MMSP 2025 Program

Room: Shougang Curling Hall

September 21, Sunday

Keynote 1: From Digital Human to Humanoids

Speaker: Dr. Xiaodong He
September 21, 9:00-9:40

Coffee Break

September 21, 9:40-10:00

Special Session: Immersive & Intelligent Speech & Audio Processing (I2SAP)

September 21, 10:00-12:00

- **Music Source Restoration**
  - Yongyi Zang (Independent Researcher)*; Zheqi Dai (The Chinese University of Hong Kong); Mark Plumbley (University of Surrey); Qiuqiang Kong (The Chinese University of Hong Kong)
- **Data-independent Beamforming for End-to-end Multichannel Multi-speaker ASR**
  - Can Cui (iFLYTECH)*; Paul Magron (INRIA); Mostafa Sadeghi (INRIA); Emmanuel Vincent (INRIA)
- **Towards Low-Latency Tracking of Multiple Speakers With Short-Context Speaker Embeddings**
  - Taous Iatariene (Orange, Université de Lorraine, CNRS, Inria, Loria)*; Alexandre Guérin (Orange); Romain Serizel (Université de Lorraine, CNRS, Inria, Loria)
- **White-box Differentiable Model of Perceived Localisation**
  - Antoine Souchaud (University of Surrey)*; Enzo De Sena (University of Surrey); Zoran Cvetkovic (King's College); Annika Neidhardt (University of Surrey); Pedro Llado (University of Surrey)
- **Frequency-Weighted Training Losses for Phoneme-Level DNN-based Speech Enhancement**
  - Nasser-Eddine Monir (Université de Lorraine, CNRS, Inria, Loria)*; Paul Magron (Université de Lorraine, CNRS, Inria, Loria); Romain Serizel (Université de Lorraine, CNRS, Inria, Loria)
- **Lightweight DNN for Full-Band Speech Denoising on Mobile Devices: Exploiting Long and Short Temporal Patterns**
  - Konstantinos Drossos (Nokia Technologies); Mikko Heikkinen (Nokia Technologies)*; Paschalis Tsiaflakis (Nokia Bell Labs)

Lunch Break

September 21, 12:00-13:30

Industry Demo

September 21, 13:30-15:30

Poster Session

September 21, 13:30-15:30

- **LSS3D: Learnable Spatial Shifting for Consistent and High-Quality 3D Generation from Single-Image**
  - Zhuojiang Cai (Beihang University); Yiheng Zhang (National University of Singapore); Meitong Guo (Tsinghua University); Mingdao Wang (Tsinghua University); Yuwang Wang (Tsinghua University)*
- **Towards Volumetric Video: a Technical Overview of Immersive Media**
  - Shi Pan (Tsinghua University)*; Hongshuai Li (University of Chinese Academy of Sciences); Zhengxian Yang (Tsinghua University); Le Wang (Migu Xinkong); Cheng Su (Migu Culture); Liqian Ma (Migu Culture); Hua Du (Beijing MEET YUAN); Borong Lin (Tsinghua University); Tao Yu (Tsinghua University)
- **PromptGS: Visual Prompting for Tiny Object Reconstruction in 3DGS Optimization**
  - Xun Wang (Beihang University); Xutao Xue (Beihang University); Xubing Kang (Beijing Aerospace Automatic Control Institute); Siyuan Li (Hangzhou International Innovation Institute Beihang University); Shayer Shabab Utsho (Beihang University); Kun Li (Tianjin University); Mengqi Ji (Beihang University)*
- **Adapting Image-to-Video Diffusion Models for Large-Motion Frame Interpolation**
  - Luoxu Jin (Waseda University)*; Hiroshi Watanabe (Waseda University)
- **Efficient Polyp Detection via Wavelet-Driven Boundary Enhancement and Temporal Consistency**
  - Hanwen Zhang (University of Electronic Science and Technology of China)*; Heqian Qiu (University of Electronic Science and Technology of China); Lanxiao Wang (University of Electronic Science and Technology of China); Chenghao Qi (University of Electronic Science and Technology of China); Ruisong Dai (University of Electronic Science and Technology of China); Hongliang Li (University of Electronic Science and Technology of China)
- **Lightweight Steel Surface Defect Detection via Knowledge Distillation**
  - Tao Lu (Northeastern University); Gaochang Wu (Northeastern University)*
- **Multimodal Federated Learning for Personalized Clothing Recommendation**
  - Xinhui Yu (The University of British Columbia)*; Sophie Liu (The University of British Columbia); Chunhua Wu (The University of British Columbia)
- **Prototype Embedding Optimization for Human-Object Interaction Detection in Livestreaming**
  - Menghui Zhang (Beijing University of Technology); Jing Zhang (Beijing University of Technology)*; Lin Chen (Beijing University of Technology); Li Zhuo (Beijing University of Technology)
- **Efficient Generative Defect Synthesis for Industrial Anomaly Detection on MVTec AD**
  - Avinash Kumar Sharma (IIT Madras Zanzibar); Tushar Shinde (IIT Madras Zanzibar)*

- **HGS_OFAT: High-fidelity Gaussian SLAM based on Optical Flow Assisted Tracking**
  - Zhenyong Li (Shandong Normal University); Shanxin Zhang (Shandong Normal University)*; Yuxiang Liu (Shandong Normal University); Chuanfen Feng (Shandong Normal University); Hui Ji (Shandong Normal University); Jiande Sun (Shandong Normal University)
- **D3Net: Dual-Path Decoupling-Distillation for Adaptive Fusion in Continual Egocentric Learning**
  - Chenghao Qi (University of Electronic Science and Technology of China)*; Heqian Qiu (University of Electronic Science and Technology of China); Zhaofeng Shi (University of Electronic Science and Technology of China); Lanxiao Wang (University of Electronic Science and Technology of China); Hanwen Zhang (University of Electronic Science and Technology of China); Xinyu Chen (University of Electronic Science and Technology of China); Hongliang Li (University of Electronic Science and Technology of China)
- **Flexibly Constrained Tucker Decomposition for High-Order Spectral Analysis**
  - Fei He (University of Electronic Science and Technology of China); Houji Du (University of Electronic Science and Technology of China); Nipon Theera-Umpon (Chiang Mai University); Yipeng Liu (University of Electronic Science and Technology of China); Ce Zhu (University of Electronic Science and Technology of China)*
- **CG-SMFNet: Consensus-Guided Selective Multimodal Fusion for Weakly Supervised Temporal Action Localization**
  - Peng Liu (Beihang University)*; Zitai Jiang (Beihang University)
- **Blind Image Super-Resolution with Local and Global Dual-Guidance**
  - Yajun Qiu (University of Electronic Science and Technology of China); Shuyuan Zhu (University of Electronic Science and Technology of China)*; Lantao Yu (Adobe Inc.); Bing Zeng (University of Electronic Science and Technology of China)
- **Real-Time View Synthesis with Multiplane Image Network using Multimodal Supervision**
  - Manu Gond (Mid Sweden University)*; Mohammadreza Shamshirgarha (Mid Sweden University); Emin Zerman (Mid Sweden University); Sebastian Knorr (HTW Berlin - University of Applied Sciences); Mårten Sjöstrom (Mid Sweden University)
- **Rethinking Document Layout Analysis through Text Clustering via Multi-Modal Graph Convolution Networks**
  - Wenxi Li (Tsinghua University)*; Chenyang Lyu (Alibaba International Digital Commerce); Wei Ji (China Telecom); Liting Zhou (Dublin City University); Cathal Gurrin (Dublin City University); Yuchen Guo (Tsinghua University)
- **Dynamic Gaussian Streams for Volumetric Video via Codebook-Based Quantization**
  - Zhehao Shen (Shanghaitech)*; Yiwen Cai (Shanghaitech); Yuanji Lu (Shanghaitech); Yu Hong (Shanghaitech); Yize Wu (Shanghaitech); Meihan Zheng (Shanghaitech); Yingliang Zhang (DGene Inc.); Lan Xu (Shanghaitech)
- **Low Latency Immersive Visual Communication with Scalable Gaussian Splatting Coding**
  - Lingyu Shi (Hebei University); Jiaqi Zou (Beijing University of Posts and Telecommunications)*; Songlin Sun (Independent Contributor); Geert van der Auwer (Qualcomm Inc); Zhu Li (University of Missouri, Kansas City)

- **Tackling Re-buffering in Adaptive Video Streaming over Dynamic Networks: A Generative AI Approach**
  - Duc Nguyen (Tohoku Institute of Technology)*
- **Explicit Residual-Based Scalable Image Coding for Humans and Machines**
  - Yui Tatsumi (Waseda University)*; Ziyue Zeng (Waseda University); Hiroshi Watanabe (Waseda University)
- **Exploring Cross-Stage Adversarial Transferability in Class-Incremental Continual Learning**
  - Jungwoo Kim (Yonsei University); Jong-Seok Lee (Yonsei University)*
- **Secure protection of 3D content through reversible geometric deformation**
  - Khélian LARVET (Univ Montpellier)*; William PUECH (Univ Montpellier); Jean-Pierre PEDEBOY (STRATEGIES)
- **Carbon-Efficient Internet Video Streaming**
  - Zichen Zhu (Rutgers University); Tian Guo (Worcester Polytechnic Institute); Sheng Wei (Rutgers University)*
- **Meta Learning-based Multimodal Recommendation with Adaptive User Modality-Aware Preference Integration**
  - Zhenchao Wu (Renmin University of China)*; Hongteng Xu (Renmin University of China); Xu Chen (Renmin University of China)
- **Meta Learning for Adaptive Disentangled User Preference Integration Toward Multimodal Recommendation**
  - Zhenchao Wu (Renmin University of China)*; Hongteng Xu (Renmin University of China); Xu Chen (Renmin University of China)
- **FPGA Accelerated One-Sided Box Filter for Edge-Preserving Image Processing**
  - Yongfei Guo (Chinese Academy of Sciences)*; Xudong Niu (Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences); Chizhi Zhang (Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences); Yuanhao Gong (Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences)
- **Anderson Accelerated Residual Solver for Total Variation Models in Image Processing**
  - Yuanhao Gong (Chinese Academy of Sciences)*; Yongfei Guo (Chinese Academy of Sciences)
- **An Exploration of User Biometric Identification In XR Applications Based On User Head Movement**
  - Owen Dossett (Miami University); Ke Lyu (Miami University); Maohong Liao (Miami University); Han Li (Miami University); Xianglong Feng (Miami University)*
- **OrthCal: Synergizing Orthogonal Contrastive Learning and Prototype Calibration for Few-Shot Class-Incremental Learning**
  - Ruisong Dai (UESTC)*; Hanwen Zhang (UESTC); Xinyu Chen (UESTC); Chenghao Qi (UESTC); Heqian Qiu (UESTC); Hongliang Li (UESTC)

Keynote 2: Echocardiography-based Computer-Aided Diagnosis of Myocardial Infarction

Speaker: Prof. Moncef Gabbouj
September 21, 15:30-16:10

Coffee Break

September 21, 16:10-16:30

Oral Session 1: Multimedia Creation, Synthesis and Application

September 21, 16:30-18:30

- **CompBench: Benchmarking and Comparing Image Generation with Large Multimodal Models**
  - JiaRui Wang (Shanghai Jiao Tong University)*; Huiyu Duan (Shanghai Jiao Tong University); Yuke Xing (Shanghai Jiao Tong University); Yiling Xu (Shanghai Jiao Tong University); Guangtao Zhai (Shanghai Jiao Tong University); Xiongkuo Min (Shanghai Jiao Tong University)
- **HyperDiff: Hypergraph Guided Diffusion Model for 3D Human Pose Estimation**
  - Bing Han (Nanjing University of Aeronautics and Astronautics)*; Yuhua Huang (Nanjing University of Aeronautics and Astronautics); Pan Gao (Nanjing University of Aeronautics and Astronautics)
- **Guided Diffusion for the Extension of Machine Vision to Human Visual Perception**
  - Takahiro Shindo (Waseda University)*; Yui Tatsumi (Waseda University); Taiju Watanabe (Waseda University); Hiroshi Watanabe (Waseda University)
- **Sphere-GAN: a GAN-based approach for saliency estimation in 360° videos**
  - Mahmoud Z. A. Wahba (University of Padova); Sara Baldoni (University of Padova)*; Federica Battisti (University of Padova)
- **Infant Cry Detection In Noisy Environment Using Blueprint Separable Convolutions and Time-Frequency Recurrent Neural Network**
  - Haolin Yu (South China University of Technology); Yanxiong Li (School of Electronic and Information Engineering South China University of Technology)*
- **IdCo: Joint Identification and Contrastive Learning for Masked Face Recognition**
  - Qingtong Xu (University of Electronic Science and Technology of China)*; Chao Zhang (Sichuan Police College); Ao Li (University of Electronic Science and Technology of China); Xiaoning Liu (University of Electronic Science and Technology of China); Ce Zhu (University of Electronic Science and Technology of China)

Reception

September 21, 18:30-21:00

September 22, Monday

Keynote 3: Optical neural network: a rival to quantum computer?

Speaker: Prof. Alexander Lvovsky
September 22, 9:00-9:40

Dialogue: Igniting the AI Spark

September 22, 9:40-10:30

Coffee Break

September 22, 9:40-10:30

Oral Session 2: Multi-modal Processing, Analysis and Synthesis

September 22, 10:30-12:30

- **DFR: A Decompose-Fuse-Reconstruct Framework for Multi-Modal Few-Shot Segmentation**
  - Shuai Chen (University of Electronic Science and Technology of China); Fanman Meng (University of Electronic Science and Technology of China)*; Xiwei Zhang (University of Electronic Science and Technology of China); Haoran Wei (University of Electronic Science and Technology of China); Chenhao Wu (University of Electronic Science and Technology of China); Qingbo Wu (University of Electronic Science and Technology of China); Hongliang Li (University of Electronic Science and Technology of China)
- **Learning 3D mesh saliency from spiral patch features**
  - Olivier Lézoray (University of Caen Normandy)*; Anass Nouri (Ibn Tofail University)
- **SpatialGeo: Boosting Spatial Reasoning in Multimodal LLMs via Geometry-Semantics Fusion**
  - Jiajie Guo (Tongji University); Qingpeng Zhu (Independent Researcher); Jin Zeng (Tongji University)*; Xiaolong Wu (Tongji University); Changyong He (Tongji University); Weida Wang (Tongji University)
- **MGFT: Multi-Geometric Fusion Transformer for Robust Point Cloud Registration**
  - Yuxiang Liu (Shandong Normal University); Shanxin Zhang (Shandong Normal University)*; Zhenyong Li (Shandong Normal University); Chuanfen Feng (Shandong Normal University); Hui Ji (Shandong Normal University); Jiande Sun (Shandong Normal University)
- **Touch-Augmented Gaussian Splatting for Enhanced 3D Scene Reconstruction**
  - Yuchen Gao (Aarhus University)*; Xiao Xu (Technical University of Munich); Eckehard Steinbach (Technical University of Munich); Daniel Lucani (Aarhus University); Qi Zhang (Aarhus University)
- **Restore Anything Anywhere: Targeted Image Restoration with Object Segmentation and Text Guidance**

- Yen Ku Yeh (National Yang Ming Chiao Tung University); Chun Hao Yang (National Tsing Hua University)*; Kun Tai Wu (National Chengchi University); Yan Tsung Peng (National Chengchi University); Chun Rong Huang (National Yang Ming Chiao Tung University); Jun Cheng Chen (Academia Sinica)

Lunch Break

September 22, 12:30-14:00

TC Meeting

September 22, 12:30-14:00

Rising Star Session

September 22, 14:00-15:30

Keynote 4: Physics-Informed Machine Learning

Speaker: Prof. Aggelos K. Katsaggelos
September 22, 15:30-16:10

Coffee Break

September 22, 16:10-16:30

Oral Session 3: Multimedia Compression, Transmission and Security

September 22, 16:30-18:50

- **Reinforcement Learning-Based Dynamic Resource Allocation for Aerial 360-Degree Video VR Streaming**
  - Jacob Chakareski (NJIT)*; Lingdong Wang (UMass Amherst); Nicholas Mastronarde (University of Buffalo)
- **Learned Image Codec with Progressive Multi-Scale Probability Model for Streaming in Unreliable Communication Channels**
  - Honglei Zhang (Nokia Technologies)*; A. Burakhan Koyuncu (Nokia Technologies); Jukka Ahonen (Nokia Technologies); Nam Le (Nokia Technologies); Nannan Zou (Nokia Technologies); Francesco Cricri (Nokia Technologies)
- **Secure INN-based Steganography via Model Smoothing and Adversarial Attacks**
  - Weixiang Zhao (Sun Yat-Sen University); Fei Shang (Sun Yat-Sen University); Jin Li (Sun Yat-Sen University); Xiangui Kang (Sun Yat-Sen University)*; Jingyang Wen (North Minzu University); Z. Jane Wang (University of British Columbia)
- **NeRFCompressor: Enhancing Dynamic Scene Representation for Efficient 6-DoF Object Transportation**
  - Jin Zhou (George Mason University)*; Mufeng Zhu (Rutgers University); Yao Liu (Rutgers University); Songqing Chen (George Mason University)
- **Cross-Modal Thermal Image Compression via RGB Side Information**

- Sayush Maharjan (UMKC)*; Raghunath Puttagunta (UMKC); Zach Button (UMKC); Zhu Li (UMKC)

❖ **Structure-Preserving Patch Decoding for Efficient Neural Video Representation**
- Taiga Hayami (Waseda University)*; Kakeru Koizumi (Waseda University); Hiroshi Watanabe (Waseda University)

❖ **Task-Aware Optimized Color Image Demosaicing**
- Lei Xiong (University of Electronic Science and Technology of China); Zihao Wang (University of Electronic Science and Technology of China); Boyuan Zhang (University of Electronic Science and Technology of China); Feiyu Chen (University of Electronic Science and Technology of China); Shuyuan Zhu (University of Electronic Science and Technology of China)*; Bing Zeng (University of Electronic Science and Technology of China)

Banquet and Award

September 22, 19:30-21:00

September 23, Tuesday

Keynote 5: Spoken Language as a Window to Cognitive Health

Speaker: Prof. Helen Meng
September 23, 9:00-9:40

Coffee Break

September 23, 9:40-10:00

Oral Session 4: Multimedia Architecture Design, Systems and User Experience

September 23, 10:00-12:20

- **Subjective Visual Quality Assessment of Compressed Light Field Images: Learning-based vs. Conventional Methods**
  - Emin Zerman (Mid Sweden University)*; Soheib Takhtardeshir (Mid Sweden University); Anthony Trioux (Xidian University); Jianlong Qin (Xidian University); Wenjie Wu (Xidian University); Roger Olsson (Mid Sweden University); Mårten Sjöström (Mid Sweden University)
- **Real-Time Distortion Detection for PTZ Camera Systems**
  - Zhuobin Yuan (University of Cincinnati)*; Rui Dai (University of Cincinnati); Rayan Alghamdi (University of Cincinnati)
- **FPG-NAS: FLOPs-Aware Gated Differentiable Neural Architecture Search for Efficient 6DoF Pose Estimation**
  - Mohamed Nassim ALI OU SALAH (University of Luxembourg)*; Peyman ROSTAMI ABENDANSARI (University of Luxembourg); Anis Kacem (University of Luxembourg); Enjie Ghorbel (Cristal Laboratory, National School of Computer Sciences (ENSI), Manouba University); Emmanuel Koumandakis (Infinite Orbits); Djamila Aouada (University of Luxembourg)
- **DBAB: A Dual-Branch Adaptive Balance Framework with Optimized Plasticity Branch for Class-Incremental Learning**
  - Xinyu Chen (University of Electronic Science and Technology of China)*; Heqian Qiu (University of Electronic Science and Technology of China); Chenghao Qi (University of Electronic Science and Technology of China); Ruisong Dai (University of Electronic Science and Technology of China); Hongliang Li (University of Electronic Science and Technology of China)
- **Latent Space Stability vs. Perceptual Sensitivity: A Study of Visual Encoders under Distortion**
  - Abderrezzaq Sendjasni (University of Poitiers)*; Mohamed-Chaker Larabi (University of Poitiers)

- **EPINET-Lite: Rethinking Mixed Convolutions for Efficient Light Field Disparity Estimation Network**
  - Ali Hassan (Mid Sweden University); Tingting Zhang (Mid Sweden University); Karen Egiazarian (Tampere University); Mårten Sjöström (Mid Sweden University)*
- **S-LAM3D: Segmentation-Guided Monocular 3D Object Detection via Feature Space Fusion**
  - Diana Sas (Technical University of Cluj-Napoca)*; Florin Oniga (Technical University of Cluj-Napoca)

Lunch Break

September 23, 12:20-13:20

Tour (Great Wall of Badaling)

September 23, 13:20-18:00