ACM MULTIMEDIA ASIA 2025

Grand Millennium Hotel

Kuala Lumpur, Malaysia

9th December ~ 12th December

Program at a Glance

Time	9/12/2025 (Tuesday)			
	Ballroom	Millennium III	Millennium IV, V & VI	Foyer
9.30-11.00	**WS4**: Affective and Value-Centered Evaluation in Multimedia UX	**Grand Challenge 2:** Multimodal Multiethnic Talking-Head Video Generation		
11.00-11.30	Coffee			
11.30-13.00	**WS4**: Affective and Value-Centered Evaluation in Multimedia UX	**Grand Challenge 1:** I3: Image-Text Interleaved Instruction Following Challenge	**WS2**: MuReC '25 Workshop	
13.00-14.00	Lunch			
14.00-15.30		**WS3**: Visual and Signal Communication Technologies in Design of Housing, Urban Spaces, Local Communities, and Human Behavior	**WS1**: HD-Vision Workshop	**WS1**: HD-Vision Workshop Poster Session
15.30-16.00	Coffee			
16.00-17.30		**WS3**: Visual and Signal Communication Technologies in Design of Housing, Urban Spaces, Local Communities, and Human Behavior	**WS1**: HD-Vision Workshop	
17.30-18.30	Setup Time for Reception			
18.30-20.00	Welcome Reception			

Time	10/12/2025 (Wednesday)			
	Ballroom	Millennium III	Millennium IV, V & VI	Foyer
8.30-9.00	Opening Ceremony			
9.00-9.30	**Keynote 1: Prof Liu Zhiyuan** Towards Artificial General Intelligence: Key Challenges and Trends			
9.30-10.00				
10.00-10.30	Industry Program	**Oral 1**: Video streaming and compression	**Oral 2**: Deepfake detection	
10.30-11.00				**Poster 1:** Beyond image
11.00-11.30	Coffee			
11.30-12.00	Industry Program	**Oral 3:** Medical language models and tools	**Oral 4:** Federated learning	
12.00-12.30				
12.30-13.00	Lunch			
13.00-13.30				
13.30-14.00				**Poster 2:** Cross-media analysis
14.00-14.30	Industry Program			
14.30-15.00				
15.00-15.30				
15.30-16.00	Coffee			
16.00-16.30	Industry Program	**Oral 5:** Diffusion models	**Oral 6:** Social Emotion	**Poster 3:** 3D and video
16.30-17.00				
17.00-17.30				
17.30-18.00				
18.00-18.30				

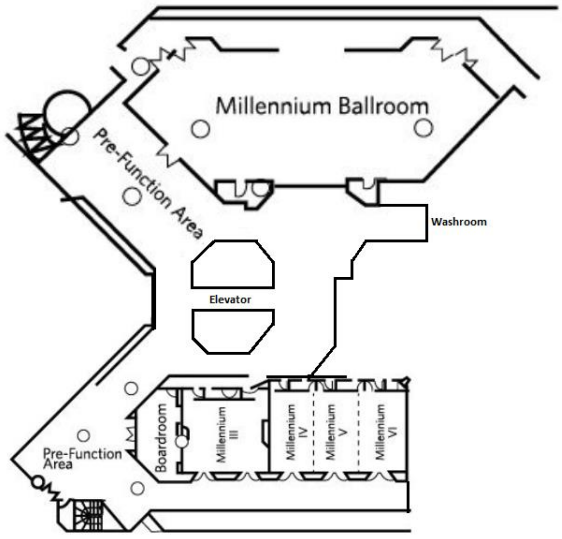
Time	11/12/2025 (Thursday)			
	Ballroom	Millennium III	Millennium IV, V & VI	Foyer
9.00-9.30	**Keynote 2: Prof Shin'ichi Satoh** Psychiatric Disease Diagnosis: The Challenge of Multimedia Technologies			
9.30-10.00				
10.00-10.30	**Best Paper Oral**			
10.30-11.00				**Poster 4:** Vertical domains
11.00-11.30	Coffee			
11.30-12.00	**Oral 7:** Captioning	**Special Session:** Privacy-Aware Multimodal Analysis and Its Applications	**Special Session:** Advances in Multimedia Security: Ensuring Trust and Integrity in the AI Era	
12.00-12.30				
12.30-13.00	Lunch			
13.00-13.30				
13.30-14.00				**Poster 5:** Multimodal fusion
14.00-14.30	**Panel:** Shaping the Future of Multimedia: Women Inspiring the Next Wave of AI Innovation	**Special Session:** Cultural Heritage Analysis and Recommendation	**Oral 8:** Content generation	
14.30-15.00				
15.00-15.30				
15.30-16.00	Coffee			
16.00-16.30	**Oral 9:** Classification and Segmentation	**Oral 10:** Multimodality	**Oral 11:** Sound, sign and text	**Poster 6:** Intelligent multimedia
16.30-17.00				
17.00-17.30				
17.30-18.00	Travel to Restaurant			
18.00-20.00	Banquet Dinner			

Time	12/12/2025 (Friday)			
	Ballroom	Millennium III	Millennium IV, V & VI	Foyer
9.00-9.30	**Keynote 3: Prof. Cathal Gurrin** Learnings From a Decade of Research into Lifelog Organisation and Retrieval			
9.30-10.00				
10.00-10.30	**Oral 12:** All about light!	**Oral 13:** Sketch and search	**Oral 14:** Emerging topics	
10.30-11.00				**Poster 7:** Everyday multimedia + **Short Paper Posters**
11.00-11.30	Coffee			
11.30-12.00	**Brave New Idea**	**Doctoral Consortium**	**Demo**	
12.00-12.30				
12.30-13.00	Closing Ceremony			
13.00-13.30	Lunch			
13.30-14.00				
14.00-14.30	Half-day Tour			
14.30-15.00				
15.00-15.30				
15.30-16.00				
16.00-16.30				
16.30-17.00				
17.00-17.30				
17.30-18.00				

Venue Map

Grand Millennium Hotel KL

Level 2



Technical Program

Tuesday (9th December)

9.30 ~ 11.00	**Workshop 4:** Affective and Value-Centered Evaluation in Multimedia UX **Grand Challenge 2:** Multimodal Multiethnic Talking-Head Video Generation
11.00 ~ 11.30	Coffee Break
11.30 ~ 13.00	**Workshop 2:** Multimodal Representation Learning and Clustering (MuReC '25) **Workshop 4:** Affective and Value-Centered Evaluation in Multimedia UX **Grand Challenge 1:** I3: Image-Text Interleaved Instruction Following Challenge
13.00 ~ 14.00	Lunch
14.00 ~ 15.30	**Workshop 1:** International Workshop on Imaging, Processing, Perception, and Reasoning for High-Dimensional Visual Data (HD-Vision) **Workshop 3:** Visual and Signal Communication Technologies in Design of Housing, Urban Spaces, Local Communities, and Human Behavior
15.30 ~ 16.00	Coffee Break
15.30 ~ 16.00	**Workshop 1:** International Workshop on Imaging, Processing, Perception, and Reasoning for High-Dimensional Visual Data (HD-Vision) **Workshop 3:** Visual and Signal Communication Technologies in Design of Housing, Urban Spaces, Local Communities, and Human Behavior
18.30 ~ 20.00	Welcome Reception

Wednesday (10th December)

8.30 ~ 9.00	Opening	
9.00 ~ 10.00	**Keynote 1:** Prof. Dr. Liu Zhiyuan, *Tsinghua University* Title: Towards Artificial General Intelligence: Key Challenges and Trends Chair: Tat-Seng Chua, *National University of Singapore*	
10.00 ~ 11.00	**Oral 1:** Video Streaming and compression Session Chair: Ding Caiwen, *University of Minnesota - Twin Cities*	
	69	Joint Optimization for Image Compression and Deblurring via Blur-Aware Guidance. *Zunian Wan (The University of Tokyo)*; Jiancheng Zhao (The University of Tokyo); Yepeng Ding (Hiroshima University); Jinfeng Guan (The University of Tokyo); Lingfeng Zhang (The University of Tokyo); Takefumi Ogawa (The University of Tokyo)*
	322	Leveraging Cross-Directional Dependency in Realtime Interactive Streaming. *Sen Lin (Northwestern University)*; Andre Chen (Northwestern University); Kevin Zhikai Chen (Northwestern University); Aleksandar Kuzmanovic (Northwestern University)*
	505	UniStream: Unifying In-Network Video Processing and Caching for Cost-Optimal Edge-Cloud Video Streaming. *Luting Cao (Nanjing University of Science and Technology); Guanyu Gao (Nanjing University of Science and Technology)**
10.00 ~ 11.00	**Oral 2:** Deepfake detection Session Chair: Mei Kuan Lim, *Monash University, Malaysia*	
	103	Time Step Generating: A Universal Synthesized Deepfake Image Detector. *Ziyue Zeng (Waseda University)*; Yupei Guo (Tokyo University of Science); Haoyuan Liu (Waseda University); Dingjie Peng (Waseda University); Hiroshi Watanabe (Waseda University)*
	221	Frequency-Enhanced Multi-Modal Consistency Learning for Audio-Visual Deepfake Detection. *Xinyu Zheng (Tongji University)*; Dongdong Zhang (Tongji University); Chengyu Sun (Tongji University)*

	290	Dynamic Routing between Multimodal Capsules for Deepfake Image Editing Detection. *Tuan Nguyen (Hamad Bin Khalifa University)*; Naseem Khan (Hamad Bin Khalifa University); Issa Khalil (Hamad Bin Khalifa University)*
	390	Adaptive Inter-Modality Attention for Enhanced Cross-Domain Deepfake Detection Transferability. *Naseem Khan (Hamad Bin Khalifa University)*; Nguyen Vu Tuan (Qatar Computing Research Institute); Issa Khalil (Qatar Computing Research Institute)*
10.30 ~ 12.30	**Poster 1:** Beyond Image Session Chair: Trung Thanh Nguyen, *Nagoya University*	
	219	Phoneme-Controlled LLM with Self-Supervised Speech Prompts for Mispronunciation Detection. *Zhengping Song (Xinjiang University); Zadeer Kadeer (Xinjiang University)*; Mulati Kahaer (Xinjiang University); XuDong Pang (Xinjiang University); Yinfeng Yu (Xinjiang University); Aishan Wumaier (Xinjiang University)*
	275	Zero-Shot Speech Recognition from Text-Only Data through Synthesized Spectrogram Refinement Using Style Truncation and Contextual Alignment Loss. *Yuan Li (Inner Mongolia University)*; Yonghe Wang (Inner Mongolia University); Zhenjie Gao (Inner Mongolia University); Feilong Bao (Inner Mongolia University)*
	323	SynSpeech: A Dataset and Benchmark for Fake Speech Detection. *Qifeng Qiu (Guangdong University of Technology); Yutian Li (Guangdong University of Technology); Lap-Kei Lee (Hong Kong Metropolitan University); Fu Lee Wang (Hong Kong Metropolitan University); Zhenguo Yang (Guangdong University of Technology)**
	348	FabasedVC: Enhancing Voice Conversion with Text Modality Fusion and Phoneme-Level SSL Features *Wenyu Wang (Xi'an Jiaotong University)*; Zhetao Hu (Xi'an Jiaotong University); Yiquan Zhou (Xi'an Jiaotong University); Jiacheng Xu (East China Normal University); Zhiyu Wu (Fudan University); Chen Li (Xi'an Jiaotong University); Shihao Li (Union Wheatland Culture and Media Ltd.)*
	356	CDFN: Credibility-Driven Fusion Network for Multimodal Sentiment Analysis. *Chengyu Liu (Shanghai DianJi University); Chengguang Liu (Nanjing University of Information Science and Technology); Xiqing Wei (Shanghai DianJi University)**

	415	Multimodal Sentiment Analysis with Contrastive Constraint Propagation. *Ming Zhang (Wuhan Institute Of Technology); Guoping Wang (Wuhan Institute Of Technology)**
11.00 ~ 11.30	Coffee Break	
11.30 ~ 12.30	**Oral 3:** Medical language models and tools Session chair: Iman Yi Liao, *University of Nottingham Malaysia*	
	91	Extending Pretrained Diffusion Models for Medical Image–Label Generation. *Yuenan Sun (Wuhan University of Science and Technology); Zhida Feng (Wuhan University of Science and Technology)*; Chen Li (Wuhan University of Science and Technology)*
	184	Inter-Patient Arrhythmia Classification via Cloud-Edge-End Multimedia Architecture. *Jia Luo (Yunnan University)*; Yi Cheng (Yunnan University); Cheng Xie (Yunnan University)*
	422	DRF-Net: An Asymmetric Lightweight Network with a Dynamic Routing Fusion Bridge for Medical Image Segmentation. *Wei Sun (Chongqing Normal University)*; Yanmin Niu (Chongqing Normal University)*
	438	Position-Enhanced Gradient Attack (PEGA) on Medical Language Models. *Nuo Xu (University of Minnesota Twin Cities)*; Christopher Stanley (, Oak Ridge National Laboratory); John Gounley (Oak Ridge National Laboratory); Heidi Hanson Oak Ridge National Laboratory); Chang Ge (University of Minnesota Twin Cities); Caiwen Ding (University of Minnesota Twin Cities)*
11.30 ~ 12.30	**Oral 4:** Federated Learning Session Chair: Zhang Tianyun, *Cleveland State University*	
	112	One-shot Multimodal Federated Learning via Diverse Synthetic Feature Optimization. *Fan Qi (Tianjin University of Technology)*; Ziqi Zhao (Tianjin University of Technology); Zixin Zhang (Tianjin University of Technology); Huaiwen Zhang (Inner Mongolia University)*

	119	FedART: Enhancing Replay in Federated Incremental Learnin. *Zijiang Tan (Beijing Normal University)*; Haodi Wang (City University of Hong Kong); Libin Jiao (China University of Mining and Technology-Beijing); Rongfang Bie (Beijing Normal University)*
	446	Task-Aware Federated Multi-Task Learning. *Lei Li (Cleveland State University); Haochen Yang (Cleveland State University); Jiacheng Guo (Cleveland State University); Hongkai Yu (Cleveland State University); Minghai Qin (Western Digital Corporation); Tianyun Zhang (Cleveland State University)**
	470	FCLHet: Spatiotemporal Knowledge Continual Learning for Federated Heterogeneous Models. Jinke Zhou (Nanjing University of Science and Technology); *Guanyu Gao (Nanjing University of Science and Technology)**
12.30 ~ 14.00	Lunch	
13.30 ~ 15.30	**Poster 2:** Cross-media analysis Session Chair: Zhihe Lu, *Hamad Bin Khalifa University*	
	13	Robust Dual Embedding Contrastive Learning for Text-to-Image Person Re-identification with Noisy Correspondence. *Jingjie Zhang (Wuhan University of Technology); Lingli Tang (Wuhan University of Technology); Jiachen Li (Wuhan University of Technology); Jinyu Xu (Wuhan University of Technology); Yanchun Ma (Wuhan Vocational College of Software and Engineering)*; Qing Xie (Wuhan University of Technology)*
	14	Dual-Space Adaptive Fusion for Self-supervised Text-guided Image Editing. *Qingyang Liu (Shanghai Jiao Tong University)*; Haonan Zhao (Shanghai Jiao Tong University); Li Niu (Shanghai Jiao Tong University)*
	45	GAOT: Generating Articulated Objects Through Text-Guided Diffusion Models. *Hao Sun (Harbin Institute of Technology)*; Fan Lei (University of New South Wales); Donglin Di (Li Auto); Shaohui Liu (Harbin Institute of Technology)*
	89	An Interactive Drawing Assistant Harmonizing User Control and Creative Freedom in Image Generation. *Jingyi Chen (Durham University); Zixin Zhang (Shanghai Jiao Tong University)**

	260	A Visual Speech Language Model for Visual Text-to-Speech Task. *Yuyue Wang (Renmin University of China)*; Xin Cheng (Renmin University of China); Yihan Wu (Renmin University of China); Xihua Wang (Renmin University of China); Jinchuan Tian (Carnegie Mellon University); Ruihua Song (Renmin University of China)*
	418	Bi-affine Semantic Fusion Generative Adversarial Networks for Text-to-Image Synthesis. *Zhiqiang Zhang (Southwest University of Science and Technology); Wenxin Yu (Southwest University of Science and Technology); Xin Cheng (Southwest University of Science and Technology)*; Yunan Zhang (Southwest University of Science and Technology)*
15.30 ~ 16.00	Coffee Break	
15.30 ~ 17.30	**Poster 3:** 3D and video Session Chair: Zhihe Lu, *Hamad Bin Khalifa University*	
	92	IFFN: Irrelevant Frames Filtering Network for Online Action Detection. *Ming-Xuan Lin (Huaqiao University)*; Hong-Bo Zhang (Huaqiao University); BoSheng Zheng (Huaqiao University); JingHua Liu (Huaqiao University); ZhenZhen Sun (Huaqiao University); Qing Lei (Huaqiao University)*
	110	In-Sensor Real-Time 3D Surface Reconstruction via Parallel Morphological Computing Using Linear Laser. *Zheng Hu (Shanghai University); Shuhuan Hao (Shanghai University); Jiajian Zheng (Shanghai University); Huixin Zhong (Xi'an Jiaotong-Liverpool University); Yanan Liu (Shanghai University)**
	144	CPR-CLIP: Cross-modal Consistent and Prompt-diverse Regularized CLIP for Action Recognition. *Hao Song (Wuhan Textile University); Yangjun Ou (Wuhan Textile University)*; Chen Wang (Wuhan Textile University)*
	304	Automated Fine-Scale Change Detection Using 3D Gaussian Splatting and VLMs *Satoshi Date (Mitsui Sumitomo Insurance Co., Ltd.); Shojiro Tsutsui (Mitsui Sumitomo Insurance Co., Ltd.); Israel Mendonça (Kumamoto University); Masayoshi Aritsugi (Kumamoto University)**

	342	DSGV: Efficient Video Representation and Compression through Dynamic-Static 2D Gaussian Splatting. *Zhihang Luo (Tongji University)**
	381	VidCog: Empowering LLM with Long Video Understanding via Human-like Temporal Cognitive Loop. *Yupeng Wu (University of Chinese Academy of Sciences(UCAS))*; Xiaoshan Yang (University of Chinese Academy of Sciences(UCAS)); Changsheng Xu (University of Chinese Academy of Sciences(UCAS))*
	409	Dual-Branch Feature Modeling and Multi-Directional Motion Perception for Video Compression. *Zhi Liu (North China University of Technology); Yangbing Wang (North China University of Technology)*; Yuan Li (North China University of Technology); Hongyuan Jing (Beijing Union University); Mengmeng Zhang (Beijing Union University)*
	440	Time-IC: Empowering MLLM with Interleaved Context for Temporal-Sensitive Video Understanding. *Henghao Zhao (Nanjing University of Science and Technology)*; Peng Huang (Nanjing University of Science and Technology); Rui Yan (Nanjing University of Science and Technology); Zechao Li (Nanjing University of Science and Technology)*
	469	DynaPlane-Lane: Dynamic Multi-Plane Geometry Learning for Robust Monocular 3D Lane Detection. *Chunying Song (Nanjing University of Science and Technology); Huafeng Liu (Nanjing University of Science and Technology); Tao Chen (Nanjing University of Science and Technology); Qiong Wang (Nanjing University of Science and Technology)**
	492	Diffusion-Driven 3D Gaussian Splatting for Occlusion-Free Egocentric Scene Reconstruction. *Roucheng Lai (Sun Yat-Sen University)*; Haijing Liu (Sun Yat-Sen University); Hefeng Wu (Sun Yat-Sen University)*
	494	Prefix-Guided Adaptation of Pretrained Segmenter for RGB-D Indoor Panoptic Segmentation. *Yibin Wang (Tongji University); Yuanyang Wang (Tongji University); Tuowei Qu (Tongji University); Jiaxuan Zhu (Tongji University); Yu Fang (Tongji University)**
	503	Hint-Guided Video Frame Interpolation for Video Compression. *Pan Tan (Portland State University)*; Wu-chi Feng (Portland State University)*

16.00 ~ 18.00	**Oral 5:** Diffusion models Session Chair: John See, *Heriot-Watt University Malaysia*	
	114	Behavior Conditional Diffusion Model for Multi-Modal Recommendation. *Mengfan Kong (National University of Defense Technology)*; Chonghao Chen (National University of Defense Technology); Zhiqiang Pan (National University of Defense Technology); Fei Cai (National University of Defense Technology); Aimin Luo (National University of Defense Technology)*
	169	CharCom: Composable Identity Control for Multi-Character Story Illustration. *Zhongsheng Wang (University of Auckland)*; Ming Lin (University of Auckland); Zhedong Lin (University of Auckland); Yaser Shakib (Bedaia.ai); Qian Liu (University of Auckland); Jiamou Liu (University of Auckland)*
	197	Unifying Generative Self-Supervised Paradigms with Diffusion Models. *Xiaoyu Yue (The University of Sydney)*; Luping Zhou (The University of Sydney)*
	280	NeuroSwift: A Lightweight Cross-Subject Framework for fMRI Visual Reconstruction of Complex Scenes. *Shiyi Zhang (Southern University of Science and Technology)*; Dong Liang (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences); Yihang Zhou (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences)*
	371	DreamSwap: Motion-Preserving Personalized Video Subject Swapping with Video Diffusion Models. *Cong Luo (Zhejiang University); Xingxi Yin (Zhejiang University); Yan Gong (Zhejiang University); Yin Zhang (Zhejiang University)**
	424	A Unified Ensemble Clustering Framework via Higher-Order Graph Diffusion and Feedback-Based Refinement. *Faria Chowdhury Mumu (Agami Education Foundation); Kife I. Bin Iqbal (Shanxi University); Liang Du (Shanxi University)**
	437	IDiffPose: Equilibrium Substitution for Lightweight Diffusion-Graph-Based Pose Estimation. *Nugroho Wicaksono (Brawijaya University)*; Lailil Muflikhah (Brawijaya University); Novanto Yudistira (Brawijaya University)*
	465	How Much To Guide: Revisiting Adaptive Guidance in Classifier-Free Guidance Text-to-Vision Diffusion Models. *Huixuan Zhang (Peking University)*; Xiaojun Wan (Peking University)*

16.00 ~ 18.00	**Oral 6:** Social Emotion Session Chair: Lizi Liao, *Singapore Management University*	
	11	Dynamic Text-guided Multi-task Transformer for Multi-modal Sentiment Analysis. *Weile Huang (Shantou University)*; Weihong Cai (Shantou University); Zhanting She (Shantou University)*
	99	Personality-Enhanced Multimodal Depression Detection in the Elderly. *Honghong Wang (Beijing Fosafer Information Technology Co., Ltd.)**
	162	Multimodal Hyperbolic Embedding and Hyperbolic Hypergraph Fusion for Emotion Recognition in Conversation. *Yao Zheng (Communication University of China)*; Guowei Chen (Communication University of China); Wenchao Song (Communication University of China); Yanchao Liu (Communication University of China); Pengzhou Zhang (Communication University of China)*
	177	Joint Fine-grained Disentanglement and Modal-agnostic Fusion Multi-task Framework for Multimodal Sentiment Analysis. *Yao Zheng (Communication University of China)*; Chi Zhang (Communication University of China); Yujun Wen (Communication University of China); Xin Qi (Communication University of China); Pengzhou Zhang (Communication University of China)*
	245	MCIHN: A Hybrid Network Model Based on Multi-path Cross-modal Interaction for Multimodal Emotion Recognition. *Haoyang Zhang (Chongqing University of Posts and Telecommunications)*; Zhou Yang (Xi'an Jiaotong University); Ke Sun (University of New South Wales); Yucai Pang (Chongqing University of Posts and Telecommunications); Guoliang Xu (Chongqing University of Posts and Telecommunications)*
	386	Speech Emotion Recognition via Multi-Level Acoustic Modeling and Cross-Modal Temporal Fusion. *Xuan Zhang (Qilu University of Technology); Peng Zhang (Qilu University of Technology)*; Jianqiang Zhang (Qilu University of Technology); Wei Zhao (Qilu University of Technology); Fuqiang Wang (Qilu University of Technology); Xiaoming Wu (Qilu University of Technology)*
	506	Robust speech emotion recognition using conditional transformer-based architecture. *Hanwook Chung (Faurecia Irystec Inc.)*; Hyunjin Yoo (Faurecia Irystec Inc.)*

Thursday (11th December)

9.00 ~ 10.00	**Keynote 2:** Prof. Dr. Shin'ichi Satoh, *National Institute of Informatics (NII)* Title: Psychiatric Disease Diagnosis: The Challenge of Multimedia Technologies Chair: Lai-Kuan Wong, *Multimedia University*	
10.00 ~ 11.00	**Best Paper Oral** Session Chair: Kiyoharu Aizawa, *The University of Tokyo*	
	27	Mixture of Group Experts for Learning Invariant Representations. *Lei Kang (Beijing Normal University); Jia Li (Beijing Normal University)*; Mi Tian (TAL Education Group); Hua Huang (Beijing Normal University)*
	136	OF-NeRF: A Subjective Benchmark of Perceptual Quality Assessment for Outward-Facing NeRF Scenes with Multiple Distortions and Diverse Viewing Trajectories. *Qian Wang (Chongqing University of Technology); Zongju Peng (Chongqing University of Technology)*; Wenhui Zou (Ningbo University); Fen Chen (Chongqing University of Technology); Kai Xu (Chongqing University of Technology); Youshuang Zhao (Chongqing University of Technology)*
	265	EliGen: Entity-Level Controlled Image Generation with Regional Attention. *Hong Zhang (Zhejiang University)*; Zhongjie Duan (ModelScope Team, Alibaba Group Inc.); Xingjun Wang (ModelScope Team, Alibaba Group Inc.); Yingda Chen (ModelScope Team, Alibaba Group Inc.); Yu Zhang (Zhejiang University)*
	334	EmoSEM: Segment and Explain Emotion Stimuli in Visual Art. *Jing Zhang (Hefei University of Technology); Dan Guo (Hefei University of Technology)*; Zhangbin Li (Hefei University of Technology); Meng Wang (Hefei University of Technology)*
10.30 ~ 12.30	**Poster 4:** Vertical domains Session Chair: Lap-Kei Lee, *Hong Kong Metropolitan University*	
	102	MS^2G-Net: A Multi-Scale and Multi-Structural Guided Network for Information-Preserving Restoration of Ancient Murals. *Shengping Xiong (Yunnan University) *; Ying Yu (Yunnan University)*
	129	An Evaluation of Open Source Data Anonymisation Tools for Medical Data. *Anjali Pullattukunnel (University of Koblenz) *; Ibraheem Al-Dhamari (University of Koblenz); Jan Jürjens (University of Koblenz)*

	155	A new dataset and method for comprehensive evaluation of visual design elements in educational slides. *Xinning Du (Northeast Normal University); Qiwen Liang (Northeast Normal University); Jianping Ren (Northeast Normal University); Ming Fang (Northeast Normal University); Shuhua Liu (Northeast Normal University)**
	166	APGNet: Adaptive Prior-Guided for Underwater Camouflaged Object Detection. *Xinxin Huang (Nanjing University of Aeronautics and Astronautics); Han Sun (Nanjing University of Aeronautics and Astronautics)*; Junmin Cai (Nanjing University of Aeronautics and Astronautics); Ningzhong Liu (Nanjing University of Aeronautics and Astronautics); Huiyu Zhou (University of Leicester)*
	174	Single Image Dehazing Network Based on Depth-Guided Sparse Attention. *Xiaolong Kou (Inner Mongolia University)*; Ru Yi (Inner Mongolia University); Youchen Wang (Inner Mongolia University); Tianzhen Chen (Inner Mongolia University)*
	247	NamingCAD: A Naming Method for Unique Identification toward CAD Modeling Learning. *Yuxin Liu (Wuhan University); Fazhi He (Wuhan University)*; Rubin Fan (Wuhan University); Zhihao Zong (Wuhan University)*
	274	Distortion-Aware Network for Zero-Reference Retinal Image Enhancement. *Tianwei Zhou (Shenzhen University); Yuhang Feng (Shenzhen University); Shaoping Zhang (Shenzhen University); Linling Li (Southern Medical University); Guanghui Yue (Shenzhen University)*; Shishun Tian (Shenzhen University); Tianfu Wang (Shenzhen University)*
	310	Guiding Image Super-Resolution with Adaptive Residual Scaling. *Hao Zhang (China University Of Geosciences)*; Jiajun Lu (CSSC Marine Technology Co., Ltd.); Li He (Wuhan Digital Engineering Institute); Xiang Li (China University of Geosciences)*
	396	Towards Better Detection in the Rain: A Task-Driven Deraining Framework with Semantic Guidance. *Ya-Han Chang (National Taipei University of Technology); Shih-Hsuan Yang (National Taipei University of Technology)**
	403	DERM: Dynamic Evidence Reliability Modeling for Multimodal Rumor Detection. *Zhuangzhuang Pan (Universiti Malaya); Yan Xia (Suzhou University of Technology); Vimala Balakrishnan (Universiti Malaya)*; Muhammad Zaiamri Zainal Abidin (Universiti Malaya); Siti Soraya Abdul Rahman (Universiti Malaya)*

	416	Hybrid Mamba-Transformer Model for Lightweight Image Super-Resolution. *Szuchi Lu (National Taipei University of Technology); Shih-Hsuan Yang (National Taipei University of Technology)**
11.00 ~ 11.30	Coffee Break	
11.30 ~ 12.30	**Oral 7:** Captioning Session chair: Yu-Tong Cheng, *Singapore Management University*	
	30	Q-Adapter: Visual Query Adapter for Extracting Textually-related Features in Video Captioning. *Junan Chen (Nagoya University)*; Trung Thanh Nguyen (Nagoya University); Takahiro Komamizu (Nagoya University); Ichiro Ide (Nagoya University)*
	50	OPCap:Object-aware Prompting Captioning. *Feiyang Huang (South China Normal University)*; Yang Cao (South China Normal University); Jingyue Zhong (South China Normal University)*
	173	DualCap: Enhancing Lightweight Image Captioning via Dual Retrieval with Similar Scenes Visual Prompts. *Binbin Li (Chinese Academy of Sciences); Guimiao Yang (Chinese Academy of Sciences)*; Zisen Qi (Chinese Academy of Sciences); Haiping Wang (Chinese Academy of Sciences); Yu Ding (Chinese Academy of Sciences)*
	250	Intrinsic Feature Rectification: Mitigating RAG Dependency by Addressing Information Loss in Image Captioning. *Hao Wu (University of Science and Technology of China); Zhihang Zhong (Shanghai Artificial Intelligence Laboratory); Xiao Sun (Shanghai Artificial Intelligence Laboratory)**
	345	SAFE-AKT: Kazakh Image-Text Retrieval via Semantic-Agnostic Feature Enhancement and Adaptive Knowledge Transfer. *Ping Hu (Xinjiang University)*; Zhiqun Cao (Xinjiang University); Mingqiang Xu (IFlytek, University Of Science And Technology); Jingjing He (Xinjiang Institute Of Ecology And Geography Chinese Academy Of Science); Lumei Zhou (Xinjiang University); Changle Ying (Xinjiang University)*

11.30 ~ 12.30	**Special Session:** Privacy-Aware Multimodal Analysis and Its Applications Session Chair: Wei-Ta Chu, *National Cheng Kung University*	
	295	DCN: Decoupled-Coupled Network for Text-based Person Search. *Zhiyuan Qi (Nanjing University of Aeronautics and Astronautics); Rong Quan (Nanjing University of Aeronautics and Astronautics); Liangxu Su (Nanjing University of Aeronautics and Astronautics); Yizhen Jia (Nanjing University of Aeronautics and Astronautics); Wentong Li (Nanjing University of Aeronautics and Astronautics); Yichao Yan (Shanghai Jiao Tong University); Jie Qin (Nanjing University of Aeronautics and Astronautics)*
	372	Mitigating Hallucinations in Large Vision-Language Models via Dual Contrastive Decoding. *Jiulong Wu (Soochow University); Yucheng Shen (Soochow University); Haixin Sun (Soochow University); Min Cao (Soochow University)*
	461	Wandering and feeling the Scenes: Body-Aware Diffusion for 3D Human Motion Generation. *Chong Zhang (East China Normal University); Jingyu Gong (East China Normal University); Shaohui Lin (East China Normal University); Yang Li (East China Normal University); Zhizhong Zhang (East China Normal University)*
11.30 ~ 12.30	**Special Session 2:** Advances in Multimedia Security: Ensuring Trust and Integrity in the AI Era Session Chair: Fan Qi, *Tianjin University of Technology*	
	293	When Hallucinated Concepts Cross Modals: Unveiling Backdoor Vulnerability in Multi-modal In-context Learning. *Guanyu Hou (University of Manchester); Jiaming He (University of Electronic Science and Technology of China); Yitong Qiao (Sun Yat-sen University); Jiachen Li (Wuhan University of Technology); Qiyang Song (Chinese Academy of Sciences); Ji Guo (University of Electronic Science and Technology of China); Zihan Wang (University of Electronic Science and Technology of China); Hongyun Wang (Huazhong University of Science and Technology); Wenbo Jiang (University of Electronic Science and Technology of China)*

	325	You Are Out of My Focus: A Defocus-Blur Backdoor Attack against Deep Learning Models. *Tao Wang (University of Electronic Science and Technology of China); Hongwei Li (University of Electronic Science and Technology of China); Wenbo Jiang (University of Electronic Science and Technology of China); Jiaming He (University of Electronic Science and Technology of China); Zihan Wang (University of Electronic Science and Technology of China); Rui Zhang (University of Electronic Science and Technology of China); Ji Guo (University of Electronic Science and Technology of China); Jiachen Li (School of Computer Science and Artificial Intelligence, Wuhan University of Technology)*
	326	Unveiling Byzantine-robust with Varied Batch Sizes across Different Clients in Federated Learning. *Yunxuan Li (School of Cyber Science and Engineering, Nanjing University of Science & Technology); Yuxin Wei (School of Cyber Science and Engineering, Nanjing University of Science & Technology); Xinjian Huang (School of Cyber Science and Engineering, Nanjing University of Science & Technology); Bo Du (School of Computer Science, Wuhan University)*
12.30 ~ 14.00	Lunch	
13.30 ~ 15.30	**Poster 5:** Multimodal fusion Session Chair: Trung Thanh Nguyen, *Nagoya University*	
	31	Multi-Scale Attention Prediction and Multi-Path Fusion Attention For Compressed Video Quality Enhancement. *Chi Cheng (Wuhan University)*; Bingyu Wu (Tufts University); Haojun Ai (Wuhan University); Xinyue Yang (Wuhan University); Pengfei Lv (Wuhan University); Feng Zhang (Wuhan University)*
	93	CSSA-Fusion: Channel Selective and Spatial Alignment Infrared-Visible Image Fusion. *Zhen Li (Chongqing Normal University)*; Zhi Zeng (Chongqing Normal University); Zhongrui Xiao (Chongqing Normal University); Wen Ming (Shenzhen University); Zhiyuan Zhang (Singapore Management University); Yibin Tian (Shenzhen University)*
	233	Spatial Momentum Networks: Physics-Guided Efficient Attention Fusion Method. *Mengmeng Yu (Shanghai University of International Business and Economics)*; Xiang Fu (Shanghai University of International Business and Economics); Caiyun Fan (Shanghai University of International Business and Economics); Huizheng Yu (Shanghai University of International Business and Economics)*

	258	SILLM4Rec: Self-Improving with Chain of Thought Enhanced Preference Optimization for Multimodal Recommendation. *Yuhao Wu (Beijing University of Posts and Telecommunications); Fang Quan (Beijing University of Posts and Telecommunications)*; Maoqi Liu (Beijing University of Posts and Telecommunications); Xiaowen Huang (Beijing Jiaotong University); Jitao Sang (Beijing Jiaotong University)*
	268	A Multi-Agent Framework for Fine-Grained Multimodal Named Entity Recognition through Check and Reasoning. *Heng-yang Lu (Jiangnan University)*; Xintong Liu (Jiangnan University); Xingda Shang (Jiangnan University); Wei Fang (Jiangnan University); Xiao-jun Wu (Jiangnan University)*
	273	Multi-modal Recommendation with Joint Content and Interaction Augmentation. *Jiajie Deng (City University of Hong Kong)*; Haokun Wen (Harbin Institute of Technology (Shenzhen) & City University of Hong Kong); Xiao Han (Zhejiang University of Technology); Xuemeng Song (Southern University of Science and Technology); Xiangyu Zhao (City University of Hong Kong)*
	316	Fusion-Augmented Deep Multi-view Clustering via Contrastive View Expansion. *Zhongyi Ma (University of Science and Technology of China); Xiaorui Jiang (University of Science and Technology of China); Yong Liao (University of Science and Technology of China)**
	336	MCTG: A Multimodal Self-Supervised Contrastive Learning Framework Based on CTG. *Jiayu Wu (Jinan University)*; Huijin Wang (Jinan University); Ziduo Yang (Jinan University); Jiadong Wu (Jinan University)*
	363	From Feature Alignment to Multimodal Fusion: A Two-Stage Primary Modality-Guided Approach for MSA. *Guoyu Ma (Qilu University of Technology); Xiaoqiang Ren (Qilu University of Technology); Yan Jiang (Qilu University of Technology)*; Hongjiao Guan (Qilu University of Technology); Bing Xu (Harbin Institute of Technology)*
	397	WP-CMA: Waypoint Prediction for Cross-modal Alignment of Vision-and-Language Navigation in Continuous Environments. *Siyang Fu (Hangzhou Normal University); Yifei Wu (Hangzhou Normal University); Ting Yu (Hangzhou Normal University)**

14.00 ~ 15.30	**Panel:** Shaping the Future of Multimedia: Women Inspiring the Next Wave of AI Innovation Moderator: Shoko Imaizumi, *Chiba University* Panelists: • Jiaying Liu, *Peking University* • Phoebe Chen, *La Trobe University* • Lizi Liao, *Singapore Management University*	
14.00 ~ 15.30	**Special Session:** Cultural Heritage Analysis and Recommendation Session Chair: Guoshuai Zhao, *Xi'an Jiaotong University*	
	427	Sequence-augmented Conversational Recommendation System Based on Diffusion Models for Personalized Cultural Exploration. *Lun Tan (Xi'an Jiaotong University); Jun Gong (Xi'an Jiaotong University); Jiakui Shen (Xi'an Jiaotong University); Yunqi Mi (Xi'an Jiaotong University); Guoshuai Zhao (Xi'an Jiaotong University); Jialie Shen (City University of London); Xueming Qian (Xi'an Jiaotong University)*
	481	DriveEval-Agent: A Closed-Loop Framework Combining Zero-Shot Benchmarking and Full Fine-Tuning for Multimodal Autonomous Driving. *Jingwen Zhao (Sun Yat-Sen University); Junwei Hu (South China Normal University); Yichu Liu (Desay SV)*
	487	TOVect: Topology-Optimized Vectorization for Intangible Cultural Heritage Thangka Element Line Art. *Anshu Hu (Wuhan University of Technology); Yifei Sun (Chomo Yarlung Tibet Co., Ltd); Jiachen Li (Wuhan University of Technology); Yanchun Ma (Wuhan Vocational College of Software and Engineering); Qing Xie (Wuhan University of Technology); Yongjian Liu (Wuhan University of Technology)*
	497	Multi-Task Learning with Adaptive Fusion for Point-of-Interest Recommendation. *Bosong Yang (Xi'an Jiaotong University); Chengxu Liu (Xi'an Jiaotong University); Boyang Yan (Xi'an Jiaotong University); Zimo Zhu (Xi'an Jiaotong University)*

14.00 ~ 15.30	**Oral 8:** Content Generation Session Chair: Chong-Wah Ngo, *Singapore Management University*	
	19	MoCA: Identity-Preserving Text-to-Video Generation via Mixture of Cross Attention. *Qi Xie (University of Science and Technology of China)*; Yongjia Ma (Li Auto); Donglin Di (Li Auto); Xuehao Gao (Northwestern Polytechnical University); Xun Yang (University of Science and Technology of China)*
	154	UniCP: A Unified Caching and Pruning Framework for Efficient Video Generation. *Wenzhang Sun (Li Auto)*; Qirui Hou (Harbin Institute of Technology); Donglin Di (Li Auto); Jiahui Yang (Harbin Institute of Technology); Yongjia Ma (Li Auto); Jianxun Cui (Harbin Institute of Technology)*
	222	BALANCE: A Fairness-Aware Framework for Privacy-Preserving Synthetic Clinical Text Generation. *Yi Tang (Wuhan University)*; Dengpan Ye (Wuhan University)*
	294	You Can Generate It Again: Data-to-Text Generation with Verification and Correction Prompting. *Xuan Ren (University of Adelaide)*; Zeyu Zhang (The Australian National University); Lingqiao Liu (University of Adelaide)*
	389	InstructTrack: Language-Guided Multi-Object Tracking with Semantic-Aware Association. *Zishun Zhou (Beihang University); Shuai Wang (Beihang University)*; Hao Sheng (Beihang University); Dazhi Yang (Beihang University); Sentan Li (Beihang University); Da Yang (Beihang University); Zhenglong Cui (Beihang University)*
	410	Stance-Driven Multimodal Controlled Statement Generation: New Task and Dataset. *Bingqian Wang (Beijing University of Posts and Telecommunications); Quan Fang (Beijing University of Posts and Telecommunications)*; Xiaoxiao Ma (Beijing University of Posts and Telecommunications)*
15.30 ~ 16.00	Coffee Break	

15.30 ~ 17.30	**Poster 6:** Intelligent multimedia Session Chair: Lap-Kei Lee, *Hong Kong Metropolitan University*	
	17	Federated Deep Incomplete Multi-View Clustering with Heterogeneity-Matching and Attention-Based Imputation. *Lang Qin (University of Science and Technology of China)*; Xiaorui Jiang (University of Science and Technology of China); Yu Gao (University of Science and Technology of China); Yong Liao (University of Science and Technology of China)*
	33	ChronoSelect: Robust Learning with Noisy Labels via Dynamics Temporal Memory. *Jianchao Wang (University of Electronic Science and Technology of China); Qingfeng Li (University of Electronic Science and Technology of China); PengCheng Zheng (University of Electronic Science and Technology of China); Xiaorong Pu (University of Electronic Science and Technology of China)*; Yazhou Ren (University of Electronic Science and Technology of China)*
	81	Test-time Filtering Boosts Training-free Zero-shot Composed Image Retrieval. *Haoyue Chong (Nanjing University of Science and Technology)*; Lunbo Li (Nanjing University of Science and Technology); Haofeng Zhang (Nanjing University of Science and Technology)*
	83	Learn Concepts from Multi-Scale Visual Information for Compositional Zero-Shot Learning. *Guanyu Wang (Peking University)*; Zhijie Tan (Peking University); Xu Chu (Peking University); Xinrong Chen (Peking University); Tong Mo (Peking University); Weiping Li (Peking University)*
	153	CCANet : A Cognition-Inspired Framework for Few-Shot Segmentation from Category-Agnostic to Category-Aware. *Long Zhou (Hefei University of Technology)*; Ronggui Wang (Hefei University of Technology); Lixia Xue (Hefei University of Technology); juan yang (Hefei University of Technology)*
	175	Pseudo-Partial Label Helps: A Simple Way to Mitigate Pseudo-Label Noise in Source-Free Domain Adaptation. *Yiran Zhang (University of Chinese Academy of Sciences)*; Qianqian Xu (Academy of Sciences); Yangbangyan Jiang (University of Chinese Academy of Sciences); Qingming Huang (University of Chinese Academy of Sciences)*

	179	FALU: A Proactive Deepfake Detection Scheme Based on Average Hashing and Mamba-Like Linear Attention U-Net. *Jian Li (Qilu University of Technology); Wei Han (Qilu University of Technology)*; Bin Ma (Qilu University of Technology); Xiaolong Li (Beijing Jiaotong University); Zhenxing Qian (Fudan University)*
	190	CARD: Control-Driven Autoregressive Reconstruction with Decoupled Learning for Multi-Class Anomaly Detection. *Yifan Wang (University of Chinese Academy of Science)*; Mingqing Wang (Tsinghua University); Boyi Sun (University of Chinese Academy of Science); Qianfan Zhao (Hangzhou Xingwuzhong Robot Co., Ltd); Lu Zhang (Chinese Academy of Science); Zhiyong Liu (Chinese Academy of Science); Xu Yang (Chinese Academy of Science); Suiwu Zheng (Chinese Academy of Science)*
	191	TriDet-MLLM: Triple-Feature Fusion Prompt Learning for AI-Generated Image Detection. *Yichen Liu (Harbin Engineering University); Jinqi Zhang (Harbin Engineering University); Yuchen Zhou (Harbin Engineering University); Rongsheng Li (Harbin Engineering University); Yanxia Wu (Harbin Engineering University); Qiao Tian Harbin Engineering University); Shang Feng (Harbin Engineering University)*
	207	AS-UAP:Attention-Shift Universal Adversarial Perturbation on Transformer-based Models. *Xiaoyu Zhang (Jinan University); Bowei Wu (Jinan University); Shujun Xie (Jinan University); Qiang Xu (Jinan University); Shuai Pang (Jinan University); Dehua Zhou (Jinan University)*
	285	AVIR: Adaptive Visual In-Document Retrieval for Efficient Multi-Page Document Question Answering. *Zongmin Li (China University of Petroleum (East China)); Yachuan Li (China University of Petroleum (East China))*; Lei Kang (Universitat Autònoma de Barcelon); Dimosthenis Karatzas (UniversitatAutònoma de Barcelona); Wenkang Ma (China University of Petroleum (East China))*
	435	Adaptive Image Acquisition Algorithms for Resource-Constrained Single-Photon Cameras. *Yeganeh Jalalpour (Portland State University)*; Wu-chi Feng (Portland State University)*

16.00 ~ 17.30	**Oral 9:** Classification and Segmentation Session Chair: Cheng Li, *University of Alberta*	
	130	SAAB: Enhancing Action Segmentation via Spatial Attention Pooling and Action-Background Classifier. *Zhenhui Lou (Zhejiang University of Technology)*; Xia Zhang (Zhejiang University of Technology); Fengjie Li (Zhejiang University of Technology); Jiangtao Ye (Zhejiang University of Technology); Xiaoping Jiang (Zhejiang University of Technology); Zhenbo Cheng (Zhejiang University of Technology)*
	148	Learning Content-enhanced Tokens for Domain Generalized Semantic Segmentation. *Shishun Tian (Shenzhen University)*; Ziqi Yang (Shenzhen University); Wenbin Zou (Shenzhen University); Yuanhao Gong (Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences); Guanghui Yue (Shenzhen University); Ting Su (Shenzhen MSU-BIT University)*
	181	Ora-NSC: A Novel Semi-supervised Approach for Oracle Bone Fragment Classification with Imbalanced Classes. *Jie Liu (Henan Normal University)*; Yunyue Hu (Henan Normal University); Weijie Gao (Henan Normal University); Yuhang Wang (Henan Normal University); Yijing Si (Henan Normal University); Yue Zhang (Henan Normal University)*
	244	Fourier-based Dual-level Perturbation for Single Domain Generalized Object Detection. *YongBing Zhang (NingXia University)*; Yan Lirong (NingXia University); Tang Xiaofen (NingXia University)*
	349	MFNet: A Deep Learning Method for Urban Perception Evaluation Using Street View Images. *Jing Jin (Beijing Technology and Business University)*; Yingying Liu (Beijing Technology and Business University); Dongming Sun (Changchun University of Chinese Medicine); Su Li (Beijing Technology and Business University); Xun Zhang (Beijing Technology and Business University)*
	421	Amodal-KAN: The First Look at Kolmogorov–Arnold Network for Amodal Instance Segmentation. *Ce Zheng (Zhejiang University of Technology); Qiong Wang (Zhejiang University of Technology); Cong Bai (Zhejiang University of Technology)**

16.00 ~ 17.30	**Oral 10:** Multimodality Session Chair: Shin'ichi Satoh, *National Institute of Informatics (NII)*	
	182	Can Multimodal Large Language Models Understand Human Values in Videos? *Yuchen Wang (Communication University of China)*; Hao Qiu (Communication University of China); He Chang (Communication University of China); Junbin Xiao (National University of Singapore); Zhulin Tao (Communication University of China); Libiao Jin (Communication University of China)*
	188	MM-HRIViT: Early Behavior Prediction for HRI using mmWave Radar and a Vision Transformer. *Xiaohang Zhang (Xinjiang University)*; Yajun Zhang (Xinjiang University); Chunhong Yue (Xinjiang University)*
	312	PMCMatcher: A Parallel Multi-Scale Cascaded Transformer-Based Network for Multimodal Feature Matching. *Yun Liao (Yunnan University); Jiayi Lyu (Yunnan University); Junhui Liu (Yunnan University); Nan Chen (Yunnan University); Zongxiao Hu (Yunnan University); Qing Duan (Yunnan University)**
	433	Global Question-Aware Multimodal Retrieval-Augmented Generation for Multimedia Multi-Hop Question Answering. *Zhixiao Shen (Nanjing University of Science and Technology)*; Jianfei Yu (Nanjing University of Science and Technology); Wenya Wang (Nanyang Technological University); Rui Xia (Nanjing University of Science and Technology)*
	451	Uni-IL: Unified Incremental Learning of Vision-Language Models via Mixture of Attribute-Guided Experts. *Fanyu Meng (Beijing Institute of Technology)*; Yufeng Zhan (Beijing Institute of Technology); Jie Zhang (The Hong Kong University of Science and Technology); Zhiyuan Wang (Beijing Institute of Technology); Yuanqing Xia (Beijing Institute of Technology, Zhongyuan University of Technology)*
	472	PENFORMER: Identifying Signature Truth with Log Normal Aided Multimodal Transformer. *Anurag Pandey (Indian Institute of Technology Mandi)*; Aditya Nigam (Indian Institute of Technology Mandi); Arnav Bhavsar (Indian Institute of Technology Mandi); Divya Acharya (Simula Research); Basu Verma (HCLTech)*

16.00 ~ 17.30	**Oral 11:** Sound, sign and text Session Chair: Jiaying Liu, *Peking University*	
	140	SeeingSounds: Learning Audio-to-Visual Alignment via Text. *Simone Carnemolla (University of Catania)*; Matteo Pennisi (University of Catania); Chiara Maria Russo (University of Catania); Simone Palazzo (University of Catania); Daniela Giordano (University of Catania); Concetto Spampinato (University of Catania)*
	319	ECTSpeech: Enhancing Efficient Speech Synthesis via Easy Consistency Tuning. *Tao Zhu (Xinjiang University)*; Yinfeng Yu (Xinjinag University); Liejun Wang (Xinjinag University); Fuchun Sun (Tsinghua university); Wendong Zheng (Tianjin University of Technology)*
	327	SELECT: Detecting Label Errors in Real-world Scene Text Data. *Wenjun Liu (NetEase); Qian Wu (NetEase)*; Yifeng Hu (NetEase); Yuke Li (NetEase)*
	341	Gloss-Free Sign Language Translation With Optical-Flow Guided Two-Stream Network. *Peidong Liu (Tianjin University)*; Lianyu Hu (Tianjin University); Tongkai Shi (Tianjin University); Fanhua Shang (Tianjin University); Jichao Feng (Tianjin University); Liang Wan (Tianjin University); Wei Feng (Tianjin University)*
	473	Guiding Task Choice in Japanese Voice Interfaces through Vocalization Cost: Click-based vs. Voice-based Selection. *Ryunosuke Shigematsu (Meiji University)*; Ryuto Oishi (Meiji University); Yuki Nakagawa (Meiji University); Satoshi Nakamura (Meiji University); Takeshi Torii (Subaru Corporation); Hideyuki Takao (Subaru Corporation)*
	479	Don’t Break the Melody: Encouraging Accurate Handwriting Practice with Sound Feedback. *Reo Hatogai (Meiji University)*; Sayuri Matsuda (Meiji University); Kento Watanabe (Meiji University); Satoshi Nakamura (Meiji University); Akiyuki Kake (Wacom Co., Ltd.)*

Friday (12th December)

9.00 ~ 10.00	**Keynote 3:** Prof. Dr. Cathal Gurrin, *Dublin City University* Title: Learnings From a Decade of Research into Lifelog Organisation and Retrieval Chair: Chee Seng Chan, *Universiti Malaya*	
10.00 ~ 11.00	**Oral 12:** All about light! Session Chair: Ven Jyn Kok, *Universiti Kebangsaan Malaysia*	
	216	HyperEnhancerNet: Lightweight Speech Enhancement with Learnable Time-Frequency Resampling. *Yihan Zhang (China FAW)*; Jialiang Wang (China FAW)*
	242	DS-IB Net: Ultra-Lightweight Weakly-Supervised Video Anomaly Detection through Synergistic Dual Streams and Information Bottleneck. *Tao Zhu (Jiangxi University of Finance and Economics)*; Qi Yu (Jiangxi Science & Technology Normal University); Heran Song (Jiangxi University of Finance and Economics); Yuheng Cheng (Jiangxi University of Finance and Economics); Shiyu Li (Jiangxi University of Finance and Economics); Yue Liu (Jiangxi University of Finance and Economics); Xinyi Tu (Jiangxi University of Finance and Economics); Kaiwen Luo (Jiangxi University of Finance and Economics)*
	395	Enhancing Low-Light Object Detection with Zero-Shot Dual-Branch Illumination-Invariant Network. *Chengpeng Li (Jiangxi Normal University); Aiwen Jiang (Jiangxi Normal University)*; Tianjia Miao (Jiangxi Normal University)*
	502	Multi-contrastive Regularization for Single Image Portrait Relighting. *Quanxing Peng (North China Electric Power University); Jipeng Zhang (North China Electric Power University)*; Wengang Cheng (North China Electric Power University)*
10.00 ~ 11.00	**Oral 13:** Sketch and search Session Chair: Yunshan Ma, *Singapore Management University*	
	94	Leveraging Pseudo-triplet and Flexible Prompt for Zero-shot Composed Image Retrieval. *Haoyue Chong (Nanjing University of Science and Technology)*; Lunbo Li (Nanjing University of Science and Technology); Haofeng Zhang (Nanjing University of Science and Technology)*

	272	STCGen:Sketch-based Text-to-Clothing Image Generation with Contour and Style Consistency. *Fei Fang (Wuhan Textile University); Heng Jiang (Wuhan Textile University)*; Jiawen Yan (Wuhan Textile University); Chunxia Xiao (Wuhan University); Ruhan He (Wuhan Textile University); Jia Chen (Wuhan Textile University); Mingfu Xiong (Wuhan Textile University); Tao Peng (Wuhan Textile University); Xinrong Hu (Wuhan Textile University)*
	393	Sketch-1-to-3: One Single Sketch to 3D Detailed Face Reconstruction. *Liting Wen (Carnegie Mellon University); Zimo Yang (Nanyang Technological University)*; Xianlin Zhang (Beijing University of Posts and Telecommunications); Chi Ding (Beijing University of Posts and Telecommunications); Mingdao Wang (Tsinghua University); Xueming Li (Beijing University of Posts and Telecommunications)*
10.00 ~ 11.00	**Oral 14:** Emerging topics Session Chair: Zhedong Zheng, *University of Macau*	
	142	Omnidirectional Spatial Modeling from Correlated Panoramas. *Xinshen Zhang (The Hong Kong Polytechnic University); Tongxi Fu (ZJTU); Xu Zheng (Hong Kong University of Science and Technology)**
	317	What Really Matters for Learning-based LiDAR-Camera Calibration. *Shujuan Huang (Beijing Jiaotong Unviersity); Chunyu Lin (Beijing Jiaotong Unviersity)*; Yao Zhao (Beijing Jiaotong Unviersity)*
	383	Towards Inclusive Gendered Embodiment: Designing AI Agents Based on Sociological Gender Theories. *Debasmita Mukherjee (University of New Brunswick)*; Kimia Poozesh (Ferdowsi University of Mashhad)*
	385	Gradient Shaping Beyond Clipping: A Functional Perspective on Update Magnitude Control. *Haochen You (Columbia University)*; Baojing Liu (Hebei Institute of Communications)*
10.30 ~ 12.30	**Poster 7:** Everyday multimedia Session Chair: Yu-Tong Cheng, *Singapore Management University*	
	39	Cloth-Changing Person Re-identification with Human Keypoint Prediction. *Nian Wang (Beihang University); Jin Zheng (Beihang University)*; Suhai Zhou (Beihang University)*

	42	SuPACape: Graph-based Category-Agnostic Pose Estimation with Super-Category and Pose Adaptivity. *Yi-Hsuan Lu (National Cheng Kung University); Wei-Ta Chu (National Cheng Kung University)**
	344	KeyRegionPose: Region-Aware Feature Interaction and Multi-Scale Token Pruning for Efficient Human Pose Estimation. *Xuanchen Wang (The University of Sydney)*; Heng Wang (The University of Sydney); Dongnan Liu (The University of Sydney); Weidong Cai (The University of Sydney)*
	394	AL-SSFGait: A Dual-Branch Framework with Aligned Silhouette-Skeleton for Gait Recognition. *Jianlou Lou (Northeast Electric Power University); Guiping Zhang (Northeast Electric Power University); Jianxun Lou (Northeast Electric Power University)**
	491	A Robust 3D CNN with Pyramidal Attention for Spatiotemporal Gait Recognition. *Jianyu Chen (Wuhan University); Qian Zhou (Wuhan University)*; Qin Zou (Wuhan University); Chao Liang (Wuhan University); Zengmin Xu (Guilin University of Electronic Technology); Gang Wu (Tarim University); Zhongyuan Wang (Wuhan University)*
10.30 ~ 12.30	**Short Paper Poster** Session Chair: Yu-Tong Cheng, *Singapore Management University*	
	20	Contracted Gram Tensor Distillation for Object Detection. *Takumi Karasawa (Institute of Science Tokyo)*; Nakamasa Inoue (Institute of Science Tokyo); Rei Kawakami (Institute of Science Tokyo)*
	120	ReSSFormer: A Recursive Sparse Structured Transformer for Scalable and Long-Context Reasoning. *Haochen You (Columbia University)*; Baojing Liu (Hebei Institute of Communications)*
	264	Rethinking Misalignment in Virtual Staining: A Preliminary Study on Using STN-Guided Supervision at Training Time to Improve H&E-to-IHC Translation. *Iman Yi Liao (University of Nottingham Malaysia Campus)*; Mustafa Mahir Suliman Fadlelbari (University of Nottingham Malaysia); Mahmoud Khattab (University of Southampton Malaysia); Amr Ahmed (Edge Hill University); Jia Wern Pan (Cancer Research Malaysia)*

	298	Breaking Distributional Assumptions in Multi-view Learning: Test-time Adaptive Fusion via Conformalized Evidence Representation. *Shijie Ding (Shanghai University)*; Xiaodong Yue (Shanghai University); Yufei Chen (Tongji University); Jie Shi (Shanghai University); Dongqi Xia (Shanghai University)*
	306	Dual-View Gradient Probes: Disentangling Uncertainty for Deep Active Learning. *Dongqi Xia (Shanghai University)*; Xiaodong Yue (Shanghai University); Yufei Chen (Tongji University); Jie Shi (Shanghai University); Shijie Ding (Shanghai University)*
	353	A Unified Framework for the Convergence and Weight Pruning in Federated Learning. *Mengchen Fan (The University of Alabama at Birmingham); Tianyun Zhang (Cleveland State University)*; Baocheng Geng (The University of Alabama at Birmingham)*
	384	Edge-Computing Thermal Object Detection with Homography-Augmented Data for Night Navigation. *Jia-Hong Lin (Industrial Technology Research Institute); Hsin-Jung Cheng (Industrial Technology Research Institute); Chih-Tsung Shen (Industrial Technology Research Institute)**
11.00 ~ 11.30	Coffee Break	
11.30 ~ 12.30	**Brave New Idea** Session chairs: John See, *Heriot-Watt University*; Zhedong Zheng, *University of Macau*	
	516	Co-Designer: A Human-AI Collaborative Framework for Deriving Physical Product Concept Design Opportunities from User-Generated Content (UGC) in Manufacturing. *Rongcong Cai (Zhaoqing University)*; Mengyao Guo (Harbin Institute of Technology, Shenzhen); Shuo Yang (Zhaoqing University); Runyuan Li (Nanjing University of the Arts); Xinru Wu (Guangdong University of Finance and Economics); Jiahui Li (Zhaoqing University)*
	518	Animating the Ephemeral: Transforming Edible Cultural Heritage into Dynamic Digital Heritage through AI and Mixed Reality. *Haowei Xiong (University of the Arts London); Kexin Nie (University of Sydney); Jiachen Zeng (Royal College of Art); Shujing Shen (University of the Arts London); Mengyao Guo (Harbin Institute of Technology, Shenzhen)**

	520	Encoding Music Score Data for Emotional Expression Prediction in Western Classical Music Pieces. *Sui Wei En (Universiti Tunku Abdul Rahman)*; Zhi Lin Chong (Universiti Tunku Abdul Rahman); John See (Heriot-Watt University Malaysia)*
11.30 ~ 12.30	**Doctoral Consortium** Session chair: Debasmita Mukherjee, University of New Brunswick	
.	DS1	MultiSensor-Home: Benchmark for Multi-modal Multi-view Action Recognition in Home Environments. *Trung Thanh Nguyen (Nagoya University)*
	DS2	Rehabilitation Game Model Using Brain-Computer Interface for Older Adults with Mild Dementia. *Iffa Nurlatifah (Universiti Kebangsaan Malaysia)*
	DS3	Biometric-Based Recovery Mechanism for Blockchain-Based Wallet. *Budoor Saeed Bawazeer (University of Technology Sydney)*
11.30 ~ 12.30	**Demo** Session chair: Chih-Yang Pee, *Multimedia University*	
	511	An Identity-Oriented System for Generated Human Video Detection. *Liang Tan (Beijing University of Posts and Telecommunications)**
	512	KuzushijiGen: A Real-Time Few-Shot Japanese Kuzushiji Generator via Differentiable Rendering. *Honghui Yuan (The University of Electro-Communications, Tokyo)*; Keiji Yanai (The University of Electro-Communications, Tokyo)*
	519	VibeMus: Proactive Agentic System for Music Personalization. *Zhiliang Guo (University of Science and Technology of China)*; Teng Tu (National University of Singapore); Yunshan Ma (Singapore Management University); Xun Yang (University of Sicence and Technology of China)*

12.30 ~ 13.00	Closing
13.00 ~ 14.00	Lunch
14.00 ~ 18.00	Half-day Tour