Gao (westlake university), Lirong Wu (Westlake University), Siyuan Li (Westlake University), and Stan Z. Li (Westlake University)

BatchFormer: Learning To Explore Sample Relationships for Robust Representation Learning 7246
Zhi Hou (The University of Sydney), Baosheng Yu (The University of Sydney), and Dacheng Tao (JD.com)

Poster 2.1: Recognition: Detection, Categorization, Retrieval

Cascade Transformers for End-to-End Person Search 7257
Rui Yu (The Pennsylvania State University), Dawei Du (Kitware, Inc.), Rodney LaLonde (Kitware, Inc.), Daniel Davila (Kitware Inc), Christopher Funk (Kitware, Inc), Anthony Hoogs (Kitware), and Brian Clipp (Kitware)

Delving Deep Into the Generalization of Vision Transformers Under Distribution Shifts 7267
Chongzhi Zhang (BUAA), Mingyuan Zhang (Beijing SenseTime Technology Development Limited), Shanghang Zhang (University of California, Berkeley), Daisheng Jin (Beijing SenseTime Technology Development Limited), Qiang Zhou (ShanDong University), Zhongang Cai (SenseTime International Pte Ltd), Haiyu Zhao (SenseTime International Pte Ltd), Xianglong Liu (BUAA), and Ziwei Liu (Nanyang Technological University)

MPViT: Multi-Path Vision Transformer for Dense Prediction 7277
Youngwan Lee (ETRI), Jonghee Kim (ETRI), Jeffrey Willette (KAIST), and Sung Ju Hwang (KAIST, AITRICS)

NFormer: Robust Person Re-Identification With Neighbor Transformer 7287
Haochen Wang (UvA), Jiayi Shen (University van Amsterdam), Yongtuo Liu (University of Amsterdam), Yan Gao (Xiaohongshu Inc), and Efstratios Gavves (University of Amsterdam)

Part-Based Pseudo Label Refinement for Unsupervised Person Re-Identification 7298
Yoonki Cho (KAIST), Woo Jae Kim (KAIST), Seunghoon Hong (KAIST), and Sung-Eui Yoon (KAIST)

Temporal Complementarity-Guided Reinforcement Learning for Image-to-Video Person Re-Identification 7309
Wei Wu (University of Science and Technology of China), Jiawei Liu (University of Science and Technology of China), Kecheng Zheng (University of Science and Technology of China), Qibin Sun (University of Science and Technology of China), and Zheng-Jun Zha (University of Science and Technology of China)

Augmented Geometric Distillation for Data-Free Incremental Person ReID 7319
Yichen Lu (Beijing University of Posts and Telecommunications), Mei Wang (Beijing University of Posts and Telecommunications), and Weihong Deng (Beijing University of Posts and Telecommunications)

Salient-to-Broad Transition for Video Person Re-Identification 7329
Shutao Bai (Institute of Computing Technology, Chinese Academy of Sciences), Bingpeng Ma (University of Chinese Academy of Sciences), Hong Chang (Chinese Academy of Sciences), Rui Huang (The Chinese University of Hong Kong, Shenzhen), and Xilin Chen (Institute of Computing Technology, Chinese Academy of Sciences)

FMCNet: Feature-Level Modality Compensation for Visible-Infrared Person Re-Identification 7339
Qiang Zhang (Xidian University), Changzhou Lai (Xidian University), Jianan Liu (XiDian University), Nianchang Huang (Xidian University), and Jungong Han (Aberystwyth University)

Graph Sampling Based Deep Metric Learning for Generalizable Person Re-Identification 7349
Shengcai Liao (Inception Institute of Artificial Intelligence) and Ling Shao (NCAI/IIAI/MBZUAI)

Implicit Sample Extension for Unsupervised Person Re-Identification 7359
Xinyu Zhang (Tongji University), Dongdong Li (University of Chinese Academy of Sciences), Zhigang Wang (Baidu Inc.), Jian Wang (Baidu), Errui Ding (Baidu Inc.), Javen Qinfeng Shi (University of Adelaide), Zhaoxiang Zhang (Chinese Academy of Sciences, China), and Jingdong Wang (Baidu)

Rethinking Reconstruction Autoencoder-Based Out-of-Distribution Detection 7369
Yibo Zhou (Institute of modern physics, Chinese Academy of Sciences)

Catching Both Gray and Black Swans: Open-Set Supervised Anomaly Detection 7378
Choubo Ding (University of Adelaide), Guansong Pang (Singapore Management University), and Chunhua Shen (University of Adelaide, Australia)

Fine-Grained Object Classification via Self-Supervised Pose Alignment 7389
Xuhui Yang (Peng Cheng Laboratory), Yaowei Wang (PengCheng Laboratory), Ke Chen (South China University of Technology), Yong Xu (South China University of Technology), and Yonghong Tian (Peng Cheng Laboratory)

Hyperbolic Vision Transformers: Combining Improvements in Metric Learning 7399
Aleksandr Ermolov (University of Trento), Leyla Mirvakhabova (Skolkovo Institute of Science and Technology), Valentin Khrulkov (Yandex), Nicu Sebe (University of Trento), and Ivan Oseledets (Skolkovo Institute of Science and Technology)

Non-Isotropy Regularization for Proxy-Based Deep Metric Learning 7410
Karsten Roth (University of Tuebingen), Oriol Vinyals (DeepMind), and Zeynep Akata (University of Tübingen)

Self-Taught Metric Learning Without Labels 7421
Sungyeon Kim (POSTECH), Dongwon Kim (POSTECH), Minsu Cho (POSTECH), and Suha Kwak (POSTECH)

Not Just Selection, but Exploration: Online Class-Incremental Continual Learning via Dual View Consistency 7432
Yanan Gu (Xidian University), Xu Yang (Xidian University), Kun Wei (Xidian University), and Cheng Deng (Xidian University)

Energy-Based Latent Aligner for Incremental Learning 7442
K J Joseph (Indian Institute of Technology, Hyderabad), Salman Khan (MBZUAI/ANU), Fahad Shahbaz Khan (MBZUAI), Rao Muhammad Anwer (MBZUAI/AALTO), and Vineeth N Balasubramanian (Indian Institute of Technology, Hyderabad)

Sketch3T: Test-Time Training for Zero-Shot SBIR 7452
Aneeshan Sain (University of Surrey), Ayan Kumar Bhunia (University of Surrey), Vaishnav Potlapalli (Independent), Pinaki Nath Chowdhury (University of Surrey), Tao Xiang (University of Surrey), and Yi-Zhe Song (University of Surrey)

The Devil Is in the Pose: Ambiguity-Free 3D Rotation-Invariant Learning via Pose-Aware Convolution 7462
Ronghan Chen (Shenyang Institute of Automation, Chinese Academy of Sciences) and Yang Cong (Chinese Academy of Sciences)

Finding Badly Drawn Bunnies 7472
Lan Yang (Beijing University of Posts and Telecommunications), Kaiyue Pang (University of Surrey), Honggang Zhang (Beijing University of Posts and Telecommunications), and Yi-Zhe Song (University of Surrey)

Generalized Category Discovery 7482
Sagar Vaze (Visual Geometry Group, University of Oxford), Kai Han (The University of Hong Kong), Andrea Vedaldi (Oxford University), and Andrew Zisserman (University of Oxford)

Recall@k Surrogate Loss With Large Batches and Similarity Mixup 7492
Yash Patel (Czech Technical University in Prague), Giorgos Tolias (Czech Technical University in Prague, Faculty of Electrical Engineering, Visual Recognition Group), and Jiří Matas (Czech Technical University, Prague)

Modeling 3D Layout for Group Re-Identification 7502
Quan Zhang (Sun Yat-sen University), Kaiheng Dang (Sun Yat-sen University), Jian-Huang Lai (Sun Yat-sen University), Zhanxiang Feng (Sun Yat-sen University), and Xiaohua Xie (Sun Yat-Sen University)

Causal Transportability for Visual Recognition 7511
Chengzhi Mao (Columbia University), Kevin Xia (Columbia University), James Wang (Columbia University), Hao Wang (Rutgers University), Junfeng Yang (Columbia University), Elias Bareinboim (Columbia University), and Carl Vondrick (Columbia University)

Attributable Visual Similarity Learning 7522
Borui Zhang (Tsinghua University), Wenzhao Zheng (Tsinghua University), Jie Zhou (Tsinghua University), and Jiwen Lu (Tsinghua University)

Bi-Level Alignment for Cross-Domain Crowd Counting 7532
Shenjian Gong (Nanjing University of Science and Technology), Shanshan Zhang (Nanjing University of Science and Technology), Jian Yang (Nanjing University of Science and Technology), Dengxin Dai (MPI for Informatics), and Bernt Schiele (MPI Informatics)

Mutual Quantization for Cross-Modal Search With Noisy Labels 7541
Erkun Yang (Xidian University), Dongren Yao (Harvard Medical School), Tongliang Liu (The University of Sydney), and Cheng Deng (Xidian University)

Task Adaptive Parameter Sharing for Multi-Task Learning 7551
Matthew Wallingford (University of Washington), Hao Li (Amazon Web Services), Alessandro Achille (Amazon Web Services), Avinash Ravichandran (Amazon), Charless Fowlkes (UC Irvine), Rahul Bhotika (Amazon), and Stefano Soatto (UCLA)

Simple Multi-Dataset Detection 7561
Xingyi Zhou (The University of Texas at Austin), Vladlen Koltun (Apple), and Philipp Krähenbühl (UT Austin)

Cross-Domain Adaptive Teacher for Object Detection 7571
Yu-Jhe Li (Carnegie Mellon University), Xiaoliang Dai (Facebook), Chih-Yao Ma (Facebook), Yen-Cheng Liu (Georgia Institute of Technology), Kan Chen (University of Southern California), Bichen Wu (Facebook Research), Zijian He (Facebook), Kris Kitani (Carnegie Mellon University), and Peter Vajda (Facebook)

Balanced and Hierarchical Relation Learning for One-Shot Object Detection 7581
Hanqing Yang (ZheJiang University), Sijia Cai (DAMO Academy, Alibaba Group), Hualian Sheng (College of Information Science and Electronic Engineering, Zhejiang University; DAMO Academy, Alibaba Group), Bing Deng (Damo Academy, Alibaba Group), Jianqiang Huang (Damo Academy, Alibaba Group), Xian-Sheng Hua (Damo Academy, Alibaba Group), Yong Tang (SHUDAO INVESTMENT GROUP CO., LTD.), and Yu Zhang (ZheJiang University)

Semantic-Aligned Fusion Transformer for One-Shot Object Detection 7591
Yizhou Zhao (Carnegie Mellon University), Xun Guo (Microsoft Research Asia), and Yan Lu (Microsoft Research Asia)

MSDN: Mutually Semantic Distillation Network for Zero-Shot Learning 7602
Shiming Chen (Huazhong University of Science and Technology), Ziming Hong (Huazhong University of Science and Technology), Guo-Sen Xie (Inception Institute of Artificial Intelligence), Wenhan Yang (NTU), Qinmu Peng (Huazhong University of Science and Technology), Kai Wang (National University of Singapore), Jian Zhao (Institute of North Electronic Equipment), and Xinge You (Huazhong University of Science and Technology)

Robust Region Feature Synthesizer for Zero-Shot Object Detection 7612
Peiliang Huang (Northwestern Polytechnical University), Junwei Han (NWPU, China), De Cheng (Xidian University), and Dingwen Zhang (NWPU)

Poster 2.1: Image & Video Synthesis and Generation

Region-Aware Face Swapping 7622
Chao Xu (Zhejiang University), Jiangning Zhang (Zhejiang University), Miao Hua (Bytedance), Qian He (Bytedance), Zili Yi (ByteDance), and Yong Liu (Zhejiang University)

High-Resolution Face Swapping via Latent Semantics Disentanglement 7632
Yangyang Xu (South China University of Technology), Bailin Deng (Cardiff University), Junle Wang (Tencent), Yanqing Jing (Tencent), Jia Pan (The University of Hong Kong), and Shengfeng He (South China University of Technology)

Rethinking Deep Face Restoration 7642
Yang Zhao (University at Buffalo), Yu-Chuan Su (Google), Chun-Te Chu (Google), Yandong Li (University of Central Florida), Marius Renn (Google LLC), Yukun Zhu (Google Inc.), Changyou Chen (University at Buffalo), and Xuhui Jia (google)

Blind Face Restoration via Integrating Face Shape and Generative Priors 7652
Feida Zhu (Tencent), Junwei Zhu (Tencent), Wenqing Chu (Tencent), Xinyi Zhang (Tencent Youtu), Xiaozhong Ji (Tencent), Chengjie Wang (Tencent), and Ying Tai (Tencent YouTu)

FENeRF: Face Editing in Neural Radiance Fields 7662
Jingxiang Sun (University of Illinois Urbana-Champaign), Xuan Wang (Tencent AI Lab), Yong Zhang (Tencent AI Lab), Xiaoyu Li (Tencent AI Lab), Qi Zhang (Tencent AI Lab), Yebin Liu (Tsinghua University), and Jue Wang (Tencent AI Lab)

TransEditor: Transformer-Based Dual-Space GAN for Highly Controllable Facial Editing 7673
Yanbo Xu (Shanghai AI Lab), Yueqin Yin (), Liming Jiang (Nanyang Technological University), Qianyi Wu (Monash University), Chengyao Zheng (SenseTime Research), Chen Change Loy (Nanyang Technological University), Bo Dai (Shanghai AI Lab), and Wayne Wu (SenseTime Research)

Pastiche Master: Exemplar-Based High-Resolution Portrait Style Transfer 7683
Shuai Yang (Nanyang Technological University), Liming Jiang (Nanyang Technological University), Ziwei Liu (Nanyang Technological University), and Chen Change Loy (Nanyang Technological University)

Self-Supervised Correlation Mining Network for Person Image Generation 7693
Zijian Wang (Beijing university of Posts and Telecommunications), Xingqun Qi (Beijing University of Posts and Telecommunications), Kun Yuan (Kuaishou Technology), and Muyi Sun (CRIPAC, Institute of Automation, Chinese Academy of Sciences)

Exploring Dual-Task Correlation for Pose Guided Person Image Generation 7703
Pengze Zhang (Sun Yat-sen University), Lingxiao Yang (Sun-Yat Sen University), Jian-Huang Lai (Sun Yat-sen University), and Xiaohua Xie (Sun Yat-sen University)

InsetGAN for Full-Body Image Generation 7713
Anna Frühstück (KAUST), Krishna Kumar Singh (Adobe Research), Eli Shechtman (Adobe Research, US), Niloy J. Mitra (University College London), Peter Wonka (KAUST), and Jingwan Lu (Adobe Research)

BodyGAN: General-Purpose Controllable Neural Human Body Generation 7723
Chaojie Yang (Beijing Momo Technology Co., Ltd.), Hanhui Li (Sun Yat-sen University), Shengjie Wu (Beijing Momo Technology Co., Ltd.), Shengkai Zhang (Beijing Momo Technology Co., Ltd.), Haonan Yan (Beijing Momo Technology Co., Ltd.), Nianhong Jiao (Beijing Momo Technology Co., Ltd.), Jie Tang (Beijing Momo Technology Co., Ltd.), Runnan Zhou (Beijing Momo Technology Co., Ltd.), Xiaodan Liang (Sun Yat-sen University), and Tianxiang Zheng (Beijing Momo Technology Co., Ltd.)

HumanNeRF: Efficiently Generated Human Radiance Field From Sparse Inputs 7733
Fuqiang Zhao (ShanghaiTech University), Wei Yang (Huazhong University of Science and Technology), Jiakai Zhang (ShanghaiTech University), Pei Lin (ShanghaiTech University), Yingliang Zhang (DGene), Jingyi Yu (Shanghai Tech University), and Lan Xu (ShanghaiTech)

Structure-Aware Flow Generation for Human Body Reshaping 7744
Jianqiang Ren (Alibaba), Yuan Yao (Alibaba Group), Biwen Lei (Alibaba), Miaomiao Cui (Alibaba-inc), and Xuansong Xie (Alibaba)

Modeling Image Composition for Complex Scene Generation 7754
Zuopeng Yang (Shanghai JiaoTong University), Daqing Liu (JD.com Inc), Chaoyue Wang (JD.com), Jie Yang (Shanghai Jiao Tong University), and Dacheng Tao (JD.com)

Local Attention Pyramid for Scene Image Generation 7764
Sang-Heon Shim (Sungkyunkwan University), Sangeek Hyun (Sungkyunkwan University), DaeHyun Bae (Sungkyunkwan University), and Jae-Pil Heo (Sungkyunkwan University)

Interactive Image Synthesis With Panoptic Layout Generation 7773
Bo Wang (Huawei Technologies), Tao Wu (Huawei Technologies Duesseldorf GmbH), Minfeng Zhu (Huawei Technologies), and Peng Du (Huawei Technologies)

iPLAN: Interactive and Procedural Layout Planning 7783
Feixiang He (University of Leeds), Yanlong Huang (University of Leeds), and He Wang (Leeds University)

E-CIR: Event-Enhanced Continuous Intensity Recovery 7793
Chen Song (The University of Texas at Austin), Qixing Huang (The University of Texas at Austin), and Chandrajit Bajaj (University of Texas, Austin)

Learning Robust Image-Based Rendering on Sparse Scene Geometry via Depth Completion 7803
Yuqi Sun (Fudan University), Shili Zhou (Fudan University), Ri Cheng (Fudan University), Weimin Tan (Fudan University), Bo Yan (Fudan University), and Lang Fu (Fudan University)

Neural Rays for Occlusion-Aware Image-Based Rendering 7814
Yuan Liu (The University of Hong Kong), Sida Peng (Zhejiang University), Lingjie Liu (Max Planck Institute for Informatics), Qianqian Wang (Cornell), Peng Wang (The University of Hong Kong), Christian Theobalt (MPI Informatik), Xiaowei Zhou (Zhejiang Univ., China), and Wenping Wang (The University of Hong Kong)

Industrial Style Transfer With Large-Scale Geometric Warping and Content Preservation 7824
Jinchao Yang (Nanjing University of Science and Technology), Fei Guo (Nanjing University of Science and Technology), Shuo Chen (RIKEN), Jun Li (Nanjing University of Science and Technology), and Jian Yang (Nanjing University of Science and Technology)

PCA-Based Knowledge Distillation Towards Lightweight and Content-Style Balanced Photorealistic Style Transfer Models 7834
Tai-Yin Chiu (University of Texas) and Danna Gurari (University of Colorado Boulder)

Commonality in Natural Images Rescues GANs: Pretraining GANs With Generic and Privacy-Free Synthetic Data 7844
Kyungjune Baek (Yonsei University) and Hyunjung Shim (Yonsei University)

Think Twice Before Detecting GAN-Generated Fake Images From Their Spectral Domain Imprints..... 7855
Chengdong Dong (Hongkong Polytechnic University), Ajay Kumar (The Hong Kong Polytechnic University), and Eryun Liu (Zhejiang University)

Robust Invertible Image Steganography 7865
Youmin Xu (Peking University), Chong Mou (Peking University Shenzhen Graduate School), Yujie Hu (Peking University Shenzhen Graduate School), Jingfen Xie (Peking University), and Jian Zhang (Peking University Shenzhen Graduate School)

Distinguishing Unseen From Seen for Generalized Zero-Shot Learning 7875
Hongzu Su (University of Electronic Science and Technology of China), Jingjing Li (University of Electronic Science and Technology of China), Zhi Chen (The University of Queensland), Lei Zhu (Shandong Normal Unversity), and Ke Lu (University of Electronic Science and Technology of China)

Few-Shot Font Generation by Learning Fine-Grained Local Styles 7885
Licheng Tang (Baidu), Yiyang Cai (University of California, Berkeley), Jiaming Liu (Baidu Inc.), Zhibin Hong (Baidu Inc.), Mingming Gong (University of Melbourne), Minhu Fan (Baidu China Co., Ltd.), Junyu Han (Baidu Inc.), Jingtuo Liu (baidu), Errui Ding (Baidu Inc.), and Jingdong Wang (Baidu)

XMP-Font: Self-Supervised Cross-Modality Pre-Training for Few-Shot Font Generation 7895
Wei Liu (Bytedance), Fangyue Liu (ByteDance), Fei Ding (Bytedance), Qian He (Bytedance), and Zili Yi (ByteDance)

Learning To Generate Line Drawings That Convey Geometry and Semantics 7905
Caroline Chan (MIT), Frédo Durand (MIT), and Phillip Isola (MIT)

Oral 2.2.1: Transfer / Low-Shot / Long-Tail Learning

Balanced MSE for Imbalanced Visual Regression 7916
Jiawei Ren (Nanyang Technological University), Mingyuan Zhang (Nanyang Technological University), Cunjun Yu (National University of Singapore), and Ziwei Liu (Nanyang Technological University)

Transferability Metrics for Selecting Source Model Ensembles 7926
Andrea Agostinelli (Google), Jasper Uijlings (Google Research), Thomas Mensink (Google Research / University of Amsterdam), and Vittorio Ferrari (Google Research)

OoD-Bench: Quantifying and Understanding Two Dimensions of Out-of-Distribution Generalization 7937
Nanyang Ye (Shanghai Jiao Tong University), Kaican Li (Huawei Noah's Ark Lab), Haoyue Bai (The Hong Kong University of Science and Technology), Runpeng Yu (Tsinghua University), Lanqing Hong (Huawei Noah's Ark Lab), Fengwei Zhou (Huawei Noah's Ark Lab), Zhenguo Li (Huawei Noah's Ark Lab), and Jun Zhu (Tsinghua University)

Robust Fine-Tuning of Zero-Shot Models 7949
Mitchell Wortsman (University of Washington), Gabriel Ilharco (University of Washington), Jong Wook Kim (OpenAI), Mike Li (Columbia University), Simon Kornblith (Google Brain), Rebecca Roelofs (Google), Raphael Gontijo Lopes (Google Brain), Hannaneh Hajishirzi, Ali Farhadi (University of Washington, Allen Institue for AI, Apple), Hongseok Namkoong (Ass Prof Columbia), and Ludwig Schmidt (Toyota Research Institute)

Joint Distribution Matters: Deep Brownian Distance Covariance for Few-Shot Classification 7962
Jiangtao Xie (Dalian University of Technology), Fei Long (Dalian University of Technology), Jiaming Lv (Dalian University of Technology), Qilong Wang (Tianjin University), and Peihua Li (Dalian University of Technology)

Learning To Learn and Remember Super Long Multi-Domain Task Sequence 7972
Zhenyi Wang (University at Buffalo), Li Shen (JD Explore Academy), Tiehang Duan (Facebook), Donglin Zhan (Columbia University), Le Fang (University at Buffalo), and Mingchen Gao (University at Buffalo, SUNY)

Learning Distinctive Margin Toward Active Domain Adaptation 7983
Ming Xie (Fudan University), Yuxi Li (Tencent), Yabiao Wang (Tencent), Zekun Luo (Tencent), Zhenye Gan (Tencent), Zhongyi Sun (Tencent), Mingmin Chi (Fudan university), Chengjie Wang (Tencent), and Pei Wang (CAS, NAOC)

DINE: Domain Adaptation From Single and Multiple Black-Box Predictors 7993
Jian Liang (NUS), Dapeng Hu (National University of Singapore), Jiashi Feng (ByteDance), and Ran He (Institute of Automation, Chinese Academy of Sciences)

Source-Free Object Detection by Learning To Overlook Domain Style 8004
Shuaifeng Li (University of Electronic Science and Technology of China), Mao Ye (University of Electronic Science and Technology of China), Xiatian Zhu (Samsung), Lihua Zhou (University of Electronic Science and Technology of China), and Lin Xiong (University of Electronic Science and Technology of China)

Towards Principled Disentanglement for Domain Generalization 8014
Hanlin Zhang (Carnegie Mellon University), Yi-Fan Zhang (NLPR, China), Weiyang Liu (University of Cambridge), Adrian Weller (University of Cambridge), Bernhard Schölkopf (MPI for Intelligent Systems, Tübingen), and Eric P. Xing (MBZUAI, CMU, and Petuum Inc.)

Exact Feature Distribution Matching for Arbitrary Style Transfer and Domain Generalization..... 8025
Yabin Zhang (Hong Kong Polytechnic University), Minghan Li (The Hong Kong Polytechnic University), Ruihuang Li (Hong Kong Polytechnic University), Kui Jia (South China University of Technology), and Lei Zhang (Hong Kong Polytechnic University, Hong Kong, China)

Causality Inspired Representation Learning for Domain Generalization 8036
Fangrui Lv (Beijing Institute of Technology), Jian Liang (Alibaba Group), Shuang Li (Beijing Institute of Technology), Bin Zang (Beijing Institute of Technology), Chi Harold Liu (Beijing Institute of Technology), Ziteng Wang (Yizhun-ai), and Di Liu (Alibaba Group)

Learning What Not To Segment: A New Perspective on Few-Shot Segmentation 8047
Chunbo Lang (Northwestern Polytechnical University), Gong Cheng (Northwestern Polytechnical University), Binfei Tu (Northwestern Polytechnical University), and Junwei Han (NWPU, China)

Towards Fewer Annotations: Active Learning via Region Impurity and Prediction Uncertainty for Domain Adaptive Semantic Segmentation 8058
Binhui Xie (Beijing Institute of Technology), Longhui Yuan (Beijing Institute of Technology), Shuang Li (Beijing Institute of Technology), Chi Harold Liu (Beijing Institute of Technology), and Xinjing Cheng (Inceptio Tech.)

ADeLA: Automatic Dense Labeling With Attention for Viewpoint Shift in Semantic Segmentation 8069
Hanxiang Ren (Stanford University), Yanchao Yang (Zhejiang University), He Wang (Peking University), Bokui Shen (Stanford), Qingnan Fan (Tencent AI Lab), Youyi Zheng (Zhejiang University), C. Karen Liu (Stanford), and Leonidas J. Guibas (Stanford University)

Oral 2.2.2: Motion, Tracking, Registration, Vision & X, and Theory

MeMOT: Multi-Object Tracking With Memory 8080
Jiarui Cai (University of Washington), Mingze Xu (Amazon), Wei Li (Amazon), Yuanjun Xiong (Amazon), Wei Xia (Amazon), Zhuowen Tu (UC San Diego), and Stefano Soatto (AWS Amazon ML)

Unsupervised Learning of Accurate Siamese Tracking 8091
Qiuhong Shen (Harbin Institute of Technology (Shenzhen)), Lei Qiao (SenseTime Group Limited), Jinyang Guo (The University of Sydney), Peixia Li (The University of Sydney), Xin Li (Peng Cheng Laboratory), Bo Li (SenseTime Group Limited), Weitao Feng (University of Sydney), Weihao Gan (SenseTime Group Limited), Wei Wu (SenseTime Group Limited), and Wanli Ouyang (The University of Sydney)

Beyond 3D Siamese Tracking: A Motion-Centric Paradigm for 3D Single Object Tracking in Point Clouds 8101
Chaoda Zheng (The Chinese University of Hong Kong, Shen Zhen), Xu Yan (The Chinese University of Hong Kong, Shenzhen), Haiming Zhang (The Chinese University of Hong Kong, Shenzhen), Baoyuan Wang (Xiaobing.ai), Shenghui Cheng (Westlake University), Shuguang Cui (The Chinese University of Hong Kong, Shenzhen), and Zhen Li (The Chinese University of Hong Kong, Shenzhen)

GMFlow: Learning Optical Flow via Global Matching 8111
Haofei Xu (Monash University), Jing Zhang (The University of Sydney), Jianfei Cai (Monash University), Hamid Rezatofighi (Monash University), and Dacheng Tao (JD.com)

GridShift: A Faster Mode-Seeking Algorithm for Image Segmentation and Object Tracking 8121
Abhishek Kumar (Kyungpook National University), Oladayo S. Ajani (Kyungpook National University), Swagatam Das (Kyungpook national University), and Rammohan Mallipeddi (Indian Statistical Institute)

SNUG: Self-Supervised Neural Dynamic Garments 8130
Igor Santesteban (Universidad Rey Juan Carlos), Miguel A. Otaduy (Universidad Rey Juan Carlos), and Dan Casas (Universidad Rey Juan Carlos)

Weakly-Supervised Action Transition Learning for Stochastic Human Motion Prediction 8141
Wei Mao (the Australian National University), Miaomiao Liu (The Australian National University), and Mathieu Salzmann (EPFL)

Multi-Objective Diverse Human Motion Prediction With Knowledge Distillation 8151
Hengbo Ma (University of California, Berkeley), Jiachen Li (Stanford University), Ramtin Hosseini (Tufts University), Masayoshi Tomizuka (MSC Lab), and Chiho Choi (Honda Research Institute US)

Context-Aware Sequence Alignment Using 4D Skeletal Augmentation 8162
Taein Kwon (ETH Zurich), Bugra Tekin (Microsoft), Siyu Tang (ETH Zurich), and Marc Pollefeys (ETH Zurich / Microsoft)

Enabling Equivariance for Arbitrary Lie Groups 8173
Lachlan E. MacDonald (Australian Institute for Machine Learning), Sameera Ramasinghe (University of Adelaide), and Simon Lucey (University of Adelaide)

RAMA: A Rapid Multicut Algorithm on GPU 8183
Ahmed Abbas (MPI-INF) and Paul Swoboda (MPI fuer Informatik, Saarbruecken)

Self-Supervised Material and Texture Representation Learning for Remote Sensing Tasks 8193
Peri Akiva (Rutgers University), Matthew Purri (Rutgers University), and Matthew Leotta (Kitware)

RCP: Recurrent Closest Point for Point Cloud 8206
Xiaodong Gu (Alibaba A.I. Labs), Chengzhou Tang (Simon Fraser University), Weihao Yuan (Alibaba A.I. Labs), Zuozhuo Dai (Alibaba A.I. Labs), Siyu Zhu (Alibaba A.I. Labs), and Ping Tan (Simon Fraser University)

Audio-Visual Speech Codecs: Rethinking Audio-Visual Speech Enhancement by Re-Synthesis 8217
Karren Yang (MIT), Dejan Marković (Facebook Reality Labs), Steven Krenn (Meta), Vasu Agrawal, and Alexander Richard (Facebook Reality Labs)

Balanced Multimodal Learning via On-the-Fly Gradient Modulation 8228
Xiaokang Peng (Renmin University of China), Yake Wei (Renmin University of China), Andong Deng (Shanghai Jiao Tong University), Dong Wang (School of Electrical Engineering and Computer Science, Oregon State University), and Di Hu (Renmin University of China)

Oral 2.2.3: 3D from Multiview & Sensors, Learning for Vision, Explainable Vision, and Privacy

Block-NeRF: Scalable Large Scene Neural View Synthesis 8238
Matthew Tancik (UC Berkeley), Vincent Casser (Waymo), Xinchen Yan (Waymo), Sabeek Pradhan (Waymo), Ben Mildenhall (Google Research), Pratul P. Srinivasan (Google Research), Jonathan T. Barron (Google Research), and Henrik Kretzschmar (Waymo)

SceneSqueezer: Learning To Compress Scene for Camera Relocalization 8249
Luwei Yang (Simon Fraser University), Rakesh Shrestha (Simon Fraser University), Wenbo Li (Zhejiang University), Shuaicheng Liu (UESTC; Megvii), Guofeng Zhang (Zhejiang University), Zhaopeng Cui (Zhejiang University), and Ping Tan (Simon Fraser University)

Light Field Neural Rendering 8259
Mohammed Suhail (University of British Columbia), Carlos Esteves (Google Research), Leonid Sigal (University of British Columbia), and Ameesh Makadia (Google Research)

Extracting Triangular 3D Models, Materials, and Lighting From Images 8270
Jacob Munkberg (NVIDIA), Jon Hasselgren (NVIDIA), Tianchang Shen (NVIDIA Development, Inc.), Jun Gao (NVIDIA, University of Toronto), Wenzheng Chen (University of Toronto), Alex Evans (NVIDIA), Thomas Müller (NVIDIA), and Sanja Fidler (University of Toronto, NVIDIA)

Super-Fibonacci Spirals: Fast, Low-Discrepancy Sampling of SO(3) 8281
Marc Alexa (TUB)

Stochastic Backpropagation: A Memory Efficient Strategy for Training Video Models 8291
Feng Cheng (University of North Carolina ch), Mingze Xu (Amazon), Yuanjun Xiong (Amazon), Hao Chen (Amazon), Xinyu Li (ByteDance), Wei Li (Amazon), and Wei Xia (Amazon)

It's All in the Teacher: Zero-Shot Quantization Brought Closer to the Teacher 8301
Kanghyun Choi (Yonsei University), Hye Yoon Lee (Yonsei University), Deokki Hong (Yonsei University), Joonsang Yu (NAVER CLOVA), Noseong Park (Yonsei University, Korea), Youngsok Kim (Yonsei University), and Jinho Lee (Yonsei University)

NLX-GPT: A Model for Natural Language Explanations in Vision and Vision-Language Tasks 8312
Fawaz Sammani (Vrije Universiteit Brussel), Tanmoy Mukherjee (VUB), and Nikos Deligiannis (Vrije Universiteit Brussel)

Explaining Deep Convolutional Neural Networks via Latent Visual-Semantic Filter Attention 8323
Yu Yang (University of California, Los Angeles), Seungbae Kim (University of California, Los Angeles), and Jungseock Joo (University of California Los Angeles)

Parameter-Free Online Test-Time Adaptation 8334
Malik Boudiaf (ÉTS Montreal), Romain Mueller (Five.ai), Ismail Ben Ayed (ETS Montreal), and Luca Bertinetto (FiveAI Ltd.)

Patch-Level Representation Learning for Self-Supervised Vision Transformers 8344
Sukmin Yun (KAIST), Hankook Lee (KAIST), Jaehyung Kim (KAIST), and Jinwoo Shin (KAIST)

Deep Spectral Methods: A Surprisingly Strong Baseline for Unsupervised Semantic Segmentation and Localization 8354
Luke Melas-Kyriazi (Oxford University), Christian Rupprecht (University of Oxford), Iro Laina (University of Oxford), and Andrea Vedaldi (Oxford University)

Mixed Differential Privacy in Computer Vision 8366
Aditya Golatkar (University of California, Los Angeles), Alessandro Achille (Amazon Web Services), Yu-Xiang Wang (UC Santa Barbara), Aaron Roth (University of Pennsylvania), Michael Kearns (University of Pennsylvania), and Stefano Soatto (AWS Amazon ML)

DPGEN: Differentially Private Generative Energy-Guided Network for Natural Image Synthesis. 8377
Jia-Wei Chen (Academia Sinica), Chia-Mu Yu (National Chiao Tung University), Ching-Chia Kao (Academia Sinica), Tzai-Wei Pang (Academia Sinica), and Chun-Shien Lu (Academia Sinica)

Local Learning Matters: Rethinking Data Heterogeneity in Federated Learning 8387
Matias Mendieta (University of Central Florida), Taojiannan Yang (University of Central Florida), Pu Wang (UNCC), Minwoo Lee (University of North Carolina at Charlotte), Zhengming Ding (Tulane University), and Chen Chen (University of Central Florida)

Poster 2.2: 3D From Multi-View & Sensors

AirObject: A Temporally Evolving Graph Embedding for Object Identification 8397
Nikhil Varma Keetha (Indian Institute of Technology (Indian School of Mines) Dhanbad, ,), Chen Wang (Carnegie Mellon University), Yuheng Qiu (Carnegie Mellon University), Kuan Xu (geekplus), and Sebastian Scherer (Carnegie Mellon University)

Voxel Set Transformer: A Set-to-Set Approach to 3D Object Detection from Point Clouds 8407
Chenhang He (The Hong Kong Polytechnic University), Ruihuang Li (Hong Kong Polytechnic University), Shuai Li (The Hong Kong Polytechnic University), and Lei Zhang (Hong Kong Polytechnic University, Hong Kong, China)

SS3D: Sparsely-Supervised 3D Object Detection From Point Cloud 8418
Chuandong Liu (School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications), Chenqiang Gao (Chongqing University of Posts and Telecommunications), Fangcen Liu (Chongqing University of Posts and Telecommunications), Jiang Liu (Facebook Inc.), Deyu Meng (Xi'an Jiaotong University), and Xinbo Gao (Chongqing University of Posts and Telecommunications)

Back to Reality: Weakly-Supervised 3D Object Detection With Shape-Guided Label Enhancement 8428
Xiuwei Xu (Tsinghua University), Yifan Wang (Tsinghua University), Yu Zheng (Tsinghua University), Yongming Rao (Tsinghua University), Jie Zhou (Tsinghua University), and Jiwen Lu (Tsinghua University)

VISTA: Boosting 3D Object Detection via Dual Cross-VIew SpaTial Attention 8438
Shengheng Deng (South China University of Technology), Zhihao Liang (South China University of Technology), Lin Sun (Samsung, USA), and Kui Jia (South China University of Technology)

Embracing Single Stride 3D Object Detector With Sparse Transformer 8448
Lue Fan (Institute of Automation, Chinese Academy of Sciences), Ziqi Pang (UIUC), Tianyuan Zhang (Carnegie Mellon University), Yu-Xiong Wang (Tsinghua University), Hang Zhao (University of Illinois at Urbana-Champaign), Feng Wang (TuSimple), Naiyan Wang (TuSimple), and Zhaoxiang Zhang (Chinese Academy of Sciences, China)

Point Density-Aware Voxels for LiDAR 3D Object Detection 8459
Jordan S. K. Hu (University of Toronto), Tianshu Kuai (University of Toronto), and Steven L. Waslander (University of Toronto)

Point-to-Voxel Knowledge Distillation for LiDAR Semantic Segmentation 8469
Yuenan Hou (Shanghai AI Lab), Xinge Zhu (The Chinese University of Hong Kong), Yuexin Ma (ShanghaiTech University), Chen Change Loy (Nanyang Technological University), and Yikang Li (Shanghai AI Lab)

Contrastive Boundary Learning for Point Cloud Segmentation 8479
Liyao Tang (University of Sydney), Yibing Zhan (Jingdong), Zhe Chen (The University of Sydney), Baosheng Yu (The University of Sydney), and Dacheng Tao (JD.com)

Stratified Transformer for 3D Point Cloud Segmentation 8490
Xin Lai (The Chinese University of Hong Kong), Jianhui Liu (The University of Hong Kong), Li Jiang (Max Planck Institute for Informatics), Liwei Wang (CUHK), Hengshuang Zhao (University of Oxford), Shu Liu (SmartMore), Xiaojuan Qi (The University of Hong Kong), and Jiaya Jia (Chinese University of Hong Kong)

No Pain, Big Gain: Classify Dynamic Point Cloud Sequences With Static Models by Fitting Feature-Level Space-Time Surfaces 8500
Jia-Xing Zhong (Department of Computer Science, University of Oxford), Kaichen Zhou (University of Oxford), Qingyong Hu (University of Oxford), Bing Wang (University of Oxford), Niki Trigoni (University of Oxford), and Andrew Markham (University of Oxford)

Point2Seq: Detecting 3D Objects As Sequences 8511
Yujing Xue (National University of Singapore), Jiageng Mao (The Chinese University of Hong Kong), Minzhe Niu (Huawei Noah's Ark Lab), Hang Xu (Huawei Noah's Ark Lab), Michael Bi Mi (Huawei International Pte Ltd), Wei Zhang (Noah's Ark Lab, Huawei Technologies), Xiaogang Wang (Chinese University of Hong Kong, Hong Kong), and Xinchao Wang (National University of Singapore)

PTTR: Relational 3D Point Cloud Object Tracking With Transformer 8521
Changqing Zhou (Nanyang Technological University), Zhipeng Luo (Nanyang Technological University), Yueru Luo (Nanyang Technological University), Tianrui Liu (Nanyang Technological University), Liang Pan (Nanyang Technological University), Zhongang Cai (SenseTime International Pte Ltd), Haiyu Zhao (SenseTime International Pte Ltd), and Shijian Lu (Nanyang Technological University)

A Unified Query-Based Paradigm for Point Cloud Understanding 8531
Zetong Yang (The Chinese University of Hong Kong), Li Jiang (Max Planck Institute for Informatics), Yanan Sun (HKUST), Bernt Schiele (MPI Informatics), and Jiaya Jia (Chinese University of Hong Kong)

PointCLIP: Point Cloud Understanding by CLIP 8542
Renrui Zhang (Shanghai AI Lab), Ziyu Guo (Peking University), Wei Zhang (Shanghai AI-Lab), Kunchang Li (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences), Xupeng Miao (Peking University), Bin Cui (Peking University), Yu Qiao (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences), Peng Gao (Chinese university of hong kong), and Hongsheng Li (The Chinese University of Hong Kong)

X-Trans2Cap: Cross-Modal Knowledge Transfer Using Transformer for 3D Dense Captioning 8553
Zhihao Yuan (The Chinese University of Hong Kong, Shenzhen), Xu Yan (The Chinese University of Hong Kong, Shenzhen), Yinghong Liao (The Chinese University of Hong Kong, Shenzhen), Yao Guo (Shanghai Jiao Tong Univerisity), Guanbin Li (Sun Yat-sen University), Shuguang Cui (The Chinese University of Hong Kong, Shenzhen), and Zhen Li (The Chinese University of Hong Kong, Shenzhen)

MVS2D: Efficient Multi-View Stereo via Attention-Driven 2D Convolutions 8564
Zhenpei Yang (The University of Texas at Austin), Zhile Ren (Apple Inc.), Qi Shan (Apple Inc.), and Qixing Huang (The University of Texas at Austin)

TransMVSNet: Global Context-Aware Multi-View Stereo Network With Transformers 8575
Yikang Ding (Tsinghua University), Wentao Yuan (Peking Universtiy), Qingtian Zhu (Peking University), Haotian Zhang (Megvii), Xiangyue Liu (Beihang University), Yuanjiang Wang (Megvii), and Xiao Liu (Megvii Inc)

RayMVSNet: Learning Ray-Based 1D Implicit Fields for Accurate Multi-View Stereo 8585
Junhua Xi (National University of Defense Technology), Yifei Shi (National University of Defense Technology), Yijie Wang (National University of Defense Technology, China), Yulan Guo (Sun Yat-sen University), and Kai Xu (National University of Defense Technology)

IterMVS: Iterative Probability Estimation for Efficient Multi-View Stereo 8596
Fangjinhua Wang (ETH Zurich), Silvano Galliani (Microsoft), Christoph Vogel (Microsoft MixedReality&AI), and Marc Pollefeys (ETH Zurich / Microsoft)

PSMNet: Position-Aware Stereo Merging Network for Room Layout Estimation 8606
Haiyan Wang (The City College of New York), Will Hutchcroft (Zillow Group), Yuguang Li (Zillow Group), Zhiqiang Wan (Zillow), Ivaylo Boyadzhiev (Zillow Group), Yingli Tian (City University of New York), and Sing Bing Kang (Zillow Group)

Non-Parametric Depth Distribution Modelling Based Depth Inference for Multi-View Stereo 8616
Jiayu Yang (The Australian National University), Jose M. Alvarez (NVIDIA), and Miaomiao Liu (The Australian National University)

Differentiable Stereopsis: Meshes From Multiple Views Using Differentiable Rendering 8625
Shubham Goel (University of California at Berkeley), Georgia Gkioxari (Facebook AI Resaerch), and Jitendra Malik (University of California at Berkeley)

Rethinking Depth Estimation for Multi-View Stereo: A Unified Representation 8635
Rui Peng (Peking University), Rongjie Wang (PCL), Zhenyu Wang (Shenzhen Graduate School, Peking University), Yawen Lai (Peking University), and Ronggang Wang (Peking University)

Efficient Multi-View Stereo by Iterative Dynamic Cost Volume 8645
Shaoqian Wang (Northwestern Polytechnical University), Bo Li (Northwestern Polytechnical University), and Yuchao Dai (Northwestern Polytechnical University)

PlaneMVS: 3D Plane Reconstruction From Multi-View Stereo 8655
Jiachen Liu (Pennsylvania State University), Pan Ji (OPPO US Research Center), Nitin Bansal (OPPO US Research Center), Changjiang Cai (InnoPeak Technology, Inc.), Qingan Yan (OPPO US Research Center), Xiaolei Huang (The Pennsylvania State University), and Yi Xu (OPPO US Research Center)

Discrete Time Convolution for Fast Event-Based Stereo 8666
Kaixuan Zhang (Southern university of Science and Technology), Kaiwei Che (SUSTech), Jianguo Zhang (Southern University of Science and Technology), Jie Cheng (Huawei Technology Co. Ltd.), Ziyang Zhang (Huawei Technology Co. Ltd.), Qinghai Guo (Huawei Technologies Co., Ltd.), and Luziwei Leng (Huawei Technologies Co., Ltd)

Stereo Magnification With Multi-Layer Images 8677
Taras Khakhulin (Skolkovo Institute of Science and Technology), Denis Korzhenkov (Yandex Research), Pavel Solovev (Samsung), Gleb Sterkin (Samsung AI Center, Moscow), Andrei-Timotei Ardelean (Skoltech), and Victor Lempitsky (Yandex)

Poster 2.2: Motion & Tracking

TransforMatcher: Match-to-Match Attention for Semantic Correspondence 8687
Seungwook Kim (POSTECH), Juhong Min (POSTECH), and Minsu Cho (POSTECH)

Probabilistic Warp Consistency for Weakly-Supervised Semantic Correspondences 8698
Prune Truong (ETH Zurich), Martin Danelljan (ETH Zurich), Fisher Yu (ETH Zurich), and Luc Van Gool (ETH Zurich)

Locality-Aware Inter– and Intra-Video Reconstruction for Self-Supervised Correspondence Learning 8709
Liulei Li (Beijing Institute of Technology), Tianfei Zhou (ETH Zurich), Wenguan Wang (Eidgenössische Technische Hochschule Zürich), Lu Yang (Beijing University of Posts and Telecommunications), Jianwu Li (Beijing Institute of Technology), and Yi Yang (UTS)

Transforming Model Prediction for Tracking 8721
Christoph Mayer (ETH Zurich), Martin Danelljan (ETH Zurich), Goutam Bhat (ETH Zurich), Matthieu Paul (ETH Zurich), Danda Pani Paudel (ETH Zürich), Fisher Yu (ETH Zurich), and Luc Van Gool (ETH Zurich)

Ranking-Based Siamese Visual Tracking 8731
Feng Tang (University of Science and Technology of China) and Qiang Ling (University of Science and Technology of China)

Correlation-Aware Deep Tracking 8741
Fei Xie (Southeast University), Chunyu Wang (Microsoft Research asia), Guangting Wang (University of Science and Technology of China), Yue Cao (Microsoft Research), Wankou Yang (Southeast University), and Wenjun Zeng (EIT Institute for Advanced Study)

Global Tracking via Ensemble of Local Trackers 8751
Zikun Zhou (Harbin Institute of Technology, Shenzhen), Jianqiu Chen (Harbin Institute of Technology, Shenzhen), Wenjie Pei (Harbin Institute of Technology, Shenzhen), Kaige Mao (Harbin Institute of Technology, Shenzhen), Hongpeng Wang (Harbin institute of technology, shenzhen), and Zhenyu He (Harbin Institute of Technology (Shenzhen))

Global Tracking Transformers 8761
Xingyi Zhou (The University of Texas at Austin), Tianwei Yin (UT Austin), Vladlen Koltun (Apple), and Philipp Krähenbühl (UT Austin)

Unified Transformer Tracker for Object Tracking 8771
Fan Ma (University of Technology Sydney), Mike Zheng Shou (National University of Singapore), Linchao Zhu (University of Technology, Sydney), Haoqi Fan (Facebook AI Research), Yilei Xu (Facebook), Yi Yang (UTS), and Zhicheng Yan (Facebook AI)

Transformer Tracking With Cyclic Shifting Window Attention 8781
Zikai Song (Huazhong University of Science and Technology), Junqing Yu (Huazhong University of Science & Technology), Yi-Ping Phoebe Chen (La Trobe University), and Wei Yang (Huazhong University of Science and Technology)

Spiking Transformers for Event-Based Single Object Tracking 8791
Jiqing Zhang (Dalian University of Technology), Bo Dong (Princeton University), Haiwei Zhang (Dalian University of Technology), Jianchuan Ding (Dalian University of Technology), Felix Heide (Princeton University), Baocai Yin (Dalian University of Technology), and Xin Yang (Dalian University of Technology)

Adiabatic Quantum Computing for Multi Object Tracking 8801
Jan-Nico Zaech (ETH Zürich), Alexander Liniger (ETH Zurich), Martin Danelljan (ETH Zurich), Dengxin Dai (ETH Zurich), and Luc Van Gool (ETH Zurich)

HiVT: Hierarchical Vector Transformer for Multi-Agent Motion Prediction 8813
Zikang Zhou (City University of Hong Kong), Luyao Ye (City University of Hong Kong), Jianping Wang (City University of Hong Kong), Kui Wu (University of Victoria), and Kejie Lu (University of Puerto Rico)

Towards Discriminative Representation: Multi-View Trajectory Contrastive Learning for Online Multi-Object Tracking 8824
En Yu (Huazhong University of Science and Technology), Zhuoling Li (Tsinghua University), and Shoudong Han (Huazhong University of Science and Technology)

TrackFormer: Multi-Object Tracking With Transformers 8834
Tim Meinhardt (TUM), Alexander Kirillov (Facebook AI Reserach), Laura Leal-Taixé (TUM), and Christoph Feichtenhofer (Facebook AI Research)

Learning of Global Objective for Network Flow in Multi-Object Tracking 8845
Shuai Li (Rochester Institute of Technology), Yu Kong (Rochester Institute of Technology), and Hamid Rezatofighi (Monash University)

LMGP: Lifted Multicut Meets Geometry Projections for Multi-Camera Multi-Object Tracking 8856
Duy M. H. Nguyen (German Research Centre for Artificial Intelligence), Roberto Henschel (Leibniz University of Hannover), Bodo Rosenhahn (Leibniz University Hannover), Daniel Sonntag (DFKI), and Paul Swoboda (MPI fuer Informatik, Saarbruecken)

Multi-Object Tracking Meets Moving UAV 8866
Shuai Liu (Dalian University of Technology), Xin Li (Peng Cheng Laboratory), Huchuan Lu (Dalian University of Technology), and You He (Naval Aviation University)

Visible-Thermal UAV Tracking: A Large-Scale Benchmark and New Baseline 8876
Pengyu Zhang (Dalian University of Technology), Jie Zhao (Dalian University of Technology), Dong Wang (Dalian University of Technology), Huchuan Lu (Dalian University of Technology), and Xiang Ruan (Tiwaki Co.Ltd.)

Unsupervised Domain Adaptation for Nighttime Aerial Tracking 8886
Junjie Ye (Tongji University), Changhong Fu (Tongji University), Guangze Zheng (Tongji University), Danda Pani Paudel (ETH Zürich), and Guang Chen (Tongji University)

Learning Optical Flow With Kernel Patch Attention 8896
Ao Luo (Megvii), Fan Yang (Group 42), Xin Li (Group 42), and Shuaicheng Liu (UESTC; Megvii)

Towards Understanding Adversarial Robustness of Optical Flow Networks 8906
Simon Schrodi (University of Freiburg), Tonmoy Saikia (University of Freiburg), and Thomas Brox (University of Freiburg)

DIP: Deep Inverse Patchmatch for High-Resolution Optical Flow 8915
Zihua Zheng (Megvii tec.), Ni Nie (Megvii), Zhi Ling (megvii), Pengfei Xiong (Tencent), Jiangyu Liu (megvii inc), Hao Wang (Megvii), and Jiankun Li (megvii)

Poster 2.2: Computer Vision Theory

On the Instability of Relative Pose Estimation and RANSAC's Role 8925
Hongyi Fan (Brown University), Joe Kileel (Princeton University), and Benjamin Kimia (Brown University)

Bootstrapping ViTs: Towards Liberating Vision Transformers From Pre-Training 8934
Haofei Zhang (Zhejiang University), Jiarui Duan (Zhejiang University), Mengqi Xue (Zhejiang University), Jie Song (Zhejiang University), Li Sun (Zhejiang University), and Mingli Song (Zhejiang University)

Global Sensing and Measurements Reuse for Image Compressed Sensing 8944
Zi-En Fan (Xi'an Jiaotong University), Feng Lian (Xi'an Jiaotong University), and Jia-Ni Quan (Xi'an Jiaotong University)

Maximum Consensus by Weighted Influences of Monotone Boolean Functions 8954
Erchuan Zhang (Edith Cowan University), David Suter (Edith Cowan University), Ruwan Tennakoon (RMIT), Tat-Jun Chin (The University of Adelaide), Alireza Bab-Hadiashar (RMIT University), Giang Truong (Edith Cowan University), and Syed Zulqarnain Gilani (Edith Cowan University)

MS2DG-Net: Progressive Correspondence Learning via Multiple Sparse Semantics Dynamic Graph.... 8963
Luanyuan Dai (Digital Fujian Research Institute of Big Data for Agriculture and Forestry, College of Computer and Information Science, Fujian Agriculture and Forestry University), Yizhang Liu (School of Software Engineering, Tongji University, Shanghai 201804), Jiayi Ma (Wuhan University), Lifang Wei (Fujian Agriculture and Forestry University), Taotao Lai (Minjiang University), Changcai Yang (Fujian Agriculture and Forestry University), and Riqing Chen (Fujian Agriculture and Forestry University)

Styleformer: Transformer Based Generative Adversarial Networks With Style Vector 8973
Jeeseung Park (MINDsLab Inc.) and Younggeun Kim (Seoul National University)

Scanline Homographies for Rolling-Shutter Plane Absolute Pose 8983
Fang Bai (Clermont Auvergne University), Agniva Sengupta (Université Clermont Auvergne), and Adrien Bartoli (Université Clermont Auvergne)

Poster 2.2: Transfer / Low-Shot / Long-Tail Learning

Generating Representative Samples for Few-Shot Classification 8993
Jingyi Xu (Stony Brook University) and Hieu Le (Amazon Robotics)

Matching Feature Sets for Few-Shot Image Classification 9004
Arman Afrasiyabi (Université Laval), Hugo Larochelle (Google), Jean-François Lalonde (Université Laval), and Christian Gagné (Université Laval)

Improving Adversarially Robust Few-Shot Image Classification With Generalizable Representations 9015
Junhao Dong (Sun Yat-sen University), Yuan Wang (Sun Yat-sen University), Jian-Huang Lai (Sun Yat-sen University), and Xiaohua Xie (Sun Yat-sen University)

Sylph: A Hypernetwork Framework for Incremental Few-Shot Object Detection 9025
Li Yin (Facebook AI), Juan M. Perez-Rua (Facebook AI), and Kevin J. Liang (Facebook)

Forward Compatible Few-Shot Class-Incremental Learning 9036
Da-Wei Zhou (Nanjing University), Fu-Yun Wang (Nanjing University), Han-Jia Ye (Nanjing University), Liang Ma (Hikvision Research Institute), Shiliang Pu (Hikvision Research Institute), and De-Chuan Zhan (Nanjing University)

Constrained Few-Shot Class-Incremental Learning 9047
Michael Hersche (IBM Research Zurich GmbH), Geethan Karunaratne (IBM Research Europe), Giovanni Cherubini (IBM Research - Zurich), Luca Benini (ETHZ, University of Bologna), Abu Sebastian (IBM Research Zurich GmbH), and Abbas Rahimi (IBM Research-Zurich)

Pushing the Limits of Simple Pipelines for Few-Shot Learning: External Data and Fine-Tuning Make a Difference 9058
Shell Xu Hu (Samsung), Da Li (Samsung), Jan Stühmer (Samsung AI Cambridge), Minyoung Kim (Samsung AI Center Cambridge), and Timothy M. Hospedales (Samsung AI Centre)

EASE: Unsupervised Discriminant Subspace Learning for Transductive Few-Shot Learning 9068
Hao Zhu (Australian National University) and Piotr Koniusz (ANU College of Engineering and Computer Science)

Few-Shot Learning With Noisy Labels 9079
Kevin J. Liang (Facebook), Samrudhdhi B. Rangrej (McGill University), Vladan Petrovic (Facebook), and Tal Hassner (Facebook AI)

Ranking Distance Calibration for Cross-Domain Few-Shot Learning 9089
Pan Li (Queen Mary University of London), Shaogang Gong (Queen Mary University of London), Chengjie Wang (Tencent), and Yanwei Fu (Fudan University)

Revisiting Learnable Affines for Batch Norm in Few-Shot Transfer Learning 9099
Moslem Yazdanpanah (UOK), Aamer Abdul Rahman (Universite de Montreal), Muawiz Chaudhary (MILA), Christian Desrosiers (ETS, Canada), Mohammad Havaei (Imagia), Eugene Belilovsky (Concordia University), and Samira Ebrahimi Kahou (École de Technologie Supérieure)

Attribute Surrogates Learning and Spectral Tokens Pooling in Transformers for Few-Shot Learning 9109
Yangji He (Fudan University), Weihan Liang (Fudan University), Dongyang Zhao (Fudan University), Hong-Yu Zhou (The University of Hong Kong), Weifeng Ge (Fudan University), Yizhou Yu (The University of Hong Kong), and Wenqiang Zhang (Fudan University)

Learning To Memorize Feature Hallucination for One-Shot Image Generation 9120
Yu Xie (Fudan University), Yanwei Fu (Fudan University), Ying Tai (Tencent YouTu), Yun Cao (Tencent), Junwei Zhu (Tencent), and Chengjie Wang (Tencent)

A Closer Look at Few-Shot Image Generation 9130
Yunqing Zhao (Singapore University of Technology and Design), Henghui Ding (ETH Zurich), Houjing Huang (ByteDance Ltd.), and Ngai-Man Cheung (Singapore University of Technology and Design)

Motion-Modulated Temporal Fragment Alignment Network for Few-Shot Action Recognition 9141
Jiamin Wu (University of Science and Technology of China), Tianzhu Zhang (University of Science and Technology of China), Zhe Zhang (Lunar Exploration and Space Engineering Center of CNSA), Feng Wu (University of Science and Technology of China), and Yongdong Zhang (University of Science and Technology of China)

Knowledge Distillation As Efficient Pre-Training: Faster Convergence, Higher Data-Efficiency, and Better Transferability 9151
Ruifei He (The University of Hong Kong), Shuyang Sun (University of Oxford), Jihan Yang (The University of Hong Kong), Song Bai (University of Oxford), and Xiaojuan Qi (The University of Hong Kong)

Transferability Estimation Using Bhattacharyya Class Separability 9162
Michal Pándy (University of Cambridge), Andrea Agostinelli (Google), Jasper Uijlings (Google Research), Vittorio Ferrari (Google Research), and Thomas Mensink (Google Research / University of Amsterdam)

Revisiting the Transferability of Supervised Pretraining: An MLP Perspective 9173
Yizhou Wang (Zhejiang University), Shixiang Tang (The University of Sydney), Feng Zhu (University of Science and Technology of China), Lei Bai (University of Sydney), Rui Zhao (SenseTime Group Limited), Donglian Qi (Zhejiang University), and Wanli Ouyang (The University of Sydney)

Task2Sim: Towards Effective Pre-Training and Transfer From Synthetic Data 9184
Samarth Mishra (Boston University), Rameswar Panda (MIT-IBM Watson AI Lab), Cheng Perng Phoo (Cornell University), Chun-Fu Richard Chen (JPMorgan Chase Bank, N.A.), Leonid Karlinsky (IBM-Research), Kate Saenko (Boston University), Venkatesh Saligrama (Boston University), and Rogerio S. Feris (MIT-IBM Watson AI Lab, IBM Research)

Which Model To Transfer? Finding the Needle in the Growing Haystack 9195
Cedric Renggli (ETH Zurich), André Susano Pinto (Google Research), Luka Rimanic (ETH Zurich), Joan Puigcerver (Google), Carlos Riquelme (Google Brain), Ce Zhang (ETH), and Mario Lučić (Google Brain)

Does Robustness on ImageNet Transfer to Downstream Tasks? 9205
Yutaro Yamada (Yale University) and Mayu Otani (CyberAgent)

What Makes Transfer Learning Work for Medical Images: Feature Reuse & Other Factors 9215
Christos Matsoukas (KTH Royal Institute of Technology), Johan Fredin Haslum (KTH Royal Institute of Technology), Moein Sorkhei (KTH Royal Institute of Technology), Magnus Söderberg (AstraZeneca), and Kevin Smith (KTH Royal Institute of Technology)

OW-DETR: Open-World Detection Transformer 9225
Akshita Gupta (IIAI), Sanath Narayan (Inception Institute of Artificial Intelligence), K J Joseph (Indian Institute of Technology, Hyderabad), Salman Khan (MBZUAI/ANU), Fahad Shahbaz Khan (MBZUAI), and Mubarak Shah (University of Central Florida)

Unseen Classes at a Later Time? No Problem 9235
Hari Chandana Kuchibhotla (Indian Institute of Technology, Hyderabad), Sumitra S Malagi (Indian Institute of Technology, Hyderabad), Shivam Chandhok (Indian Institute of Technology, Hyderabad), and Vineeth N Balasubramanian (Indian Institute of Technology, Hyderabad)

Continual Object Detection via Prototypical Task Correlation Guided Gating Mechanism 9245
Binbin Yang (Sun Yat-sen University), Xinchi Deng (Sun Yat-sen University), Han Shi (The Hong Kong University of Science and Technology), Changlin Li (Monash University), Gengwei Zhang (Sun Yat-sen University), Hang Xu (Huawei Noah's Ark Lab), Shen Zhao (Sun Yat-Sen University), Liang Lin (Sun Yat-sen University), and Xiaodan Liang (Sun Yat-sen University)

On Generalizing Beyond Domains in Cross-Domain Continual Learning 9255
Christian Simon (Australian National University), Masoud Faraki (NEC Labs America), Yi-Hsuan Tsai (Phiar Technologies), Xiang Yu (NEC Labs), Samuel Schulter (NEC Laboratories America), Yumin Suh (NEC Labs America), Mehrtash Harandi (Monash University), and Manmohan Chandraker (UC San Diego)

Online Continual Learning on a Contaminated Data Stream With Blurry Task Boundaries 9265
Jihwan Bang (Naver Corp.), Hyunseo Koh (Gwangju Institute of Science and Technology), Seulki Park (Seoul National University), Hwanjun Song (NAVER AI Lab), Jung-Woo Ha (NAVER CLOVA AI Lab), and Jonghyun Choi (Yonsei University)

DyTox: Transformers for Continual Learning With DYnamic TOken eXpansion 9275
Arthur Douillard (Heuritech / Sorbonne University), Alexandre Ramé (Heuritech), Guillaume Couairon (Facebook AI Research), and Matthieu Cord (Sorbonne University)

Self-Sustaining Representation Expansion for Non-Exemplar Class-Incremental Learning 9286
Kai Zhu (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)

En-Compactness: Self-Distillation Embedding & Contrastive Generation for Generalized Zero-Shot Learning 9296
Xia Kong (Xiamen University), Zuodong Gao (Xiamen University), Xiaofan Li (Xiamen University), Ming Hong (xiamen university), Jun Liu (Tencent), Chengjie Wang (Tencent), Yuan Xie (East China Normal University), and Yanyun Qu (XMU)

VGSE: Visually-Grounded Semantic Embeddings for Zero-Shot Learning 9306
Wenjia Xu (Institute of Electronics, Chinese Academy of Sciences), Yongqin Xian (ETH Zurich), Jiuniu Wang (City University of Hong Kong), Bernt Schiele (MPI Informatics), and Zeynep Akata (University of Tübingen)

Siamese Contrastive Embedding Network for Compositional Zero-Shot Learning 9316
Xiangyu Li (Xidian University), Xu Yang (Xidian University), Kun Wei (Xidian University), Cheng Deng (Xidian University), and Muli Yang (Xidian University)

KG-SP: Knowledge Guided Simple Primitives for Open World Compositional Zero-Shot Learning 9326
Shyamgopal Karthik (University of Tübingen), Massimiliano Mancini (University of Tübingen), and Zeynep Akata (University of Tübingen)

Non-Generative Generalized Zero-Shot Learning via Task-Correlated Disentanglement and Controllable Samples Synthesis 9336
Yaogong Feng (Beijing Jiaotong University), Xiaowen Huang (Beijing Jiaotong University), Pengbo Yang (Beijing Jiaotong University), Jian Yu (Beijing Jiaotong University), and Jitao Sang (Beijing Jiaotong University, China)

WALT: Watch and Learn 2D Amodal Representation From Time-Lapse Imagery 9346
N. Dinesh Reddy (Carnegie mellon University), Robert Tamburo (missing), and Srinivasa G. Narasimhan (Carnegie Mellon University, USA)

Poster 2.2: Recognition: Detection, Categorization, Retrieval

Omni-DETR: Omni-Supervised Object Detection With Transformers 9357
Pei Wang (UC San Diego), Zhaowei Cai (Amazon), Hao Yang (Amazon), Gurumurthy Swaminathan (Amazon), Nuno Vasconcelos (UC San Diego), Bernt Schiele (MPI Informatics), and Stefano Soatto (UCLA)

DESTR: Object Detection With Split Transformer 9367
Liqiang He (Oregon State University) and Sinisa Todorovic (Oregon State U)

A Dual Weighting Label Assignment Scheme for Object Detection 9377
Shuai Li (The Hong Kong Polytechnic University), Chenhang He (The Hong Kong Polytechnic University), Ruihuang Li (Hong Kong Polytechnic University), and Lei Zhang (Hong Kong Polytechnic University, Hong Kong, China)

Entropy-Based Active Learning for Object Detection With Progressive Diversity Constraint 9387
Jiaxi Wu (Beihang University), Jiaxin Chen (Beihang University), and Di Huang (Beihang University, China)

Localization Distillation for Dense Object Detection 9397
Zhaohui Zheng (Nankai University), Rongguang Ye (Tianjin University), Ping Wang (Tianjin University), Dongwei Ren (Harbin Institute of Technology), Wangmeng Zuo (Harbin Institute of Technology, China), Qibin Hou (Nankai University), and Ming-Ming Cheng (Nankai University)

Group R-CNN for Weakly Semi-Supervised Object Detection With Points 9407
Shilong Zhang (University of Science and Technology of China), Zhuoran Yu (Georgia Institute of Technology), Liyang Liu (Tencent), Xinjiang Wang (SenseTime Group Ltd.), Aojun Zhou (Sensetime&tetras.ai), and Kai Chen (SenseTime Research)

Overcoming Catastrophic Forgetting in Incremental Object Detection via Elastic Response Distillation 9417
Tao Feng (Alibaba Group), Mang Wang (Alibaba Group), and Hangjie Yuan (Zhejiang University)

CREAM: Weakly Supervised Object Localization via Class RE-Activation Mapping 9427
Jilan Xu (Fudan University), Junlin Hou (Fudan University), Yuejie Zhang (Fudan University), Rui Feng (Fudan University), Rui-Wei Zhao (Fudan University), Tao Zhang (Shanghai University of Finance and Economics), Xuequan Lu (Deakin University), and Shang Gao (Deakin University)

One Loss for Quantization: Deep Hashing With Discrete Wasserstein Distributional Matching 9437
Khoa D. Doan (Virginia Tech), Peng Yang (Baidu), and Ping Li (Baidu Research)

PSTR: End-to-End One-Step Person Search With Transformers 9448
Jiale Cao (Tianjin University), Yanwei Pang (Tianjin University), Rao Muhammad Anwer (MBZUAI/AALTO), Hisham Cholakkal (MBZUAI), Jin Xie (Chongqing University), Mubarak Shah (University of Central Florida), and Fahad Shahbaz Khan (MBZUAI)

Protecting Celebrities From DeepFake With Identity Consistency Transformer 9458
Xiaoyi Dong (University of Science and Technology of China), Jianmin Bao (Microsoft Research Asia), Dongdong Chen (Microsoft Cloud AI), Ting Zhang (MSRA), Weiming Zhang (University of Science and Technology of China), Nenghai Yu (University of Science and Technology of China), Dong Chen (Microsoft Research Asia), Fang Wen (Microsoft Research Asia), and Baining Guo (MSR Asia)

MDAN: Multi-Level Dependent Attention Network for Visual Emotion Analysis 9469
Liwen Xu (Tencent Music Entertainment), Zhengtao Wang (Tencent Music Entertainment), Bin Wu (Tencent Music Entertainment), and Simon Lui (Tencent Music Entertainment)

Contextual Similarity Distillation for Asymmetric Image Retrieval 9479
Hui Wu (University of Science and Technology of China), Min Wang (Institute of Artificial Intelligence, Hefei Comprehensive National Science Center), Wengang Zhou (University of Science and Technology of China), Houqiang Li (Huawei Cloud & AI), and Qi Tian (University of Science and Technology of China)

Improving Visual Grounding With Visual-Linguistic Verification and Iterative Reasoning 9489
Li Yang (NLPR & UCAS), Yan Xu (Chinese University of Hong Kong), Chunfeng Yuan (NLPR), Wei Liu (National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences), Bing Li (National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Sciences, ,), and Weiming Hu (Institute of AutomationChinese Academy of Sciences)

MPC: Multi-View Probabilistic Clustering 9499
Junjie Liu (Zhejiang University), Junlong Liu (Alibaba DAMO Academy), Shaotian Yan (Zhejiang University), Rongxin Jiang (Zhejiang University), Xiang Tian (Zhejiang University), Boxuan Gu (Zhejiang University), Yaowu Chen (Zhejiang University), Chen Shen (Alibaba Group), and Jianqiang Huang (Alibaba Group)

Text Spotting Transformers 9509
Xiang Zhang (University of California San Diego), Yongwen Su (Shanghai Jiao Tong University), Subarna Tripathi (Intel Labs), and Zhuowen Tu (UC San Diego)

Represent, Compare, and Learn: A Similarity-Aware Framework for Class-Agnostic Counting 9519
Min Shi (Huazhong University of Science and Technology), Hao Lu (Huazhong University of Science and Technology), Chen Feng (Huazhong University of Science and Technology), Chengxin Liu (Huazhong University of Science and Technology), and Zhiguo Cao (Huazhong Univ. of Sci.&Tech.)

Reflection and Rotation Symmetry Detection via Equivariant Learning 9529
Ahyun Seo (POSTECH), Byungjin Kim (Postech), Suha Kwak (POSTECH), and Minsu Cho (POSTECH)

Learning To Imagine: Diversify Memory for Incremental Learning Using Unlabeled Data 9539
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)

A Simple Episodic Linear Probe Improves Visual Recognition in the Wild 9549
Yuanzhi Liang (UTS), Linchao Zhu (University of Technology, Sydney), Xiaohan Wang (Zhejiang University), and Yi Yang (UTS)

Cross Domain Object Detection by Target-Perceived Dual Branch Distillation 9560
Mengzhe He (Shenzhen Institutes of Advanced Technology Chinese Academy of Science), Yali Wang (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences), Jiaxi Wu (Beihang University), Yiru Wang (SenseTime Group Limited), Hanqing Li (SenseTime Group Limited), Bo Li (SenseTime Group Limited), Weihao Gan (SenseTime Group Limited), Wei Wu (SenseTime Group Limited), and Yu Qiao (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences)

Multi-Granularity Alignment Domain Adaptation for Object Detection 9571
Wenzhang Zhou (University of Chinese Academy of Sciences), Dawei Du (Kitware, Inc.), Libo Zhang (Institute of Software Chinese Academy of Sciences), Tiejian Luo (University of Chinese Academy of Sciences), and Yanjun Wu (Institute of Software Chinese Academy of Sciences)

Expanding Low-Density Latent Regions for Open-Set Object Detection 9581
Jiaming Han (Wuhan University), Yuqiang Ren (Tencent), Jian Ding (Wuhan University), Xingjia Pan (Tencent), Ke Yan (Tencent), and Gui-Song Xia (Wuhan University)

Class-Incremental Learning With Strong Pre-Trained Models 9591
Tz-Ying Wu (UCSD), Gurumurthy Swaminathan (Amazon), Zhizhong Li (Amazon), Avinash Ravichandran (Amazon), Nuno Vasconcelos (UCSD, USA), Rahul Bhotika (Amazon), and Stefano Soatto (UCLA)

ProposalCLIP: Unsupervised Open-Category Object Proposal Generation via Exploiting CLIP Cues 9601
Hengcan Shi (Monash University), Munawar Hayat (Monash University), Yicheng Wu (Monash University), and Jianfei Cai (Monash University)

Poster 2.2: Self-, Semi-, Meta-, & Unsupervised Learning

Self-Supervised Models Are Continual Learners 9611
Enrico Fini (University of Trento), Victor G. Turrisi da Costa (University of Trento), Xavier Alameda-Pineda (INRIA), Elisa Ricci (University of Trento), Karteek Alahari (Inria), and Julien Mairal (INRIA)

The Two Dimensions of Worst-Case Training and Their Integrated Effect for Out-of-Domain Generalization 9621
Zeyi Huang (Carnegie Mellon University), Haohan Wang (Carnegie Mellon University), Dong Huang (Carnegie Mellon University), Yong Jae Lee (University of Wisconsin-Madison), and Eric P. Xing (MBZUAI, CMU, and Petuum Inc.)

Beyond Supervised vs. Unsupervised: Representative Benchmarking and Analysis of Image Representation Learning 9632
Matthew Gwilliam (University of Maryland) and Abhinav Shrivastava (University of Maryland)

SimMIM: A Simple Framework for Masked Image Modeling 9643
Zhenda Xie (Tsinghua University), Zheng Zhang (MSRA), Yue Cao (Microsoft Research), Yutong Lin (Xi'an Jiaotong University), Jianmin Bao (Microsoft Research Asia), Zhuliang Yao (Tsinghua University), Qi Dai (Microsoft Research), and Han Hu (Microsoft Research Asia)

Semantic-Aware Auto-Encoders for Self-Supervised Representation Learning 9654
Guangrun Wang (University of Oxford), Yansong Tang (Tsinghua University), Liang Lin (Sun Yat-sen University), and Philip H.S. Torr (University of Oxford)

UniCon: Combating Label Noise Through Uniform Selection and Contrastive Learning 9666
Nazmul Karim (University of Central Florida), Mamshad Nayeem Rizve (University of Central Florida), Nazanin Rahnavard (University of Central Florida), Ajmal Mian (University of Western Australia), and Mubarak Shah (University of Central Florida)

Contrastive Conditional Neural Processes 9677
Zesheng Ye (University of New South Wales, Sydney) and Lina Yao (University of New South Wales)

One-Bit Active Query With Contrastive Pairs 9687
Yuhang Zhang (Shanghai Jiao Tong University), Xiaopeng Zhang (Huawei Cloud EI), Lingxi Xie (Huawei Inc.), Jie Li (Shanghai Jiao Tong University), Robert C. Qiu (Shanghai Jiao Tong University), Hengtong Hu (Hefei University of Technology), and Qi Tian (Huawei Cloud & AI)

HCSC: Hierarchical Contrastive Selective Coding 9696
Yuanfan Guo (SJTU), Minghao Xu (Mila - Quebec AI Institute), Jiawen Li (bytedance.com), Bingbing Ni (Shanghai Jiao Tong University), Xuanyu Zhu (bytedance), Zhenbang Sun (Bytedance), and Yi Xu (Shanghai Jiao Tong University)

Motion-Aware Contrastive Video Representation Learning via Foreground-Background Merging 9706
Shuangrui Ding (Shanghai Jiao Tong University), Maomao Li (Tencent AIlab), Tianyu Yang (Tencent AI Lab), Rui Qian (The Chinese University of Hong Kong), Haohang Xu (Shanghai Jiao Tong University), Qingyi Chen (University of Michigan), Jue Wang (Tencent AI Lab), and Hongkai Xiong (Shanghai Jiao Tong University)

Hierarchical Self-Supervised Representation Learning for Movie Understanding 9717
Fanyi Xiao (Amazon), Kaustav Kundu (Amazon), Joseph Tighe (Amazon), and Davide Modolo (Amazon)

Anomaly Detection via Reverse Distillation From One-Class Embedding 9727
Hanqiu Deng (University of Alberta) and Xingyu Li (University of Alberta)

Unsupervised Representation Learning for Binary Networks by Joint Classifier Learning 9737
Dahyun Kim (Upstage AI Research) and Jonghyun Choi (Yonsei University)

DC-SSL: Addressing Mismatched Class Distribution in Semi-Supervised Learning 9747
Zhen Zhao (University of Sydney), Luping Zhou (University of Sydney), Yue Duan (Nanjing University), Lei Wang (University of Wollongong, Australia), Lei Qi (Southeast University), and Yinghuan Shi (Nanjing University)

Learning To Collaborate in Decentralized Learning of Personalized Models 9756
Shuangtong Li (University of Science and Technology of China (USTC), ,), Tianyi Zhou (University of Washington), Xinmei Tian (USTC), and Dacheng Tao (JD.com)

Highly-Efficient Incomplete Large-Scale Multi-View Clustering With Consensus Bipartite Graph 9766
Siwei Wang (National University of Defense Technology), Xinwang Liu (National University of Defense Technology), Li Liu (National University of Defense Technology), Wenxuan Tu (NUDT), Xinzhong Zhu (Zhejiang Normal University), Jiyuan Liu (National University of Defense Technology), Sihang Zhou (NUDT), and En Zhu (National University of Defense Technology)

DASO: Distribution-Aware Semantics-Oriented Pseudo-Label for Imbalanced Semi-Supervised Learning 9776
Youngtaek Oh (KAIST), Dong-Jin Kim (UC Berkeley), and In So Kweon (KAIST, Korea)

Global Convergence of MAML and Theory-Inspired Neural Architecture Search for Few-Shot Learning 9787
Haoxiang Wang (University of Illinois, Urbana-Champaign), Yite Wang (University of Illinois at Urbana-Champaign), Ruoyu Sun (University of Illinois at Urbana-Champaign), and Bo Li (UIUC)

Semi-Supervised Object Detection via Multi-Instance Alignment With Global Class Prototypes.... 9799
Aoxue Li (Noah's Ark Lab), Peng Yuan (Huawei Noah's Ark Lab), and Zhenguo Li (Huawei Noah's Ark Lab)

Unbiased Teacher v2: Semi-Supervised Object Detection for Anchor-Free and Anchor-Based Detectors 9809
Yen-Cheng Liu (Georgia Institute of Technology), Chih-Yao Ma (Facebook), and Zsolt Kira (Georgia Institute of Technology)

Spectral Unsupervised Domain Adaptation for Visual Recognition 9819
Jingyi Zhang (Nanyang Technological University), Jiaxing Huang (Nanyang Technological University), Zichen Tian (Nanyang Technological University), and Shijian Lu (Nanyang Technological University)

DATA: Domain-Aware and Task-Aware Self-Supervised Learning 9831
Qing Chang (University of Chinese Academy of Sciences), Junran Peng (Chinese Academy of Sciences), Lingxi Xie (Huawei Inc.), Jiajun Sun (University of Chinese Academy of Sciences), Haoran Yin (University of Chinese Academy of Sciences), Qi Tian (Huawei Cloud & AI), and Zhaoxiang Zhang (Chinese Academy of Sciences, China)

Dynamic Kernel Selection for Improved Generalization and Memory Efficiency in Meta-Learning 9841
Arnav Chavan (Indian Institute of Technology - Dhanbad), Rishabh Tiwari (Indian Institute of Technology (ISM) - Dhanbad), Udbhav Bamba (Indian Institute of Technology, ISM Dhanbad), and Deepak K. Gupta (Indian Institute of Technology (ISM) - Dhanbad)

DeepDPM: Deep Clustering With an Unknown Number of Clusters 9851
Meitar Ronen (Ben-Gurion University), Shahaf E. Finder (Ben-Gurion University), and Oren Freifeld (Ben-Gurion University)

PLAD: Learning To Infer Shape Programs With Pseudo-Labels and Approximate Distributions ... 9861
R. Kenny Jones (Brown University), Homer Walke (University of California - Berkeley), and Daniel Ritchie (Brown University)

Robust Outlier Detection by De-Biasing VAE Likelihoods 9871
Kushal Chauhan (Google Research India), Barath Mohan U (Indian Institute of Science), Pradeep Shenoy (Google Research India), Manish Gupta (Google Research India), and Devarajan Sridharan (Indian Institute of Science)

Image-to-Lidar Self-Supervised Distillation for Autonomous Driving Data 9881
Corentin Sautier (valeo.ai), Gilles Puy (valeo.ai), Spyros Gidaris (valeo.ai), Alexandre Boulch (valeo.ai), Andrei Bursuc (valeo.ai), and Renaud Marlet (Ecole des Ponts ParisTech)

CrossPoint: Self-Supervised Cross-Modal Contrastive Learning for 3D Point Cloud Understanding 9892
Mohamed Afham (University of Moratuwa), Isuru Dissanayake (University of Moratuwa), Dinithi Dissanayake (University of Moratuwa), Amaya Dharmasiri (University of Moratuwa), Kanchana Thilakarathna (University of Sydney), and Ranga Rodrigo (University of Moratuwa)

Cross-Domain Correlation Distillation for Unsupervised Domain Adaptation in Nighttime Semantic Segmentation 9903
Huan Gao (Tianjin University), Jichang Guo (Tianjin University), Guoli Wang (Horizon Robotics), and Qian Zhang (Horizon Robotics)

DAFormer: Improving Network Architectures and Training Strategies for Domain-Adaptive Semantic Segmentation 9914
Lukas Hoyer (ETH Zurich), Dengxin Dai (ETH Zurich), and Luc Van Gool (ETH Zurich)

WildNet: Learning Domain Generalized Semantic Segmentation From the Wild 9926
Suhyeon Lee (Yonsei University), Hongje Seong (Yonsei University), Seongwon Lee (Yonsei university), and Euntai Kim (Yonsei University)

UCC: Uncertainty Guided Cross-Head Co-Training for Semi-Supervised Semantic Segmentation 9937
Jiashuo Fan (Tsinghua University), Bin Gao (Huawei Noah's Ark Lab), Huan Jin (Huawei Noah's Ark Lab), and Lihui Jiang (Huawei Technologies Co., Ltd)

Semi-Supervised Semantic Segmentation With Error Localization Network 9947
Donghyeon Kwon (POSTECH) and Suha Kwak (POSTECH)

Unbiased Subclass Regularization for Semi-Supervised Semantic Segmentation 9958
Dayan Guan (Nanyang Technological University), Jiaxing Huang (Nanyang Technological University), Aoran Xiao (Nanyang Technological University), and Shijian Lu (Nanyang Technological University)

Integrative Few-Shot Learning for Classification and Segmentation 9969
Dahyun Kang (POSTECH) and Minsu Cho (POSTECH)

GanOrCon: Are Generative Models Useful for Few-Shot Segmentation? 9981
Oindrila Saha (University of Massachusetts Amherst), Zezhou Cheng (University of Massachusetts, Amherst), and Subhransu Maji (University of Massachusetts, Amherst)

SphericGAN: Semi-Supervised Hyper-Spherical Generative Adversarial Networks for Fine-Grained Image Synthesis 9991
Tianyi Chen (South China University of Technology), Yunfei Zhang (South China University of Technology), Xiaoyang Huo (south china university of technology), Si Wu (South China University of Technology), Yong Xu (South China University of Technology), and Hau San Wong (City University of Hong Kong)

CoordGAN: Self-Supervised Dense Correspondences Emerge From GANs 10001
Jiteng Mu (University of California, San Diego), Shalini De Mello (NVIDIA Research), Zhiding Yu (NVIDIA), Nuno Vasconcelos (UCSD, USA), Xiaolong Wang (UCSD), Jan Kautz (NVIDIA), and Sifei Liu (NVIDIA)

Poster 2.2: Privacy and Federated Learning

GradViT: Gradient Inversion of Vision Transformers 10011
Ali Hatamizadeh (NVIDIA Corporation), Hongxu Yin (NVIDIA), Holger R. Roth (NVIDIA), Wenqi Li (NVIDIA), Jan Kautz (NVIDIA), Daguang Xu (NVIDIA Corporation), and Pavlo Molchanov (NVIDIA)

Deep 3D-to-2D Watermarking: Embedding Messages in 3D Meshes and Extracting Them From 2D Renderings 10021
Innfarn Yoo (Google Research), Huiwen Chang (Google), Xiyang Luo (Google), Ondrej Stava (Google), Ce Liu (Microsoft), Peyman Milanfar (Google), and Feng Yang (Google Research)

CD2-pFed: Cyclic Distillation-Guided Channel Decoupling for Model Personalization in Federated Learning 10031
Yiqing Shen (Shanghai Jiao Tong University), Yuyin Zhou (Johns Hopkins University), and Lequan Yu (The University of Hong Kong)

APRIL: Finding the Achilles' Heel on Privacy for Vision Transformers 10041
Jiahao Lu (Institute of Automation, Chinese Academy of Sciences), Xi Sheryl Zhang (Institute of Automation, Chinese Academy of Sciences), Tianli Zhao (Institute of Automation, Chinese Academy of Sciences;University of Chinese Academy of Sciences), Xiangyu He (Institute of Automation, Chinese Academy of Sciences), and Jian Cheng (Chinese Academy of Sciences, China)

Rethinking Architecture Design for Tackling Data Heterogeneity in Federated Learning 10051
Liangqiong Qu (Stanford University), Yuyin Zhou (Johns Hopkins University), Paul Pu Liang (Carnegie Mellon University), Yingda Xia (Alibaba USA Inc.), Feifei Wang (Stanford University), Ehsan Adeli (Stanford University), Li Fei-Fei (Stanford University), and Daniel Rubin

Robust Federated Learning With Noisy and Heterogeneous Clients 10062
Xiuwen Fang (Wuhan University) and Mang Ye (Wuhan University)

Federated Learning With Position-Aware Neurons 10072
Xin-Chun Li (Nanjing University), Yi-Chu Xu (Nanjing University), Shaoming Song (Huawei Noah's Ark Lab), Bingshuai Li (Huawei Noah's Ark Lab), Yinchuan Li (Huawei Noah's Ark Lab), Yunfeng Shao (Huawei Noah's Ark Lab), and De-Chuan Zhan (Nanjing University)

Layer-Wised Model Aggregation for Personalized Federated Learning 10082
Xiaosong Ma (Southeast University), Jie Zhang (The Hong Kong Polytechnic University), Song Guo (The Hong Kong Polytechnic University), and Wenchao Xu (The Hong Kong Polytechnic University)

FedCor: Correlation-Based Active Client Selection Strategy for Heterogeneous Federated Learning 10092
Minxue Tang (Duke University), Xuefei Ning (Tsinghua University), Yitu Wang (Duke University), Jingwei Sun (Duke University), Yu Wang (tsinghua university), Hai Li (Duke University), and Yiran Chen (Duke University)

FedDC: Federated Learning With Non-IID Data via Local Drift Decoupling and Correction 10102
Liang Gao (Nathonal University of Defense Technology), Huazhu Fu (IHPC, ASTAR), Li Li (University of Macau), Yingwen Chen (National University of Defense Technology), Ming Xu (National University of Defense Technology), and Cheng-Zhong Xu (University of Macau)

Differentially Private Federated Learning With Local Regularization and Sparsification 10112
Anda Cheng (CASIA), Peisong Wang (Institute of Automation, Chinese Academy of Sciences), Xi Sheryl Zhang (Institute of Automation, Chinese Academy of Sciences), and Jian Cheng (Chinese Academy of Sciences, China)

Auditing Privacy Defenses in Federated Learning via Generative Gradient Leakage 10122
Zhuohang Li (The University of Tennessee, Knoxville), Jiaxin Zhang (Oak Ridge National Laboratory), Luyang Liu (Google Research), and Jian Liu (The University of Tennessee, Knoxville)

Learn From Others and Be Yourself in Heterogeneous Federated Learning 10133
Wenke Huang (Wuhan University), Mang Ye (Wuhan University), and Bo Du (Wuhan University)

RSCFed: Random Sampling Consensus Federated Semi-Supervised Learning 10144
Xiaoxiao Liang (The Hong Kong University of Science and Technology), Yiqun Lin (The Hong Kong University of Science and Technology), Huazhu Fu (IHPC, ASTAR), Lei Zhu (The Hong Kong University of Science and Technology (Guangzhou)), and Xiaomeng Li (The Hong Kong University of Science and Technology)

Federated Class-Incremental Learning 10154
Jiahua Dong (Shenyang Institute of Automation, Chinese Academy of Sciences), Lixu Wang (Northwestern University), Zhen Fang (University of Technology Sydney), Gan Sun (Chinese Academy of Sciences), Shichao Xu (Northwestern University), Xiao Wang (Northwestern University), and Qi Zhu (Northwestern University)

Fine-Tuning Global Model via Data-Free Knowledge Distillation for Non-IID Federated Learning 10164
Lin Zhang (Peking University), Li Shen (JD Explore Academy), Liang Ding (University of Sydney), Dacheng Tao (JD.com), and Ling-Yu Duan (Peking University)

FedCorr: Multi-Stage Federated Learning for Label Noise Correction 10174
Jingyi Xu (Singapore University of Technology and Design), Zihan Chen (Singapore University of Technology and Design), Tony Q.S. Quek (Singapore University of Technology and Design), and Kai Fong Ernest Chong (Singapore University of Technology and Design)

ResSFL: A Resistance Transfer Framework for Defending Model Inversion Attack in Split Federated Learning 10184
Jingtao Li (Arizona State University), Adnan Siraj Rakin (Arizona State University), Xing Chen (Arizona State University), Zhezhi He (Shanghai Jiao Tong University), Deliang Fan (Arizona State University), and Chaitali Chakrabarti (Arizona State University)

Poster 2.2: Explainable Computer Vision

Cycle-Consistent Counterfactuals by Latent Transformations 10193
Saeed Khorram (Oregon State University) and Li Fuxin (Oregon State University)

Consistent Explanations by Contrastive Learning 10203
Vipin Pillai (UMBC), Soroush Abbasi Koohpayegani (University of Maryland Baltimore County), Ashley Ouligian (Northrop Grumman), Dennis Fong (Northrop Grumman), and Hamed Pirsiavash (University of California Davis)

Towards Better Understanding Attribution Methods 10213
Sukrut Rao (Max Planck Institute for Informatics), Moritz Böhle (Max-Planck-Institute for Informatics), and Bernt Schiele (MPI Informatics)

Proto2Proto: Can You Recognize the Car, the Way I Do? 10223
Monish Keswani (Indian institute of Science), Sriranjani Ramakrishnan (NA), Nishant Reddy (IIT Hyderabad), and Vineeth N Balasubramanian (Indian Institute of Technology, Hyderabad)

Do Explanations Explain? Model Knows Best 10234
Ashkan Khakzar (Technical University of Munich), Pedram Khorsandi (Sharif University of Science and Technology), Rozhin Nobahari (University of Montreal), and Nassir Navab (TU Munich, Germany)

HINT: Hierarchical Neuron Concept Explainer 10244
Andong Wang (The University of Hong Kong), Wei-Ning Lee (HKU), and Xiaojuan Qi (The University of Hong Kong)

Deformable ProtoPNet: An Interpretable Image Classifier Using Deformable Prototypes 10255
Jon Donnelly (University of Maine), Alina Jade Barnett (Duke University), and Chaofan Chen (University of Maine)

What Do Navigation Agents Learn About Their Environment? 10266
Kshitij Dwivedi (Goethe University, Frankfurt am Main, Germany), Gemma Roig (Goethe University Frankfurt), Aniruddha Kembhavi (Allen Institute for Artificial Intelligence), and Roozbeh Mottaghi (Allen Institute for AI)

A Framework for Learning Ante-Hoc Explainable Models via Concepts 10276
Anirban Sarkar (IIT Hyderabad), Deepak Vijaykeerthy (IBM Research), Anindya Sarkar (Washington University in St. Louis), and Vineeth N Balasubramanian (Indian Institute of Technology, Hyderabad)

Exploiting Explainable Metrics for Augmented SGD 10286
Mahdi S. Hosseini (University of New Brunswick), Mathieu Tuli (University of Toronto), and Konstantinos N. Plataniotis (UofT)

FAM: Visual Explanations for the Feature Representations From Deep Convolutional Networks 10297
Yuxi Wu (Hikvision Research Instiute), Changhuai Chen (Hikvision Research Institute), Jun Che (Hikvision Research Institute), and Shiliang Pu (Hikvision Research Institute)

Interactive Disentanglement: Learning Concepts by Interacting With Their Prototype Representations 10307
Wolfgang Stammer (TU Darmstadt), Marius Memmel (Technical University of Darmstadt), Patrick Schramowski (TU Darmstadt), and Kristian Kersting (TU Darmstadt)

B-Cos Networks: Alignment Is All We Need for Interpretability 10319
Moritz Böhle (Max-Planck-Institute for Informatics), Mario Fritz (CISPA Helmholtz Center for Information Security), and Bernt Schiele (MPI Informatics)

The Flag Median and FlagIRLS 10329
Nathan Mankovich (Colorao State University), Emily J. King (Colorado State University), Chris Peterson (Colorado State University), and Michael Kirby (Colorado State University)

Poster 2.2: Transparency, Fairness, Accountability, Privacy & Ethics in Vision

Learning Fair Classifiers With Partially Annotated Group Labels 10338
Sangwon Jung (Seoul National University), Sanghyuk Chun (NAVER AI Lab), and Taesup Moon (Seoul National University)

Estimating Structural Disparities for Face Models 10348
Shervin Ardeshir (Netflix), Cristina Segalin (Netflix), and Nathan Kallus (Cornell University)

Estimating Example Difficulty Using Variance of Gradients 10358
Chirag Agarwal (Adobe), Daniel D'souza (ProQuest LLC), and Sara Hooker (Google)

Fairness-Aware Adversarial Perturbation Towards Bias Mitigation for Deployed Deep Models .. 10369
Zhibo Wang (Zhejiang University), Xiaowei Dong (Wuhan University), Henry Xue (Ant Group), Zhifei Zhang (Adobe Research), Weifeng Chiu (Antgroup), Tao Wei (Ant Group), and Kui Ren (Zhejiang University)

Fair Contrastive Learning for Facial Attribute Classification 10379
Sungho Park (yonsei university), Jewook Lee (Yonsei University), Pilhyeon Lee (Yonsei University), Sunhee Hwang (LG Uplus), Dohyung Kim (SK Inc.), and Hyeran Byun (Yonsei University)

Leveraging Adversarial Examples To Quantify Membership Information Leakage 10389
Ganesh Del Grosso (INRIA), Hamid Jalalzai (Telecom ParisTech), Georg Pichler (TU Wien), Catuscia Palamidessi (Laboratoire d'informatique de l'École polytechnique), and Pablo Piantanida (CNRS Université Paris-Saclay)

Leveling Down in Computer Vision: Pareto Inefficiencies in Fair Deep Classifiers 10400
Dominik Zietlow (Max Planck Institute for Intelligent Systems), Michael Lohaus (University of Tübingen), Guha Balakrishnan (Amazon Web Services (AWS), ,), Matthäus Kleindessner (Amazon), Francesco Locatello (Amazon Lablet), Bernhard Schölkopf (Amazon / MPI Intelligent Systems), and Chris Russell (Amazon)

Deep Unlearning via Randomized Conditionally Independent Hessians 10412
Ronak Mehta (University of Wisconsin-Madison), Sourav Pal (University of Wisconsin - Madison), Vikas Singh (University of Wisconsin Madison), and Sathya N. Ravi (University of Illinois at Chicago)

Equivariance Allows Handling Multiple Nuisance Variables When Analyzing Pooled Neuroimaging Datasets 10422
Vishnu Suresh Lokhande (University of Wisconsin-Madison), Rudrasis Chakraborty (Butlr), Sathya N. Ravi (University of Illinois at Chicago), and Vikas Singh (University of Wisconsin Madison)

A Study on the Distribution of Social Biases in Self-Supervised Learning Visual Models 10432
Kirill Sirotkin (Universidad Autónoma de Madrid), Pablo Carballeira (Universidad Autonoma de Madrid), and Marcos Escudero-Viñolo (Universidad Autónoma de Madrid)

Poster 2.2: Vision & X

Cross-Modal Perceptionist: Can Face Geometry Be Gleaned From Voices? 10442
Cho-Ying Wu (University of Southern California), Chin-Cheng Hsu (University of Southern California), and Ulrich Neumann (USC)

Learning Hierarchical Cross-Modal Association for Co-Speech Gesture Generation 10452
Xian Liu (The Chinese University of Hong Kong), Qianyi Wu (Monash University), Hang Zhou (The Chinese University of Hong Kong), Yinghao Xu (Chinese University of Hong Kong), Rui Qian (The Chinese University of Hong Kong), Xinyi Lin (Zhejiang University), Xiaowei Zhou (Zhejiang University), Wayne Wu (SenseTime Research), Bo Dai (Shanghai AI Lab), and Bolei Zhou (UCLA)

SEEG: Semantic Energized Co-Speech Gesture Generation 10463
Yuanzhi Liang (UTS), Qianyu Feng (UTS), Linchao Zhu (University of Technology, Sydney), Li Hu (Alibaba), Pan Pan (Alibaba Group), and Yi Yang (UTS)

Mix and Localize: Localizing Sound Sources in Mixtures 10473
Xixi Hu (University of Texas at Austin), Ziyang Chen (University of Michigan), and Andrew Owens (University of Michigan)

Reading To Listen at the Cocktail Party: Multi-Modal Speech Separation 10483
Akam Rahimi (University of Oxford), Triantafyllos Afouras (University of Oxford), and Andrew Zisserman (University of Oxford)

IntentVizor: Towards Generic Query Guided Interactive Video Summarization 10493
Guande Wu (New York University), Jianzhe Lin (University of British Columbia), and Claudio T. Silva (NYU)

M3L: Language-Based Video Editing via Multi-Modal Multi-Level Transformers 10503
Tsu-Jui Fu (UCSB), Xin Eric Wang (University of California, Santa Cruz), Scott T. Grafton (UCSB), Miguel P. Eckstein (University of California, Santa Barbara), and William Yang Wang (UC Santa Barbara)

Finding Fallen Objects via Asynchronous Audio-Visual Integration 10513
Chuang Gan (MIT-IBM Watson AI Lab), Yi Gu (Shanghai Jiao Tong University), Siyuan Zhou (mit/ibm), Jeremy Schwartz (MIT), Seth Alter (MIT), James Traer (Massachusetts Institute of Technology), Dan Gutfreund (IBM), Joshua B. Tenenbaum (MIT), Josh H. McDermott (Massachusetts Institute of Technology), and Antonio Torralba (MIT)

Weakly Paired Associative Learning for Sound and Image Representations via Bimodal Associative Memory 10524
Sangmin Lee (KAIST), Hyung-Il Kim (ETRI), and Yong Man Ro (KAIST)

Egocentric Deep Multi-Channel Audio-Visual Active Speaker Localization 10534
Hao Jiang (Facebook), Calvin Murdock (Facebook), and Vamsi Krishna Ithapu (Facebook Reality Labs)

Audio-Visual Generalised Zero-Shot Learning With Cross-Modal Attention and Language 10543
Otniel-Bogdan Mercea (University of Tübingen), Lukas Riesch (University of Tübingen), A. Sophia Koepke (University of Tübingen), and Zeynep Akata (University of Tübingen)

It's Time for Artistic Correspondence in Music and Video 10554
Dídac Surís (Columbia University), Carl Vondrick (Columbia University), Bryan Russell (Adobe Research), and Justin Salamon (Adobe Research)

Self-Supervised Object Detection From Audio-Visual Correspondence 10565
Triantafyllos Afouras (University of Oxford), Yuki M. Asano (University of Amsterdam), Francois Fagan (Facebook), Andrea Vedaldi (Oxford University), and Florian Metze (Facebook AI)

More Than Words: In-the-Wild Visually-Driven Prosody for Text-to-Speech 10577
Michael Hassid (Google), Michelle Tadmor Ramanovich (Google), Brendan Shillingford (Deepmind), Miaosen Wang (DeepMind), Ye Jia (Google), and Tal Remez (Google)

ObjectFolder 2.0: A Multisensory Object Dataset for Sim2Real Transfer 10588
Ruohan Gao (Stanford University), Zilin Si (Carnegie Mellon University), Yen-Yu Chang (Stanford University), Samuel Clarke (Stanford University), Jeannette Bohg (Stanford), Li Fei-Fei (Stanford University), Wenzhen Yuan, and Jiajun Wu (Stanford University)

A Probabilistic Graphical Model Based on Neural-Symbolic Reasoning for Visual Relationship Detection 10599
Dongran Yu (jilin university), Bo Yang (Jilin University), Qianhao Wei (Jilin University), Anchen Li (Jilin University), and Shirui Pan (Monash University)

Oral 3.1.1: Image & Video Synthesis and Generation (I)

Diffusion Autoencoders: Toward a Meaningful and Decodable Representation 10609
Konpat Preechakul (Vidyasirimedhi Institute of Science and Technology), Nattanat Chatthee (Vidyasirimedhi Institute of Science and Technology), Suttisak Wizadwongsa (Vidyasirimedhi Institute of Science and Technology), and Supasorn Suwajanakorn (Vidyasirimedhi Institute of Science and Technology)

Polymorphic-GAN: Generating Aligned Samples Across Multiple Domains With Learned Morph Maps 10620
Seung Wook Kim (University of Toronto, NVIDIA), Karsten Kreis (NVIDIA), Daiqing Li (NVIDIA), Antonio Torralba (MIT), and Sanja Fidler (University of Toronto, NVIDIA)

Polarity Sampling: Quality and Diversity Control of Pre-Trained Generative Networks via Singular Values 10631
Ahmed Imtiaz Humayun (Rice University), Randall Balestriero (Facebook AI Research), and Richard Baraniuk (Rice University)

Ensembling Off-the-Shelf Models for GAN Training 10641
Nupur Kumari (Carnegie Mellon University), Richard Zhang (Adobe), Eli Shechtman (Adobe Research, US), and Jun-Yan Zhu (Carnegie Mellon University)

Marginal Contrastive Correspondence for Guided Image Generation 10653
Fangneng Zhan (Nanyang Technological University), Yingchen Yu (Nanyang Technological University), Rongliang Wu (Nanyang Technological University), Jiahui Zhang (Nanyang Technological University), Shijian Lu (Nanyang Technological University), and Changgong Zhang (JD)

GRAM: Generative Radiance Manifolds for 3D-Aware Image Generation 10663
Yu Deng (Tsinghua University), Jiaolong Yang (Microsoft Research), Jianfeng Xiang (University of Science and Technology of China), and Xin Tong (Microsoft)

High-Resolution Image Synthesis With Latent Diffusion Models 10674
Robin Rombach (Heidelberg University), Andreas Blattmann (Heidelberg University), Dominik Lorenz (Heidelberg University), Patrick Esser (University Heidelberg), and Björn Ommer (University of Munich)

Vector Quantized Diffusion Model for Text-to-Image Synthesis 10686
Shuyang Gu (University of Science and Technology of China), Dong Chen (Microsoft Research Asia), Jianmin Bao (Microsoft Research Asia), Fang Wen (Microsoft Research Asia), Bo Zhang (hkust), Dongdong Chen (Microsoft Cloud AI), Lu Yuan (Microsoft), and Baining Guo (MSR Asia)

ManiTrans: Entity-Level Text-Guided Image Manipulation via Token-Wise Semantic Alignment and Generation 10697
Jianan Wang (Fudan University), Guansong Lu (Huawei Noah's Ark Lab), Hang Xu (Huawei Noah's Ark Lab), Zhenguo Li (Huawei Noah's Ark Lab), Chunjing Xu (Huawei Noah's Ark Lab), and Yanwei Fu (Fudan University)

Dataset Distillation by Matching Training Trajectories 10708
George Cazenavette (Carnegie Mellon University), Tongzhou Wang (MIT), Antonio Torralba (MIT), Alexei A. Efros (UC Berkeley), and Jun-Yan Zhu (Carnegie Mellon University)

Continual Predictive Learning From Videos 10718
Geng Chen (Shanghai Jiao tong University), Wendong Zhang (Shanghai Jiao Tong University), Han Lu (Shanghai Jiao Tong University), Siyu Gao (Shanghai JiaoTong University), Yunbo Wang (Shanghai Jiao Tong University), Mingsheng Long (Tsinghua University), and Xiaokang Yang (Shanghai Jiao Tong University of China)

Motion-Adjustable Neural Implicit Video Representation 10728
Long Mai (Portland State University) and Feng Liu (Portland State University)

Splicing ViT Features for Semantic Appearance Transfer 10738
Narek Tumanyan (Weizmann Institute of Science), Omer Bar-Tal (Weizmann Institute of Science), Shai Bagon (Weizmann Institute of Science), and Tali Dekel (Weizmann Institute of Science)

MAT: Mask-Aware Transformer for Large Hole Image Inpainting 10748
Wenbo Li (The Chinese University of Hong Kong), Zhe Lin (Adobe Research), Kun Zhou (CUHK-Shenzhen), Lu Qi (The Chinese University of Hong Kong), Yi Wang (Shanghai AI Laboratory), and Jiaya Jia (Chinese University of Hong Kong)

Day-to-Night Image Synthesis for Training Nighttime Neural ISPs 10759
Abhijith Punnappurath (Samsung AI Center Toronto), Abdullah Abuolaim (York University), Abdelrahman Abdelhamed (Samsung), Alex Levinshtein (Samsung AI Centre Toronto), and Michael S. Brown (York University)

Smooth-Swap: A Simple Enhancement for Face-Swapping With Smoothness 10769
Jiseob Kim (Seoul National University), Jihoon Lee (Intermodalics BVBA), and Byoung-Tak Zhang (Seoul National University)

Few-Shot Head Swapping in the Wild 10779
Changyong Shu (Baidu. inc), Hemao Wu (South China University of Technology), Hang Zhou (The Chinese University of Hong Kong), Jiaming Liu (Baidu Inc.), Zhibin Hong (Baidu Inc.), Changxing Ding (South China University of Technology), Junyu Han (Baidu Inc.), Jingtuo Liu (baidu), Errui Ding (Baidu Inc.), and Jingdong Wang (Baidu)

ClothFormer: Taming Video Virtual Try-On in All Module 10789
Jianbin Jiang (iQIYI, Inc.), Tan Wang (iQIYI, Inc.), He Yan (iQiYi, Inc.), and Junhui Liu (iQIYI Inc)

Oral 3.1.2: Deep Learning Architectures & Techniques

A-ViT: Adaptive Tokens for Efficient Vision Transformer 10799
Hongxu Yin (NVIDIA), Arash Vahdat (NVIDIA), Jose M. Alvarez (NVIDIA), Arun Mallya (NVIDIA), Jan Kautz (NVIDIA), and Pavlo Molchanov (NVIDIA)

MetaFormer Is Actually What You Need for Vision 10809
Weihao Yu (NUS), Mi Luo (SEA AI Lab), Pan Zhou (Sea AI Lab), Chenyang Si (Sea AI Labs), Yichen Zhou (Sea AI Lab), Xinchao Wang (National University of Singapore), Jiashi Feng (ByteDance), and Shuicheng Yan (Sea AI Labs)

Reversible Vision Transformers 10820
Karttikeya Mangalam (UC Berkeley), Haoqi Fan (Facebook AI Research), Yanghao Li (Facebook AI Research), Chao-Yuan Wu (Facebook), Bo Xiong (Facebook AI Research), Christoph Feichtenhofer (Facebook AI Research), and Jitendra Malik (University of California at Berkeley)

Learned Queries for Efficient Local Attention 10831
Moab Arar (Tel Aviv University), Ariel Shamir (The Interdisciplinary Center), and Amit H. Bermano (Tel-Aviv University)

Shunted Self-Attention via Multi-Scale Token Aggregation 10843
Sucheng Ren (South China University of Technology), Daquan Zhou (National University of Singapore), Shengfeng He (South China University of Technology), Jiashi Feng (ByteDance), and Xinchao Wang (National University of Singapore)

Automatic Relation-Aware Graph Network Proliferation 10853
Shaofei Cai (Institute of Computing Technology, Chinese Academy of Sciences), Liang Li (Institute of Computing Technology, Chinese Academy of Sciences), Xinzhe Han (University of Chinese Academy of Sciences), Jiebo Luo (U. Rochester), Zheng-Jun Zha (University of Science and Technology of China), and Qingming Huang (University of Chinese Academy of Sciences)

β-DARTS: Beta-Decay Regularization for Differentiable Architecture Search 10864
Peng Ye (fudan university), Baopu Li (BAIDU USA LLC), Yikang Li (Shanghai AI Lab), Tao Chen (Fudan University), Jiayuan Fan (Fudan University), and Wanli Ouyang (The University of Sydney)

Distribution Consistent Neural Architecture Search 10874
Junyi Pan (Tencent), Chong Sun (Tencent Wechat), Yizhou Zhou (Tencent), Ying Zhang (Tencent), and Chen Li (Tencent)

Training-Free Transformer Architecture Search 10884
Qinqin Zhou (Xiamen University), Kekai Sheng (Youtu Lab, Tencent Inc.), Xiawu Zheng (Xiamen University), Ke Li (Tencent), Xing Sun (Tencent), Yonghong Tian (Peking University), Jie Chen (Peking University), and Rongrong Ji (Xiamen University, China)

TeachAugment: Data Augmentation Optimization Using Teacher Knowledge 10894
Teppei Suzuki (Denso IT Laboratory)

Knowledge Distillation via the Target-Aware Transformer 10905
Sihao Lin (RMIT University), Hongwei Xie (Alibaba), Bing Wang (Alibaba group), Kaicheng Yu (Alibaba Inc.), Xiaojun Chang (University of Technology Sydney), Xiaodan Liang (Sun Yat-sen University), and Gang Wang (Alibaba A.I. Labs)

Knowledge Distillation: A Good Teacher Is Patient and Consistent 10915
Lucas Beyer (Google Brain), Xiaohua Zhai (Google Brain), Amélie Royer (IST Austria), Larisa Markeeva (DeepMind), Rohan Anil (Google Brain), and Alexander Kolesnikov (Google Brain)

An Image Patch Is a Wave: Phase-Aware Vision MLP 10925
Yehui Tang (Peking University), Kai Han (Noah's Ark Lab, Huawei Technologies), Jianyuan Guo (Noah's Ark Lab, Huawei Technologies), Chang Xu (University of Sydney), Yanxi Li (University of Sydney), Chao Xu (Peking University), and Yunhe Wang (Huawei Technologies)

Dynamic MLP for Fine-Grained Image Classification by Leveraging Geographical and Temporal Information 10935
Lingfeng Yang (Nanjing University of Science and Technology), Xiang Li (Nanjing University of Science and Technology), Renjie Song (Megvii Inc.), Borui Zhao (Megvii Technology), Juntian Tao (Nanjing University of Science and Technology), Shihao Zhou (megvii), Jiajun Liang (Megvii), and Jian Yang (Nanjing University of Science and Technology)

Controllable Dynamic Multi-Task Architectures 10945
Dripta S. Raychaudhuri (University of California, Riverside), Yumin Suh (NEC Labs America), Samuel Schulter (NEC Laboratories America), Xiang Yu (NEC Labs), Masoud Faraki (NEC Labs America), Amit K. Roy-Chowdhury (University of California, Riverside), and Manmohan Chandraker (UC San Diego)

Grounded Language-Image Pre-Training 10955
Liunian Harold Li (UCLA), Pengchuan Zhang (Microsoft Research AI), Haotian Zhang (University of Washington), Jianwei Yang (Microsoft Research), Chunyuan Li (Microsoft Research), Yiwu Zhong (University of Wisconsin-Madison), Lijuan Wang (Microsoft), Lu Yuan (Microsoft), Lei Zhang (International Digital Economy Academy (IDEA), ,), Jenq-Neng Hwang (University of Washington), Kai-Wei Chang (University of California, Los Angeles), and Jianfeng Gao (Microsoft Research)

ZZ-Net: A Universal Rotation Equivariant Architecture for 2D Point Clouds 10966
Georg Bökman (Chalmers University of Technology), Fredrik Kahl (Chalmers), and Axel Flinth (Chalmers University of Technology)

CADTransformer: Panoptic Symbol Spotting Transformer for CAD Drawings 10976
Zhiwen Fan (University of Texas at Austin), Tianlong Chen (Unversity of Texas at Austin), Peihao Wang (University of Texas at Austin), and Zhangyang Wang (University of Texas at Austin)

Oral 3.1.3: Human Pose Estimation & Tracking, Localization, and Object Pose Estimation

Adversarial Parametric Pose Prior 10987
Andrey Davydov (EPFL), Anastasia Remizova (School of Computer and Communication Sciences, EPFL), Victor Constantin (EPFL), Sina Honari (EPFL), Mathieu Salzmann (EPFL), and Pascal Fua (EPFL, Switzerland)

Temporal Feature Alignment and Mutual Information Maximization for Video-Based Human Pose Estimation 10996
Zhenguang Liu (Zhejiang University), Runyang Feng (Zhejiang Gongshang University), Haoming Chen (Zhejiang Gongshang University), Shuang Wu (Nanyang Technological University), Yixing Gao (Jilin University), Yunjun Gao (Zhejiang University), and Xiang Wang (National University of Singapore)

PoseTriplet: Co-Evolving 3D Human Pose Estimation, Imitation, and Hallucination Under Self-Supervision 11007
Kehong Gong (National University of Singapore), Bingbing Li (Nanyang Technological University), Jianfeng Zhang (NUS), Tao Wang (national university of singapore), Jing Huang (Huawei International Pte Ltd), Michael Bi Mi (Huawei International Pte Ltd), Jiashi Feng (ByteDance), and Xinchao Wang (National University of Singapore)

Generalizable Human Pose Triangulation 11018
Kristijan Bartol (University of Zagreb Faculty of Electrical Engineering and computing), David Bojanić (University of Zagreb Faculty of Electrical Engineering and computing), Tomislav Petković (University of Zagreb, Faculty of Electrical Engineering and Computing), and Tomislav Pribanić (University of Zagreb Faculty of Electrical Engineering and computing)

GLAMR: Global Occlusion-Aware Human Mesh Recovery With Dynamic Cameras 11028
Ye Yuan (Carnegie Mellon University), Umar Iqbal (NVIDIA Research), Pavlo Molchanov (NVIDIA), Kris Kitani (Carnegie Mellon University), and Jan Kautz (NVIDIA)

Bailando: 3D Dance Generation by Actor-Critic GPT With Choreographic Memory 11040
Li Siyao (Nanyang Technological University), Weijiang Yu (Sun Yat-Sen University), Tianpei Gu (UCLA), Chunze Lin (Sense Time), Quan Wang (Sensetime), Chen Qian (SenseTime), Chen Change Loy (Nanyang Technological University), and Ziwei Liu (Nanyang Technological University)

Contextual Instance Decoupling for Robust Multi-Person Pose Estimation 11050
Dongkai Wang (Peking University) and Shiliang Zhang (Peking University)

End-to-End Multi-Person Pose Estimation With Transformers 11059
Dahu Shi (Hikvision Research Institute), Xing Wei (Xi'an Jiaotong University), Liangqi Li (Hikvision Research Institute), Ye Ren (Hikvision Research Institute), and Wenming Tan (Hikvision Research Institute)

Meta Agent Teaming Active Learning for Pose Estimation 11069
Jia Gong (Singapore University of Technology and Design), Zhipeng Fan (NYU TANDON SCHOOL OF ENGINEERING), Qiuhong Ke (The University of Melbourne), Hossein Rahmani (Lancaster University), and Jun Liu (Singapore University of Technology and Design)

Keypoint Transformer: Solving Joint Identification in Challenging Hands and Object Interactions for Accurate 3D Pose Estimation 11080
Shreyas Hampali (TU Graz), Sayan Deb Sarkar (Mercedes Benz R & D India Pvt. Ltd.), Mahdi Rad (TU Graz), and Vincent Lepetit (Ecole des Ponts ParisTech)

Not All Tokens Are Equal: Human-Centric Visual Analysis via Token Clustering Transformer .. 11091
Wang Zeng (The Chinese University of Hong Kong), Sheng Jin (The University of Hong Kong), Wentao Liu (Sensetime), Chen Qian (SenseTime), Ping Luo (The University of Hong Kong), Wanli Ouyang (The University of Sydney), and Xiaogang Wang (Chinese University of Hong Kong, Hong Kong)

Occlusion-Robust Face Alignment Using a Viewpoint-Invariant Hierarchical Network Architecture 11102
Congcong Zhu (Shanghai University), Xintong Wan (Shanghai University), Shaorong Xie (Shanghai University), Xiaoqiang Li (Shanghai University), and Yinzheng Gu (HYCloud)

LASER: LAtent SpacE Rendering for 2D Visual Localization 11112
Zhixiang Min (Stevens Institute of Technology), Naji Khosravan (Zillow Group), Zachary Bessinger (Zillow Group), Manjunath Narayana (Zillow group), Sing Bing Kang (Zillow Group), Enrique Dunn (Stevens Institute of Technology), and Ivaylo Boyadzhiev (Zillow Group)

Learning To Detect Scene Landmarks for Camera Localization 11122
Tien Do (University of Minnesota), Ondrej Miksik (Microsoft), Joseph DeGol (Microsoft), Hyun Soo Park (The University of Minnesota), and Sudipta N. Sinha (Microsoft)

Geometric Transformer for Fast and Robust Point Cloud Registration 11133
Zheng Qin (National University of Defense Technology), Hao Yu (Technical University of Munich), Changjian Wang (National University Of Defense Technology), Yulan Guo (Sun Yat-sen University), Yuxing Peng (National University of Defense Technology), and Kai Xu (National University of Defense Technology)

ARCS: Accurate Rotation and Correspondence Search 11143
Liangzu Peng (Johns Hopkins University), Manolis C. Tsakiris (ShanghaiTech University), and René Vidal (Johns Hopkins University, USA)

FisherMatch: Semi-Supervised Rotation Regression via Entropy-Based Filtering 11154
Yingda Yin (Peking University), Yingcheng Cai (Peking University), He Wang (Peking University), and Baoquan Chen (Peking University)

Uni6D: A Unified CNN Framework Without Projection Breakdown for 6D Pose Estimation 11164
Xiaoke Jiang (Sensetime Group Limited), Donghai Li (SenseTime), Hao Chen (SenseTime Research), Ye Zheng (Institute of Computing Technology, Chinese Academy of Sciences; University of Chinese Academy of Sciences), Rui Zhao (SenseTime Group Limited), and Liwei Wu (SenseTime Research)

Poster 3.1: Image & Video Synthesis and Generation

OSSGAN: Open-Set Semi-Supervised Image Generation 11175
Kai Katsumata (The University of Tokyo), Duc Minh Vo (The University of Tokyo), and Hideki Nakayama (The University of Tokyo)

Attribute Group Editing for Reliable Few-Shot Image Generation 11184
Guanqi Ding (University of Chinese Academy of Science), Xinzhe Han (University of Chinese Academy of Sciences), Shuhui Wang (VIPL, ICT, Chinese academic of science), Shuzhe Wu (Beijing Huawei Cloud Computing Technologies Co., Ltd.), Xin Jin (Huawei Technologies), Dandan Tu (Huawei), and Qingming Huang (University of Chinese Academy of Sciences)

Few Shot Generative Model Adaption via Relaxed Spatial Structural Alignment 11194
Jiayu Xiao (Institute of Computing Technology, Chinese Academy of Sciences), Liang Li (Institute of Computing Technology, Chinese Academy of Sciences), Chaofei Wang (Tsinghua University), Zheng-Jun Zha (University of Science and Technology of China), and Qingming Huang (University of Chinese Academy of Sciences)

Semantic-Shape Adaptive Feature Modulation for Semantic Image Synthesis 11204
Zhengyao Lv (Harbin Institute of Technology), Xiaoming Li (Harbin Institute of Technology), Zhenxing Niu (Alibaba Group-Machine Intelligence Technology), Bing Cao (Tianjin University), and Wangmeng Zuo (Harbin Institute of Technology, China)

Retrieval-Based Spatially Adaptive Normalization for Semantic Image Synthesis 11214
Yupeng Shi (Harbin Institute of Technology), Xiao Liu (Tencent), Yuxiang Wei (Harbin Institute of Technology), Zhongqin Wu (Tomorrow Advancing Life Education Group), and Wangmeng Zuo (Harbin Institute of Technology, China)

Generative Flows With Invertible Attentions 11224
Rhea Sanjay Sukthanker (ETH Zurich), Zhiwu Huang (Singapore Management University), Suryansh Kumar (ETH Zurich), Radu Timofte (ETH Zurich), and Luc Van Gool (ETH Zurich)

Style-Structure Disentangled Features and Normalizing Flows for Diverse Icon Colorization 11234
Yuan-kui Li (National Yang Ming Chiao Tung University), Yun-Hsuan Lien (National Yang Ming Chiao Tung University), and Yu-Shuen Wang (National Yang Ming Chiao Tung University)

SemanticStyleGAN: Learning Compositional Generative Priors for Controllable Image Synthesis and Editing 11244
Yichun Shi (ByteDance), Xiao Yang (Bytedance AI Lab), Yangyue Wan (ByteDance), and Xiaohui Shen (ByteDance)

Manifold Learning Benefits GANs 11255
Yao Ni (Australian National University), Piotr Koniusz (ANU College of Engineering and Computer Science), Richard Hartley (Australian National University, Australia), and Richard Nock (Data61-CSIRO)

DO-GAN: A Double Oracle Framework for Generative Adversarial Networks 11265
*Aye Phyu Phyu Aung (Nanyang Technological University), Xinrun Wang (Nanyang Technological University), Runsheng Yu (Hong Kong University of Science and Technology), Bo An (Nanyang Technological University), Senthilnath Jayavelu (Institute for Infocomm Research , A*STAR, Singapore), and Xiaoli Li (Institute for Infocomm Research , A*STAR, Singapore/Nanyang Technological University)*

Improving GAN Equilibrium by Raising Spatial Awareness 11275
Jianyuan Wang (Chinese University of Hong Kong), Ceyuan Yang (Chinese University of Hong Kong), Yinghao Xu (Chinese University of Hong Kong), Yujun Shen (Dept. of IE, CUHK), Hongdong Li (Australian National University, Australia), and Bolei Zhou (UCLA)

Feature Statistics Mixing Regularization for Generative Adversarial Networks 11284
Junho Kim (NAVER AI Lab), Yunjey Choi (NAVER AI Lab), and Youngjung Uh (Yonsei University)

StyleSwin: Transformer-Based GAN for High-Resolution Image Generation 11294
Bowen Zhang (University of Science and Technology of China), Shuyang Gu (University of Science and Technology of China), Bo Zhang (hkust), Jianmin Bao (Microsoft Research Asia), Dong Chen (Microsoft Research Asia), Fang Wen (Microsoft Research Asia), Yong Wang (University of Science and Technology of China), and Baining Guo (MSR Asia)

MaskGIT: Masked Generative Image Transformer 11305
Huiwen Chang (Google), Han Zhang (Google), Lu Jiang (Google Research), Ce Liu (Microsoft), and William T. Freeman (Google)

StyTr2: Image Style Transfer With Transformers 11316
Yingying Deng (Institute of AutomationChinese Academy of Sciences), Fan Tang (Jilin University), Weiming Dong (NLPR, Institute of Automation, Chinese Academy of Sciences), Chongyang Ma (Kuaishou Technology), Xingjia Pan (Tencent), Lei Wang (Criminal Investigation Police University of China), and Changsheng Xu (CASIA)

Style Transformer for Image Inversion and Editing 11327
Xueqi Hu (East China Normal University), Qiusheng Huang (East China Normal University), Zhengyi Shi (East China Normal University), Siyuan Li (East China Normal University), Changxin Gao (Huazhong University of Science and Technology), Li Sun (East China Normal University), and Qingli Li (East China Normal University)

Reduce Information Loss in Transformers for Pluralistic Image Inpainting 11337
Qiankun Liu (University of Science and Technology of China), Zhentao Tan (University of Science and Technology of China), Dongdong Chen (Microsoft Cloud AI), Qi Chu (University of Science and Technology of China), Xiyang Dai (Microsoft), Yinpeng Chen (Microsoft), Mengchen Liu (Microsoft), Lu Yuan (Microsoft), and Nenghai Yu (University of Science and Technology of China)

Incremental Transformer Structure Enhanced Image Inpainting With Masking Positional Encoding 11348
Qiaole Dong (Fudan University), Chenjie Cao (Fudan University), and Yanwei Fu (Fudan University)

UniCoRN: A Unified Conditional Image Repainting Network 11359
Jimeng Sun (Beijing University of Posts and Telecommunications), Shuchen Weng (Peking University), Zheng Chang (Beijing University of Posts and Telecommunications), Si Li (Beijing University of Posts and Telecommunications), and Boxin Shi (Peking University)

High-Fidelity GAN Inversion for Image Attribute Editing 11369
Tengfei Wang (HKUST), Yong Zhang (Tencent AI Lab), Yanbo Fan (Tencent AI Lab), Jue Wang (Tencent AI Lab), and Qifeng Chen (HKUST)

HyperInverter: Improving StyleGAN Inversion via Hypernetwork 11379
Tan M. Dinh (VinAI Research), Anh Tuan Tran (VinAI), Rang Nguyen (VinAI Research), and Binh-Son Hua (VinAI Research)

Spatially-Adaptive Multilayer Selection for GAN Inversion and Editing 11389
Gaurav Parmar (Carnegie Mellon University), Yijun Li (Adobe Research), Jingwan Lu (Adobe Research), Richard Zhang (Adobe), Jun-Yan Zhu (Carnegie Mellon University), and Krishna Kumar Singh (Adobe Research)

On Aliased Resizing and Surprising Subtleties in GAN Evaluation 11400
Gaurav Parmar (Carnegie Mellon University), Richard Zhang (Adobe), and Jun-Yan Zhu (Carnegie Mellon University)

Dual-Path Image Inpainting With Auxiliary GAN Inversion 11411
Wentao Wang (Shanghai Jiao Tong University), Li Niu (Shanghai Jiao Tong University), Jianfu Zhang (RIKEN AIP), Xue Yang (Shanghai Jiao Tong University), and Liqing Zhang (Shanghai Jiao Tong University)

InOut: Diverse Image Outpainting via GAN Inversion 11421
Yen-Chi Cheng (Carnegie Mellon University), Chieh Hubert Lin (University of California at Merced), Hsin-Ying Lee (Snap Inc), Jian Ren (Snap Inc), Sergey Tulyakov (Snap Inc.), and Ming-Hsuan Yang (University of California at Merced)

Diverse Plausible 360-Degree Image Outpainting for Efficient 3DCG Background Creation 11431
Naofumi Akimoto (Keio University), Yuhi Matsuo (Keio University), and Yoshimitsu Aoki (Keio University)

Contextual Outpainting With Object-Level Contrastive Learning 11441
Jiacheng Li (University of Science and Technology of China), Chang Chen (Huawei Noah's Ark Lab), and Zhiwei Xiong (University of Science and Technology of China)

RePaint: Inpainting Using Denoising Diffusion Probabilistic Models 11451
Andreas Lugmayr (ETH Zurich), Martin Danelljan (ETH Zurich), Andres Romero (ETH Zürich), Fisher Yu (ETH Zurich), Radu Timofte (ETH Zurich), and Luc Van Gool (ETH Zurich)

Perception Prioritized Training of Diffusion Models 11462
Jooyoung Choi (Seoul National University), Jungbeom Lee (Seoul National University), Chaehun Shin (Seoul National University), Sungwon Kim (Seoul National University), Hyunwoo Kim (LG AI Research), and Sungroh Yoon (Seoul National University)

Dynamic Dual-Output Diffusion Models 11472
Yaniv Benny (Tel Aviv University) and Lior Wolf (Tel Aviv University, Israel)

Generating High Fidelity Data From Low-Density Regions Using Diffusion Models 11482
Vikash Sehwag (Princeton University), Caner Hazirbas (Facebook AI), Albert Gordo (Facebook), Firat Ozgenel (Meta), and Cristian Canton (Facebook AI)

Global Context With Discrete Diffusion in Vector Quantised Modelling for Image Generation 11492
Minghui Hu (Nanyang Technological University), Yujie Wang (Sensetime Research), Tat-Jen Cham (Nanyang Technological University), Jianfei Yang (Nanyang Technological University), and P.N. Suganthan (Nanyang Technological University)

Bridging Global Context Interactions for High-Fidelity Image Completion 11502
Chuanxia Zheng (Monash University), Tat-Jen Cham (Nanyang Technological University), Jianfei Cai (Monash University), and Dinh Phung (Monash University)

Autoregressive Image Generation Using Residual Quantization 11513
Doyup Lee (POSTECH), Chiheon Kim (Kakao Brain), Saehoon Kim (Kakao Brain), Minsu Cho (POSTECH), and Wook-Shin Han (POSTECH)

Arbitrary-Scale Image Synthesis 11523
Evangelos Ntavelis (ETH Zurich & CSEM), Mohamad Shahbazi (ETH Zurich), Iason Kastanis (CSEM SA), Radu Timofte (ETH Zurich), Martin Danelljan (ETH Zurich), and Luc Van Gool (ETH Zurich)

Cluster-Guided Image Synthesis With Unconditional Models 11533
Markos Georgopoulos (Imperial College London), James Oldfield (Queen Mary University of London), Grigorios G. Chrysos (EPFL), and Yannis Panagakis (University of Athens)

Poster 3.1: Segmentation, Grouping and Shape Analysis

Dynamic Prototype Convolution Network for Few-Shot Semantic Segmentation 11543
Jie Liu (University of Amsterdam), Yanqi Bao (Northeastern University), Guo-Sen Xie (Inception Institute of Artificial Intelligence), Huan Xiong (Mohamed bin Zayed University of Artificial Intelligence (MBZUAI), ,), Jan-Jakob Sonke (The Netherlands Cancer Institute), and Efstratios Gavves (University of Amsterdam)

Generalized Few-Shot Semantic Segmentation 11553
Zhuotao Tian (The Chinese University of Hong Kong), Xin Lai (The Chinese University of Hong Kong), Li Jiang (The Chinese University of Hong Kong), Shu Liu (SmartMore), Michelle Shu (Cornell University), Hengshuang Zhao (University of Oxford), and Jiaya Jia (Chinese University of Hong Kong)

Learning Non-Target Knowledge for Few-Shot Semantic Segmentation 11563
Yuanwei Liu (Northwestern Polytechnical University), Nian Liu (Inception Institute of Artificial Intelligence), Qinglong Cao (Northwestern Polytechnical University), Xiwen Yao (Northwestern Polytechnical University), Junwei Han (NWPU, China), and Ling Shao (NCAI/IIAI/MBZUAI)

Decoupling Zero-Shot Semantic Segmentation 11573
Jian Ding (Wuhan University), Nan Xue (Wuhan University), Gui-Song Xia (Wuhan University), and Dengxin Dai (MPI for Informatics)

Class-Balanced Pixel-Level Self-Labeling for Domain Adaptive Semantic Segmentation 11583
Ruihuang Li (Hong Kong Polytechnic University), Shuai Li (The Hong Kong Polytechnic University), Chenhang He (The Hong Kong Polytechnic University), Yabin Zhang (Hong Kong Polytechnic University), Xu Jia (Dalian University of Technology), and Lei Zhang (Hong Kong Polytechnic University, Hong Kong, China)

ContrastMask: Contrastive Learning To Segment Every Thing 11594
Xuehui Wang (Shanghai Jiao Tong University), Kai Zhao (UCLA), Ruixin Zhang (Tencent), Shouhong Ding (Tencent), Yan Wang (East China Normal University), and Wei Shen (Shanghai Jiao Tong University)

The Neurally-Guided Shape Parser: Grammar-Based Labeling of 3D Shape Regions With Approximate Inference 11604
R. Kenny Jones (Brown University), Aalia Habib (Brown University), Rana Hanocka (University of Chicago), and Daniel Ritchie (Brown University)

AutoGPart: Intermediate Supervision Search for Generalizable 3D Part Segmentation 11614
Xueyi Liu (Tsinghua University), Xiaomeng Xu (Tsinghua University), Anyi Rao (The Chinese University of Hong Kong), Chuang Gan (MIT-IBM Watson AI Lab), and Li Yi (Tsinghua University)

APES: Articulated Part Extraction From Sprite Sheets 11625
Zhan Xu (University of Massachusetts Amherst), Matthew Fisher (Adobe Research), Yang Zhou (Adobe Research), Deepali Aneja (Adobe), Rushikesh Dudhat (University of Massachusetts Amherst), Li Yi (Tsinghua University), and Evangelos Kalogerakis (UMass Amherst)

GASP, a Generalized Framework for Agglomerative Clustering of Signed Graphs and Its Application to Instance Segmentation 11635
Alberto Bailoni (European Molecular Biology Laboratory Heidelberg), Constantin Pape (University of Heidelberg), Nathan Hütsch (Heidelberg University), Steffen Wolf (MRC Laboratory of Molecular Biology), Thorsten Beier (None), Anna Kreshuk (European Molecular Biology Laboratory), and Fred A. Hamprecht (Heidelberg Collaboratory for Image Processing)

CycleMix: A Holistic Strategy for Medical Image Segmentation From Scribble Supervision 11646
Ke Zhang (Fudan University) and Xiahai Zhuang (Fudan University)

Cross-Patch Dense Contrastive Learning for Semi-Supervised Segmentation of Cellular Nuclei in Histopathologic Images 11656
Huisi Wu (Shenzhen University), Zhaoze Wang (Shenzhen University), Youyi Song (The Hong Kong Polytechnic University), Lin Yang (Hong Kong Polytechnic University), and Jing Qin (The Hong Kong Polytechnic University)

C-CAM: Causal CAM for Weakly Supervised Semantic Segmentation on Medical Image 11666
Zhang Chen (Xi'an Jiaotong University), Zhiqiang Tian (Xi'an Jiaotong University), Jihua Zhu (Xi'an Jiaotong University), Ce Li (Lanzhou University of Technology), and Shaoyi Du (Xi'an Jiaotong Unviersity)

CRIS: CLIP-Driven Referring Image Segmentation 11676
Zhaoqing Wang (University of Sydney), Yu Lu (Beijing University of Posts and Telecommunications), Qiang Li (Kuaishou Technology), Xunqiang Tao (OPPO Research Institution), Yandong Guo (OPPO Research Institute), Mingming Gong (University of Melbourne), and Tongliang Liu (The University of Sydney)

MatteFormer: Transformer-Based Image Matting via Prior-Tokens 11686
GyuTae Park (Naver Webtoon), SungJoon Son (NAVER WEBTOON), JaeYoung Yoo (Naver Webtoon), SeHo Kim (Naver Webtoon), and Nojun Kwak (Seoul National University)

Boosting Robustness of Image Matting With Context Assembling and Strong Data Augmentation 11697
Yutong Dai (The University of Adelaide), Brian Price (Adobe), He Zhang (Adobe), and Chunhua Shen (University of Adelaide, Australia)

Pyramid Grafting Network for One-Stage High Resolution Saliency Detection 11707
Chenxi Xie (Beihang University), Changqun Xia (Peng Cheng Laboratory), Mingcan Ma (Beihang University), Zhirui Zhao (Beihang University), Xiaowu Chen (Beihang University), and Jia Li (Beihang University)

Multi-Source Uncertainty Mining for Deep Unsupervised Saliency Detection 11717
Yifan Wang (Dalian University of Technology), Wenbo Zhang (Dalian University of Technology), Lijun Wang (Dalian University of Technology), Ting Liu (Alibaba), and Huchuan Lu (Dalian University of Technology)

Modeling Motion With Multi-Modal Features for Text-Based Video Segmentation 11727
Wangbo Zhao (Northwestern Polytechnical University), Kai Wang (National University of Singapore), Xiangxiang Chu (Meituan), Fuzhao Xue (National University of Singapore), Xinchao Wang (National University of Singapore), and Yang You (National University of Singapore)

GAT-CADNet: Graph Attention Network for Panoptic Symbol Spotting in CAD Drawings 11737
Zhaohua Zheng (Technical University of Munich), Jianfang Li (alibaba), Lingjie Zhu (Alibaba Group), Honghua Li (Alibaba A.I. Labs), Frank Petzold (Technical University of Munich), and Ping Tan (Simon Fraser University)

Bending Graphs: Hierarchical Shape Matching Using Gated Optimal Transport 11747
Mahdi Saleh (Technical University of Munich), Shun-Cheng Wu (Technical University of Munich), Luca Cosmo (University of Venice), Nassir Navab (TU Munich, Germany), Benjamin Busam (Technical University of Munich), and Federico Tombari (Google, TU Munich)

CAPRI-Net: Learning Compact CAD Shapes With Adaptive Primitive Assembly 11758
Fenggen Yu (Simon Fraser University), Zhiqin Chen (Simon Fraser University), Manyi Li (Shandong University), Aditya Sanghi (Autodesk; Autodesk AI Lab), Hooman Shayani (Autodesk AI Lab), Ali Mahdavi-Amiri (Simon Fraser University), and Hao Zhang (Simon Fraser University)

RIM-Net: Recursive Implicit Fields for Unsupervised Learning of Hierarchical Shape Structures 11769
Chengjie Niu (National University of Defense Technology), Manyi Li (Shandong University), Kai Xu (National University of Defense Technology), and Hao Zhang (Simon Fraser University)

Discovering Objects That Can Move 11779
Zhipeng Bao (Carnegie Mellon University), Pavel Tokmakov (Toyota Research Institute), Allan Jabri (UC Berkeley), Yu-Xiong Wang (University of Illinois at Urbana-Champaign), Adrien Gaidon (Toyota Research Institute), and Martial Hebert (Carnegie Mellon School of Computer Science)

PatchFormer: An Efficient Point Transformer With Patch Attention 11789
Cheng Zhang (Hangzhou Dianzi Universuty), Haocheng Wan (Hangzhou Dianzi University), Xinyi Shen (University college London), and Zizhao Wu (Hangzhou Dianzi University)

Panoptic-PHNet: Towards Real-Time and High-Precision LiDAR Panoptic Segmentation via Clustering Pseudo Heatmap 11799
Jinke Li (UISEE Technology (Beijing) Co., Ltd., ,), Xiao He (UISEE Technology (Beijing) Co., Ltd., ,), Yang Wen (UISEE Technology (Beijing) Co., Ltd., ,), Yuan Gao (UISEE Technology (Beijing) Co., Ltd., ,), Xiaoqiang Cheng (UISEE Technology (Beijing) Co., Ltd., ,), and Dan Zhang (UISEE Technology (Beijing) Co., Ltd., ,)

SemAffiNet: Semantic-Affine Transformation for Point Cloud Segmentation 11809
Ziyi Wang (Tsinghua University), Yongming Rao (Tsinghua University), Xumin Yu (Tsinghua University), Jie Zhou (Tsinghua University), and Jiwen Lu (Tsinghua University)

An MIL-Derived Transformer for Weakly Supervised Point Cloud Segmentation 11820
Cheng-Kun Yang (National Taiwan University), Ji-Jia Wu (National Yang Ming Chiao Tung University), Kai-Syun Chen (National Yang Ming Chiao Tung University), Yung-Yu Chuang (National Taiwan University), and Yen-Yu Lin (National Yang Ming Chiao Tung University)

Weakly Supervised Segmentation on Outdoor 4D Point Clouds With Temporal Matching and Spatial Graph Propagation 11830
*Hanyu Shi (Nanyang technological university), Jiacheng Wei (Nanyang Technological University), Ruibo Li (Nanyang Technological University), Fayao Liu (Institute for Infocomm Research, A*STAR), and Guosheng Lin (Nanyang Technological University)*

Point2Cyl: Reverse Engineering 3D Objects From Point Clouds to Extrusion Cylinders 11840
Mikaela Angelina Uy (Stanford University), Yen-Yu Chang (Stanford University), Minhyuk Sung (KAIST), Purvi Goel (Brown University), Joseph G. Lambourne (Autodesk AI Lab), Tolga Birdal (Stanford University), and Leonidas J. Guibas (Stanford University)

Poster 3.1: Deep Learning Architectures & Techniques

Demystifying the Neural Tangent Kernel From a Practical Perspective: Can It Be Trusted for Neural Architecture Search Without Training? 11851
Jisoo Mok (Seoul National University), Byunggook Na (Seoul National University), Ji-Hoon Kim (Naver), Dongyoon Han (NAVER AI Lab), and Sungroh Yoon (Seoul National University)

BaLeNAS: Differentiable Architecture Search via the Bayesian Learning Rule 11861
Miao Zhang (Aalborg University), Shirui Pan (Monash University), Xiaojun Chang (University of Technology Sydney), Steven Su (University of Technology Sydney), Jilin Hu (Aalborg University), Gholamreza Reza Haffari (Monash University, Australia), and Bin Yang (Aalborg University)

Arch-Graph: Acyclic Architecture Relation Predictor for Task-Transferable Neural Architecture Search 11871
Minbin Huang (Sun-Yat-sen University), Zhijian Huang (Sun-Yat-sen University), Changlin Li (Monash University), Xin Chen (The University of Hong Kong), Hang Xu (Huawei Noah's Ark Lab), Zhenguo Li (Huawei Noah's Ark Lab), and Xiaodan Liang (Sun Yat-sen University)

Shapley-NAS: Discovering Operation Contribution for Neural Architecture Search 11882
Han Xiao (Tsinghua University), Ziwei Wang (Tsinghua University), Zheng Zhu (Tsinghua University), Jie Zhou (Tsinghua University), and Jiwen Lu (Tsinghua University)

GreedyNASv2: Greedier Search With a Greedy Path Filter 11892
Tao Huang (SenseTime Research), Shan You (SenseTime), Fei Wang (University of Science and Technology of China), Chen Qian (SenseTime), Changshui Zhang (Tsinghua University), Xiaogang Wang (Chinese University of Hong Kong, Hong Kong), and Chang Xu (University of Sydney)

Neural Architecture Search With Representation Mutual Information 11902
Xiawu Zheng (Xiamen University), Xiang Fei (Xiamen University), Lei Zhang (Xiamen University), Chenglin Wu (fuzhi.ai), Fei Chao (Xiamen University), Jianzhuang Liu (Huawei Noah's Ark Lab), Wei Zeng (Peking University, China), Yonghong Tian (Peking University), and Rongrong Ji (Xiamen University, China)

Performance-Aware Mutual Knowledge Distillation for Improving Neural Architecture Search . 11912
Pengtao Xie (UC San Diego) and Xuefeng Du (Unversity of Wisconsin, Madison)

Knowledge Distillation With the Reused Teacher Classifier 11923
Defang Chen (Zhejiang University), Jian-Ping Mei (Zhejiang University of Technology), Hailin Zhang (Zhejiang University), Can Wang (Zhejiang University), Yan Feng (Zhejiang University), and Chun Chen (Zhejiang university)

Self-Distillation From the Last Mini-Batch for Consistency Regularization 11933
Yiqing Shen (Shanghai Jiao Tong University), Liwu Xu (OPPO Research Institute), Yuzhe Yang (OPPO Research Institute), Yaqian Li (OPPO Research Institute), and Yandong Guo (OPPO Research Institute)

Decoupled Knowledge Distillation 11943
Borui Zhao (Megvii Technology), Quan Cui (Waseda University), Renjie Song (Megvii Inc.), Yiyu Qiu (Tsinghua University), and Jiajun Liang (Megvii)

Scaling Up Your Kernels to 31x31: Revisiting Large Kernel Design in CNNs 11953
Xiaohan Ding (Tsinghua University), Xiangyu Zhang (Megvii Technology), Jungong Han (Aberystwyth University), and Guiguang Ding (Tsinghua University, China)

A ConvNet for the 2020s 11966
Zhuang Liu (UC Berkeley), Hanzi Mao (Facebook AI Research), Chao-Yuan Wu (Facebook), Christoph Feichtenhofer (Facebook AI Research), Trevor Darrell (UC Berkeley), and Saining Xie (Facebook AI Research)

Beyond Fixation: Dynamic Window Visual Transformer 11977
Pengzhen Ren (NorthWest University), Changlin Li (Monash University), Guangrun Wang (University of Oxford), Yun Xiao (NorthWest University), Qing Du (South China Univercity of Technology), Xiaodan Liang (Sun Yat-sen University), and Xiaojun Chang (University of Technology Sydney)

Lite Vision Transformer With Enhanced Self-Attention 11988
Chenglin Yang (Johns Hopkins University), Yilin Wang (Adobe), Jianming Zhang (Adobe Research), He Zhang (Adobe), Zijun Wei (Adobe), Zhe Lin (Adobe Research), and Alan Yuille (Johns Hopkins University)

Swin Transformer V2: Scaling Up Capacity and Resolution 11999
Ze Liu (USTC), Han Hu (Microsoft Research Asia), Yutong Lin (Xi'an Jiaotong University), Zhuliang Yao (Tsinghua University), Zhenda Xie (Tsinghua University), Yixuan Wei (Tsinghua University), Jia Ning (Huazhong University of Science and Technology), Yue Cao (Microsoft Research), Zheng Zhang (MSRA), Li Dong (Microsoft Research), Furu Wei (Microsoft Research Asia), and Baining Guo (MSR Asia)

The Principle of Diversity: Training Stronger Vision Transformers Calls for Reducing All Levels of Redundancy 12010
Tianlong Chen (Unversity of Texas at Austin), Zhenyu Zhang (University of Science and Technology of China), Yu Cheng (Microsoft Research), Ahmed Awadallah (Microsoft), and Zhangyang Wang (University of Texas at Austin)

MulT: An End-to-End Multitask Learning Transformer 12021
Deblina Bhattacharjee (EPFL), Tong Zhang (EPFL), Sabine Süsstrunk (EPFL), and Mathieu Salzmann (EPFL)

Towards Robust Vision Transformer 12032
Xiaofeng Mao (Alibaba Group), Gege Qi (Alibaba), Yuefeng Chen (Alibaba Group), Xiaodan Li (Alibaba Group), Ranjie Duan (Swinburne University of Technology), Shaokai Ye (EPFL), Yuan He (Alibaba Group), and Hui Xue (Alibaba)

DearKD: Data-Efficient Early Knowledge Distillation for Vision Transformers 12042
Xianing Chen (ShanghaiTech University), Qiong Cao (JD.com), Yujie Zhong (University of Oxford), Jing Zhang (The University of Sydney), Shenghua Gao (Shanghaitech University), and Dacheng Tao (JD.com)

MSG-Transformer: Exchanging Local Spatial Information by Manipulating Messenger Tokens .. 12053
Jiemin Fang (Huazhong University of Science and Technology), Lingxi Xie (Huawei Inc.), Xinggang Wang (Huazhong University of Science and Technology), Xiaopeng Zhang (Huawei Cloud EI), Wenyu Liu (Huazhong University of Science and Technology), and Qi Tian (Huawei Cloud & AI)

NomMer: Nominate Synergistic Context in Vision Transformer for Visual Recognition 12063
Hao Liu (Tencent), Xinghua Jiang (Tencent), Xin Li (Tencent), Zhimin Bao (Tencent), Deqiang Jiang (Tencent), and Bo Ren (Tencent)

TopFormer: Token Pyramid Transformer for Mobile Semantic Segmentation 12073
Wenqiang Zhang (Huazhong University of Science and Technology), Zilong Huang (Tencent), Guozhong Luo (Tencent), Tao Chen (Fudan University), Xinggang Wang (Huazhong University of Science and Technology), Wenyu Liu (Huazhong University of Science and Technology), Gang Yu (Tencent), and Chunhua Shen (University of Adelaide, Australia)

Multi-Scale High-Resolution Vision Transformer for Semantic Segmentation 12084
Jiaqi Gu (University of Texas at Austin), Hyoukjun Kwon (Facebook), Dilin Wang (Facebook), Wei Ye (Facebook), Meng Li (Facebook Inc), Yu-Hsin Chen (Meta), Liangzhen Lai (Facebook Inc), Vikas Chandra (Facebook), and David Z. Pan (University of Texas at Austin)

Scaling Vision Transformers 12094
Xiaohua Zhai (Google Brain), Alexander Kolesnikov (Google Brain), Neil Houlsby (Google), and Lucas Beyer (Google Brain)

Bridged Transformer for Vision and Point Cloud 3D Object Detection 12104
Yikai Wang (Tsinghua University), TengQi Ye (ByteDance), Lele Cao (Tsinghua University), Wenbing Huang (Tsinghua University), Fuchun Sun (Tsinghua University), Fengxiang He (JD.com), and Dacheng Tao (JD.com)

CSWin Transformer: A General Vision Transformer Backbone With Cross-Shaped Windows 12114
Xiaoyi Dong (University of Science and Technology of China), Jianmin Bao (Microsoft Research Asia), Dongdong Chen (Microsoft Cloud AI), Weiming Zhang (University of Science and Technology of China), Nenghai Yu (University of Science and Technology of China), Lu Yuan (Microsoft), Dong Chen (Microsoft Research Asia), and Baining Guo (MSR Asia)

TransMix: Attend To Mix for Vision Transformers 12125
Jie-Neng Chen (Johns Hopkins University), Shuyang Sun (University of Oxford), Ju He (Johns Hopkins University), Philip H.S. Torr (University of Oxford), Alan Yuille (Johns Hopkins University), and Song Bai (University of Oxford)

MiniViT: Compressing Vision Transformers With Weight Multiplexing 12135
Jinnian Zhang (University of Wisconsin Madison), Houwen Peng (Microsoft Research), Kan Wu (Sun Yat-sen University), Mengchen Liu (Microsoft), Bin Xiao (Microsoft), Jianlong Fu (Microsoft Research), and Lu Yuan (Microsoft)

Fine-Tuning Image Transformers Using Learnable Memory 12145
Mark Sandler (Google), Andrey Zhmoginov (Google), Max Vladymyrov (Google), and Andrew Jackson (Google)

Patch Slimming for Efficient Vision Transformers 12155
Yehui Tang (Peking University), Kai Han (Noah's Ark Lab, Huawei Technologies), Yunhe Wang (Huawei Technologies), Chang Xu (University of Sydney), Jianyuan Guo (Noah's Ark Lab, Huawei Technologies), Chao Xu (Peking University), and Dacheng Tao (The University of Sydney)

CMT: Convolutional Neural Networks Meet Vision Transformers 12165
Jianyuan Guo (Noah's Ark Lab, Huawei Technologies), Kai Han (Huawei Noah's Ark Lab), Han Wu (University of Sydney), Yehui Tang (Peking University), Xinghao Chen (Huawei Noah's Ark Lab), Yunhe Wang (Huawei Technologies), and Chang Xu (University of Sydney)

Multimodal Token Fusion for Vision Transformers 12176
Yikai Wang (Tsinghua University), Xinghao Chen (Huawei Noah's Ark Lab), Lele Cao (Tsinghua University), Wenbing Huang (Tsinghua University), Fuchun Sun (Tsinghua University), and Yunhe Wang (Huawei Technologies)

Poster 3.1: Efficient Learning & Inference

CAFE: Learning To Condense Dataset by Aligning Features 12186
Kai Wang (National University of Singapore), Bo Zhao (The University of Edinburgh), Xiangyu Peng (Tsinghua University), Zheng Zhu (National University of Singapore), Shuo Yang (University of Technology Sydney), Shuo Wang (Alibaba Group), Guan Huang (Institute of Automation, Chinese Academy of Sciences), Hakan Bilen (University of Edinburgh), Xinchao Wang (National University of Singapore), and Yang You (National University of Singapore)

Lite-MDETR: A Lightweight Multi-Modal Detector 12196
Qian Lou (Samsung Research America), Yen-Chang Hsu (Samsung Research America), Burak Uzkent (Stanford University), Ting Hua (Samsung Research America), Yilin Shen (Samsung Research America), and Hongxia Jin (Samsung Research America)

DeeCap: Dynamic Early Exiting for Efficient Image Captioning 12206
Zhengcong Fei (Chinese Academy of Sciences, Institute of Computing Technology), Xu Yan (VIPL, ICT, Chinese Academy of Sciences), Shuhui Wang (VIPL, ICT, Chinese academic of science), and Qi Tian (Huawei Cloud & AI)

Searching the Deployable Convolution Neural Networks for GPUs 12217
Linnan Wang (Brown University), Chenhan Yu (NVIDIA), Satish Salian (NVIDIA), Slawomir Kierat (NVIDIA), Szymon Migacz (NVIDIA), and Alex Fit Florea (NVIDIA)

Active Learning by Feature Mixing 12227
Amin Parvaneh (University of Adelaide), Ehsan Abbasnejad (The University of Adelaide), Damien Teney (University of Adelaide), Gholamreza Reza Haffari (Monash University, Australia), Anton van den Hengel (University of Adelaide), and Javen Qinfeng Shi (University of Adelaide)

When To Prune? A Policy Towards Early Structural Pruning 12237
Maying Shen (NVIDIA), Pavlo Molchanov (NVIDIA), Hongxu Yin (NVIDIA), and Jose M. Alvarez (NVIDIA)

Contrastive Dual Gating: Learning Sparse Features With Contrastive Learning 12247
Jian Meng (Arizona State University), Li Yang (Arizona State University), Jinwoo Shin (KAIST), Deliang Fan (Arizona State University), and Jae-sun Seo (Arizona State University)

How Well Do Sparse ImageNet Models Transfer? 12256
Eugenia Iofinova (IST Austria), Alexandra Peste (Institute of Science and Technology Austria), Mark Kurtz (Neural Magic), and Dan Alistarh (IST Austria & NeuralMagic)

Rep-Net: Efficient On-Device Learning via Feature Reprogramming 12267
Li Yang (Arizona State University), Adnan Siraj Rakin (Arizona State University), and Deliang Fan (Arizona State University)

CHEX: CHannel EXploration for CNN Model Compression 12277
Zejiang Hou (Princeton University), Minghai Qin (Western Digital Research), Fei Sun (Alibaba DAMO Academy), Xiaolong Ma (Northeastern University), Kun Yuan (Alibaba Group (US), ,), Yi Xu (Alibaba Group), Yen-Kuang Chen (Alibaba DAMO Academy), Rong Jin (alibaba group), Yuan Xie (Alibaba DAMO Academy), and Sun-Yuan Kung (Princeton University)

HODEC: Towards Efficient High-Order DEcomposed Convolutional Neural Networks 12289
Miao Yin (Rutgers University), Yang Sui (Rutgers University), Wanzhao Yang (Rutgers University), Xiao Zang (Rutgers University, New Brunswick), Yu Gong (Rutgers University), and Bo Yuan (rutgers university)

AdaViT: Adaptive Vision Transformers for Efficient Image Recognition 12299
Lingchen Meng (Fudan University), Hengduo Li (University of Maryland, College Park), Bor-Chun Chen (Facebook AI), Shiyi Lan (University of Maryland), Zuxuan Wu (UMD), Yu-Gang Jiang (Fudan University), and Ser-Nam Lim (Meta AI)

Cross-Image Relational Knowledge Distillation for Semantic Segmentation 12309
Chuanguang Yang (Institute of Computing Technology, Chinese Academy of Sciences), Helong Zhou (Beijing Horizon Information Technology Co., Ltd), Zhulin An (Institute of Computing Technology, Chinese Academy of Sciences), Xue Jiang (Wuhan University), Yongjun Xu (Institute of Computing Technology, Chinese Academy of Sciences), and Qian Zhang (Horizon Robotics)

Mr.BiQ: Post-Training Non-Uniform Quantization Based on Minimizing the Reconstruction Error 12319
Yongkweon Jeon (Samsung Research), Chungman Lee (Samsung Research), Eulrang Cho (Samsung Research), and Yeonju Ro (Samsung Research)

IntraQ: Learning Synthetic Images With Intra-Class Heterogeneity for Zero-Shot Network Quantization 12329
Yunshan Zhong (xiamen university), Mingbao Lin (Xiamen University, China), Gongrui Nan (Xmu), Jianzhuang Liu (Huawei Noah's Ark Lab), Baochang Zhang (Beihang University), Yonghong Tian (Peking University), and Rongrong Ji (Xiamen University, China)

DECORE: Deep Compression With Reinforcement Learning 12339
Manoj Alwani (Element Inc), Yang Wang (Element), and Vashisht Madhavan (Element Inc.)

Towards Efficient and Scalable Sharpness-Aware Minimization 12350
Yong Liu (National University of Singapore), Siqi Mai (National University of Singapore), Xiangning Chen (University of California, Los Angeles), Cho-Jui Hsieh (UCLA), and Yang You (National University of Singapore)

AEGNN: Asynchronous Event-Based Graph Neural Networks 12361
Simon Schaefer (TUM), Daniel Gehrig (University of Zurich & ETH Zurich), and Davide Scaramuzza (University of Zurich & ETH Zurich, Switzerland)

DiSparse: Disentangled Sparsification for Multitask Model Compression 12372
Xinglong Sun (University of Illinois at Urbana-Champaign), Ali Hassani (University of Oregon), Zhangyang Wang (University of Texas at Austin), Gao Huang (Tsinghua), and Humphrey Shi (U of Oregon/UIUC)

Multi-Modal Extreme Classification 12383
Anshul Mittal (IIT Delhi), Kunal Dahiya (IIT Delhi), Shreya Malani (Microsoft), Janani Ramaswamy (Microsoft), Seba Kuruvilla (MICROSOFT), Jitendra Ajmera (Microsoft, India), Keng-hao Chang (Microsoft), Sumeet Agarwal (Department of Electrical Engineering, IIT Delhi), Purushottam Kar (IIT Kanpur), and Manik Varma (Microsoft Research)

A Sampling-Based Approach for Efficient Clustering in Large Datasets 12393
Georgios Exarchakis (IHU Strasbourg), Omar Oubari (Sorbonne Université), and Gregor Lenz (Institut de la Vision)

Come-Closer-Diffuse-Faster: Accelerating Conditional Diffusion Models for Inverse Problems Through Stochastic Contraction 12403
Hyungjin Chung (KAIST), Byeongsu Sim (Department of Mathematical Sciences, KAIST), and Jong Chul Ye (Kim Jaechul Graduate School of AI, KAIST, Korea)

Learnable Lookup Table for Neural Network Quantization 12413
Longguang Wang (National University of Defense Technology), Xiaoyu Dong (The University of Tokyo / RIKEN AIP), Yingqian Wang (National University of Defense Technology), Li Liu (National University of Defense Technology), Wei An (National University of Defense Technology), and Yulan Guo (National University of Defense Technology)

Instance-Aware Dynamic Neural Network Quantization 12424
Zhenhua Liu (Peking University), Yunhe Wang (Huawei Technologies), Kai Han (Noah's Ark Lab, Huawei Technologies), Siwei Ma (Peking University, China), and Wen Gao (PKU)

Training High-Performance Low-Latency Spiking Neural Networks by Differentiation on Spike Representation 12434
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)

Fire Together Wire Together: A Dynamic Pruning Approach With Self-Supervised Mask Prediction 12444
Sara Elkerdawy (University of Alberta), Mostafa Elhoushi (Huawei Technologies), Hong Zhang (University of Alberta), and Nilanjan Ray (University of Alberta)

Wavelet Knowledge Distillation: Towards Efficient Image-to-Image Translation 12454
Linfeng Zhang (Tsinghua University), Xin Chen (Intel Corp.), Xiaobing Tu (Kuaishou Technology), Pengfei Wan (Kuaishou Technology), Ning Xu (Kuaishou Technology), and Kaisheng Ma (Tsinghua University)

PokeBNN: A Binary Pursuit of Lightweight Accuracy 12465
Yichi Zhang (Cornell University), Zhiru Zhang (Cornell Univeristy), and Lukasz Lew (Google)

Automated Progressive Learning for Efficient Training of Vision Transformers 12476
Changlin Li (Monash University), Bohan Zhuang (Monash University), Guangrun Wang (University of Oxford), Xiaodan Liang (Sun Yat-sen University), Xiaojun Chang (University of Technology Sydney), and Yi Yang (Zhejiang University)

DeltaCNN: End-to-End CNN Inference of Sparse Frame Differences in Videos 12487
Mathias Parger (Graz University of Technology), Chengcheng Tang (Facebook Reality Labs), Christopher D. Twigg (Meta), Cem Keskin (Facebook), Robert Wang (Facebook Reality Labs), and Markus Steinberger (Graz University of Technology)

Channel Balancing for Accurate Quantization of Winograd Convolutions 12497
Vladimir Chikin (Huawei Noah's Ark Lab) and Vladimir Kryzhanovskiy (Huawei Noah's Ark Lab)

ClusterGNN: Cluster-Based Coarse-To-Fine Graph Neural Network for Efficient Feature Matching 12507
Yan Shi (Tsinghua University), Jun-Xiong Cai (Tsinghua University), Yoli Shavit (Faculty of Engineering, Bar Ilan University), Tai-Jiang Mu (Tsinghua University), Wensen Feng (Huawei Technologies), and Kai Zhang (Tsinghua University)

Interspace Pruning: Using Adaptive Filter Representations To Improve Training of Sparse CNNs 12517
Paul Wimmer (Luebeck University / Robert Bosch GmbH), Jens Mehnert (Robert Bosch GmbH), and Alexandru Condurache (Bosch)

AlignQ: Alignment Quantization With ADMM-Based Correlation Preservation 12528
Ting-An Chen (National Taiwan University), De-Nian Yang (Academia Sinica), and Ming-Syan Chen (National Taiwan University)

TVConv: Efficient Translation Variant Convolution for Layout-Aware Visual Processing 12538
Jierun Chen (The Hong Kong University of Science and Technology), Tianlang He (The Hong Kong University of Science and Technology), Weipeng Zhuo (The Hong Kong University of Science and Technology), Li Ma (The Hong Kong University of Science and Technology), Sangtae Ha (University of Colorado at Boulder), and S.-H. Gary Chan (The Hong Kong University of Science and Technology)

SplitNets: Designing Neural Architectures for Efficient Distributed Computing on Head-Mounted Systems 12549
Xin Dong (Harvard Univeristy), Barbara De Salvo (Facebook), Meng Li (Facebook Inc), Chiao Liu (Facebook Reality Lab), Zhongnan Qu (ETH Zurich), H.T. Kung (Harvard University), and Ziyun Li (Facebook Inc)

TO-FLOW: Efficient Continuous Normalizing Flows With Temporal Optimization Adjoint With Moving Speed 12560
Shian Du (South China University of Technology), Yihong Luo (The Hong Kong University of Science and Technology), Wei Chen (South China University of Technology), Jian Xu (School of Mathematics, South China University of Technology (SCUT)), and Delu Zeng (South China University of Technology)

Poster 3.1: Physics-Based Vision and Shape-From-X

DiLiGenT102: A Photometric Stereo Benchmark Dataset With Controlled Shape and Material Variation 12571
Jieji Ren (Shanghai Jiao Tong University), Feishi Wang (Peking University), Jiahao Zhang (Peking University), Qian Zheng (Nanyang Technological University), Mingjun Ren (Shanghai Jiao Tong University), and Boxin Shi (Peking University)

Universal Photometric Stereo Network Using Global Lighting Contexts 12581
Satoshi Ikehata (National Institute of Informatics)

Uncertainty-Aware Deep Multi-View Photometric Stereo 12591
Berk Kaya (ETH Zurich), Suryansh Kumar (ETH Zurich), Carlos Oliveira (CVL Zurich), Vittorio Ferrari (Google Research), and Luc Van Gool (ETH Zurich)

Fast Light-Weight Near-Field Photometric Stereo 12602
Daniel Lichy (University of Maryland), Soumyadip Sengupta (University of Washington), and David W. Jacobs (University of Maryland, USA)

Glass Segmentation Using Intensity and Spectral Polarization Cues 12612
Haiyang Mei (Dalian University of Technology), Bo Dong (Princeton University), Wen Dong (Dalian University of Technology), Jiaxi Yang (Dalian University of Technology), Seung-Hwan Baek (POSTECH), Felix Heide (Princeton University), Pieter Peers (College of William & Mary), Xiaopeng Wei (Dalian University of Technology), and Xin Yang (Dalian University of Technology)

Shape From Polarization for Complex Scenes in the Wild 12622
Chenyang Lei (HKUST), Chenyang Qi (The Hong Kong University of Science and Technology), Jiaxin Xie (The Hong Kong university of science and technology), Na Fan (HKUST), Vladlen Koltun (Apple), and Qifeng Chen (HKUST)

Deep Depth From Focus With Differential Focus Volume 12632
Fengting Yang (Pennsylvania State University), Xiaolei Huang (The Pennsylvania State University), and Zihan Zhou (Penn State University)

Optimal LED Spectral Multiplexing for NIR2RGB Translation 12642
Lei Liu (Shanghai Jiao Tong University), Yuze Chen (Shanghai Jiao Tong University), Junchi Yan (Shanghai Jiao Tong University), and Yinqiang Zheng (The University of Tokyo)

Shape From Thermal Radiation: Passive Ranging Using Multi-Spectral LWIR Measurements 12651
Yasuto Nagase (Nara Institute of Science and Technology), Takahiro Kushida (Nara Institute of Science and Technology), Kenichiro Tanaka (Ritsumeikan University), Takuya Funatomi (Nara Institute of Science and Technology), and Yasuhiro Mukaigawa (NAIST)

NAN: Noise-Aware NeRFs for Burst-Denoising 12662
Naama Pearl (University of Haifa), Tali Treibitz (University of Haifa), and Simon Korman (University of Haifa)

Estimating Fine-Grained Noise Model via Contrastive Learning 12672
Yunhao Zou (Beijing Institute of Technology) and Ying Fu (Beijing Institute of Technology)

Real-Time Hyperspectral Imaging in Hardware via Trained Metasurface Encoders 12682
Maksim Makarenko (KAUST), Arturo Burguete-Lopez (KAUST), Qizhou Wang (KAUST), Fedor Getman (KAUST), Silvio Giancola (KAUST), Bernard Ghanem (KAUST), and Andrea Fratalocchi (KAUST)

MNSRNet: Multimodal Transformer Network for 3D Surface Super-Resolution 12693
Wuyuan Xie (Shenzhen University), Tengcong Huang (Shenzhen University), and Miaohui Wang (Shenzhen University)

PhyIR: Physics-Based Inverse Rendering for Panoramic Indoor Images 12703
Zhen Li (Northwestern Polytechnical University), Lingli Wang (Realsee), Xiang Huang (Realsee), Cihui Pan (Realsee), and Jiaqi Yang (Northwestern Polytechnical University)

Poster 3.1: Visual Reasoning

Neural Shape Mating: Self-Supervised Object Assembly With Adversarial Shape Priors 12714
Yun-Chun Chen (University of Toronto, Vector Institute), Haoda Li (University of Toronto), Dylan Turpin (University of Toronto), Alec Jacobson (University of Toronto), and Animesh Garg (University of Toronto, Vector Institute, Nvidia)

Learning To Anticipate Future With Dynamic Context Removal 12724
Xinyu Xu (Shanghai Jiao Tong University), Yong-Lu Li (Shanghai Jiao Tong University), and Cewu Lu (Shanghai Jiao Tong University)

Self-Supervised Spatial Reasoning on Multi-View Line Drawings 12735
Siyuan Xiang (New York University), Anbang Yang (New York University), Yanfei Xue (New York University), Yaoqing Yang (UC Berkeley), and Chen Feng (New York University)

Contextual Debiasing for Visual Recognition With Causal Mechanisms 12745
Ruyang Liu (SECE, Shenzhen Graduate School, Peking University), Hao Liu (Tencent), Ge Li (SECE, Shenzhen Graduate School, Peking University), Haodi Hou (tencent), TingHao Yu (Tencent AI Lab), and Tao Yang (Tencent AILab)

Poster 3.1: 3D From Multi-View & Sensors

Relative Pose From a Calibrated and an Uncalibrated Smartphone Image 12756
Yaqing Ding (Nanjing University of Science and Technology), Daniel Barath (ETH Zürich), Jian Yang (Nanjing University of Science and Technology), and Zuzana Kukelova (Czech Technical University in Prague)

Exploiting Rigidity Constraints for LiDAR Scene Flow Estimation 12766
Guanting Dong (University of Science and Technology of China), Yueyi Zhang (University of Science and Technology of China), Hanlin Li (University of Science and Technology of China), Xiaoyan Sun (University of Science and Technology of China), and Zhiwei Xiong (University of Science and Technology of China)

NICE-SLAM: Neural Implicit Scalable Encoding for SLAM 12776
Zihan Zhu (Zhejiang University), Songyou Peng (ETH Zurich and MPI-IS), Viktor Larsson (ETH Zurich), Weiwei Xu (Zhejiang unviersity), Hujun Bao (Zhejiang University), Zhaopeng Cui (Zhejiang University), Martin R. Oswald (ETH Zurich), and Marc Pollefeys (ETH Zurich / Microsoft)

NinjaDesc: Content-Concealing Visual Descriptors via Adversarial Learning 12787
Tony Ng (Imperial College London), Hyo Jin Kim (Facebook), Vincent T. Lee (Facebook), Daniel DeTone (Facebook), Tsun-Yi Yang (Facebook), Tianwei Shen (Facebook), Eddy Ilg (Facebook Reality Labs), Vassileios Balntas (Facebook Reality Labs), Krystian Mikolajczyk (Imperial College London), and Chris Sweeney (Facebook Reality Labs)

ScaleNet: A Shallow Architecture for Scale Estimation 12798
Axel Barroso-Laguna (Imperial College London), Yurun Tian (Imperial College London), and Krystian Mikolajczyk (Imperial College London)

Camera Pose Estimation Using Implicit Distortion Models 12809
Linfei Pan (ETHZ), Marc Pollefeys (ETH Zurich / Microsoft), and Viktor Larsson (ETH Zurich)

GIFS: Neural Implicit Function for General Shape Representation 12819
Jianglong Ye (UCSD), Yuntao Chen (TuSimple), Naiyan Wang (TuSimple), and Xiaolong Wang (UCSD)

Learning Deep Implicit Functions for 3D Shapes With Dynamic Code Clouds 12830
Tianyang Li (Tsinghua University), Xin Wen (Tsinghua University and JD.com), Yu-Shen Liu (Tsinghua University), Hua Su (Kuaishou Technology), and Zhizhong Han (Wayne State University)

SPAMs: Structured Implicit Parametric Models 12841
Pablo Palafox (Technical University Munich), Nikolaos Sarafianos (Facebook Reality Labs), Tony Tung (Facebook Reality Labs), and Angela Dai (Technical University of Munich)

Deblur-NeRF: Neural Radiance Fields From Blurry Images 12851
Li Ma (The Hong Kong University of Science and Technology), Xiaoyu Li (Tencent AI Lab), Jing Liao (City University of Hong Kong), Qi Zhang (Tencent AI Lab), Xuan Wang (Tencent AI Lab), Jue Wang (Tencent AI Lab), and Pedro V. Sander (HKUST)

Panoptic Neural Fields: A Semantic Object-Aware Neural Scene Representation 12861
Abhijit Kundu (Google), Kyle Genova (Google Research), Xiaoqi Yin (Google), Alireza Fathi (Google), Caroline Pantofaru (Google Research), Leonidas J. Guibas (Stanford University), Andrea Tagliasacchi (Google Brain and University of Toronto), Frank Dellaert (Georgia Tech), and Thomas Funkhouser (Google Research)

Depth-Supervised NeRF: Fewer Views and Faster Training for Free 12872
Kangle Deng, Andrew Liu (Google), Jun-Yan Zhu (Carnegie Mellon University), and Deva Ramanan (Carnegie Mellon University)

Dense Depth Priors for Neural Radiance Fields From Sparse Input Views 12882
Barbara Roessle (Technical University of Munich), Jonathan T. Barron (Google Research), Ben Mildenhall (Google Research), Pratul P. Srinivasan (Google Research), and Matthias Nießner (Technical University of Munich)

EfficientNeRF - Efficient Neural Radiance Fields 12892
Tao Hu (Chinese University of Hong Kong), Shu Liu (SmartMore), Yilun Chen (Chinese University of Hong Kong), Tiancheng Shen (The Chinese University of Hong Kong), and Jiaya Jia (Chinese University of Hong Kong)

InfoNeRF: Ray Entropy Minimization for Few-Shot Neural Volume Rendering 12902
Mijeong Kim (Seoul National Univ), Seonguk Seo (Seoul National University), and Bohyung Han (Seoul National University)

Mega-NERF: Scalable Construction of Large-Scale NeRFs for Virtual Fly-Throughs 12912
Haithem Turki (Carnegie Mellon University), Deva Ramanan (Carnegie Mellon University), and Mahadev Satyanarayanan (Carnegie Mellon University)

Urban Radiance Fields 12922
Konstantinos Rematas (Google), Andrew Liu (Google), Pratul P. Srinivasan (Google Research), Jonathan T. Barron (Google Research), Andrea Tagliasacchi (Google Brain and University of Toronto), Thomas Funkhouser (Google Research), and Vittorio Ferrari (Google Research)

Hallucinated Neural Radiance Fields in the Wild 12933
Xingyu Chen (Xi'an Jiaotong University), Qi Zhang (Tencent AI Lab), Xiaoyu Li (Tencent AI Lab), Yue Chen (Xi'an Jiaotong University), Ying Feng (Tencent AI Lab), Xuan Wang (Tencent AI Lab), and Jue Wang (Tencent AI Lab)

Towards Multimodal Depth Estimation From Light Fields 12943
Titus Leistner (Universität Heidelberg), Radek Mackowiak (Heidelberg University), Lynton Ardizzone (Heidelberg University), Ullrich Köthe (University of Heidelberg), and Carsten Rother (University of Heidelberg)

Degradation-Agnostic Correspondence From Resolution-Asymmetric Stereo 12952
Xihao Chen (University of Science and Technology of China), Zhiwei Xiong (University of Science and Technology of China), Zhen Cheng (University of Science and Technology of China), Jiayong Peng (University of Science and Technology of China), Yueyi Zhang (University of Science and Technology of China), and Zheng-Jun Zha (University of Science and Technology of China)

Uniform Subdivision of Omnidirectional Camera Space for Efficient Spherical Stereo Matching 12962
Donghun Kang (KAIST), Hyeonjoong Jang (KAIST), Jungeon Lee (Korea Advanced Institute of Science and Technology), Chong-Min Kyung (Korea Advanced Institute of Science and Technology), and Min H. Kim (KAIST)

Attention Concatenation Volume for Accurate and Efficient Stereo Matching 12971
Gangwei Xu (Huazhong University of Science and Technology), Junda Cheng (Huazhong University of Science and Technology), Peng Guo (Huazhong University of Science and Technology), and Xin Yang (Huazhong University of Science and Technology)

Generalized Binary Search Network for Highly-Efficient Multi-View Stereo 12981
Zhenxing Mi (The Hong Kong University of Science and Technology), Chang Di (Dalian University of Technology), and Dan Xu (The Hong Kong University of Science and Technology)

Revisiting Domain Generalized Stereo Matching Networks From a Feature Consistency Perspective 12991
Jiawei Zhang (Beihang University), Xiang Wang (Beihang University), Xiao Bai (Beihang University), Chen Wang (Beihang University), Lei Huang (Beihang University), Yimin Chen (Beihang University), Lin Gu (RIKENAIP / The University of Tokyo), Jun Zhou (Griffith University), Tatsuya Harada (The University of Tokyo / RIKEN), and Edwin R. Hancock (University of York, UK)

GraftNet: Towards Domain Generalized Stereo Matching With a Broad-Spectrum and Task-Oriented Feature 13002
Biyang Liu (College of Information Science & Electronic Engineering, Zhejiang University), Huimin Yu (TBD), and Guodong Qi (Zhejiang University)

ITSA: An Information-Theoretic Approach to Automatic Shortcut Avoidance and Domain Generalization in Stereo Matching Networks 13012
WeiQin Chuah (RMIT University), Ruwan Tennakoon (RMIT), Reza Hoseinnezhad (RMIT University), Alireza Bab-Hadiashar (RMIT), and David Suter (Edith Cowan University)

ActiveZero: Mixed Domain Learning for Active Stereovision With Zero Annotation 13023
Isabella Liu (UCSD), Edward Yang (UC San Diego), Jianyu Tao (UCSD), Rui Chen (Tsinghua University), Xiaoshuai Zhang (UC San Diego), Qing Ran (Alibaba DAMO), Zhu Liu (Alibaba), and Hao Su (UCSD)

FoggyStereo: Stereo Matching With Fog Volume Representation 13033
Chengtang Yao (Beijing Institute of Technology) and Lidong Yu (NIO)

Poster 3.1: Pose Estimation & Tracking

Multi-Person Extreme Motion Prediction 13043
Wen Guo (INRIA), Xiaoyu Bie (INRIA), Xavier Alameda-Pineda (INRIA), and Francesc Moreno-Noguer (IRI)

Learning Local-Global Contextual Adaptation for Multi-Person Pose Estimation 13055
Nan Xue (Wuhan University), Tianfu Wu (NC State University), Gui-Song Xia (Wuhan University), and Liangpei Zhang (Wuhan University)

AdaptPose: Cross-Dataset Adaptation for 3D Human Pose Estimation by Learnable Motion Generation 13065
Mohsen Gholami (University of British Columbia), Bastian Wandt (University of British Columbia), Helge Rhodin (UBC), Rabab Ward (University of British Columbia), and Z. Jane Wang (University of British Columbia)

Single-Stage Is Enough: Multi-Person Absolute 3D Pose Estimation 13076
Lei Jin (Beijing University of Posts and Telecommunications), Chenyang Xu (Beijing University of Posts and Telecommunications), Xiaojuan Wang (Beijing University of Posts and Telecommunications), Yabo Xiao (Beijing University of Posts and Telecommunications), Yandong Guo (OPPO Research Institute), Xuecheng Nie (NUS), and Jian Zhao (Institute of North Electronic Equipment)

Distribution-Aware Single-Stage Models for Multi-Person 3D Pose Estimation 13086
Zitian Wang (Beihang University), Xuecheng Nie (NUS), Xiaochao Qu (MTlab, Meitu Inc.), Yunpeng Chen (National University of Singapore), and Si Liu (Beihang University)

Trajectory Optimization for Physics-Based Reconstruction of 3D Human Pose From Monocular Video 13096
Erik Gärtner (Lund University), Mykhaylo Andriluka (Google), Hongyi Xu (Google), and Cristian Sminchisescu (Google)

Ray3D: Ray-Based 3D Human Pose Estimation for Monocular Absolute 3D Localization 13106
Yu Zhan (AiBee Inc), Fenghai Li (Beijing Technology and Business University), Renliang Weng (Aibee Inc), and Wongun Choi (AiBee Inc)

Lite Pose: Efficient Architecture Design for 2D Human Pose Estimation 13116
Yihan Wang (Tsinghua University), Muyang Li (Carnegie Mellon University), Han Cai (Massachusetts Institute of Technology), Wei-Ming Chen (MIT), and Song Han (MIT)

Location-Free Human Pose Estimation 13127
Xixia Xu (Beijing Jiaotong university), Yingguo Gao (Peking University), Ke Yan (Tencent), Xue Lin (Beijing Jiaotong University), and Qi Zou (Beijing Jiaotong University)

MHFormer: Multi-Hypothesis Transformer for 3D Human Pose Estimation 13137
Wenhao Li (Peking University), Hong Liu (Peking University Shenzhen Graduate School), Hao Tang (ETH Zurich), Pichao Wang (Alibaba Group), and Luc Van Gool (ETH Zurich)

Estimating Egocentric 3D Human Pose in the Wild With External Weak Supervision 13147
Jian Wang (Max Planck Institute for Informatics), Lingjie Liu (Max Planck Institute for Informatics), Weipeng Xu (Reality Labs Research), Kripasindhu Sarkar (Max Plank Institute for Informatics), Diogo Luvizon (Max Planck Institute for Informatics), and Christian Theobalt (MPI Informatik)

Physical Inertial Poser (PIP): Physics-Aware Real-Time Human Motion Tracking From Sparse Inertial Sensors 13157
Xinyu Yi (Tsinghua University), Yuxiao Zhou (Tsinghua University), Marc Habermann (Max Planck Institute for Informatics), Soshi Shimada (MPI for Informatics), Vladislav Golyanik (MPI for Informatics), Christian Theobalt (MPI Informatik), and Feng Xu (Tsinghua University)

PoseKernelLifter: Metric Lifting of 3D Human Pose Using Sound 13169
Zhijian Yang (Samsung AI Center New York), Xiaoran Fan (Samsung), Volkan Isler (University of Minnesota, Twin Cities), and Hyun Soo Park (The University of Minnesota)

Differentiable Dynamics for Articulated 3D Human Motion Reconstruction 13180
Erik Gärtner (Lund University), Mykhaylo Andriluka (Google), Erwin Coumans (Google), and Cristian Sminchisescu (Google)

COAP: Compositional Articulated Occupancy of People 13191
Marko Mihajlovic (ETH Zurich), Shunsuke Saito (Facebook), Aayush Bansal (Carnegie Mellon University), Michael Zollhöfer (Facebook Reality Labs), and Siyu Tang (ETH Zurich)

Capturing Humans in Motion: Temporal-Attentive 3D Human Pose and Shape Estimation From Monocular Video 13201
Wen-Li Wei (Academia Sinica), Jen-Chun Lin (Academia Sinica), Tyng-Luh Liu (Academia Sinica), and Hong-Yuan Mark Liao (Institute of Information Science, Academia Sinica, Taiwan)

SC2-PCR: A Second Order Spatial Compatibility for Efficient and Robust Point Cloud Registration 13211
Zhi Chen (Huazhong University of Science and Technology), Kun Sun (China University of Geosciences (Wuhan)), Fan Yang (Huazhong University of Science and Technology), and Wenbing Tao (Huazhong University of Science and Technology)

MixSTE: Seq2seq Mixed Spatio-Temporal Encoder for 3D Human Pose Estimation in Video 13222
Jinlu Zhang (Wuhan University), Zhigang Tu (Wuhan University), Jianyu Yang (Soochow University), Yujin Chen (Technical University of Munich), and Junsong Yuan (State University of New York at Buffalo, USA)

Putting People in Their Place: Monocular Regression of 3D People in Depth 13233
Yu Sun (Harbin Institute of technology), Wu Liu (AI Research of JD.com), Qian Bao (AI Research of JD.com), Yili Fu (HIT), Tao Mei (AI Research of JD.com), and Michael J. Black (Max Planck Institute for Intelligent Systems)

FLAG: Flow-Based 3D Avatar Generation From Sparse Observations 13243
Sadegh Aliakbarian (Microsoft), Pashmina Cameron (Microsoft), Federica Bogo (Meta), Andrew Fitzgibbon (Microsoft), and Thomas J. Cashman (Microsoft)

GOAL: Generating 4D Whole-Body Motion for Hand-Object Grasping 13253
Omid Taheri (Max Planck Institute for Intelligent Systems), Vasileios Choutas (ETH Zurich), Michael J. Black (Max Planck Institute for Intelligent Systems), and Dimitrios Tzionas (Max Planck Institute for Intelligent Systems)

Capturing and Inferring Dense Full-Body Human-Scene Contact 13264
Chun-Hao P. Huang (Max Planck Institute for Intelligent Systems), Hongwei Yi (Max Planck Institute for Intelligent Systems), Markus Höschle (Max Planck Institute), Matvey Safroshkin (Max Planck Institute for Intellegent systems), Tsvetelina Alexiadis (Max-Planck Institute for Intelligent Systems), Senya Polikovsky (Max Planck Institute for Intelligent Systems), Daniel Scharstein (Middlebury College), and Michael J. Black (Max Planck Institute for Intelligent Systems)

BodyMap: Learning Full-Body Dense Correspondence Map 13276
Anastasia Ianina (Facebook Reality Labs), Nikolaos Sarafianos (Facebook Reality Labs), Yuanlu Xu (Facebook Reality Labs), Ignacio Rocco (Facebook AI Research), and Tony Tung (Facebook Reality Labs)

ICON: Implicit Clothed Humans Obtained From Normals 13286
Yuliang Xiu (Max Planck Institute for Intelligent Systems), Jinlong Yang (Max Planck Institute for Intelligent Systems), Dimitrios Tzionas (Max Planck Institute for Intelligent Systems), and Michael J. Black (Max Planck Institute for Intelligent Systems)

Oral 3.2.1: Security, Transparency, Fairness, Accountability, Privacy & Ethics in Vision

Adversarial Texture for Fooling Person Detectors in the Physical World 13297
Zhanhao Hu (Tsinghua University), Siyuan Huang (Tsinghua University), Xiaopei Zhu (Tsinghua University), Fuchun Sun (Tsinghua), Bo Zhang (Tsinghua University), and Xiaolin Hu (Tsinghua University)

Infrared Invisible Clothing: Hiding From Infrared Detectors at Multiple Angles in Real World 13307
Xiaopei Zhu (Tsinghua University), Zhanhao Hu (Tsinghua University), Siyuan Huang (Tsinghua University), Jianmin Li (Tsinghua University), and Xiaolin Hu (Tsinghua University)

Enhancing Classifier Conservativeness and Robustness by Polynomiality 13317
Ziqi Wang (Delft University of Technology) and Marco Loog (Delft University of Technology & University of Copenhagen)

Backdoor Attacks on Self-Supervised Learning 13327
Aniruddha Saha (University of Maryland Baltimore County), Ajinkya Tejankar (UMBC), Soroush Abbasi Koohpayegani (University of Maryland Baltimore County), and Hamed Pirsiavash (University of California Davis)

Towards Practical Deployment-Stage Backdoor Attack on Deep Neural Networks 13337
Xiangyu Qi (Princeton University), Tinghao Xie (Zhejiang University), Ruizhe Pan (Zhejiang University), Jifeng Zhu (Tencent), Yong Yang (Tencent), and Kai Bu (Zhejiang University)

Few-Shot Backdoor Defense Using Shapley Estimation 13348
Jiyang Guan (Institute of Automation, Chinese Academy of Sciences), Zhuozhuo Tu (The University of Sydney), Ran He (Institute of Automation, Chinese Academy of Sciences), and Dacheng Tao (JD.com)

Better Trigger Inversion Optimization in Backdoor Scanning 13358
Guanhong Tao (Purdue University), Guangyu Shen (Purdue University), Yingqi Liu (Purdue University), Shengwei An (Purdue University), Qiuling Xu (Purdue University), Shiqing Ma (Rutgers University), Pan Li (Purdue University), and Xiangyu Zhang (Purdue University)

Bandits for Structure Perturbation-Based Black-Box Attacks To Graph Neural Networks With Theoretical Guarantees 13369
Binghui Wang (Illinois Institute of Technology), Youqi Li (Beijing Institute of Technology), and Pan Zhou (Huazhong University of Science and Technology)

Improving Robustness Against Stealthy Weight Bit-Flip Attacks by Output Code Matching 13378
Ozan Özdenizci (Graz University of Technology) and Robert Legenstein (Graz University of Technology)

LAS-AT: Adversarial Training With Learnable Attack Strategy 13388
Xiaojun Jia (Institute of Information EngineeringChinese Academy of Sciences), Yong Zhang (Tencent AI Lab), Baoyuan Wu (The Chinese University of Hong Kong, Shenzhen; Shenzhen Research Institute of Big Data), Ke Ma (UCAS), Jue Wang (Tencent AI Lab), and Xiaochun Cao (Chinese Academy of Sciences)

Subspace Adversarial Training 13399
Tao Li (Shanghai Jiao Tong University), Yingwen Wu (Shanghai Jiaotong University), Sizhe Chen (Shanghai Jiao Tong University), Kun Fang (Shanghai Jiao Tong Unviersity), and Xiaolin Huang (Shanghai Jiao Tong University)

Pyramid Adversarial Training Improves ViT Performance 13409
Charles Herrmann (Google), Kyle Sargent (Google), Lu Jiang (Google Research), Ramin Zabih (Google), Huiwen Chang (Google), Ce Liu (Microsoft), Dilip Krishnan (Google), and Deqing Sun (Google)

Fingerprinting Deep Neural Networks Globally via Universal Adversarial Perturbations 13420
Zirui Peng (Shanghai JiaoTong University), Shaofeng Li (Shanghai Jiao Tong University), Guoxing Chen (Shanghai Jiao Tong University), Cheng Zhang (The Ohio State University), Haojin Zhu (Shanghai Jiao Tong University), and Minhui Xue (University of Adelaide)

Robust Image Forgery Detection Over Online Social Network Shared Images 13430
Haiwei Wu (University of Macau), Jiantao Zhou (University of Macau), Jinyu Tian (University of Macau), and Jun Liu (University of Macau)

Quantifying Societal Bias Amplification in Image Captioning 13440
Yusuke Hirota (Osaka University), Yuta Nakashima (Osaka University), and Noa Garcia (Osaka University)

Oral 3.2.2: Image & Video Synthesis and Generation (II); Video Analysis & Understanding

Drop the GAN: In Defense of Patches Nearest Neighbors As Single Image Generative Models ... 13450
Niv Granot (Weizmann Institute of Science), Ben Feinstein (Weizmann Institute of Science), Assaf Shocher (Weizmann Institute of Science), Shai Bagon (Weizmann Institute of Science), and Michal Irani (Weizmann Institute, Israel)

GAN-Supervised Dense Visual Alignment 13460
William Peebles (University of California, Berkeley), Jun-Yan Zhu (Carnegie Mellon University), Richard Zhang (Adobe), Antonio Torralba (MIT), Alexei A. Efros (UC Berkeley), and Eli Shechtman (Adobe Research, US)

Look Closer To Supervise Better: One-Shot Font Generation via Component-Based Discriminator 13472
Yuxin Kong (South China University of Technology), Canjie Luo (South China University of Technology), Weihong Ma (South China University of Technology), Qiyuan Zhu (zhuqiyuan), Shenggao Zhu (Huawei), Nicholas Yuan (Huawei Cloud), and Lianwen Jin (South China University of Technology)

Text2Mesh: Text-Driven Neural Stylization for Meshes 13482
Oscar Michel (University of Chicago), Roi Bar-On (University of Chicago), Richard Liu (University of Chicago), Sagie Benaim (University of Copenhagen), and Rana Hanocka (University of Chicago)

StyleSDF: High-Resolution 3D-Consistent Image and Geometry Generation 13493
Roy Or-El (University of Washington), Xuan Luo (University of Washington), Mengyi Shan (University of Washington), Eli Shechtman (Adobe Research, US), Jeong Joon Park (Stanford University), and Ira Kemelmacher-Shlizerman (University of Washington)

Physical Simulation Layer for Accurate 3D Modeling 13504
Mariem Mezghanni (École Polytechnique), Théo Bodrito (Mines Paristech), Malika Boulkenafed (None), and Maks Ovsjanikov (Ecole polytechnique)

Fourier PlenOctrees for Dynamic Radiance Field Rendering in Real-Time 13514
Liao Wang (ShanghaiTech University), Jiakai Zhang (ShanghaiTech University), Xinhang Liu (ShanghaiTech University), Fuqiang Zhao (ShanghaiTech University), Yanshun Zhang (Dgene), Yingliang Zhang (DGene), Minye Wu (KU Leuven), Jingyi Yu (Shanghai Tech University), and Lan Xu (ShanghaiTech)

Neural Texture Extraction and Distribution for Controllable Person Image Synthesis 13525
Yurui Ren (Shenzhen Graduate School, Peking University), Xiaoqing Fan (Shenzhen Graduate School, Peking University), Ge Li (SECE, Shenzhen Graduate School, Peking University), Shan Liu (Tencent America), and Thomas H. Li (Advanced Institute of Information Technology, Peking University)

I M Avatar: Implicit Morphable Head Avatars From Videos 13535
Yufeng Zheng (ETH Zurich), Victoria Fernández Abrevaya (Max Planck Institute), Marcel C. Bühler (ETH Zürich), Xu Chen (ETH Zürich), Michael J. Black (Max Planck Institute for Intelligent Systems), and Otmar Hilliges (ETH Zurich)

E2V-SDE: From Asynchronous Events to Fast and Continuous Video Reconstruction via Neural Stochastic Differential Equations ..... N/A
Jongwan Kim (Seoul National University, Republic of Korea), DongJin Lee (Seoul National University, Republic of Korea), Byunggook Na (Seoul National University, Republic of Korea), Seongsik Park (Korea Institute of Science and Technology), Jeonghee Jo (Seoul National University, Republic of Korea), and Sungroh Yoon (Seoul National University; Korea Institute of Science and Technology)

RCL: Recurrent Continuous Localization for Temporal Action Detection 13556
Qiang Wang (Alibaba Group), Yanhao Zhang (Alibaba Group), Yun Zheng (Alibaba group), and Pan Pan (Alibaba Group)

Self-Supervised Predictive Convolutional Attentive Block for Anomaly Detection 13566
Nicolae-Cătălin Ristea (University Politehnica of Bucharest), Neelu Madan (Aalborg University), Radu Tudor Ionescu (University of Bucharest), Kamal Nasrollahi (Aalborg University, Denmark), Fahad Shahbaz Khan (MBZUAI), Thomas B. Moeslund (Aalborg University), and Mubarak Shah (University of Central Florida)

MeMViT: Memory-Augmented Multiscale Vision Transformer for Efficient Long-Term Video Recognition 13577
Chao-Yuan Wu (Facebook), Yanghao Li (Facebook AI Research), Karttikeya Mangalam (UC Berkeley), Haoqi Fan (Facebook AI Research), Bo Xiong (Facebook AI Research), Jitendra Malik (University of California at Berkeley), and Christoph Feichtenhofer (Facebook AI Research)

TubeR: Tubelet Transformer for Video Action Detection 13588
Jiaojiao Zhao (University of Amsterdam), Yanyi Zhang (Rutgers University), Xinyu Li (ByteDance), Hao Chen (Amazon), Bing Shuai (Amazon), Mingze Xu (Amazon), Chunhui Liu (Amazon), Kaustav Kundu (Amazon), Yuanjun Xiong (Amazon), Davide Modolo (Amazon), Ivan Marsic (Rutgers University), Cees G. M. Snoek (University of Amsterdam), and Joseph Tighe (Amazon)

MixFormer: End-to-End Tracking With Iterative Mixed Attention 13598
Yutao Cui (Nanjing University), Cheng Jiang (Nanjing University), Limin Wang (Nanjing University), and Gangshan Wu (Nanjing University)

Oral 3.2.3: Recognition, Learning for Vision, and Robot Vision

DN-DETR: Accelerate DETR Training by Introducing Query DeNoising 13609
Feng Li (Hong Kong Univerity of Science and Technology), Hao Zhang (hkust), Shilong Liu (Tsinghua University), Jian Guo (IDEA Research), Lionel M. Ni (The Hong Kong University of Science and Technology (Guangzhou), ,), and Lei Zhang (International Digital Economy Academy (IDEA), ,)

Proper Reuse of Image Classification Features Improves Object Detection 13618
Cristina Vasconcelos (Google), Vighnesh Birodkar (Google), and Vincent Dumoulin (Google)

Boosting 3D Object Detection by Simulating Multimodality on Point Clouds 13628
Wu Zheng (The Chinese University of Hong Kong), Mingxuan Hong (The Chinese University of Hong Kong), Li Jiang (Max Planck Institute for Informatics), and Chi-Wing Fu (The Chinese University of Hong Kong)

TransVPR: Transformer-Based Place Recognition With Multi-Level Attention Aggregation 13638
Ruotong Wang (Xi'an Jiaotong University), Yanqing Shen (Xi'an Jiaotong University), Weiliang Zuo (Xi'an Jiaotong University), Sanping Zhou (Xi'an Jiaotong University), and Nanning Zheng (Xi'an Jiaotong University)

Disentangling Visual Embeddings for Attributes and Objects 13648
Nirat Saini (University of Maryland), Khoi Pham (University of Maryland, College Park), and Abhinav Shrivastava (University of Maryland)

QueryDet: Cascaded Sparse Query for Accelerating High-Resolution Small Object Detection 13658
Chenhongyi Yang (University of Edinburgh), Zehao Huang (TuSimple), and Naiyan Wang (TuSimple)

Unknown-Aware Object Detection: Learning What You Don't Know From Videos in the Wild ... 13668
Xuefeng Du (Unversity of Wisconsin, Madison), Xin Wang (Microsoft Research), Gabriel Gozum (University of Wisconsin - Madison), and Yixuan Li (University of Wisconsin-Madison)

Interpretable Part-Whole Hierarchies and Conceptual-Semantic Relationships in Neural Networks 13679
Nicola Garau (UNITN), Niccolò Bisagno (Università di Trento), Zeno Sambugaro (University of Trento), and Nicola Conci (UNITN)

Can Neural Nets Learn the Same Model Twice? Investigating Reproducibility and Double Descent From the Decision Boundary Perspective 13689
Gowthami Somepalli (University of Maryland, College Park), Liam Fowl (University of Maryland), Arpit Bansal (University of Maryland - College Park), Ping Yeh-Chiang (University of Maryland - College Park), Yehuda Dar (Rice University), Richard Baraniuk (Rice University), Micah Goldblum (University of Maryland), and Tom Goldstein (University of Maryland, College Park)

Calibrating Deep Neural Networks by Pairwise Constraints 13699
Jiacheng Cheng (University of California, San Diego) and Nuno Vasconcelos (UC San Diego)

Lifelong Graph Learning 13709
Chen Wang (Carnegie Mellon University), Yuheng Qiu (Carnegie Mellon University), Dasong Gao (Carnegie Mellon University), and Sebastian Scherer (Carnegie Mellon University)

OrphicX: A Causality-Inspired Latent Variable Model for Interpreting Graph Neural Networks.. 13719
Wanyu Lin (The Hong Kong Polytechnic University), Hao Lan (University of Toronto), Hao Wang (Rutgers University), and Baochun Li (University of Toronto)

Coarse-To-Fine Q-Attention: Efficient Learning for Visual Robotic Manipulation via Discretisation 13729
Stephen James (UC Berkeley), Kentaro Wada (Imperial College London), Tristan Laidlow (Imperial College London), and Andrew J. Davison (Imperial College London)

Dual Task Learning by Leveraging Both Dense Correspondence and Mis-Correspondence for Robust Change Detection With Imperfect Matches 13739
Jin-Man Park (KAIST), Ue-Hwan Kim (GIST), Seon-Hoon Lee (KAIST), and Jong-Hwan Kim (KAIST)

Cross-View Transformers for Real-Time Map-View Semantic Segmentation 13750
Brady Zhou (UT Austin) and Philipp Krähenbühl (UT Austin)

Poster 3.2: Video Analysis & Understanding

UnweaveNet: Unweaving Activity Stories 13760
Will Price (University of Bristol), Carl Vondrick (Columbia University), and Dima Damen (University of Bristol)

Weakly-Supervised Online Action Segmentation in Multi-View Instructional Videos 13770
Reza Ghoddoosian (University of Texas at Arlington), Isht Dwivedi (Honda Research Institute USA), Nakul Agarwal (Honda Research Institute USA), Chiho Choi (Honda Research Institute US), and Behzad Dariush (Honda Research Institute US)

Audio-Adaptive Activity Recognition Across Video Domains 13781
Yunhua Zhang (University of Amsterdam), Hazel Doughty (University of Amsterdam), Ling Shao (NCAI/IIAI/MBZUAI), and Cees G. M. Snoek (University of Amsterdam)

Frame-Wise Action Representations for Long Videos via Sequence Contrastive Learning 13791
Minghao Chen (Zhejiang University), Fangyun Wei (Microsoft Research Asia), Chong Li (Microsoft Research Asia), and Deng Cai (ZJU)

Image Based Reconstruction of Liquids From 2D Surface Detections 13801
Florian Richter (UCSD), Ryan K. Orosco (University of California San Diego), and Michael C. Yip (UC San Diego)

Learning From Untrimmed Videos: Self-Supervised Video Representation Learning With Hierarchical Consistency 13811
Zhiwu Qing (Huazhong University of Science and Technology), Shiwei Zhang (DAMO Academy, Alibaba Group), Ziyuan Huang (National University of Singapore), Yi Xu (Alibaba Group), Xiang Wang (Huazhong University of Science and Technology), Mingqian Tang (Alibaba Group), Changxin Gao (Huazhong University of Science and Technology), Rong Jin (alibaba group), and Nong Sang (Huazhong University of Science and Technology)

How Do You Do It? Fine-Grained Action Understanding With Pseudo-Adverbs 13822
Hazel Doughty (University of Amsterdam) and Cees G. M. Snoek (University of Amsterdam)

Programmatic Concept Learning for Human Motion Description and Synthesis 13833
Sumith Kulal (Stanford University), Jiayuan Mao (MIT), Alex Aiken (Stanford University), and Jiajun Wu (Stanford University)

Learning To Recognize Procedural Activities With Distant Supervision 13843
Xudong Lin (Columbia University), Fabio Petroni (Facebook AI Research), Gedas Bertasius (UNC Chapel Hill), Marcus Rohrbach (Facebook AI Research), Shih-Fu Chang (Columbia University), and Lorenzo Torresani (Facebook AI Research)

Implicit Motion Handling for Video Camouflaged Object Detection 13854
Xuelian Cheng (Monash University), Huan Xiong (Mohamed bin Zayed University of Artificial Intelligence (MBZUAI), ,), Deng-Ping Fan (ETH Zurich), Yiran Zhong (Australian National University), Mehrtash Harandi (Monash University), Tom Drummond (University of Melbourne), and Zongyuan Ge (Monash)

Dynamic Scene Graph Generation via Anticipatory Pre-Training 13864
Yiming Li (Zheng Zhou University), Xiaoshan Yang (CASIA), and Changsheng Xu (CASIA)

Learning To Refactor Action and Co-Occurrence Features for Temporal Action Localization 13874
Kun Xia (Xi'an Jiaotong University), Le Wang (Xi'an Jiaotong University), Sanping Zhou (Xi'an Jiaotong University), Nanning Zheng (Xi'an Jiaotong University), and Wei Tang (University of Illinois at Chicago)

OCSampler: Compressing Videos to One Clip With Single-Step Sampling 13884
Jintao Lin (Nanjing university), Haodong Duan (Chinese University of HongKong), Kai Chen (SenseTime Research), Dahua Lin (The Chinese University of Hong Kong), and Limin Wang (Nanjing University)

A Hybrid Egocentric Activity Anticipation Framework via Memory-Augmented Recurrent and One-Shot Representation Forecasting 13894
Tianshan Liu (The Hong Kong Polytechnic University) and Kin-Man Lam (The Hong Kong Polytechnic University)

TubeFormer-DeepLab: Video Mask Transformer 13904
Dahun Kim (KAIST), Jun Xie (Google), Huiyu Wang (JHU), Siyuan Qiao (Google), Qihang Yu (Johns Hopkins University), Hong-Seok Kim (Google), Hartwig Adam (Google), In So Kweon (KAIST), and Liang-Chieh Chen (Google Inc.)

ASM-Loc: Action-Aware Segment Modeling for Weakly-Supervised Temporal Action Localization 13915
Bo He (University of Maryland), Xitong Yang (University of Maryland), Le Kang (Baidu), Zhiyu Cheng (BAIDU USA LLC), Xin Zhou (Baidu USA), and Abhinav Shrivastava (University of Maryland)

STRPM: A Spatiotemporal Residual Predictive Model for High-Resolution Video Prediction 13926
Zheng Chang (University of Chinese Academy of Sciences), Xinfeng Zhang (University of Chinese Academy of Sciences), Shanshe Wang (Peking University), Siwei Ma (Peking University, China), and Wen Gao (PKU)

Look for the Change: Learning Object States and State-Modifying Actions From Untrimmed Web Videos 13936
Tomáš Souček (Czech Technical University), Jean-Baptiste Alayrac (DeepMind), Antoine Miech (DeepMind), Ivan Laptev (INRIA Paris), and Josef Sivic (Czech Technical University)

End-to-End Compressed Video Representation Learning for Generic Event Boundary Detection 13947
Congcong Li (UCAS), Xinyao Wang (Bytedance), Longyin Wen (Bytedance Inc.), Dexiang Hong (University of Chinese Academy of Sciences), Tiejian Luo (University of Chinese Academy of Sciences), and Libo Zhang (Institute of Software Chinese Academy of Sciences)

Contextualized Spatio-Temporal Contrastive Learning With Self-Supervision 13957
Liangzhe Yuan (Google Research), Rui Qian (Cornell University), Yin Cui (Google), Boqing Gong (Google), Florian Schroff (Google Inc.), Ming-Hsuan Yang (Google Research), Hartwig Adam (Google), and Ting Liu (Google Research)

Deep Anomaly Discovery From Unlabeled Videos via Normality Advantage and Self-Paced Refinement 13967
Guang Yu (National University of Defense Technology), Siqi Wang (National University of Defense Technology), Zhiping Cai (NUDT), Xinwang Liu (National University of Defense Technology), Chuanfu Xu (National University of Defense Technology), and Chengkun Wu (National University of Defense Technology)

A Deeper Dive Into What Deep Spatiotemporal Networks Encode: Quantifying Static vs. Dynamic Information 13979
Matthew Kowal (York University), Mennatullah Siam (York University), Md Amirul Islam (Ryerson University), Neil D. B. Bruce (University of Guelph), Richard P. Wildes (York University), and Konstantinos G. Derpanis (York University)

Long-Short Temporal Contrastive Learning of Video Transformers 13990
Jue Wang (Facebook AI), Gedas Bertasius (UNC Chapel Hill), Du Tran (Facebook AI), and Lorenzo Torresani (Facebook AI Research)

Scene Consistency Representation Learning for Video Scene Segmentation 14001
Haoqian Wu (Shenzhen University), Keyu Chen (tencent), Yanan Luo (tencent), Ruizhi Qiao (Tencent Youtu Lab), Bo Ren (Tencent), Haozhe Liu (King Abdullah University of Science and Technology), Weicheng Xie (Shenzhen University), and Linlin Shen (Shenzhen University)

Unsupervised Pre-Training for Temporal Action Localization Tasks 14011
Can Zhang (Peking University), Tianyu Yang (Tencent AI Lab), Junwu Weng (Tencent AI Lab), Meng Cao (Peking University), Jue Wang (Tencent AI Lab), and Yuexian Zou (Peking University)

Contrastive Learning for Unsupervised Video Highlight Detection 14022
Taivanbat Badamdorj (University of Alberta), Mrigank Rochan (Huawei Noah's Ark Lab), Yang Wang (University of Manitoba; Huawei Technologies Canada), and Li Cheng (ECE dept., University of Alberta)

Deformable Video Transformer 14033
Jue Wang (Facebook AI) and Lorenzo Torresani (Facebook AI Research)

Recurring the Transformer for Video Action Recognition 14043
Jiewen Yang (TCL Corporate Research (HK) Co., Ltd, ,), Xingbo Dong (Yonsei University; TCL Research AI Lab; Monash University), Liujun Liu (Fudan University), Chao Zhang (TCL Corporate Research), Jiajun Shen (TCL Research), and Dahai Yu (TCL Corporate Research (HK) Co., Ltd, ,)

Poster 3.2: Recognition: Detection, Categorization, Retrieval

Open-Vocabulary One-Stage Detection With Hierarchical Visual-Language Knowledge Distillation 14054
Zongyang Ma (Institute of Automation, Chinese Academy of Sciences), Guan Luo (National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences; School of Artificial Intelligence, University of Chinese Academy of Sciences), Jin Gao (Institute of Automation, Chinese Academy of Sciences), Liang Li (The Brain Science Center, Beijing Institute of Basic Medical Sciences), Yuxin Chen (Institute of Automation, Chinese Academy of Sciences), Shaoru Wang (NLPR&UCAS), Congxuan Zhang (Nanchang Hangkong University), and Weiming Hu (Institute of AutomationChinese Academy of Sciences)

Learning To Prompt for Open-Vocabulary Object Detection With Vision-Language Model 14064
Yu Du (Tsinghua University), Fangyun Wei (Microsoft Research Asia), Zihe Zhang (Tsinghua University), Miaojing Shi (King's College London), Yue Gao (Microsoft Research Asia), and Guoqi Li (Tsinghua University)

Sign Language Video Retrieval With Free-Form Textual Queries 14074
Amanda Duarte (Barcelona Supercomputing Center), Samuel Albanie (University of Cambridge), Xavier Giró-i-Nieto (Universitat Politecnica de Catalunya), and Gül Varol (Ecole des Ponts ParisTech)

FashionVLP: Vision Language Transformer for Fashion Retrieval With Feedback 14085
Sonam Goenka (Amazon), Zhaoheng Zheng (University of Southern California), Ayush Jaiswal (Amazon.com Inc.), Rakesh Chada (Amazon), Yue Wu (Amazon.com Inc.), Varsha Hedau (Amazon), and Pradeep Natarajan (Amazon.com Inc.)

Pushing the Performance Limit of Scene Text Recognizer Without Human Annotation 14096
Caiyuan Zheng (Northwestern Polytechnical University), Hui Li (Samsung Research and Development Institute China Xi'an), Seon-Min Rhee (Samsung Electronics), Seungju Han (Samsung Advanced Institute of Technology), Jae-Joon Han (Samsung), and Peng Wang (Northwestern Polytechnical University)

ESCNet: Gaze Target Detection With the Understanding of 3D Scenes 14106
Jun Bao (Hangzhou Dianzi University), Buyu Liu (NEC Labs), and Jun Yu (HDU)

Interactive Multi-Class Tiny-Object Detection 14116
Chunggi Lee (Lunit), Seonwook Park (Lunit Inc.), Heon Song (Lunit Inc.), Jeongun Ryu (Lunit), Sanghoon Kim (Lunit Inc.), Haejoon Kim (Lunit), Sérgio Pereira (Lunit Inc.), and Donggeun Yoo (Lunit)

Weakly Supervised Rotation-Invariant Aerial Object Detection Network 14126
Xiaoxu Feng (Northwestern Polytechnical University), Xiwen Yao (Northwestern Polytechnical University), Gong Cheng (Northwestern Polytechnical University), and Junwei Han (NWPU, China)

Large Loss Matters in Weakly Supervised Multi-Label Classification 14136
Youngwook Kim (Seoul National University), Jae Myung Kim (Seoul National University), Zeynep Akata (University of Tübingen), and Jungwoo Lee (Seoul National University)

MetaFSCIL: A Meta-Learning Approach for Few-Shot Class Incremental Learning 14146
Zhixiang Chi (Huawei Noah's Ark Laboratory), Li Gu (Huawei Canada), Huan Liu (McMaster University), Yang Wang (Huawei Noah's Ark Laboratory), Yuanhao Yu (University of Manitoba; Huawei Technologies Canada), and Jin Tang (Huawei Noah's Ark Laboratory)

FreeSOLO: Learning To Segment Objects Without Annotations 14156
Xinlong Wang (University of Adelaide), Zhiding Yu (NVIDIA), Shalini De Mello (NVIDIA Research), Jan Kautz (NVIDIA), Anima Anandkumar (Caltech), Chunhua Shen (University of Adelaide, Australia), and Jose M. Alvarez (NVIDIA)

Revisiting AP Loss for Dense Object Detection: Adaptive Ranking Pair Selection 14167
Dongli Xu (University of Electronic Science and Technology of China), Jinhong Deng (University of Electronic Science and Technology of China), and Wen Li (University of Electronic Science and Technology of China)

SIOD: Single Instance Annotated per Category per Image for Object Detection 14177
Hanjun Li (Sun Yat-sen University), Xingjia Pan (Tencent), Ke Yan (Tencent), Fan Tang (Jilin University), and Wei-Shi Zheng (Sun Yat-sen University, China)

Towards Robust Adaptive Object Detection Under Noisy Annotations 14187
Xinyu Liu (City University of Hong Kong), Wuyang Li (City University of Hong Kong), Qiushi Yang (City University of Hong Kong), Baopu Li (BAIDU USA LLC), and Yixuan Yuan (City University of Hong Kong)

Task-Specific Inconsistency Alignment for Domain Adaptive Object Detection 14197
Liang Zhao (Nanjing University) and Limin Wang (Nanjing University)

Salvage of Supervision in Weakly Supervised Object Detection 14207
Lin Sui (Nanjing University), Chen-Lin Zhang (4Paradigm, Inc), and Jianxin Wu (Nanjing University)

Label, Verify, Correct: A Simple Few Shot Object Detection Method 14217
Prannay Kaul (University of Oxford), Weidi Xie (Shanghai Jiao Tong University), and Andrew Zisserman (University of Oxford)

Background Activation Suppression for Weakly Supervised Object Localization 14228
Pingyu Wu (USTC), Wei Zhai (University of Science and Technology of China), and Yang Cao (University of Science and Technology of China)

Bridging the Gap Between Classification and Localization for Weakly Supervised Object Localization 14238
Eunji Kim (Seoul National University), Siwon Kim (Seoul National University), Jungbeom Lee (Seoul National University), Hyunwoo Kim (LG AI Research), and Sungroh Yoon (Seoul National University)

Divide and Conquer: Compositional Experts for Generalized Novel Class Discovery 14248
Muli Yang (Xidian University), Yuehua Zhu (Xidian University), Jiaping Yu (Xidian University), Aming Wu (Xidian University), and Cheng Deng (Xidian University)

Cloth-Changing Person Re-Identification From a Single Image With Gait Prediction and Regularization 14258
Xin Jin (University of Science and Technology of China), Tianyu He (Alibaba Group), Kecheng Zheng (University of Science and Technology of China), Zhiheng Yin (University of Michigan), Xu Shen (Alibaba Group), Zhen Huang (University of Science and Technology of China), Ruoyu Feng (University of Science and Technology of China), Jianqiang Huang (Damo Academy, Alibaba Group), Zhibo Chen (University of Science and Technology of China), and Xian-Sheng Hua (Damo Academy, Alibaba Group)

Lifelong Unsupervised Domain Adaptive Person Re-Identification With Coordinated Anti-Forgetting and Adaptation 14268
Zhipeng Huang (University of Science and Technology of China), Zhizheng Zhang (Microsoft Research), Cuiling Lan (Microsoft Research), Wenjun Zeng (EIT Institute for Advanced Study), Peng Chu (Microsoft), Quanzeng You (Microsoft Azure Computer Vision), Jiang Wang (Microsoft), Zicheng Liu (Microsoft), and Zheng-Jun Zha (University of Science and Technology of China)

Unleashing Potential of Unsupervised Pre-Training With Intra-Identity Regularization for Person Re-Identification 14278
Zizheng Yang (University of Science and Technology of China), Xin Jin (University of Science and Technology of China), Kecheng Zheng (University of Science and Technology of China), and Feng Zhao (University of Science and Technology of China)

Learning With Twin Noisy Labels for Visible-Infrared Person Re-Identification 14288
Mouxing Yang (Sichuan University), Zhenyu Huang (Sichuan University), Peng Hu (College of Computer Science, Sichuan University), Taihao Li (zhejianglab), Jiancheng Lv (Sichuan University), and Xi Peng (College of Computer Science, Sichuan Univerisity)

Towards Total Recall in Industrial Anomaly Detection 14298
Karsten Roth (University of Tuebingen), Latha Pemula (Amazon), Joaquin Zepeda (Amazon), Bernhard Schölkopf (Amazon / MPI Intelligent Systems), Thomas Brox (University of Freiburg), and Peter Gehler (Amazon)

H2FA R-CNN: Holistic and Hierarchical Feature Alignment for Cross-Domain Weakly Supervised Object Detection 14309
Yunqiu Xu (University of Technology Sydney), Yifan Sun (Baidu Research), Zongxin Yang (Zhejiang University), Jiaxu Miao (Zhejiang University), and Yi Yang (UTS)

Geometric and Textural Augmentation for Domain Gap Reduction 14320
Xiao-Chang Liu (University of Bath), Yong-Liang Yang (University of Bath), and Peter Hall (University of Bath)

General Incremental Learning With Domain-Aware Categorical Representations 14331
Jiangwei Xie (ShanghaiTech University), Shipeng Yan (ShanghaiTech University), and Xuming He (ShanghaiTech University)

DST: Dynamic Substitute Training for Data-Free Black-Box Attack 14341
Wenxuan Wang (Fudan University), Xuelin Qian (Fudan University), Yanwei Fu (Fudan University), and Xiangyang Xue (Fudan University)

ART-Point: Improving Rotation Robustness of Point Cloud Classifiers via Adversarial Rotation 14351
Ruibin Wang (Peking University), Yibo Yang (Peking University), and Dacheng Tao (JD.com)

Poster 3.2: Self-, Semi-, Meta-, & Unsupervised Learning

Label Matching Semi-Supervised Object Detection 14361
Binbin Chen (Hikvision Research Institute), Weijie Chen (Zhejiang University), Shicai Yang (Hikvision Research Institute), Yunyi Xuan (Hikvision Research Institute), Jie Song (Zhejiang University), Di Xie (Hikvision Research Institute), Shiliang Pu (Hikvision Research Institute), Mingli Song (Zhengjiang University), and Yueting Zhuang (Zhejiang University)

Multidimensional Belief Quantification for Label-Efficient Meta-Learning 14371
Deep Shankar Pandey (Rochester Institute of Technology) and Qi Yu (Rochester Institute of Technology)

Propagation Regularizer for Semi-Supervised Learning With Extremely Scarce Labeled Samples 14381
Noo-ri Kim (Sungkyunkwan University) and Jee-Hyong Lee (Sungkyunkwan University)

Learning To Affiliate: Mutual Centralized Learning for Few-Shot Classification 14391
Yang Liu (Zhejiang University), Weifeng Zhang (Zhejiang University), Chao Xiang (Zhejiang University), Tu Zheng (FABU), Deng Cai (ZJU), and Xiaofei He (Zhejiang University)

Class-Aware Contrastive Semi-Supervised Learning 14401
Fan Yang (Tsinghua University), Kai Wu (tencent), Shuyi Zhang (Tencent Youtu Lab), Guannan Jiang (CATL), Yong Liu (Tencent), Feng Zheng (SUSTech), Wei Zhang (CATL), Chengjie Wang (Tencent), and Long Zeng (Tsinghua University)

Exploring the Equivalence of Siamese Self-Supervised Learning via a Unified Gradient Framework 14411
Chenxin Tao (Tsinghua University), Honghui Wang (Tsinghua University), Xizhou Zhu (SenseTime), Jiahua Dong (Zhejiang University), Shiji Song (Department of Automation, Tsinghua University), Gao Huang (Tsinghua), and Jifeng Dai (SenseTime)

Dual Temperature Helps Contrastive Learning Without Many Negative Samples: Towards Understanding and Simplifying MoCo 14421
Chaoning Zhang (KAIST), Kang Zhang (KAIST), Trung X. Pham (KAIST), Axi Niu (Northwestern Polytechnical University), Zhinan Qiao (University of North Texas), Chang D. Yoo (KAIST), and In So Kweon (KAIST)

Learning Where To Learn in Cross-View Self-Supervised Learning 14431
Lang Huang (The University of Tokyo), Shan You (SenseTime), Mingkai Zheng (The University of Sydney), Fei Wang (University of Science and Technology of China), Chen Qian (SenseTime), and Toshihiko Yamasaki (The University of Tokyo)

Dist-PU: Positive-Unlabeled Learning From a Label Distribution Perspective 14441
Yunrui Zhao (School of Computer Science and Technology, University of Chinese Academy of Sciences), Qianqian Xu (Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences), Yangbangyan Jiang (Institute of Information Engineering, Chinese Academy of Sciences; University of Chinese Academy of Sciences), Peisong Wen (Institute of Computing Technology, Chinese Academy of Sciences; University of Chinese Academy of Sciences), and Qingming Huang (University of Chinese Academy of Sciences)

SimMatch: Semi-Supervised Learning With Similarity Matching 14451
Mingkai Zheng (The University of Sydney), Shan You (SenseTime), Lang Huang (The University of Tokyo), Fei Wang (University of Science and Technology of China), Chen Qian (SenseTime), and Chang Xu (University of Sydney)

Active Teacher for Semi-Supervised Object Detection 14462
Peng Mi (xmu.edu.cn), Jianghang Lin (Xiamen University), Yiyi Zhou (Xiamen University), Yunhang Shen (Xiamen University), Gen Luo (Xiamen University), Xiaoshuai Sun (Xiamen University), Liujuan Cao (Xiamen University), Rongrong Fu (Huawei), Qiang Xu (huawei), and Rongrong Ji (Xiamen University, China)

Not All Labels Are Equal: Rationalizing the Labeling Costs for Training Object Detection 14472
Ismail Elezi (Technical University of Munich (TUM), ,), Zhiding Yu (NVIDIA), Anima Anandkumar (Caltech), Laura Leal-Taixé (TUM), and Jose M. Alvarez (NVIDIA)

Self-Supervised Learning of Object Parts for Semantic Segmentation 14482
Adrian Ziegler (TUM) and Yuki M. Asano (University of Amsterdam)

MUM: Mix Image Tiles and UnMix Feature Tiles for Semi-Supervised Object Detection 14492
JongMok Kim (SNUAILAB), JooYoung Jang (SNUAILAB), Seunghyeon Seo (Seoul National University), Jisoo Jeong (Seoul National University), Jongkeun Na (SNUAILAB), and Nojun Kwak (Seoul National University)

Scale-Equivalent Distillation for Semi-Supervised Object Detection 14502
Qiushan Guo (The University of Hong Kong), Yao Mu (The University of Hong Kong), Jianyu Chen (Tsinghua University), Tianqi Wang (The University of Hong Kong), Yizhou Yu (The University of Hong Kong), and Ping Luo (The University of Hong Kong)

A Self-Supervised Descriptor for Image Copy Detection 14512
Ed Pizzi (Facebook), Sreya Dutta Roy (Facebook Inc), Sugosh Nagavara Ravindra (Facebook), Priya Goyal (Facebook Inc.), and Matthijs Douze (Facebook AI Research)

Self-Supervised Transformers for Unsupervised Object Discovery Using Normalized Cut 14523
Yangtao Wang (Univ. Grenoble Alpes), Xi Shen (École des Ponts ParisTech), Shell Xu Hu (Samsung), Yuan Yuan (MIT), James L. Crowley (Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, LIG, 38000 Grenoble, France), and Dominique Vaufreydaz (Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, LIG, 38000 Grenoble, France)

CAD: Co-Adapting Discriminative Features for Improved Few-Shot Classification 14534
Philip Chikontwe (DGIST (Daegu Gyeongbuk Institute of Science and Technology), ,), Soopil Kim (DGIST), and Sang Hyun Park (DGIST)

Semi-Supervised Few-Shot Learning via Multi-Factor Clustering 14544
Jie Ling (Sun Yat-sen University), Lei Liao (Sun Yat-Sen University), Meng Yang (Sun Yat-sen University), and Jia Shuai (Sun Yat-Sen University)

CoSSL: Co-Learning of Representation and Classifier for Imbalanced Semi-Supervised Learning 14554
Yue Fan (Max Planck Institute for Informatics), Dengxin Dai (MPI for Informatics), Anna Kukleva (MPI for Informatics), and Bernt Schiele (MPI Informatics)

Safe-Student for Safe Deep Semi-Supervised Learning With Unseen-Class Unlabeled Data 14565
Rundong He (Shandong University), Zhongyi Han (Shandong University), Xiankai Lu (Shandong University), and Yilong Yin (Shandong University)

A Simple Data Mixing Prior for Improving Self-Supervised Learning 14575
Sucheng Ren (South China University of Technology), Huiyu Wang (JHU), Zhengqi Gao (MIT), Shengfeng He (South China University of Technology), Alan Yuille (Johns Hopkins University), Yuyin Zhou (Johns Hopkins University), and Cihang Xie (University of California, Santa Cruz)

DETReg: Unsupervised Pretraining With Region Priors for Object Detection 14585
Amir Bar (Tel Aviv University), Xin Wang (Microsoft Research), Vadim Kantorov (INRIA), Colorado J. Reed (UC Berkeley), Roei Herzig (Tel Aviv University), Gal Chechik (NVIDIA), Anna Rohrbach (UC Berkeley), Trevor Darrell (UC Berkeley), and Amir Globerson (Tel Aviv University, Google)

Sound and Visual Representation Learning With Multiple Pretraining Tasks 14596
Arun Balajee Vasudevan (ETH Zurich), Dengxin Dai (ETH Zurich), and Luc Van Gool (ETH Zurich)

UniVIP: A Unified Framework for Self-Supervised Visual Pre-Training 14607
Zhaowen Li (National Laboratory of Pattern RecognitionInstitute of AutomationChinese Academy of Sciences), Yousong Zhu (Institute of Automation, Chinese Academy of Sciences), Fan Yang (Sensetime), Wei Li (SenseTime Research), Chaoyang Zhao (National Laboratory of Pattern Recognition, CASIA), Yingying Chen (CASIA), Zhiyang Chen (Institute of Automation, Chinese Academy of Sciences), Jiahao Xie (Nanyang Technological University), Liwei Wu (SenseTime Research), Rui Zhao (SenseTime Group Limited), Ming Tang (Institute of Automation, Chinese Academy of Sciences), and Jinqiao Wang (Institute of Automation, Chinese Academy of Sciences)

Weakly Supervised Object Localization As Domain Adaption 14617
Lei Zhu (Beijing University of Posts and Telecommunications), Qi She (Bytedance AI Lab), Qian Chen (University of Science and Technology of China), Yunfei You (Peking University), Boyu Wang (University of Western Ontario), and Yanye Lu (Peking University)

Debiased Learning From Naturally Imbalanced Pseudo-Labels 14627
Xudong Wang (UC Berkeley / ICSI), Zhirong Wu (Microsoft Research), Long Lian (UC Berkeley / ICSI), and Stella X. Yu (UC Berkeley / ICSI)

Towards Discovering the Effectiveness of Moderately Confident Samples for Semi-Supervised Learning 14638
Hui Tang (South China University of Technology) and Kui Jia (South China University of Technology)

Masked Feature Prediction for Self-Supervised Visual Pre-Training 14648
Chen Wei (Johns Hopkins University), Haoqi Fan (Facebook AI Research), Saining Xie (Facebook AI Research), Chao-Yuan Wu (Facebook), Alan Yuille (Johns Hopkins University), and Christoph Feichtenhofer (Facebook AI Research)

Contrastive Learning for Space-Time Correspondence via Self-Cycle Consistency 14659
Jeany Son (GIST)

Id-Free Person Similarity Learning 14669
Bing Shuai (Amazon), Xinyu Li (ByteDance), Kaustav Kundu (Amazon), and Joseph Tighe (Amazon)

End-to-End Semi-Supervised Learning for Video Action Detection 14680
Akash Kumar (University of Central Florida) and Yogesh Singh Rawat (University of Central Florida)

Probabilistic Representations for Video Contrastive Learning 14691
Jungin Park (Yonsei University), Jiyoung Lee (NAVER AI LAB), Ig-Jae Kim (KIST), and Kwanghoon Sohn (Yonsei Univ.)

Interact Before Align: Leveraging Cross-Modal Knowledge for Domain Adaptive Action Recognition 14702
Lijin Yang (The University of Tokyo), Yifei Huang (The University of Tokyo), Yusuke Sugano (The University of Tokyo), and Yoichi Sato (University of Tokyo)

BEVT: BERT Pretraining of Video Transformers 14713
Rui Wang (Fudan University), Dongdong Chen (Microsoft Cloud AI), Zuxuan Wu (UMD), Yinpeng Chen (Microsoft), Xiyang Dai (Microsoft), Mengchen Liu (Microsoft), Yu-Gang Jiang (Fudan University), Luowei Zhou (Microsoft), and Lu Yuan (Microsoft)

Generative Cooperative Learning for Unsupervised Video Anomaly Detection 14724
M. Zaigham Zaheer (Electronics and Telecommunications Research Institute), Arif Mahmood (Information Technology University), M. Haris Khan (Muhammad Bin Zayed University of Artificial Intelligence), Mattia Segu (ETH Zurich), Fisher Yu (ETH Zurich), and Seung-Ik Lee (Electronics and Telecommunication Research Institute (ETRI), ,)

When Does Contrastive Visual Representation Learning Work? 14735
Elijah Cole (Caltech), Xuan Yang (Google), Kimberly Wilber (Google), Oisin Mac Aodha (University of Edinburgh), and Serge Belongie (University of Copenhagen)

The Norm Must Go On: Dynamic Unsupervised Domain Adaptation by Normalization 14745
M. Jehanzeb Mirza (Graz University of Technology), Jakub Micorek (Graz University of Technology), Horst Possegger (Graz University of Technology), and Horst Bischof (Graz University of Technology)

What Matters for Meta-Learning Vision Regression Tasks? 14756
Ning Gao (Bosch Center for Artificial Intelligence (BCAI), ,), Hanna Ziesche (Bosch Center for AI), Ngo Anh Vien (Bosch Center for Artificial Intelligence), Michael Volpp (Karlsruhe Institute for Technology), and Gerhard Neumann (Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany, ,)

Poster 3.2: Robot Vision

IFOR: Iterative Flow Minimization for Robotic Object Rearrangement 14767
Ankit Goyal (Princeton University), Arsalan Mousavian (NVIDIA), Chris Paxton (NVIDIA), Yu-Wei Chao (NVIDIA), Brian Okorn (Carnegie Mellon University), Jia Deng (Princeton University), and Dieter Fox (NVIDIA)

TCTrack: Temporal Contexts for Aerial Tracking 14778
Ziang Cao (Tongji University), Ziyuan Huang (National University of Singapore), Liang Pan (Nanyang Technological University), Shiwei Zhang (DAMO Academy, Alibaba Group), Ziwei Liu (Nanyang Technological University), and Changhong Fu (Tongji University)

AKB-48: A Real-World Articulated Object Knowledge Base 14789
Liu Liu (Shanghai JiaoTong University), Wenqiang Xu (Shanghai Jiao Tong University), Haoyuan Fu (Shanghai Jiao Tong University), Sucheng Qian (Shanghai Jiao Tong University), Qiaojun Yu (Shanghai Jiaotong University), Yang Han (Shanghai Jiao Tong University), and Cewu Lu (Shanghai Jiao Tong University)

3DAC: Learning Attribute Compression for Point Clouds 14799
Guangchi Fang (Sun Yat-sen University), Qingyong Hu (University of Oxford), Hanyun Wang (Information Engineering University), Yiling Xu (Shanghai Jiao Tong University), and Yulan Guo (Sun Yat-sen University)

Simple but Effective: CLIP Embeddings for Embodied AI 14809
Apoorv Khandelwal (Allen Institute for AI), Luca Weihs (Allen Institute for Artificial Intelligence), Roozbeh Mottaghi (Allen Institute for AI), and Aniruddha Kembhavi (Allen Institute for Artificial Intelligence)

Multi-Robot Active Mapping via Neural Bipartite Graph Matching 14819
Kai Ye (Peking University), Siyan Dong (Shandong University), Qingnan Fan (Tencent AI Lab), He Wang (Peking University), Li Yi (Tsinghua University), Fei Xia (Google Inc), Jue Wang (Tencent AI Lab), and Baoquan Chen (Peking University)

Continuous Scene Representations for Embodied AI 14829
Samir Yitzhak Gadre (Columbia University), Kiana Ehsani (Allen Institute for Artificial Intelligence), Shuran Song (Columbia University), and Roozbeh Mottaghi (Allen Institute for AI)

Interactron: Embodied Adaptive Object Detection 14840
Klemen Kotar (Allen Institute for AI) and Roozbeh Mottaghi (Allen Institute for AI)

Online Learning of Reusable Abstract Models for Object Goal Navigation 14850
Tommaso Campari (University of Padova), Leonardo Lamanna (Fondazione Bruno Kessler and University of Brescia), Paolo Traverso (Fondazione Bruno Kessler (FBK), ,), Luciano Serafini (Fondazione Bruno Kessler), and Lamberto Ballan (University of Padova)

RNNPose: Recurrent 6-DoF Object Pose Refinement With Robust Correspondence Field Estimation and Pose Optimization 14860
Yan Xu (Chinese University of Hong Kong), Kwan-Yee Lin (SenseTime Research), Guofeng Zhang (Zhejiang University), Xiaogang Wang (Chinese University of Hong Kong, Hong Kong), and Hongsheng Li (The Chinese University of Hong Kong)

UDA-COPE: Unsupervised Domain Adaptation for Category-Level Object Pose Estimation 14871
Taeyeop Lee (KAIST), Byeong-Uk Lee (KAIST), Inkyu Shin (KAIST), Jaesung Choe (KAIST), Ukcheol Shin (KAIST), In So Kweon (KAIST), and Kuk-Jin Yoon (KAIST)

Symmetry and Uncertainty-Aware Object SLAM for 6DoF Object Pose Estimation 14881
Nathaniel Merrill (University of Delaware), Yuliang Guo (BOSCH Research North America), Xingxing Zuo (Zhejiang University), Xinyu Huang (BOSCH Research North America), Stefan Leutenegger (TU Munich), Xi Peng (University of Delaware), Liu Ren (BOSCH Research North America), and Guoquan Huang (University of Delaware)

Upright-Net: Learning Upright Orientation for 3D Point Cloud 14891
Xufang Pang (Shenzhen Institute of Artificial Intelligence and Robotics for Society (AIRS)), Feng Li (Shenzhen Institute of Artificial Intelligence and Robotics for Society), Ning Ding (The Chinese University of Hong Kong, Shenzhen), and Xiaopin Zhong (Shenzhen University)

Poster 3.2: Computer Vision for Social Good

DeepFake Disrupter: The Detector of DeepFake Is My Friend 14900
Xueyu Wang (The University of Sydney), Jiajun Huang (The University of Sydney), Siqi Ma (University of Queensland), Surya Nepal (CSIRO), and Chang Xu (University of Sydney)

HybridCR: Weakly-Supervised 3D Point Cloud Semantic Segmentation via Hybrid Contrastive Regularization 14910
Mengtian Li (East China Normal University), Yuan Xie (East China Normal University), Yunhang Shen (Xiamen University), Bo Ke (Tencent Youtu Lab), Ruizhi Qiao (Tencent Youtu Lab), Bo Ren (Tencent), Shaohui Lin (East China Normal University), and Lizhuang Ma (East China Normal University)

Open-Domain, Content-Based, Multi-Modal Fact-Checking of Out-of-Context Images via Online Resources 14920
Sahar Abdelnabi (CISPA Helmholtz Center for Information Security), Rakibul Hasan (CISPA), and Mario Fritz (CISPA Helmholtz Center for Information Security)

Leveraging Real Talking Faces via Self-Supervision for Robust Forgery Detection 14930
Alexandros Haliassos (Imperial College London), Rodrigo Mira (Imperial College London), Stavros Petridis (Imperial College London), and Maja Pantic (Imperial College London)

Poster 3.2: Adversarial Attack & Defense

Transferable Sparse Adversarial Attack 14943
Ziwen He (Institute of Automation), Wei Wang (Center for Research on Intelligent Perception and Computing, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences), Jing Dong (Chinese Academy of Sciences), and Tieniu Tan (NLPR, China)

Segment and Complete: Defending Object Detectors Against Adversarial Patch Attacks With Robust Patch Detection 14953
Jiang Liu (Johns Hopkins University), Alexander Levine (University of Maryland), Chun Pong Lau (Johns Hopkins University), Rama Chellappa (Johns Hopkins University), and Soheil Feizi (University of Maryland)

Stochastic Variance Reduced Ensemble Adversarial Attack for Boosting the Adversarial Transferability 14963
Yifeng Xiong (Huazhong University of Science and Technology), Jiadong Lin (Huazhong University of Science and Technology), Min Zhang (Huazhong University of Science and Technology), John E. Hopcroft (Cornell University), and Kun He (Huazhong University of Science and Technology)

Improving Adversarial Transferability via Neuron Attribution-Based Attacks 14973
Jianping Zhang (The Chinese University of Hong Kong), Weibin Wu (Sun Yat-sen University), Jen-tse Huang (The Chinese University of Hong Kong), Yizhan Huang (The Chinese University of Hong Kong), Wenxuan Wang (The Chinese University of Hong Kong), Yuxin Su (Sun Yat-sen University), and Michael R. Lyu (The Chinese University of Hong Kong)

Complex Backdoor Detection by Symmetric Feature Differencing 14983
Yingqi Liu (Purdue University), Guangyu Shen (Purdue University), Guanhong Tao (Purdue University), Zhenting Wang (Rutgers University), Shiqing Ma (Rutgers University), and Xiangyu Zhang (Purdue University)

Protecting Facial Privacy: Generating Adversarial Identity Masks via Style-Robust Makeup Transfer 14994
Shengshan Hu (Huazhong University of Science and Technology), Xiaogeng Liu (Huazhong University of Science and Technology), Yechao Zhang (Huazhong University of Science and Technology), Minghui Li (Huazhong University of Science and Technology), Leo Yu Zhang (Deakin University), Hai Jin (Huazhong University of Science and Technology), and Libing Wu (Wuhan University)

Zero-Query Transfer Attacks on Context-Aware Object Detectors 15004
Zikui Cai (University of California, Riverside), Shantanu Rane (PARC), Alejandro E. Brito (PARC), Chengyu Song (University of California, Riverside), Srikanth V. Krishnamurthy (University of California, Riverside), Amit K. Roy-Chowdhury (University of California, Riverside), and M. Salman Asif (University of California, Riverside)

360-Attack: Distortion-Aware Perturbations From Perspective-Views 15015
Yunjian Zhang (Institute of Information Engineering, Chinese Academy of Sciences), Yanwei Liu (Institute of Information Engineering, Chinese Academy of Sciences), Jinxia Liu (Zhejiang Wanli University), Jingbo Miao (Institute of Information EngineeringChinese Academy of Sciences), Antonios Argyriou (University of Thessaly), Liming Wang (State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences), and Zhen Xu (State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences)

Label-Only Model Inversion Attacks via Boundary Repulsion 15025
Mostafa Kahla (Virginia Tech), Si Chen (Virginia Tech), Hoang Anh Just (Virginia Tech), and Ruoxi Jia (Virginia Tech)

Merry Go Round: Rotate a Frame and Fool a DNN 15034
Daksh Thapar (Indian Institute of Technology, Mandi), Aditya Nigam (IIT mandi), and Chetan Arora (Indian Institute of Technology Delhi)

Cross-Modal Transferable Adversarial Attacks From Images to Videos 15044
Zhipeng Wei (Fudan University), Jingjing Chen (Fudan University), Zuxuan Wu (UMD), and Yu-Gang Jiang (Fudan University)

BppAttack: Stealthy and Efficient Trojan Attacks Against Deep Neural Networks via Image Quantization and Contrastive Adversarial Learning 15054
Zhenting Wang (Rutgers University), Juan Zhai (Rutgers University), and Shiqing Ma (Rutgers University)

Investigating Top-k White-Box and Transferable Black-Box Attack 15064
Chaoning Zhang (KAIST), Philipp Benz (KAIST), Adil Karjauv (KAIST), Jae Won Cho (KAIST), Kang Zhang (KAIST), and In So Kweon (KAIST)

Boosting Black-Box Attack With Partially Transferred Conditional Adversarial Distribution 15074
Yan Feng (Tsinghua University), Baoyuan Wu (The Chinese University of Hong Kong, Shenzhen; Shenzhen Research Institute of Big Data), Yanbo Fan (Tencent AI Lab), Li Liu (Shenzhen Research Institute of Big Data, the chinese university of hong kong shenzhen), Zhifeng Li (Tencent AI Lab), and Shu-Tao Xia (Tsinghua University)

Practical Evaluation of Adversarial Robustness via Adaptive Auto Attack 15084
Ye Liu (University of Electronic Science and Technology of China), Yaya Cheng (University of Electronic Science and Technology of China), Lianli Gao (The University of Electronic Science and Technology of China), Xianglong Liu (BUAA), Qilong Zhang (University of Electronic Science and Technology of China), and Jingkuan Song (UESTC)

Towards Efficient Data Free Black-Box Adversarial Attack 15094
Jie Zhang (Zhejiang University), Bo Li (Nanjing University), Jianghe Xu (Tencent Youtu Lab), Shuang Wu (Tencent), Shouhong Ding (Tencent), Lei Zhang (ZheJiang University), and Chao Wu (Zhejiang University)

Masking Adversarial Damage: Finding Adversarial Saliency for Robust and Sparse Network 15105
Byung-Kwan Lee (KAIST), Junho Kim (KAIST), and Yong Man Ro (KAIST)

Certified Patch Robustness via Smoothed Vision Transformers 15116
Hadi Salman (MIT), Saachi Jain (MIT), Eric Wong (MIT), and Aleksander Madry (MIT)

Towards Practical Certifiable Patch Defense With Vision Transformer 15127
Zhaoyu Chen (Fudan University), Bo Li (Nanjing University), Jianghe Xu (Tencent Youtu Lab), Shuang Wu (Tencent), Shouhong Ding (Tencent), and Wenqiang Zhang (Fudan University)

On Adversarial Robustness of Trajectory Prediction for Autonomous Vehicles 15138
Qingzhao Zhang (University of Michigan, Ann Arbor), Shengtuo Hu (University of Michigan), Jiachen Sun (University of Michigan), Qi Alfred Chen (UC Irvine), and Z. Morley Mao (University of Michigan)

3DeformRS: Certifying Spatial Deformations on Point Clouds 15148
Gabriel Pérez S. (Universidad Nacional De Colombia), Juan C. Pérez (KAUST), Motasem Alfarra (KAUST), Silvio Giancola (KAUST), and Bernard Ghanem (KAUST)

Stereoscopic Universal Perturbations Across Different Architectures and Datasets 15159
Zachary Berger (UCLA), Parth Agrawal (UCLA), Tian Yu Liu (UCLA), Stefano Soatto (UCLA), and Alex Wong (University of California, Los Angeles)

Aug-NeRF: Training Stronger Neural Radiance Fields With Triple-Level Physically-Grounded Augmentations 15170
Tianlong Chen (Unversity of Texas at Austin), Peihao Wang (University of Texas at Austin), Zhiwen Fan (University of Texas at Austin), and Zhangyang Wang (University of Texas at Austin)

Bounded Adversarial Attack on Deep Content Features 15182
Qiuling Xu (Purdue University), Guanhong Tao (Purdue University), and Xiangyu Zhang (Purdue University)

DEFEAT: Deep Hidden Feature Backdoor Attacks by Imperceptible Perturbation and Latent Representation Constraints 15192
Zhendong Zhao (Institute of Information Engineering, Chinese Academy of Sciences), Xiaojun Chen (Institute of Information Engineering, CAS), Yuexin Xuan (Institute of Information Engineering, Chinese Academy of Sciences), Ye Dong (Institute of Information Engineering, Chinese Academy of Sciences), Dakui Wang (Institute of Information Engineering, CAS), and Kaitai Liang (Delft University of Technology)

Two Coupled Rejection Metrics Can Tell Adversarial Examples Apart 15202
Tianyu Pang (Tsinghua University), Huishuai Zhang (Microsoft), Di He (Microsoft Research), Yinpeng Dong (Tsinghua University), Hang Su (Tsinghua Univiersity), Wei Chen (Chinese Academy of Sciences), Jun Zhu (Tsinghua University), and Tie-Yan Liu (Microsoft Research)

Give Me Your Attention: Dot-Product Attention Considered Harmful for Adversarial Patch Robustness 15213
Giulio Lovisotto (University of Oxford), Nicole Finnie (Bosch Center for Artificial Intelligence), Mauricio Munoz (Bosch Center for Artificial Intelligence), Chaithanya Kumar Mummadi (Bosch Center for Artificial Intelligence), and Jan Hendrik Metzen (Bosch Center for Artificial Intelligence)

Improving the Transferability of Targeted Adversarial Examples Through Object-Based Diverse Input 15223
Junyoung Byun (KAIST), Seungju Cho (KAIST), Myung-Joon Kwon (KAIST), Hee-Seon Kim (KAIST), and Changick Kim (KAIST)

Adversarial Eigen Attack on Black-Box Models 15233
Linjun Zhou (Tsinghua University), Peng Cui (Tsinghua University), Xingxuan Zhang (Tsinghua University), Yinan Jiang (China Academy of Electronics and Information Technology), and Shiqiang Yang (Tsinghua University)

Appearance and Structure Aware Robust Deep Visual Graph Matching: Attack, Defense and Beyond 15242
Qibing Ren (Shanghai Jiao Tong University), Qingquan Bao (Shanghai Jiao Tong University), Runzhong Wang (Shanghai Jiao Tong University), and Junchi Yan (Shanghai Jiao Tong University)

Enhancing Adversarial Training With Second-Order Statistics of Weights 15252
Gaojie Jin (University of Liverpool), Xinping Yi (University of Liverpool), Wei Huang (University of Liverpool), Sven Schewe (University of Liverpool), and Xiaowei Huang (Liverpool University)

Towards Data-Free Model Stealing in a Hard Label Setting 15263
Sunandini Sanyal (Indian Institute of Science, Bengaluru), Sravanti Addepalli (Indian Institute of Science), and R. Venkatesh Babu (Indian Institute of Science)

Robust Structured Declarative Classifiers for 3D Point Clouds: Defending Adversarial Attacks With Implicit Gradients 15273
Kaidong Li (University of Kansas), Ziming Zhang (Worcester Polytechnic Institute), Cuncong Zhong (University of Kansas), and Guanghui Wang (Ryerson University)

DTA: Physical Camouflage Attacks Using Differentiable Transformation Network 15284
Naufal Suryanto (Pusan National University), Yongsu Kim (SmartM2M), Hyoeun Kang (Pusan National University), Harashta Tatimma Larasati (Pusan National University, Institut Teknologi Bandung), Youngyeo Yun (Pusan National University), Thi-Thu-Huong Le (Pusan National University), Hunmin Yang (Agency for Defense Development), Se-Yoon Oh (Agency for Defense Development), and Howon Kim (Pusan National University)

Frequency-Driven Imperceptible Adversarial Attack on Semantic Similarity 15294
Cheng Luo (Shenzhen University), Qinliang Lin (Shenzhen University), Weicheng Xie (Shenzhen University), Bizhu Wu (Shenzhen University), Jinheng Xie (Shenzhen University), and Linlin Shen (Shenzhen University)

Enhancing Adversarial Robustness for Deep Metric Learning 15304
Mo Zhou (Johns Hopkins University) and Vishal M. Patel (Johns Hopkins University)

Shape-Invariant 3D Adversarial Point Clouds 15314
Qidong Huang (University of Science and Technology of China), Xiaoyi Dong (University of Science and Technology of China), Dongdong Chen (Microsoft Cloud AI), Hang Zhou (Simon Fraser University), Weiming Zhang (University of Science and Technology of China), and Nenghai Yu (University of Science and Technology of China)

Shadows Can Be Dangerous: Stealthy and Effective Physical-World Adversarial Attack by Natural Phenomenon 15324
Yiqi Zhong (Harbin Institute of Technology), Xianming Liu (Harbin Institute of Technology), Deming Zhai (Harbin Institute of Technolgy), Junjun Jiang (Harbin Institute of Technology), and Xiangyang Ji (Tsinghua University)

Exploring Effective Data for Surrogate Training Towards Black-Box Attack 15334
Xuxiang Sun (Northwestern Polytechnical University), Gong Cheng (Northwestern Polytechnical University), Hongda Li (Northwestern Polytechnical University), Lei Pei (Northwestern Polytechnical University), and Junwei Han (NWPU, China)

NICGSlowDown: Evaluating the Efficiency Robustness of Neural Image Caption Generation Models 15344
Simin Chen (UTD), Zihe Song (University of Texas at Dallas), Mirazul Haque (University of Texas at Dallas), Cong Liu (University of Texas at Dallas), and Wei Yang (University of Texas at Dallas)

Dual-Key Multimodal Backdoors for Visual Question Answering 15354
Matthew Walmer (University of Maryland), Karan Sikka (SRI International), Indranil Sur (SRI International), Abhinav Shrivastava (University of Maryland), and Susmit Jha (SRI International)

Proactive Image Manipulation Detection 15365
Vishal Asnani (Michigan State University), Xi Yin (Facebook AI), Tal Hassner (Facebook AI), Sijia Liu (Michigan State University), and Xiaoming Liu (Michigan State University)

Poster 3.2: Vision & Language

ADAPT: Vision-Language Navigation With Modality-Aligned Action Prompts 15375
Bingqian Lin (Sun Yat-sen University), Yi Zhu (Huawei Noah's Ark Lab), Zicong Chen (Sun Yat-sen University), Xiwen Liang (Sun Yat-sen University), Jianzhuang Liu (Huawei Noah's Ark Lab), and Xiaodan Liang (Sun Yat-sen University)

EnvEdit: Environment Editing for Vision-and-Language Navigation 15386
Jialu Li (University of North Carolina at Chapel Hill), Hao Tan (UNC, Chapel Hill), and Mohit Bansal (University of North Carolina at Chapel Hill)

HOP: History-and-Order Aware Pre-Training for Vision-and-Language Navigation 15397
Yanyuan Qiao (University of Adelaide), Yuankai Qi (The University of Adelaide), Yicong Hong (Australian National University), Zheng Yu (University of Adelaide), Peng Wang (Northwestern Polytechnical University), and Qi Wu (University of Adelaide)

Less Is More: Generating Grounded Navigation Instructions From Landmarks 15407
Su Wang (Google AI Language), Ceslee Montgomery (Google), Jordi Orbay (Google), Vighnesh Birodkar (Google), Aleksandra Faust (Google Brain), Izzeddin Gur (Google), Natasha Jaques (UC Berkeley), Austin Waters (Google), Jason Baldridge (Google Inc.), and Peter Anderson (Google)

Bridging the Gap Between Learning in Discrete and Continuous Environments for Vision-and-Language Navigation 15418
Yicong Hong (Australian National University), Zun Wang (Australia National University), Qi Wu (University of Adelaide), and Stephen Gould (Australian National University, Australia)

Reinforced Structured State-Evolution for Vision-Language Navigation 15429
Jinyu Chen (Beihang University), Chen Gao (Beihang University), Erli Meng (Xiaomi AI Lab), Qiong Zhang (Xiaomi Technology), and Si Liu (Beihang University)

Cross-Modal Map Learning for Vision and Language Navigation 15439
Georgios Georgakis (University of Pennsylvania), Karl Schmeckpeper (University of Pennsylvania), Karan Wanchoo (University of Pennsylvania), Soham Dan (University of Pennsylvania), Eleni Miltsakaki (University of Pennsylvania), Dan Roth (UPenn), and Kostas Daniilidis (University of Pennsylvania)

Counterfactual Cycle-Consistent Learning for Instruction Following and Generation in Vision-Language Navigation 15450
Hanqing Wang (Beijing Institute of Technology), Wei Liang (Beijing Institute of Technology), Jianbing Shen (Inception Institute of Artificial Intelligence), Luc Van Gool (ETH Zurich), and Wenguan Wang (Eidgenössische Technische Hochschule Zürich)

One Step at a Time: Long-Horizon Vision-and-Language Navigation With Milestones 15461
Chan Hee Song (The Ohio State University), Jihyung Kil (The Ohio State University), Tai-Yu Pan (The Ohio State University), Brian M. Sadler (Army Research Laboratory, USA), Wei-Lun Chao (The Ohio State University), and Yu Su (The Ohio State University)

Expanding Large Pre-Trained Unimodal Models With Multimodal Information Injection for Image-Text Multimodal Classification 15471
Tao Liang (Shenzhen Jinritoutiao Technology Co., Ltd), Guosheng Lin (Nanyang Technological University), Mingyang Wan (Bytedance), Tianrui Li (School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu, 611756, China), Guojun Ma (ByteDance), and Fengmao Lv (Southwest Jiaotong University)

Shifting More Attention to Visual Backbone: Query-Modulated Refinement Networks for End-to-End Visual Grounding 15481
Jiabo Ye (East China Normal University), Junfeng Tian (Alibaba Group), Ming Yan (Alibaba Group), Xiaoshan Yang (CASIA), Xuwu Wang (Fudan University), Ji Zhang (Alibaba Inc.), Liang He (ECNU), and Xin Lin (ECNU)

Pseudo-Q: Generating Pseudo Language Queries for Visual Grounding 15492
Haojun Jiang (Tsinghua University), Yuanze Lin (University of Washington), Dongchen Han (Tsinghua University), Shiji Song (Department of Automation, Tsinghua University), and Gao Huang (Tsinghua)

Multi-View Transformer for 3D Visual Grounding 15503
Shijia Huang (The Chinese University of Hong Kong), Yilun Chen (Chinese University of Hong Kong), Jiaya Jia (Chinese University of Hong Kong), and Liwei Wang (CUHK)

Multi-Modal Dynamic Graph Transformer for Visual Grounding 15513
Sijia Chen (University of Toronto) and Baochun Li (University of Toronto)

Weakly-Supervised Generation and Grounding of Visual Descriptions With Conditional Generative Models 15523
Effrosyni Mavroudi (Johns Hopkins University) and René Vidal (Johns Hopkins University, USA)

Weakly Supervised Temporal Sentence Grounding With Gaussian-Based Contrastive Proposal Learning 15534
Minghang Zheng (Peking University), Yanjie Huang (Beijing Institute of Technology), Qingchao Chen (Peking University), Yuxin Peng (Peking University), and Yang Liu (Peking University)

Visual Abductive Reasoning 15544
Chen Liang (Zhejiang University), Wenguan Wang (Eidgenössische Technische Hochschule Zürich), Tianfei Zhou (ETH Zurich), and Yi Yang (Zhejiang University)

Query and Attention Augmentation for Knowledge-Based Explainable Reasoning 15555
Yifeng Zhang (University of Minnesota, Twin Cities), Ming Jiang (University of Minnesota), and Qi Zhao (University of Minnesota)

REX: Reasoning-Aware and Grounded Explanation 15565
Shi Chen (University of Minnesota) and Qi Zhao (University of Minnesota)

Not All Relations Are Equal: Mining Informative Labels for Scene Graph Generation 15575
*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)*

Unsupervised Vision-Language Parsing: Seamlessly Bridging Visual Scene Graphs With Language Structures via Dependency Relationships 15586
Chao Lou (ShanghaiTech University), Wenjuan Han (Beijing Institute for General Artificial Intelligence), Yuhuan Lin (Tsinghua University), and Zilong Zheng (UCLA)

Scene Graph Expansion for Semantics-Guided Image Outpainting 15596
Chiao-An Yang (National Taiwan University), Cheng-Yo Tan (Rice University), Wan-Cyuan Fan (National Taiwan University), Cheng-Fu Yang (National Taiwan University), Meng-Lin Wu (Qualcomm Technologies, Inc.), and Yu-Chiang Frank Wang (National Taiwan University)

VisualHow: Multimodal Problem Solving 15606
Jinhui Yang (University of Minnesota), Xianyu Chen (University of Minnesota, Twin Cities), Ming Jiang (University of Minnesota), Shi Chen (University of Minnesota), Louis Wang (University of Minnesota, Twin Cities), and Qi Zhao (University of Minnesota)

FLAVA: A Foundational Language and Vision Alignment Model 15617
Amanpreet Singh (Facebook), Ronghang Hu (Facebook), Vedanuj Goswami (Facebook AI Research), Guillaume Couairon (Facebook AI Research), Wojciech Galuba (Facebook), Marcus Rohrbach (Facebook AI Research), and Douwe Kiela (Facebook AI Research)

Multi-Modal Alignment Using Representation Codebook 15630
Jiali Duan (Amazon), Liqun Chen (Amazon), Son Tran (Amazon A9), Jinyu Yang (University of Texas at Arlington), Yi Xu (Amazon), Belinda Zeng (Amazon), and Trishul Chilimbi (Amazon)

Negative-Aware Attention Framework for Image-Text Matching 15640
Kun Zhang (University of Science and Technology of China), Zhendong Mao (University of Science and Technology of China), Quan Wang (Beijing University of Posts and Telecommunications), and Yongdong Zhang (University of Science and Technology of China)

Vision-Language Pre-Training With Triple Contrastive Learning 15650
Jinyu Yang (University of Texas at Arlington), Jiali Duan (Amazon), Son Tran (Amazon A9), Yi Xu (Amazon), Sampath Chanda (Amazon), Liqun Chen (Amazon), Belinda Zeng (Amazon), Trishul Chilimbi (Amazon), and Junzhou Huang (University of Texas at Arlington)

Vision-Language Pre-Training for Boosting Scene Text Detectors 15660
Sibo Song (Alibaba Group), Jianqiang Wan (Alibaba Group), Zhibo Yang (Alibaba Group), Jun Tang (Alibaba Group), Wenqing Cheng (Huazhong University of Science and Technology), Xiang Bai (Huazhong University of Science and Technology), and Cong Yao (Alibaba DAMO Academy)

COTS: Collaborative Two-Stream Vision-Language Pre-Training Model for Cross-Modal Retrieval 15671
Haoyu Lu (Renmin University of China), Nanyi Fei (Renmin University of China), Yuqi Huo (Renmin University of China), Yizhao Gao (Renmin University of China), Zhiwu Lu (Renmin University of China), and Ji-Rong Wen (Renmin University of China)

Poster 3.2: 3D From Multi-View & Sensors

NeurMiPs: Neural Mixture of Planar Experts for View Synthesis 15681
Zhi-Hao Lin (National Taiwan University), Wei-Chiu Ma (MIT), Hao-Yu Hsu (National Taiwan University), Yu-Chiang Frank Wang (National Taiwan University), and Shenlong Wang (UIUC)

FWD: Real-Time Novel View Synthesis With Forward Warping and Depth 15692
Ang Cao (University of Michigan), Chris Rockwell (University of Michigan), and Justin Johnson (University of Michigan)

SOMSI: Spherical Novel View Synthesis With Soft Occlusion Multi-Sphere Images 15704
Tewodros Habtegebrial (Technische Universität Kaiserslautern), Christiano Gava (German Research Center for Artificial Intelligence (DFKI GmbH)), Marcel Rogge (TU Kaiserslautern), Didier Stricker (DFKI), and Varun Jampani (Google)

Fast, Accurate and Memory-Efficient Partial Permutation Synchronization 15714
Shaohan Li (University of Minnesota), Yunpeng Shi (Princeton University), and Gilad Lerman (University of Minnesota)

Learning To Find Good Models in RANSAC 15723
Daniel Barath (ETH Zürich), Luca Cavalli (ETH Zurich), and Marc Pollefeys (ETH Zurich / Microsoft)

Optimizing Elimination Templates by Greedy Parameter Search 15733
Evgeniy Martyushev (South Ural State University), Jana Vráblíková (Charles University), and Tomas Pajdla (Czech Technical University in Prague)

GPU-Based Homotopy Continuation for Minimal Problems in Computer Vision 15744
Chiang-Heng Chien (Brown University), Hongyi Fan (Brown University), Ahmad Abdelfattah (UTK), Elias Tsigaridas (INRIA Paris), Stanimire Tomov (University of Tennessee), and Benjamin Kimia (Brown University)

HARA: A Hierarchical Approach for Robust Rotation Averaging 15756
Seong Hun Lee (University of Zaragoza) and Javier Civera (Universidad de Zaragoza)

RAGO: Recurrent Graph Optimizer for Multiple Rotation Averaging 15766
Heng Li (Simon Fraser University), Zhaopeng Cui (Zhejiang University), Shuaicheng Liu (UESTC; Megvii), and Ping Tan (Simon Fraser University)

A Unified Model for Line Projections in Catadioptric Cameras With Rotationally Symmetric Mirrors 15776
Pedro Miraldo (MERL Mitsubishi Electric Research Laboratories) and José Pedro Iglesias (Chalmers)

ELSR: Efficient Line Segment Reconstruction With Planes and Points Guidance 15786
Dong Wei (Wuhan University), Yi Wan (Wuhan University), Yongjun Zhang (Wuhan University), Xinyi Liu (Wuhan University), Bin Zhang (Wuhan University), and Xiqi Wang (wuhan university)

Self-Supervised Neural Articulated Shape and Appearance Models 15795
Fangyin Wei (Princeton University), Rohan Chabra (University of North Carolina at Chapel Hill), Lingni Ma (Facebook Reality Labs), Christoph Lassner (Meta Reality Labs Research), Michael Zollhöfer (Facebook Reality Labs), Szymon Rusinkiewicz (Princeton University), Chris Sweeney (Facebook Reality Labs), Richard Newcombe (Facebook), and Mira Slavcheva (Reality Labs Research)

Virtual Elastic Objects 15806
Hsiao-yu Chen (University of Texas at Austin), Edith Tretschk (Max-Planck-Institut für Informatik), Tuur Stuyck (Meta Reality Labs Research), Petr Kadlecek (Meta Reality Labs Research), Ladislav Kavan (Reality Labs Research), Etienne Vouga (The University of Texas at Austin), and Christoph Lassner (Meta Reality Labs Research)

Decoupling Makes Weakly Supervised Local Feature Better 15817
Kunhong Li (Sun Yat-sen University), Longguang Wang (National University of Defense Technology), Li Liu (National University of Defense Technology), Qing Ran (Alibaba DAMO), Kai Xu (National University of Defense Technology), and Yulan Guo (Sun Yat-sen University)

JoinABLe: Learning Bottom-Up Assembly of Parametric CAD Joints 15828
Karl D.D. Willis (Autodesk Research), Pradeep Kumar Jayaraman (Autodesk Research), Hang Chu (Autodesk AI Lab), Yunsheng Tian (Massachusetts Institute of Technology), Yifei Li (MIT CSAIL), Daniele Grandi (Autodesk Research), Aditya Sanghi (Autodesk; Autodesk AI Lab), Linh Tran (Imperial College London / Autodesk AI Lab), Joseph G. Lambourne (Autodesk AI Lab), Armando Solar-Lezama (MIT), and Wojciech Matusik (MIT)

ImplicitAtlas: Learning Deformable Shape Templates in Medical Imaging 15840
Jiancheng Yang (Shanghai Jiao Tong University), Udaranga Wickramasinghe (EPFL, Switzerland), Bingbing Ni (Shanghai Jiao Tong University), and Pascal Fua (EPFL, Switzerland)

DoubleField: Bridging the Neural Surface and Radiance Fields for High-Fidelity Human Reconstruction and Rendering 15851
Ruizhi Shao (Tsinghua University), Hongwen Zhang (Tsinghua University), He Zhang (Beihang university), Mingjia Chen (East China Normal University), Yan-Pei Cao (Y-tech, Kuaishou Technology), Tao Yu (Tsinghua University), and Yebin Liu (Tsinghua University)

Surface-Aligned Neural Radiance Fields for Controllable 3D Human Synthesis 15862
Tianhan Xu (The University of Tokyo), Yasuhiro Fujita (Preferred Networks, Inc.), and Eiichi Matsumoto (Preferred Networks, Inc.)

Structured Local Radiance Fields for Human Avatar Modeling 15872
Zerong Zheng (Tsinghua University), Han Huang (OPPO Research Institute), Tao Yu (Tsinghua University), Hongwen Zhang (Tsinghua University), Yandong Guo (OPPO Research Institute), and Yebin Liu (Tsinghua University)

High-Fidelity Human Avatars From a Single RGB Camera 15883
Hao Zhao (Tianjin university), Jinsong Zhang (Tianjin University), Yu-Kun Lai (Cardiff University), Zerong Zheng (Tsinghua University), Yingdi Xie (VRC Inc.), Yebin Liu (Tsinghua University), and Kun Li (Tianjin University)

Forecasting Characteristic 3D Poses of Human Actions 15893
Christian Diller (Technical University of Munich), Thomas Funkhouser (Google Research), and Angela Dai (Technical University of Munich)

Virtual Correspondence: Humans as a Cue for Extreme-View Geometry 15903
Wei-Chiu Ma (MIT), Anqi Joyce Yang (University of Toronto), Shenlong Wang (UIUC), Raquel Urtasun (Uber ATG), and Antonio Torralba (MIT)

BEHAVE: Dataset and Method for Tracking Human Object Interactions 15914
Bharat Lal Bhatnagar (University of Tübingen, MPI informatik), Xianghui Xie (Saarland University), Ilya A. Petrov (University of Tübingen), Cristian Sminchisescu (Google), Christian Theobalt (MPI Informatik), and Gerard Pons-Moll (University of Tübingen)

Primitive3D: 3D Object Dataset Synthesis From Randomly Assembled Primitives 15926
Xinke Li (National University of Singapore), Henghui Ding (ETH Zurich), Zekun Tong (National University of Singapore), Yuwei Wu (National University of Singapore), and Yeow Meng Chee (National University of Singapore)

RGB-Multispectral Matching: Dataset, Learning Methodology, Evaluation 15937
Fabio Tosi (University of Bologna), Pierluigi Zama Ramirez (University of Bologna), Matteo Poggi (University of Bologna), Samuele Salti (University of Bologna), Stefano Mattoccia (University of Bologna), and Luigi Di Stefano (University of Bologna)

NPBG++: Accelerating Neural Point-Based Graphics 15948
Ruslan Rakhimov (Skoltech), Andrei-Timotei Ardelean (Skoltech), Victor Lempitsky (Yandex), and Evgeny Burnaev (Skoltech)

Depth-Guided Sparse Structure-From-Motion for Movies and TV Shows 15959
Sheng Liu (University at Buffalo), Xiaohan Nie (Amazon), and Raffay Hamid (Amazon)

Motion-From-Blur: 3D Shape and Motion Estimation of Motion-Blurred Objects in Videos 15969
Denys Rozumnyi (ETH Zurich / CTU in Prague), Martin R. Oswald (ETH Zurich), Vittorio Ferrari (Google Research), and Marc Pollefeys (ETH Zurich / Microsoft)

Oral 4.1.1: Representation Learning

Masked Autoencoders Are Scalable Vision Learners 15979
Kaiming He (Facebook AI Research), Xinlei Chen (FAIR), Saining Xie (Facebook AI Research), Yanghao Li (Facebook AI Research), Piotr Dollár (FAIR), and Ross Girshick (FAIR)

Learning ABCs: Approximate Bijective Correspondence for Isolating Factors of Variation With Weak Supervision 15989
Kieran A. Murphy (University of Pennsylvania), Varun Jampani (Google), Srikumar Ramalingam (Google), and Ameesh Makadia (Google Research)

Bayesian Invariant Risk Minimization 16000
Yong Lin (HKUST), Hanze Dong (HKUST), Hao Wang (Rutgers University), and Tong Zhang (The Hong Kong University of Science and Technology)

Crafting Better Contrastive Views for Siamese Representation Learning 16010
Xiangyu Peng (National University of Singapore), Kai Wang (National University of Singapore), Zheng Zhu (Tsinghua University), Mang Wang (Alibaba Group), and Yang You (National University of Singapore)

Rethinking Minimal Sufficient Representation in Contrastive Learning 16020
Haoqing Wang (Peking University), Xun Guo (Microsoft Research Asia), Zhi-Hong Deng (Peking University), and Yan Lu (Microsoft Research Asia)

Multi-Level Feature Learning for Contrastive Multi-View Clustering 16030
Jie Xu (University of Electronic Science and Technology of China), Huayi Tang (University of Electronic Science and Technology of China), Yazhou Ren (University of Electronic Science and Technology of China), Liang Peng (University of Electronic Science and Technology of China), Xiaofeng Zhu (University of Electronic Science and Technology of China), and Lifang He (Lehigh University)

Point-Level Region Contrast for Object Detection Pre-Training 16040
Yutong Bai (Johns Hopkins University), Xinlei Chen (FAIR), Alexander Kirillov (Facebook AI Reserach), Alan Yuille (Johns Hopkins University), and Alexander C. Berg (University of North Carolina, USA)

Class-Incremental Learning by Knowledge Distillation With Adaptive Feature Consolidation 16050
Minsoo Kang (Seoul National University), Jaeyoo Park (Seoul National University), and Bohyung Han (Seoul National University)

A Stitch in Time Saves Nine: A Train-Time Regularizing Loss for Improved Neural Network Calibration 16060
Ramya Hebbalaguppe (Indian Institute of Technology, Delhi), Jatin Prakash (Indian Institute of Technology, Delhi), Neelabh Madan (Indian Institute of Technology Delhi), and Chetan Arora (Indian Institute of Technology Delhi)

SLIC: Self-Supervised Learning With Iterative Clustering for Human Action Videos 16070
Salar Hosseini Khorasgani (University of Toronto), Yuxuan Chen (University of Toronto), and Florian Shkurti (University of Toronto)

Omnivore: A Single Model for Many Visual Modalities 16081
Rohit Girdhar (Facebook AI Research), Mannat Singh (Facebook AI Research), Nikhila Ravi (Facebook AI Research), Laurens van der Maaten (Facebook), Armand Joulin (Facebook AI Research), and Ishan Misra (Facebook AI Research)

DPICT: Deep Progressive Image Compression Using Trit-Planes 16092
Jae-Han Lee (Korea University), Seungmin Jeon (Korea University), Kwang Pyo Choi (Samsung Electronics), Youngo Park (Samsung Electronics Co., LTd.), and Chang-Su Kim (Korea university)

Efficient Geometry-Aware 3D Generative Adversarial Networks 16102
Eric R. Chan (Stanford University), Connor Z. Lin (Stanford University), Matthew A. Chan (Stanford), Koki Nagano (NVIDIA), Boxiao Pan (Stanford University), Shalini De Mello (NVIDIA Research), Orazio Gallo (NVIDIA Research), Leonidas J. Guibas (NVIDIA), Jonathan Tremblay (Nvidia), Sameh Khamis (NVIDIA), Tero Karras (Stanford University), and Gordon Wetzstein (Stanford University)

Geometric Anchor Correspondence Mining With Uncertainty Modeling for Universal Domain Adaptation 16113
Liang Chen (Peking University), Yihang Lou (huawei), Jianzhong He (Huawei), Tao Bai (Huawei), and Minghua Deng (Peking University)

Scaling Vision Transformers to Gigapixel Images via Hierarchical Self-Supervised Learning 16123
Richard J. Chen (Harvard Medical School), Chengkuan Chen (Columbia University), Yicong Li (Tsinghua-Berkeley Shenzhen Institute, Tsinghua University.), Tiffany Y. Chen (Pathology, Brigham and Women's Hospital, Harvard Medical School), Andrew D. Trister (Bill & Melinda Gates Foundation), Rahul G. Krishnan (University of Toronto), and Faisal Mahmood (Pathology, Brigham and Women's Hospital, Harvard Medical School)

Versatile Multi-Modal Pre-Training for Human-Centric Perception 16135
Fangzhou Hong (Nanyang Technological University), Liang Pan (Nanyang Technological University), Zhongang Cai (SenseTime International Pte Ltd), and Ziwei Liu (Nanyang Technological University)

Bridging Video-Text Retrieval With Multiple Choice Questions 16146
Yuying Ge (The University of Hong Kong), Yixiao Ge (Tencent), Xihui Liu (UC Berkeley), Dian Li (Tencent.com), Ying Shan (Tencent), Xiaohu Qie (Tencent), and Ping Luo (The University of Hong Kong)

Integrating Language Guidance Into Vision-Based Deep Metric Learning 16156
Karsten Roth (University of Tuebingen), Oriol Vinyals (DeepMind), and Zeynep Akata (University of Tübingen)

Oral 4.1.2: Computational Photography

NeRF in the Dark: High Dynamic Range View Synthesis From Noisy Raw Images 16169
Ben Mildenhall (Google Research), Peter Hedman (Google Research), Ricardo Martin-Brualla (Google), Pratul P. Srinivasan (Google Research), and Jonathan T. Barron (Google Research)

DIVeR: Real-Time and Accurate Neural Radiance Fields With Deterministic Integration for Volume Rendering 16179
Liwen Wu (University of Illinois Urbana-Champaign), Jae Yong Lee (University of Illinois), Anand Bhattad (UIUC), Yu-Xiong Wang (University of Illinois at Urbana-Champaign), and David Forsyth (Univeristy of Illinois at Urbana-Champaign)

HumanNeRF: Free-Viewpoint Rendering of Moving People From Monocular Video 16189
Chung-Yi Weng (University of Washington), Brian Curless (University of Washington), Pratul P. Srinivasan (Google Research), Jonathan T. Barron (Google Research), and Ira Kemelmacher-Shlizerman (University of Washington)

Neural Reflectance for Shape Recovery With Shadow Handling 16200
Junxuan Li (The Australian National University) and Hongdong Li (Australian National University, Australia)

Visual Vibration Tomography: Estimating Interior Material Properties From Monocular Video .. 16210
Berthy T. Feng (California Institute of Technology), Alexander C. Ogren (Caltech), Chiara Daraio (California Institute of Technology), and Katherine L. Bouman (Caltech)

Dancing Under the Stars: Video Denoising in Starlight 16220
Kristina Monakhova (UC Berkeley), Stephan R. Richter (Intel Labs), Laura Waller (UC Berkeley), and Vladlen Koltun (Apple)

BACON: Band-Limited Coordinate Networks for Multiscale Scene Representation 16231
David B. Lindell (Stanford University), Dave Van Veen (Stanford University), Jeong Joon Park (Stanford University), and Gordon Wetzstein (Stanford University)

Practical Stereo Matching via Cascaded Recurrent Network With Adaptive Correlation 16242
Jiankun Li (megvii), Peisen Wang (Megvii Inc), Pengfei Xiong (Tencent), Tao Cai (Megvii Inc), Ziwei Yan (Beihang University), Lei Yang (Megvii Inc.), Jiangyu Liu (megvii inc), Haoqiang Fan (Megvii Incface++), and Shuaicheng Liu (UESTC; Megvii)

3D Photo Stylization: Learning To Generate Stylized Novel Views From a Single Image 16252
Fangzhou Mu (University of Wisconsin-Madison), Jian Wang (Snap), Yicheng Wu (Snap Inc.), and Yin Li (University of Wisconsin-Madison)

BokehMe: When Neural Rendering Meets Classical Rendering 16262
Juewen Peng (Huazhong University of Science and Technology), Zhiguo Cao (Huazhong Univ. of Sci.&Tech.), Xianrui Luo (Huazhong University of Science and Technology), Hao Lu (Huazhong University of Science and Technology), Ke Xian (Huazhong University of Science and Technology), and Jianming Zhang (Adobe Research)

Deblurring via Stochastic Refinement 16272
Jay Whang (The University of Texas at Austin), Mauricio Delbracio (Google Research), Hossein Talebi (Google), Chitwan Saharia (Google), Alexandros G. Dimakis (UT Austin), and Peyman Milanfar (Google)

Learning to Deblur Using Light Field Generated and Real Defocus Images 16283
Lingyan Ruan (City University of Hong Kong), Bin Chen (Max-Planck-Institut für Informatik), Jizhou Li (Stanford University), and Miuling Lam (City University of Hong Kong)

Towards Layer-Wise Image Vectorization 16293
Xu Ma (Northeastern University), Yuqian Zhou (Adobe), Xingqian Xu (UIUC), Bin Sun (Northeastern University), Valerii Filev (Picsart), Nikita Orlov (PicsArt), Yun Fu (Northeastern University), and Humphrey Shi (U of Oregon/UIUC)

Dual-Shutter Optical Vibration Sensing 16303
Mark Sheinin (Carnegie mellon University), Dorian Chan (Carnegie Mellon University), Matthew O'Toole (Carnegie Mellon University), and Srinivasa G. Narasimhan (Carnegie Mellon University, USA)

Fisher Information Guidance for Learned Time-of-Flight Imaging 16313
Jiaqu Li (Nanjing University), Tao Yue (Nanjing University), Sijie Zhao (Nanjing University), and Xuemei Hu (Nanjing University)

Autofocus for Event Cameras 16323
Shijie Lin (The University of Hong Kong), Yinqiang Zhang (The University of Hong Kong), Lei Yu (Wuhan University), Bin Zhou (State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science and Engineering, Beihang University), Xiaowei Luo (City University of Hong Kong), and Jia Pan (The University of Hong Kong)

Adaptive Gating for Single-Photon 3D Imaging 16333
Ryan Po (Carnegie Mellon University), Adithya Pediredla (CMU), and Ioannis Gkioulekas (Carnegie Mellon University)

LiDAR Snowfall Simulation for Robust 3D Object Detection 16343
Martin Hahner (ETH Zurich), Christos Sakaridis (ETH Zurich), Mario Bijelic (Princeton University), Felix Heide (Princeton University), Fisher Yu (ETH Zurich), Dengxin Dai (MPI for Informatics), and Luc Van Gool (ETH Zurich)

Oral 4.1.3: Vision & Language

MERLOT Reserve: Neural Script Knowledge Through Vision and Language and Sound 16354
Rowan Zellers (University of Washington), Jiasen Lu (Allen Institute for AI), Ximing Lu (Allen Institute for AI), Youngjae Yu (Seoul National University Vision and Learning Lab), Yanpeng Zhao (University of Edinburgh), Mohammadreza Salehi (University of Washington), Aditya Kusupati (University of Washington), Jack Hessel (Allen Institute for AI), Ali Farhadi (University of Washington, Allen Institue for AI, Apple), and Yejin Choi (University of Washington)

Joint Video Summarization and Moment Localization by Cross-Task Sample Transfer 16367
Hao Jiang (Peking University) and Yadong Mu (Peking University)

Towards General Purpose Vision Systems: An End-to-End Task-Agnostic Vision-Language Architecture 16378
Tanmay Gupta (Allen Institute for Artificial Intelligence), Amita Kamath (Allen Institute for Artificial Intelligence), Aniruddha Kembhavi (Allen Institute for Artificial Intelligence), and Derek Hoiem (University of Illinois at Urbana-Champaign)

Disentangling Visual and Written Concepts in CLIP 16389
Joanna Materzyńska (MIT), Antonio Torralba (MIT), and David Bau (Northeastern University)

CLIP-Event: Connecting Text and Images With Event Structures 16399
Manling Li (University of Illinois at Urbana-Champaign), Ruochen Xu (Microsoft), Shuohang Wang (Microsoft), Luowei Zhou (Microsoft), Xudong Lin (Columbia University), Chenguang Zhu (Microsoft), Michael Zeng (Microsoft), Heng Ji (University of Illinois at Urbana-Champaign), and Shih-Fu Chang (Columbia University)

Robust Cross-Modal Representation Learning With Progressive Self-Distillation 16409
Alex Andonian (MIT), Shixing Chen (Amazon), and Raffay Hamid (Amazon)

TubeDETR: Spatio-Temporal Video Grounding With Transformers 16421
Antoine Yang (Inria), Antoine Miech (DeepMind), Josef Sivic (Czech Technical University), Ivan Laptev (INRIA Paris), and Cordelia Schmid (Inria/Google)

3D-SPS: Single-Stage 3D Visual Grounding via Referred Point Progressive Selection 16433
Junyu Luo (Beihang University), Jiahui Fu (Beihang University), Xianghao Kong (Beihang University), Chen Gao (Beihang University), Haibing Ren (Meituan), Hao Shen (Meituan), Huaxia Xia (Meituan), and Si Liu (Beihang University)

3DJCG: A Unified Framework for Joint Dense Captioning and Visual Grounding on 3D Point Clouds 16443
Daigang Cai (Beihang University), Lichen Zhao (Beihang University), Jing Zhang (Beihang University), Lu Sheng (Beihang University), and Dong Xu (University of Sydney)

Globetrotter: Connecting Languages by Connecting Images 16453
Dídac Surís (Columbia University), Dave Epstein (UC Berkeley), and Carl Vondrick (Columbia University)

Unsupervised Vision-and-Language Pre-Training via Retrieval-Based Multi-Granular Alignment..... 16464
Mingyang Zhou (University of California, Davis), Licheng Yu (Facebook), Amanpreet Singh (Facebook), Mengjiao Wang (Facebook), Zhou Yu (Columbia), and Ning Zhang (Facebook)

WebQA: Multihop and Multimodal QA 16474
Yingshan Chang (Carnegie Mellon University), Mridu Narang (Carnegie Mellon University), Hisami Suzuki (Microsoft Corporation), Guihong Cao (Microsoft), Jianfeng Gao (Microsoft Corporation), and Yonatan Bisk (Microsoft Research)

PartGlot: Learning Shape Part Segmentation From Language Reference Games 16484
Juil Koo (KAIST), Ian Huang (Stanford University), Panos Achlioptas (Stanford University), Leonidas J. Guibas (Stanford University), and Minhyuk Sung (KAIST)

DF-GAN: A Simple and Effective Baseline for Text-to-Image Synthesis 16494
Ming Tao (Nanjing University Of Posts And Telecommunications), Hao Tang (ETH Zurich), Fei Wu (Nanjing University of Posts and Telecommunications), Xiao-Yuan Jing (School of Computer, Wuhan University), Bing-Kun Bao (Nanjing University of Posts and Telecommunications), and Changsheng Xu (CASIA)

L-Verse: Bidirectional Generation Between Image and Text 16505
Taehoon Kim (LG AI Research), Gwangmo Song (LG AI Research), Sihaeng Lee (LG AI Research), Sangyun Kim (LG AI Research), Yewon Seo (LG AI Research), Soonyoung Lee (LG AI Research), Seung Hwan Kim (LG AI Research), Honglak Lee (LG AI Research), and Kyunghoon Bae (LG AI Research)

Think Global, Act Local: Dual-Scale Graph Transformer for Vision-and-Language Navigation ... 16516
Shizhe Chen (INRIA), Pierre-Louis Guhur (Inria), Makarand Tapaswi (Wadhwani AI, IIIT Hyderbad), Cordelia Schmid (Inria/Google), and Ivan Laptev (INRIA Paris)

LaTr: Layout-Aware Transformer for Scene-Text VQA 16527
Ali Furkan Biten (Computer Vision Center), Ron Litman (Amazon), Yusheng Xie (Amazon), Srikar Appalaraju (Amazon), and R. Manmatha (Amazon)

Learning Program Representations for Food Images and Cooking Recipes 16538
Dim P. Papadopoulos (MIT), Enrique Mora (Nestlé), Nadiia Chepurko (MIT), Kuan Wei Huang (Massachusetts Institute of Technology), Ferda Ofli (Qatar Computing Research Institute, HBKU), and Antonio Torralba (MIT)

Poster 4.1: Representation Learning

On the Importance of Asymmetry for Siamese Representation Learning 16549
Xiao Wang (Purdue University), Haoqi Fan (Facebook AI Research), Yuandong Tian (Facebook), Daisuke Kihara (Purdue University), and Xinlei Chen (FAIR)

Leverage Your Local and Global Representations: A New Self-Supervised Learning Strategy 16559
Tong Zhang (EPFL), Congpei Qiu (Xi'an Jiaotong University), Wei Ke (Xi'an Jiaotong University), Sabine Süsstrunk (EPFL), and Mathieu Salzmann (EPFL)

Exploring Set Similarity for Dense Self-Supervised Representation Learning 16569
Zhaoqing Wang (University of Sydney), Qiang Li (Kuaishou Technology), Guoxin Zhang (Kuaishou), Pengfei Wan (Kuaishou Technology), Wen Zheng (Kuaishou Technology), Nannan Wang (Xidian University), Mingming Gong (University of Melbourne), and Tongliang Liu (The University of Sydney)

Align Representations With Base: A New Approach to Self-Supervised Learning 16579
Shaofeng Zhang (Shanghai Jiao Tong University), Lyn Qiu (Shanghai Jiao Tong University), Feng Zhu (University of Science and Technology of China), Junchi Yan (Shanghai Jiao Tong University), Hengrui Zhang (University of Illinois at Chicago), Rui Zhao (SenseTime Group Limited), Hongyang Li (SenseTime), and Xiaokang Yang (Shanghai Jiao Tong University of China)

Identifying Ambiguous Similarity Conditions via Semantic Matching 16589
Han-Jia Ye (Nanjing University), Yi Shi (Nanjing University), and De-Chuan Zhan (Nanjing University)

Node Representation Learning in Graph via Node-to-Neighbourhood Mutual Information Maximization 16599
Wei Dong (School of Computer Science and Engineering, Northwestern Polytechnical University), Junsheng Wu (School of Software and Microelectronics, Northwestern Polytechnical University), Yi Luo (Northwest Polytechnical University), Zongyuan Ge (Monash), and Peng Wang (University of Wollongong)

Instance-Dependent Label-Noise Learning With Manifold-Regularized Transition Matrix Estimation 16609
De Cheng (Xidian University), Tongliang Liu (The University of Sydney), Yixiong Ning (Xidian University), Nannan Wang (Xidian University), Bo Han (HKBU / RIKEN), Gang Niu (RIKEN), Xinbo Gao (Chongqing University of Posts and Telecommunications), and Masashi Sugiyama (RIKEN/The University of Tokyo)

Unsupervised Visual Representation Learning by Online Constrained K-Means 16619
Qi Qian (Alibaba Group), Yuanhong Xu (Alibaba Group), Juhua Hu (University of Washington), Hao Li (Alibaba Group), and Rong Jin (alibaba group)

Rethinking the Augmentation Module in Contrastive Learning: Learning Hierarchical Augmentation Invariance With Expanded Views 16629
Junbo Zhang (Tsinghua University) and Kaisheng Ma (Tsinghua University)

Use All the Labels: A Hierarchical Multi-Label Contrastive Learning Framework 16639
Shu Zhang (Salesforce Research), Ran Xu (Salesforce Research), Caiming Xiong (Salesforce Research), and Chetan Ramaiah (Salesforce Research)

Robust Contrastive Learning Against Noisy Views 16649
Ching-Yao Chuang (MIT), R Devon Hjelm (Microsoft Research), Xin Wang (Microsoft Research), Vibhav Vineet (Microsoft Research), Neel Joshi (Microsoft Research), Antonio Torralba (MIT), Stefanie Jegelka (MIT), and Yale Song (Microsoft Research)

On Learning Contrastive Representations for Learning With Noisy Labels 16661
Li Yi (University of Western Ontario), Sheng Liu (NYU), Qi She (Bytedance AI Lab), A. Ian McLeod (University of Western Ontario), and Boyu Wang (University of Western Ontario)

Directional Self-Supervised Learning for Heavy Image Augmentations 16671
Yalong Bai (JD AI Research), Yifan Yang (Microsoft), Wei Zhang (JD AI Research), and Tao Mei (AI Research of JD.com)

Continual Learning for Visual Search With Backward Consistent Feature Embedding 16681
Timmy S. T. Wan (National Taiwan University), Jun-Cheng Chen (Academia Sinica), Tzer-Yi Wu (ucfunnel Co. Ltd.), and Chu-Song Chen (National Taiwan University)

Probing Representation Forgetting in Supervised and Unsupervised Continual Learning 16691
MohammadReza Davari (Mila and Concordia University), Nader Asadi (Mila and Concordia University), Sudhir Mudur (Concordia University), Rahaf Aljundi (Toyota Motor Europe), and Eugene Belilovsky (Mila)

Mimicking the Oracle: An Initial Phase Decorrelation Approach for Class Incremental Learning 16701
Yujun Shi (NUS), Kuangqi Zhou (National University of Singapore), Jian Liang (NUS), Zihang Jiang (National University of Singapore), Jiashi Feng (ByteDance), Philip H.S. Torr (University of Oxford), Song Bai (University of Oxford), and Vincent Y. F. Tan (NUS)

Bring Evanescent Representations to Life in Lifelong Class Incremental Learning 16711
Marco Toldo (University of Padova) and Mete Ozay (Samsung Research UK)

Unsupervised Learning of Debiased Representations With Pseudo-Attributes 16721
Seonguk Seo (Seoul National University), Joon-Young Lee (Adobe Research), and Bohyung Han (Seoul National University)

A Conservative Approach for Unbiased Learning on Unknown Biases 16731
Myeongho Jeon (Seoul National University), Daekyung Kim (Monitorcorp & Seoul National University), Woochul Lee (Seoul National University), Myungjoo Kang (Seoul National University), and Joonseok Lee (Google Research & Seoul National University)

Evading the Simplicity Bias: Training a Diverse Set of Models Discovers Solutions With Superior OOD Generalization 16740
Damien Teney (University of Adelaide), Ehsan Abbasnejad (The University of Adelaide), Simon Lucey (University of Adelaide), and Anton van den Hengel (University of Adelaide)

Co-Advise: Cross Inductive Bias Distillation 16752
Sucheng Ren (South China University of Technology), Zhengqi Gao (MIT), Tianyu Hua (Tsinghua University), Zihui Xue (The University of Texas at Austin), Yonglong Tian (MIT), Shengfeng He (South China University of Technology), and Hang Zhao (Tsinghua University)

PixMix: Dreamlike Pictures Comprehensively Improve Safety Measures 16762
Dan Hendrycks (UC Berkeley), Andy Zou (UC Berkeley), Mantas Mazeika (UIUC), Leonard Tang (NVIDIA), Bo Li (UIUC), Dawn Song (UC Berkeley), and Jacob Steinhardt (UC Berkeley)

RegionCLIP: Region-Based Language-Image Pretraining 16772
Yiwu Zhong (University of Wisconsin-Madison), Jianwei Yang (Microsoft Research), Pengchuan Zhang (Microsoft Research AI), Chunyuan Li (Microsoft Research), Noel Codella (Microsoft), Liunian Harold Li (UCLA), Luowei Zhou (Microsoft), Xiyang Dai (Microsoft), Lu Yuan (Microsoft), Yin Li (University of Wisconsin-Madison), and Jianfeng Gao (Microsoft Research)

Uni-Perceiver: Pre-Training Unified Architecture for Generic Perception for Zero-Shot and Few-Shot Tasks 16783
Xizhou Zhu (SenseTime), Jinguo Zhu (Xi'an Jiaotong University), Hao Li (Chinese University of Hong Kong), Xiaoshi Wu (Chinese University of Hong Kong), Hongsheng Li (The Chinese University of Hong Kong), Xiaohua Wang (Xi'an Jiaotong University), and Jifeng Dai (SenseTime)

Conditional Prompt Learning for Vision-Language Models 16795
Kaiyang Zhou (Nanyang Technological University), Jingkang Yang (Nanyang Technological University), Chen Change Loy (Nanyang Technological University), and Ziwei Liu (Nanyang Technological University)

Poster 4.1: Scene Analysis & Understanding

Noisy Boundaries: Lemon or Lemonade for Semi-Supervised Instance Segmentation? 16805
Zhenyu Wang (Tsinghua University), Yali Li (Tsinghua University), and Shengjin Wang (Tsinghua University)

Partial Class Activation Attention for Semantic Segmentation 16815
Sun-Ao Liu (University of Science and Technology of China), Hongtao Xie (University of Science and Technology of China), Hai Xu (University of Science and Technology of China), Yongdong Zhang (University of Science and Technology of China), and Qi Tian (Huawei Cloud & AI)

Learning Affinity From Attention: End-to-End Weakly-Supervised Semantic Segmentation With Transformers 16825
Lixiang Ru (Wuhan University), Yibing Zhan (Jingdong), Baosheng Yu (The University of Sydney), and Bo Du (Wuhan University)

Towards Noiseless Object Contours for Weakly Supervised Semantic Segmentation 16835
Jing Li (Chinese Academy of Sciences, China), Junsong Fan (Chinese Academy of Sciences, China), and Zhaoxiang Zhang (Chinese Academy of Sciences, China)

Class Similarity Weighted Knowledge Distillation for Continual Semantic Segmentation 16845
Minh Hieu Phan (University of Wollongong), The-Anh Ta (FPT Software, AIC), Son Lam Phung (University of Wollongong, VinAI), Long Tran-Thanh (University of Warwick), and Abdesselam Bouzerdoum (University of Wollongong)

Structural and Statistical Texture Knowledge Distillation for Semantic Segmentation 16855
Deyi Ji (DAMO Academy, Alibaba Group), Haoran Wang (Xidian University), Mingyuan Tao (Damo Academy, Alibaba Group), Jianqiang Huang (Alibaba Group), Xian-Sheng Hua (Damo Academy, Alibaba Group), and Hongtao Lu (Shanghai Jiao Tong University)

L2G: A Simple Local-to-Global Knowledge Transfer Framework for Weakly Supervised Semantic Segmentation 16865
Peng-Tao Jiang (Nankai University), Yuqi Yang (Nankai University), Qibin Hou (Nankai University), and Yunchao Wei (UTS)

Weakly Supervised Semantic Segmentation Using Out-of-Distribution Data 16876
Jungbeom Lee (Seoul National University), Seong Joon Oh (NAVER AI Lab), Sangdoo Yun (NAVER AI LAB), Junsuk Choe (Sogang University), Eunji Kim (Seoul National University), and Sungroh Yoon (Seoul National University)

Tree Energy Loss: Towards Sparsely Annotated Semantic Segmentation 16886
Zhiyuan Liang (Beijing Institute of technology), Tiancai Wang (Megvii Technology), Xiangyu Zhang (Megvii Technology), Jian Sun (Megvii Technology), and Jianbing Shen (Inception Institute of Artificial Intelligence)

Bending Reality: Distortion-Aware Transformers for Adapting to Panoramic Semantic Segmentation 16896
Jiaming Zhang (Karlsruhe Institute of Technology), Kailun Yang (Karlsruhe Institute of Technology), Chaoxiang Ma (Karlsruhe institute of technology), Simon Reiß (Karlsruhe Institute of Technology), Kunyu Peng (KIT), and Rainer Stiefelhagen (Karlsruhe Institute of Technology)

MM-TTA: Multi-Modal Test-Time Adaptation for 3D Semantic Segmentation 16907
Inkyu Shin (KAIST), Yi-Hsuan Tsai (Phiar Technologies), Bingbing Zhuang (NEC Labs), Samuel Schulter (NEC Laboratories America), Buyu Liu (NEC Labs), Sparsh Garg (NEC Labs America), In So Kweon (KAIST), and Kuk-Jin Yoon (KAIST)

NightLab: A Dual-Level Architecture With Hardness Detection for Segmentation at Night 16917
Xueqing Deng (University of California, Merced), Peng Wang (Bytedance USA LLC.), Xiaochen Lian (ByteDance), and Shawn Newsam (UC Merced)

Fast Point Transformer 16928
Chunghyun Park (POSTECH), Yoonwoo Jeong (POSTECH), Minsu Cho (POSTECH), and Jaesik Park (POSTECH)

RigidFlow: Self-Supervised Scene Flow Learning on Point Clouds by Local Rigidity Prior 16938
Ruibo Li (Nanyang Technological University), Chi Zhang (Nanyang Technological University), Guosheng Lin (Nanyang Technological University), Zhe Wang (SenseTime Research), and Chunhua Shen (University of Adelaide, Australia)

ConDor: Self-Supervised Canonicalization of 3D Pose for Partial Shapes 16948
Rahul Sajnani (Robotics Research Center, KCIS, International Institute of Information Technology, Hyderabad), Adrien Poulenard (Stanford), Jivitesh Jain (International Institute of Information Technology, Hyderabad), Radhika Dua (KAIST), Leonidas J. Guibas (Stanford University), and Srinath Sridhar (Brown University)

DisARM: Displacement Aware Relation Module for 3D Detection 16959
Yao Duan (National University of Defense Technology), Chenyang Zhu (National University of Defense Technology), Yuqing Lan (National University of Defense Technology), Renjiao Yi (National University of Defense Technology), Xinwang Liu (National University of Defense Technology), and Kai Xu (National University of Defense Technology)

Learning Object Context for Novel-View Scene Layout Generation 16969
Xiaotian Qiao (Xidian University), Gerhard P. Hancke (City University of Hong Kong), and Rynson W.H. Lau (City University of Hong Kong)

Weakly but Deeply Supervised Occlusion-Reasoned Parametric Road Layouts 16979
Buyu Liu (NEC Labs), Bingbing Zhuang (NEC Labs), and Manmohan Chandraker (UC San Diego)

Beyond Cross-View Image Retrieval: Highly Accurate Vehicle Localization Using Satellite Image 16989
Yujiao Shi (ANU) and Hongdong Li (Australian National University, Australia)

Raw High-Definition Radar for Multi-Task Learning 17000
Julien Rebut (Valeo AI), Arthur Ouaknine (Telecom Paris), Waqas Malik (Valeo), and Patrick Pérez (Valeo.ai)

Zero Experience Required: Plug & Play Modular Transfer Learning for Semantic Visual Navigation 17010
Ziad Al-Halah (UT Austin), Santhosh Kumar Ramakrishnan (University of Texas at Austin), and Kristen Grauman (Facebook AI Research & UT Austin)

UKPGAN: A General Self-Supervised Keypoint Detector 17021
Yang You (Shanghai Jiao Tong University), Wenhai Liu (Shanghai Jiao Tong University), Yanjie Ze (Shanghai Jiao Tong University), Yong-Lu Li (Shanghai Jiao Tong University), Weiming Wang (shanghai jiaotong university), and Cewu Lu (Shanghai Jiao Tong University)

Cannot See the Forest for the Trees: Aggregating Multiple Viewpoints To Better Classify Objects in Videos 17031
Sukjun Hwang (Yonsei University), Miran Heo (Yonsei University), Seoung Wug Oh (Adobe Research), and Seon Joo Kim (Yonsei University)

Poster 4.1: Navigation & Autonomous Driving

Rethinking Efficient Lane Detection via Curve Modeling 17041
Zhengyang Feng (Shanghai Jiao Tong University), Shaohua Guo (Shanghai Jiao Tong University), Xin Tan (Shanghai Jiao Tong University), Ke Xu (City University of Hong Kong), Min Wang (Sensetime Group), and Lizhuang Ma (Shanghai Jiao Tong University)

Exploiting Temporal Relations on Radar Perception for Autonomous Driving 17050
Peizhao Li (Brandeis University), Pu Wang (MERL), Karl Berntorp (Mitsubishi Electric Research Labs), and Hongfu Liu (Brandeis University)

Towards Robust and Adaptive Motion Forecasting: A Causal Representation Perspective 17060
Yuejiang Liu (EPFL), Riccardo Cadei (EPFL), Jonas Schweizer (EPFL), Sherwin Bahmani (TU Darmstadt), and Alexandre Alahi (EPFL)

BE-STI: Spatial-Temporal Integrated Network for Class-Agnostic Motion Prediction With Bidirectional Enhancement 17072
Yunlong Wang (Tsinghua University), Hongyu Pan (DAMO Academy, Alibaba Group), Jun Zhu (DAMO Academy, Alibaba Group), Yu-Huan Wu (Nankai University), Xin Zhan (DAMO Academy, Alibaba Group), Kun Jiang (Tsinghua University), and Diange Yang (Tsinghua University)

ScePT: Scene-Consistent, Policy-Based Trajectory Predictions for Planning 17082
Yuxiao Chen (Nvidia Corporation), Boris Ivanovic (Stanford University), and Marco Pavone (Stanford University)

Stochastic Trajectory Prediction via Motion Indeterminacy Diffusion 17092
Tianpei Gu (UCLA), Guangyi Chen (MBZUAI), Junlong Li (Tsinghua University), Chunze Lin (Sense Time), Yongming Rao (Tsinghua University), Jie Zhou (Tsinghua University), and Jiwen Lu (Tsinghua University)

Vehicle Trajectory Prediction Works, but Not Everywhere 17102
Mohammadhossein Bahari (EPFL), Saeed Saadatnejad (EPFL), Ahmad Rahimi (Sharif University of Technology), Mohammad Shaverdikondori (Sharif University of Technology), Amir Hossein Shahidzadeh (Sharif University of Technology), Seyed-Mohsen Moosavi-Dezfooli (Imperial College London), and Alexandre Alahi (EPFL)

LTP: Lane-Based Trajectory Prediction for Autonomous Driving 17113
Jingke Wang (Alibaba Damo Acadamy), Tengju Ye (Alibaba Damo Acadamy), Ziqing Gu (Alibaba Damo Acadamy), and Junbo Chen (Alibaba Damo Acadamy)

ONCE-3DLanes: Building Monocular 3D Lane Detection 17122
Fan Yan (Fudan University), Ming Nie (Fudan University), Xinyue Cai (Huawei Noah's Ark Lab), Jianhua Han (Huawei Noah's Ark Lab), Hang Xu (Huawei Noah's Ark Lab), Zhen Yang (Huawei Noah's Ark Lab), Chaoqiang Ye (Huawei), Yanwei Fu (Fudan University), Michael Bi Mi (Huawei International Pte Ltd), and Li Zhang (Fudan University)

Towards Driving-Oriented Metric for Lane Detection Models 17132
Takami Sato (University of California, Irvine) and Qi Alfred Chen (UC Irvine)

Eigenlanes: Data-Driven Lane Descriptors for Structurally Diverse Lanes 17142
Dongkwon Jin (Korea university), Wonhui Park (Korea university), Seong-Gyun Jeong (42dot.ai), Heeyeon Kwon (42dot), and Chang-Su Kim (Korea university)

LIFT: Learning 4D LiDAR Image Fusion Transformer for 3D Object Detection 17151
Yihan Zeng (Shanghai Jiao Tong University), Da Zhang (UC Santa Barbara), Chunwei Wang (Shanghai Jiao Tong University), Zhenwei Miao (DAMO Academy, Alibaba Group), Ting Liu (Alibaba), Xin Zhan (DAMO Academy, Alibaba Group), Dayang Hao (A), and Chao Ma (Shanghai Jiao Tong University)

DeepFusion: Lidar-Camera Deep Fusion for Multi-Modal 3D Object Detection 17161
Yingwei Li (Johns Hopkins University), Adams Wei Yu (Google Brain), Tianjian Meng (The University of Texas at Austin), Ben Caine (Google), Jiquan Ngiam (Google Brain), Daiyi Peng (Google Brain), Junyang Shen (Google), Yifeng Lu (Google Brain), Denny Zhou (Google Brain), Quoc V. Le (Google Brain), Alan Yuille (Johns Hopkins University), and Mingxing Tan (Waymo)

A Versatile Multi-View Framework for LiDAR-Based 3D Object Detection With Guidance From Panoptic Segmentation 17171
Hamidreza Fazlali (Huawei Noah's Ark Lab), Yixuan Xu (Huawei Noah's Ark Lab), Yuan Ren (Huawei), and Bingbing Liu (Huawei Noah's Ark Lab, Canada)

Forecasting From LiDAR via Future Object Detection 17181
Neehar Peri (Carnegie Mellon University), Jonathon Luiten (RWTH Aachen University), Mengtian Li (Carnegie Mellon University), Aljoša Ošep (TUM Munich), Laura Leal-Taixé (TUM), and Deva Ramanan (Carnegie Mellon University)

RIDDLE: Lidar Data Compression With Range Image Deep Delta Encoding 17191
Xuanyu Zhou (Waymo), Charles R. Qi (Waymo), Yin Zhou (Waymo), and Dragomir Anguelov (Waymo)

Learning From All Vehicles 17201
Dian Chen (UT Austin) and Philipp Krähenbühl (UT Austin)

Is Mapping Necessary for Realistic PointGoal Navigation? 17211
Ruslan Partsey (Ukrainian Catholic University), Erik Wijmans (Georgia Tech), Naoki Yokoyama (Georgia Institute of Technology), Oles Dobosevych (Ukrainian Catholic University), Dhruv Batra (Georgia Tech & Facebook AI Research), and Oleksandr Maksymets (Facebook AI Research)

Symmetry-Aware Neural Architecture for Embodied Visual Exploration 17221
Shuang Liu (RIKEN AIP) and Takayuki Okatani (Tohoku University/RIKEN AIP)

Coopernaut: End-to-End Driving With Cooperative Perception for Networked Vehicles 17231
Jiaxun Cui (The University of Texas at Austin), Hang Qiu (Stanford University), Dian Chen (UT Austin), Peter Stone (University of Texas at Austin and Sony AI), and Yuke Zhu (University of Texas - Austin)

Topology Preserving Local Road Network Estimation From Single Onboard Camera Image 17242
Yigit Baran Can (ETH Zurich), Alexander Liniger (ETH Zurich), Danda Pani Paudel (ETH Zürich), and Luc Van Gool (ETH Zurich)

Coupling Vision and Proprioception for Navigation of Legged Robots 17252
Zipeng Fu (Carnegie Mellon University), Ashish Kumar (UC Berkeley), Ananye Agarwal (Carnegie Mellon University), Haozhi Qi (UC Berkeley), Jitendra Malik (University of California at Berkeley), and Deepak Pathak (Carnegie Mellon University)

Pyramid Architecture for Multi-Scale Processing in Point Cloud Segmentation 17263
Dong Nie (UNC), Rui Lan (Alibaba Inc), Ling Wang (Alibaba Group), and Xiaofeng Ren (alibaba group)

3D-VField: Adversarial Augmentation of Point Clouds for Domain Generalization in 3D Object Detection 17274
Alexander Lehner (Technical University of Munich, BMW Group), Stefano Gasperini (Technical University of Munich, BMW Group), Alvaro Marcos-Ramiro (BMW Group), Michael Schmidt (BMW Group), Mohammad-Ali Nikouei Mahani (BMW), Nassir Navab (TU Munich, Germany), Benjamin Busam (Technical University of Munich), and Federico Tombari (Google, TU Munich)

Generating Useful Accident-Prone Driving Scenarios via a Learned Traffic Prior 17284
Davis Rempe (Stanford University), Jonah Philion (University of Toronto, NVIDIA), Leonidas J. Guibas (Stanford University), Sanja Fidler (University of Toronto, NVIDIA), and Or Litany (NVIDIA)

SelfD: Self-Learning Large-Scale Driving Policies From the Web 17295
Jimuyang Zhang (Boston University), Ruizhao Zhu (Boston University), and Eshed Ohn-Bar (Boston University)

Towards Real-World Navigation With Deep Differentiable Planners 17306
Shu Ishida (University of Oxford) and João F. Henriques (University of Oxford)

Privacy Preserving Partial Localization 17316
Marcel Geppert (ETH Zurich), Viktor Larsson (ETH Zurich), Johannes L. Schönberger (Microsoft), and Marc Pollefeys (ETH Zurich / Microsoft)

Efficient Large-Scale Localization by Global Instance Recognition 17327
Fei Xue (Cambridge University), Ignas Budvytis (Department of Engineering, University of Cambridge), Daniel Olmeda Reino (Toyota Motor Europe), and Roberto Cipolla (University of Cambridge)

CrossLoc: Scalable Aerial Localization Assisted by Multimodal Synthetic Data 17337
Qi Yan (UBC), Jianhao Zheng (EPFL), Simon Reding (EPFL), Shanci Li (EPFL), and Iordan Doytchinov (EPFL)

Poster 4.1: Low-Level Vision

Bilateral Video Magnification Filter 17348
Shoichiro Takeda (NTT Corporation), Kenta Niwa (NTT), Mariko Isogawa (NTT), Shinya Shimizu (NTT Corporation), Kazuki Okami (NTT TechnoCross Corporation), and Yushi Aono (NTT Corporation)

Neural Data-Dependent Transform for Learned Image Compression 17358
Dezhao Wang (Peking University), Wenhan Yang (City University of Hong Kong), Yueyu Hu (Peking University), and Jiaying Liu (Peking University)

Towards Bidirectional Arbitrary Image Rescaling: Joint Optimization and Cycle Idempotence ... 17368
Zhihong Pan (Baidu USA LLC), Baopu Li (BAIDU USA LLC), Dongliang He (Baidu), Mingde Yao (University of Science and Technology of China), Wenhao Wu (Baidu), Tianwei Lin (Baidu Inc), Xin Li (Baidu), and Errui Ding (Baidu Inc.)

Deep Generalized Unfolding Networks for Image Restoration 17378
Chong Mou (Peking University Shenzhen Graduate School), Qian Wang (Peking University Shenzhen Graduate School), and Jian Zhang (Peking University Shenzhen Graduate School)

Look Back and Forth: Video Super-Resolution With Explicit Temporal Difference Modeling 17390
Takashi Isobe (Tsinghua University), Xu Jia (Dalian University of Technology), Xin Tao (Kuaishou), Changlin Li (Kuaishou), Ruihuang Li (Hong Kong Polytechnic University), Yongjie Shi (Peking University), Jing Mu (Kuaishou Technology), Huchuan Lu (Dalian University of Technology), and Yu-Wing Tai (Kuaishou Technology / HKUST)

XYDeblur: Divide and Conquer for Single Image Deblurring 17400
Seo-Won Ji (Korea University), Jeongmin Lee (Korea University), Seung-Wook Kim (Korea University), Jun-Pyo Hong (Korea University), Seung-Jin Baek (Korea University), Seung-Won Jung (Korea University), and Sung-Jea Ko (Korea University)

Abandoning the Bayer-Filter To See in the Dark 17410
Xingbo Dong (Yonsei University; TCL Research AI Lab; Monash University), Wanyan Xu (Fuzhou University), Zhihui Miao (Fuzhou University), Lan Ma (TCL Corporate Research), Chao Zhang (Anhui University), Jiewen Yang (Yonsei University), Zhe Jin (TCL Corporate Research (HK) Co., Ltd, ,), Andrew Beng Jin Teoh (TCL Corporate Research), and Jiajun Shen (TCL Research)

RSTT: Real-Time Spatial Temporal Transformer for Space-Time Video Super-Resolution 17420
Zhicheng Geng (The University of Texas at Austin), Luming Liang (Microsoft), Tianyu Ding (Microsoft), and Ilya Zharkov (Microsoft)

All-in-One Image Restoration for Unknown Corruption 17431
Boyun Li (College of Computer Science, Sichuan University), Xiao Liu (Tomorrow Advancing Life Education Group), Peng Hu (College of Computer Science, Sichuan University), Zhongqin Wu (Tomorrow Advancing Life Education Group), Jiancheng Lv (Sichuan University), and Xi Peng (College of Computer Science, Sichuan Univerisity)

Modeling sRGB Camera Noise With Normalizing Flows 17442
Shayan Kousha (York University), Ali Maleky (York University), Michael S. Brown (York University), and Marcus A. Brubaker (York University)

A Differentiable Two-Stage Alignment Scheme for Burst Image Reconstruction With Large Shift 17451
Shi Guo (The Hong Kong Polytechnic University), Xi Yang (The Hong Kong Polytechnic University), Jianqi Ma (the Hong Kong Polytechnic University), Gaofeng Ren (Alibaba), and Lei Zhang (Hong Kong Polytechnic University, Hong Kong, China)

Video Frame Interpolation Transformer 17461
Zhihao Shi (McMaster University), Xiangyu Xu (Sea AI Lab), Xiaohong Liu (Shanghai Jiao Tong University), Jun Chen (McMaster University), and Ming-Hsuan Yang (University of California at Merced)

The Devil Is in the Details: Window-Based Attention for Image Compression 17471
Renjie Zou (Institute of AutomationChinese Academy of Sciences), Chunfeng Song (CASIA, UCAS, CRIPAC), and Zhaoxiang Zhang (Chinese Academy of Sciences, China)

Mask-Guided Spectral-Wise Transformer for Efficient Hyperspectral Image Reconstruction 17481
Yuanhao Cai (Tsinghua Univisity, Tsinghua Shenzhen International Graduate School), Jing Lin (Tsinghua Univisity, Tsinghua Shenzhen International Graduate School), Xiaowan Hu (Tsinghua Univisity, Tsinghua Shenzhen International Graduate School), Haoqian Wang (Tsinghua Shenzhen International Graduate School, Tsinghua University), Xin Yuan (Westlake University), Yulun Zhang (ETH Zurich), Radu Timofte (ETH Zurich), and Luc Van Gool (ETH Zurich)

RestoreFormer: High-Quality Blind Face Restoration From Undegraded Key-Value Pairs 17491
Zhouxia Wang (The University of Hong Kong), Jiawei Zhang (Sensetime Research), Runjian Chen (The University of Hong Kong), Wenping Wang (The University of Hong Kong), and Ping Luo (The University of Hong Kong)

AdaInt: Learning Adaptive Intervals for 3D Lookup Tables on Real-Time Image Enhancement .. 17501
Canqian Yang (Shanghai Jiao Tong University), Meiguang Jin (Alibaba Group), Xu Jia (Dalian University of Technology), Yi Xu (Shanghai Jiao Tong University), and Ying Chen (Alibaba Group)

HerosNet: Hyperspectral Explicable Reconstruction and Optimal Sampling Deep Network for Snapshot Compressive Imaging 17511
Xuanyu Zhang (Peking University Shenzhen Graduate School), Yongbing Zhang (Harbin Institute of Technology (Shenzhen), ,), Ruiqin Xiong (Peking University), Qilin Sun (The Chinese University of Hong Kong, Shenzhen), and Jian Zhang (Peking University Shenzhen Graduate School)

HDNet: High-Resolution Dual-Domain Learning for Spectral Compressive Imaging 17521
Xiaowan Hu (Tsinghua Univisity, Tsinghua Shenzhen International Graduate School), Yuanhao Cai (Tsinghua Univisity, Tsinghua Shenzhen International Graduate School), Jing Lin (Tsinghua Univisity, Tsinghua Shenzhen International Graduate School), Haoqian Wang (Tsinghua Shenzhen International Graduate School, Tsinghua University), Xin Yuan (Westlake University), Yulun Zhang (ETH Zurich), Radu Timofte (ETH Zurich), and Luc Van Gool (ETH Zurich)

Learning To Zoom Inside Camera Imaging Pipeline 17531
Chengzhou Tang (Simon Fraser University), Yuqiang Yang (University of Electronic Science and Technology of China), Bing Zeng (University of Electronic Science and Technology of China), Ping Tan (Simon Fraser University), and Shuaicheng Liu (UESTC; Megvii)

Towards an End-to-End Framework for Flow-Guided Video Inpainting 17541
Zhen Li (Nankai University), Cheng-Ze Lu (Nankai University), Jianhua Qin (Hisilicon Technologies Co., Ltd.), Chun-Le Guo (Nankai University), and Ming-Ming Cheng (Nankai University)

Context-Aware Video Reconstruction for Rolling Shutter Cameras 17551
Bin Fan (Northwestern Polytechnical University), Yuchao Dai (Northwestern Polytechnical University), Zhiyuan Zhang (Northwestern Polytechnical University), Qi Liu (Northwestern Polytechnical University), and Mingyi He (Northwestern Polytechnical University)

CVF-SID: Cyclic Multi-Variate Function for Self-Supervised Image Denoising by Disentangling Noise From Image 17562
Reyhaneh Neshatavar (Seoul National University), Mohsen Yavartanoo (Seoul National University), Sanghyun Son (Seoul National University), and Kyoung Mu Lee (Seoul National University)

Global Matching With Overlapping Attention for Optical Flow Estimation 17571
Shiyu Zhao (Rutgers University), Long Zhao (Rutgers University), Zhixing Zhang (Rutgers University), Enyu Zhou (Sensetime), and Dimitris Metaxas (Rutgers)

CRAFT: Cross-Attentional Flow Transformer for Robust Optical Flow 17581
*Xiuchao Sui (IHPC), Shaohua Li (IHPC, A*STAR), Xue Geng (I2R, A*STAR), Yan Wu (A*STAR Institute for Infocomm Research), Xinxing Xu (Institute of High Performance Computing, A*Star), Yong Liu (IHPC), Rick Goh (Institute of High Performance Computing, A*STAR), and Hongyuan Zhu (Institute for Infocomm, Research Agency for Science, Technology and Research (A*STAR) Singapore)*

Unified Multivariate Gaussian Mixture for Efficient Neural Image Compression 17591
Xiaosu Zhu (University of Electronic Science and Technology of China), Jingkuan Song (UESTC), Lianli Gao (The University of Electronic Science and Technology of China), Feng Zheng (SUSTech), and Heng Tao Shen (University of Electronic Science and Technology of China (UESTC))

Video Demoiréing With Relation-Based Temporal Consistency 17601
Peng Dai (The University of Hong Kong), Xin Yu (The University of Hong Kong), Lan Ma (TCL Corporate Research), Baoheng Zhang (The University of HongKong), Jia Li (Sun Yat-Sen University), Wenbo Li (The Chinese University of Hong Kong), Jiajun Shen (TCL Research), and Xiaojuan Qi (The University of Hong Kong)

Noise2NoiseFlow: Realistic Camera Noise Modeling Without Clean Images 17611
Ali Maleky (York University), Shayan Kousha (York University), Michael S. Brown (York University), and Marcus A. Brubaker (York University)

Deep Constrained Least Squares for Blind Image Super-Resolution 17621
Ziwei Luo (Megvii), Haibin Huang (Kuaishou Technology), Lei Yu (Megvii), Youwei Li (Megvii), Haoqiang Fan (Megvii Incface++), and Shuaicheng Liu (UESTC; Megvii)

Learning Multiple Adverse Weather Removal via Two-Stage Knowledge Learning and Multi-Contrastive Regularization: Toward a Unified Model 17632
Wei-Ting Chen (National Taiwan University), Zhi-Kai Huang (National Taiwan University), Cheng-Che Tsai (National Taiwan University), Hao-Hsiang Yang (National Taiwan University), Jian-Jiun Ding (National Taiwan University), and Sy-Yen Kuo (National Taiwan University)

Unsupervised Homography Estimation With Coplanarity-Aware GAN 17642
Mingbo Hong (Sichuan Unviersity), Yuhang Lu (University of South Carolina), Nianjin Ye (Megvii), Chunyu Lin (Beijing Jiaotong University), Qijun Zhao (Sichuan University), and Shuaicheng Liu (UESTC; Megvii)

Poster 4.1: Computational Photography

Attentive Fine-Grained Structured Sparsity for Image Restoration 17652
Junghun Oh (Seoul National University), Heewon Kim (Seoul National University), Seungjun Nah (NVIDIA), Cheeun Hong (Seoul National University), Jonghyun Choi (Yonsei University), and Kyoung Mu Lee (Seoul National University)

Uformer: A General U-Shaped Transformer for Image Restoration 17662
Zhendong Wang (University of Science and Technology of China), Xiaodong Cun (Tencent AI Lab), Jianmin Bao (University of Science and Technology of China), Wengang Zhou (University of Science and Technology of China), Jianzhuang Liu (Huawei Noah's Ark Lab), and Houqiang Li (University of Science and Technology of China)

Bringing Old Films Back to Life 17673
Ziyu Wan (City University of Hong Kong), Bo Zhang (hkust), Dongdong Chen (Microsoft Cloud AI), and Jing Liao (City University of Hong Kong)

Learning sRGB-to-Raw-RGB De-Rendering With Content-Aware Metadata 17683
Seonghyeon Nam (York University), Abhijith Punnappurath (Samsung AI Center Toronto), Marcus A. Brubaker (York University), and Michael S. Brown (York University)

SNR-Aware Low-Light Image Enhancement 17693
Xiaogang Xu (The Chinese University of Hong Kong), Ruixing Wang (The Chinese University of Hong Kong), Chi-Wing Fu (The Chinese University of Hong Kong), and Jiaya Jia (Chinese University of Hong Kong)

AP-BSN: Self-Supervised Denoising for Real-World Images via Asymmetric PD and Blind-Spot Network 17704
Wooseok Lee (Seoul National University), Sanghyun Son (Seoul National University), and Kyoung Mu Lee (Seoul National University)

Synthetic Aperture Imaging With Events and Frames 17714
Wei Liao (WuHan University), Xiang Zhang (Wuhan University), Lei Yu (Wuhan University), Shijie Lin (The University of Hong Kong), Wen Yang (Wuhan University), and Ning Qiao (Chengdu SynSense Tech. Co. Ltd.)

Ev-TTA: Test-Time Adaptation for Event-Based Object Recognition 17724
Junho Kim (Seoul National University), Inwoo Hwang (Seoul National University), and Young Min Kim (Seoul National University)

Time Lens++: Event-Based Frame Interpolation With Parametric Non-Linear Flow and Multi-Scale Fusion 17734
Stepan Tulyakov (Huawei), Alfredo Bochicchio (Huawei), Daniel Gehrig (University of Zurich & ETH Zurich), Stamatios Georgoulis (Huawei), Yuanyou Li (Huawei), and Davide Scaramuzza (University of Zurich & ETH Zurich, Switzerland)

Unifying Motion Deblurring and Frame Interpolation With Events 17744
Xiang Zhang (Wuhan University) and Lei Yu (Wuhan University)

EvUnroll: Neuromorphic Events Based Rolling Shutter Image Correction 17754
Xinyu Zhou (Peking University), Peiqi Duan (Peking University), Yi Ma (Peking University), and Boxin Shi (Peking University)

Learning Adaptive Warping for Real-World Rolling Shutter Correction 17764
Mingdeng Cao (Tsinghua University), Zhihang Zhong (The University of Tokyo), Jiahao Wang (Tsinghua University), Yinqiang Zheng (The University of Tokyo), and Yujiu Yang (Tsinghua University)

Neural Global Shutter: Learn To Restore Video From a Rolling Shutter Camera With Global Reset Feature 17773
Zhixiang Wang (The University of Tokyo; National Institute of Informatics), Xiang Ji (The University of Tokyo), Jia-Bin Huang (Facebook), Shin'ichi Satoh (National Institute of Informatics), Xiao Zhou (Hefei Normal University), and Yinqiang Zheng (The University of Tokyo)

TimeReplayer: Unlocking the Potential of Event Cameras for Video Interpolation 17783
Weihua He (Tsinghua University), Kaichao You (Tsinghua Univ), Zhendong Qiao (Huawei, Multimedia Technology Dept), Xu Jia (Dalian University of Technology), Ziyang Zhang (HUAWEI TECHNOLOGIES CO.LTD), Wenhui Wang (Tsinghua University), Huchuan Lu (Dalian University of Technology), Yaoyuan Wang (Huawei Technologies Co., Ltd.), and Jianxing Liao (HUAWEI TECHNOLOGIES CO.LTD)

Optimizing Video Prediction via Video Frame Interpolation 17793
Yue Wu (HKUST), Qiang Wen (HKUST), and Qifeng Chen (HKUST)

Reference-Based Video Super-Resolution Using Multi-Camera Video Triplets 17803
Junyong Lee (POSTECH), Myeonghee Lee (POSTECH), Sunghyun Cho (POSTECH), and Seungyong Lee (POSTECH)

Memory-Augmented Non-Local Attention for Video Super-Resolution 17813
Jiyang Yu (JD AI Research), Jingen Liu (JD AI Research), Liefeng Bo (JD Finance), and Tao Mei (AI Research of JD.com)

Optical Flow Estimation for Spiking Camera 17823
Liwen Hu (Peking University), Rui Zhao (Peking University), Ziluo Ding (Peking University), Lei Ma (Peking University), Boxin Shi (Peking University), Ruiqin Xiong (Peking University), and Tiejun Huang (Peking University)

Compressive Single-Photon 3D Cameras 17833
Felipe Gutierrez-Barragan (University of Wisconsin-Madison), Atul Ingle (University of Wisconsin-Madison), Trevor Seets (University of Wisconsin-Madison), Mohit Gupta (University of Wisconsin-Madison, USA), and Andreas Velten (University of Wisconsin - Madison)

Single-Photon Structured Light 17844
Varun Sundar (University of Wisconsin Madison), Sizhuo Ma (University of Wisconsin-Madison), Aswin C. Sankaranarayanan (Carnegie Mellon University), and Mohit Gupta (University of Wisconsin-Madison, USA)

All-Photon Polarimetric Time-of-Flight Imaging 17855
Seung-Hwan Baek (POSTECH) and Felix Heide (Princeton University)

Holocurtains: Programming Light Curtains via Binary Holography 17865
Dorian Chan (Carnegie Mellon University), Srinivasa G. Narasimhan (Carnegie Mellon University, USA), and Matthew O'Toole (Carnegie Mellon University)

Poster 4.1: Vision & Language

Towards Implicit Text-Guided 3D Shape Generation 17875
Zhengzhe Liu (The Chinese University of Hong Kong), Yi Wang (Shanghai AI Laboratory), Xiaojuan Qi (The University of Hong Kong), and Chi-Wing Fu (The Chinese University of Hong Kong)

Towards Language-Free Training for Text-to-Image Generation 17886
Yufan Zhou (University at Buffalo), Ruiyi Zhang (Adobe Research), Changyou Chen (University at Buffalo), Chunyuan Li (Microsoft Research), Chris Tensmeyer (Adobe Research), Tong Yu (Adobe Research), Jiuxiang Gu (Adobe Research), Jinhui Xu (SUNY Buffalo), and Tong Sun (Adobe Research)

ZeroCap: Zero-Shot Image-to-Text Generation for Visual-Semantic Arithmetic 17897
Yoad Tewel (Tel-Aviv University), Yoav Shalev (Tel Aviv University), Idan Schwartz (Technion), and Lior Wolf (Tel Aviv University, Israel)

EMScore: Evaluating Video Captioning via Coarse-Grained and Fine-Grained Embedding Matching 17908
Yaya Shi (University of Science and Technology of China), Xu Yang (Southeast University), Haiyang Xu (Alibaba DAMO Academy), Chunfeng Yuan (NLPR), Bing Li (National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Sciences, ,), Weiming Hu (Institute of AutomationChinese Academy of Sciences), and Zheng-Jun Zha (University of Science and Technology of China)

Hierarchical Modular Network for Video Captioning 17918
Hanhua Ye (University of Chinese Academy of Sciences), Guorong Li (University of Chinese Academy of Sciences), Yuankai Qi (The University of Adelaide), Shuhui Wang (VIPL, ICT, Chinese academic of science), Qingming Huang (University of Chinese Academy of Sciences), and Ming-Hsuan Yang (University of California at Merced)

SwinBERT: End-to-End Transformers With Sparse Attention for Video Captioning 17928
Kevin Lin (Microsoft), Linjie Li (Microsoft), Chung-Ching Lin (Microsoft), Faisal Ahmed (Microsoft), Zhe Gan (Microsoft), Zicheng Liu (Microsoft), Yumao Lu (Microsoft), and Lijuan Wang (Microsoft)

End-to-End Generative Pretraining for Multimodal Video Captioning 17938
Paul Hongsuck Seo (Google), Arsha Nagrani (Google), Anurag Arnab (Google), and Cordelia Schmid (Google)

Beyond a Pre-Trained Object Detector: Cross-Modal Textual and Visual Context for Image Captioning 17948
Chia-Wen Kuo (Georgia Institute of Technology) and Zsolt Kira (Georgia Institute of Technology)

Scaling Up Vision-Language Pre-Training for Image Captioning 17959
Xiaowei Hu (Microsoft), Zhe Gan (Microsoft), Jianfeng Wang (Microsoft), Zhengyuan Yang (Microsoft), Zicheng Liu (Microsoft), Yumao Lu (Microsoft), and Lijuan Wang (Microsoft)

Comprehending and Ordering Semantics for Image Captioning 17969
Yehao Li (JD AI Research), Yingwei Pan (JD AI Research), Ting Yao (JD AI Research), and Tao Mei (AI Research of JD.com)

NOC-REK: Novel Object Captioning With Retrieved Vocabulary From External Knowledge 17979
Duc Minh Vo (The University of Tokyo), Hong Chen (The University of Tokyo), Akihiro Sugimoto (NII), and Hideki Nakayama (The University of Tokyo)

Injecting Semantic Concepts Into End-to-End Image Captioning 17988
Zhiyuan Fang (Arizona State University), Jianfeng Wang (Microsoft), Xiaowei Hu (Microsoft), Lin Liang (Microsoft), Zhe Gan (Microsoft), Lijuan Wang (Microsoft), Yezhou Yang (Arizona State University), and Zicheng Liu (Microsoft)

DIFNet: Boosting Visual Information Flow for Image Captioning 17999
Mingrui Wu (Xiamen University), Xuying Zhang (Xiamen University), Xiaoshuai Sun (Xiamen University), Yiyi Zhou (Xiamen University), Chao Chen (tencent), Jiaxin Gu (Tencent), Xing Sun (Tencent), and Rongrong Ji (Xiamen University, China)

VisualGPT: Data-Efficient Adaptation of Pretrained Language Models for Image Captioning 18009
Jun Chen (King Abdullah University of Science and Technology), Han Guo (CMU), Kai Yi (King Abdullah University of Science and Technology), Boyang Li (Nanyang Technological University), and Mohamed Elhoseiny (KAUST)

Show, Deconfound and Tell: Image Captioning With Causal Inference 18020
Bing Liu (China University of Mining and Technology), Dong Wang (China University of Mining and Technology), Xu Yang (Southeast University), Yong Zhou (China University of Mining and Technology), Rui Yao (China University of Mining and Technology), Zhiwen Shao (China University of Mining and Technology), and Jiaqi Zhao (China University of Mining and Technology)

EI-CLIP: Entity-Aware Interventional Contrastive Learning for E-Commerce Cross-Modal Retrieval 18030
Haoyu Ma (University of California, Irvine), Handong Zhao (Adobe Research), Zhe Lin (Adobe Research), Ajinkya Kale (Adobe), Zhangyang Wang (University of Texas at Austin), Tong Yu (Adobe Research), Jiuxiang Gu (Adobe Research), Sunav Choudhary (Adobe Research), and Xiaohui Xie (University of California, Irvine)

CLIPstyler: Image Style Transfer With a Single Text Condition 18041
Gihyun Kwon (KAIST) and Jong Chul Ye (Kim Jaechul Graduate School of AI, KAIST, Korea)

HairCLIP: Design Your Hair by Text and Reference Image 18051
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), Zhentao Tan (University of Science and Technology of China), Lu Yuan (Microsoft), Weiming Zhang (University of Science and Technology of China), and Nenghai Yu (University of Science and Technology of China)

DenseCLIP: Language-Guided Dense Prediction With Context-Aware Prompting 18061
Yongming Rao (Tsinghua University), Wenliang Zhao (Tsinghua University), Guangyi Chen (MBZUAI), Yansong Tang (Tsinghua University), Zheng Zhu (Tsinghua University), Guan Huang (Institute of Automation, Chinese Academy of Sciences), Jie Zhou (Tsinghua University), and Jiwen Lu (Tsinghua University)

On Guiding Visual Attention With Language Specification 18071
Suzanne Petryk (UC Berkeley), Lisa Dunlap (UC Berkeley), Keyan Nasseri (UC Berkeley), Joseph Gonzalez (UC Berkeley), Trevor Darrell (UC Berkeley), and Anna Rohrbach (UC Berkeley)

UTC: A Unified Transformer With Inter-Task Contrastive Learning for Visual Dialog 18082
Cheng Chen (Fudan university), Zhenshan Tan (Fudan University), Qingrong Cheng (Fudan University), Xin Jiang (Huawei Noah's Ark Lab), Qun Liu (Huawei Noah's Ark Lab), Yudong Zhu (Huawei Noah's Ark Lab), and Xiaodong Gu (Fudan University)

Text-to-Image Synthesis Based on Object-Guided Joint-Decoding Transformer 18092
Fuxiang Wu (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences), Liu Liu (The University of Sydney), Fusheng Hao (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences), Fengxiang He (JD.com), and Jun Cheng (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences)

LiT: Zero-Shot Transfer With Locked-Image Text Tuning 18102
Xiaohua Zhai (Google Brain), Xiao Wang (Google), Basil Mustafa (Google), Andreas Steiner (Google Brain), Daniel Keysers (Google), Alexander Kolesnikov (Google Brain), and Lucas Beyer (Google Brain)

GroupViT: Semantic Segmentation Emerges From Text Supervision 18113
Jiarui Xu (University of California San Diego), Shalini De Mello (NVIDIA Research), Sifei Liu (NVIDIA), Wonmin Byeon (NVIDIA Research), Thomas Breuel (NVIDIA), Jan Kautz (NVIDIA), and Xiaolong Wang (UCSD)

ReSTR: Convolution-Free Referring Image Segmentation Using Transformers 18124
Namyup Kim (POSTECH), Dongwon Kim (POSTECH), Cuiling Lan (Microsoft Research), Wenjun Zeng (EIT Institute for Advanced Study), and Suha Kwak (POSTECH)

LAVT: Language-Aware Vision Transformer for Referring Image Segmentation 18134
Zhao Yang (University of Oxford), Jiaqi Wang (The Chinese University of Hong Kong), Yansong Tang (Tsinghua University), Kai Chen (SenseTime Research), Hengshuang Zhao (University of Oxford), and Philip H.S. Torr (University of Oxford)

An Empirical Study of Training End-to-End Vision-and-Language Transformers 18145
Zi-Yi Dou (University of California, Los Angeles), Yichong Xu (Microsoft), Zhe Gan (Microsoft), Jianfeng Wang (Microsoft), Shuohang Wang (Microsoft), Lijuan Wang (Microsoft), Chenguang Zhu (Microsoft), Pengchuan Zhang (Microsoft Research AI), Lu Yuan (Microsoft), Nanyun Peng (UCLA), Zicheng Liu (Microsoft), and Michael Zeng (Microsoft)

Are Multimodal Transformers Robust to Missing Modality? 18156
Mengmeng Ma (University of Delaware), Jian Ren (Snap Inc.), Long Zhao (Rutgers University), Davide Testuggine (Facebook), and Xi Peng (University of Delaware)

Poster 4.1: Image & Video Synthesis and Generation

Text to Image Generation With Semantic-Spatial Aware GAN 18166
Wentong Liao (Leibniz University Hannover), Kai Hu (Leibniz University Hannover), Michael Ying Yang (University of Twente), and Bodo Rosenhahn (Leibniz University Hannover)

StyleT2I: Toward Compositional and High-Fidelity Text-to-Image Synthesis 18176
Zhiheng Li (University of Rochester), Martin Renqiang Min (NEC Labs America-Princeton), Kai Li (NEC LABORATORIES AMERICA, INC), and Chenliang Xu (University of Rochester)

Blended Diffusion for Text-Driven Editing of Natural Images 18187
Omri Avrahami (The Hebrew University of Jerusalem), Dani Lischinski (The Hebrew University of Jerusalem), and Ohad Fried (IDC Herzliya)

Make It Move: Controllable Image-to-Video Generation With Text Descriptions 18198
Yaosi Hu (Wuhan University), Chong Luo (MSRA), and Zhenzhong Chen (Wuhan University)

Predict, Prevent, and Evaluate: Disentangled Text-Driven Image Manipulation Empowered by Pre-Trained Vision-Language Model 18208
Zipeng Xu (University of Trento), Tianwei Lin (Baidu Inc), Hao Tang (ETH Zurich), Fu Li (Baidu), Dongliang He (Baidu), Nicu Sebe (University of Trento), Radu Timofte (ETH Zurich), Luc Van Gool (ETH Zurich), and Errui Ding (Baidu Inc.)

A Style-Aware Discriminator for Controllable Image Translation 18218
Kunhee Kim (Pohang University of Science and Technology), Sanghun Park (Pohang University of Science and Technology), Eunyeong Jeon (Pohang University of Science and Technology), Taehun Kim (Pohang University of Science and Technology), and Daijin Kim (Pohang University of Science and Technology)

Alleviating Semantics Distortion in Unsupervised Low-Level Image-to-Image Translation via Structure Consistency Constraint 18228
Jiaxian Guo (The University of Sydney), Jiachen Li (Shanghai Jiao Tong University), Huan Fu (Alibaba Group), Mingming Gong (University of Melbourne), Kun Zhang (Carnegie Mellon University), and Dacheng Tao (JD.com)

Exploring Patch-Wise Semantic Relation for Contrastive Learning in Image-to-Image Translation Tasks 18239
Chanyong Jung (KAIST), Gihyun Kwon (KAIST), and Jong Chul Ye (Kim Jaechul Graduate School of AI, KAIST, Korea)

FlexIT: Towards Flexible Semantic Image Translation 18249
Guillaume Couairon (Facebook AI Research), Asya Grechka (meero), Jakob Verbeek (Facebook), Holger Schwenk (Facebook AI Research), and Matthieu Cord (Sorbonne University)

Modulated Contrast for Versatile Image Synthesis 18259
Fangneng Zhan (Nanyang Technological University), Jiahui Zhang (Nanyang Technological University), Yingchen Yu (Nanyang Technological University), Rongliang Wu (Nanyang Technological University), and Shijian Lu (Nanyang Technological University)

QS-Attn: Query-Selected Attention for Contrastive Learning in I2I Translation 18270
Xueqi Hu (East China Normal University), Xinyue Zhou (East China Normal University), Qiusheng Huang (East China Normal University), Zhengyi Shi (East China Normal University), Li Sun (East China Normal University), and Qingli Li (East China Normal University)

Self-Supervised Dense Consistency Regularization for Image-to-Image Translation 18280
Minsu Ko (SAIT (Samsung Advanced Institute of Technology), ,), Eunju Cha (Samsung Advanced Institute of Technology), Sungjoo Suh (Samsung Advanced Institute of Technology), Huijin Lee (SAMSUNG), Jae-Joon Han (Samsung), Jinwoo Shin (KAIST), and Bohyung Han (Seoul National University)

Maximum Spatial Perturbation Consistency for Unpaired Image-to-Image Translation 18290
Yanwu Xu (University of Pittsburgh), Shaoan Xie (Carnegie Mellon University), Wenhao Wu (Baidu), Kun Zhang (Carnegie Mellon University), Mingming Gong (University of Melbourne), and Kayhan Batmanghelich (University of Pittsburgh)

InstaFormer: Instance-Aware Image-to-Image Translation With Transformer 18300
Soohyun Kim (Korea University), Jongbeom Baek (Korea University), Jihye Park (Korea University), Gyeongnyeon Kim (Korea University), and Seungryong Kim (Korea University)

Unsupervised Image-to-Image Translation With Generative Prior 18311
Shuai Yang (Nanyang Technological University), Liming Jiang (Nanyang Technological University), Ziwei Liu (Nanyang Technological University), and Chen Change Loy (Nanyang Technological University)

StylizedNeRF: Consistent 3D Scene Stylization As Stylized NeRF via 2D-3D Mutual Learning ... 18321
Yi-Hua Huang (Institute of Computing Technology, Chinese Academy of Sciences), Yue He (Institute of Computing Technology, Chinese Academy of Sciences), Yu-Jie Yuan (Institute of Computing Technology, Chinese Academy of Sciences), Yu-Kun Lai (Cardiff University), and Lin Gao (Institute of Computing Technology, Chinese Academy of Sciences)

NeRF-Editing: Geometry Editing of Neural Radiance Fields 18332
Yu-Jie Yuan (Institute of Computing Technology, Chinese Academy of Sciences), Yang-Tian Sun (Institute of Computing Technology, Chinese Academy of Sciences), Yu-Kun Lai (Cardiff University), Yuewen Ma (Alibaba Group), Rongfei Jia (Alibaba Group), and Lin Gao (Institute of Computing Technology, Chinese Academy of Sciences)

GeoNeRF: Generalizing NeRF With Geometry Priors 18344
Mohammad Mahdi Johari (Idiap Research Institute, École polytechnique fédérale de Lausanne (EPFL)), Yann Lepoittevin (ams OSRAM), and François Fleuret (University of Geneva)

Ray Priors Through Reprojection: Improving Neural Radiance Fields for Novel View Extrapolation 18355
Jian Zhang (Alibaba Group), Yuanqing Zhang (Zhejiang University), Huan Fu (Alibaba Group), Xiaowei Zhou (Zhejiang University), Bowen Cai (Alibaba Group), Jinchi Huang (Alibaba Group), Rongfei Jia (Alibaba Group), Binqiang Zhao (Alibaba), and Xing Tang (Alibaba Group)

AR-NeRF: Unsupervised Learning of Depth and Defocus Effects From Natural Images With Aperture Rendering Neural Radiance Fields 18366
Takuhiro Kaneko (NTT Corporation)

HDR-NeRF: High Dynamic Range Neural Radiance Fields 18377
Xin Huang (Northwestern Polytechnical University), Qi Zhang (Tencent AI Lab), Ying Feng (Tencent AI Lab), Hongdong Li (Australian National University, Australia), Xuan Wang (Tencent AI Lab), and Qing Wang (Northwestern Polytechnical University)

NeRFReN: Neural Radiance Fields With Reflections 18388
Yuan-Chen Guo (Tsinghua University), Di Kang (Tencent), Linchao Bao (Tencent AI Lab), Yu He (BIMSA), and Song-Hai Zhang (Tsinghua University)

Neural Point Light Fields 18398
Julian Ost (Algolux), Issam Laradji (ServiceNow), Alejandro Newell (Princeton University), Yuval Bahat (Princeton University), and Felix Heide (Princeton University)

3D-Aware Image Synthesis via Learning Structural and Textural Representations 18409
Yinghao Xu (Chinese University of Hong Kong), Sida Peng (Zhejiang University), Ceyuan Yang (Chinese University of Hong Kong), Yujun Shen (Dept. of IE, CUHK), and Bolei Zhou (UCLA)

GIRAFFE HD: A High-Resolution 3D-Aware Generative Model 18419
Yang Xue (University of California, Davis), Yuheng Li (UW Madison), Krishna Kumar Singh (Adobe Research), and Yong Jae Lee (University of Wisconsin-Madison)

Multi-View Consistent Generative Adversarial Networks for 3D-Aware Image Synthesis 18429
Xuanmeng Zhang (University of Technology Sydney), Zhedong Zheng (National University of Singapore), Daiheng Gao (Alibaba), Bang Zhang (Damo Academy Alibaba), Pan Pan (Alibaba Group), and Yi Yang (UTS)

Bi-Level Doubly Variational Learning for Energy-Based Latent Variable Models 18439
Ge Kan (Beihang University), Jinhu Lü (Beihang University, Beijing, China), Tian Wang (Beihang University), Baochang Zhang (Beihang University), Aichun Zhu (Nanjing Tech University), Lei Huang (Beihang University), Guodong Guo (Baidu), and Hichem Snoussi (University of Troyes, France)

High-Resolution Image Harmonization via Collaborative Dual Transformations 18449
Wenyan Cong (Shanghai Jiao Tong University), Xinhao Tao (Harbin Institute of Technology), Li Niu (Shanghai Jiao Tong University), Jing Liang (Shanghai Jiao Tong University), Xuesong Gao (Tianjin University), Qihao Sun (Hisense Group Holdings Co., Ltd.), and Liqing Zhang (Shanghai Jiao Tong University)

Brain-Supervised Image Editing 18459
Keith M. Davis (University of Helsinki), Carlos de la Torre-Ortiz (University of Helsinki), and Tuukka Ruotsalo (University of Copenhagen and University of Helsinki)

Poster 4.1: Vision & Graphics

De-Rendering 3D Objects in the Wild 18469
Felix Wimbauer (University of Oxford), Shangzhe Wu (University of Oxford), and Christian Rupprecht (University of Oxford)

Neural Fields As Learnable Kernels for 3D Reconstruction 18479
Francis Williams (NVIDIA), Zan Gojcic (NVIDIA), Sameh Khamis (Nvidia), Denis Zorin (New York University), Joan Bruna (Courant Institute of Mathematical Sciences, NYU, USA), Sanja Fidler (University of Toronto, NVIDIA), and Or Litany (NVIDIA)

HyperStyle: StyleGAN Inversion With HyperNetworks for Real Image Editing 18490
Yuval Alaluf (Tel Aviv University), Omer Tov (Tel Aviv University), Ron Mokady (Tel Aviv University), Rinon Gal (Tel Aviv University), and Amit Bermano (Tel Aviv University)

3PSDF: Three-Pole Signed Distance Function for Learning Surfaces With Arbitrary Topologies... 18501
Weikai Chen (Tencent America), Cheng Lin (Tencent), Weiyang Li (Tencent), and Bo Yang (Tencent America)

Pop-Out Motion: 3D-Aware Image Deformation via Learning the Shape Laplacian 18511
Jihyun Lee (KAIST), Minhyuk Sung (KAIST), Hyunjin Kim (KAIST), and Tae-Kyun Kim (Imperial College London)

Deep Image-Based Illumination Harmonization 18521
Zhongyun Bao (Wuhan university), Chengjiang Long (Meta Reality Labs), Gang Fu (Wuhan University), Daquan Liu (Wuhan University), Yuanzhen Li (Wuhan university), Jiaming Wu (Wuhan University), and Chunxia Xiao (Wuhan University)

Glass: Geometric Latent Augmentation for Shape Spaces 18531
Sanjeev Muralikrishnan (UCL), Siddhartha Chaudhuri (Adobe Research), Noam Aigerman (Adobe), Vladimir G. Kim (Adobe), Matthew Fisher (Adobe Research), and Niloy J. Mitra (university college london)

PhotoScene: Photorealistic Material and Lighting Transfer for Indoor Scenes 18541
Yu-Ying Yeh (UC San Diego), Zhengqin Li (Meta), Yannick Hold-Geoffroy (Adobe Research), Rui Zhu (University of California San Diego), Zexiang Xu (Adobe Research), Miloš Hašan (Adobe Research), Kalyan Sunkavalli (Adobe Research), and Manmohan Chandraker (UC San Diego)

Neural Template: Topology-Aware Reconstruction and Disentangled Generation of 3D Meshes . 18551
Ka-Hei Hui (The Chinese University of Hong Kong), Ruihui Li (Hunan University), Jingyu Hu (The Chinese University of Hong Kong), and Chi-Wing Fu (The Chinese University of Hong Kong)

Neural Mesh Simplification 18562
Rolandos Alexandros Potamias (Imperial College London), Stylianos Ploumpis (Huawei Technologies Co. Ltd), and Stefanos Zafeiriou (Imperial College London)

SkinningNet: Two-Stream Graph Convolutional Neural Network for Skinning Prediction of Synthetic Characters 18572
Albert Mosella-Montoro (Universitat Politècnica de Catalunya) and Javier Ruiz-Hidalgo (Universitat Politècnica de Catalunya)

CLIP-Forge: Towards Zero-Shot Text-To-Shape Generation 18582
Aditya Sanghi (Autodesk; Autodesk AI Lab), Hang Chu (Autodesk AI Lab), Joseph G. Lambourne (Autodesk AI Lab), Ye Wang (Autodesk Research), Chin-Yi Cheng (Autodesk Research), Marco Fumero (La Sapienza, University of Rome), and Kamal Rahimi Malekshan (Autodesk)

UNIST: Unpaired Neural Implicit Shape Translation Network 18593
Qimin Chen (Simon Fraser University), Johannes Merz (Simon Fraser University), Aditya Sanghi (Autodesk; Autodesk AI Lab), Hooman Shayani (Autodesk AI Lab), Ali Mahdavi-Amiri (Simon Fraser University), and Hao Zhang (Simon Fraser University)

CoNeRF: Controllable Neural Radiance Fields 18602
Kacper Kania (Warsaw University of Technology), Kwang Moo Yi (University of British Columbia), Marek Kowalski (Microsoft), Tomasz Trzciński (Tooploox, Warsaw University of Technology, Jagiellonian University of Cracow), and Andrea Tagliasacchi (Google Brain and University of Toronto)

Neural Points: Point Cloud Representation With Neural Fields for Arbitrary Upsampling 18612
Wanquan Feng (University of Science and Technology of China), Jin Li (Guilin University Of Electronic Technology), Hongrui Cai (University of Science and Technology of China), Xiaonan Luo (Guilin University of Electronic Technology), and Juyong Zhang (University of Science and Technology of China)

Modeling Indirect Illumination for Inverse Rendering 18622
Yuanqing Zhang (Zhejiang University), Jiaming Sun (Zhejiang University), Xingyi He (Zhejiang University), Huan Fu (Alibaba Group), Rongfei Jia (Alibaba Group), and Xiaowei Zhou (Zhejiang University)

Neural Head Avatars From Monocular RGB Videos 18632
Philip-William Grassal (Heidelberg University), Malte Prinzler (Heidelberg University), Titus Leistner (Universität Heidelberg), Carsten Rother (University of Heidelberg), Matthias Nießner (Technical University of Munich), and Justus Thies (Max Planck Institute for Intelligent Systems)

DeepCurrents: Learning Implicit Representations of Shapes With Boundaries 18644
David Palmer (MIT), Dmitriy Smirnov (MIT), Stephanie Wang (UC San Diego), Albert Chern (University of California, San Diego), and Justin Solomon (MIT)

Oral 4.2.1: Biometrics, Face & Gestures, and Medical Image Analysis

Escaping Data Scarcity for High-Resolution Heterogeneous Face Hallucination 18655
Yiqun Mei (University of Illinois), Pengfei Guo (Johns Hopkins University), and Vishal M. Patel (Johns Hopkins University)

AnyFace: Free-Style Text-To-Face Synthesis and Manipulation 18666
Jianxin Sun (Center for Research on Intelligent Perception and Computing, NLPR, CASIA), Qiyao Deng (Institute of Automation, Chinese Academy of Sciences (CASIA); University of Chinese Academy of Sciences, ,), Qi Li (CASIA), Muyi Sun (CRIPAC, Institute of Automation, Chinese Academy of Sciences), Min Ren (Center for Research on Intelligent Perception and Computing (CRIPAC), Institute of Automation, Chinese Academy of Sciences (CASIA), University of Chinese Academy of Sciences (UCAS)), and Zhenan Sun (Chinese of Academy of Sciences)

General Facial Representation Learning in a Visual-Linguistic Manner 18676
Yinglin Zheng (School of Informatics, Xiamen University), Hao Yang (Microsoft Research Asia), Ting Zhang (Microsoft Research Asia), Jianmin Bao (Microsoft Research Asia), Dongdong Chen (Microsoft Cloud AI), Yangyu Huang (Microsoft Research Asia), Lu Yuan (Microsoft), Dong Chen (Microsoft Research Asia), Ming Zeng (School of Informatics, Xiamen University), and Fang Wen (Microsoft Research Asia)

Self-Supervised Learning of Adversarial Example: Towards Good Generalizations for Deepfake Detection 18689
Liang Chen (University of Adelaide), Yong Zhang (Tencent AI Lab), Yibing Song (Tencent), Lingqiao Liu (University of Adelaide), and Jue Wang (Tencent AI Lab)

Detecting Deepfakes With Self-Blended Images 18699
Kaede Shiohara (The University of Tokyo) and Toshihiko Yamasaki (The University of Tokyo)

3D Shape Variational Autoencoder Latent Disentanglement via Mini-Batch Feature Swapping for Bodies and Faces 18709
Simone Foti (University College London), Bongjin Koo (University College London), Danail Stoyanov (UCL), and Matthew J. Clarkson (University College London)

Evaluation-Oriented Knowledge Distillation for Deep Face Recognition 18719
Yuge Huang (Tencent YouTu), Jiaxiang Wu (Tencent), Xingkun Xu (Tencent), and Shouhong Ding (Tencent)

AdaFace: Quality Adaptive Margin for Face Recognition 18729
Minchul Kim (Michigan State University), Anil K. Jain (Michigan State University), and Xiaoming Liu (Michigan State University)

Moving Window Regression: A Novel Approach to Ordinal Regression 18739
Nyeong-Ho Shin (Korea University), Seon-Ho Lee (MCL, Korea University), and Chang-Su Kim (Korea university)

FaceFormer: Speech-Driven 3D Facial Animation With Transformers 18749
Yingruo Fan (The University of Hong Kong), Zhaojiang Lin (Hong Kong University of Science and Technology), Jun Saito (Adobe), Wenping Wang (The University of Hong Kong), and Taku Komura (The University of Hong Kong)

Neural Emotion Director: Speech-Preserving Semantic Control of Facial Expressions in "In-the-Wild" Videos 18759
Foivos Paraperas Papantoniou (National Technical University of Athens), Panagiotis P. Filntisis (National Technical University of Athens), Petros Maragos (National Technical University of Athens), and Anastasios Roussos (Institute of Computer Science, Foundation for Research and Technology Hellas)

Deep Decomposition for Stochastic Normal-Abnormal Transport 18769
Peirong Liu (UNC Chapel Hill), Yueh Lee (UNC), Stephen Aylward (Kitware), and Marc Niethammer (UNC)

DTFD-MIL: Double-Tier Feature Distillation Multiple Instance Learning for Histopathology Whole Slide Image Classification 18780
Hongrun Zhang (University of Liverpool), Yanda Meng (University of Liverpool), Yitian Zhao (Cixi Institute of Biomedical Engineering, Ningbo Institute of Industrial Technology, Chinese Academy of Sciences), Yihong Qiao (intellicloud), Xiaoyun Yang (Remark Holdings), Sarah E. Coupland (University of Liverpool), and Yalin Zheng (University of Liverpool)

Node-Aligned Graph Convolutional Network for Whole-Slide Image Representation and Classification 18791
Yonghang Guan (ShanghaiTech University), Jun Zhang (Tencent), Kuan Tian (Tencent), Sen Yang (Sichuan University), Pei Dong (Tencent), Jinxi Xiang (tencent), Wei Yang (Tencent AI Lab), Junzhou Huang (University of Texas at Arlington), Yuyao Zhang (ShanghaiTech University), and Xiao Han (Tencent)

Temporal Context Matters: Enhancing Single Image Prediction With Disease Progression Representations 18802
Aishik Konwer (Stony Brook University), Xuan Xu (Stony Brook University), Joseph Bae (Stony Brook University), Chao Chen (Stony Brook University), and Prateek Prasanna (Stony Brook University)

Oral 4.2.2: Scene & Shape Analysis and Understanding

VRDFormer: End-to-End Video Visual Relation Detection With Transformers 18814
Sipeng Zheng (Renmin University of China), Shizhe Chen (INRIA), and Qin Jin (Renmin University of China)

Video K-Net: A Simple, Strong, and Unified Baseline for Video Segmentation 18825
Xiangtai Li (Peking University), Wenwei Zhang (NTU), Jiangmiao Pang (CUHK), Kai Chen (SenseTime Research), Guangliang Cheng (Sensetime Group Limited), Yunhai Tong (Peking University), and Chen Change Loy (Nanyang Technological University)

Visual Acoustic Matching 18836
Changan Chen (University of Texas at Austin), Ruohan Gao (Stanford University), Paul Calamia (Facebook Reality Labs Research), and Kristen Grauman (Facebook AI Research & UT Austin)

The Devil Is in the Labels: Noisy Label Correction for Robust Scene Graph Generation 18847
Lin Li (Zhejiang University), Long Chen (Columbia University), Yifeng Huang (Zhejiang University), Zhimeng Zhang (Zhejiang University), Songyang Zhang (University of Rochester), and Jun Xiao (Zhejiang University)

Learning Multiple Dense Prediction Tasks From Partially Annotated Data 18857
Wei-Hong Li (University of Edinburgh), Xialei Liu (Nankai University), and Hakan Bilen (University of Edinburgh)

PONI: Potential Functions for ObjectGoal Navigation With Interaction-Free Learning 18868
Santhosh Kumar Ramakrishnan (University of Texas at Austin), Devendra Singh Chaplot (Facebook AI Research), Ziad Al-Halah (UT Austin), Jitendra Malik (University of California at Berkeley), and Kristen Grauman (Facebook AI Research & UT Austin)

Continual Stereo Matching of Continuous Driving Scenes With Growing Architecture 18879
Chenghao Zhang (National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, China), Kun Tian (Institute of Automation, Chinese Academy of Sciences), Bin Fan (University of Science and Technology Beijing), Gaofeng Meng (Chinese Academy of Sciences), Zhaoxiang Zhang (Chinese Academy of Sciences, China), and Chunhong Pan (Institute of Automation, Chinese Academy of Sciences)

FIFO: Learning Fog-Invariant Features for Foggy Scene Segmentation 18889
Sohyun Lee (POSTECH), Taeyoung Son (POSTECH), and Suha Kwak (POSTECH)

Both Style and Fog Matter: Cumulative Domain Adaptation for Semantic Foggy Scene Understanding 18900
Xianzheng Ma (Wuhan university), Zhixiang Wang (The University of Tokyo), Yacheng Zhan (Wuhan University), Yinqiang Zheng (The University of Tokyo), Zheng Wang (Wuhan University), Dengxin Dai (MPI for Informatics), and Chia-Wen Lin (National Tsing Hua University)

Equivariant Point Cloud Analysis via Learning Orientations for Message Passing 18910
Shitong Luo (HeliXon Research), Jiahan Li (Peking University), Jiaqi Guan (UIUC), Yufeng Su (UIUC), Chaoran Cheng (University of Illinois at Urbana-Champaign), Jian Peng (University of Illinois at Urbana-Champaign), and Jianzhu Ma (Institute for Artificial Intelligence, Peking University)

Surface Representation for Point Clouds 18920
Haoxi Ran (Northeastern University), Jun Liu (Tencent), and Chengjie Wang (Tencent)

Not All Points Are Equal: Learning Highly Efficient Point-Based Detectors for 3D LiDAR Point Clouds 18931
Yifan Zhang (National University of Defense Technology), Qingyong Hu (University of Oxford), Guoquan Xu (National University of Defense Technology), Yanxin Ma (National University of Defense Technology), Jianwei Wan (NUDT), and Yulan Guo (Sun Yat-sen University)

3D Common Corruptions and Data Augmentation 18941
Oğuzhan Fatih Kar (EPFL), Teresa Yeo (EPFL), Andrei Atanov (EPFL), and Amir Zamir (Swiss Federal Institute of Technology (EPFL), ,)

INS-Conv: Incremental Sparse Convolution for Online 3D Segmentation 18953
Leyao Liu (Tsinghua University), Tian Zheng (Tsinghua University), Yun-Jou Lin (Oppo), Kai Ni (HoloMatic TechnologyBeijingCo., Ltd.), and Lu Fang (Tsinghua University)

How Much Does Input Data Type Impact Final Face Model Accuracy? 18963
Jiahao Luo (UCSC), Fahim Hasan Khan (University of California, Santa Cruz), Issei Mori (University of California Santa Cruz), Akila de Silva (University of California, Santa Cruz), Eric Sandoval Ruezga (AVIS), Minghao Liu (UCSC), Alex Pang (ucsc), and James Davis (UC Santa Cruz)

Oral 4.2.3: Datasets & Evaluation, Action & Event Recognition, and Visual Question Answering

Ego4D: Around the World in 3,000 Hours of Egocentric Video 18973
Kristen Grauman (Facebook AI Research & UT Austin), Andrew Westbury (Facebook AI Research), Eugene Byrne (Meta), Zachary Chavis (University of Minnesota), Antonino Furnari (University of Catania), Rohit Girdhar (Facebook AI Research), Jackson Hamburger (Sensorum Health), Hao Jiang (Facebook), Miao Liu (Georgia Institute of Technology), Xingyu Liu (Carnegie Mellon University), Miguel Martin (Facebook), Tushar Nagarajan (UT Austin & Facebook AI Research), Ilija Radosavovic (UC Berkeley), Santhosh Kumar Ramakrishnan (University of Texas at Austin), Fiona Ryan (Georgia Institute of Technology), Jayant Sharma (University of Minnesota), Michael Wray (University of Bristol), Mengmeng Xu (KAUST), Eric Zhongcong Xu (National University of Singapore), Chen Zhao (KAUST), Siddhant Bansal (IIIT, Hyderabad), Dhruv Batra (Georgia Tech & Facebook AI Research), Vincent Cartillier (Georgia Tech), Sean Crane (Carnegie Mellon University), Tien Do (University of Minnesota), Morrie Doulaty (Facebook), Akshay Erapalli (Facebook (Meta), ,), Christoph Feichtenhofer (Facebook AI Research), Adriano Fragomeni (University of Bristol), Qichen Fu (Carnegie Mellon University), Abrham Gebreselasie (Carnegie Mellon University), Cristina González (Universidad de los Andes), James Hillis (Facebook Reality Labs), Xuhua Huang (Carnegie Mellon University), Yifei Huang (The University of Tokyo), Wenqi Jia (Georgia Institute of Technology), Weslie Khoo (Indiana University), Jáchym Kolář (Meta), Satwik Kottur (Facebook AI), Anurag Kumar (Facebook Research), Federico Landini (Brno University of Technology), Chao Li (Facebook Reality Labs), Yanghao Li (Facebook AI Research), Zhenqiang Li (The University of Tokyo), Karttikeya Mangalam (UC Berkeley), Raghava Modhugu (International Institute of Information Techonology), Jonathan Munro (University of Bristol), Tullie Murrell (Facebook AI Research), Takumi Nishiyasu (The University of Tokyo), Will Price (University of Bristol), Paola Ruiz (Universidad de los Andes), Merey Ramazanova (KAUST), Leda Sari (Facebook), Kiran Somasundaram (Facebook Reality Labs), Audrey Southerland (Georgia Institute of Technology), Yusuke Sugano (The University of Tokyo), Ruijie Tao (National University of Singapore), Minh Vo (Facebook Reality Labs), Yuchen Wang (Indiana University), Xindi Wu (Carnegie Mellon University), Takuma Yagi (The University of Tokyo), Ziwei Zhao (Indiana University), Yunyi Zhu (National University of Singapore), Pablo Arbeláez (Universidad de los Andes), David Crandall (Indiana University), Dima Damen (University of Bristol), Giovanni Maria Farinella (University of Catania, Italy), Christian Fuegen (Facebook), Bernard Ghanem (KAUST), Vamsi Krishna Ithapu (Facebook Reality Labs), C. V. Jawahar (IIIT-Hyderabad), Hanbyul Joo (Facebook AI Research), Kris Kitani (Carnegie Mellon University), Haizhou Li (The Chinese University of Hong Kong (Shenzhen), ,), Richard Newcombe (Facebook), Aude Oliva (MIT), Hyun Soo Park (The University of Minnesota), James M. Rehg (Georgia Institute of Technology), Yoichi Sato (University of Tokyo), Jianbo Shi (University of Pennsylvania), Mike Zheng Shou (National University of Singapore), Antonio Torralba (MIT), Lorenzo Torresani (Facebook AI Research), Mingfei Yan (Meta), and Jitendra Malik (University of California at Berkeley)

TransRAC: Encoding Multi-Scale Temporal Correlation With Transformers for Repetitive Action Counting 18991
Huazhang Hu (Shanghaitech University), Sixun Dong (Shanghaitech Univercity), Yiqun Zhao (ShanghaiTech University), Dongze Lian (ShanghaiTech University), Zhengxin Li (ShanghaiTech University), and Shenghua Gao (Shanghaitech University)

Animal Kingdom: A Large and Diverse Dataset for Animal Behavior Understanding 19001
Xun Long Ng (Singapore University of Technology and Design), Kian Eng Ong (Singapore University of Technology and Design), Qichen Zheng (Singapore University of Technology and Design), Yun Ni (Singapore University of Technology and Design), Si Yong Yeo (SUTD), and Jun Liu (Singapore University of Technology and Design)

vCLIMB: A Novel Video Class Incremental Learning Benchmark 19013
Andrés Villa (Pontificia Universidad Católica de Chile), Kumail Alhamoud (KAUST), Victor Escorcia (Samsung AI Center), Fabian Caba (Adobe Research), Juan León Alcázar (KAUST), and Bernard Ghanem (KAUST)

Opening Up Open World Tracking 19023
Yang Liu (Technical University of Munich), Idil Esen Zulfikar (RWTH Aachen University), Jonathon Luiten (RWTH Aachen University), Achal Dave (Amazon), Deva Ramanan (Carnegie Mellon University), Bastian Leibe (RWTH Aachen University), Aljoša Ošep (TUM Munich), and Laura Leal-Taixé (TUM)

Bongard-HOI: Benchmarking Few-Shot Visual Reasoning for Human-Object Interactions 19034
Huaizu Jiang (Northeastern University), Xiaojian Ma (University of California, Los Angeles), Weili Nie (NVIDIA), Zhiding Yu (NVIDIA), Yuke Zhu (University of Texas - Austin), and Anima Anandkumar (Caltech)

CNN Filter DB: An Empirical Investigation of Trained Convolutional Filters 19044
Paul Gavrikov (Offenburg University) and Janis Keuper (Offenburg University)

Failure Modes of Domain Generalization Algorithms 19055
Tigran Galstyan (YerevaNN), Hrayr Harutyunyan (USC Information Sciences Institute), Hrant Khachatrian (YerevaNN), Greg Ver Steeg (USC Information Sciences Institute), and Aram Galstyan (USC Information Sciences Institute)

A Comprehensive Study of Image Classification Model Sensitivity to Foregrounds, Backgrounds, and Visual Attributes 19065
Mazda Moayeri (University of Maryland), Phillip Pope, Yogesh Balaji, and Soheil Feizi (University of Maryland)

Grounding Answers for Visual Questions Asked by Visually Impaired People 19076
Chongyan Chen (University of Texas at Austin), Samreen Anjum (University of Colorado Boulder), and Danna Gurari (University of Colorado Boulder)

Learning To Answer Questions in Dynamic Audio-Visual Scenarios 19086
Guangyao Li (Renmin University of China), Yake Wei (Renmin University of China), Yapeng Tian (University of Rochester), Chenliang Xu (University of Rochester), Ji-Rong Wen (Renmin University of China), and Di Hu (Renmin University of China)

Episodic Memory Question Answering 19097
Samyak Datta (Georgia Tech), Sameer Dharur (Georgia Tech), Vincent Cartillier (Georgia Tech), Ruta Desai (Facebook Reality Labs), Mukul Khanna (IIIT-Hyderabad), Dhruv Batra (Georgia Tech & Facebook AI Research), and Devi Parikh (Georgia Tech & Facebook AI Research)

ScanQA: 3D Question Answering for Spatial Scene Understanding 19107
Daichi Azuma (Kyoto University), Taiki Miyanishi (Advanced Telecommunications Research Institute International (ATR)), Shuhei Kurita (RIKEN AIP), and Motoaki Kawanabe (Advanced Telecommunications Research Institute International (ATR))

Learning Part Segmentation Through Unsupervised Domain Adaptation From Synthetic Vehicles 19118
Qing Liu (Johns Hopkins University), Adam Kortylewski (Max Planck Institute for Informatics), Zhishuai Zhang (Waymo), Zizhang Li (Zhejiang University), Mengqi Guo (National University of Singapore), Qihao Liu (Johns Hopkins University), Xiaoding Yuan (Johns Hopkins University), Jiteng Mu (University of California, San Diego), Weichao Qiu (Johns Hopkins University), and Alan Yuille (Johns Hopkins University)

BTS: A Bi-Lingual Benchmark for Text Segmentation in the Wild 19130
Xixi Xu (Fudan University), Zhongang Qi (Tencent), Jianqi Ma (the Hong Kong Polytechnic University), Honglun Zhang (Applied Research Center, Tencent PCG), Ying Shan (Tencent), and Xiaohu Qie (Tencent)

Poster 4.2: Representation Learning

Unified Contrastive Learning in Image-Text-Label Space 19141
Jianwei Yang (Microsoft Research), Chunyuan Li (Microsoft Research), Pengchuan Zhang (Microsoft Research AI), Bin Xiao (Microsoft), Ce Liu (Microsoft), Lu Yuan (Microsoft), and Jianfeng Gao (Microsoft Research)

AlignMixup: Improving Representations by Interpolating Aligned Features 19152
Shashanka Venkataramanan (INRIA), Ewa Kijak (IRISA), Laurent Amsaleg (CNRS-IRISA), and Yannis Avrithis (Athena RC)

On the Road to Online Adaptation for Semantic Image Segmentation 19162
Riccardo Volpi (Naver Labs), Pau De Jorge (University of Oxford), Diane Larlus (Naver Labs Europe), and Gabriela Csurka (Naver Labs Europe)

ADAS: A Direct Adaptation Strategy for Multi-Target Domain Adaptive Semantic Segmentation 19174
Seunghun Lee (DGIST), Wonhyeok Choi (DGIST), Changjae Kim (DGIST), Minwoo Choi (DGIST), and Sunghoon Im (DGIST)

Kernelized Few-Shot Object Detection With Efficient Integral Aggregation 19185
Shan Zhang (Australian National University), Lei Wang (University of Wollongong, Australia), Naila Murray (Naver Labs), and Piotr Koniusz (ANU College of Engineering and Computer Science)

Neural Mean Discrepancy for Efficient Out-of-Distribution Detection 19195
Xin Dong (Harvard Univeristy), Junfeng Guo (The University of Texas at Dallas), Ang Li (Google DeepMind), Wei-Te Ting (Harvard), Cong Liu (University of Texas at Dallas), and H.T. Kung (Harvard University)

A Structured Dictionary Perspective on Implicit Neural Representations 19206
Gizem Yüce (ETH Zürich), Guillermo Ortiz-Jiménez (EPFL), Beril Besbinar (EPFL), and Pascal Frossard (EPFL)

LARGE: Latent-Based Regression Through GAN Semantics 19217
Yotam Nitzan (Tel Aviv University), Rinon Gal (Tel Aviv University), Ofir Brenner (Tel Aviv University), and Daniel Cohen-Or (Tel Aviv University)

Rethinking Controllable Variational Autoencoders 19228
Huajie Shao (College of William and Mary), Yifei Yang (Zhejiang University), Haohong Lin (Carnegie Mellon University), Longzhong Lin (Zhejiang University), Yizhuo Chen (Zhejiang University), Qinmin Yang (Zhejiang University), and Han Zhao (University of Illinois at Urbana-Champaign)

Learning Canonical F-Correlation Projection for Compact Multiview Representation 19238
Yun-Hao Yuan (Yangzhou University), Jin Li (Yangzhou University), Yun Li (Yangzhou University), Jipeng Qiang (Yangzhou University), Yi Zhu (Yangzhou University), Xiaobo Shen (Nanjing University of Science and Technology), and Jianping Gou (Jiangsu University)

Cross-Architecture Self-Supervised Video Representation Learning 19248
Sheng Guo (Ant Group), Zihua Xiong (antgroup), Yujie Zhong (University of Oxford), Limin Wang (Nanjing University), Xiaobo Guo (antgroup), Bing Han (antgroup), and Weilin Huang (Malong LLC)

Improving Video Model Transfer With Dynamic Representation Learning 19258
Yi Li (UC San Diego) and Nuno Vasconcelos (UC San Diego)

Self-Supervised Image Representation Learning With Geometric Set Consistency 19270
Nenglun Chen (The University of Hong Kong; Microsoft Research Asia), Lei Chu (MSRA), Hao Pan (Microsoft Research), Yan Lu (Microsoft Research Asia), and Wenping Wang (The University of Hong Kong)

HLRTF: Hierarchical Low-Rank Tensor Factorization for Inverse Problems in Multi-Dimensional Imaging 19281
Yisi Luo (School of Mathematical Sciences, University of Electronic Science and Technology of China), Xi-Le Zhao (School of Mathematical Sciences, University of Electronic Science and Technology of China), Deyu Meng (Xi'an Jiaotong University), and Tai-Xiang Jiang (School of Economic Information Engineering, Southwestern University of Finance and Economics)

Point-BERT: Pre-Training 3D Point Cloud Transformers With Masked Point Modeling 19291
Xumin Yu (Tsinghua University), Lulu Tang (Tsingua University), Yongming Rao (Tsinghua University), Tiejun Huang (Peking University), Jie Zhou (Tsinghua University), and Jiwen Lu (Tsinghua University)

DiGS: Divergence Guided Shape Implicit Neural Representation for Unoriented Point Clouds ... 19301
Yizhak Ben-Shabat (Technion), Chamin Hewa Koneputugodage (Australian National University), and Stephen Gould (Australian National University, Australia)

Neural Convolutional Surfaces 19311
Luca Morreale (University College London), Noam Aigerman (Adobe), Paul Guerrero (Adobe), Vladimir G. Kim (Adobe), and Niloy J. Mitra (University College London)

Representing 3D Shapes With Probabilistic Directed Distance Fields 19321
Tristan Aumentado-Armstrong (Samsung Toronto AIC), Stavros Tsogkas (Samsung AI Centre Toronto), Sven Dickinson (Samsung AI Research Center Toronto), and Allan D. Jepson (Samsung Toronto AIC)

H4D: Human 4D Modeling by Learning Neural Compositional Representation 19333
Boyan Jiang (Fudan University), Yinda Zhang (Google), Xingkui Wei (Fudan University), Xiangyang Xue (Fudan University), and Yanwei Fu (Fudan University)

Learning Memory-Augmented Unidirectional Metrics for Cross-Modality Person Re-Identification 19344
Jialun Liu (Jilin University), Yifan Sun (Baidu Research), Feng Zhu (University of Technology Sydney), Hongbin Pei (Xian Jiao Tong University), Yi Yang (Zhejiang University), and Wenhui Li (Jilin University)

Contrastive Regression for Domain Adaptation on Gaze Estimation 19354
Yaoming Wang (Shanghai Jiao Tong University), Yangzhou Jiang (Shanghai Jiao Tong University), Jin Li (Shanghai Jiao Tong University), Bingbing Ni (Shanghai Jiao Tong University), Wenrui Dai (Shanghai Jiao Tong University), Chenglin Li (Shanghai Jiao Tong University), Hongkai Xiong (Shanghai Jiao Tong University), and Teng Li (Anhui University)

Forward Compatible Training for Large-Scale Embedding Retrieval Systems 19364
Vivek Ramanujan (University of Washington), Pavan Kumar Anasosalu Vasu (Apple Inc.), Ali Farhadi (University of Washington, Allen Institue for AI, Apple), Oncel Tuzel (Apple), and Hadi Pouransari (Apple)

Improving Subgraph Recognition With Variational Graph Information Bottleneck 19374
Junchi Yu (Institute of Automation, Chinese Academy of Sciences), Jie Cao (Institute of Automation, Chinese Academy of Sciences), and Ran He (Institute of Automation, Chinese Academy of Sciences)

Learning Soft Estimator of Keypoint Scale and Orientation With Probabilistic Covariant Loss 19384
Pei Yan (Huazhong University of Science and Technology), Yihua Tan (Huazhong University of Science and Technology), Shengzhou Xiong (Huazhong University of Science and Technology), Yuan Tai (Huazhong University of Science and Technology), and Yansheng Li (Wuhan University)

Few-Shot Keypoint Detection With Uncertainty Learning for Unseen Species 19394
Changsheng Lu (The Australian National University) and Piotr Koniusz (ANU College of Engineering and Computer Science)

Poster 4.2: Scene Analysis and Understanding

Stacked Hybrid-Attention and Group Collaborative Learning for Unbiased Scene Graph Generation 19405
Xingning Dong (Shandong University), Tian Gan (Shandong University), Xuemeng Song (Shandong University), Jianlong Wu (Shandong University), Yuan Cheng (Ant Group), and Liqiang Nie (Shandong University)

Structured Sparse R-CNN for Direct Scene Graph Generation 19415
Yao Teng (Nanjing University) and Limin Wang (Nanjing University)

PPDL: Predicate Probability Distribution Based Loss for Unbiased Scene Graph Generation 19425
Wei Li (Nankai University), Haiwei Zhang (Nankai University), Qijie Bai (Nankai University), Guoqing Zhao (Mashang Consumer Finance Co., Ltd), Ning Jiang (Mashang Consumer Finance Co., Ltd.), and Xiaojie Yuan (Nankai Univeristy)

RU-Net: Regularized Unrolling Network for Scene Graph Generation 19435
Xin Lin (South China University of Technology), Changxing Ding (South China University of Technology), Jing Zhang (The University of Sydney), Yibing Zhan (Jingdong), and Dacheng Tao (JD.com)

Fine-Grained Predicates Learning for Scene Graph Generation 19445
Xinyu Lyu (University of Electronic Science and Technology of China), Lianli Gao (The University of Electronic Science and Technology of China), Yuyu Guo (UESTC), Zhou Zhao (Zhejiang University), Hao Huang (Kuaishou), Heng Tao Shen (University of Electronic Science and Technology of China (UESTC)), and Jingkuan Song (UESTC)

HL-Net: Heterophily Learning Network for Scene Graph Generation 19454
Xin Lin (South China University of Technology), Changxing Ding (South China University of Technology), Yibing Zhan (Jingdong), Zijian Li (South China University of Technology), and Dacheng Tao (JD.com)

SGTR: End-to-End Scene Graph Generation With Transformer 19464
Rongjie Li (ShanghaiTech University), Songyang Zhang (ShanghaiTech University), and Xuming He (ShanghaiTech University)

Classification-Then-Grounding: Reformulating Video Scene Graphs As Temporal Bipartite Graphs 19475
Kaifeng Gao (Zhejiang University), Long Chen (Columbia University), Yulei Niu (Columbia University), Jian Shao (Zhejiang University), and Jun Xiao (Zhejiang University)

RelTransformer: A Transformer-Based Long-Tail Visual Relationship Recognition 19485
Jun Chen (King Abdullah University of Science and Technology), Aniket Agarwal (IIT Roorkee), Sherif Abdelkarim (KAUST), Deyao Zhu (King Abdullah University of Science and Technology), and Mohamed Elhoseiny (KAUST)

Spatial Commonsense Graph for Object Localisation in Partial Scenes 19496
Francesco Giuliari (University of Genoa; Istituto Italiano di Tecnologia (IIT), ,), Geri Skenderi (University of Verona), Marco Cristani (University of Verona), Yiming Wang (Istituto Italiano di Tecnologia), and Alessio Del Bue (Istituto Italiano di Tecnologia (IIT), ,)

"The Pedestrian Next to the Lamppost" Adaptive Object Graphs for Better Instantaneous Mapping 19506
Avishkar Saha (University of Surrey), Oscar Mendez (University of Surrey), Chris Russell (Amazon), and Richard Bowden (University of Surrey)

Category-Aware Transformer Network for Better Human-Object Interaction Detection 19516
Leizhen Dong (Huazhong University of Science and Technology), Zhimin Li (Huazhong University of Science and Technology), Kunlun Xu (Huazhong University of Science and Technology), Zhijun Zhang (Huazhong University of Science and Technology), Luxin Yan (Huazhong University of Science and Technology), Sheng Zhong (Huazhong University of Science and Technology), and Xu Zou (Huazhong University of Science and Technology)

Exploring Structure-Aware Transformer Over Interaction Proposals for Human-Object Interaction Detection 19526
Yong Zhang (The Chinese University of Hong Kong, Shenzhen), Yingwei Pan (JD AI Research), Ting Yao (JD AI Research), Rui Huang (The Chinese University of Hong Kong, Shenzhen), Tao Mei (AI Research of JD.com), and Chang-Wen Chen (The Hong Kong Polytechnic University)

Distillation Using Oracle Queries for Transformer-Based Human-Object Interaction Detection.... 19536
Xian Qu (South China University of Technology), Changxing Ding (South China University of Technology), Xingao Li (South China University of Technology), Xubin Zhong (South China University of Technology), and Dacheng Tao (JD.com)

Human-Object Interaction Detection via Disentangled Transformer 19546
Desen Zhou (Baidu, Inc.), Zhichao Liu (shanghaitech university), Jian Wang (Baidu), Leshan Wang (ShanghaiTech University), Tao Hu (baidu), Errui Ding (Baidu Inc.), and Jingdong Wang (Baidu)

MSTR: Multi-Scale Transformer for End-to-End Human-Object Interaction Detection 19556
Bumsoo Kim (Kakao Brain), Jonghwan Mun (Kakao Brain), Kyoung-Woon On (Kakao Brain), Minchul Shin (Kakao Brain), Junhyun Lee (Korea University), and Eun-Sol Kim (Hanyang University)

GaTector: A Unified Framework for Gaze Object Prediction 19566
Binglu Wang (Xi'an University of Architecture and Technology), Tao Hu (Xi'an University of Architecture and Technology), Baoshan Li (Xi'an University of Architecture and Technology), Xiaojuan Chen (Xi'an University of Architecture and Technology), and Zhijie Zhang (Xi'an University of Architecture and Technology)

STCrowd: A Multimodal Dataset for Pedestrian Perception in Crowded Scenes 19576
Peishan Cong (ShanghaiTech), Xinge Zhu (The Chinese University of Hong Kong), Feng Qiao (RWTH Aachen University), Yiming Ren (shanghaitech), Xidong Peng (Shanghaitech University), Yuenan Hou (Shanghai AI Lab), Lan Xu (ShanghaiTech), Ruigang Yang (University of Kentucky, USA), Dinesh Manocha (University of Maryland at College Park), and Yuexin Ma (ShanghaiTech University)

Crowd Counting in the Frequency Domain 19586
Weibo Shu (City University of Hong Kong), Jia Wan (City University of Hong Kong), Kay Chen Tan (Hong Kong Polytechnic University), Sam Kwong (City Univeristy of Hong Kong), and Antoni B. Chan (City University of Hong Kong, Hong, Kong)

Boosting Crowd Counting via Multifaceted Attention 19596
Hui Lin (Xi'an Jiaotong University), Zhiheng Ma (Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences), Rongrong Ji (Xiamen University, China), Yaowei Wang (PengCheng Laboratory), and Xiaopeng Hong (Harbin Institute of Technology)

Rethinking Spatial Invariance of Convolutional Networks for Object Counting 19606
Zhi-Qi Cheng (Carnegie Mellon University), Qi Dai (Microsoft Research), Hong Li (Carnegie Mellon University), Jingkuan Song (UESTC), Xiao Wu (Southwest Jiaotong University), and Alexander G. Hauptmann (Carnegie Mellon University)

Cerberus Transformer: Joint Semantic, Affordance and Attribute Parsing 19617
Xiaoxue Chen (Tsinghua University), Tianyu Liu (The Hong Kong University of Science and Technology), Hao Zhao (Intel Labs China), Guyue Zhou (Tsinghua University), and Ya-Qin Zhang (Tsinghua University)

Collaborative Transformers for Grounded Situation Recognition 19627
Junhyeong Cho (POSTECH), Youngseok Yoon (POSTECH), and Suha Kwak (POSTECH)

Poster 4.2: Computational Photography

Deep Stereo Image Compression via Bi-Directional Coding 19637
Jianjun Lei (Tianjin University), Xiangrui Liu (Tianjin University), Bo Peng (Tianjin University), Dengchao Jin (Tianjin University), Wanqing Li (University of Wollongong), and Jingxiao Gu (CalmCar Vehicle Vision System)

RFNet: Unsupervised Network for Mutually Reinforcing Multi-Modal Image Registration and Fusion 19647
Han Xu (Wuhan University), Jiayi Ma (Wuhan University), Jiteng Yuan (Wuhan University), Zhuliang Le (Wuhan University), and Wei Liu (Tencent)

Semi-Supervised Wide-Angle Portraits Correction by Multi-Scale Transformer 19657
Fushun Zhu (Sichuan University), Shan Zhao (Megvii), Peng Wang (Megvii), Hao Wang (Megvii), Hua Yan (Sichuan University), and Shuaicheng Liu (UESTC; Megvii)

Semi-Supervised Learning of Semantic Correspondence With Pseudo-Labels 19667
Jiwon Kim (Korea University), Kwangrok Ryoo (Korea University), Junyoung Seo (Korea University), Gyuseong Lee (Korea University), Daehwan Kim (Samsung electro mechanics), Hansang Cho (Samsung Electro-Mechanics), and Seungryong Kim (Korea University)

SCS-Co: Self-Consistent Style Contrastive Learning for Image Harmonization 19678
Yucheng Hang (Tsinghua University), Bin Xia (Tsinghua university), Wenming Yang (Tsinghua University), and Qingmin Liao (Tsinghua Univeristy)

Automatic Color Image Stitching Using Quaternion Rank-1 Alignment 19688
Jiaxue Li (University of Macau) and Yicong Zhou (University of Macau)

SpaceEdit: Learning a Unified Editing Space for Open-Domain Image Color Editing 19698
Jing Shi (university of rochester), Ning Xu (Adobe Research), Haitian Zheng (University of Rochester), Alex Smith (Adobe Inc.), Jiebo Luo (U. Rochester), and Chenliang Xu (University of Rochester)

Degree-of-Linear-Polarization-Based Color Constancy 19708
Taishi Ono (Sony Corporation), Yuhi Kondo (Sony Corporation), Legong Sun (Sony Corporation), Teppei Kurita (Sony Group Corporation), and Yusuke Moriuchi (Sony Group Corporation)

Point Cloud Color Constancy 19718
Xiaoyan Xing (Tsinghua University), Yanlin Qian (Honor), Sibo Feng (Huawei Technologies Co., Ltd.), Yuhan Dong (Tsinghua University), and Jiří Matas (CMP CTU FEE)

Boosting View Synthesis With Residual Transfer 19728
Xuejian Rong (Facebook), Jia-Bin Huang (Facebook), Ayush Saraf (Facebook), Changil Kim (Facebook), and Johannes Kopf (Facebook)

Deep Hyperspectral-Depth Reconstruction Using Single Color-Dot Projection 19738
Chunyu Li (Tokyo Institute of Technology), Yusuke Monno (Tokyo Institute of Technology), and Masatoshi Okutomi (Tokyo Institute of Technology)

Quantization-Aware Deep Optics for Diffractive Snapshot Hyperspectral Imaging 19748
Lingen Li (Beijing Institute of Technology), Lizhi Wang (Beijing Institute of Technology), Weitao Song (Beijing Institute of Technology), Lei Zhang (Beijing Institute of Technology), Zhiwei Xiong (University of Science and Technology of China), and Hua Huang (Beijing Normal University)

PIE-Net: Photometric Invariant Edge Guided Network for Intrinsic Image Decomposition 19758
Partha Das (University of Amsterdam), Sezer Karaoglu (University of Amsterdam), and Theo Gevers (University of Amsterdam)

Multimodal Material Segmentation 19768
Yupeng Liang (Kyoto University), Ryosuke Wakaki (Kyoto University), Shohei Nobuhara (Kyoto University), and Ko Nishino (Kyoto University)

Occlusion-Aware Cost Constructor for Light Field Depth Estimation 19777
Yingqian Wang (National University of Defense Technology), Longguang Wang (National University of Defense Technology), Zhengyu Liang (National University of Denfense Technology), Jungang Yang (National University of Defense Technology), Wei An (National University of Defense Technology), and Yulan Guo (National University of Defense Technology)

Learning Neural Light Fields With Ray-Space Embedding 19787
Benjamin Attal (Carnegie Mellon University), Jia-Bin Huang (Facebook), Michael Zollhöfer (Facebook Reality Labs), Johannes Kopf (Facebook), and Changil Kim (Facebook)

Acquiring a Dynamic Light Field Through a Single-Shot Coded Image 19798
Ryoya Mizuno (Nagoya University), Keita Takahashi (Nagoya University), Michitaka Yoshida (Osaka University), Chihiro Tsutake (Nagoya University), Toshiaki Fujii (Nagoya University), and Hajime Nagahara (Osaka University)

Gravitationally Lensed Black Hole Emission Tomography 19809
Aviad Levis (Caltech), Pratul P. Srinivasan (Google Research), Andrew A. Chael (Princeton University), Ren Ng (UC Berkeley), and Katherine L. Bouman (Caltech)

Deep Saliency Prior for Reducing Visual Distraction 19819
Kfir Aberman (Google), Junfeng He (Google Inc.), Yossi Gandelsman (BAIR), Inbar Mosseri (Google, Inc.), David E. Jacobs (Google), Kai Kohlhoff (Google), Yael Pritch (Google), and Michael Rubinstein (Google)

Personalized Image Aesthetics Assessment With Rich Attributes 19829
Yuzhe Yang (OPPO Research Institute), Liwu Xu (OPPO Research Institute), Leida Li (Xidian University), Nan Qie (OPPO Research Institute), Yaqian Li (OPPO Research Institute), Peng Zhang (OPPO Research Institute), and Yandong Guo (OPPO Research Institute)

Artistic Style Discovery With Independent Components 19838
Xin Xie (Dalian University of Technology), Yi Li (Dalian University of Technology), Huaibo Huang (Institute of Automation, Chinese Academy of Sciences), Haiyan Fu (Dalian University of Technology), Wanwan Wang (InsightOne Tech Co, Ltd), and Yanqing Guo (Dalian University of Technology)

Poster 4.2: Action and Event Recognition

Bridge-Prompt: Towards Ordinal Action Understanding in Instructional Videos 19848
Muheng Li (Tsinghua University), Lei Chen (Tsinghua University), Yueqi Duan (Tsinghua University), Zhilan Hu (Huawei), Jianjiang Feng (Tsinghua University), Jie Zhou (Tsinghua University), and Jiwen Lu (Tsinghua University)

SVIP: Sequence VerIfication for Procedures in Videos 19858
Yicheng Qian (ShanghaiTech University), Weixin Luo (Meituan), Dongze Lian (ShanghaiTech University), Xu Tang (Xiaohongshu Inc.), Peilin Zhao (Tencent AI Lab), and Shenghua Gao (Shanghaitech University)

Set-Supervised Action Learning in Procedural Task Videos via Pairwise Order Consistency 19871
Zijia Lu (Northeastern University) and Ehsan Elhamifar (Northeastern University)

Exploring Denoised Cross-Video Contrast for Weakly-Supervised Temporal Action Localization..... 19882
Jingjing Li (University of Alberta), Tianyu Yang (Tencent AI Lab), Wei Ji (University of Alberta), Jue Wang (Tencent AI Lab), and Li Cheng (ECE dept., University of Alberta)

GateHUB: Gated History Unit With Background Suppression for Online Action Detection 19893
Junwen Chen (Rochester Institute of Technology), Gaurav Mittal (Microsoft), Ye Yu (Microsoft), Yu Kong (Rochester Institute of Technology), and Mei Chen (Microsoft)

E2(GO)MOTION: Motion Augmented Event Stream for Egocentric Action Recognition 19903
Chiara Plizzari (Politecnico di Torino), Mirco Planamente (Politecnico di Torino), Gabriele Goletto (Politecnico di Torino), Marco Cannici (Politecnico di Milano), Emanuele Gusso (Polytechnic of Turin), Matteo Matteucci (Politecnico di Milano), and Barbara Caputo (Politecnico di Torino)

Hybrid Relation Guided Set Matching for Few-Shot Action Recognition 19916
Xiang Wang (Huazhong University of Science and Technology), Shiwei Zhang (DAMO Academy, Alibaba Group), Zhiwu Qing (Huazhong University of Science and Technology), Mingqian Tang (Alibaba Group), Zhengrong Zuo (Huazhong University of Science and Technology), Changxin Gao (Huazhong University of Science and Technology), Rong Jin (alibaba group), and Nong Sang (Huazhong University of Science and Technology)

Spatio-Temporal Relation Modeling for Few-Shot Action Recognition 19926
Anirudh Thatipelli (International Institute of Information Technology, Hyderabad), Sanath Narayan (Inception Institute of Artificial Intelligence), Salman Khan (MBZUAI/ANU), Rao Muhammad Anwer (MBZUAI/AALTO), Fahad Shahbaz Khan (MBZUAI), and Bernard Ghanem (KAUST)

Alignment-Uniformity Aware Representation Learning for Zero-Shot Video Classification 19936
Shi Pu (Tencent), Kaili Zhao (Beijing University of Posts and Telecommunications), and Mao Zheng (Tencent)

Cross-Modal Representation Learning for Zero-Shot Action Recognition 19946
Chung-Ching Lin (Microsoft), Kevin Lin (Microsoft), Lijuan Wang (Microsoft), Zicheng Liu (Microsoft), and Linjie Li (Microsoft)

Cross-Modal Background Suppression for Audio-Visual Event Localization 19957
Yan Xia (Zhejiang University) and Zhou Zhao (Zhejiang University)

Fine-Grained Temporal Contrastive Learning for Weakly-Supervised Temporal Action Localization 19967
Junyu Gao (CASIA), Mengyuan Chen (Institute of Automation, Chinese Academy of Sciences), and Changsheng Xu (CASIA)

An Empirical Study of End-to-End Temporal Action Detection 19978
Xiaolong Liu (Huazhong University of Science and Technology), Song Bai (University of Oxford), and Xiang Bai (Huazhong University of Science and Technology)

Everything at Once – Multi-Modal Fusion Transformer for Video Retrieval 19988
Nina Shvetsova (Goethe University Frankfurt), Brian Chen (Columbia University), Andrew Rouditchenko (MIT CSAIL), Samuel Thomas (IBM Research AI), Brian Kingsbury (IBM Research), Rogerio S. Feris (MIT-IBM Watson AI Lab, IBM Research), David Harwath (MIT CSAIL), James Glass (Massachusetts Institute of Technology), and Hilde Kuehne (University of Frankfurt)

DirecFormer: A Directed Attention in Transformer Approach to Robust Action Recognition 19998
Thanh-Dat Truong (University of Arkansas), Quoc-Huy Bui (University of Science, VNU-HCM), Chi Nhan Duong (Concordia University), Han-Seok Seo (University of Arkansas), Son Lam Phung (University of Wollongong, VinAI), Xin Li (West Virginia University), and Khoa Luu (University of Arkansas)

MS-TCT: Multi-Scale Temporal ConvTransformer for Action Detection 20009
Rui Dai (INRIA), Srijan Das (Stony Brook University), Kumara Kahatapitiya (Stony Brook University), Michael S. Ryoo (Stony Brook/Google), and François Brémond (Inria Sophia Antipolis, France)

Uncertainty-Guided Probabilistic Transformer for Complex Action Recognition 20020
Hongji Guo (Rensselaer Polytechnic Institute), Hanjing Wang (Renselaer Polytechnic Institute), and Qiang Ji (Rensselaer Polytechnic Institute)

AdaFocus V2: End-to-End Training of Spatial Dynamic Networks for Video Recognition 20030
Yulin Wang (Tsinghua University), Yang Yue (Tsinghua University), Yuanze Lin (University of Washington), Haojun Jiang (Tsinghua University), Zihang Lai (CMU), Victor Kulikov (Picsart), Nikita Orlov (PicsArt), Humphrey Shi (U of Oregon/UIUC), and Gao Huang (Tsinghua)

UBoCo: Unsupervised Boundary Contrastive Learning for Generic Event Boundary Detection ... 20041
Hyolim Kang (Yonsei University), Jinwoo Kim (Yonsei University), Taehyun Kim (Yonsei Univ.), and Seon Joo Kim (Yonsei University)

Detector-Free Weakly Supervised Group Activity Recognition 20051
Dongkeun Kim (POSTECH), Jinsung Lee (POSTECH), Minsu Cho (POSTECH), and Suha Kwak (POSTECH)

Multi-Grained Spatio-Temporal Features Perceived Network for Event-Based Lip-Reading 20062
Ganchao Tan (University of Science and Technology of China), Yang Wang (University of Science and Technology of China), Han Han (University of Science and Technology of China), Yang Cao (University of Science and Technology of China), Feng Wu (University of Science and Technology of China), and Zheng-Jun Zha (University of Science and Technology of China)

Efficient Two-Stage Detection of Human-Object Interactions With a Novel Unary-Pairwise Transformer 20072
Frederic Z. Zhang (Australian National University), Dylan Campbell (University of Oxford), and Stephen Gould (Australian National University, Australia)

Interactiveness Field in Human-Object Interactions 20081
Xinpeng Liu (Shanghai Jiao Tong University), Yong-Lu Li (Shanghai Jiao Tong University), Xiaoqian Wu (Shanghai Jiao Tong University), Yu-Wing Tai (Kuaishou Technology / HKUST), Cewu Lu (Shanghai Jiao Tong University), and Chi-Keung Tang (Hong Kong University of Science and Technology)

GEN-VLKT: Simplify Association and Enhance Interaction Understanding for HOI Detection 20091
Yue Liao (Beihang University), Aixi Zhang (Alibaba Group), Miao Lu (Alibaba), Yongliang Wang (Alibaba.inc), Xiaobo Li (Alibaba), and Si Liu (Beihang University)

Object-Relation Reasoning Graph for Action Recognition 20101
Yangjun Ou (Wuhan Univercity), Li Mi (Wuhan University), and Zhenzhong Chen (Wuhan University)

UBnormal: New Benchmark for Supervised Open-Set Video Anomaly Detection 20111
Andra Acsintoae (University of Bucharest), Andrei Florescu (University of Bucharest), Mariana-Iuliana Georgescu (University of Bucharest), Tudor Mare (SecurifAI), Paul Sumedrea (University of Bucharest), Radu Tudor Ionescu (University of Bucharest), Fahad Shahbaz Khan (MBZUAI), and Mubarak Shah (University of Central Florida)

Decoupling and Recoupling Spatiotemporal Representation for RGB-D-Based Motion Recognition..... 20122
Benjia Zhou (Macau University of Science and Technology), Pichao Wang (Alibaba Group), Jun Wan (NLPR, CASIA), Yanyan Liang (Macau University of Science and Technology, Macao SAR, China), Fan Wang (Alibaba Group), Du Zhang (Macau University of Science and Technology), Zhen Lei (NLPR, CASIA, China), Hao Li (Alibaba Group), and Rong Jin (alibaba group)

SPAct: Self-Supervised Privacy Preservation for Action Recognition 20132
Ishan Rajendrakumar Dave (University of Central Florida), Chen Chen (University of Central Florida), and Mubarak Shah (University of Central Florida)

Unsupervised Action Segmentation by Joint Representation Learning and Online Clustering 20142
Sateesh Kumar (Retrocausal), Sanjay Haresh (Retrocausal, Inc), Awais Ahmed (Retrocausal), Andrey Konin (Retrocausal), M. Zeeshan Zia (Retrocausal, Inc.), and Quoc-Huy Tran (Retrocausal, Inc.)

InfoGCN: Representation Learning for Human Skeleton-Based Action Recognition 20154
Hyung-gun Chi (Purdue University), Myoung Hoon Ha (KAIST), Seunggeun Chi (Purdue University), Sang Wan Lee (KAIST), Qixing Huang (The University of Texas at Austin), and Karthik Ramani (Purdue University, USA)

Learning Video Representations of Human Motion From Synthetic Data 20165
Xi Guo (State Key Laboratory of Software Development Environment, Beihang University), Wei Wu (SenseTime Group Limited), Dongliang Wang (SenseTime Group Limited), Jing Su (SenseTime Group Limited), Haisheng Su (SenseTime Group Limited), Weihao Gan (SenseTime Group Limited), Jian Huang (School of Software, Beihang University), and Qin Yang (State Key Laboratory of Software Development Environment, Beihang University)

Learnable Irrelevant Modality Dropout for Multimodal Action Recognition on Modality-Specific Annotated Videos 20176
Saghir Alfasly (Shenzhen University), Jian Lu (Shenzhen University), Chen Xu (Shenzhen University), and Yuru Zou (Shenzhen University)

Poster 4.2: Biometrics

EyePAD++: A Distillation-Based Approach for Joint Eye Authentication and Presentation Attack Detection Using Periocular Images 20186
Prithviraj Dhar (Johns Hopkins University), Amit Kumar (Facebook), Kirsten Kaplan (Meta), Khushi Gupta (Facebook Inc), Rakesh Ranjan (Facebook), and Rama Chellappa (Johns Hopkins University)

Gait Recognition in the Wild With Dense 3D Representations and a Benchmark 20196
Jinkai Zheng (Hangzhou Dianzi University), Xinchen Liu (AI Research of JD.com), Wu Liu (AI Research of JD.com), Lingxiao He (AI Research of JD.com), Chenggang Yan (Hangzhou Dianzi University), and Tao Mei (AI Research of JD.com)

Camera-Conditioned Stable Feature Generation for Isolated Camera Supervised Person Re-IDentification 20206
Chao Wu (Sun Yat-sen University), Wenhang Ge (Sun Yat-sen University), Ancong Wu (Sun Yat-sen University), and Xiaobin Chang (Sun Yat-Sen University)

Lagrange Motion Analysis and View Embeddings for Improved Gait Recognition 20217
Tianrui Chai (Beihang University), Annan Li (Beijing University of Aeronautics and Astronautics), Shaoxiong Zhang (Beihang University), Zilong Li (University of Chinese Academy of Sciences), and Yunhong Wang (State Key Laboratory of Virtual Reality Technology and System, Beihang University, Beijing, China)

DeepFace-EMD: Re-Ranking Using Patch-Wise Earth Mover's Distance Improves Out-of-Distribution Face Identification 20227
Hai Phan (Carnegie Mellon University) and Anh Nguyen (Auburn University)

Learning Second Order Local Anomaly for General Face Forgery Detection 20238
Jianwei Fei (Nanjing University of Information Science and Technology), Yunshu Dai (Nanjing University of Information Science and Technology), Peipeng Yu (NUIST), Tianrun Shen (Nanjing University), Zhihua Xia (Jinan University), and Jian Weng (Jinan University)

PatchNet: A Simple Face Anti-Spoofing Framework via Fine-Grained Patch Recognition 20249
Chien-Yi Wang (Microsoft), Yu-Ding Lu (UC Merced), Shang-Ta Yang (Microsoft), and Shang-Hong Lai (Microsoft)

Face2Exp: Combating Data Biases for Facial Expression Recognition 20259
Dan Zeng (Southern University of Science and Technology), Zhiyuan Lin (Southern University of Science and Technology), Xiao Yan (Southern University of Science and Technology), Yuting Liu (sichuan university), Fei Wang (Southern University of Science and Technology), and Bo Tang (Southern University of Science and Technology)

Local-Adaptive Face Recognition via Graph-Based Meta-Clustering and Regularized Adaptation..... 20269
Wenbin Zhu (Microsoft), Chien-Yi Wang (Microsoft), Kuan-Lun Tseng (Microsoft), Shang-Hong Lai (Microsoft), and Baoyuan Wang (Xiaobing.AI)

Poster 4.2: Face and Gestures

EMOCA: Emotion Driven Monocular Face Capture and Animation 20279
Radek Daněček (Max Planck Institute for Intelligent Systems), Michael J. Black (Max Planck Institute for Intelligent Systems), and Timo Bolkart (Max Planck Institute for Intelligent Systems)

Robust Egocentric Photo-Realistic Facial Expression Transfer for Virtual Reality 20291
Amin Jourabloo, Fernando De la Torre (Carnegie Mellon), Jason Saragih (Facebook), Shih-En Wei (Facebook), Stephen Lombardi (Facebook), Te-Li Wang, Danielle Belko (Facebook), Autumn Trimble (Facebook Reality Labs Pittsburgh), and Hernan Badino (Facebook)

FaceVerse: A Fine-Grained and Detail-Controllable 3D Face Morphable Model From a Hybrid Dataset 20301
Lizhen Wang (Tsinghua University), Zhiyuan Chen (Ant Financial Services Group), Tao Yu (Tsinghua University), Chenguang Ma (Ant Financial Services Group), Liang Li (Ant Financial Service Group), and Yebin Liu (Tsinghua University)

ImFace: A Nonlinear 3D Morphable Face Model With Implicit Neural Representations 20311
Mingwu Zheng (Beihang University), Hongyu Yang (Beihang University), Di Huang (Beihang University, China), and Liming Chen (Ecole Centrale de Lyon)

Physically-Guided Disentangled Implicit Rendering for 3D Face Modeling 20321
Zhenyu Zhang (Tencent), Yanhao Ge (Tencent YouTu), Ying Tai (Tencent YouTu), Weijian Cao (Tencent Youtu Lab), Renwang Chen (Tencent), Kunlin Liu (University of Science and Technology of China), Hao Tang (ETH Zurich), Xiaoming Huang (Tencent Youtu), Chengjie Wang (Tencent), Zhifeng Xie (Shanghai University), and Dongjin Huang (Shanghai University)

RigNeRF: Fully Controllable Neural 3D Portraits 20332
ShahRukh Athar (Stony Brook University), Zexiang Xu (Adobe Research), Kalyan Sunkavalli (Adobe Research), Eli Shechtman (Adobe Research, US), and Zhixin Shu (Adobe Research)

HeadNeRF: A Real-Time NeRF-Based Parametric Head Model 20342
Yang Hong (University of Science and Technology of China), Bo Peng (University of Science and Technology of China), Haiyao Xiao (University of Science and Technology of China), Ligang Liu (University of Science and Technology of China), and Juyong Zhang (University of Science and Technology of China)

Sparse to Dense Dynamic 3D Facial Expression Generation 20353
Naima Otberdout (University of Lille, CRIStAL), Claudio Ferrari (University of Parma), Mohamed Daoudi (IMT Lille Douai), Stefano Berretti (University of Florence, Italy), and Alberto Del Bimbo (University of Florence)

Learning To Listen: Modeling Non-Deterministic Dyadic Facial Motion 20363
Evonne Ng (UC Berkeley), Hanbyul Joo (Facebook AI Research), Liwen Hu (Pinscreen), Hao Li (Pinscreen), Trevor Darrell (UC Berkeley), Angjoo Kanazawa (University of California Berkeley), and Shiry Ginosar (UC Berkeley)

Speech Driven Tongue Animation 20374
Salvador Medina (Carnegie Mellon University), Denis Tome (EpicGames), Carsten Stoll (Epic Games), Mark Tiede (Haskins Laboratories), Kevin Munhall (Queen's University), Alexander G. Hauptmann (Carnegie Mellon University), and Iain Matthews (Carnegie Mellon University)

Knowledge-Driven Self-Supervised Representation Learning for Facial Action Unit Recognition 20385
Yanan Chang (University of Science and Technology of China) and Shangfei Wang (University of Science and Technology of China)

gDNA: Towards Generative Detailed Neural Avatars 20395
Xu Chen (ETH Zürich), Tianjian Jiang (ETH Zurich), Jie Song (ETH Zurich), Jinlong Yang (Max Planck Institute for Intelligent Systems), Michael J. Black (Max Planck Institute for Intelligent Systems), Andreas Geiger (University of Tuebingen), and Otmar Hilliges (ETH Zurich)

GraFormer: Graph-Oriented Transformer for 3D Pose Estimation 20406
Weixi Zhao (University of Chinese Academy of Sciences), Weiqiang Wang (University of Chinese Academy of Sciences), and Yunjie Tian (University of Chinese Academy of Science)

Uncertainty-Aware Adaptation for Self-Supervised 3D Human Pose Estimation 20416
Jogendra Nath Kundu (Indian Institute of Science), Siddharth Seth (Indian Institute of Science), Pradyumna YM (Indian Institute of Science), Varun Jampani (Google), Anirban Chakraborty (Indian Institute of Science), and R. Venkatesh Babu (Indian Institute of Science)

Towards Diverse and Natural Scene-Aware 3D Human Motion Synthesis 20428
Jingbo Wang (The Chinese University of HongKong), Yu Rong (The Chinese University of Hong Kong), Jingyuan Liu (Hong Kong Univerisity of Science and Technology), Sijie Yan (Chinese University of Hong Kong), Dahua Lin (The Chinese University of Hong Kong), and Bo Dai (Shanghai AI Lab)

PINA: Learning a Personalized Implicit Neural Avatar From a Single RGB-D Video Sequence ... 20438
Zijian Dong (ETH Zurich), Chen Guo (ETH Zurich), Jie Song (ETH Zurich), Xu Chen (ETH Zürich), Andreas Geiger (University of Tuebingen), and Otmar Hilliges (ETH Zurich)

The Wanderings of Odysseus in 3D Scenes 20449
Yan Zhang (ETH Zurich) and Siyu Tang (ETH Zurich)

OSSO: Obtaining Skeletal Shape From Outside 20460
Marilyn Keller (Max Planck Institute for Intelligent Systems), Silvia Zuffi (IMATI-CNR), Michael J. Black (Max Planck Institute for Intelligent Systems), and Sergi Pujades (Université Grenoble Alpes)

LiDARCap: Long-Range Marker-Less 3D Human Motion Capture With LiDAR Point Clouds 20470
Jialian Li (Xiamen University), Jingyi Zhang (Xiamen University), Zhiyong Wang (Xiamen University), Siqi Shen (Xiamen University), Chenglu Wen (Xiamen University), Yuexin Ma (ShanghaiTech University), Lan Xu (ShanghaiTech), Jingyi Yu (Shanghai Tech University), and Cheng Wang (Xiamen University)

Unimodal-Concentrated Loss: Fully Adaptive Label Distribution Learning for Ordinal Regression 20481
Qiang Li (Hikvision Research Institute), Jingjing Wang (Hikvision Research Institute), Zhaoliang Yao (Hikvision Research Institute), Yachun Li (Hikvision Research Institute), Pengju Yang (Hikvision Research Institute), Jingwei Yan (Hikvision Research Institute), Chunmao Wang (Hikvision Research Institute), and Shiliang Pu (Hikvision Research Institute)

Spatial-Temporal Parallel Transformer for Arm-Hand Dynamic Estimation 20491
Shuying Liu (NetEase Games), Wenbin Wu (NetEase Games), Jiaxian Wu (NetEase Games), and Yue Lin (NetEase Games AI Lab)

LISA: Learning Implicit Shape and Appearance of Hands 20501
Enric Corona (IRI), Tomas Hodan (Facebook Reality Labs), Minh Vo (Facebook Reality Labs), Francesc Moreno-Noguer (IRI), Chris Sweeney (Facebook Reality Labs), Richard Newcombe (Facebook), and Lingni Ma (Facebook Reality Labs)

MobRecon: Mobile-Friendly Hand Mesh Reconstruction From Monocular Image 20512
Xingyu Chen (Kuaishou Technology), Yufeng Liu (SEU-ALLEN Joint Center, Institute for Brain and Intelligence, Southeast University), Yajiao Dong (Kuaishou Technology), Xiong Zhang (Live YY, Baidu Inc.), Chongyang Ma (Kuaishou Technology), Yanmin Xiong (Kwai), Yuan Zhang (Kuaishou Technology), and Xiaoyan Guo (Kuaishou Technology)

Mining Multi-View Information: A Strong Self-Supervised Framework for Depth-Based 3D Hand Pose and Mesh Estimation 20523
Pengfei Ren (Beijing University of Posts and Telecommunications), Haifeng Sun (Beijing university of posts and telecommunications), Jiachang Hao (Beijing University of Posts and Telecommunications), Jingyu Wang (Beijing University of Posts and Telecommunications), Qi Qi (Beijing University of Posts and Telecommunications), and Jianxin Liao (beijing university of posts and telecommunications)

Low-Resource Adaptation for Personalized Co-Speech Gesture Generation 20534
Chaitanya Ahuja (Carnegie Mellon University), Dong Won Lee (Carnegie Mellon University), and Louis-Philippe Morency (Carnegie Mellon University)

D-Grasp: Physically Plausible Dynamic Grasp Synthesis for Hand-Object Interactions 20545
Sammy Christen (ETH Zurich), Muhammed Kocabas (Max Planck Institute for Intelligent Systems / ETH Zurich), Emre Aksan (ETH Zurich), Jemin Hwangbo (KAIST), Jie Song (ETH Zurich), and Otmar Hilliges (ETH Zurich)

Poster 4.2: Medical, Biological and Cell Microscopy

Synthetic Generation of Face Videos With Plethysmograph Physiology 20555
Zhen Wang (UCLA), Yunhao Ba (UCLA), Pradyumna Chari (UCLA), Oyku Deniz Bozkurt (UCLA), Gianna Brown (Visual Machines Group), Parth Patwa (UCLA), Niranjan Vaddi (UCLA), Laleh Jalilian (UCLA), and Achuta Kadambi (UCLA)

Contour-Hugging Heatmaps for Landmark Detection 20565
James McCouat (University of Oxford) and Irina Voiculescu (University of Oxford)

Which Images To Label for Few-Shot Medical Landmark Detection? 20574
Quan Quan (Institute of Computing Technology, Chinese Acadamy of Sciences), Qingsong Yao (Institute of Computing Technology, CAS), Jun Li (Institute of Computing Technology, Chinese Academy of Sciences), and S. Kevin Zhou (USTC)

Self-Supervised Bulk Motion Artifact Removal in Optical Coherence Tomography Angiography 20585
Jiaxiang Ren (Stony Brook University), Kicheon Park (Stony Brook Univ), Yingtian Pan (Stony Brook University), and Haibin Ling (Stony Brook University)

Multi-Marginal Contrastive Learning for Multi-Label Subcellular Protein Localization 20594
Ziyi Liu (Wuhan University), Zengmao Wang (Wuhan University), and Bo Du (Wuhan University)

Transformer-Empowered Multi-Scale Contextual Matching and Aggregation for Multi-Contrast MRI Super-Resolution 20604
Guangyuan Li (Yantai University), Jun Lv (Yantai University), Yapeng Tian (University of Rochester), Qi Dou (The Chinese University of Hong Kong), Chengyan Wang (Fudan University), Chenliang Xu (University of Rochester), and Jing Qin (The Hong Kong Polytechnic University)

Harmony: A Generic Unsupervised Approach for Disentangling Semantic Content From Parameterized Transformations 20614
Mostofa Rafid Uddin (Carnegie Mellon University), Gregory Howe (Carnegie Mellon), Xiangrui Zeng (Carnegie Mellon University), and Min Xu (Carnegie Mellon University)

Cross-Modal Clinical Graph Transformer for Ophthalmic Report Generation 20624
Mingjie Li (Monash University), Wenjia Cai (State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University), Karin Verspoor (RMIT University), Shirui Pan (Monash University), Xiaodan Liang (Sun Yat-sen University), and Xiaojun Chang (University of Technology Sydney)

BoostMIS: Boosting Medical Image Semi-Supervised Learning With Adaptive Pseudo Labeling and Informative Active Annotation 20634
Wenqiao Zhang (NUS), Lei Zhu (NUS Graduate School Integrative Sciences and Engineering Programme), James Hallinan (NUH), Shengyu Zhang (National University of Singapore), Andrew Makmur (National University Hospital Singapore), Qingpeng Cai (National University of Singapore), and Beng Chin Ooi (NUS)

Incremental Cross-View Mutual Distillation for Self-Supervised Medical CT Synthesis 20645
Chaowei Fang (Xidian University), Liang Wang (Xidian University), Dingwen Zhang (NWPU), Jun Xu (Nankai University), Yixuan Yuan (City University of Hong Kong), and Junwei Han (Northwestern Polytechnical University)

Towards Low-Cost and Efficient Malaria Detection 20655
Waqas Sultani (Information Technology University), Wajahat Nawaz (Information Technology University), Syed Javed (Information Technology University.), Muhammad Sohail Danish (Information Technology University, Lahore), Asma Saadia (Central Park Medical College), and Mohsen Ali (Information Technology University. Lahore)

ACPL: Anti-Curriculum Pseudo-Labelling for Semi-Supervised Medical Image Classification 20665
Fengbei Liu (University of Adelaide), Yu Tian (Australian Institute for Machine Learning, University of Adelaide), Yuanhong Chen (Australian Institute for Machine Learning, University of Adelaide), Yuyuan Liu (University of Adelaide), Vasileios Belagiannis (Otto von Guericke University Magdeburg), and Gustavo Carneiro (University of Adelaide)

Multimodal Dynamics: Dynamical Fusion for Trustworthy Multimodal Classification 20675
Zongbo Han (Tianjin University), Fan Yang (Tencent AI Lab), Junzhou Huang (University of Texas at Arlington), Changqing Zhang (Tianjin university), and Jianhua Yao (Tencent AI Lab)

M3T: Three-Dimensional Medical Image Classifier Using Multi-Plane and Multi-Slice Transformer 20686
Jinseong Jang (Yonsei University) and Dosik Hwang (Yonsei University)

Self-Supervised Pre-Training of Swin Transformers for 3D Medical Image Analysis 20698
Yucheng Tang (Vanderbilt University), Dong Yang (NVIDIA Corporation), Wenqi Li (NVIDIA), Holger R. Roth (NVIDIA), Bennett Landman (Vanderbilt University), Daguang Xu (NVIDIA Corporation), Vishwesh Nath (NVIDIA), and Ali Hatamizadeh (NVIDIA Corporation)

HyperSegNAS: Bridging One-Shot Neural Architecture Search With 3D Medical Image Segmentation Using HyperNet 20709
Cheng Peng (Johns Hopkins University), Andriy Myronenko (NVIDIA), Ali Hatamizadeh (NVIDIA Corporation), Vishwesh Nath (NVIDIA), Md Mahfuzur Rahman Siddiquee (Arizona State University), Yufan He (NVidia Corporation), Daguang Xu (NVIDIA Corporation), Rama Chellappa (Johns Hopkins University), and Dong Yang (NVIDIA Corporation)

DArch: Dental Arch Prior-Assisted 3D Tooth Instance Segmentation With Weak Annotations 20720
Liangdong Qiu (Shenzhen Research Institute of Big Data, the Chinese University of Hong Kong (Shenzhen)), Chongjie Ye (The Chinese University of Hong Kong, Shenzhen), Pei Chen (the Chinese University of Hong Kong (Shenzhen)), Yunbi Liu (School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen), Xiaoguang Han (Shenzhen Research Institute of Big Data, the Chinese University of Hong Kong (Shenzhen)), and Shuguang Cui (The Chinese University of Hong Kong, Shenzhen)

Clean Implicit 3D Structure From Noisy 2D STEM Images 20730
Hannah Kniesel (Ulm University), Timo Ropinski (Ulm University), Tim Bergner (central facility electron microscopy university ulm), Kavitha Shaga Devan (Universität Ulm), Clarissa Read (Ulm University), Paul Walther (Ulm University), Tobias Ritschel (UCL), and Pedro Hermosilla (Ulm University)

Vox2Cortex: Fast Explicit Reconstruction of Cortical Surfaces From 3D MRI Scans With Geometric Deep Neural Networks 20741
Fabian Bongratz (Technical University of Munich), Anne-Marie Rickmann (Ludwig Maximilians University Munich), Sebastian Pölsterl (Ludwig-Maximilians Universitat), and Christian Wachinger (Technical University of Munich)

Aladdin: Joint Atlas Building and Diffeomorphic Registration Learning With Pairwise Alignment 20752
Zhipeng Ding (Department of Computer Science, University of North Carolina at Chapel Hill) and Marc Niethammer (UNC)

Learning Optimal K-Space Acquisition and Reconstruction Using Physics-Informed Neural Networks 20762
Wei Peng (CMVS, University of Oulu), Li Feng (Memorial Sloan Kettering Cancer Center), Guoying Zhao (University of Oulu), and Fang Liu (Harvard University)

NODEO: A Neural Ordinary Differential Equation Based Optimization Framework for Deformable Image Registration 20772
Yifan Wu (University of Pennsylvania), Tom Z. Jiahao (University of Pennsylvania), Jiancong Wang (University of Pennsylvania), Paul A. Yushkevich (University of Pennsylvania), M. Ani Hsieh (University of Pennsylvania), and James C. Gee (University of Pennsylvania)

SMPL-A: Modeling Person-Specific Deformable Anatomy 20782
Hengtao Guo (Rensselaer Polytechnic Institute), Benjamin Planche (United Imaging Intelligence), Meng Zheng (United Imaging Intelligence), Srikrishna Karanam (Adobe Research), Terrence Chen (United Imaging Intelligence), and Ziyan Wu (United Imaging Intelligence)

DiRA: Discriminative, Restorative, and Adversarial Learning for Self-Supervised Medical Image Analysis 20792
Fatemeh Haghighi (Arizona State University), Mohammad Reza Hosseinzadeh Taher (Arizona State University), Michael B. Gotway (Mayo Clinic), and Jianming Liang (Arizona State University, USA)

Affine Medical Image Registration With Coarse-To-Fine Vision Transformer 20803
Tony C. W. Mok (Hong Kong University of Science and Technology) and Albert C. S. Chung (Hong Kong University of Science and Technology)

Topology-Preserving Shape Reconstruction and Registration via Neural Diffeomorphic Flow 20813
Shanlin Sun (University of California, Irvine), Kun Han (University of California Irvine), Deying Kong (university of california, irvine), Hao Tang (University of California Irvine), Xiangyi Yan (University of California, Irvine), and Xiaohui Xie (University of California, Irvine)

Generalizable Cross-Modality Medical Image Segmentation via Style Augmentation and Dual Normalization 20824
Ziqi Zhou (Nanjing University), Lei Qi (Southeast University), Xin Yang (Shenzhen University), Dong Ni (Shenzhen University, China), and Yinghuan Shi (Nanjing University)

Closing the Generalization Gap of Cross-Silo Federated Medical Image Segmentation 20834
An Xu (University of Pittsburgh), Wenqi Li (NVIDIA), Pengfei Guo (Johns Hopkins University), Dong Yang (NVIDIA Corporation), Holger R. Roth (NVIDIA), Ali Hatamizadeh (NVIDIA Corporation), Can Zhao (Nvidia), Daguang Xu (NVIDIA Corporation), Heng Huang (University of Pittsburgh), and Ziyue Xu (NVIDIA)

FIBA: Frequency-Injection Based Backdoor Attack in Medical Image Analysis 20844
Yu Feng (Northwestern Polytechnical University), Benteng Ma (Northwestern Polytechnical University), Jing Zhang (The University of Sydney), Shanshan Zhao (JD.COM), Yong Xia (Northwestern Polytechnical University, Research & Development Institute of Northwestern Polytechnical University in Shenzhen), and Dacheng Tao (JD.com)

Surpassing the Human Accuracy: Detecting Gallbladder Cancer From USG Images With Curriculum Learning 20854
Soumen Basu (IIT Delhi), Mayank Gupta (IIT Delhi), Pratyaksha Rana (PGIMER Chandigarh), Pankaj Gupta (PGIMER Chandigarh), and Chetan Arora (Indian Institute of Technology Delhi)

CellTypeGraph: A New Geometric Computer Vision Benchmark 20865
Lorenzo Cerrone (HCI/IWR uni heildelberg), Athul Vijayan (Max Planck Institute for Plant Breeding Research, Koln), Tejasvinee Mody (Technical University of Munich), Kay Schneitz (Technical University of Munich), and Fred A. Hamprecht (Heidelberg Collaboratory for Image Processing)

ContIG: Self-Supervised Multimodal Contrastive Learning for Medical Imaging With Genetics .. 20876
Aiham Taleb (Hasso-plattner-institute), Matthias Kirchler (Hasso Plattner Institute), Remo Monti (Hasso Plattner Institute), and Christoph Lippert (Hasso Plattner Insitute for Digital Engineering, Universität Potsdam)

Poster 4.2: Datasets and Evaluation

FERV39k: A Large-Scale Multi-Scene Dataset for Facial Expression Recognition in Videos 20890
Yan Wang (Fudan University), Yixuan Sun (Fudan University), Yiwen Huang (Fudan University), Zhongying Liu (Fudan University), Shuyong Gao (Fudan University), Wei Zhang (Fudan University), Weifeng Ge (Fudan University), and Wenqiang Zhang (Fudan University)

Multi-Dimensional, Nuanced and Subjective – Measuring the Perception of Facial Expressions... 20900
De'Aira Bryant (Georgia Institute of Technology), Siqi Deng (Amazon), Nashlie Sephus (AWS), Wei Xia (Amazon), and Pietro Perona (California Institute of Technology)

DAD-3DHeads: A Large-Scale Dense, Accurate and Diverse Dataset for 3D Head Alignment From a Single Image 20910
Tetiana Martyniuk (Ukrainian Catholic University), Orest Kupyn (Ukrainian Catholic University), Yana Kurlyak (Ukrainian Catholic University), Igor Krashenyi (Piñata Farms), Jiří Matas (CMP CTU FEE), and Viktoriia Sharmanska (University of Sussex and Imperial College London)

OakInk: A Large-Scale Knowledge Repository for Understanding Hand-Object Interaction 20921
Lixin Yang (Shanghai Jiao Tong University), Kailin Li (Shanghai Jiao Tong University), Xinyu Zhan (Shanghai Jiao Tong University), Fei Wu (Shanghai Jiao Tong University), Anran Xu (Shanghai Jiao Tong University), Liu Liu (Shanghai JiaoTong University), and Cewu Lu (Shanghai Jiao Tong University)

PoseTrack21: A Dataset for Person Search, Multi-Object Tracking and Multi-Person Pose Tracking 20931
Andreas Döring (University of Bonn), Di Chen, Shanshan Zhang (Nanjing University of Science and Technology), Bernt Schiele (MPI Informatics), and Jürgen Gall (University of Bonn)

Learning Modal-Invariant and Temporal-Memory for Video-Based Visible-Infrared Person Re-Identification 20941
Xinyu Lin (The Chinese University of Hong Kong (Shenzhen), ,), Jinxing Li (Harbin Institute of Technology, Shenzhen), Zeyu Ma (Harbin Institute of Technology, Shenzhen), Huafeng Li (Kunming University of Science and Technology), Shuang Li (Kunming University of Science and Technology), Kaixiong Xu (Kunming university of science and technology), Guangming Lu (Harbin Institute of Technology, Shenzhen), and David Zhang (Chinese University of Hong Kong (Shenzhen), ,)

JRDB-Act: A Large-Scale Dataset for Spatio-Temporal Action, Social Group and Activity Detection 20951
Mahsa Ehsanpour (The University of Adelaide), Fatemeh Saleh (Samsung AI Center), Silvio Savarese (Stanford University), Ian Reid (University of Adelaide, Australia), and Hamid Rezatofighi (Monash University)

DanceTrack: Multi-Object Tracking in Uniform Appearance and Diverse Motion 20961
Peize Sun (The University of Hong Kong), Jinkun Cao (Carnegie Mellon University), Yi Jiang (Bytedance), Zehuan Yuan (Bytedance.Inc), Song Bai (University of Oxford), Kris Kitani (Carnegie Mellon University), and Ping Luo (The University of Hong Kong)

Egocentric Prediction of Action Target in 3D 20971
Yiming Li (New York University), Ziang Cao (Tongji University), Andrew Liang (New York University), Benjamin Liang (New York University), Luoyao Chen (New York University), Hang Zhao (Tsinghua University), and Chen Feng (New York University)

HOI4D: A 4D Egocentric Dataset for Category-Level Human-Object Interaction 20981
Yunze Liu (Tsinghua University), Yun Liu (Tsinghua University), Che Jiang (TsingHua University), Kangbo Lyu (Tsinghua University), Weikang Wan (Peking University), Hao Shen (Peking University), Boqiang Liang (Peking University), Zhoujie Fu (Tsinghua University), He Wang (Peking University), and Li Yi (Tsinghua University)

Amodal Panoptic Segmentation 20991
Rohit Mohan (University of Freiburg) and Abhinav Valada (University of Freiburg)

Large-Scale Video Panoptic Segmentation in the Wild: A Benchmark 21001
Jiaxu Miao (Zhejiang University), Xiaohan Wang (Zhejiang University), Yu Wu (Princeton University), Wei Li (Zhejiang University), Xu Zhang (Zhejiang University), Yunchao Wei (UTS), and Yi Yang (UTS)

YouMVOS: An Actor-Centric Multi-Shot Video Object Segmentation Dataset 21012
Donglai Wei (Boston College), Siddhant Kharbanda (Aalto University), Sarthak Arora (Boston University), Roshan Roy (Birla Institute of Technology and Science, Pilani), Nishant Jain (iit roorkee), Akash Palrecha (Pixxel), Tanav Shah (IIT ROORKEE), Shray Mathur (Birla Institute of Technology and Science), Ritik Mathur (BITS Pilani, pilani campus), Abhijay Kemkar (Indian Institute of Technology Roorkee, India), Anirudh Chakravarthy (Carnegie Mellon University), Zudi Lin (Harvard University), Won-Dong Jang (Harvard University), Yansong Tang (Tsinghua University), Song Bai (University of Oxford), James Tompkin (Brown University), Philip H.S. Torr (University of Oxford), and Hanspeter Pfister (Harvard University)

The DEVIL Is in the Details: A Diagnostic Evaluation Benchmark for Video Inpainting 21022
Ryan Szeto (University of Michigan Ann Arbor) and Jason J. Corso (University of Michigan)

3MASSIV: Multilingual, Multimodal and Multi-Aspect Dataset of Social Media Short Videos 21032
Vikram Gupta (ShareChat), Trisha Mittal (University of Maryland), Puneet Mathur (University of Maryland), Vaibhav Mishra (ShareChat, India), Mayank Maheshwari (ShareChat Labs), Aniket Bera (University of Maryland, College Park), Debdoot Mukherjee (Sharechat), and Dinesh Manocha (University of Maryland at College Park)

AxIoU: An Axiomatically Justified Measure for Video Moment Retrieval 21044
Riku Togashi (CyberAgent, Inc., Waseda University), Mayu Otani (CyberAgent), Yuta Nakashima (Osaka University), Esa Rahtu (Tampere University), Janne Heikkilä (University of Oulu, Finland), and Tetsuya Sakai (Waseda University)

A Large-Scale Comprehensive Dataset and Copy-Overlap Aware Evaluation Protocol for Segment-Level Video Copy Detection 21054
Sifeng He (Ant Group), Xudong Yang (AntGroup), Chen Jiang (Antgroup), Gang Liang (Ant Group), Wei Zhang (Ant Financial Services Group), Tan Pan (Ant Group), Qing Wang (Ant Group), Furong Xu (Ant Group), Chunguang Li (ant group), JinXiong Liu (), Hui Xu (ant group), Kaiming Huang (ant group), Yuan Cheng (Ant Group), Feng Qian (ANT GROUP), Xiaobo Zhang (Ant Financial), and Lei Yang (Ant Group)

Assembly101: A Large-Scale Multi-View Video Dataset for Understanding Procedural Activities 21064
Fadime Sener (Facebook Reality Labs), Dibyadip Chatterjee (National University of Singapore), Daniel Shelepov (Facebook Reality Labs), Kun He (Facebook Reality Labs), Dipika Singhania (National University of Singapore), Robert Wang (Facebook Reality Labs), and Angela Yao (National University of Singapore)

Optimal Correction Cost for Object Detection Evaluation 21075
Mayu Otani (CyberAgent), Riku Togashi (CyberAgent, Inc., Waseda University), Yuta Nakashima (Osaka University), Esa Rahtu (Tampere University), Janne Heikkilä (University of Oulu, Finland), and Shin'ichi Satoh (National Institute of Informatics)

GrainSpace: A Large-Scale Dataset for Fine-Grained and Domain-Adaptive Recognition of Cereal Grains 21084
Lei Fan (University Of New South Wales), Yiwen Ding (Gauture), Dongdong Fan (GAUTURE), Donglin Di (Baidu Times Technology (Beijing) Co., Ltd.), Maurice Pagnucco (UNSW), and Yang Song (University of New South Wales)

ABO: Dataset and Benchmarks for Real-World 3D Object Understanding 21094
Jasmine Collins (UC Berkeley), Shubham Goel (University of California at Berkeley), Kenan Deng (Amazon Inc.), Achleshwar Luthra (BITS Pilani), Leon Xu (UC Berkeley), Erhan Gundogdu (Amazon), Xi Zhang (Amazon Inc.), Tomas F. Yago Vicente (Amazon Inc.), Thomas Dideriksen (Amazon.com), Himanshu Arora (Amazon), Matthieu Guillaumin (Amazon.com), and Jitendra Malik (University of California at Berkeley)

Improving Segmentation of the Inferior Alveolar Nerve Through Deep Label Propagation 21105
Marco Cipriano (Università degli Studi di Modena e Reggio Emilia), Stefano Allegretti (Università degli Studi di Modena e Reggio Emilia), Federico Bolelli (Università degli Studi di Modena e Reggio Emilia), Federico Pollastri (Università degli Studi di Modena e Reggio Emilia), and Costantino Grana (University of Modena and Reggio Emilia)

ZeroWaste Dataset: Towards Deformable Object Segmentation in Cluttered Scenes 21115
Dina Bashkirova (Boston University), Mohamed Abdelfattah (The American University in Cairo (AUC), ,), Ziliang Zhu (Boston University), James Akl (Worcester Polytechnic Institute), Fadi Alladkani (Worcester Polytechnic Institute), Ping Hu (Boston University), Vitaly Ablavsky (University of Washington), Berk Calli (Worcester Polytechnic Institute), Sarah Adel Bargal (Boston University), and Kate Saenko (Boston University)

DynamicEarthNet: Daily Multi-Spectral Satellite Dataset for Semantic Change Segmentation 21126
Aysim Toker (TUM), Lukas Kondmann (German Aerospace Center), Mark Weber (Technical University Munich), Marvin Eisenberger (TU Munich), Andrés Camero (German Aerospace Center), Jingliang Hu (German Aerospace Center), Ariadna Pregel Hoderlein (TUM), Çağlar Şenaras (Planet Labs), Timothy Davis (Planet Labs), Daniel Cremers (TU Munich), Giovanni Marchisio (Planet), Xiao Xiang Zhu (Technical University of Munich (TUM); German Aerospace Center, (DLR),), and Laura Leal-Taixé (TUM)

Open Challenges in Deep Stereo: The Booster Dataset 21136
Pierluigi Zama Ramirez (University of Bologna), Fabio Tosi (University of Bologna), Matteo Poggi (University of Bologna), Samuele Salti (University of Bologna), Stefano Mattoccia (University of Bologna), and Luigi Di Stefano (University of Bologna)

No-Reference Point Cloud Quality Assessment via Domain Adaptation 21147
Qi Yang (Shanghai Jiao Tong University), Yipeng Liu (Shanghai Jiao Tong University), Siheng Chen (Shanghai Jiao Tong University), Yiling Xu (Shanghai Jiao Tong University), and Jun Sun (SJTU)

Exploring Endogenous Shift for Cross-Domain Detection: A Large-Scale Benchmark and Perturbation Suppression Network 21157
Renshuai Tao (Beihang University), Hainan Li (Beihang University), Tianbo Wang (Beihang University), Yanlu Wei (Beihang University), Yifu Ding (Beihang University), Bowei Jin (IFLYTEK (Suzhou) Technology Co., Ltd, ,), Hongping Zhi (iFlytek), Xianglong Liu (BUAA), and Aishan Liu (Beihang University)

How Good Is Aesthetic Ability of a Fashion Model? 21168
Xingxing Zou (Laboratory for Artificial Intelligence in Design, The Hong Kong Polytechnic University), Kaicheng Pang (Hong Kong Polytechnic University), Wen Zhang (Amazon.com), and Waikeung Wong (Institute of Textiles and Clothing, The Hong Kong Polytechnic University)

Instance-Wise Occlusion and Depth Orders in Natural Scenes 21178
Hyunmin Lee (POSTECH) and Jaesik Park (POSTECH)

PhoCaL: A Multi-Modal Dataset for Category-Level Object Pose Estimation With Photometrically Challenging Objects 21190
Pengyuan Wang (TUM), HyunJun Jung (Technical University of Munich), Yitong Li (Technical University of Munich), Siyuan Shen (TUM), Rahul Parthasarathy Srikanth (Technical University of Munich), Lorenzo Garattoni (Toyota-Europe), Sven Meier (Toyota Motor Europe), Nassir Navab (TU Munich, Germany), and Benjamin Busam (Technical University of Munich)

Replacing Labeled Real-Image Datasets With Auto-Generated Contours 21200
Hirokatsu Kataoka (National Institute of Advanced Industrial Science and Technology (AIST), ,), Ryo Hayamizu (National Institute of Advanced Industrial Science and Technology (AIST), ,), Ryosuke Yamada (National Institute of Advanced Industrial Science and Technology (AIST), Tokyo Denki University (TDU),), Kodai Nakashima (National Institute of Advanced Industrial Science and Technology (AIST), ,), Sora Takashima (Tokyo Institute of Technology), Xinyu Zhang (Tokyo Institute of Technology), Edgar Josafat Martinez-Noriega (AIST), Nakamasa Inoue (Tokyo Institute of Technology), and Rio Yokota (Tokyo Institute of Technology, AIST- Tokyo Tech Real World Big-Data Computation Open Innovation Laboratory (RWBC- OIL), National Institute of Advanced Industrial Science and Technology (AIST),)

V2C: Visual Voice Cloning 21210
Qi Chen (University of Adelaide), Mingkui Tan (South China University of Technology), Yuankai Qi (The University of Adelaide), Jiaqiu Zhou (South China University of Technology), Yuanqing Li (South China University of Technology), and Qi Wu (University of Adelaide)

M5Product: Self-Harmonized Contrastive Learning for E-Commercial Multi-Modal Pretraining . 21220
Xiao Dong (Sun Yat-sen University), Xunlin Zhan (Sun Yat-Sen University), Yangxin Wu (Sun Yat-sen University), Yunchao Wei (UTS), Michael C. Kampffmeyer (UiT The Arctic University of Norway), Xiaoyong Wei (Sichuan University), Minlong Lu (Alibaba Group), Yaowei Wang (PengCheng Laboratory), and Xiaodan Liang (Sun Yat-sen University)

It Is Okay To Not Be Okay: Overcoming Emotional Bias in Affective Image Captioning by Contrastive Data Collection 21231
Youssef Mohamed (KAUST), Faizan Farooq Khan (IIIT, Hyderabad), Kilichbek Haydarov (KAUST), and Mohamed Elhoseiny (KAUST)

From Representation to Reasoning: Towards Both Evidence and Commonsense Reasoning for Video Question-Answering 21241
Jiangtong Li (Shanghai Jiao Tong University), Li Niu (Shanghai Jiao Tong University), and Liqing Zhang (Shanghai Jiao Tong University)

Point Cloud Pre-Training With Natural 3D Structures 21251
Ryosuke Yamada (National Institute of Advanced Industrial Science and Technology (AIST), Tokyo Denki University (TDU)), Hirokatsu Kataoka (National Institute of Advanced Industrial Science and Technology (AIST)), Naoya Chiba (Waseda Univ. / Omron Sinic X), Yukiyasu Domae (AIST), and Tetsuya Ogata (Waseda University)

The Auto Arborist Dataset: A Large-Scale Benchmark for Multiview Urban Forest Monitoring Under Domain Shift 21262
Sara Beery (Caltech), Guanhang Wu (Google), Trevor Edwards (Google), Filip Pavetic (Google), Bo Majewski (Google), Shreyasee Mukherjee (Google), Stanley Chan (Google LLC), John Morgan (ManageXR), Vivek Rathod (Google), and Jonathan Huang (Google)

AutoMine: An Unmanned Mine Dataset 21276
Yuchen Li (Waytous; BNU-HKBU UIC; HKBU), Zixuan Li (Waytous), Siyu Teng (missing), Yu Zhang (Sun Yat-sen University), Yuhang Zhou (ZHOU YUHANG), Yuchang Zhu (Sun Yat-sen University), Dongpu Cao (Dongpu Cao), Bin Tian (Institute of Automation, Chinese Academy of Sciences), Yunfeng Ai (University of Chinese Academy of Sciences), Zhe Xuanyuan (Beijing Normal University-Hong Kong Baptist University United International College), and Long Chen (Institute of Automation, Chinese Academy of Science)

SmartPortraits: Depth Powered Handheld Smartphone Dataset of Human Portraits for State Estimation, Reconstruction and Synthesis 21286
Anastasiia Kornilova (Skolkovo Institute of Science and Technology), Marsel Faizullin (Skolkovo Institute of Science and Technology), Konstantin Pakulev (Skoltech), Andrey Sadkov (Skolkovo Institute of Science and Technology), Denis Kukushkin (Skolkovo Institute of Science and Technology), Azat Akhmetyanov (Skolkovo Institute of Science and Technology), Timur Akhtyamov (Skolkovo Institute of Science and Technology), Hekmat Taherinejad (Skolkovo Institute of Science and Technology), and Gonzalo Ferrer (Skolkovo Institute of Science and Technology)

BigDatasetGAN: Synthesizing ImageNet With Pixel-Wise Annotations 21298
Daiqing Li (NVIDIA), Huan Ling (University of Toronto, NVIDIA), Seung Wook Kim (University of Toronto, NVIDIA), Karsten Kreis (NVIDIA), Sanja Fidler (University of Toronto, NVIDIA), and Antonio Torralba (MIT)

Rope3D: The Roadside Perception Dataset for Autonomous Driving and Monocular 3D Object Detection Task 21309
Xiaoqing Ye (baidu), Mao Shu (Baidu Inc), Hanyu Li (Baidu), Yifeng Shi (baidu), Yingying Li (Baidu), Guangjie Wang (Baidu), Xiao Tan (Baidu Inc.), and Errui Ding (Baidu Inc.)

Unifying Panoptic Segmentation for Autonomous Driving 21319
Oliver Zendel (AIT Austrian Institute of Technology), Matthias Schörghuber (AIT Austrian Institute of Technology), Bernhard Rainer (AIT Austrian Institute of Technology), Markus Murschitz (AIT Austrian Institute of Technology), and Csaba Beleznai (Austrian Institute of Technology, Austria)

DAIR-V2X: A Large-Scale Dataset for Vehicle-Infrastructure Cooperative 3D Object Detection.... 21329
Haibao Yu (Tsinghua University), Yizhen Luo (Tsinghua University), Mao Shu (Baidu Inc), Yiyi Huo (University of Chinese Academy of Sciences), Zebang Yang (Tsinghua Unversity), Yifeng Shi (Baidu Inc), Zhenglong Guo (baidu.inc), Hanyu Li (Baidu), Xing Hu (Baidu Inc), Jirui Yuan (Tsinghua University), and Zaiqing Nie (Tsinghua University)

SHIFT: A Synthetic Driving Dataset for Continuous Multi-Task Domain Adaptation 21339
Tao Sun (ETH Zurich), Mattia Segu (ETH Zurich), Janis Postels (ETH Zurich), Yuxuan Wang (ETH Zurich), Luc Van Gool (ETH Zurich), Bernt Schiele (MPI Informatics), Federico Tombari (Google, TU Munich), and Fisher Yu (ETH Zurich)

Ithaca365: Dataset and Driving Perception Under Repeated and Challenging Weather Conditions 21351
Carlos A. Diaz-Ruiz (Cornell University), Youya Xia (Cornell University), Yurong You (Cornell University), Jose Nino (Cornell University), Junan Chen (Cornell University), Josephine Monica (Cornell University), Xiangyu Chen (Cornell University), Katie Luo (Cornell University), Yan Wang (Waymo), Marc Emond (Cornell University), Wei-Lun Chao (The Ohio State University), Bharath Hariharan (Cornell University), Kilian Q. Weinberger (Cornell University), and Mark Campbell (Cornell University)

Demos: CVPR Demos

Scenic: A JAX Library for Computer Vision Research and Beyond 21361
Mostafa Dehghani (Google Brain), Alexey Gritsenko (Google Brain), Anurag Arnab (Google), Matthias Minderer (Google Research), and Yi Tay (Google)

DeepLIIF: An Online Platform for Quantification of Clinical Pathology Slides 21367
Parmida Ghahremani (Stony Brook University), Joseph Marino (Memorial Sloan Kettering Cancer Center), Ricardo Dodds (MSKCC), and Saad Nadeem (Memorial Sloan Kettering Cancer Center)

VL-InterpreT: An Interactive Visualization Tool for Interpreting Vision-Language Transformers 21374
Estelle Aflalo (Intel Corp), Meng Du (UCLA/Intel Labs), Shao-Yen Tseng (Intel), Yongfei Liu (ShanghaiTech), Chenfei Wu (Microsoft), Nan Duan (Microsoft Research), and Vasudev Lal (Intel Corp)

GeoEngine: A Platform for Production-Ready Geospatial Research 21384
Sagar Verma (CentraleSupelec), Siddharth Gupta (Granular AI), Hal Shin (Granular.ai), Akash Panigrahi (Granular AI), Shubham Goswami (Granular Data Inc), Shweta Pardeshi (Granular AI), Natanael Exe (Granular AI), Ujwal Dutta (Granular.ai), Tanka Raj Joshi (Granular.au), and Nitin Bhojwani (Granular)

Talking Face Generation With Multilingual TTS 21393
Hyoung-Kyu Song (MINDsLab Inc.), Sang Hoon Woo (MINDsLab Inc.), Junhyeok Lee (MINDsLab Inc.), Seungmin Yang (MINDsLab Inc.), Hyunjae Cho (MINDsLab Inc.), Youseong Lee (Seoul National University), Dongho Choi (Supertone Inc.), and Kang-wook Kim (MINDsLab Inc.)

Real-Time, Accurate, and Consistent Video Semantic Segmentation via Unsupervised Adaptation and Cross-Unit Deployment on Mobile Device 21399
Hyojin Park (Qualcomm AI Research), Alan Yessenbayev (Qualcomm), Tushar Singhal (Qualcomm AI Research), Navin Kumar Adhikari (Qualcomm Technologies, Inc.), Yizhe Zhang (Nanjing University of Science and Technology), Shubhankar Mangesh Borse (Qualcomm AI Research), Hong Cai (Qualcomm AI Research), Nilesh Prasad Pandey (Qualcomm AI Research), Fei Yin (Qualcomm AI Research), Frank Mayer (Qualcomm AI Research), Balaji Calidas (Qualcomm AI Research), and Fatih Porikli (Qualcomm AI Research)

BigDL 2.0: Seamless Scaling of AI Pipelines From Laptops to Distributed Cluster 21407
Jason Jinquan Dai (Intel), Ding Ding (Intel), Dongjie Shi (Intel), Shengsheng Huang (Intel), Jiao Wang (Intel), Xin Qiu (Intel), Kai Huang (Intel), Guoqiong Song (Intel), Yang Wang (Intel), Qiyuan Gong (Intel), Jiaming Song (Intel), Shan Yu (Intel), Le Zheng (Intel), Yina Chen (Intel), Junwei Deng (Intel), and Ge Song (Intel)

Interactive Segmentation and Visualization for Tiny Objects in Multi-Megapixel Images 21415
Chengyuan Xu (University of California, Santa Barbara), Boning Dong (University of California, Santa Barbara), Noah Stier (University of California, Santa Barbara), Curtis McCully (Las Cumbres Observatory), D. Andrew Howell (Las Cumbres Observatory), Pradeep Sen (UC Santa Barbara), and Tobias Höllerer (UCSB)

A Low-Cost & Real-Time Motion Capture System 21421
Anargyros Chatzitofis (AC CODEWHEEL LTD), Georgios Albanis (University of Thessaly), Nikolaos Zioulis (Independent Researcher), and Spyridon Thermos (AC Codewheel Ltd)

PyMiceTracking: An Open-Source Toolbox for Real-Time Behavioral Neuroscience Experiments 21427
Richardson Menezes (UFRN), Aron de Miranda (UFRN), and Helton Maia (Brain Institute, Federal University of Rio Grande do Norte, Natal)

Effective Conditioned and Composed Image Retrieval Combining CLIP-Based Features 21434
Alberto Baldrati (Università degli Studi di Firenze), Marco Bertini (University of Florence), Tiberio Uricchio (University of Florence), and Alberto Del Bimbo (University of Florence)

VisCUIT: Visual Auditor for Bias in CNN Image Classifier 21443
Seongmin Lee (Georgia Institute of Technology), Zijie J. Wang (Georgia Tech), Judy Hoffman (Georgia Tech), and Duen Horng Polo Chau (Georgia Institute of Technology)

DetectorDetective: Investigating the Effects of Adversarial Examples on Object Detectors 21452
Sivapriya Vellaichamy (Georgia Institute of Technology), Matthew Hull (Georgia Institue of Technology), Zijie J. Wang (Georgia Institute of Technology), Nilaksh Das (Georgia Institute of Technology), ShengYun Peng (Georgia Institute of Technology), Haekyu Park (Georgia Institute of Technology), and Duen Horng Polo Chau (Georgia Institute of Technology)

V-Doc: Visual Questions Answers With Documents 21460
Yihao Ding (USYD), Zhe Huang (The University of Sydney), Runlin Wang (The University of Sydney), YanHang Zhang (The University of Sydney), Xianru Chen (The University of Sydney), Yuzhong Ma (The University of Sydney), Hyunsuk Chung (FortifyEdge), and Soyeon Caren Han (University of Sydney)

Clustering Plotted Data by Image Segmentation 21467
Tarek Naous (American University of Beirut), Srinjay Sarkar (VinAI Research), Abubakar Abid (Stanford), and James Zou (Stanford University)

Author Index