

# **2025 IEEE International Workshop on Multimedia Signal Processing (MMSP 2025)**

**Beijing, China  
21-23 September 2025**



**IEEE Catalog Number: CFP25MSP-POD  
ISBN: 979-8-3315-9242-4**

**Copyright © 2025 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:	CFP25MSP-POD
ISBN (Print-On-Demand):	979-8-3315-9242-4
ISBN (Online):	979-8-3315-9241-7
ISSN:	2163-3517

**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](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

CURRAN ASSOCIATES INC.  
**proceedings**  
.com

# Contents

A01-01	Infant Cry Detection In Noisy Environment Using Blueprint Separable Convolutions and Time-Frequency Recurrent Neural Network	1
A01-02	Anderson Accelerated Residual Solver for Total Variation Models in Image Processing	7
A01-03	DFR: A Decompose-Fuse-Reconstruct Framework for Multi-Modal Few-Shot Segmentation	12
A01-04	Prototype Embedding Optimization for Human-Object Interaction Detection in Livestreaming	18
A01-05	SpatialGeo: Boosting Spatial Reasoning in Multimodal LLMs via Geometry-Semantics Fusion	24
A01-06	Learning 3D mesh saliency from spiral patch features	30
A01-07	Guided Diffusion for the Extension of Machine Vision to Human Visual Perception	36
A01-08	FPGA Accelerated One-Sided Box Filter for Edge-Preserving Image Processing	42
A01-09	Blind Image Super-Resolution with Local and Global Dual-Guidance	48
A01-10	CompBench: Benchmarking and Comparing Image Generation with Large Multimodal Models	54
A01-11	Real-Time View Synthesis with Multiplane Image Network using Multimodal Supervision	60
A01-12	Sphere-GAN: a GAN-based approach for saliency estimation in 360° videos	66
A01-13	HyperDiff: Hypergraph Guided Diffusion Model for 3D Human Pose Estimation	72
A01-14	Multimodal Federated Learning for Personalized Clothing Recommendation	78
A01-15	CG-SMFNet: Consensus-Guided Selective Multimodal Fusion for Weakly Supervised Temporal Action Localization	84
A01-16	Explicit Residual-Based Scalable Image Coding for Humans and Machines	90

A01-17	Subjective Visual Quality Assessment of Compressed Light Field Images: Learning-based vs. Conventional Methods	96
A01-18	Secure protection of 3D content through reversible geometric deformation	102
A01-19	Towards Volumetric Video: a Technical Overview of Immersive Media	108
A01-20	Secure INN-based Steganography via Model Smoothing and Adversarial Attacks	114
A01-21	EPINET-Lite: Rethinking Mixed Convolutions for Efficient Light Field Disparity Estimation Network	120
A01-22	Touch-Augmented Gaussian Splatting for Enhanced 3D Scene Reconstruction	126
A01-23	Real-Time Distortion Detection for PTZ Camera Systems	132
A01-24	Music Source Restoration	138
A01-25	IdCo: Joint Identification and Contrastive Learning for Masked Face Recognition	144
A01-26	Flexibly Constrained Tucker Decomposition for High-Order Spectral Analysis	150
A01-27	D3Net: Dual-Path Decoupling-Distillation for Adaptive Fusion in Continual Egocentric Learning	156
A01-28	FPG-NAS: FLOPs-Aware Gated Differentiable Neural Architecture Search for Efficient 6DoF Pose Estimation	162
A01-29	Lightweight DNN for Full-Band Speech Denoising on Mobile Devices: Exploiting Long and Short Temporal Patterns	168
A01-30	Meta Learning for Adaptive Disentangled User Preference Integration Toward Multimodal Recommendation	174
A01-31	Task-Aware Optimized Color Image Demosaicing	180
A01-32	Efficient Polyp Detection via Wavelet-Driven Boundary Enhancement and Temporal Consistency	186
A01-33	Exploring Cross-Stage Adversarial Transferability in Class-Incremental Continual Learning	192
A01-34	Restore Anything Anywhere: Targeted Image Restoration with Object Segmentation and Text Guidance	198
A01-35	OrthCal: Synergizing Orthogonal Contrastive Learning and Prototype Calibration for Few-Shot Class-Incremental Learning	204

A01-36	DBAB: A Dual-Branch Adaptive Balance Framework with Optimized Plasticity Branch for Class-Incremental Learning	210
A01-37	Structure-Preserving Patch Decoding for Efficient Neural Video Representation	216
A01-38	S-LAM3D: Segmentation-Guided Monocular 3D Object Detection via Feature Space Fusion	222
A01-39	PromptGS: Visual Prompting for Tiny Object Reconstruction in 3DGS Optimization	228
A01-40	An Exploration of User Biometric Identification In XR Applications Based On User Head Movement	234
A01-41	Data-independent Beamforming for End-to-end Multichannel Multi-speaker ASR	240
A01-42	Towards Low-Latency Tracking of Multiple Speakers With Short-Context Speaker Embeddings	246
A01-43	Efficient Generative Defect Synthesis for Industrial Anomaly Detection on MVTec AD	252
A01-44	Low Latency Immersive Visual Communication with Scalable Gaussian Splatting Coding	258
A01-45	White-box Differentiable Model of Perceived Localisation	264
A01-46	Tackling Re-buffering in Adaptive Video Streaming over Dynamic Networks: A Generative AI Approach	269
A01-47	Meta Learning-based Multimodal Recommendation with Adaptive User Modality-Aware Preference Integration	274
A01-48	Adapting Image-to-Video Diffusion Models for Large-Motion Frame Interpolation	280
A01-49	MGFT: Multi-Geometric Fusion Transformer for Robust Point Cloud Registration	286
A01-50	HGS_OFAT: High-fidelity Gaussian SLAM based on Optical Flow Assisted Tracking	292
A01-51	Learned Image Codec with Progressive Multi-Scale Probability Model for Streaming in Unreliable Communication Channels	298
A01-52	Carbon-Efficient Internet Video Streaming	304
A01-53	Frequency-Weighted Training Losses for Phoneme-Level DNN-based Speech Enhancement	310

A01-54	NeRFCompressor: Enhancing Dynamic Scene Representation for Efficient 6-DoF Object Transportation	316
A01-55	Cross-Modal Thermal Image Compression via RGB Side Information	322
A01-56	Reinforcement Learning-Based Dynamic Resource Allocation for Aerial 360° Video VR Streaming	328
A01-57	Latent Space Stability vs. Perceptual Sensitivity: A Study of Visual Encoders under Distortion	334
A01-58	Dynamic Gaussian Streams for Volumetric Video via Codebook-Based Quantization	340
A01-59	Lightweight Steel Surface Defect Detection via Knowledge Distillation	346
A01-60	Rethinking Document Layout Analysis through Text Clustering via Multi-Modal Graph Convolution Networks	352
A01-61	LSS3D: Learnable Spatial Shifting for Consistent and High-Quality 3D Generation from Single-Image	358
	Author Index	364