

2024 IEEE International Conference on Multimedia and Expo Workshops (ICMEW 2024)

**Niagara Falls, Ontario, Canada
15-19 July 2024**



**IEEE Catalog Number: CFP24IEW-POD
ISBN: 979-8-3503-7982-2**

**Copyright © 2024 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: | CFP24IEW-POD |
| ISBN (Print-On-Demand): | 979-8-3503-7982-2 |
| ISBN (Online): | 979-8-3503-7981-5 |
| ISSN: | 2330-7927 |

Additional Copies of This Publication Are Available From:

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

CURRAN ASSOCIATES INC.
proceedings
.com

TABLE OF CONTENTS

| | |
|---|----|
| Semi-Supervised Acoustic Scene Classification Under Domain Shift Using an Attention Module and Angular Loss..... | 1 |
| <i>Michael Neri, Marco Carli</i> | |
| 3DMIT: 3D Multi-Modal Instruction Tuning for Scene Understanding..... | 7 |
| <i>Zeju Li, Chao Zhang, Xiaoyan Wang, Ruilong Ren, Yifan Xu, Ruifei Ma, Xiangde Liu, Rong Wei</i> | |
| Compressive Feature Selection for Remote Visual Multi-Task Inference..... | 12 |
| <i>Saeed Ranjbar Alvar, Ivan V. Bajic</i> | |
| Visibility-Aware Human Mesh Recovery Via Balancing Dense Correspondence and Probability Model | 18 |
| <i>Yanjun Wang, Wenjia Wang, Jun Ling, Rong Xie, Li Song</i> | |
| Q-Boost: On Visual Quality Assessment Ability of Low-Level Multi-Modality Foundation Models | 24 |
| <i>Zicheng Zhang, Haoning Wu, Zhongpeng Ji, Chunyi Li, Erli Zhang, Wei Sun, Xiaohong Liu, Xiongkuo Min, Fengyu Sun, Shangling Jui, Weisi Lin, Guangtao Zhai</i> | |
| Compression Without Compromise: Optimizing Point Cloud Object Detection with Bottleneck Architectures for Split Computing | 30 |
| <i>Vinay Kashyap, Nilesh Ahuja, Omesh Tickoo</i> | |
| AFC: Asymmetrical Feature Coding for Multi-Task Machine Intelligence | 36 |
| <i>Yuan Zhang, Hanming Wang, Yunlong Li, Lu Yu</i> | |
| Towards Task-Compatible Compressible Representations..... | 42 |
| <i>Anderson De Andrade, Ivan V. Bajic</i> | |
| Equipped with Monocular Depth Estimation and Intelligent Wake-Up Vision Based Tracking System for a Human-Following Mobile Robot | 48 |
| <i>Tsung-Han Tsai, Chun-Lin Lee</i> | |
| Self-Supervised Learning Via Multi-Transformation Classification for Action Recognition..... | 50 |
| <i>Duc-Quang Vu, Ngan Le, Jia-Ching Wang</i> | |
| Chinese Ancient Painting Figure Face Restoration and Its Application in a Q&A Interaction System..... | 56 |
| <i>Rui Li, Yifan Wei, Haopeng Lu, Siwei Ma, Zhenyu Liu, Hui Liu, Qianying Wang, Yaqiang Wu, Jianrong Tan</i> | |
| An SoC Based Hardware Accelerator for Blind Assistive System..... | 62 |
| <i>Tsung-Han Tsai, Chun-Yu Chen</i> | |
| LFCAVE: Interactive 3D Space with Multiple Light Field Displays | 64 |
| <i>Haopeng Lu, Wenkang Shan, Yuhuai Zhang, Li Song, Xinfeng Zhang, Siwei Ma, Liuxin Zhang, Wen Gao</i> | |
| An End-To-End Channel-Adaptive Feature Compression Approach in Device-Edge Co-Inference Systems..... | 66 |
| <i>Yuan Ouyang, Ping Wang, Lijun He, Fan Li</i> | |
| Joint Modal Circular Complementary Attention for Multimodal Aspect-Based Sentiment Analysis..... | 72 |
| <i>Hao Liu, Lijun He, Jiayi Liang</i> | |

| | |
|---|-----|
| Learning Discriminative and Robust Representations for UAV-View Skeleton-Based Action Recognition | 78 |
| <i>Shaofan Sun, Jiahang Zhang, Guo Tang, Chuanmin Jia, Jiaying Liu</i> | |
| Rate Control Optimizing Model for Constraining Over-Saturated Live Streaming Quality | 84 |
| <i>Huiwen Ren, Zhao Wang, Jiexi Wang, Yuwen He, Siwei Ma, Li Zhang, Wen Gao</i> | |
| Lightweight Texture-Guided Fast Partition Method for Luma and Chroma Intra Coding in VVC | 90 |
| <i>Zhikai Liu, Zhidao Zhou, Fan Liang, Wei Sun</i> | |
| Automatic Malleefowl Mound Detection Using LiDAR-Based Ground and Habitat Features with Planar Terrain Modelling | 96 |
| <i>Nazia Hossain, Manzur Murshed, Mohammad Awrangjeb, Singarayer Florentine, Marc Irvin, Shyh Wei Teng</i> | |
| Multimodal Semantic-Aware Automatic Colorization with Diffusion Prior | 102 |
| <i>Han Wang, Xinning Chai, Yiwu Wang, Yuhong Zhang, Rong Xie, Li Song</i> | |
| LLM-SAP: Large Language Models Situational Awareness-Based Planning | 108 |
| <i>Liman Wang, Hanyang Zhong</i> | |
| Real-Time Human Motion Transfer System for Holographic Displays | 114 |
| <i>Wenkang Shan, Haopeng Lu, Chuanmin Jia, Xinfeng Zhang, Siwei Ma, Yaqiang Wu, Wen Gao</i> | |
| Creating and Experiencing 3D Immersion Using Generative 2D Diffusion: An Integrated Framework | 116 |
| <i>Ziming He, Xiaomin Zou, Pengfei Wu, Ling Fan, Xiaomei Li</i> | |
| Real-Time Interaction with Animated Human Figures in Chinese Ancient Paintings | 122 |
| <i>Yifan Wei, Wenkang Shan, Qi Zhang, Liuxin Zhang, Jian Zhang, Siwei Ma</i> | |
| StyleSelf: Style-Controllable High-Fidelity Conversational Virtual Avatars Generation | 128 |
| <i>Yilin Guo, Ruoke Yan, Yaqiang Wu, Siwei Ma</i> | |
| Impact of Prioritized HTTP/3 Transport on Low-Latency Live Streaming | 134 |
| <i>Ayşe B. Demir, Mervegul Parlak, Zafer Gurel, Deniz Ugur, Ali C. Beğen</i> | |
| Blender-NeRF: A Monocular Dynamic Human Body Explicit Reconstruction and Rendering Method | 140 |
| <i>Shuo Chen, Wu Liu, Binbin Yan, Xinzhu Sang, Alicia Li, Xiangcheng Yi</i> | |
| Region-Of-Interest-Based Video Coding for Machines | 146 |
| <i>Olgierd Stankiewicz, Tomasz Grajek, Sławomir Mackowiak, Jakub Stankowski, Sławomir Rózek, Mateusz Lorkiewicz, Maciej Wawrzyniak, Marek Domanski</i> | |
| Optimizing Quality and Energy Efficiency in WebRTC with ML-Powered Adaptive FEC | 152 |
| <i>Jason Gerard, David C. Bonilla, Abdelhak Bentaleb, Sandra Céspedes</i> | |
| Dual Attribute-Spatial Relation Alignment for 3D Visual Grounding | 158 |
| <i>Yue Xu, Kaizhi Yang, Kai Cheng, Jiebo Luo, Xuejin Chen</i> | |
| Improving Acoustic Scene Classification Via Self-Supervised and Semi-Supervised Learning with Efficient Audio Transformer | 164 |
| <i>Yuzhe Liang, Wenxi Chen, Anbai Jiang, Yihong Qiu, Xinhui Zheng, Wen Huang, Bing Han, Yanmin Qian, Pingyi Fan, Wei-Qiang Zhang, Cheng Lu, Jia Liu, Xie Chen</i> | |
| Decoupling Classification and Localization of CLIP | 170 |
| <i>Muyang Yi, Zhaozhi Xie, Yuwen Yang, Chang Liu, Yue Ding, Hongtao Lu</i> | |

| | |
|---|-----|
| NeuProofreader: An Interactive Proofreading System with Suggestive Prompts for Connectomics | 176 |
| <i>Yixiong Liu, Qihua Chen, Xuejin Chen</i> | |
| Language-Guided Zero-Shot Object Counting..... | 178 |
| <i>Mingjie Wang, Song Yuan, Zhuohang Li, Longlong Zhu, Eric Buys, Minglun Gong</i> | |
| Low-Complexity Video PSNR Measurement in Real-Time Communication Products | 184 |
| <i>Yu-Chen Sun, Jie Dong, Ahmed Fouad, Jian Zhou, Roger Zhou, Shyam Sadhwani</i> | |
| Beyond Aligned Target Face: StyleGAN-Based Face-Swapping Via Inverted Identity Learning..... | 188 |
| <i>Yuanhang Li, Qi Mao, Libiao Jin</i> | |
| Multimodal Guidance Network for Missing-Modality Inference in Content Moderation..... | 194 |
| <i>Zhuokai Zhao, Harish Palani, Tianyi Liu, Lena Evans, Ruth Toner</i> | |
| AI-Assisted Content Creation of Naked-Eye 3D Effects on Curved LED Screen: Enhancing Artistic Expression and Creativity..... | 198 |
| <i>Yeming Li, Junrong Song, David Keiman Yip</i> | |
| I3FNET: Instance-Aware Feature Fusion for Few-Shot Point Cloud Generation from Single Image..... | 203 |
| <i>Pu Ching, Wen-Cheng Chen, Min-Chun Hu</i> | |
| Optimizing an Open VVC Encoder for Low Delay Remote Desktop Applications..... | 209 |
| <i>Anastasia Henkel, Benjamin Bross, Jens Brandenburg, Adam Wieckowski, Detlev Marpe, Andoni Morales, Sergio Sanchez</i> | |
| Characteristics of Visual Complexity: Calligraphic Fonts Vs. Printed Fonts | 215 |
| <i>Yuchen Wang, Ruimin Lyu</i> | |
| Visual-Language Alignment for Background Subtraction..... | 221 |
| <i>Jiahe Liu, Dandan Zhu, Sajid Javed</i> | |
| Adaptive Intra Period Size for Deep Learning-Based Screen Content Video Coding..... | 228 |
| <i>Yuyang Wu, Liang Xie, Shangkun Sun, Wei Gao, Yiqiang Yan</i> | |
| Aesthetic Assessment of Movie Still Frame for Various Field of Views | 234 |
| <i>Xin Jin, Jinyu Wang, Wenbo Yuan, Yihang Bo, Heng Huang, Yiran Zhang, Bao Peng, Peng Xu, Xin Song, Hanbing Yang</i> | |
| Leveraging Multimodal Knowledge for Spatio-Temporal Action Localization | 240 |
| <i>Keke Chen, Zhewei Tu, Xiangbo Shu</i> | |
| Optimizing Facial Landmark Estimation for Embedded Systems Through Iterative Autolabeling and Model Pruning | 245 |
| <i>Yu-Hsi Chen, I-Hsuan Tai</i> | |
| Efficient Facial Landmark Detection for Embedded Systems..... | 251 |
| <i>Ji-Jia Wu</i> | |
| Semi-Supervised Acoustic Scene Classification Under Domain Shift with MixMatch and Information Bottleneck Optimization..... | 257 |
| <i>Yongpeng Yan, Wuyang Liu, Yi Chai, Yanzhen Ren</i> | |
| PartCLIP: How Does CLIP Assist Mechanical Part Image Retrieval?..... | 261 |
| <i>Shangbo Mao, Dongyun Lin, Aiyuan Guo, Yiqun Li</i> | |

| | |
|---|-----|
| Attribute-Aware Network for Pedestrian Attribute Recognition | 266 |
| <i>Zesen Wu, Mang Ye, Shuoyi Chen, Bo Du</i> | |
| Enhancing Visual Wake Word Spotting with Pretrained Model and Feature Balance Scaling | 272 |
| <i>Xuandong Huang, Shangfei Wang, Jinghao Yan, Kai Tang, Pengfei Hu</i> | |
| SemiPL: A Semi-Supervised Method for Event Sound Source Localization | 278 |
| <i>Yue Li, Baiqiao Yin, Jinfu Liu, Jiajun Wen, Jiaying Lin, Mengyuan Liu</i> | |
| HDBN: A Novel Hybrid Dual-Branch Network for Robust Skeleton-Based Action Recognition | 284 |
| <i>Jinfu Liu, Baiqiao Yin, Jiaying Lin, Jiajun Wen, Yue Li, Mengyuan Liu</i> | |
| VidBot: Intelligent Video Learning Tool for Content Mining and Playback Traffic Statistics | 290 |
| <i>Qinhua Xie, Weicong Liu, Fan Yuan, Jifan Shi, Ziyu Liu, Yanbing Zhang</i> | |
| Summary on the Chat-Scenario Chinese Lipreading (ChatCLR) Challenge..... | 293 |
| <i>Chen-Yue Zhang, Hang Chen, Jun Du, Sabato Marco Siniscalchi, Ya Jiang, Chin-Hui Lee</i> | |
| An Intra- And Inter-Frame Sequence Model with Discrete Cosine Transform for Streaming Speech Enhancement | 299 |
| <i>Yuewei Zhang, Huanbin Zou, Jie Zhu</i> | |
| Body-Part Guided Animal Pose Estimation | 303 |
| <i>Jiyong Rao, Tianyang Xu, Xiaoning Song, Zhenhua Feng, Xiao-Jun Wu</i> | |
| A Survey on Backbones for Deep Video Action Recognition | 309 |
| <i>Zixuan Tang, Youjun Zhao, Yuhang Wen, Mengyuan Liu</i> | |
| Intelligent Music Chord Recognition and Evaluation Based on Convolution and Attention | 315 |
| <i>Shuo Wang, Xiaobing Li, Qingwen Zhou, Yun Tie, Yan Gao, Xinran Zhang</i> | |
| SFMViT: Slowfast Meet ViT in Chaotic World..... | 321 |
| <i>Jiaying Lin, Jiajun Wen, Mengyuan Liu, Yue Li, Jinfu Liu, Baiqiao Yin</i> | |
| Anatomically-Informed Vector Quantization Variational Auto-Encoder for Text to Motion Generation | 327 |
| <i>Lian Chen, Zehai Niu, Qingyuan Liu, Jinbao Wang, Jian Xue, Ke Lu</i> | |
| Enhancing Lip Reading with Multi-Scale Video and Multi-Encoder | 333 |
| <i>He Wang, Pengcheng Guo, Xucheng Wan, Huan Zhou, Lei Xie</i> | |
| “AI Life” and Human Fear: From Phenomenological Insights to Digital Creation..... | 339 |
| <i>Jiaying Fu, Tianyue Gong, Jialin Gu, Tiange Zhou</i> | |
| Popular Hooks: A Multimodal Dataset of Musical Hooks for Music Understanding and Generation | 345 |
| <i>Xinda Wu, Jiaming Wang, Jiaying Yu, Tiejiao Zhang, Kejun Zhang</i> | |
| The WuShu Database for Cursive Script Character and Style Recognition | 351 |
| <i>Xinrui Shan, Kejun Zhang, Lyukesheng Shen, Bolin Wang</i> | |
| Unveiling Soil-Vegetation Interactions: Reflection Relationships and an Attention-Based Deep Learning Approach for Carbon Estimation..... | 357 |
| <i>Dristi Datta, Manoranjan Paul, Manzur Murshed, Shyh Wei Teng, Leigh M. Schmidtke</i> | |
| A Hybrid Multi-Perspective Complementary Model for Human Skeleton-Based Action Recognition | 363 |
| <i>Linze Li, Youwei Zhou, Jiannan Hu, Cong Wu, Tianyang Xu, Xiao-Jun Wu</i> | |

| | |
|---|-----|
| Enhancing Video Grounding with Dual-Path Modality Fusion on Animal Kingdom Datasets..... | 369 |
| <i>Chengpeng Xiong, Zhengxuan Chen, Nuoer Long, Kin-Seong Un, Zhuolin Li, Shaobin Chen, Tao Tan, Chan-Tong Lam, Yue Sun</i> | |
| A Multimodal Behavior Recognition Network with Interconnected Architectures..... | 375 |
| <i>Nuoer Long, Kin-Seong Un, Chengpeng Xiong, Zhuolin Li, Shaobin Chen, Tao Tan, Chan-Tong Lam, Yue Sun</i> | |
| Segmentation-Based Parametric Painting..... | 381 |
| <i>Manuel Ladron De Guevara, Matt Fisher, Aaron Hertzmann</i> | |
| MemoMusic 4.0: Personalized Emotion Music Generation Conditioned by Valence and Arousal as Virtual Tokens | 387 |
| <i>Luntian Mou, Yihan Sun, Yunhan Tian, Ruichen He, Feng Gao, Zijin Li, Ramesh Jain</i> | |
| A Micro-Expression Recognition System with Event Cameras | 393 |
| <i>Peilin Xiao, Yueyi Zhang, Dachun Kai, Yansong Peng, Zheyu Zhang, Xiaoyan Sun</i> | |
| Robust Person Re-Identification Approach with Deep Learning and Optimized Feature Extraction..... | 395 |
| <i>Jian Ding, Linze Li, Rongchang Li, Cong Wu, Tianyang Xu, Xiao-Jun Wu</i> | |
| Summary of the 2024 Low-Power Efficient and Accurate Facial-Landmark Detection for Embedded Systems..... | 401 |
| <i>Yu-Shu Ni, Han-Chun Chen, Chia-Chi Tsai, Chih-Cheng Chen, Po-Yu Chen, Hsien-Kai Kuo, Jun-Ying Hunag, Po-Chi Hu, Jenq-Neng Hwang, Jiun-In Guo</i> | |
| AJA-Pose: A Framework for Animal Pose Estimation Based on VHR Network Architecture..... | 407 |
| <i>Austin Kaburia Kibaara, Joan Kabura, Antony Gitau, Ciira Maina</i> | |
| MtYOLO: A Multi-Task Model to Concurrently Obtain the Vital Characteristics of Individuals Or Animals | 413 |
| <i>Kian Eng Ong, Sivaji Retta, Ramarajulu Srinivasan, Shawn Tan, Jun Liu</i> | |
| Using Large Language Models to Understand Leadership Perception and Expectation..... | 417 |
| <i>Yundi Zhang, Xin Wang, Ziyi Zhang, Xueying Wang, Xiaohan Ma, Yingying Wu, Han-Wu-Shuang Bao, Xiyang Zhang</i> | |
| The NERCSLIP-USTC System for Semi-Supervised Acoustic Scene Classification of ICME 2024 Grand Challenge..... | 424 |
| <i>Qing Wang, Guirui Zhong, Hengyi Hong, Lei Wang, Mingqi Cai, Xin Fang, Ya Jiang, Jun Du</i> | |
| High-Fidelity 3D Model Generation with Relightable Appearance from Single Freehand Sketches and Text Guidance..... | 428 |
| <i>Tianrun Chen, Runlong Cao, Ankang Lu, Tao Xu, Xiaoling Zhang, Mao Papa, Ming Zhang, Lingyun Sun, Ying Zang</i> | |
| Learning to Learn Multiview Detection by Camera-Aware Attention..... | 434 |
| <i>Hung-Min Hsu, Zhongwei Cheng, Xinyu Yuan, Lin Chen</i> | |
| An Enhanced Multimodal Negative Feedback Detection Framework with Target Retrieval in Thai Spoken Audio | 438 |
| <i>Pantid Chantangphol, Sattaya Singkul, Thanawat Lodkaew, Nattasit Maharattanamalai, Atthakorn Petchsod, Theerat Sakdejayont, Tawunrat Chalothorn</i> | |
| Semi-Supervised Acoustic Scene Classification with Test-Time Adaptation | 445 |
| <i>Wen Huang, Anbai Jiang, Bing Han, Xinhui Zheng, Yihong Qiu, Wenxi Chen, Yuzhe Liang, Pingyi Fan, Wei-Qiang Zhang, Cheng Lu, Xie Chen, Jia Liu, Yanmin Qian</i> | |

| | |
|--|-----|
| Attribute Vision Transformer for UAV-Human Re-Identification | 450 |
| <i>Hao Ni, Yuke Li, Ping Lai, Pengpeng Zeng, Hangyu Guo, Lianli Gao</i> | |
| Pedestrian Attributes Recognition for UAV-Human..... | 456 |
| <i>Hao Ni, Ping Lai, Yuke Li, Pengpeng Zeng, Haonan Zhang, Jingkuan Song</i> | |
| Assistant Referee System in Da-Qiang(Pike) Competition..... | 461 |
| <i>Chia-Chun Yen, Show-Po Guo, Tsi-Uí Ik</i> | |
| Dual-Phase MSQNET for Species-Specific Animal Activity Recognition | 469 |
| <i>An Yu, Jeremy Varghese, Ferhat Demirkiran, Peter Buonaiuto, Xin Li, Ming-Ching Chang</i> | |
| Exploring Semi-Supervised, Subcategory Classification and Subwords Alignment for Visual Wake Word Spotting..... | 475 |
| <i>Shifu Xiong, Lirong Dai</i> | |
| Multi-Modal Knowledge Transfer for Target Speaker Lipreading with Improved Audio-Visual Pretraining and Cross-Lingual Fine-Tuning | 481 |
| <i>Genshun Wan, Zhongfu Ye</i> | |
| The WHU Wake Word Lipreading System for the 2024 Chat-Scenario Chinese Lipreading Challenge..... | 487 |
| <i>Haoxu Wang, Cancan Li, Fei Su, Juan Liu, Hongbin Suo, Ming Li</i> | |

Author Index