

## Offline Session W1

Paper id	Paper title
mmfp2069	PEARL: Preprocessing Enhanced Adversarial Robust Learning of Image Deraining for Semantic Segmentation
mmfp2074	A Novel Temporal Channel Enhancement and Contextual Excavation Network for Temporal Action Localization
mmfp2124	Karma: Adaptive Video Streaming via Causal Sequence Modeling
mmfp2217	Toward Zero-shot Character Recognition: A Gold Standard Dataset with Radical-level Annotations
mmfp2219	Rethinking Neural Style Transfer: Generating Personalized and Watermarked Stylized Images
mmfp2220	Interactive Image Style Transfer Guided by Graffiti
mmfp2271	High Fidelity Face Swapping via Semantics Disentanglement and Structure Enhancement
mmfp2279	FedCE: Personalized Federated Learning Method based on Clustering Ensembles
mmfp2330	DuDoINet: Dual-Domain Implicit Network for Multi-Modality MR Image Arbitrary-scale Super-Resolution
mmfp2338	Points-to-3D: Bridging the Gap between Sparse Points and Shape-Controllable Text-to-3D Generation
mmfp2389	Scene-Generalizable Interactive Segmentation of Radiance Fields
mmfp2462	Resolve Domain Conflicts for Generalizable Remote Physiological Measurement
mmfp2496	Robust Image Steganography against General Scaling Attacks
mmfp2507	A Novel Deep Video Watermarking Framework with Enhanced Robustness to H.264/AVC Compression
mmfp2512	Personalized Single Image Reflection Removal Network through Adaptive Cascade Refinement
mmfp2525	Two-stage Content-Aware Layout Generation for Poster Designs
mmfp2536	Lite-MKD: A Multi-modal Knowledge Distillation Framework for Lightweight Few-shot Action Recognition
mmfp2558	Graph Convolutional Incomplete Multi-modal Hashing
mmfp2627	Progressive Spatio-temporal Perception for Audio-Visual Question Answering
mmfp2660	ROAD: Robust Unsupervised Domain Adaptation with Noisy Labels
mmfp2690	Secondary Labeling: A Novel Labeling Strategy for Image Manipulation Detection
mmfp2698	Digging into Depth Priors for Outdoor Neural Radiance Fields
mmfp2718	Echoes: Unsupervised Debiasing via Pseudo-bias Labeling in an Echo Chamber
mmfp2724	A Hardware-efficient Unified Motion Estimation for Video Coding
mmfp2726	SUG: Single-dataset Unified Generalization for 3D Point Cloud Classification
mmfp2751	RAIRNet: Region-Aware Identity Rectification for Face Forgery Detection
mmfp2753	A Generalized Physical-knowledge-guided Dynamic Model for Underwater Image Enhancement
mmfp2812	G <sup>S</sup> -DUN: Gradient Guided Deep Unfolding Network for Image Compressive Sensing
mmfp2814	SIFDetectCracker: An Adversarial Attack Against Fake Voice Detection Based on Speaker-Irrelative Features
mmfp2924	3DStyle-Diffusion: Pursuing Fine-grained Text-driven 3D Stylization with 2D Diffusion Models
mmfp2942	Collaborative Fraud Detection: How Collaboration Impacts Fraud Detection
mmfp2949	Audio-Visual Segmentation by Exploring Cross-Modal Mutual Semantics
mmfp2996	Locate and Verify: A Two-Stream Network for Improved Deepfake Detection
mmfp3034	My Brother Helps Me: Node Injection Based Adversarial Attack on Social Bot Detection
mmfp3052	High-Order Tensor Recovery Coupling Multilayer Subspace Priors With Application in Video Restoration
mmfp3060	MV-Diffusion: Motion-aware Video Diffusion Model
mmfp3061	Efficiency-Optimized Lightweight Video Diffusion Models
mmfp3115	Semantic-aware Consistency Network for Cloth-changing Person Re-identification
mmfp3137	Dynamic View Synthesis with Spatio-Temporal Feature Warping from Sparse Views
mmfp3152	TikTalk: A Video-Based Dialogue Dataset for Multi-Modal Chitchat in Real World
mmfp3193	MUP: Multi-granularity Unified Perception for Panoramic Activity Recognition
mmfp3194	Fearless Luminance Adaptation: a Macro-Micro-Hierarchical Transformer for Exposure Correction
mmfp3207	Designing Loving-Kindness Meditation in Virtual Reality for Long-Distance Romantic Relationships
mmfp3285	IN/ACTIVE: A Distance-Technology-Mediated Stage for Performer-Audience Telepresence and Environmental Control
mmfp3311	LDRM: Degradation Rectify Model for Low-light Imaging via Color-Monochrome Cameras
mmfp3358	Physical Invisible Backdoor Based on Camera Imaging
mmfp3360	Learning Intra and Inter-Camera Invariance for Isolated Camera Supervised Person Re-identification
mmfp3361	Hierarchical Masked 3D Diffusion Model for Video Outpainting
mmfp3457	STIRER: A Unified Model for Low-Resolution Scene Text Image Recovery and Recognition
mmfp3558	Localization-assisted Uncertainty Score Disentanglement Network for Action Quality Assessment
mmfp3592	Improving Rumor Detection by Class-based Adversarial Domain Adaptation
mmfp3615	Event-based Motion Deblurring with Modality-Aware Decomposition and Recomposition
mmfp3665	StegaDDPM: Generative Image Steganography based on Denoising Diffusion Probabilistic Model
mmfp3676	Deep Neural Network Watermarking against Model Extraction Attack
mmfp3739	ControlStyle: Text-Driven Stylized Image Generation Using Diffusion Priors
mmfp3773	High Visual-Fidelity Learned Video Compression
mmfp3832	Factorized Omnidirectional Representation based Vision GNN for Anisotropic 3D Multimodal MR Image Segmentation
mmfp3922	HCSN-Net: Single Image Desnowing with Hybrid Color Space Transformation
mmfp3926	SSPU-Net: A Structure Sensitive Point Cloud Upsampling Network with Multi-Scale Spatial Refinement
mmfp4010	Entropy-based Optimization on Individual and Global Predictions for Semi-Supervised Learning
mmfp4051	MALIANG: Handwritten Chemical Structure Image to Structure-Specific Markup Using Random Conditional Guided Decoder
mmfp4077	Digital Twins Fuzzy System Based on Time Series Forecasting Model LFTformer
mmfp4088	Augmented Digital Twins for Predictive Automatic Regulation and Fault Alarm in Sewage Plan
mmfp4132	PI-NeRF: A Partial-Invertible Neural Radiance Fields for Pose Estimation
mmfp4204	Towards Fast and Stable Federated Learning: Confronting Heterogeneity via Knowledge Anchor.
mmfp1149	Towards Real-Time Neural Video Codec for Cross-Platform Application Using Calibration Information
mmfp1921	Cuing Without Sharing: A Federated Cued Speech Recognition Framework via Mutual Knowledge Distillation
mmfp1952	Universal Defensive Underpainting Patch: Making Your Text Invisible to Optical Character Recognition
mmfp1960	WRAP: Watermarking Approach Robust Against Film-coating upon Printed Photographs

mmfp1997 End-to-end XY Separation for Single Image Blind Deblurring  
mmfp2031 Moby: Empowering 2D Models for Efficient Point Cloud Analytics on the Edge  
mmfp2081 Non-Local Geometry and Color Gradient Aggregation Graph Model for No-Reference Point Cloud Quality Assessment  
mmfp2135 Deconfounded Multimodal Learning for Spatio-temporal Video Grounding  
mmfp2210 A Multiple Prediction Mechanisms Ensemble for Complex Remote Sensing Scenes  
mmfp2211 Video Inverse Tone Mapping Network with Luma and Chroma Mapping  
mmfp2223 S3DS: Self-supervised Learning of 3D Skeletons from Single View Images  
mmfp2232 Autistic Spectrum Disorders Diagnose with Graph Neural Networks  
mmfp2263 Cross-view Resolution and Frame Rate Joint Enhancement for Binocular Video  
mmfp2281 Automatic Human Scene Interaction through Contact Estimation and Motion Adaptation  
mmfp2285 Face-Driven Zero-Shot Voice Conversion with Memory-based Face-Voice Alignment  
mmfp2299 Generalized Universal Domain Adaptation with Generative Flow Networks  
mmfp2314 Recurrent Self-Supervised Video Denoising with Denser Receptive Field  
mmfp2402 CLG-INet: Coupled Local-Global Interactive Network for Image Restoration  
mmfp2412 Taming the Power of Diffusion Models for High-Quality Virtual Try-On with Appearance Flow  
mmfp2436 Semi-Supervised Panoptic Narrative Grounding  
mmfp2693 Multi-Part Token Transformer with Dual Contrastive Learning for Fine-grained Image Classification  
mmfp2695 Reinforcement Learning-based Adversarial Attacks on Object Detectors using Reward Shaping  
mmfp2696 Progressive domain-style translation for Nighttime Tracking  
mmfp2703 DiffDance: Cascaded Human Motion Diffusion Model for Dance Generation  
mmfp2706 The Silent Manipulator: A Practical and Inaudible Backdoor Attack against Speech Recognition Systems  
mmfp2730 Benign Shortcut for Debiasing: Fair Visual Recognition via Intervention with Shortcut Features  
mmfp2787 Do We Need Model Fine-tuning? Towards an Efficient and Scalable Watermarking Scheme for Deep Neural Networks  
mmfp2790 Who is Speaking Actually? Robust and Versatile Speaker Traceability for Voice Conversion  
mmfp2822 Spatio-Temporal Catcher: a Self-Supervised Transformer for Deepfake Video Detection  
mmfp2854 Bridging Trustworthiness and Open-World Learning: An Exploratory Neural Approach for Enhancing Interpretability, Generalization, and Robustness  
mmfp2955 Toward Intelligent Interactive Design: A Generation Framework Based on Cross-domain Fashion Elements  
mmfp3019 Learning Dynamic Point Cloud Compression via Hierarchical Inter-frame Block Matching  
mmfp3021 Faster Video Moment Retrieval with Point-Level Supervision  
mmfp3037 Automatic Asymmetric Embedding Cost Learning via Generative Adversarial Networks  
mmfp3113 Model-Contrastive Learning for Backdoor Elimination  
mmfp3155 Combating Online Misinformation Videos: Characterization, Detection, and Future Directions  
mmfp3191 Toward Human Perception-Centric Video Thumbnail Generation  
mmfp3235 Predictive Sampling for Efficient Pairwise Subjective Image Quality Assessment  
mmfp3257 Flexible and Secure Watermarking for Latent Diffusion Model  
mmfp3314 SIEGE: Self-Supervised Incremental Deep Graph Learning for Ethereum Phishing Scam Detection  
mmfp3340 SepMark: Deep Separable Watermarking for Unified Source Tracing and Deepfake Detection  
mmfp3343 Exploiting Fine-Grained DCT Representations for Hiding Image-Level Messages within JPEG Images  
mmfp3417 SDDNet: Style-guided Dual-layer Disentanglement Network for Shadow Detection  
mmfp3626 360-Degree Panorama Generation from Few Unregistered NFOV Images  
mmfp3721 Filling in the Blank: Rationale-Augmented Prompt Tuning for TextVQA  
mmfp3722 When Masked Image Modeling Meets Source-free Unsupervised Domain Adaptation: Dual-Level Masked Network for Semantic Segmentation  
mmfp3806 OraclePoints: A Hybrid Neural Representation for Oracle Character  
mmfp3840 CoCa: A Connectivity-Aware Cascade Framework for Histology Gland Segmentation  
mmfp3950 Data-Scarce Animal Face Alignment via Bi-Directional Cross-Species Knowledge Transfer  
mmfp3969 A Model-Agnostic Semantic-Quality Compatible Framework based on Self-Supervised Semantic Decoupling  
mmfp3982 AbCoRD: Exploiting multimodal generative approach for Aspect-based Complaint and Rationale Detection  
mmfp4103 Edge-Assisted On-Device Model Update for Video Analytics in Adverse Environments

## Offline Session W2

Paper id	Paper title
mmfp1081	Follow-me: Deceiving Trackers with Fabricated Paths
mmfp0021	MLIC: Multi-Reference Entropy Model for Learned Image Compression
mmfp0031	Sample Less, Learn More: Efficient Action Recognition via Frame Feature Restoration
mmfp0036	Prompted Contrast with Masked Motion Modeling: Towards Versatile 3D Action Representation Learning
mmfp0039	DAWN: Direction-aware Attention Wavelet Network for Image Deraining
mmfp0093	Cross-Architecture Distillation for Face Recognition
mmfp0136	LHAct: Rectifying Extremely Low and High Activations for Out-of-Distribution Detection
mmfp0158	WaterFlow: Heuristic Normalizing Flow for Underwater Image Enhancement and Beyond
mmfp0180	DiffBFR: Bootstrapping Diffusion Model for Blind Face Restoration
mmfp0243	Active CT Reconstruction with a Learned Sampling Policy
mmfp0251	Hierarchical Dynamic Image Harmonization
mmfp0286	Separable Modulation Network for Efficient Image Super-Resolution
mmfp0298	S-OmniMVS: Incorporating Sphere Geometry into Omnidirectional Stereo Matching
mmfp0300	Learning Spectral-wise Correlation for Spectral Super-Resolution: Where Similarity Meets Particularity
mmfp0330	Context-Aware Talking-Head Video Editing
mmfp0340	Stroke-based Neural Painting and Stylization with Dynamically Predicted Painting Region
mmfp0429	Ada-DQA: Adaptive Diverse Quality-aware Feature Acquisition for Video Quality Assessment
mmfp0557	TSSAT: Two-Stage Statistics-Aware Transformation for Artistic Style Transfer
mmfp0575	Learning Discriminative Feature Representation for Open Set Action Recognition
mmfp0588	Exploring the Adversarial Robustness of Video Object Segmentation via One-shot Adversarial Attacks
mmfp0589	Towards Decision-based Sparse Attacks on Video Recognition
mmfp0594	Recurrent Spike-based Image Restoration under General Illumination
mmfp0617	Controllable Face Sketch-Photo Synthesis with Flexible Generative Priors
mmfp0632	Learning Non-Uniform-Sampling for Ultra-High-Definition Image Enhancement
mmfp0637	Feature Decoupling-Recycling Network for Fast Interactive Segmentation
mmfp0698	Improving Anomaly Segmentation with Multi-Granularity Cross-Domain Alignment
mmfp0723	CPNet: Cartoon Parsing with Pixel and Part Correlation
mmfp0785	Self-Reference Image Super-Resolution via Pre-trained Diffusion Large Model and Window Adjustable Transformer
mmfp0830	EasyNet: An Easy Network for 3D Industrial Anomaly Detection
mmfp0852	Restoration of Multiple Image Distortions using a Semi-dynamic Deep Neural Network
mmfp0896	JAVP: Joint-Aware Video Processing with Edge-Cloud Collaboration for DNN Inference
mmfp0901	Uncovering the Unseen: Discover Hidden Intentions by Micro-Behavior Graph Reasoning
mmfp0910	Adaptive Feature Swapping for Unsupervised Domain Adaptation
mmfp0912	Synthesizing Videos from Images for Image-to-Video Adaptation
mmfp0932	Orthogonal Temporal Interpolation for Zero-Shot Video Recognition
mmfp0952	Moiré Backdoor Attack (MBA): A Novel Trigger for Pedestrian Detectors in the Physical World
mmfp0973	A Baseline Investigation: Transformer-based Cross-view Baseline for Text-based Person Search
mmfp1017	Disentangle Propagation and Restoration for Efficient Video Recovery
mmfp1036	TextPainter: Multimodal Text Image Generation with Visual-harmony and Text-comprehension for Poster Design
mmfp1037	AutoPoster: A Highly Automatic and Content-aware Design System for Advertising Poster Generation
mmfp1061	FastLLVE: Real-Time Low-Light Video Enhancement with Intensity-Aware Look-Up Table
mmfp1177	RecolorNeRF: Layer Decomposed Radiance Fields for Efficient Color Editing of 3D Scenes
mmfp1199	GridFormer: Towards Accurate Table Structure Recognition via Grid Prediction
mmfp1202	Decoupled Cross-Scale Cross-View Interaction for Stereo Image Enhancement in the Dark
mmfp1212	FedCD: A Classifier Debaised Federated Learning Framework for Non-IID Data
mmfp1247	A Capture to Registration Framework for Realistic Image Super-Resolution in the Industry Environment
mmfp1258	SEAM: Searching Transferable Mixed-Precision Quantization Policy through Large Margin Regularization
mmfp1436	Ada3Diff: Defending against 3D Adversarial Point Clouds via Adaptive Diffusion
mmfp1491	Make-It-4D: Synthesizing a Consistent Long-Term Dynamic Scene Video from a Single Image
mmfp1554	Adversarial Bootstrapped Question Representation Learning for Knowledge Tracing
mmfp1557	Dance with You: The Diversity Controllable Dancer Generation via Diffusion Models.
mmfp1625	SMM:Self-supervised Multi-Illumination Color Constancy Model with Multiple Pretext Tasks
mmfp1666	Normality Learning-based Graph Anomaly Detection via Multi-Scale Contrastive Learning
mmfp1679	DeNoL: A Few-Shot-Sample-Based Decoupling Noise Layer for Cross-channel Watermarking Robustness
mmfp1698	ProTegO: Protect Text Content against OCR Extraction Attack
mmfp1701	Unsupervised Visible-Infrared Person ReID by Collaborative Learning with Neighbor-Guided Label Refinement
mmfp1794	Efficient Hierarchical Multi-view Fusion Transformer for 3D Human Pose Estimation
mmfp1797	Implicit Obstacle Map-driven indoor navigation model for robust obstacle avoidance
mmfp1860	Visual Redundancy Removal of Composite Images via Multimodal Learning
mmfp1878	MFR-Net: Multi-faceted Responsive Listening Head Generation via Denoising Diffusion Model
mmfp1897	Unfolding Once is Enough: A Deployment-Friendly Transformer Unit for Super-Resolution
mmfp1962	Slowfast Diversity-aware Prototype Learning for Egocentric Action Recognition
mmfp1963	CLE Diffusion: Controllable Light Enhancement Diffusion Model
mmfp1999	UniSinger: Unified End-to-End Singing Voice Synthesis With Cross-Modality Information Matching
mmfp2013	An Intelligent Learning Approach to Achieve Near-Second Low-Latency Live Video Streaming under Highly Fluctuating Networks
mmfp2051	Relit-NeuLF: Efficient Relighting and Novel View Synthesis via Neural 4D Light Field

mmfp0106 Exploring Shape Embedding for Cloth-Changing Person Re-Identification via 2D-3D Correspondences  
mmfp0111 SD-Net: Spatially-Disentangled Point Cloud Completion Network  
mmfp0217 CHAIN: Exploring Global-Local Spatio-Temporal Information for Improved Self-Supervised Video Hashing  
mmfp0252 Up to Thousands-fold Storage Saving: Towards Efficient Data-Free Distillation of Large-Scale Visual Classifiers  
mmfp0316 Improving Few-shot Image Generation by Structural Discrimination and Textural Modulation  
mmfp0357 CoP: Chain-of-Pose for Image Animation in Large Pose Changes  
mmfp0361 Speech-Driven 3D Face Animation with Composite and Regional Facial Movements  
mmfp0395 Incremental Few Shot Semantic Segmentation via Class-agnostic Mask Proposal and Language-driven Classifier  
mmfp0471 Style-Controllable Generalized Person Re-identification  
mmfp0475 SimulFlow: Simultaneously Extracting Feature and Identifying Target for Unsupervised Video Object Segmentation  
mmfp0523 SimHMR: A Simple Query-based Framework for Parameterized Human Mesh Reconstruction  
mmfp0569 SeeDS: Semantic Separable Diffusion Synthesizer for Zero-shot Food Detection  
mmfp0596 Fine-Grained Music Plagiarism Detection: Revealing Plagiarists through Bipartite Graph Matching and a Comprehensive Large-Scale Dataset  
mmfp0613 Unlearnable Examples Give a False Sense of Security: Piercing through Unexploitable Data with Learnable Examples  
mmfp0671 When Measures are Unreliable: Imperceptible Adversarial Perturbations toward Top- $\$k$  Multi-Label Learning  
mmfp0718 Open-Scenario Domain Adaptive Object Detection in Autonomous Driving  
mmfp0745 Bilevel Generative Learning for Low-Light Vision  
mmfp0766  $\$IS^2\$Net$ : Intra-domain Semantic and Inter-domain Style Enhancement for Semi-supervised Medical Domain Generalization  
mmfp0786 Degeneration-Tuning: Using Scrambled Grid shield Unwanted Concepts from Stable Diffusion  
mmfp0827 Hardware-friendly Scalable Image Super Resolution with Progressive Structured Sparsity  
mmfp0942 Kernel Dimension Matters: to Activate Available Kernels for Real-time Video Super-Resolution  
mmfp1004 Securing Fixed Neural Network Steganography  
mmfp1008 A lightweight collective-attention network for change detection  
mmfp1086 Taking a Part for the Whole: An Archetype-agnostic Framework for Voice-Face Association  
mmfp1101 Personalized Image Aesthetics Assessment with Attribute-guided Fine-grained Feature Representation  
mmfp1135 ParliRobo: Participant Lightweight AI Robots for Massively Multiplayer Online Games (MMOGs)  
mmfp1178 FedVQA: Personalized Federated Visual Question Answering over Heterogeneous Scenes  
mmfp1207 UMMAFormer: A Universal Multimodal-adaptive Transformer Framework for Temporal Forgery Localization  
mmfp1294 Recurrent Multi-scale Transformer for High-resolution Salient Object Detection  
mmfp1365 Prototypical Cross-domain Knowledge Transfer for Cervical Dysplasia Visual Inspection  
mmfp1420 Practical Deep Dispersed Watermarking with Synchronization and Fusion  
mmfp1466 BiFPro: A Bidirectional Facial-data Protection Framework against DeepFake  
mmfp1488 PatchBackdoor: Backdoor Attack against Deep Neural Networks without Model Modification  
mmfp1505 FCBoost-Net: A Generative Network for Synthesizing Multiple Collocated Outfits via Fashion Compatibility Boosting  
mmfp1556 IDDR-NGP: Incorporating Detectors for Distractors Removal with Instant Neural Radiance Field  
mmfp1600 Recognizing High-Speed Moving Objects with Spike Camera  
mmfp1611 Rethinking Neighborhood Consistency Learning on Unsupervised Domain Adaptation  
mmfp1628 AniPixel: Towards Animatable Pixel-Aligned Human Avatar  
mmfp1669 Versatile Face Animator: Driving Arbitrary 3D Facial Avatar in RGBD Space  
mmfp1690 Efficient Bilateral Cross-Modality Cluster Matching for Unsupervised Visible-Infrared Person ReID  
mmfp1696 On Physically Occluded Fake Identity Document Detection  
mmfp1729 Frequency Perception Network for Camouflaged Object Detection  
mmfp1758 LocalPose: Object Pose Estimation with Local Geometry Guidance  
mmfp1820 Text-to-Image Diffusion Models can be Easily Backdoored through Multimodal Data Poisoning  
mmfp1823 MCUNeRF: Packing NeRF into an MCU with 1MB Memory  
mmfp1889 RD-FGFS: A Rule-Data Hybrid Framework for Fine-grained Footstep Sound Synthesis from Visual Guidance  
mmfp1890 Multi-scale Target-Aware Framework for Constrained Splicing Detection and Localization  
mmfp1895 FashionDiff: A Controllable Diffusion Model Using Pairwise Fashion Elements for Intelligent Design  
mmfp1899 InspirNET: An Unsupervised Generative Adversarial Network with Controllable Fine-grained Texture Disentanglement for Fashion Generation