

ACM ICMR 2024 Full Program

<i>MON 10 JUN</i>	<i>WORKSHOPS DAY 1</i>	<i>Room</i>
08:30 – 09:00	Registration Open	
09:00 – 17:30	LSC 2024	Dusit Laguna
	<p>The 7th Lifelog Search Challenge. Cathal Gurrin, Björn Þór Jónsson, Duc-Tien Dang-Nguyen, Jakub Lokoc, Klaus Schoeffmann, Minh-Triet Tran, Steve Hodges, Graham Healy, Luca Rossetto, Werner Bailer</p> <p>Expert session at 13:30, novice session at 16:00.</p>	
09:00 – 12:30	AI-SIPM 2024	Informal
	<p>International Workshop on Artificial Intelligence for Signal, Image Processing and Multimedia. Mahasak Ketcham, Kanyalag Phodong, Patiyuth Pramkeaw, Worawut Yimyam, Narumol Chumuang, Pokpong Songmuang, Thittaporn Ganokratanaa</p>	
09:00 – 12:30	MORE 2024	Suriyan
	<p>Workshop on Multimedia Object Re-ID: Advancements, Challenges, and Opportunities. Zhedong Zheng, Yaxiong Wang, Xuelin Qian, Zhun Zhong, Zheng Wang, Liang Zheng</p>	
09:00 – 17:30	MAD 2024	Chantra
	<p>3rd ACM International Workshop on Multimedia AI against Disinformation. Bogdan Ionescu, Dan-Cristian Stanciu, Luca Cuccovillo, Giorgos Kordopatis-Zilos, Symeon Papadopoulos, Adrian Popescu, Roberto Caldelli</p>	
09:00 – 12:30	AIQAM 2024	Dara
	<p>The 1st ACM Workshop in AI-Powered Question Answering Systems for Multimedia. Tai Tan Mai, Ly-Duyen Tran, Quang-Linh Tran, Cathal Gurrin, Duc-Tien Dang-Nguyen, Van-Tu Ninh</p>	
10:30 – 11:00	Coffee Break	
12:30 – 13:30	Lunch	
13:30 – 17:30	ICDAR 2024	Suriyan
	<p>The 5th International Workshop on Intelligent Cross-Data Analysis and Retrieval. Minh-Son Dao, Michael A. Riegler, Duc-Tien Dang-Nguyen, Hanh-Nhi Tran, RAGE Uday Kiran, Takahiro Komamizu</p>	
13:30 – 17:30	MUWS 2024	Dara
	<p>The 3rd International Workshop on Multimodal Human Understanding for the Web and Social Media. Marc A. Kastner, Gullal S. Cheema, Sherzod Hakimov, Noa Garcia</p>	
17:30 – 19:00	Welcome Reception	

<i>TUE 11 JUN</i>	<i>MAIN CONFERENCE DAY 1</i>	<i>Room</i>
08:30 – 09:15	Registration Open	
09:15 – 09:30	Welcome to ICMR 2024	Dusit Laguna
09:30 – 10:30	Keynote 1 Chair: Shin'ichi Satoh	Dusit Laguna
	Multi-Modal Retrieval in Computer Vision. Mubarak Shah (University of Florida, FL, USA)	
10:30 – 11:00	Coffee Break	
	Session 1 (Best Paper Session) Chair: Luca Rossetto Presentation time per paper: 12 min (3 min Q/A)	Dusit Laguna
11:00 – 11:15	Known-item Search in Video: An Eye Tracking-based Study. Lucas Joos, Bastian Jäckl, Daniel A. Keim, Maximilian T. Fischer, Ladislav Peška, Jakub Lokoč	
11:15 – 11:30	Multi-modal Entity Alignment via Position-enhanced Multi-label Propagation. Wei Tang, Yuanyi Wang	
11:30 – 11:45	Identification of Speaker Roles and Situation Types in News Videos. Gullal Singh Cheema, Judi Arafat, Chiao-I Tseng, John A. Bateman, Ralph Ewerth, Eric Müller-Budack	
11:45 – 12:00	Component-Level Oracle Bone Inscription Retrieval. Zhikai Hu, Yiu-ming Cheung, Yonggang Zhang, Peiyang Zhang, Pui-ling Tang	
	Session 2 (Deep Learning) Chair: Konstantin Schall Presentation time per paper: 12 min (3 min Q/A)	Dusit Laguna
12:00 – 12:15	HashNeck is a Boosting Tool for Deep Learning to Hashing. Hua Gao, ChenChen Hu, Guang Han, Jiafa Mao, Wei Huang, Kaiyuan Wan	
12:15 – 12:30	Prompt Expending for Single Positive Multi-Label Learning with Global Unannotated Categories. Zhongnian Li, Peng Ying, Meng Wei, Tongfeng Sun, Xinzheng Xu	
12:30 – 12:45	GSD-GNN: Generalizable and Scalable Algorithms for Decoupled Graph Neural Networks. Yunfeng Yu, Longlong Lin, Qiyu Liu, Zeli Wang, Xi Ou, Tao Jia	
12:45 – 13:00	Causal Inference-based Few-Shot Class-Incremental Learning. Weiwei Zhou, Guoqiang Xiao, Michael S. Lew, Song Wu	
13:00 – 14:00	Lunch	
13:00 – 14:00	Diversity Lunch	
	Diversity in Multimedia. Phoebe Chen (La Trobe University, Australia)	on invite
14:00 – 16:30	Poster Boaster Chair: Duc-Tien Dang-Nguyen Presentation time per paper: 1 min / 1 slide	Dusit Laguna
	Authors of papers in the on-site poster, demo, and reproducibility session (17:00 – 18:30) present a 1-minute teaser introduction to their poster.	
14:00 – 16:30	Tutorial 1	Suryan
	Fine-Tuning Large Language Models for Private Document Retrieval. Frank Sommers, Alisa Kongthon, and Sarawoot Kongyoung	

16:30 – 17:00	Coffee Break	
17:00 – 18:30	Posters Chair: Duc-Tien Dang-Nguyen Presented on-site (A1 Portrait format)	Dusit Laguna
	CoDancers: Music-Driven Coherent Group Dance Generation with Choreographic Unit. Kaixing Yang, Xulong Tang, Ran Diao, Hongyan Liu , Jun He, Zhaoxin Fan	
	Federated Multi-Task Learning on Non-IID Data Silos: An Experimental Study. Yuwen Yang , Yuxiang Lu, Suizhi Huang, Shalayiding Sirejiding, Hongtao Lu, Yue Ding	
	Refracting Once is Enough: Neural Radiance Fields for Novel-View Synthesis of Real Refractive Objects. Xiaoqian Liang , Jianji Wang, Yuanliang Lu, Xubin Duan, Xichun Liu, Nanning Zheng	
	MFVG: A Visual Grounding Network with Multi-scale Fusion. Peijia Chen , Ke Qi, Xi Tao, Wenhao Xu, Jingdong Zhang	
	SFAM: Lightweight Spectrum Unreferenced Attention Network. Xuanhao Qi , Min Zhi, Yanjun Yin, Ping Ping, Yuening Zhang	
	FaceX: Understanding Face Attribute Classifiers through Summary Model Explanations. Ioannis Sarridis , Christos Koutlis, Symeon Papadopoulos, Christos Diou	
	Pattern4Ego: Learning Egocentric Video Representation Using Cross-video Activity Patterns. Ruihai Wu , Yourong Zhang, Yu Qi, Andy Guanhong Chen, Hao Dong	
	Neural Parametric Human Hand Modeling with Point Cloud Representation. Jian Yang , Weize Quan, Zhen Shen, Dong-Ming Yan, Huai-Yu Wu	
	HybridHash: Hybrid Convolutional and Self-Attention Deep Hashing for Image Retrieval. Chao He , Hongxi Wei	
	ML²MG-VLCR: A Multimodal LLM Guided Zero-shot Method for Visio-Linguistic Compositional Reasoning with Autoregressive Generative Language Model. Ziyu Gong, Chengcheng Mai , Yihua Huang	
	DualStyle3D: Real-time Exemplar-based Artistic Portrait View Synthesis Based on Radiance Field. Runlai Hao , Jinlong Li, Qingqiu Chen, Huanhuan Chen	
	TrustGo: Trust Mining and Multi-semantic Regularization in Social Recommendation. Shenghao Liu, Yuqin Lan , Xianjun Deng, Lingzhi Yi, Chenlu Zhu, Laurence Yang, Jong Park	
	STDG: Semi-Teacher-Student Training Paradigm for Depth-guided One-stage Scene Graph Generation. Xukun Zhou, Zhenbo Song, Jun He , Hongyan Liu, Zhaoxin Fan	

	<p>BFIDet: A YOLOv7-improved Vehicle and Pedestrian Detector via Balancing Feature Integration. <u>Anrui Wang</u>, Libo Weng, Fei Gao</p>	
	<p>Self-Supervised Multi-Label Classification with Global Context and Local Attention. Chun-Yen Chen, <u>Mei-Chen Yeh</u></p>	
	<p>End-to-End Thai Text-to-Speech with Linguistic Unit. Kontawat Wisetpaitoon, <u>Sattaya Singkul</u>, Theerat Sakdejayont, Tawunrat Chalothorn</p>	
	<p>Enhancing Interactive Retrieval With Query Rewriting Using Large Language Models and Vision Language Models. Hongyi Zhu, Jia-Hong Huang, <u>Stevan Rudinac</u>, Evangelos Kanoulas</p>	
	<p>Lifelong Visible-Infrared Person Re-Identification via a Tri-Token Transformer with a Query-Key Mechanism. <u>Yitong Xing</u>, Guoqiang Xiao, Michael S. Lew, Song Wu</p>	
	<p>Wireless Capsule Endoscope Low-light Image Enhancement with Balanced Brightness and Saturation. <u>Wenzhuo Li</u>, Yinghui Wang, Wei Li, Liangyi Huang, Kamoliddin Shukurov, Mingfeng Wang</p>	
	<p>CLIPping the Deception: Adapting Vision-Language Models for Universal Deepfake Detection. <u>Sohail Ahmed Khan</u>, Duc-Tien Dang-Nguyen</p>	
	<p>RGB-D Video Object Segmentation via Enhanced Multi-store Feature Memory. <u>Boyue Xu</u>, Ruichao Hou, Tongwei Ren, Gangshan Wu</p>	
	<p>AdOCteRA: Adaptive Optimization Constraints for improved Text-guided Retrieval of Apartments. Ali Abdari, <u>Alex Falcon</u>, Giuseppe Serra</p>	
	<p>Parametric CAD Primitive Retrieval via Multi-Modal Fusion and Deep Hashing. <u>Minyang Xu</u>, Yunzhong Lou, Weijian Ma, Xueyang Li, Xiangdong Zhou</p>	
	<p>Knowledge Distillation for Single Image Super-Resolution via Contrastive Learning. <u>Cencen Liu</u>, Dongyang Zhang, Ke Qin</p>	
	<p>Proactive Privacy and Intellectual Property Protection of Multimedia Retrieval Models in Edge Intelligence. <u>Peihao Li</u>, Jie Huang, Shuaishuai Zhang, Chunyang Qi</p>	
	<p>Directly Locating Actions in Video with Single Frame Annotation. <u>Haoran Tong</u>, Xinyan Liu, Guorong Li, Laiyun Qing</p>	
	<p>Content-Based Exclusion Queries in Keyword-Based Image Retrieval. <u>Eisaku Yoshikawa</u>, Keishi Tajima</p>	
	<p>Unveiling Global Narratives: A Multilingual Twitter Dataset of News Media on the Russo-Ukrainian Conflict. <u>Sherzod Hakimov</u>, Gullal S. Cheema</p>	
	<p>Near-Miss Accident Prediction on the Edge: A Real-Time System for Safer Driving. <u>Minh-Son Dao</u>, Koji Zettsu</p>	
	<p>FedPAM: Federated Personalized Augmentation Model for Text-to-Image Retrieval.</p>	

	<u>Yueying Feng</u> , Fan Ma, Wang Lin, Chang Yao, Jingyuan Chen, Yi Yang	
	Overview of the Grand Challenge on Detecting Cheapfakes at ACM ICMR 2024. <u>Duc-Tien Dang-Nguyen</u> , Sohail Ahmed Khan, Michael Riegler, Pål Halvorsen, Anh-Duy Tran, Minh-Son Dao, Minh-Triet Tran	
	Detecting Out-of-Context Media with LLaMa-Adapter V2 and RoBERTa: An Effective Method for Cheapfakes Detection. Hoa-Vien Vo-Hoang, Long-Khanh Pham, <u>Minh-Son Dao</u>	
	A Generative Adaptive Context Learning Framework for Large Language Models in Cheapfake Detection. <u>Long-Khanh Pham</u> , Hoa-Vien Vo-Hoang, Anh-Duy Tran	
	TeGA: A Text-Guided Generative-based Approach in Cheapfake Detection. Anh-Thu Le, Minh-Dat Nguyen, Minh-Son Dao, Anh-Duy Tran, <u>Duc-Tien Dang-Nguyen</u>	
	A Unified Network for Detecting Out-Of-Context Information Using Generative Synthetic Data. Van-Loc Nguyen, Bao-Tin Nguyen, Thanh-Son Nguyen, Duc-Tien Dang-Nguyen, <u>Minh-Triet Tran</u>	
	Enhancing Cheapfake Detection: An Approach Using Prompt Engineering and Interleaved Text-Image Model. <u>Dang Vu</u> , Minh-Nhat Nguyen, Quoc-Trung Nguyen	
	A Multi-Stage Deep Learning Approach Incorporating Text-Image and Image-Image Comparisons for Cheapfake Detection. <u>Jangwon Seo</u> , Hyo-Seok Hwang, Jiyoung Lee, Minhyeok Lee, Wonsuk Kim, Junhee Seok	
	Contrastive pre-training with Multi-Level Alignment for Grounded Multimodal Named Entity Recognition. Xigang Bao, Mengyuan Tian, Luyao Wang, Zhiyuan Zha, Biao Qin	
17:00 – 18:30	Demos Chair: Klaus Schoeffmann Presented as live-demonstrations and posters (A1 Portrait)	Suryan and Chantra
	OpenLifelogCam - A Low-Cost Open-Source Wearable Camera Platform. <u>Luca Rossetto</u>	
	CarAI: Car Inspection with Artificial Intelligence. <u>Panumate Chetprayoon</u> , Sakol Tasanangam, Gayatri Tirumalasetty, Thanatwit Angsarawanee, Paveen Virameteekul, Wadeepas Lertwatanawanich, Theerat Sakdejyont	
	AI Batting Buddy: A Computational and Kinematic Approach for Enhancing Batting Performance and Analysis in Baseball. <u>Kuo-Yu Liu</u> , Ting-Yu Guo, Ta-Shan Pan, Ping-Yi Tung, Yi-Rou Lin	
	A Web Demo Interface for Super-Resolution Reconstruction with Parametric Regularization Loss. <u>Supatta Viriyavisuthisakul</u> , Parinya Sanguansat, <u>Toshihiko Yamasaki</u>	
	MemoriLens: A Low-cost Lifelog Camera using Raspberry Pi Zero. Quang-Linh Tran, Binh Nguyen, Gareth J. F. Jones, <u>Cathal Gurrin</u>	

	<p>3DMSE: An Interactive 3D Media Search Engine. <u>Maria Eirini Pegia</u>, Dimitris Georgalis, Nick Pantelidis, Björn Þór Jónsson, Anastasia Moutzidou, Sotiris Diplaris, Ilias Gialampoukidis, Stefanos Vrochidis, Ioannis Kompatsiaris</p>	
	<p>Conversational Image Search: A Sketch-based Approach. Daniel-Dracos Braghis, <u>Haiming Liu</u></p>	
	<p>RE-IDVIS: Person Re-Identification System based on Interactive Visualization. <u>Wang Xia</u>, Guodao Sun, Zihao Zhu, Pan Liang, Sujia Zhu, Yiming Wu, Haoran Liang, Ronghua Liang</p>	
17:00 – 18:30	<p>Reproducibility Papers Chair: Klaus Schoeffmann Presented as posters (A1 Portrait format)</p>	Suryan and Chandra
	<p>Reproducibility Companion Paper: Recommendation of Mix-and-Match Clothing by Modeling Indirect Personal Compatibility. Shuiying Liao, Yujuan Ding, P. Y. Mok, Qiushi Huang, Jialun Cao</p>	<i>virtual</i>
	<p>Reproducibility Companion Paper: Stable Diffusion for Content-Style Disentanglement in Art Analysis. <u>Yankun Wu</u>, Yuta Nakashima, Noa Garcia, Sheng Li, Zhaoyang Zeng</p>	
	<p>Reproducibility Companion Paper of “MMSF: A Multimodal Sentiment-Fused Method to Recognize Video Speaking Style”. <u>Fan Yu</u>, Beibei Zhang, Yaqun Fang, Jia Bei, Tongwei Ren, Jiyi Li, Luca Rossetto</p>	

<i>WED 12 JUN</i>	<i>MAIN CONFERENCE DAY 2</i>	<i>Room</i>
08:30 – 09:30	Registration Open	
09:30 – 10:30	Keynote 2 Chair: Rachada Kongkachandra AI-Based Video Analytics. Supavadee Aramvith (Chulalongkorn University, Thailand)	Dusit Laguna
10:30 – 11:00	Coffee Break	
	Session 3 (Deepfake Detection and Security) Chair: Shin'ichi Satoh Presentation time per paper: 12 min (3 min Q/A)	Dusit Laguna
11:00 – 11:15	CodeDetector: Revealing Forgery Traces with Codebook for Generalized Deepfake Detection. <u>Jiaxin Li</u> , Zhihan Yu, Guibo Luo, Yuesheng Zhu	
11:15 – 11:30	Generative Data Augmentation with Liveness Information Preserving for Face Anti-Spoofing. <u>Changgu Chen</u> , Yang Li, Jian Zhang, Jiali Liu, Changbo Wang	
11:30 – 11:45	Retrieval-Augmented Audio Deepfake Detection. <u>Zuheng Kang</u> , Yayun He, Botao Zhao, Xiaoyang Qu, Junqing Peng, Jing Xiao, Jianzong Wang	
11:45 – 12:00	Context or Clutter? Efficiently Matching Objects Across Scenes. <u>Albatool Wazzan</u> , Imtiaz Ahmad, Stephen Macneil, Richard Souvenir	
	Session 5 (Doctoral Symposium) Chair: Stevan Rudinac Presentation time per paper: 30 min (30 min Q/A)	Suryan
11:00 – 12:00	Multi-modal Video Summarization. <u>Jia-Hong Huang</u>	
12:00 – 13:00	Multimodality in Media Retrieval. <u>Maria Eirini Pegia</u>	
	Session 4 (Image and Video Analysis) Chair: Chong-Wah Ngo Presentation time per paper: 12 min (3 min Q/A)	Dusit Laguna
12:00 – 12:15	Enhancing Class-Incremental Learning for Image Classification via Bidirectional Transport and Selective Momentum. <u>Feifei Fu</u> , Yizhao Gao, Zhiwu Lu	
12:15 – 12:30	Smart Fitting Room: A One-stop Framework for Matching-aware Virtual Try-On. Mingzhe Yu, <u>Yunshan Ma</u> , Lei Wu, Kai Cheng, Xue Li, Lei Meng, Tat-Seng Chua	
12:30 – 12:45	Vector-Aware Anisotropic Gauge Equivariant Mesh Convolution Network for 3D Aneurysm Detection. <u>Xudong Ru</u> , Haichuan Zhao, Xingce Wang, Zhongke Wu, Shaolong Liu, Yi-Cheng Zhu, Alejandro F. Frangi	
12:45 – 13:00	Semantic-guided RGB-Thermal Crowd Counting with Segment Anything Model. <u>Yaqun Fang</u> , Yi Shi, Jia Bei, Tongwei Ren	
13:00 – 14:00	Lunch	

	Session 6 (Retrieval Frameworks) Chair: Jakub Lokoc Presentation time per paper: 12 min (3 min Q/A)	Dusit Laguna
14:00 – 14:15	ELSEIR: A Privacy-Preserving Large-Scale Image Retrieval Framework for Outsourced Data Sharing. <u>Zixin Tang</u> , Haihui Fan, Xiaoyan Gu, Yang Li, Bo Li, Xin Wang	
14:15 – 14:30	BeatDance: A Beat-Based Model-Agnostic Contrastive Learning Framework for Music-Dance Retrieval. Kaixing Yang, Xukun Zhou, Xulong Tang, Ran Diao, Hongyan Liu, <u>Jun He</u> , Zhaoxin Fan	
14:30 – 14:45	Anchor-aware Deep Metric Learning for Audio-visual Retrieval. <u>Donghuo Zeng</u> , Yanan Wang, Kazushi Ikeda, Yi Yu	
14:45 – 15:00	Improving Video Corpus Moment Retrieval with Partial Relevance Enhancement. <u>Danyang Hou</u> , Liang Pang, Huawei Shen, Xueqi Chen	
15:00 – 15:15	A Parallel Transformer Framework for Video Moment Retrieval. <u>Thao-Nhu Nguyen</u> , Zongyao Li, Yamazaki Satoshi, <u>Jianquan Liu</u> , Cathal Gurrin	
14:00 – 17:00	Tutorial 2	Suryan
	Multimedia Retrieval in and for XR. Maria Pegia, Sotiris Diplaris, Stefanos Vrochidis, Heiko Schuldt, Florian Spiess, Rahel Arnold, and Werner Bailer	
15:30 – 16:00	Coffee Break	
	Session 7 (Pose Detection) Chair: Cathal Gurrin Presentation time per paper: 12 min (3 min Q/A)	Dusit Laguna
16:00 – 16:15	ExpoGenius: Robust Personalized Human Image Generation using Diffusion Model for Exposure Variation and Pose Transfer. <u>Depei Liu</u> , Hongjie Fan, Junfei Liu	
16:15 – 16:30	PoseRec: 3D Human Pose Driven Online Advertisement Recommendation for Micro-videos. Zhaoxin Fan, Fengxin Li, Hongyan Liu, <u>Jun He</u> , Xiaoyong Du	
16:30 – 16:45	Monocular Expressive 3D Human Reconstruction of Multiple People. Zhenghao Zhao, Hao Tang, Joy Wan, <u>Yan Yan</u>	
16:45 – 17:00	ACR-Pose: Adversarial Canonical Representation Reconstruction Network for Category Level 6D Object Pose Estimation. Zhaoxin Fan, Zhenbo Song, Zhicheng Wang, Jian Xu, Kejian Wu, Hongyan Liu, <u>Jun He</u>	
17:00 – 18:30	Virtual Posters	Online
	Multimodal Prototype-Enhanced Network for Few-Shot Action Recognition. Xinzhe Ni, Yong Liu, Hao Wen, Yatai Ji, Jing Xiao, Yujiu Yang	
	Triadic Elastic Structure Representation for Open-Set Incremental 3D Object Retrieval. Yang Xu, Yifan Feng, Lin Bie	
	Dynamic Segmentation for Efficient Retrieval of Podcasts: The Repping Algorithm. Stephan Repp, Ernst Georg Haffner	

	Progressive Multi-modal Conditional Prompt Tuning. Xiaoyu Qiu, Hao Feng, Yuechen Wang, Wengang Zhou, Houqiang Li	
	Fine-grained Semantics-aware Representation Learning for Text-based Person Retrieval. Di Wang, Feng Yan, Yifeng Wang, Lin Zhao, Xiao Liang, Haodi Zhong, Ronghua Zhang	
	CMFF-Face: Attention-Based Cross-Modal Feature Fusion for High-Quality Audio-Driven Talking Face Generation. Guangzhe Zhao, Yanan Liu, Xueping Wang, Feihu Yan	
	Fingerprinting in EEG Model IP Protection Using Diffusion Model. Tianyi Wang, Sheng-hua Zhong	
	MarginFinger: Controlling Generated Fingerprint Distance to Classification boundary Using Conditional GANs. Weixing Liu, Shenghua Zhong	
	Improve Deep Hashing with Language Guidance for Unsupervised Image Retrieval. Chuang Zhao, Hefei Li, Shijie Lu, Yuxuan Shi, Jiazhong Chen, Ping Li	
	Exploiting Degradation Prior for Personalized Federated Learning in Real-World Image Super-Resolution. Yue Yang, Liangjun Ke	
	QAVidCap: Enhancing Video Captioning through Question Answering Techniques. Hui Liu, Xiaojun Wan	
	Targeted Universal Adversarial Attack on Deep Hash Networks. Fanlei Meng, Xiangru Chen, Yuan Cao	
	Secure Verification Encrypted Image Retrieval Scheme with Addition Homomorphic Bitmap Index. Mingyue Li, Yuting Zhu, Ruizhong Du, Chunfu Jia	
	A Novel Auxiliary Task Framework in 3D Human Pose Estimation for Opera Videos. Xingquan Cai, Haoyu Zhang, Shanshan He, Haoyu Song, Haiyan Sun	
	Dynamic Soft Labeling for Visual Semantic Embedding. Jiaao Yu, Yunlai Ding, Junyu Dong, Yuezun Li	
	Navigating Style Variations in Scene Text Image Super-Resolution through Multi-Scale Perception. Feifei Xu, Ziheng Yu	
	NeurNCD: Novel Class Discovery via Implicit Neural Representation. Junming Wang, Yi Shi	
	Image-to-Point Registration via Cross-Modality Correspondence Retrieval. Lin Bie, Siqi Li, Kai Cheng	
	TWIST: Text-only Weakly Supervised Scene Text Spotting Using Pseudo Labels. Lilong Wen, Xiu Tang, Dongxiang Zhang	
	A Graph Convolution Network with a POS-aware Filter and Context Enhancement Mechanism for Event Detection. Xintao Jiao, Jiansheng Chen, Jiale Liu	

	<p>Pseudo Content Hallucination for Unpaired Image Captioning. Huixia Ben, Shuo Wang, Meng Wang, Richang Hong</p>	
	<p>Deep Image Clustering Based on Curriculum Learning and Density Information. Haiyang Zheng, Ruilin Zhang, Hongpeng Wang</p>	
	<p>Text Adversarial Defense via Granular-Ball Sample Enhancement. Zeli Wang, Jian Li, Shuyin Xia, Longlong Lin, Guoyin Wang</p>	
	<p>GB_{RAIN}: Combating Textual Label Noise by Granular-ball based Robust Training. Zeli Wang, Tuo Zhang, Shuyin Xia, Longlong Lin, Guoyin Wang</p>	
	<p>Multi-view Counterfactual Contrastive Learning for Fact-checking Fake News Detection. Yongcheng Zhang, Lingou Kong, Sheng Tian, Hao Fei, Changpeng Xiang, Huan Wang, Xiaomei Wei</p>	
	<p>A Lightweight Surface Defect Segmentation Network with External Semantics and High-frequency Information. Tianpeng Zhang, Xuesong Jiang</p>	
	<p>A Causal View for Multi-Interest User Modeling in News Recommendation. Mei Yu, Xiaoxi Zhou, Mankun Zhao, Tianyi Xu, Yue Zhao, Ruiguo Yu, Xuwei Li</p>	
	<p>Comment-aided Video-Language Alignment via Contrastive Pre-training for Short-form Video Humor Detection. Yang Liu, Tongfei Shen, Dong Zhang, Qingying Sun, Shoushan Li, Guodong Zhou</p>	
	<p>VEC-MNER: Hybrid Transformer with Visual-Enhanced Cross-Modal Multi-level Interaction for Multimodal NER. Pengfei Wei, Hongjun Ouyang, Qintai Hu, Bi Zeng, Guang Feng, Qingpeng Wen</p>	
	<p>S2F-Net: Shared-Specific Fusion Network for Infrared and Visible Image Fusion. Yijing Zhao, Yuchao Xia, Yi Ding, Yumeng Liu, Shuai Liu, Hongan Wang</p>	
	<p>CGI-MRE: A Comprehensive Genetic-Inspired Model For Multimodal Relation Extraction. Pengfei Wei, Zhaokang Huang, Hongjun Ouyang, Qintai Hu, Bi Zeng, Guang Feng</p>	
	<p>Speak From Heart: An Emotion-Guided LLM-Based Multimodal Method for Emotional Dialogue Generation. Chenxiao Liu, Zheyong Xie, Sirui Zhao, Jin Zhou, Tong Xu, Minglei Li, Enhong Chen</p>	
	<p>Multi-Source Augmentation and Composite Prompts for Visual Recognition with Missing Modality. Zhirui Kuai, Yulu Zhou, Qi Xie, Li Kuang</p>	
	<p>TriMPL: Masked Multi-Prompt Learning with Knowledge Mixing for Vision-Language Few-shot Learning. Xiangyu Liu, Yanlei Shang, Yong Chen</p>	
	<p>Visibility-guided Human Body Reconstruction from Uncalibrated Multi-view Cameras. Zhenyu Xie, Huanyu He, Gui Zou, Jie Wu, Guoliang Liu, Jun Zhao, Yingxue Wang, Hui Lin, Weiyao Lin</p>	

	SamCap: Energy-based Controllable Image Captioning by Gradient-Based Sampling. Yuchen Niu, Min Zhu, Zhihua Wei	
	PTAN: Principal Token-aware Adjacent Network for Compositional Temporal Grounding. Zhuoyuan Wei, Xun Jiang, Zheng Wang, Fumin Shen, Xing Xu	
	A Hybrid Few-Shot Image Classification Framework Combining Gaussian Modeling and Label Propagation. Chao Ye, Qian Wang, Lanfang Dong	
	Intra and Inter-modality Incongruity Modeling and Adversarial Contrastive Learning for Multimodal Fake News Detection. Siqi Wei, Bin Wu	
	Team HUGE: Image-Text Matching via Hierarchical and Unified Graph Enhancing. Bo Li, You Wu, Zhixin Li	
	When Handcrafted Filter Meets CNN: A Lightweight Conv-Filter Mixer Network for Efficient Image Super-Resolution. Zhijian Wu, Wenhui Liu, Dingjiang Huang	
	Modality-specific and -shared Contrastive Learning for Sentiment Analysis. Dahuang Liu, Jiuxiang You, Guobo Xie, Lap Kei Lee, Fu Lee Wang, Zhenguo Yang	
	Pyramidal Cross-Modal Transformer with Sustained Visual Guidance for Multi-Label Image Classification. Zhuohua Li, Ruyun Wang, Fuqing Zhu, Jizhong Han, Songlin Hu	
	Low-Light Image Enhancement via Weighted Low-Rank Tensor Regularized Retinex Model. Weipeng Yang, Hongxia Gao, Wenbin Zou, Tongtong Liu, Shasha Huang, Jianliang Ma	
	Multi-view Subspace Clustering via An Adaptive Consensus Graph Filter. Lai Wei, Shanshan Song	
	MSI: Multi-modal Recommendation via Superfluous Semantics Discarding and Interaction Preserving. Yi Li, Qingmeng Zhu, Changwen Zheng, Jiangmeng Li	
	Modeling Multi-Task Joint Training of Aggregate Networks for Multi-Modal Sarcasm Detection. Lisong Ou, Zhixin Li	
	Deep Scaling Factor Quantization Network for Large-scale Image Retrieval. Ziqing Deng, Zhihui Lai, Yujuan Ding, Heng Kong, Xu Wu	
	RetrievalMMT: Retrieval-Constrained Multi-Modal Prompt Learning for Multi-Modal Machine Translation. Yan Wang, Yawen Zeng, Junjie Liang, Xiaofen Xing, Jin Xu, Xiangmin Xu	
	SBCR: Stochasticity Beats Content Restriction Problem in Training and Tuning Free Image Editing. Jiancheng Huang, Mingfu Yan, Yifan Liu, Shifeng Chen	
	SkeletonFormer: Point Cloud Completion with Dynamic Selective Skeleton Points. Beiqi Liu, Fuqing Duan, Junli Zhao	

	<p>Research on Epilepsy Classification Model Based on Variational Mode Quadratic Decomposition. Chen Huang, Zhijun Fan, Kui Xiao, Yan Zhang, Shihui Wang, Jianhua Song, Wei Wu</p>	
	<p>Semi-Parametric Style Transfer with Multi-Perspective Feature Fusion and Information-Guided Alignment. Tianlong Zhang, Jing Lv, Ming Yang</p>	
	<p>Discovering Multi-Relational Integration for Knowledge Tracing with Retentive Networks. Zhou Linhao, Zhong Shenghua, Xiao Zhijiao</p>	
	<p>DeepEnhancer: Temporally Consistent Focal Transformer for Comprehensive Video Enhancement. Qin Jiang, Qinglin Wang, Lihua Chi, Wentao Ma, Feng Li, Jie Liu</p>	
	<p>Multidimensional Semantic Disentanglement Network for Clothes-Changing Person Re-Identification. Yongkang Ding, Anqi Wang, Liyan Zhang</p>	
	<p>UBiSS: A Unified Framework for Bimodal Semantic Summarization of Videos. Yuting Mei, Linli Yao, Qin Jin</p>	
	<p>G-SAP: Graph-based Structure-Aware Prompt Learning over Heterogeneous Knowledge for Commonsense Reasoning. Ruiting Dai, Yuqiao Tan, Lisi Mo, Shuang Liang, Guohao Huo, Jiayi Luo, Yao Cheng</p>	
	<p>Subspace Clustering with A Hybrid Adaptive Graph Filter. Lai Wei, Mingyuan Xi</p>	
	<p>Improving Data Augmentation for Robust Visual Question Answering with Effective Curriculum Learning. Yuhang Zheng, Zhen Wang, Long Chen</p>	
	<p>CLCP: Realtime Text-Image Retrieval for Retailing via Pre-trained Clustering and Priority Queue. Shuyang Zhang, Liangwu Wei, Qingyu Wang, Yuntao Wei, Yanzhi Song</p>	
	<p>A Knowledge-Driven Approach to Enhance Topic Modeling with Multi-Modal Representation Learning. Hongzhang Mu, Shuili Zhang, Hongbo Xu</p>	
	<p>Robust Video Hashing with Non-negative Tensor Factorization for Copy Detection. Mengzhu Yu, Zhenjun Tang, Huijiang Zhuang, Xiaoping Liang, Zhixin Li, Xianquan Zhang</p>	
	<p>Unifying Pictorial and Textual Features for Screen Content Image Quality Evaluation. Yihua Chen, Xiaoping Liang, Mengzhu Yu, Zhenjun Tang</p>	
	<p>CLIP-ProbCR: CLIP-based Probability Embedding Combination Retrieval. Mingyong Li, Zongwei Zhao, Xiaolong Jiang, Zheng Jiang</p>	
	<p>Sketch-aided Interactive Fusion Point Cloud Place Recognition. Ruonan Zhang, Xiaohang Liu, Ge Li, Thomas.H Li, Pengjun Zhao</p>	
	<p>TIM: Temporal Interaction Model in Notification System. Huxiao Ji, Haitao Yang, Linchuan Li, Shunyu Zhang, Cunyi Zhang, Xuanping Li, Wenwu Ou</p>	

	<p>Local Deep Learning Quantization for Approximate Nearest Neighbor Search. Quan Li, Xike Xie, Chao Wang, Jiali Weng</p>	
	<p>DiffHarmony: Latent Diffusion Model Meets Image Harmonization. Pengfei Zhou, Fangxiang Feng, Xiaojie Wang</p>	
	<p>Octree-Retention Fusion: A High-Performance Context Model for Point Cloud Geometry Compression. Zhikang Zhang, Zhongjie Zhu, Yongqiang Bai, Ming Wang, Zhijing Yu</p>	
	<p>A GAN based Video Summarization Method with Representation Loss. Zhuo Lei, Qiang Yu, Lidan Shou, Shengquan Li, Yunqing Mao</p>	
	<p>Extending CLIP for Text-to-font Retrieval. Qinghua Sun, Jia Cui, Zhenyu Gu</p>	
	<p>YawnNet: A Visual-Centric Approach for Yawning Detection. Ruoxi Sun, Xinyu Yang, Cong Qian, Chenyu Zhu, Wei Sui, Zeyd Boukhers, Cong Yang</p>	
	<p>CLTalk: Speech-Driven 3D Facial Animation with Contrastive Learning. Xitie Zhang, Suping Wu</p>	
	<p>PiCoGen: Generate Piano Covers with a Two-stage Approach. Chih-Pin Tan, Shuen-Huei Guan, Yi-Hsuan Yang</p>	
19:30 – 22:00	<p>Conference Banquet</p>	

THU 13 JUN	MAIN CONFERENCE DAY 3	Room
08:30 – 09:30	Registration Open	
09:30 – 10:30	Panel Chair: Alan Smeaton The LLM Wrecking Ball: Are We About to Lose Decades of Work in Multimedia because of MM-LLMs? Confirmed participants: Kai Barthel, Chong Wah Ngo, Heiko Schuldt, Stefanos Vrochidis	Dusit Laguna
09:30 – 13:00	Tutorial 3 Compact Visual Data Representation for Multimedia Search and Analytics. Shiqi Wang and Xinfeng Zhang	Suryan
10:30 – 11:00	Coffee Break	
	Session 8 (Multimedia Applications) Chair: Klaus Schoeffmann Presentation time per paper: 12 min (3 min Q/A)	Dusit Laguna
11:00 – 11:15	Bringing Video Browsing to Virtual Reality: Empirical Evaluation of a Novel Multimedia Drawer. <u>Florian Spiess</u> , Nicolas Scharowski, Ariane Haller, Zgjim Memeti, Heiko Schuldt, Florian Brühlmann	
11:15 – 11:30	An Exploration Graph with Continuous Refinement for Efficient Multimedia Retrieval. Nico Hezel, <u>Kai Uwe Barthel</u> , Konstantin Schall, Klaus Jung	
11:30 – 11:45	FEST: A Multi-way Framework with Enhanced Spatial-Temporal Modeling for Traffic Forecasting. <u>Yilin Li</u> , Tszyin Guo, Ying Qiao, Zitong Bo, Hongan Wang	
11:45 – 12:00	Improving Interpretable Embeddings for Ad-hoc Video Search with Generative Captions and Multi-word Concept Bank. Jiaxin Wu, <u>Chong Wah Ngo</u> , Wing Kwong Chan	
	Session 9 (Brave New Ideas) Chair: Thittaporn Ganokratanaa Presentation time per paper: 12 min (3 min Q/A)	Dusit Laguna
12:00 – 12:15	Reconciling the Rift Between Recognition and Recall: Insights from a Video Memorability Drawing Experiment. Lorin Sweeney, Graham Healy, <u>Alan Smeaton</u>	
12:15 – 12:30	Creating Sorted Grid Layouts with Gradient-based Optimization. Kai Uwe Barthel, Florian Tim Barthel, <u>Peter Eisert</u> , Nico Hezel, Konstantin Schall	
12:30 – 12:45	Mapping the Audio Landscape for Innovative Music Sample Generation. <u>Christian Limberg</u> , Zhe Zhang	
13:00 – 14:00	Lunch	
	Session 10 (Object Detection and Sentiment Analysis) Chair: Werner Bailer Presentation time per paper: 12 min (3 min Q/A)	Dusit Laguna
14:00 – 14:15	Enhancing Visible-Infrared Person Re-identification with Modality- and Instance-aware Visual Prompt Learning. <u>Ruiqi Wu</u> , Bingliang Jiao, Wenxuan Wang, Meng Liu, Peng Wang	
14:15 – 14:30	Learning from Reduced Labels for Long-Tailed Data. <u>Meng Wei</u> , Zhongnian Li, Yong Zhou, Xinzheng Xu	

14:30 – 15:45	Calibration & Reconstruction: Deep Integrated Language for Referring Image Segmentation. Yichen Yan, Xingjian He, Sihan Chen, Jing Liu	
14:45 – 15:00	Retrieving Emotional Stimuli in Artworks. <u>Tianwei Chen</u> , Noa Garcia, Liangzhi Li, Yuta Nakashima	
15:00 – 15:15	A Sentimental Prompt Framework with Visual Text Encoder for Multimodal Sentiment Analysis. <u>Shizhou Huang</u> , Bo Xu, Changqun Li, Jiabo Ye, Xin Lin	
14:00 – 17:30	Tutorial 4	Suryan
	Detecting Misinformation in Photos Utilizing Reverse Image Search. Vinh Dang, Thanh-Son Nguyen, Minh Triet Tran, and Duc-Tien Dang-Nguyen	
15:30 – 16:00	Closing	Dusit Laguna
16:00 – 16:30	Exit Coffee	

FRI 14 JUN	WORKSHOPS DAY 2	<i>Room</i>
08:30 – 09:00	Registration Open	
09:30 – 13:00	RESEARCH2BIZ 2024	Dusit
	Translating Research to Business. João Costa Ribeiro, Fredrik Håland Jensen, Minh-Triet Tran, Duc-Tien Dang-Nguyen, Cathal Gurrin	
09:30 – 17:30	MVRMLM 2024	Laguna
	International Workshop on Multimodal Video Retrieval and Multimodal Language Modelling. Hui Wang, Josef Kittler, Mark Gales, Rob Cooper, Maurice Mulvenna, Wing Ng, Yang Hua, Richard Gault	
10:30 – 11:00	Coffee Break	
13:00 – 14:00	Lunch	
15:30 – 16:00	Coffee Break	