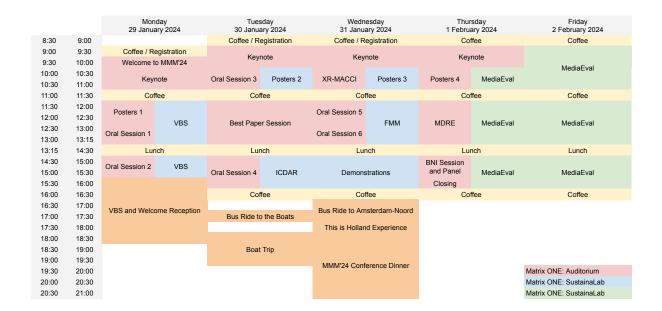# 30th International Conference on Multimedia Modeling
Amsterdam, The Netherlands, 29 January - 2 February, 2024

		Monday 29 January 2024	Tuesday 30 January 2024	Wednesday 31 January 2024	Thursday 1 February 2024	Friday 2 February 2024
8:30	9:00		Coffee / Registration	Coffee / Registration	Coffee	Coffee
9:00	9:30	Coffee / Registration	Keynote	Keynote	Keynote	MediaEval
9:30	10:00	Welcome to MMM'24				
10:00	10:30	Keynote	Oral Session 3 / Posters 2	XR-MACCI / Posters 3	Posters 4 / MediaEval	
10:30	11:00					
11:00	11:30	Coffee	Coffee	Coffee	Coffee	Coffee
11:30	12:00	Posters 1 / VBS	Best Paper Session	Oral Session 5 / FMM	MDRE / MediaEval	MediaEval
12:00	12:30					
12:30	13:00	Oral Session 1		Oral Session 6		
13:00	13:15					
13:15	14:30	Lunch	Lunch	Lunch	Lunch	Lunch
14:30	15:00	Oral Session 2 / VBS	Oral Session 4 / ICDAR	Demonstrations	BNI Session and Panel / MediaEval	MediaEval
15:00	15:30					
15:30	16:00	VBS and Welcome Reception			Closing	
16:00	16:30		Coffee	Coffee	Coffee	Coffee
16:30	17:00		Bus Ride to the Boats	Bus Ride to Amsterdam-Noord		
17:00	17:30					
17:30	18:00			This is Holland Experience		
18:00	18:30					
18:30	19:00		Boat Trip			
19:00	19:30			MMM'24 Conference Dinner		
19:30	20:00					
20:00	20:30					
20:30	21:00					

Matrix ONE: Auditorium
Matrix ONE: SustainaLab
Matrix ONE: SustainaLab

Monday, 29 January 2024

Keynote

- *Visual Analytics: Machine Learning and the Human in the Loop*, Anna Vilanova

Poster Session 1

- *Cross-Modal Hash Retrieval with Category Semantics*, Mengying Xu, Hanjiang Lai, Jian Yin

- *ARE-CAM: An interpretable approach to quantitatively evaluating the adversarial robustness of deep models based on CAM*, Zituo Li, Jianbin Sun, Yuqi Qin, Lunhao Ju, Kewei Yang

- *Structure-aware Adaptive Hybrid Interaction Modeling for Image-Text Matching*, Wei Liu, Jiahuan Wang, Chao Wang, Yan Peng, Shaorong Xie

- *MAMixer: Multivariate Time Series Forecasting via Multi-Axis Mixing*, Liu Yongyu, Lin Guoliang, Lai Hanjiang, Pan Yan

- *A Detail-guided Multi-source Fusion Network for Remote Sensing Object Detection*, Xiaoting Li, Shouhong Wan, Hantao Zhang, Peiquan Jin

- *Bi-directional Interaction and Dense Aggregation Network for RGB-D Salient Object Detection*, Kang Yi, Haoran Tang, Hongyu Bai, Yinjie Wang, Jing Xu, Ping Li

- *Multi-head Hashing with Orthogonal Decomposition for Cross-modal Retrieval*, Wei Liu, Jun Li, Zhijian Wu, Jianhua Xu, Bo Yang

- *A Coarse and Fine Grained Masking Approach for Video-grounded Dialogue*, Feifei Xu, Wang Zhou, Tao Sun, Jiahao Lu, Ziheng Yu, Guangzhen Li

- *Two-Stage Reasoning Network with Modality Decomposition for Text VQA*, Ling Shengrong, You Sisi, Bao Bing-Kun

- *Pseudo-label based Unsupervised Momentum Representation Learning for Multi-domain Image Retrieval*, Mingyuan Ge, Jianan Shui, Junyu Chen, Mingyong Li

- *Enhancing Generative Generalized Zero Shot Learning via Multi-Space Constraints and Adapative Integration*, Zhe Kong, Neng Gao, Yifei Zhang, Yuhan Liu

- *Foreground Feature Enhancement and Peak & Background Suppression for Fine-Grained Visual Classification*, Die Yu, Zhaoyan Fang, Yong Jiang

- *Point Cloud Classification via Learnable Memory Bank*, Lisa Liu, William Wang, Pingping Cai

- *Adversarially Regularized Low-Light Image Enhancement*, William Wang, Lisa Liu, Pingping Cai

- *Quality Scalable Video Coding based on Neural Representation*, Qian Cao, Dongdong Zhang, Chengyu Sun

- *MVD-NeRF: Resolving Shape-Radiance Ambiguity via Mitigating View Dependency*, Yangjie Cao, Bo Wang, Zhen-qiang Li, Jie Li

- *Learning Complementary Instance Representation with Parallel Adaptive Graph-Based Network for Action Detection*, Yanyan Jiao, Wenzhu Yang, Wenjie Xing

- *Global-to-Local Feature Mining Network for RGB-Infrared Person Re-Identification*, Qiang Chen, Fuxiao He, Guoqiang Xiao

- *Prior-Knowledge-Free Video Frame Interpolation with Bidirectional Regularized Implicit Neural Representations*, Yuanjian He, Weile Zhang, Junyuan Deng, Yulai Cong

- *Dual-Fisheye Image Stitching via Unsupervised Deep Learning*, Zhanjie Jin, Anming Dong, Jiguo Yu, Shuxiang Dong, You Zhou

VBS: Video Browser Showdown

- *Waseda Meisei SoftBank at Video Browser Showdown 2024*, Takayuki Hori, Kazuya Ueki, Yuma Suzuki, Hiroki Takushima, Hayato Tanoue, Haruki Sato, Takumi Takada, Aiswariya Manoj Kumar

- *Exploring Multimedia Vector Spaces with vitrivr-VR*, Florian Spiess, Luca Rossetto, Heiko Schuldt

- *A new Retrieval Engine for vitrivr*, Ralph Gasser, Rahel Arnold, Fynn Firouz Faber, Heiko Schuldt, Raphael Waltenspül, Luca Rossetto

- *VISIONE 5.0: Enhanced User Interface and AI Models for VBS2024*, Giuseppe Amato, Paolo Bolettieri, Fabio Carrara, Fabrizio Falchi, Claudio Gennaro, Nicola Messina, Lucia Vadicamo, Claudio Vairo

- *PraK Tool: An Interactive Search Tool Based on Video Data Services*, Jakub Lokoc, Zuzana Vopálková, Michael Stroh, Raphael Buchmueller, Udo Schlegel

- *Exquisitor at the Video Browser Showdown 2024: Relevance Feedback Meets Conversational Search*, Omar Shahbaz Khan, Hongyi Zhu, Ujjwal Sharma, Evangelos Kanoulas, Stevan Rudinac, Björn Þór Jónsson

- *VERGE in VBS 2024*, Nick Pantelidis, Maria Pegia, Damianos Galanopoulos, Konstantinos Apostolidis, Klearchos Stavrothanasopoulos, Anastasia Moumtzidou, Konstantinos Gkountakos, Ilias Gialampoukidis, Stefanos Vrochidis, Vasileios Mezaris, Ioannis Kompatsiaris, Björn Þór Jónsson

- *Optimizing the Interactive Video Retrieval Tool Vibro for the Video Browser Showdown 2024*, Konstantin Schall, Nico Hezel, Kai Uwe Barthel, Klaus Jung

- *diveXplore at the Video Browser Showdown 2024*, Klaus Schöffmann, Sahar Nasirihaghighi

- *Leveraging LLMs and Generative Models for Interactive Known-Item Video Search*, Zhixin MA, Jiaxin WU, Chong Wah NGO

- *TalkSee: Interactive Video Retrieval Engine Using Large Language Model*, Guihe Gu, Zhengqian Wu, Jiangshan He, Lin Song, Zhongyuan Wang, Chao Liang

- *VideoCLIP 2: An Interactive CLIP-based Video Retrieval System for Novice Users at VBS2024*, Thao-Nhu Nguyen, Le Minh Quang, Cathal Gurrin, Binh T. Nguyen, Graham Healy

- *ViewsInsight: Enhancing Video Retrieval for VBS 2024 with a User-Friendly Interaction Mechanism*, Gia-Huy Vuong, Van-Son Ho, Tien-Thanh Nguyen-Dang, Xuan-Dang Thai, Tu-Khiem Le, Minh-Khoi Pham, Van-Tu Ninh, Cathal Gurrin, Minh-Triet Tran

Oral Session 1: Multimedia Fusion Methods

- *Fusion Boundary and Gradient Enhancement Network for Camouflage Object Detection*, Guangrui Liu, Wei Wu

- *Co-speech Gesture Generation with Variational Auto Encoder*, Shihichi Ka, Koichi Shinoda

- *Exploring Multi-Modal Fusion for Image Manipulation Detection and Localization*, Konstantinos Triaridis, Vasileios Mezaris

Oral Session 2: Audio, Image, Video Processing, Coding and Compression

- *Joint Image Data Hiding and Rate-Distortion Optimization in Neural Compressed Latent Representations*, Chen-Hsiu Huang, Ja-Ling Wu

- *Advancing Incremental Few-shot Semantic Segmentation via Semantic-guided Relation Alignment and Adaptation*, Yuan Zhou, Xin Chen, Yanrong Guo, Jun Yu, Richang Hong, Qi Tian

- *Semantic Importance-Based Deep Image Compression Using A Generative Approach*, Xi Gu, Yuanyuan Xu, Kun Zhu

- *CT-MVSNet: Efficient Multi-View Stereo with Cross-scale Transformer*, Wang Sicheng, Jiang Hao, Xiang Lei

Tuesday, 30 January 2024

Keynote

- *What Multimodal Foundation Models Cannot Perceive*, Cees Snoek

Oral Session 3: Multimodal Analysis for Retrieval Applications

- *MRHF: Multi-stage Retrieval and Hierarchical Fusion for Textbook Question Answering*, Peide Zhu, Zhen Wang, Manabu Okumura, Jie Yang

- *Time-Quality Tradeoff of MuseHash Query Processing Performance*, Maria-Eirini Pegia, Ferran Agulló López, Anastasia Moumtzidou, Alberto Gutierrez-Torre, Björn Þór Jónsson, Josep Lluís Berral-García, Ilias Gialampoukidis, Stefanos Vrochidis, Ioannis Kompatsiaris

- *A Language-based solution to enable Metaverse Retrieval*, Ali Abdari, Alex Falcon, Giuseppe Serra

Poster Session 2

- *Spatiotemporal Representation Enhanced ViT for Video Recognition*, Min Li, Fengfa Li, Bo Meng, Ruwen Bai, Junxing Ren, Zihao Huang, Chenghua Gao

- *ACT: Action-associated and Target-related Representations for Object Navigation*, Youkai Wang, Wansen Wu, Ting Liu, Yong Peng, Yue Hu

- *PANDA: Prompt-based Context- and Indoor-aware Pretraining for Vision and Language Navigation*, Ting Liu, Wansen Wu, Youkai Wang, Yue Hu, Kai Xu, Quanjun Yin

- *Cross-Modal Semantic Alignment Learning for Text-based Person Search*, Wenjun Gan, Jiawei Liu, Yangchun Zhu, Yong Wu, Guozhi Zhao, Zheng-Jun Zha

- *Contextual Augmentation with Bias Adaptive for Few-shot Video Object Segmentation*, ShuaiWei Wang, Zhao Liu, Jie Lei, Zunlei Feng, Juan Xu, Xuan Li, Ronghua Liang

- *MAVAR-SE: Multi-scale Audio-Visual Association Representation Network for End-to-end Speaker Extraction*, Shilong Yu, Chenhui Yang

- *Gait Recognition Based on Temporal Gait Information Enhancing*, Qizhen Chen, Xin Chen, Xiaoling Deng, Yubin Lan

- *MoCap-Video Data Retrieval with Deep Cross-Modal Learning*, Lu Zhang, Jingliang Peng, Na Lv

- *DFGait: Decomposition Fusion Representation Learning for Multimodal Gait Recognition*, Jianbo Xiong, Shinan Zou, Jin Tang

- *TICondition: Expanding Control Capabilities for Text-to-Image Generation with Multi-Modal Conditions*, Yuhang Yang, Xiao Yan, Sanyuan Zhang

- *Image Clustering and Generation with HDGMVAE-I*, Yongqi Liu, Jiashuang Zhou, Xiaoqin Du

- *From Skulls to Faces: A Deep Generative Framework for Realistic 3D Craniofacial Reconstruction*, Yehong Pan, Jian Wang, Guihong Liu, Qiushuo Wu, Yazi Zheng, Xin Lan, Weibo Liang, Jiancheng Lv, Yuan Li

- *CLF-Net: A Few-shot Cross-Language Font Generation Method*, Qianqian Jin, Fazhi He, Wei Tang

- *Multi-dimensional Fusion and Consistency for Semi-supervised Medical Image Segmentation*, Yixing Lu, Zhaoxin Fan, Min Xu

- *Lightweight Image Captioning Model Based on Knowledge Distillation*, Zhenlei Cui, Zhenhua Tang, Jianze Li, Kai Chen

- *MSMV-UNet : A 2.5D Stroke Lesion Segmentation Method based on Multi-slice Feature Fusion*, Jingjing Xie, JiXuan Hong, Manjin Sheng, Chenhui Yang

- *Localization and Local Motion Magnification of Pulsatile Regions in Endoscopic Surgery Videos*, Honglei Zheng, Wenkang Fan, Yinran Chen, Xiongbiao Luo

- *A Custom GAN-based Robust Algorithm for Medical Image Watermarking*, Kun Zhang, Chunling Gao, Shuangyuan Yang

- *A Secure and Fair Federated Learning Protocol under the Universal Composability Framework*, Qiuxian Li, Quanxing Zhou, Hongfa Ding

- *MobileViT-FocR: MobileViT with Fixed-One-Centre Loss and Gradient Reversal for Generalised Fake Face Detection*, Ting Peng, Yihang Zhou, Rong Sun, Yizhi Luo, Yuqi Li

- *Exploring Imperceptible Adversarial Examples in YCbCr Color Space*, Pei Chen, Zhiyong Feng, Meng Xing, Yiming Zhang, Jinqing Zheng

- *Fractional-order image moments and applications*, Liyun Xu, Min Zhang

Best Paper Session

- *Multi-Task Collaborative Network for Image-text Retrieval*, Xueyang Qin, Lishuang Li, Jing Hao, Meiling Ge, Jiayi Huang, Guangyao Pang

- *Object-based Spatio-Temporal Heterogeneous Network for VideoQA*, Feifei Xu, Zheng Zhong, Yitao Zhu, Guangzhen Li, Yingchen Zhou

- *Sustainable Commercial Fishery Control using Multimedia Forensics Data from Non-trusted, Mobile Edge Nodes*, Aril Bernhard Ovesen, Tor-Arne Schmidt Nordmo, Michael Alexander Riegler, Pål Halvorsen, Dag Johansen

- *Adapting Pretrained Large-Scale Vision Models for Face Forgery Detection*, Lantao Wang, Chao Ma

Oral Session 4: Multimedia for e-Health and Medical Applications

- *Where are Biases? Adversarial Debiasing with Spurious Feature Visualization*, Chi-Yu Chen, Pu Ching, Pei-Hsin Huang, Min-Chun Hu

- *NearbyPatchCL: Leveraging Nearby Patches for Self-Supervised Patch-Level Multi-Class Classification in Whole-Slide Images*, Gia-Bao Le, Van-Tien Nguyen, Trung-Nghia Le, Minh-Triet Tran

- *Fourier Transformer for Joint Super-Resolution and Reconstruction of MR Image*, Jiacheng Chen, Fei Wu, Wanliang Wang, Haoxin Shen

ICDAR: Special Session on Intelligent Cross-Data Analysis and Retrieval

- *Towards Cross-modal Point Cloud Retrieval for Indoor Scenes*, FuYang Yu, Zhen Wang, Dongyuan Li, Peide Zhu, Xiaohui Liang, Xiaochuan Wang, Manabu Okumura

- *Correlation visualization under missing values: a comparison between imputation and direct parameter estimation methods*, Nhat-Hao Pham, Khanh-Linh Vo, Mai Anh Vu, Thu Nguyen, Michael Riegler, Pål Halvorsen, Binh T. Nguyen

- *IFI: Interpreting for Improving: a Multimodal Transformer with an Interpretability Technique for Recognition of Risk Events.*, Rupayan Mallick, Jenny Benois-Pineau, Akka Zemmari

- *OOKPIK - A Collection of Out-of-Context Image-Caption Pairs*, Kha-Luan Pham, Minh-Khoi Nguyen-Nhat, Anh-Huy Dinh, Quang-Tri Le, Manh-Thien Nguyen, Anh-Duy Tran, Minh-Triet Tran, Duc-Tien Dang-Nguyen

- *LUMOS-DM: Landscape-based Multimodal Scene Retrieval Enhanced by Diffusion Model*, Viet-Tham Huynh, Trong-Thuan Nguyen, Quang-Thuc Nguyen, Mai-Khiem Tran, Tam V. Nguyen, Minh-Triet Tran

Wednesday, 31 January 2024

Keynote

- *From structure to function: how do neural properties influence information transfer?*, Fleur Zeldenrust

XR-MACCI: Special Session on eXtended Reality and Multimedia - Advancing Content Creation and Interaction

- *Mining Landmark Images for Scene Reconstruction from Weakly Annotated Video Collections*, Helmut Neuschmied, Werner Bailer

- *A framework for 3D modeling of construction sites using aerial imagery and semantic NeRFs*, Panagiotis Vrachnos, Marios Krestenitis, Ilias Koulalis, Konstantinos Ioannidis, Stefanos Vrochidis

- *Multimodal 3D Object Retrieval*, Maria-Eirini Pegia, Björn Þór Jónsson, Anastasia Moumtzidou, Sotiris Diplaris, Ilias Gialampoukidis, Stefanos Vrochidis, Ioannis Kompatsiaris

- *An Integrated System for Spatio-Temporal Summarization of 360-degrees Videos*, Ioannis Kontostathis, Evlampios Apostolidis, Vasileios Mezaris

Poster Session 3

- *A Purified Stacking Ensemble Framework for Cytology Classification*, Linyi Qian, Qian Huang, Yulin Chen, Junzhou Chen

- *C3-PO: A Convolutional Neural Network for COVID Onset Prediction from Cough Sounds*, Xiangyu Chen, Md Ayshik Rahman Khan, Md Rakibul Hasan, Tom Gedeon, Md Zakir Hossain

- *GSUNet: A Brain Tumor Segmentation Method Based On 3D Ghost Shuffle U-Net*, JiXuan Hong, JingJing Xie, ChenHui Yang

- *"Car or Bus?" CLearSeg: CLIP-enhanced Discrimination among Resembling Classes for Few-Shot Semantic Segmentation*, Anqi Zhang, Guangyu Gao, Zhuocheng Lv, Yukun An

- *Face Forgery Detection via Texture and Saliency Enhancement*, Sizheng Guo, Haozhe Yang, Xianming Lin

- *Self-distillation Enhanced Vertical Wavelet Spatial Attention for Person Re-identification*, Yuxuan Zhang, Huibin Tan, Long Lan, Xiao Teng, Jing Ren, Yongjun Zhang

- *HPattack: An Effective Adversarial Attack for Human Parsing*, Xin Dong, Rui Wang, Sanyi Zhang, Lihua Jing

- *Dynamic-Static Graph Convolutional Network for Video-Based Facial Expression Recognition*, Fahong Wang, Zhao Liu, Jie Lei, Zeyu Zou, Wentao Han, Juan Xu, Xuan Li, Zunlei Feng, Ronghua Liang

- *Audio-Visual Segmentation By Leveraging Multi-Scaled Features Learning*, Sze An Peter Tan, Guangyu Gao, Jia Zhao

- *Semantic Transition Detection for Self-Supervised Video Scene Segmentation*, Lu Chen, Jiawei Tan, Pingan Yang, Hongxing Wang

- *FGENet:Fine-Grained Extraction Network for Congested Crowd Counting*, Hao-Yuan Ma, Li Zhang, Xiang-Yi Wei

- *Learning Collaborative Reinforcement Attention for 3D Face Reconstruction and Dense Alignment*, Zhengwei Yang, Yange Wang, Lei Ma, Xiangzheng Li

- *MoPE: Mixture of Pooling Experts Framework for Image-Text Retrieval*, Jiangfeng Li, Bowen Wang, Yongrui Qin, Chenxi Zhang, Gang Yu, Qinpei Zhao

- *SSK-Yolo:Global feature-driven small object detection network for images*, Bei Liu, Jian Zhang, Tianwen Yuan, Peng Huang, Chengwei Feng, Minghe Li

- *MetaVSR: A Novel Approach to Video Super-Resolution for Arbitrary Magnification*, Zixuan Hong, Weipeng Cao, Zhiwu Xu, Zhenru Chen, Xi Tao, Zhong Ming, Chuqing Cao, Liang Zheng

- *TNT-Net: Point Cloud Completion by Transformer in Transformer*, Xiaohai Zhang, Jinming Zhang, Jianliang Li, Ming Chen

- *CESegNet:Context-Enhancement Semantic Segmentation Network Based on Transformer*, Xu Chen, Zhibin Zhang

- *Non-Local Spatial-Wise and Global Channel-Wise Transformer for Efficient Image Super-Resolution*, Xiang Gao, Sining Wu, Fan Wang, Xiaopeng Hu

- *Multi-scale Decomposition Dehazing with Polarimetric Vision*, Tongwei Ma, Lilian Zhang, Bo Sun, Chen Fan

- *SM-GAN: Single-stage and Multi-object Text Guided Image Editing*, Ruichen Li, Lei Wu, Pei Dong, Minggang He

- *SEAS-Net: Segment Exchange Augmentation for Semi-Supervised Brain Tumor Segmentation*, Jing Zhang, Wei Wu

- *Super-Resolution-Assisted Feature Refined Extraction for Small Objects in Remote Sensing Images*, Lihua Du, Wei Wu, Chen Li

- *CA-GAN: Conditional Adaptive Generative Adversarial Network for Text-to-Image Synthesis*, Junpeng Liu, Hengkang Bao

Oral Session 5: Cultural, Educational and Social Multimedia Applications

- *Find the Cliffhanger: Multi-Modal Trailerness in Soap Operas*, Carlo Bretti, Pascal Mettes, Hendrik Vincent Koops, Daan Odijk, Nanne van Noord

- *A Multidimensional Taxonomy Model for Music Tangible User Interfaces*, Adriano Baratè, Luca Andrea Ludovico

- *Using Saliency and Cropping to Improve Video Memorability*, Vaibhav Mudgal, Qingyang Wang, Lorin Sweeney, Alan Smeaton

- *Prototype-Enhanced Hypergraph Learning for Heterogeneous Information Networks*, Shuai Wang, Jiayi Shen, Athanasios Efthymiou, Stevan Rudinac, Monika Kackovic, Nachoem Wijnberg, Marcel Worring

FMM: Special Session on Foundation Models for Multimedia

- *Removing Stray-Light for Wild-Field Fundus Image Fusion based on Large Generative Models*, Jun Wu, Mingxin He, Yang Liu, Jingjie Lin, Zeyu Huang, Dayong Ding

- *Training-free Region Prediction with Stable Diffusion*, Yuma Honbu, Keiji Yanai

- *Mitigating Fine-Grained Hallucination by Fine-Tuning Large Vision-Language Models with Caption Rewrites*, Lei Wang, Jiabang He, Shenshen Li, Ning Liu, Ee-Peng Lim

- *GDTNet: A Synergistic Dilated Transformer and CNN by Gate Attention for Abdominal Multi-organ Segmentation*, Can Zhang, Zhiqiang Wang, Yuan Zhang, Xuanya Li, Kai Hu

- *Fine-Grained Multi-Modal Fundus Image Generation Based on Diffusion Models for Glaucoma Classification*, Xinyue Liu, Gang Yang, Yang Zhou, Yajie Yang, Weichen Huang, Dayong Ding, Jun Wu

- *Adapting Pretrained Large-Scale Vision Models for Face Forgery Detection*, Lantao Wang, Chao Ma

Oral Session 6: Contrastive and Adversarial Learning for Multimedia

- *Drive-CLIP: Cross-modal Contrastive Safety-Critical Driving Scenario Representation Learning and Zero-shot Driving Risk Analysis*, Wenbin Gan, Minh-Son Dao, Koji Zettsu

- *Hierarchical Supervised Contrastive Learning for Multimodal Sentiment Analysis*, Kezhou Chen, Shuo Wang, Yanbin Hao

Demonstrations

- *Implementation of Melody Slot Machines*, Masatoshi Hamanaka

- *E2Evideo: End to End Video and Image Pre-processing and Analysis Tool*, Faiga Alawad, Pål Halvorsen, Michael A. Riegler

- *Augmented Reality Photo Presentation and Content-based Image Retrieval on Mobile Devices with AR-Explorer*, Loris Sauter, Tim Bachmann, Heiko Schuldt, Luca Rossetto

- *Facilitating the Production of Well-tailored Video Summaries for Sharing on Social Media*, Evlampios Apostolidis, Konstantinos Apostolidis, Vasileios Mezaris

- *AI-Based Cropping of Soccer Videos for Different Social Media Representations*, Mehdi Houshmand Sarkhoosh, Sayed Mohammad Majidi Dorcheh, Cise Midoglu, Saeed Shafiee Sabet, Tomas Kupka, Dag Johansen, Michael A. Riegler, Pål Halvorsen

- *Few-shot Object Detection as a Service: Facilitating Training and Deployment for Domain Experts*, Werner Bailer, Mihai Dogariu, Bogdan Ionescu, Hannes Fassold

- *DatAR: Supporting Neuroscience Literature Exploration by Finding Relations between Topics in Augmented Reality*, Boyu Xu, Ghazaleh Tanhaei, Lynda Hardman, Wolfgang Hürst

- *EmoAda:A Multimodal Emotion Interaction and Psychological Adaptation System*, Tengteng Dong, Fangyuan Liu, Xinke Wang, Yishun Jiang, Xiwei Zhang, Xiao Sun

Thursday, 1 February 2024

Keynote

- *Visual and Multimodal Disinformation Detection*, Ioannis (Yiannis) Kompatsiaris, Information Technologies Institute, CERTH

Poster Session 4

- *Adversarially Robust Deepfake Detection via Adversarial Feature Similarity Learning*, Sarwar Khan, Jun-Cheng Chen, Wen-Hung Liao, Chu-song Chen

- *Part-aware Prompt Tuning For Weakly Supervised Referring Expression Grounding*, Chenlin Zhao, Jiabo Ye, Yaguang Song, Ming Yan, Xiaoshan Yang, Changsheng Xu

- *SCFormer: A Vision Transformer with Split Channel in Sitting Posture Recognition*, Kedi Qiu, Shoudong Shi, Tianxiang Zhao, Yongfang Ye

- *Dive into Coarse-to-Fine Strategy in Single Image Deblurring*, Zebin Li, Jianping Luo

- *YOLOv5-SRR: Enhancing YOLOv5 for Effective Underwater Target Detection*, Jinyu Shi, Wenjie Wu

- *PMGCN:Preserving measuring mapping prototype graph calibration network for few-shot learning*, Zhengye Shen, Guangtong Lu, Qian Qiao, Fanzhang Li

- *A lightweight local attention network for image super-resolution*, Feng Chen, Xin Song, Liang Zhu

- *Domain Adaptation for Speaker Verification Based on Self-Supervised Learning with Adversarial Training*, Qiulin Li, Junhao Qiang, Qun Yang

- *Hierarchical Bi-Directional Temporal Context Mining for Improved Video Compression*, Zijian Lin, Jianping Luo

- *High Capacity Reversible Data Hiding in Encrypted Images Based on Pixel Value Preprocessing and Block Classification*, Tao Zhang, Ju Zhang, Yicheng Zou, Yu Zhang

- *Improving Small License Plate Detection with Bidirectional Vehicle-plate Relation*, Songkang Dai, Song-Lu Chen, Qi Liu, Chao Zhu, Yan Liu, Feng Chen, Xu-Cheng Yin

- *Irregular License Plate Recognition via Global Information Integration*, Yuan-Yuan Liu, Qi Liu, Song-Lu Chen, Feng Chen, Xu-Cheng Yin

- *DPM-Det: Diffusion Model Object Detection Based on DPM-Solver++ Guided Sampling*, Jingzhi Zhang, Xudong Li, Linghui Sun, Chengjie Bai

- *Deep self-supervised subspace clustering with triple loss*, Xiaotong Bu, Jiwen Dong, Mengjiao Zhang, Guang Feng, Xizhan Gao, Sijie Niu

- *LigCDnet:Remote Sensing Image Cloud Detection Based on Lightweight Framework*, Baotong Su, Wenguang Zheng

- *LRATNet: Local-Relationship-Aware Transformer Network for Table Structure Recognition*, Guangjie Yang, Dajian Zhong, Yu-jie Xiong, Hongjian Zhan

- *ASF-Conformer: Audio Scoring Conformer with FFC for Speaker Verification in Noisy Environments*, Xiran Zhang, Haiyan Liu, Caixia Liu, Haiyang Zhang, Zhiwei Huo

- *Differentiable Neural Architecture Search Based on Efficient Architecture for Lightweight Image Super-Resolution*, Chunyin Sheng, Xiang Gao, Xiaopeng Hu, Fan Wang

- *RDC-YOLOv5:Improved Safety Helmet Detection in Adverse Weather*, Dexu Yao, Aimin Li, Deqi Liu, Mengfan Cheng

- *A Region Based Non-overlapping Reference Speech Estimation Method for Speaker Extraction*, Yiru Zhang, Zeke Li, Bijing Liu, Haiwei Fan, Yong Yang, Qun Yang

- *Adaptive Token Selection and Fusion Network for Multimodal Sentiment Analysis*, Xiang Li, Ming Lu, Ziming Guo, Xiaoming Zhang

- *MC-TCMNER: A Multi-Modal Fusion Model Combining Contrast Learning Method for Traditional Chinese Medicine NER*, Shan Cao, Qingfeng Wu

- *Multi-Modal Video Topic Segmentation with Dual-Contrastive Domain Adaptation*, Linzi Xing, Quan Tran, Fabian Caba, Franck Dernoncourt, Seunghyun Yoon, Zhaowen Wang, Trung Bui, Giuseppe Carenini

- *Unsupervised Multi-Collaborative Learning Network for 3D Face Reconstruction*, Wenlong Lu, Suping Wu, Xitie Zhang, Shengjia Zhang

- *Self-Supervised Edge Structure Learning for Multi-View Stereo and Parallel Optimization*, Pan Li, Suping Wu, Xitie Zhang, Yuxin Peng, Boyang Zhang, Bin Wang

MDRE: Special Session on Multimedia Datasets for Repeatable Experimentation

- *RESET: Relational Similarity Extension for V3C1 Video Dataset*, Patrik Veselý, Ladislav Peška

- *A New Benchmark and OCR-free Method for Document Image Topic Classification*, Zhen Wang, Peide Zhu, Fuyang Yu, Manabu Okumura

- *The Rach3 Dataset: Towards Data-Driven Analysis of Piano Performance Rehearsal*, Carlos Cancino-Chacón, Ivan Pilkov

- *WikiMuTe: A web-sourced dataset of semantic descriptions for music audio*, Benno Weck, Holger Kirchhoff, Peter Grosche, Xavier Serra

- *PDTW150K: A Dataset for Patent Drawing Retrieval*, Chan-Ming Hsu, Tse-Hung Lin, Yu-Hsien Chen, Chih-Yi Chiu

- *Interactive Question Answering for Multimodal Lifelog Retrieval*, Ly-Duyen Tran, Liting Zhou, Binh Nguyen, Cathal Gurrin

- *Event Recognition in Laparoscopic Gynecology Videos with Hybrid Transformers*, Sahar Nasirihaghighi, Negin Ghamsarian, Heinrich Husslein, Klaus Schoeffmann

- *GreenScreen: A Multimodal Dataset for Detecting Corporate Greenwashing in the Wild*, Ujjwal Sharma, Stevan Rudinac, Joris Demmers, Willemijn van Dolen, Marcel Worring

Panel: Brave New Ideas – Are There Any Left?

- *Mutant Texts: A Technique for Uncovering Unexpected Inconsistencies in Large-Scale Vision-language Models*, Mingliang Liang, Zhuoran Liu, Martha Larson

- *Exploring Artificial Intelligence for Advancing Performance Processes and Events in Io3MT*, Romulo Vieira, Debora Muchaluat-Saade, Pablo Cesar