

Teaser Session M1

Paper id	Paper title
mmfp1843	Margin MCC: Chance-Robust Metric for Video Boundary Detection with Allowed Margin
mmfp1850	Hypergraph-Enhanced Hashing for Unsupervised Cross-Modal Retrieval via Robust Similarity Guidance
mmfp1940	Multimodal Color Recommendation in Vector Graphic Documents
mmfp2189	On the Importance of Spatial Relations for Few-shot Action Recognition
mmfp2506	Exploiting Time-Frequency Conformers For General Music Enhancement
mmfp2687	Multi-stage Factorized Spatio-Temporal Representation for RGB-D Action and Gesture Recognition
mmfp2945	MM-AU: Towards Multimodal understanding of advertisement videos
mmfp3035	Addressing Scalability for Real-time Multiuser Holo-portation: Introducing and Assessing a Multipoint Control Unit (MCU) for Volumetric Video
mmfp3062	Open-Vocabulary Object Detection via Scene Graph Discovery
mmfp3173	Chain of Propagation Prompting for Node Classification
mmfp3247	Peering into The Sketch: Ultra-Low Bitrate Face Compression for Joint Human and Machine Perception
mmfp3271	FDCNet: Feature Drift Compensation Network for Class-Incremental Weakly Supervised Object Localization
mmfp3330	ReCo: A Dataset for Residential Community Layout Planning
mmfp3617	Learning Event-Specific Localization Preferences for Audio-Visual Event Localization
mmfp3766	Unbalanced Multi-view Deep Learning
mmfp4068	Redundancy-aware Transformer for Video Question Answering
mmfp4104	That's What I Said: Fully-Controllable Talking Face Generation
mmfp4137	SEAR: Semantically-grounded Audio Representations
mmfp2370	Incomplete Multi-View Clustering with Regularized Hierarchical Graph
mmfp2517	Hierarchical Visual Attribute Learning in the Wild
mmfp2540	Self-Relational Graph Convolution Network for Skeleton-Based Action Recognition
mmfp2659	Cross-modality Representation Interactive Learning for Multimodal Sentiment Analysis
mmfp2723	Unsupervised Domain Adaptation for Video Object Grounding with Cascaded Debiasing Learning
mmfp3239	SCLAV: Supervised Cross-modal Contrastive Learning for Audio-Visual Coding
mmfp3292	MTSN: Multiscale Temporal Similarity Network for Temporal Action Localization
mmfp3371	Video-based Visible-Infrared Person Re-Identification via Style Disturbance Defense and Dual Interaction
mmfp3392	Cross-Silo Prototypical Calibration for Federated Learning with Non-IID Data
mmfp3639	Class-level Structural Relation Modeling and Smoothing for Visual Representation Learning
mmfp3892	Tile Classification Based Viewport Prediction with Multi-modal Fusion Transformer
mmfp3999	MIEP: Channel Pruning with Multi-granular Importance Estimation for Object Detection
mmfp4004	Long Short-Term Graph Memory Against Class-imbalanced Over-smoothing
mmfp2024	Reinforcement Graph Clustering with Unknown Cluster Number

All papers are presented also in the Poster Session M

Teaser Session M2

Paper id	Paper title
mmfp2487	TIVA-KG: A Multimodal Knowledge Graph with Text, Image, Video and Audio
mmfp1917	Co-Salient Object Detection with Semantic-Level Consensus Extraction and Dispersion
mmfp0138	Audio-Visual Spatial Integration and Recursive Attention for Robust Sound Source Localization
mmfp0187	Beyond Domain Gap: Exploiting Subjectivity in Sketch-Based Person Retrieval
mmfp0352	Concerto: Client-server Orchestration for Real-Time Video Analytics
mmfp0365	TwinStar: A Practical Multi-path Transmission Framework for Ultra-Low Latency Video Delivery
mmfp0859	Enhancing Visibility in Nighttime Haze Images Using Guided APSF and Gradient Adaptive Convolution
mmfp0895	Improving Scene Graph Generation with Superpixel-Based Interaction Learning
mmfp0922	Gaze Analysis System for Immersive 360 Video for Preservice Teacher Education
mmfp0991	Neural Image Popularity Assessment with Retrieval-augmented Transformer
mmfp1082	Layout Sequence Prediction From Noisy Mobile Modality
mmfp1323	Cross-Lingual Transfer of Large Language Model by Visually-Derived Supervision Toward Low-Resource Languages
mmfp1378	PetalView: Fine-grained Location and Orientation Extraction of Street-view Images via Cross-view Local Search
mmfp1638	Scale-space Tokenization for Improving the Robustness of Vision Transformers
mmfp1713	MSECNet: Accurate and Robust Normal Estimation for 3D Point Clouds by Multi-Scale Edge Conditioning
mmfp1747	Enhancing Fake News Detection in Social Media via Label Propagation on Cross-modal Tweet Graph
mmfp1767	Object Detection Difficulty: Suppressing Over-aggregation for Faster and Better Video Object Detection
mmfp1772	Event-guided Frame Interpolation and Dynamic Range Expansion of Single Rolling Shutter Image
mmfp1837	GCL: Gradient-Guided Contrastive Learning for Medical Image Segmentation with Multi-Perspective Meta Labels
mmfp0137	Suspected Objects Matter: Rethinking Model's Prediction for One-stage Visual Grounding
mmfp0215	SemanticRT: A Large-Scale Dataset and Method for Robust Semantic Segmentation in Multispectral Images
mmfp0230	A Closer Look at Classifier in Adversarial Domain Generalization
mmfp0259	MVFlow: Deep Optical Flow Estimation of Compressed Videos with Motion Vector Prior
mmfp0821	Depth-aided Camouflaged Object Detection
mmfp0883	Synthesizing Long-Term Human Motions with Diffusion Models via Coherent Sampling
mmfp1399	UER: A Heuristic Bias Addressing Approach for Online Continual Learning
mmfp1560	CPU: Codebook Lookup Transformer with Knowledge Distillation for Point Cloud Upsampling
mmfp1584	Parsing is All You Need for Accurate Gait Recognition in the Wild
mmfp2205	Propagation is All You Need: A New Framework for Representation Learning and Classifier Training on Graphs
mmfp2369	Unambiguous Object Tracking by Exploiting Target Cues

All papers are presented also in the Poster Session M