

mmfp1458 Distortion-aware Transformer in 360° Salient Object Detection

mmfp0351 Symmetrical Linguistic Feature Distillation with CLIP for Scene Text Recognition

mmfp2962 SpaceCLIP: A Vision-Language Pretraining Framework With Spatial Reconstruction On Text

mmfp2194 Improving Cross-Modal Recipe Retrieval with Component-Aware Prompted CLIP Embedding

mmfp2990 Dynamic Contrastive Learning with Pseudo-samples Intervention for Weakly Supervised Joint Video MR and HC

mmfp0595 RAMM: Retrieval-augmented Biomedical Visual Question Answering with Multi-modal Pre-training

mmfp2004 RTQ: Rethinking Video-language Understanding Based on Image-text Model

mmfp0769 SUR-adapter: Enhancing Text-to-Image Pre-trained Diffusion Models with Large Language Models

mmfp2341 Face Encryption via Frequency-Restricted Identity-Agnostic Attacks

mmfp0156 Emotion-Prior Awareness Network for Emotional Video Captioning

mmfp2089 TE-KWS: Text-Informed Speech Enhancement for Noise-Robust Keyword Spotting

mmfp2952 A Prior Instruction Representation Framework for Remote Sensing Image-text Retrieval

mmfp2312 PromptMTopic: Unsupervised Multimodal Topic Modeling of Memes using Large Language Models

mmfp2176 Dynamic Low-Rank Instance Adaptation for Universal Neural Image Compression

mmfp1407 Layout-Guided High-Faithfulness Text-to-Image Generation

mmfp0135 POAR: Towards Open-World Pedestrian Attribute Recognition

mmfp3300 PointCRT: Detecting Backdoor in 3D Point Cloud via Corruption Robustness

mmfp0983 Blind Image Super-resolution with Rich Texture-Aware Codebook

mmfp0264 V2Depth: Monocular Depth Estimation via Feature-Level Virtual-View Simulation and Refinement

mmfp1828 GCMA: Generative Cross-Modal Transferable Adversarial Attacks from Images to Videos

mmfp0235 AdaBrowse: Adaptive Video Browser for Efficient Continuous Sign Language Recognition

mmfp2415 Dynamic Triple Reweighting Network for Automatic Femoral Head Necrosis Diagnosis from Computed Tomogra

mmfp0711 Category-Level Articulated Object 9D Pose Estimation via Reinforcement Learning

mmfp0661 RetouchingFFHQ: A Large-scale Dataset for Fine-grained Face Retouching Detection

mmfp3313 Slow-Fast Time Parameter Aggregation Network for Class-Incremental Lip Reading

mmfp2572 Text-based Person Search without Parallel Image-Text Data

mmfp2551 Exploring Inconsistent Knowledge Distillation for Object Detection with Data Augmentation

mmfp1857 CARIS: Context-Aware Referring Image Segmentation

mmfp1811 Ground-to-Aerial Person Search: Benchmark Dataset and Approach

mmfp1444 Sparse Sharing Relation Network for Panoptic Driving Perception

mmfp3091 Multi-Granularity Interactive Transformer Hashing for Cross-modal Retrieval

mmfp3733 Equivariant Learning for Out-of-Distribution Cold-start Recommendation

mmfp0541 Vision-Guided Composed Image Retrieval

mmfp1799 Your Negative May not Be True Negative: Boosting Image-Text Matching with False Negative Elimination

mmfp1106 A Tale of Two Graphs: Freezing and Denoising Graph Structures for Multimodal Recommendation

mmfp2662 ProtoHPE: Prototype-guided High-frequency Patch Enhancement for Visible-Infrared Person Re-identification

mmfp1768 Online Distillation-enhanced Multi-modal Transformer for Sequential Recommendation

mmfp0221 Zero-shot Micro-video Classification with Neural Variational Inference in Graph Prototype Network

mmfp2843 Joint Searching and Grounding: Multi-Granularity Video Content Retrieval

mmfp3136 Making Users Indistinguishable: Attribute-wise Unlearning in Recommender Systems

mmfp4047 Prior-Guided Accuracy-Bias Tradeoff Learning for CTR Prediction in Multimedia Recommendation

mmfp2360 GoRec: A Generative Cold-start Recommendation Framework

mmfp2120 Prototype-guided Knowledge Transfer for Federated Unsupervised Cross-modal Hashing

mmfp0847 EAT: An Enhancer for Aesthetics-Oriented Transformers

mmfp3609 UnifiedGesture: A Unified Gesture Synthesis Model for Multiple Skeletons

mmfp0209 Towards Explainable In-the-Wild Video Quality Assessment: a Database and a Language-Prompted Approach

mmfp1847 Sketch Input Method Editor: A Comprehensive Dataset and Methodology for Systematic Input Recognition

mmfp0758 StableVQA: A Deep No-Reference Quality Assessment Model for Video Stability

mmfp1033 Spatial-angular Quality-aware Representation Learning for Blind Light Field Image Quality Assessment

mmfp1019 Light-VQA: A Multi-Dimensional Quality Assessment Model for Low-Light Video Enhancement

mmfp1447 Capturing Co-existing Distortions in User-Generated Content for No-reference Video Quality Assessment

mmfp3771 Understanding User Behavior in Volumetric Video Watching: Dataset, Analysis and Prediction

mmfp1232 AesCLIP: Multi-Attribute Contrastive Learning for Image Aesthetics Assessment

mmfp1284 Semantics-Enriched Cross-Modal Alignment for Complex-Query Video Moment Retrieval

mmfp0232 NightHazeFormer: Single Nighttime Haze Removal Using Prior Query Transformer

mmfp2196 FSNet: Frequency Domain Guided Superpixel Segmentation Network for Complex Scenes

mmfp0570 Zero-Shot Learning by Harnessing Adversarial Samples

mmfp0228 Sequential Affinity Learning for Video Restoration

mmfp0349 Beat: Bi-directional One-to-Many Embedding Alignment for Text-based Person Retrieval

mmfp2307 Transformer-based Point Cloud Generation Network

mmfp1770 Isolation and Induction: Training Robust Deep Neural Networks against Model Stealing Attacks

mmfp1523 Filling the Information Gap between Video and Query for Language-Driven Moment Retrieval

mmfp0534 Improving Semi-Supervised Semantic Segmentation with Dual-Level Siamese Structure Network

mmfp3872 Focusing on Flexible Masks: A Novel Framework for Panoptic Scene Graph Generation with Relation Constraint

mmfp0837 CCMB: A Large-scale Chinese Cross-modal Benchmark

mmfp0902 CPLFormer: Cross-scale Prototype Learning Transformer for Image Snow Removal

mmfp2825 Video Entailment via Reaching a Structure-Aware Cross-modal Consensus

mmfp0129 Cerebrovascular Segmentation in TOF-MRA with Topology Regularization Adversarial Model

mmfp2717 Hierarchical Reasoning Network with Contrastive Learning for Few-Shot Human-Object Interaction Recognition

mmfp1374 Uncertainty-Driven Dynamic Degradation Perceiving and Background Modeling for Efficient Single Image Denoising

mmfp1107 DAE-Talker: High Fidelity Speech-Driven Talking Face Generation with Diffusion Autoencoder

mmfp2786 Spatio-Temporal Branching for Motion Prediction using Motion Increments

mmfp1409 Generative Neutral Features-Disentangled Learning for Facial Expression Recognition

mmfp1354 Deep Algorithm Unrolling with Registration Embedding for Pansharpening

mmfp0426 DAOT: Domain-Agnostically Aligned Optimal Transport for Domain-Adaptive Crowd Counting

mmfp1757 Partial Annotation-based Video Moment Retrieval via Iterative Learning

mmfp3485 Style Transfer Meets Super-Resolution: Advancing Unpaired Infrared-to-Visible Image Translation with Detail Enhancement

mmfp0206 Mind the Gap: Improving Success Rate of Vision-and-Language Navigation by Revisiting Oracle Success Route

mmfp1099 Feature-Suppressed Contrast for Self-Supervised Food Pre-training

mmfp4087 Learning from Easy to Hard Pairs: Multi-step Reasoning Network for Human-Object Interaction Detection

mmfp0270 Separate and Locate: Rethink the Text in Text-based Visual Question Answering

mmfp3005 Improving Zero-shot Visual Question Answering via Large Language Models with Reasoning Question Prompts

mmfp0548 Adaptive Decoupled Pose Knowledge Distillation

mmfp0682 Biased-Predicate Annotation Identification via Unbiased Visual Predicate Representation

mmfp3735 Zero-Shot Object Detection by Semantics-Aware DETR with Adaptive Contrastive Loss

mmfp2607 Rethinking Missing Modality Learning from a Decoding Perspective

mmfp1681 Improving the Transferability of Adversarial Examples with Arbitrary Style Transfer

mmfp3048 Mixup-Augmented Temporally Debaised Video Grounding with Multimodal Disentanglement

mmfp3816 Learning Semantics-Grounded Vocabulary Representation for Video-Text Retrieval

mmfp1928 Learning a Graph Neural Network with Cross Modality Interaction for Image Fusion

mmfp0583 COPA : Efficient Vision-Language Pre-training through Collaborative Object- and Patch-Text Alignment

mmfp0088 Towards Unified Text-based Person Retrieval: A Large-scale Multi-Attribute and Language Search Benchmark

mmfp0562 Towards Real-Time Sign Language Recognition and Translation on Edge Devices

mmfp2760 Enhancing Visually-Rich Document Understanding via Layout Structure Modeling

mmfp1028 Non-Exemplar Class-Incremental Learning via Adaptive Old Class Reconstruction

mmfp0412 CLIP-Count: Towards Text-Guided Zero-Shot Object Counting

mmfp2095 Self-Supervised Cross-Language Scene Text Editing

mmfp1780 Learning Implicit Entity-object Relations by Bidirectional Generative Alignment for Multimodal NER

mmfp2246 MORE: A Multimodal Object-Entity Relation Extraction Dataset with a Benchmark Evaluation

mmfp1439 Weakly-supervised Video Scene Graph Generation via Unbiased Cross-modal Learning

mmfp2873 Reducing Intrinsic and Extrinsic Data Biases for Moment Localization with Natural Language

mmfp0075 VioLET: Vision-Language Efficient Tuning with Collaborative Multi-modal Gradients
mmfp0729 Mirror-NeRF: Learning Neural Radiance Fields for Mirrors with Whitted-Style Ray Tracing
mmfp1045 Semi-supervised Deep Multi-view Stereo
mmfp1377 Dual-Modal Attention-Enhanced Text-Video Retrieval with Triplet Partial Margin Contrastive Learning
mmfp1868 Temporal Sentence Grounding in Streaming Videos
mmfp2870 Modality-agnostic Augmented Multi-Collaboration Representation for Semi-supervised Heterogenous Face Recog
mmfp3690 Swin-UNIT: Transformer-based GAN for High-resolution Unpaired Image Translation
mmfp1678 PixelFace+: Towards Controllable Face Generation and Manipulation with Text Descriptions and Segmentation
mmfp0732 LiFT: Transfer Learning in Vision-Language Models for Downstream Adaptation and Generalization
mmfp1712 VCMaster: Generating Diverse and Fluent Live Video Comments Based on Multimodal Contexts
mmfp2265 Whether you can locate or not? Interactive Referring Expression Generation
mmfp3172 Iterative Learning with Extra and Inner Knowledge for Long-tail Dynamic Scene Graph Generation
mmfp3372 Improving Image Captioning through Visual and Semantic Mutual Promotion
mmfp1049 Fine-Grained Spatiotemporal Motion Alignment for Contrastive Video Representation Learning
mmfp2206 Better Integrating Vision and Semantics for Improving Few-shot Classification
mmfp1875 Multi-Domain Lifelong Visual Question Answering via Self-Critical Distillation
mmfp1091 Relation Triplet Construction for Cross-modal Text-to-Video Retrieval
mmfp2032 HSVLT: Hierarchical Scale-Aware Vision-Language Transformer for Multi-Label Image Classification
mmfp0101 Depth-Aware Sparse Transformer for Video-Language Learning
mmfp0309 Invariant Meets Specific: A Scalable Harmful Memes Detection Framework
mmfp1147 A Method of Micro-Geometric Details Preserving in Surface Reconstruction from Gradient
mmfp3624 Progressive Positive Association Framework for Image and Text Retrieval
mmfp3743 Globally-Robust Instance Identification and Locally-Accurate Keypoint Alignment for Multi-Person Pose Estimati
mmfp0066 Unlocking the Power of Cross-Dimensional Semantic Dependency for Image-Text Matching
mmfp1487 Dark Knowledge Balance Learning for Unbiased Scene Graph Generation
mmfp0040 Orthogonal Uncertainty Representation of Data Manifold for Robust Long-Tailed Learning
mmfp2202 Topological Structure Learning for Weakly-Supervised Out-of-Distribution Detection
mmfp3226 Efficient Spatio-Temporal Video Grounding with Semantic-Guided Feature Decomposition
mmfp2396 Prior Knowledge-driven Dynamic Scene Graph Generation with Causal Inference
mmfp3627 ATM: Action Temporality Modeling for Video Question Answering
mmfp3013 CLIP-Hand3D: Exploiting 3D Hand Pose Estimation via Context-Aware Prompting
mmfp4120 A Multitask Framework for Graffiti-to-Image Translation
mmfp3480 Adaptive Contrastive Learning for Learning Robust Representations under Label Noise
mmfp0960 Distilling Vision-Language Foundation Models: A Data-Free Approach via Prompt Diversification
mmfp3063 Real20M: A Large-scale E-commerce Dataset for Cross-domain Retrieval
mmfp4038 Zero-TextCap: Zero-shot Framework for Text-based Image Captioning
mmfp2066 Adversarial Training of Deep Neural Networks Guided by Texture and Structural Information
mmfp3119 TeViS: Translating Text Synopses to Video Storyboards
mmfp0913 Chain-of-Look Prompting for Verb-centric Surgical Triplet Recognition in Endoscopic Videos
mmfp0923 Dense Object Grounding in 3D Scenes
mmfp0934 Uniformly Distributed Category Prototype-Guided Vision-Language Framework for Long-Tail Recognition
mmfp1309 Beyond Generic: Enhancing Image Captioning with Real-World Knowledge using Vision-Language Pre-Training
mmfp1364 Toward High Quality Facial Representation Learning
mmfp1789 HAAN: Human Action Aware Network for Multi-label Temporal Action Detection
mmfp2233 Exploring Coarse-to-Fine Action Token Localization and Interaction for Fine-grained Video Action Recognition
mmfp2282 Semantic-Aware Generator and Low-level Feature Augmentation for Few-shot Image Generation
mmfp2291 Self-PT: Adaptive Self-Prompt Tuning for Low-Resource Visual Question Answering
mmfp2373 SAUNet: Spatial-Attention Unfolding Network for Image Compressive Sensing
mmfp2758 CONICA: A Contrastive Image Captioning Framework with Robust Similarity Learning
mmfp3002 Enhancing Vision-Language Pre-Training with Jointly Learned Questioner and Dense Captioner
mmfp3815 Deconfounded Visual Question Generation with Causal Inference
mmfp4108 Null-text Guidance in Diffusion Models is Secretly a Cartoon-style Creator
mmfp1456 Triple Correlations-Guided Label Supplementation for Unbiased Video Scene Graph Generation
mmfp3653 Probability Distribution Based Frame-supervised Language-driven Action Localization
mmfp4094 LUNA: Language as Continuing Anchors for Referring Expression Comprehension

mmfp0914 Prompt Me Up: Unleashing the Power of Alignments for Multimodal Entity and Relation Extraction
mmfp0935 Language-Guided Visual Aggregation Network for Video Question Answering
mmfp0065 Resource Constrained Model Compression via Minimax Optimization for Spiking Neural Networks
mmfp0564 Semi-Supervised Convolutional Vision Transformer with Bi-Level Uncertainty Estimation for Medical Image Segmentation
mmfp1209 Enhancing Multi-modal Multi-hop Question Answering via Structured Knowledge and Unified Retrieval-Generation
mmfp3303 Local Consensus Enhanced Siamese Network with Reciprocal Loss for Two-view Correspondence Learning
mmfp3576 Pro-Cap: Leveraging a Frozen Vision-Language Model for Hateful Meme Detection
mmfp3927 Prototype-guided Cross-modal Completion and Alignment for Incomplete Text-based Person Re-identification
mmfp2602 Language-guided Human Motion Synthesis with Atomic Actions
mmfp0410 Avatar Knowledge Distillation: Self-ensemble Teacher Paradigm with Uncertainty
mmfp1784 Constructing Holistic Spatio-Temporal Scene Graph for Video Semantic Role Labeling
mmfp0189 SelfTalk: A Self-Supervised Commutative Training Diagram to Comprehend 3D Talking Faces
mmfp0887 Zero-shot Skeleton-based Action Recognition via Mutual Information Estimation and Maximization
mmfp3661 Contrast-augmented Diffusion Model with Fine-grained Sequence Alignment for Markup-to-Image Generation
mmfp2187 Guided Image Synthesis via Initial Image Editing in Diffusion Model
mmfp3632 External Knowledge Dynamic Modeling for Image-text Retrieval
mmfp3474 Seeing in Flowing: Adapting CLIP for Action Recognition with Motion Prompts Learning
mmfp0174 Enhancing Real-Time Super Resolution with Partial Convolution and Efficient Variance Attention
mmfp0701 HSIC-based Moving Weight Averaging for Few-Shot Open-Set Object Detection
mmfp2516 Exploiting Low-confidence Pseudo-labels for Source-free Object Detection
mmfp3076 Bridging Language and Geometric Primitives for Zero-shot Point Cloud Segmentation
mmfp3334 Graph Spectral Perturbation for 3D Point Cloud Contrastive Learning
mmfp2714 Retrieval-based Knowledge Augmented Vision Language Pre-training
mmfp3084 ACQ: Few-shot Backdoor Defense via Activation Clipping and Quantizing
mmfp2960 Underwater Image Enhancement by Transformer-based Diffusion Model with Non-uniform Sampling for Skip Str
mmfp4000 Faster Attention with Heterogeneous RGCN for Medical ICD Coding Generation
mmfp0938 Semi-supervised Semantic Segmentation with Mutual Knowledge Distillation
mmfp0052 Shift Pruning: Equivalent Weight Pruning for CNN via Differentiable Shift Operator
mmfp0192 Improving Human-Object Interaction Detection via Virtual Image Learning
mmfp0505 ZRIGF: An Innovative Multimodal Framework for Zero-Resource Image-Grounded Dialogue Generation
mmfp0660 Diffused Fourier Network for Video Action Segmentation
mmfp1187 Rethinking the Localization in Weakly Supervised Object Localization
mmfp1421 Hyperspectral Image Denoising with Spectrum Alignment
mmfp3748 Training Multimedia Event Extraction With Generated Images and Captions
mmfp1353 BMI-Net: A Brain-inspired Multimodal Interaction Network for Image Aesthetic Assessment
mmfp0409 Towards Accurate Lip-to-Speech Synthesis in-the-Wild
mmfp2260 Towards End-to-End Unsupervised Saliency Detection with Self-Supervised Top-Down Context
mmfp2939 PoSynDA: Multi-Hypothesis Pose Synthesis Domain Adaptation for Enhanced 3D Human Pose Estimation
mmfp0472 All in One: Exploring Unified Vision-Language Tracking with Multi-Modal Alignment
mmfp4139 Scene-text Oriented Visual Entailment: Task, Dataset and Solution
mmfp1973 QA-CLIMS: Question-Answer Cross Language Image Matching for Weakly Supervised Semantic Segmentation
mmfp1509 Confidence-Aware Contrastive Learning for Semantic Segmentation
mmfp1312 Hierarchical Prompt Learning Using CLIP for Multi-label Classification with Single Positive Labels
mmfp0295 Reservoir Computing Transformer for Image-Text Retrieval
mmfp1686 Model Inversion Attack via Dynamic Memory Learning
mmfp2055 AdaCLIP: Towards Pragmatic Multimodal Video Retrieval
mmfp3029 Do Vision-Language Transformers Exhibit Visual Commonsense? An Empirical Study of VCR
mmfp1317 Semi-supervised Domain Adaptation via Joint Contrastive Learning with Sensitivity
mmfp2070 LocLoc: Low-level Cues and Local-area Guides for Weakly Supervised Object Localization
mmfp2394 Expand BERT Representation with Visual Information via Grounded Language Learning with Multimodal Partial
mmfp1347 Food-500 Cap: A Fine-Grained Food Caption Benchmark for Evaluating Vision-Language Models
mmfp2411 Enhancing Sentence Representation with Visually-supervised Multimodal Pre-training
mmfp0722 Cross-modal Contrastive Learning for Multimodal Fake News Detection
mmfp2459 Visual Captioning at Will: Describing Images and Videos Guided by a Few Stylized Sentences
mmfp3921 Category-Specific Prompts for Animal Action Recognition with Pretrained Vision-Language Models

mmfp2453 Scene Graph Masked Variational Autoencoders for 3D Scene Generation
mmfp1445 AvatarFusion: Zero-shot Generation of Clothing-Decoupled 3D Avatars Using 2D Diffusion
mmfp2931 KeyPosS: Plug-and-Play Facial Landmark Detection through GPS-Inspired True-Range Multilateration
mmfp3907 WormTrack: Dataset and Benchmark for Multi-Object Tracking in Worm Crowds
mmfp2390 Relational Contrastive Learning for Scene Text Recognition
mmfp2503 MaTCR: Modality-Aligned Thought Chain Reasoning for Multimodal Task-Oriented Dialogue Generation
mmfp0421 Fine-grained Pseudo Labels for Scene Text Recognition
mmfp0628 VPA: Fully Test-Time Visual Prompt Adaptation
mmfp0844 Unsupervised Domain Adaptation for Referring Semantic Segmentation
mmfp1652 OCSKB: An Object Component Sketch Knowledge Base for Fast 6D Pose Estimation
mmfp3054 Fine-Grained Visual Prompt Learning of Vision-Language Models for Image Recognition
mmfp1583 General Debiasing for Multimodal Sentiment Analysis
mmfp0448 Feeling Positive? Predicting Emotional Image Similarity from Brain Signals
mmfp3935 Multimodal Physiological Signals Fusion for Online Emotion Recognition
mmfp0054 Learning from More: Combating Uncertainty Cross-multidomain for Facial Expression Recognition
mmfp2935 MindDiffuser: Controlled Image Reconstruction from Human Brain Activity with Semantic and Structural Diffusion
mmfp3332 Pretrained Implicit-Ensemble Transformer for Open-Set Authentication on Multimodal Mobile Biometrics
mmfp1588 Revisiting Disentanglement and Fusion on Modality and Context in Conversational Multimodal Emotion Recognition
mmfp0275 Sensing Micro-Motion Human Patterns using Multimodal mmRadar and Video Signal for Affective and Psychological
mmfp2683 Unlocking the Power of Multimodal Learning for Emotion Recognition in Conversation
mmfp0068 Emo-DNA: Emotion Decoupling and Alignment Learning for Cross-Corpus Speech Emotion Recognition
mmfp2791 Emotionally Situated Text-to-Speech Synthesis in User-Agent Conversation
mmfp3165 Multimodal Adaptive Emotion Transformer with Flexible Modality Inputs on A Novel Dataset with Continuous Latent
mmfp1691 Graph to Grid: Learning Deep Representations for Multimodal Emotion Recognition
mmfp0484 Multimodal Prompt Transformer with Hybrid Contrastive Learning for Emotion Recognition in Conversation
mmfp2588 SAAML: A Framework for Semi-supervised Affective Adaptation via Metric Learning
mmfp2946 Learning Shared Semantic Information from Multimodal Bio-signals for Brain-Muscle Modulation Analysis
mmfp0789 Auditory Attention Decoding with Task-Related Multi-View Contrastive Learning
mmfp2163 Progressive Visual Content Understanding Network for Image Emotion Classification
mmfp2138 Few-shot Multimodal Sentiment Analysis Based on Multimodal Probabilistic Fusion Prompts
mmfp2832 MEDIC: A Multimodal Empathy Dataset in Counseling
mmfp2880 Towards Adaptable Graph Representation Learning: An Adaptive Multi-Graph Contrastive Transformer
mmfp2740 MetaFBP: Learning to Learn High-Order Predictor for Personalized Facial Beauty Prediction
mmfp4001 CMCU-CSS: Enhancing Naturalness via Commonsense-based Multi-modal Context Understanding in Conversation
mmfp3687 Multi-label Emotion Analysis in Conversation via Multimodal Knowledge Distillation
mmfp1717 Facial Auto Rigging from 4D Expressions via Skinning Decomposition
mmfp2930 MAE-DFER: Efficient Masked Autoencoder for Self-supervised Dynamic Facial Expression Recognition
mmfp2527 EmotionKD: A Cross-Modal Knowledge Distillation Framework for Emotion Recognition Based on Physiological Signals
mmfp2798 UniSA: Unified Generative Framework for Sentiment Analysis
mmfp2819 Patch-Aware Representation Learning for Facial Expression Recognition
mmfp4224 COVES: A Cognitive-Affective Deep Model that Personalizes Math Problem Difficulty in Real Time and Improves
mmfp3691 Efficient Labelling of Affective Video Datasets via Few-Shot & Multi-Task Contrastive Learning
mmfp1386 Pedestrian-specific Bipartite-aware Similarity Learning for Text-based Person Retrieval
mmfp0012 Generalizable Label Distribution Learning
mmfp2769 Dual Dynamic Proxy Hashing Network for Long-tailed Image Retrieval
mmfp0486 Hybrid Interaction Temporal Knowledge Graph Embedding Based on Householder Transformations
mmfp0994 C²MR: Continual Cross-Modal Retrieval for Streaming Multi-modal Data
mmfp3301 When Perceptual Authentication Hashing Meets Neural Architecture Search

mmfp2083 GraphMedia:Communication-Balanced Graph Searching for Billion-scale Social Media Access
mmfp1131 Learning Style-Invariant Robust Representation for Generalizable Visual Instance Retrieval
mmfp2071 Hierarchical Category-Enhanced Prototype Learning for Imbalanced Temporal Recommendation
mmfp2860 In-processing User Constrained Dominant Sets for User-Oriented Fairness in Recommender Systems
mmfp0601 Learning Comprehensive Representations with Richer Self for Text-to-Image Person Re-Identification
mmfp3241 Doubly Intention Learning for Cold-start Recommendation with Uncertainty-aware Stochastic Meta Process
mmfp1476 $\mathcal{D}iVa$: An Iterative Framework to Harvest More Diverse and Valid Labels from User Comments for Music
mmfp2921 Contrastive Intra- and Inter-Modality Generation for Enhancing Incomplete Multimedia Recommendation
mmfp1026 Differentially Private Sparse Mapping for Privacy-Preserving Cross Domain Recommendation
mmfp2653 Handling Label Uncertainty for Camera Incremental Person Re-Identification
mmfp3341 Towards Deconfounded Image-text Matching with Causal Inference
mmfp2801 Enhancing Adversarial Robustness of Multi-modal Recommendation via Modality Balancing
mmfp0327 Enhancing Domain-Invariant Parts for Generalized Zero-Shot Learning
mmfp2380 DCEL: Deep Cross-Modal Evidential Learning for Text-Based Person Retrieval
mmfp2273 Tran-GCN: Multi-label Pattern Image Retrieval via Transformer Driven Graph Convolutional Network
mmfp3289 AdvCLIP: Downstream-agnostic Adversarial Examples in Multimodal Contrastive Learning
mmfp0152 Personalized Behavior-Aware Transformer for Multi-Behavior Sequential Recommendation
mmfp2397 A Contrastive Learning Framework for Dual-Target Cross-Domain Recommendation
mmfp1865 Self-Distillation Dual-Memory Online Hashing with Hash Centers for Streaming Data Retrieval
mmfp4170 Unsupervised Hashing with Contrastive Learning by Exploiting Similarity Knowledge and Hidden Structure of Data
mmfp1803 Giving text more imagination space for image-text matching
mmfp4011 Modal-aware Bias Constrained Contrastive Learning for Multimodal Recommendation
mmfp2927 Precise Target-Oriented Attack against Deep Hashing-based Retrieval
mmfp0865 Conversational Composed Retrieval with Iterative Sequence Refinement
mmfp2562 Hi-SIGIR: Hierarchical Semantic-Guided Image-to-image Retrieval via Scene Graph
mmfp4136 Pareto Invariant Representation Learning for Multimedia Recommendation
mmfp1277 Hashing One With All
mmfp2026 ChinaOpen: A Dataset for Open-world Multimodal Learning
mmfp0840 A Reinforcement Learning-Based Automatic Video Editing Method Using Pre-trained Vision-Language Model
mmfp2403 Knowledge Prompt-tuning for Sequential Recommendation
mmfp3281 Learning Occlusion Disentanglement with Fine-grained Localization for Occluded Person Re-identification
mmfp3141 Interactive Interior Design Recommendation via Coarse-to-fine Multimodal Reinforcement Learning
mmfp2833 Towards Visual Taxonomy Expansion
mmfp1672 Enhancing Product Representation with Multi-form Interactions for Multimodal Conversational Recommendation
mmfp0776 Stepwise Refinement Short Hashing for Image Retrieval
mmfp2237 Knowledge Decomposition and Replay: A Novel Cross-modal Image-Text Retrieval Continual Learning Method
mmfp0441 Striking a Balance: Unsupervised Cross-Domain Crowd Counting via Knowledge Diffusion
mmfp3016 Task-Adversarial Adaptation for Multi-modal Recommendation
mmfp3503 Counterfactual Cross-modality Reasoning for Weakly Supervised Video Moment Localization
mmfp1218 Pre-training and Transferring Multi-modal Interest-aware Sequence Representation for Recommendation
mmfp1547 Attributes Grouping and Mining Hashing for Fine-Grained Image Retrieval
mmfp0867 Semantic-Guided Feature Distillation for Multimodal Recommendation
mmfp0779 Multi-View Graph Convolutional Network for Multimedia Recommendation