mmfp2670	Mutual Information-driven Triple Interaction Network for Efficient Image Dehazing
mmfp0137	Suspected Objects Matter: Rethinking Model's Prediction for One-stage Visual Grounding
mmfp2540	Self-Relational Graph Convolution Network for Skeleton-Based Action Recognition
mmfp0391	Exploring Correlations in Degraded Spatial Identity Features for Blind Face Restoration
mmfp3371	Video-based Visible-Infrared Person Re-Identification via Style Disturbance Defense and Dual Interaction
mmfp1378	PetalView: Fine-grained Location and Orientation Extraction of Street-view Images via Cross-view Local Searc
mmfp2099	Shifted GCN-GAT and Cumulative-Transformer based Social Relation Recognition for Long Videos
mmfp1880	Causal Intervention for Sparse-View Gait Recognition
mmfp2945	MM-AU:Towards Multimodal understanding of advertisement videos
mmfp1399	UER: A Heuristic Bias Addressing Approach for Online Continual Learning
mmfp1281	Clip Fusion with Bi-level Optimization for Human Mesh Reconstruction from Monocular Videos
mmfp1584	Parsing is All You Need for Accurate Gait Recognition in the Wild
mmfp3890	Multi-Scale Similarity Aggregation for Dynamic Metric Learning
mmfp0815	RefineTAD: Learning Proposal-free Refinement for Temporal Action Detection
mmfp1368	Video Infringement Detection via Feature Disentanglement and Mutual Information Maximization
mmfp1143	Pseudo Object Replay and Mining for Incremental Object Detection
mmfp1171	Informative Classes Matter: Towards Unsupervised Domain Adaptive Nighttime Semantic Segmentation
mmfp1502	View while Moving: Efficient Video Recognition in Long-untrimmed Videos
mmfp3046	PMVC: Data Augmentation-Based Prosody Modeling for Expressive Voice Conversion
mmfp0458	Alleviating Spatial Misalignment and Motion Interference for UAV-based Video Recognition
mmfp3020	Learning Causality-inspired Representation Consistency for Video Anomaly Detection
mmfp1935	M2ATS: A Real-world Multimodal Air Traffic Situation Benchmark Dataset and Beyond
mmfp1298	Federated Learning with Label-Masking Distillation
mmfp3280	Painterly Image Harmonization using Diffusion Model
mmfp0433	Exploring Hyperspectral Histopathology Image Segmentation from A Deformable Perspective
mmfp3810	Uncertainty-Aware Variate Decomposition for Self-supervised Blind Image Deblurring
mmfp3239	SCLAV: Supervised Cross-modal Contrastive Learning for Audio-Visual Coding
mmfp2780	Cross-Modal and Multi-Attribute Face Recognition: A Benchmark
mmfp0230	A Closer Look at Classifier in Adversarial Domain Generalization
mmfp0814	Mixture-of-Experts Learner for Single Long-Tailed Domain Generalization
mmfp0403	Robust Spectral Embedding Completion Based Incomplete Multi-view Clustering
mmfp2461	SA-GDA: Spectral Augmentation for Graph Domain Adaptation
mmfp0504	CONVERT:Contrastive Graph Clustering with Reliable Augmentation
mmfp0188	High-order Complementarity Induced Fast Multi-View Clustering with Enhanced Tensor Rank Minimization
mmfp1137	DealMVC: Dual Contrastive Calibration for Multi-view Clustering
mmfp2178	Bidomain Modeling Paradigm for Pansharpening
mmfp1083	Learning High-frequency Feature Enhancement and Alignment for Pan-sharpening
mmfp3429	Distribution Consistency based Fast Anchor Imputation for Incomplete Multi-view Clustering
mmfp0819	Visual Causal Scene Refinement for Video Question Answering
mmfp1090	Parameter-Efficient Transfer Learning for Audio-Visual-Language Tasks
mmfp3330	ReCo: A Dataset for Residential Community Layout Planning
mmfp1293	Point-aware Interaction and CNN-induced Refinement Network for RGB-D salient object detection
mmfp2354	Multi-view Self-Expressive Subspace Clustering Network
mmfp2659	Cross-modality Representation Interactive Learning for Multimodal Sentiment Analysis
mmfp1558	Entropy Neural Estimation for Graph Contrastive Learning
mmfp3356	Cross-modal and Cross-medium Adversarial Attack for Audio
mmfp1240	Unsupervised Multiplex Graph learning with Complementary and Consistent Information

mmfp1837 GCL: Gradient-Guided Contrastive Learning for Medical Image Segmentation with Multi-Perspective Meta Lab
mmfp1376 Multi-Spectral Image Stitching via Spatial Graph Reasoning
mmfp2205 Propagation is All You Need: A New Framework for Representation Learning and Classifier Training on Graph
mmfp1408 Cross-modal Unsupervised Domain Adaptation for 3D Semantic Segmentation via Bidirectional Fusion-then-D
mmfp1254 AcFormer: An Aligned and Compact Transformer for Multimodal Sentiment Analysis
mmfp1246 Freq-HD: an Interpretable Frequency-based High-Dynamics Affective Clip Selection Method for in-the-Wild Fa
mmfp1528 StyleEDL: Style-Guided High-order Attention Network for Image Emotion Distribution Learning
mmfp3888 Variance-Aware Bi-Attention Expression Transformer for Open-Set Facial Expression Recognition in the Wild
mmfp3229 AffectFAL: Federated Active Affective Computing with Non-IID data
mmfp2242 ASTDF-Net: Attention-Based Spatial-Temporal Dual-Stream Fusion Network for EEG-Based Emotion Recogni
mmfp2287 M^{3}Net: Multi-view Encoding, Matching, and Fusion for Few-shot Fine-grained Action Recognition
mmfp0099 CUCL: Codebook for Unsupervised Continual Learning
mmfp1812 Regress Before Construct: Regress Autoencoder for Point Cloud Self-supervised Learning
mmfp2385 CropCap: Embedding Visual Cross-Partition Dependency for Image Captioning
mmfp1801 Generalizing Face Forgery Detection via Uncertainty Learning
mmfp1767 Object Detection Difficulty: Suppressing Over-aggregation for Faster and Better Video Object Detection
mmfp2450 Mutual-guided Dynamic Network for Image Fusion
mmfp0358 Frequency Representation Integration for Camouflaged Object Detection
mmfp1313 DecenterNet: Bottom-Up Human Pose Estimation Via Decentralized Pose Representation
mmfp0895 Improving Scene Graph Generation with Superpixel-Based Interaction Learning
mmfp1649 Lifelong Scene Text Recognizer via Expert Modules
mmfp1643 CoMoSpeech: One-Step Speech and Singing Voice Synthesis via Consistency Model
mmfp2006 Exploring Motion Cues for Video Test-Time Adaptation
mmfp2985 Perceiving Ambiguity and Semantics without Recognition: An Efficient and Effective Ambiguous Scene Text De
mmfp1327 Single-stage Multi-human Parsing via Point Sets and Center-based Offsets
mmfp0507 Partitioned Saliency Ranking with Dense Pyramid Transformers
mmfp1716 CenterLPS: Segment Instances by Centers for LiDAR Panoptic Segmentation
mmfp2594 Boosting Few-shot 3D Point Cloud Segmentation via Query-Guided Enhancement
mmfp0084 PiPa: Pixel- and Patch-wise Self-supervised Learning for Domain Adaptative Semantic Segmentation
mmfp2378 Weakly-Supervised Text Instance Segmentation
mmfp1929 PNT-Edge: Towards Robust Edge Detection with Noisy Labels by Learning Pixel-level Noise Transitions
mmfp3214 Video Frame Interpolation with Flow Transformer
mmfp1130 DUSA: Decoupled Unsupervised Sim2Real Adaptation for Vehicle-to-Everything Collaborative Perception
mmfp0080 Explicify Neural Implicit Fields for Efficient Dynamic Human Avatar Modeling via a Neural Explicit Surface
mmfp0259 MVFlow: Deep Optical Flow Estimation of Compressed Videos with Motion Vector Prior
mmfp0266 Uncertainty-Guided Spatial Pruning Architecture for Efficient Frame Interpolation
mmfp2534 Learning Generalized Representations for Open Set Temporal Action Localization
mmfp2369 Unambiguous Object Tracking by Exploiting Target Cues
mmfp2944 Masked Text Modeling: A Self-Supervised Pre-training Method for Scene Text Detection
mmfp1063 Object Part Parsing with Hierarchical Dual Transformer
mmfp0465 Towards Robust Real-Time Scene Text Detection: From Semantic to Instance Representation Learning
mmfp1834 Towards Flexible and Universal: A Novel Endpoint-based Framework for Vessel Structural Information Extracti
mmfp3271 FDCNet: Feature Drift Compensation Network for Class-Incremental Weakly Supervised Object Localization
mmfp2250 Collaborative Learning of Diverse Experts for Source-free Universal Domain Adaptation
mmfp3585 Read Ten Lines at One Glance: Line-Aware Semi-Autoregressive Transformer for Multi-Line Handwritten Math
mmfp0187 Beyond Domain Gap: Exploiting Subjectivity in Sketch-Based Person Retrieval
mmfp1288 Rethinking Pseudo-Label-Based Unsupervised Person Re-ID with Hierarchical Prototype-based Graph

mmfp2953 Single Domain Generalization via Unsupervised Diversity Probe
mmfp1636 PBFormer: Capturing Complex Scene Text Shape with Polynomial Band Transformer
mmfp1967 DANet: Multi-scale UAV Target Detection with Dynamic Feature Perception and Scale-aware Knowledge Distil
mmfp2162 A Unified Query-based Paradigm for Camouflaged Instance Segmentation
mmfp0506 Unite-Divide-Unite: Joint Boosting Trunk and Structure for High-accuracy Dichotomous Image Segmentation
mmfp2712 Exploring High-Correlation Source Domain Information for Multi-Source Domain Adaptation in Semantic Segm
mmfp3055 Deep Image Harmonization in Dual Color Spaces
mmfp0964 Pixel Adapter: A Graph-Based Post-Processing Approach for Scene Text Image Super-Resolution
mmfp1470 Where and How: Mitigating Confusion in Neural Radiance Fields from Sparse Inputs
mmfp0366 One-stage Low-resolution Text Recognition with High-resolution Knowledge Transfer
mmfp0424 Calibration-based Dual Prototypical Contrastive Learning Approach for Domain Generalization Semantic Segm
mmfp0921 Skeleton MixFormer: Multivariate Topology Representation for Skeleton-based Action Recognition
mmfp1898 Mask Again: Masked Knowledge Distillation for Masked Video Modeling
mmfp0522 Human-Object-Object Interaction: Towards Human-Centric Complex Interaction Detection
mmfp2189 On the Importance of Spatial Relations for Few-shot Action Recognition
mmfp0900 CgT-GAN: CLIP-guided Text GAN for Image Captioning
mmfp1903 Fine-grained Key-Value Memory Enhanced Predictor for Video Representation Learning
mmfp3001 Train One, Generalize to All: Generalizable Semantic Segmentation from Single-Scene to All Adverse Scenes
mmfp0580 All-in-one Multi-degradation Image Restoration Network via Hierarchical Degradation Representation
mmfp0653 NPF-200: A Multi-Modal Eye Fixation Dataset and Method for Non-Photorealistic Videos
mmfp0657 LandmarkGait: Intrinsic Human Parsing for Gait Recognition
mmfp0767 Patchmatch Stereo++: Patchmatch Binocular Stereo with Continuous Disparity Optimization
mmfp2345 Consistency-aware Feature Learning for Hierarchical Fine-grained Visual Classification
mmfp2888 FSR-Net: Deep Fourier Network for Shadow Removal
mmfp2972 Multi-Speed Global Contextual Subspace Matching for Few-Shot Action Recognition
mmfp2351 Lightweight Super-Resolution Head for Human Pose Estimation
mmfp2506 Exploiting Time-Frequency Conformers For General Music Enhancement
mmfp2296 Exploring Dual Representations in Large-Scale Point Clouds: A Simple Weakly Supervised Semantic Segmen
mmfp1115 Foreground/Background-Masked Interaction Learning for Spatio-temporal Action Detection
mmfp2487 TIVA-KG: A Multimodal Knowledge Graph with Text, Image, Video and Audio
mmfp1747 Enhancing Fake News Detection in Social Media via Label Propagation on Cross-modal Tweet Graph
mmfp1382 Cooperative Colorization: Exploring Latent Cross-Domain Priors for NIR Image Spectrum Translation
mmfp0965 ALA: Naturalness-aware Adversarial Lightness Attack
mmfp0991 Neural Image Popularity Assessment with Retrieval-augmented Transformer
mmfp2980 A Figure Skating Jumping Dataset for Replay-Guided Action Quality Assessment
mmfp0859 Enhancing Visibility in Nighttime Haze Images Using Guided APSF and Gradient Adaptive Convolution
mmfp0373 Rethinking Voice-Face Correlation: A Geometry View
mmfp0908 Dynamic Grouped Interaction Network for Low-Light Stereo Image Enhancement
mmfp1876 PVG: Progressive Vision Graph for Vision Recognition
mmfp2080 StylePrompter: All Styles Need Is Attention
mmfp2849 Improving Federated Person Re-Identification through Feature-Aware Proximity and Aggregation
mmfp3487 Transformer-based Open-world Instance Segmentation with Cross-task Consistency Regularization
mmfp3787 Cross-Illumination Video Anomaly Detection Benchmark
mmfp1796 Practical Edge Detection via Robust Asynchronous Learning
mmfp1713 MSECNet: Accurate and Robust Normal Estimation for 3D Point Clouds by Multi-Scale Edge Conditioning
mmfp0683 Efficient Parallel Multi-Scale Detail and Semantic Encoding Network for Lightweight Semantic Segmentation
mmfp1542 Multi-Frame Self-Supervised Depth Estimation with Multi-Scale Feature Fusion in Dynamic Scenes

mmfp3247 Peering into The Sketch: Ultra-Low Bitrate Face Compression for Joint Human and Machine Perception
mmfp3292 MTSN: Multiscale Temporal Similarity Network for Temporal Action Localization
mmfp0428 Disentangling Multi-view Representations Beyond Inductive Bias
mmfp0418 Implicit Decouple Network for Efficient Pose Estimation
mmfp2082 Occluded Skeleton-Based Human Action Recognition with Dual Inhibition Training
mmfp2871 P2I-NET: Mapping Camera Pose to Image via Adversarial Learning for New View Synthesis in Real Indoor Env
mmfp1388 IRCasTRF: Inverse Rendering by Optimizing Cascaded Tensorial Radiance Fields, Lighting, and Materials Fro
mmfp1684 Noise-Robust Continual Test-Time Domain Adaptation
mmfp2087 TIRDet: Mono-Modality Thermal InfraRed Object Detection Based on Prior Thermal-To-Visible Translation
mmfp3142 HARP: Let Object Detector Undergo Hyperplasia to Counter Adversarial Patches
mmfp1638 Scale-space Tokenization for Improving the Robustness of Vision Transformers
mmfp1843 Margin MCC: Chance-Robust Metric for Video Boundary Detection with Allowed Margin
mmfp2061 Exploring the Knowledge Transferred by Response-Based Teacher-Student Distillation
mmfp2180 Selecting Learnable Training Samples is All DETRs Need in Crowded Pedestrian Detection
mmfp3532 Data-Efficient Masked Video Modeling for Self-supervised Action Recognition
mmfp0169 DeNoising-MOT: Towards Multiple Object Tracking with Severe Occlussions
mmfp1917 Co-Salient Object Detection with Semantic-Level Consensus Extraction and Dispersion
mmfp3621 BLAT: Bootstrapping Language-Audio Pre-training based on AudioSet Tag-guided Synthetic Data
mmfp0024 A Simple Baseline for Open-World Tracking via Self-training
mmfp0796 VTLayout: A Multi-Modal Approach for Video Text Layout
mmfp4137 SEAR: Semantically-grounded Audio Representations
mmfp0175 DocDiff: Document Enhancement via Residual Diffusion Models
mmfp3441 POV: Prompt-Oriented View-agnostic Learning for Egocentric Hand-Object Interaction in the Multi-view World
mmfp2526 GraMMaR: Ground-aware Motion Model for 3D Human Motion Reconstruction
mmfp3442 SpeechTripleNet: End-to-End Disentangled Speech Representation Learning for Content, Timbre and Prosody
mmfp2853 Generating Explanations for Embodied Action Decision from Visual Observation
mmfp3213 Scene-aware Human Pose Generation using Transformer
mmfp3790 Dynamic Compositional Graph Convolutional Network for Efficient Composite Human Motion Prediction
mmfp0490 Diffusion-Augmented Depth Prediction with Sparse Annotations
mmfp2405 Chaos to Order: A Label Propagation Perspective on Source-Free Domain Adaptation
mmfp2247 Beware of Overcorrection: Scene-induced Commonsense Graph for Scene Graph Generation
mmfp0279 Scene Text Segmentation with Text-Focused Transformers
mmfp3999 MIEP: Channel Pruning with Multi-granular Importance Estimation for Object Detection
mmfp0155 Disentangled Representation Learning with Causality for Unsupervised Domain Adaptation
mmfp3877 Localized and Balanced Efficient Incomplete Multi-view Clustering
mmfp0658 Interpolation Normalization for Contrast Domain Generalization
mmfp3182 Multi-teacher Self-training for Semi-supervised Node Classification with Noisy Labels
mmfp4004 Long Short-Term Graph Memory Against Class-imbalanced Over-smoothing
mmfp3639 Class-level Structural Relation Modeling and Smoothing for Visual Representation Learning
mmfp3258 Unified Multi-modal Unsupervised Representation Learning for Skeleton-based Action Understanding
mmfp3974 Little Strokes Fell Great Oaks: Boosting the Hierarchical Features for Multi-exposure Image Fusion
mmfp0666 Triple-Granularity Contrastive Learning for Deep Multi-View Subspace Clustering
mmfp3199 CTCP: Cross Transformer and CNN for Pansharpening
mmfp3173 Chain of Propagation Prompting for Node Classification
mmfp1308 Efficient Multi-View Graph Clustering with Local and Global Structure Preservation
mmfp1292 Scalable Incomplete Multi-View Clustering with Structure Alignment
mmfp3157 Unifying Two-Stream Encoders with Transformers for Cross-Modal Retrieval

mmfp3766 Unbalanced Multi-view Deep Learning
mmfp2370 Incomplete Multi-View Clustering with Regularized Hierarchical Graph
mmfp2722 On Regularizing Multiple Clusterings for Ensemble Clustering by Graph Tensor Learning
mmfp1772 Event-guided Frame Interpolation and Dynamic Range Expansion of Single Rolling Shutter Image
mmfp0963 Learnable Graph Filter for Multi-view Clustering
mmfp3392 Cross-Silo Prototypical Calibration for Federated Learning with Non-IID Data
mmfp1211 CALM: An Enhanced Encoding and Confidence Evaluating Framework for Trustworthy Multi-view Learning
mmfp3610 Curriculum-Listener: Consistency- and Complementarity-Aware Audio-Enhanced Temporal Sentence Groundi
mmfp2818 Quality-Aware RGBT Tracking via Supervised Reliability Learning and Weighted Residual Guidance
mmfp1968 Event-Enhanced Multi-Modal Spiking Neural Network for Dynamic Obstacle Avoidance
mmfp2687 Multi-stage Factorized Spatio-Temporal Representation for RGB-D Action and Gesture Recognition
mmfp1807 M3R: Masked Token Mixup and Cross-Modal Reconstruction for Zero-Shot Learning
mmfp4068 Redundancy-aware Transformer for Video Question Answering
mmfp1314 Frequency-based Zero-Shot Learning with Phase Augmentation
mmfp2214 Uni-paint: A Unified Framework for Multimodal Image Inpainting with Pretrained Diffusion Model
mmfp2444 UniNeXt: Exploring A Unified Architecture for Vision Recognition
mmfp3335 MCG-MNER: A Multi-Granularity Cross-Modality Generative Framework for Multimodal NER with Instruction
mmfp1731 U2Net: A General Framework with Spatial-Spectral-Integrated Double U-Net for Image Fusion
mmfp0095 Modal-aware Visual Prompting for Incomplete Multi-modal Brain Tumor Segmentation
mmfp1526 Where to Find Fascinating Inter-Graph Supervision: Imbalanced Graph Classification with Kernel Information E
mmfp1358 pmBQA: Projection-based Blind Point Cloud Quality Assessment via Multimodal Learning
mmfp2788 Dropping Pathways Towards Deep Multi-View Graph Subspace Clustering Networks
mmfp2183 Multi-view Graph Clustering via Efficient Global-Local Spectral Embedding Fusion
mmfp3456 Debunking Free Fusion Myth: Online Multi-view Anomaly Detection with Disentangled Product-of-Experts Mod
mmfp0904 Domain-irrelevant Feature Learning for Generalizable Pan-sharpening
mmfp0821 Depth-aided Camouflaged Object Detection
mmfp0215 SemanticRT: A Large-Scale Dataset and Method for Robust Semantic Segmentation in Multispectral Images
mmfp0406 MEAformer: Multi-modal Entity Alignment Transformer for Meta Modality Hybrid
mmfp1120 Multi-Modal and Multi-Scale Temporal Fusion Architecture Search for Audio-Visual Video Parsing
mmfp1736 Incorporating Domain Knowledge Graph into Multimodal Movie Genre Classification with Self-Supervised Atter
mmfp3185 Multi-scale Spatial-Spectral Attention Guided Fusion Network for Pansharpening
mmfp1301 Modality Profile - A New Critical Aspect to be Considered When Generating RGB-D Salient Object Detection T
mmfp0715 TMac: Temporal Multi-Modal Graph Learning for Acoustic Event Classification
mmfp0787 UAVTracker: Fast multiple object tracking from UAV-captured videos based on optical flow
mmfp1615 ScribbleVC: Scribble-supervised Medical Image Segmentation with Vision-Class Embedding
mmfp1423 Temporally Efficient Gabor Transformer for Unsupervised Video Object Segmentation
mmfp0344 Beyond First Impressions: Integrating Joint Multi-modal Cues for Comprehensive 3D Representation
mmfp2517 Hierarchical Visual Attribute Learning in the Wild
mmfp3144 Hierarchical Semantic Enhancement Network for Multimodal Fake News Detection
mmfp3323 Towards Balanced Active Learning for Multimodal Classification
mmfp3617 Learning Event-Specific Localization Preferences for Audio-Visual Event Localization
mmfp1233 Object Segmentation by Mining Cross-Modal Semantics
mmfp0369 PSNEA: Pseudo-Siamese Network for Entity Alignment between Multi-modal Knowledge Graphs
mmfp1464 Federated Deep Multi-View Clustering with Global Self-Supervision
mmfp0138 Audio-Visual Spatial Integration and Recursive Attention for Robust Sound Source Localization
mmfp1850 Hypergraph-Enhanced Hashing for Unsupervised Cross-Modal Retrieval via Robust Similarity Guidance
mmfp2024 Reinforcement Graph Clustering with Unknown Cluster Number

mmfp0070	Cultural Self-Adaptive Multimodal Gesture Generation Based on Multiple Culture Gesture Dataset
mmfp0089	DPNET: Dynamic Poly-attention Network for Trustworthy Multi-modal Classification
mmfp3892	Tile Classification Based Viewport Prediction with Multi-modal Fusion Transformer
mmfp0399	Semantic-based Selection, Synthesis, and Supervision for Few-shot Learning
mmfp2329	Exploring Universal Principles for Graph Contrastive Learning: A Statistical Perspective
mmfp2839	Text-to-Audio Generation using Instruction Guided Latent Diffusion Model
mmfp4062	DRIN: Dynamic Relation Interactive Network for Multimodal Entity Linking
mmfp2906	MVCIR-net: Multi-view Clustering Information Reinforcement Network
mmfp2349	Preserving Local and Global Information: An Effective Metric-based Subspace Clustering
mmfp0846	FeaCo: Reaching Robust Feature-Level Consensus in Noisy Pose Conditions
mmfp1323	Cross-Lingual Transfer of Large Language Model by Visually-Derived Supervision Toward Low-Resource Lanç
mmfp1461	ALEX: Towards Effective Graph Transfer Learning with Noisy Labels
mmfp3971	Skeletal Spatial-Temporal Semantics Guided Homogeneous-Heterogeneous Multimodal Network for Action Re
mmfp1580	Unveiling the Power of CLIP in Unsupervised Visible-Infrared Person Re-Identification
mmfp1957	DTF-Net: Category-Level Pose Estimation and Shape Reconstruction via Deformable Template Field
mmfp2129	Text-Only Training for Visual Storytelling
mmfp3331	Saliency Prototype for RGB-D and RGB-T Salient Object Detection
mmfp1035	PAIF: Perception-Aware Infrared-Visible Image Fusion for Attack-Tolerant Semantic Segmentation
mmfp2002	Cross-Modal Graph Attention Network for Entity Alignment
mmfp3333	Intra- and Inter-Modal Curriculum for Multimodal Learning
mmfp1680	Graph based Spatial-temporal Fusion for Multi-modal Person Re-identification
mmfp1817	Transferring CLIP's Knowledge into Zero-Shot Point Cloud Semantic Segmentation
mmfp3163	Bio-Inspired Audiovisual Multi-Representation Integration via Self-Supervised Learning
mmfp2824	DLFusion: Painting-Depth Augmenting-LiDAR for Multimodal Fusion 3D Object Detection
mmfp2596	Automatic Network Architecture Search for RGB-D Semantic Segmentation
mmfp0380	Attentive Alignment Network for Multispectral Pedestrian Detection
mmfp1144	FedAA: Using Non-sensitive Modalities to Improve Federated Learning while Preserving Image Privacy
mmfp2723	Unsupervised Domain Adaptation for Video Object Grounding with Cascaded Debiasing Learning
mmfp2900	RAHNet: Retrieval Augmented Hybrid Network for Long-tailed Graph Classification
mmfp4104	That's What I Said: Fully-Controllable Talking Face Generation
mmfp3317	Event-Diffusion: Event-Based Image Reconstruction and Restoration with Diffusion Models
mmfp0285	Mask to reconstruct: Cooperative Semantics Completion for Video-text Retrieval
mmfp0533	Self-Contrastive Graph Diffusion Network
mmfp2747	Cross-modal & Cross-domain Learning for Unsupervised LiDAR Semantic Segmentation
mmfp3494	Multi-View Representation Learning via View-Aware Modulation
mmfp2795	Uni-Dual: A Generic Unified Dual-Task Medical Self-Supervised Learning Framework
mmfp3107	Towards Better Multi-modal Keyphrase Generation via Visual Entity Enhancement and Multi-granularity Image
mmfp4026	Multi-modal Social Bot Detection: Learning Homophilic and Heterophilic Connections Adaptively
mmfp1560	CPU: Codebook Lookup Transformer with Knowledge Distillation for Point Cloud Upsampling
mmfp3770	Your tone speaks louder than your face! Modality Order Infused Multi-modal Sarcasm Detection
mmfp2752	Fine-Grained Multimodal Named Entity Recognition and Grounding with a Generative Framework
mmfp0670	SkipStreaming: Pinpointing User-Perceived Redundancy in Correlated Web Video Streaming through the Lens
mmfp0883	Synthesizing Long-Term Human Motions with Diffusion Models via Coherent Sampling
mmfp1082	Layout Sequence Prediction From Noisy Mobile Modality
mmfp1912	Graph-Based Video-Language Learning with Multi-Grained Audio-Visual Alignment
mmfp2361	Advancing Video Question Answering with a Multi-modal and Multi-layer Question Enhancement Network
mmfp0296	Motion-Decoupled Spiking Transformer for Audio-Visual Zero-Shot Learning

mmfp1940	Multimodal Color Recommendation in Vector Graphic Documents
mmfp3062	Open-Vocabulary Object Detection via Scene Graph Discovery
mmfp3948	Universal Domain Adaptive Network Embedding for Node Classification
mmfp3146	Uncertainty-Guided End-to-End Audio-Visual Speaker Diarization for Far-field Recordings
mmfp3600	Induction Network: Audio-Visual Modality Gap-Bridging for Self-Supervised Sound Source Localization
mmfp2155	HELIOS: Hyper-Relational Schema Modeling from Knowledge Graphs
mmfp3680	Breaking the Barrier Between Pre-training and Fine-tuning: A Hybrid Prompting Model for Knowledge-Based V
mmfp3827	OccluBEV: Occlusion Aware Spatiotemporal Modeling for Multi-view 3D Object Detection
mmfp0850	Hermes: Leveraging Implicit Inter-Frame Correlation for Bandwidth-Efficient Mobile Volumetric Video Streamin
mmfp0254	Reparo: QoE-Aware Live Video Streaming in Low Rate Networks by Intelligent Frame Recovery
mmfp3591	DeepSVC: Deep Scalable Video Coding for Both Machine and Human Vision
mmfp0352	Concerto: Client-server Orchestration for Real-Time Video Analytics
mmfp4065	Think before You Leap: Content-Aware Low-Cost Edge-Assisted Video Semantic Segmentation
mmfp0365	TwinStar: A Practical Multi-path Transmission Framework for Ultra-Low Latency Video Delivery
mmfp3035	Addressing Scalability for Real-time Multiuser Holo-portation: Introducing and Assessing a Multipoint Control U
mmfp3837	ELFIC: A Learning-based Flexible Image Codec with Rate-Distortion-Complexity Optimization
mmfp0972	Mamba: Bringing Multi-dimensional ABR to WebRTC
mmfp3143	PDE-based Progressive Prediction Framework for Attribute Compression of 3D Point Clouds
mmfp0922	Gaze Analysis System for Immersive 360 Video for Preservice Teacher Education