2024 IEEE 7th International Conference on Multimedia Information Processing and Retrieval (MIPR 2024)

San Jose, California, USA
7-9 August 2024

IEEE Catalog Number: CFP24K85-POD
ISBN: 979-8-3503-5143-9

Copyright © 2024 by the Institute of Electrical and Electronics Engineers, Inc.
All Rights Reserved

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved.

****** This is a print representation of what appears in the IEEE Digital Library. Some format issues inherent in the e-media version may also appear in this print version.***

IEEE Catalog Number:	CFP24K85-POD
ISBN (Print-On-Demand):	979-8-3503-5143-9
ISBN (Online):	979-8-3503-5142-2
ISSN:	2770-4327

Additional Copies of This Publication Are Available From:

Curran Associates, Inc
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: (845) 758-0400
Fax: (845) 758-2633
E-mail: curran@proceedings.com
Web: www.proceedings.com

2024 IEEE 7th International Conference on Multimedia Information Processing and Retrieval (MIPR)

MIPR 2024

Table of Contents

Message from the Organizing Committee ..... **xix**
Organizing Committee ..... **xxii**
Program Committee ..... **xxiv**

Session 1: Visual Detection and Tracking

GHOI: A Green Human-Object-Interaction Detector 1
Tsung-Shan Yang (University of Southern California, USA), Yun-Cheng Wang (University of Southern California, USA), Chengwei Wei (University of Southern California, USA), and C.-C. Jay Kuo (University of Southern California, USA)

Real-Time Lane-Wise Traffic Monitoring in Optimal ROIs 8
Mei Qiu (Purdue University, USA), Wei Lin (Purdue University, USA), Lauren Christopher (Purdue University, USA), Stanley Chien (Purdue University, USA), Yaobin Chen (Purdue University, USA), and Shu Hu (Purdue University, USA)

Prohibited Object Detection and Liquid Component Classification Based on Different Types of X-ray Images 15
Hongping Zhi (Tsinghua University, China), Lifeng Sun (Tsinghua University, China), and Xu Wang (Iflytek Research Institute, Iflytek (Suzhou) Technology Co., Ltd., China)

SkyDataNet: an Object Detection Algorithm with 2D Gaussian Loss for UAV-Based Aerial Images 21
Mehmet Akif Ozkanoglu (Bilkent University, Turkiye), Ali C. Begen (Ozyegin University, Turkiye), and Sedat Ozer (California State Polytechnic University, USA)

Target-Aware Siamese Networks Based on Masked Attention Mechanism for Visual Object Tracking 28
Yao-Hui Su (National Cheng Kung University, Taiwan), Ming-Der Shieh (National Cheng Kung University, Taiwan), and Chia-Chi Tsai (National Cheng Kung University, Taiwan)

PM-VIS+: High-Performance Video Instance Segmentation without Video Annotation 35
Zhangjing Yang (Nanjing Audit University, China), Dun Liu (Nanjing Audit University, China), Xin Wang (University at Albany, State University of New York (SUNY), USA), Zhe Li (Lucid Motors, USA), Barathwaj Anandan (Lucid Motors, USA), and Yi Wu (Lucid Motors, USA)

Session 2: Generative and Foundation Models in Multimedia

A Framework for Generating Images and Hashtags for Social Media Posts for Artificial Influencers 42
Raju Shrestha (Oslo Metropolitan University, Norway) and Hanne Korneliussen (Oslo Metropolitan University, Norway)

Automatic Visual Citation Generation for Text-to-Image Generation 49
Ning Xu (Adeia Inc., USA) and Serhad Doken (Adeia Inc., USA)

Enhancing Local LLM Performance through Heterogeneous Multi-Device Computing 55
Ryan Metcalfe (Intel Corporation), Garth Long (Intel Corporation), Charlie L. Wang (Intel Corporation), and Iole Moccagatta (Intel Corporation)

Text-Driven Synchronized Diffusion Video and Audio Talking Head Generation 61
Zhenfei Zhang (University at Albany, State University of New York), Tsung-Wei Huang (Dolby Laboratories), Guan-Ming Su (Dolby Laboratories), Ming-Ching Chang (University at Albany, State University of New York), and Xin Li (University at Albany, State University of New York)

10x Future of Filmmaking Empowered by AIGC 68
Haohong Wang (TCL Research America, USA), Daniel Smith (TCL Research America, USA), and Malgorzata Kudelska (TCL Research America, USA)

Segformer++: Efficient Token-Merging Strategies for High-Resolution Semantic Segmentation 75
Daniel Kienzle (University of Augsburg, Germany), Marco Kantonis (University of Augsburg, Germany), Robin Schön (University of Augsburg, Germany), and Rainer Lienhart (University of Augsburg, Germany)

Session 3: Large Vision Language Models

Green Image Label Transfer 82
Haiyi Li (University of Southern California, USA), Xuejing Lei (University of Southern California, USA), Xinyu Wang (University of Southern California, USA), and C.-C. Jay Kuo (University of Southern California, USA)

GenCheck: A LoRA-Adapted Multimodal Large Language Model for Check Analysis 88
Fei Zhao (The University of Alabama at Birmingham, USA), Jiawen Chen (Beijing University of Technology, China), Bin Huang (Beijing Hua Yu Xin Zhang Technologies Co., Ltd, China), Chengcui Zhang (The University of Alabama at Birmingham, USA), Gary Warner (The University of Alabama at Birmingham, USA), Rushi Chen (Shihezi University, China), Shaorou Tang (Beijing University of Technology, China), Yuanfei Ma (Beijing University of Technology, China), and Zixi Nan (University of Chinese Academy of Sciences, China)

Leveraging Semantic Segmentation for Image Manipulation Detection and Localization 95
Yuwei Chen (University at Albany (SUNY), USA), Ming-Ching Chang (University at Albany (SUNY), USA), and Xin Li (University at Albany (SUNY), USA)

GeoVQA: A Comprehensive Multimodal Geometry Dataset for Secondary Education 102
Avinash Anand (IIIT Delhi, India), Raj Jaiswal (IIIT Delhi, India), Abhishek Dharmadhikari (IIIT Delhi, MIDAS Lab, India), Atharva Marathe (IIIT Delhi, MIDAS Lab, India), Harsh Popat (IIIT Delhi, India), Harshil Mital (IIIT Delhi, India), Ashwin R Nair (IIIT Delhi, India), Kritarth Prasad (IIIT Delhi, India), Sidharth Kumar (IIT Roorkee, India), Astha Verma (NUS, Singapore), Rajiv Ratn Shah (IIIT Delhi, India), and Roger Zimmermann (NUS, Singapore)

Session 4: Vision and Content Understanding

ProxeGraph: Scene Graph Generation Utilizing Proxemics for Smart Homes 109
Debaleen Das Spandan (Missouri State University, USA) and Razib Iqbal (Missouri State University, USA)

HOI as Embeddings: Advancements of Model Representation Capability in Human-Object Interaction Detection 116
Junwen Chen (The University of Electro-Communications, Japan), Yingcheng Wang (The University of Electro-Communications, Japan), and Keiji Yanai (The University of Electro-Communications, Japan)

Lightweight Schemes Fusion for Heatmap-Based Human Pose Estimation 123
Sheng-Jhou Lu (Yuan Ze University, Taiwan), Hung-Wei Lee (Yuan Ze University, Taiwan), Yu-Ming Han (Yuan Ze University, Taiwan), Ji-Min Zhou (Yuan Ze University, Taiwan), Ying Liu (Santa Clara University, USA), and Huang-Chia Shih (Yuan Ze University, Taiwan)

Anomaly Detection in Video using Compression 127
Micheal Smith (Sandia National Laboratories, USA), Renee Gooding (Sandia National Laboratories, USA), Jon Bisila (Sandia National Laboratories, USA), and Christina Ting (Sandia National Laboratories, USA)

SSLCT: A Convolutional Transformer for Synthetic Speech Localization 134
Kratika Bhagtani (Purdue University, USA), Amit Kumar Singh Yadav (Purdue University, USA), Paolo Bestagini (Politecnico di Milano, Italy), and Edward J. Delp (Purdue University, USA)

Playlist Continuation of Cold-Start Songs 141
Chun-Han Cheng (National Taiwan University), Ting-Yu Wei (National Taiwan University), and Homer H. Chen (National Taiwan University)

Session 5: Multimodal Media Retrieval and Perceptual Multimedia

Improved Standard-Based Motion Parallax Measurement in Mixed Reality 148
Hung-Jui Guo (The University of Texas at Dallas, USA) and Balakrishnan Prabhakaran (The University of Texas at Dallas, USA)

Blended RAG: Improving RAG (Retriever-Augmented Generation) Accuracy with Semantic Search and Hybrid Query-Based Retrievers 155
Kunal Sawarkar (IBM), Abhilasha Mangal (IBM), and Shivam Raj Solanki (IBM)

Automated Thematic Composer Classification Using Segment Retrieval 162
Jacob Galajda (University of Central Florida, USA) and Kien Hua (University of Central Florida, USA)

DRM-SN: Detecting Reused Multimedia Content on Social Networks 169
Wen-Shiang Li (National Chung Cheng University, Taiwan), Yao-Cheng Lu (National Chung Cheng University, Taiwan), Wen-Kai Hsiao (National Chung Cheng University, Taiwan), Yu-Yao Tseng (National Chung Cheng University, Taiwan), and Ming-Hung Wang (National Chung Cheng University, Taiwan)

Uncertainty-Guided Appearance-Motion Association Network for Out-of-Distribution Action Detection 176
Xiang Fang (Nanyang Technological University, Singapore), Arvind Easwaran (Nanyang Technological University, Singapore), and Blaise Genest (CNRS@CREATE, Singapore)

FastLearn: A Rapid Learning Agent for Chat Models to Acquire Latest Knowledge 183
Chenhan Fu (Zhejiang University, China), Guoming Wang (Zhejiang University, China), Rongxing Lu (University of New Brunswick, Canada), and Siliang Tang (Zhejiang University, China)

Session 6: Multimodal Integration and Analysis

Enhancement of Neonatal Lung Pathology Classification Using Multi-view Feature Representation 190
Ryan Tan (Toronto Metropolitan University), Thanh Hong-Phuoc (Toronto Metropolitan University), Lei Gao (Toronto Metropolitan University), Randy Tan (Toronto Metropolitan University), Sagarjit Aujla (Toronto Metropolitan University), Adel Mohamed (Mount Sinai Hospital, Canada), Ling Guan (Toronto Metropolitan University), Karthikeyan Umapathy (Toronto Metropolitan University), and Naimul Khan (Toronto Metropolitan University)

Holistic Visual-Textual Sentiment Analysis with Prior Models 196
Junyu Chen (University of Rochester, USA), Jie An (University of Rochester, USA), Hanjia Lyu (University of Rochester, USA), Christopher Kanan (University of Rochester, USA), and Jiebo Luo (University of Rochester, USA)

Radio Map Estimation (RME) with Deep Progressive Network 203
Jashia Mitayeegiri (University of North Texas, USA), Shaohua Dong (University of North Texas, USA), Chenxi Qiu (University of North Texas, USA), Qing Yang (University of North Texas, USA), Xinrong Li (University of North Texas, USA), Heng Fan (University of North Texas, USA), and Yan Huang (University of North Texas, USA)

CM-ASAP: Cross-Modality Adaptive Sensing and Perception for Efficient Hand Gesture Recognition 207
Soheil Hor (Electrical Engineering Department, Stanford University), Mostafa El-Khamy (Device Solutions Research America, Samsung Semiconductor Inc.), Yanlin Zhou (Device Solutions Research America, Samsung Semiconductor, Inc.), Amin Arbabian (Electrical Engineering Department, Stanford University), and SukHwan Lim (Device Solutions Research America, Samsung Semiconductor, Inc.)

Mitigating Privacy Threats Without Degrading Visual Quality of VR Applications: Using Re-Identification Attack as a Case Study 214
Yu-Szu Wei (National Tsing Hua University, Taiwan), Yuan-Chun Sun (National Tsing Hua University, Taiwan), Shin-Yi Zheng (National Tsing Hua University, Taiwan), Hsun-Fu Hsu (National Yang Ming Chiao Tung University, Taiwan), Chun-Ying Huang (National Yang Ming Chiao Tung University, Taiwan), and Cheng-Hsin Hsu (National Tsing Hua University, Taiwan)

Device-Agnostic Remote Range-of-Motion Assessment using Data Abstraction 221
Omeed Ashtiani (University of Texas at Dallas, USA), Meghana Spurthi Maadugundu (University of Texas at Dallas, USA), Minhas Kamal (University of Texas at Dallas, USA), and Balakrishnan Prabhakaran (University of Texas at Dallas, USA)

Session 7: Short Paper Session (1)

Retrieval Augmented Structured Generation: Business Document Information Extraction As Tool Use 227
Franz Louis Cesista (Expedock Software Inc., USA), Rui Aguiar (Expedock Software Inc., USA), Jason Kim (Expedock Software Inc., USA), and Joshua Paolo Acilo (Expedock Software Inc., USA)

Will Neural 3D Object Representations be the Silver Bullet for Improving VR Experience in HMDs? 231
Charlie Hsu (National Experimental High School at Hsin-Chu Science Park, Taiwan), Yuan-Chun Sun (National Tsing Hua University, Taiwan), Kuan-Yu Lee (National Tsing Hua University, Taiwan), and Chun-Ying Huang (National Yang Ming Chiao Tung University, Taiwan)

Frame-Level Latent Embedding using Weak Labels for Multi-View Action Recognition 235
Vijay John (RIKEN, Japan) and Yasutomo Kawanishi (RIKEN, Japan)

A Deep Features Based Approach Using Modified ResNet50 and Gradient Boosting for Visual Sentiments Classification 239
Muhammad Arslan (Loyola University Chicago, USA), Muhammad Mubeen (University of the People, USA), Arslan Akram (The Superior University Lahore, Pakistan), Saadullah Farooq Abbasi (University of Birmingham, U.K.), Muhammad Salman Ali (The Superior University Lahore, Pakistan), and Muhammad Usman Tariq (The Superior University Lahore, Pakistan)

Dental X-ray Segmentation and Auto Implant Design Based on Convolutional Neural Network 243
Yang Xing (Boston University, USA), Peixi Liao (Boston University, USA), Reem AwdhE Alasleh (Boston University, USA), Vissuta Khampatee (Boston University, USA), and Farshid Alizadeh-Shabdiz (Boston University, USA)

Session 8: Machine/Deep Learning/Data Mining

Joint HDR Denoising and Fusion on Mobile Devices 247
Jie Cai (OPPO AI Center, USA), Yuan Lin (OPPO AI Center, USA), Jiang Li (OPPO AI Center, USA), Jiaming Ding (OPPO AI Center, USA), Ling Ouyang (OPPO AI Center, USA), Chiu Man Ho (OPPO AI Center, USA), and Zibo Meng (OPPO AI Center, USA)

MU-MAE: Multimodal Masked Autoencoders-Based One-Shot Learning 253
Rex Liu (University of California, USA) and Xin Liu (University of California, USA)

Structured Pruning for Multi-Task Deep Neural Networks 260
Siddhant Garg (Adobe Inc.), Lijun Zhang (University of Massachusetts Amherst), and Hui Guan (University of Massachusetts Amherst)

UU-Mamba: Uncertainty-Aware U-Mamba for Cardiac Image Segmentation 267
Ting Yu Tsai (University at Albany, State University of New York), Li Lin (Purdue University), Shu Hu (Purdue University), Ming-Ching Chang (University at Albany, State University of New York), Hongtu Zhu (University of North Carolina at Chapel Hill), and Xin Wang (University at Albany, State University of New York)

Unveiling Statistical Significance of Online Regression over Multiple Datasets 274
Mohammad Abu-Shaira (The University of North Texas, USA) and Weishi Shi (The University of North Texas, USA)

Macro-AUC-Driven Active Learning Strategy for Multi-Label Classification Enhancement 280
Minghao Li (University of North Texas, USA), Junjie Qiu (University of North Texas, USA), and Weishi Shi (University of North Texas, USA)

Session 9: User Experience and Engagement

Viewing Comfort Enhancement on Head-Mounted Displays Using Stereo Disparity Control 287
Dae Yeol Lee (Dolby Laboratories, Inc., USA), Geonsun Lee (University of Maryland, USA), and Guan-Ming Su (Dolby Laboratories, Inc., USA)

Multi-Task Decision-Making for Multi-User 360° Video Processing over Wireless Networks 294
Babak Badnava (University of Kansas), Jacob Chakareski (New Jersey Institute of Technology), and Morteza Hashemi (University of Kansas)

ExCEDA: Unlocking Attention Paradigms in Extended Duration E-Classrooms by Leveraging Attention-Mechanism Models 301
Avinash Anand (IIIT New Delhi, India), Avni Mittal (Indian Institute of Technology Mandi, India), Laavanaya Dhawan (Netaji Subhas University of Technology New Delhi, India), Juhi Krishnamurthy (Adobe, India), Mahisha Ramesh (IIIT Delhi, India), Naman Lal (IIIT Delhi, India), Astha Verma (IIIT New Delhi, India), Pijush Bhuyan (IIIT New Delhi, India), Himani Himani (IIIT New Delhi, India), Rajiv Shah (IIIT New Delhi, India), Roger Zimmermann (NUS, Singapore), and Shin'ichi Satoh (National Institute of Informatics, Japan)

Pulse of the Crowd: Quantifying Crowd Energy through Audio and Video Analysis 308
Avinash Anand (MIDAS Lab, IIIT Delhi, India), Sarthak Jain (MIDAS Lab, IIIT Delhi, India), Shashank Sharma (MIDAS Lab, IIIT Delhi, India), Akhil P. Dominic (MIDAS Lab, IIIT Delhi, India), Aman Gupta (MIDAS Lab, IIIT Delhi, India), Ashta Verma (MIDAS Lab, IIIT Delhi, India), Raj Jaiswal (MIDAS Lab, IIIT Delhi, India), Naman Lal (MIDAS Lab, IIIT Delhi, India), Rajiv Ratn Shah (MIDAS Lab, IIIT Delhi, India), and Roger Zimmermann (NUS, Singapore)

Plastic Surgery Image Classification and Generation 315
Yiwei Han (University of Rochester, USA), Kaiyi Qi (University of Rochester, USA), and Jiebo Luo (University of Rochester, USA)

SelfEngage: Self-Attentive Engagement Prediction with Neural Cellular Automata 321
Avinash Anand (IIIT Delhi), Avni Mittal (IIT Mandi), Pijush Bhuyan (IIIT Delhi), Raj Jaiswal (IIIT Delhi), Laavanaya Dhawan (NSUT Delhi), Sidharth Kumar (IIT Roorkee), Naman Lal (Midas Labs), Astha Verma (IIIT Delhi), Himani Himani (IIIT Delhi), Rajiv Ratn Shah (IIIT Delhi), Roger Zimmermann (NUS Singapore), and Shin'ichi Satoh (NII Tokyo)

Session 10: Multimedia Features and Datasets

CU-Mamba: Selective State Space Models with Channel Learning for Image Restoration 328
Rui Deng (Stanford University, USA) and Tianpei Gu (Krea AI, USA)

OmniDet: Omnidirectional Object Detection via Fisheye Camera Adaptation 335
Chih-Chung Hsu (Institute of Data Science, National Cheng Kung University, Taiwan), Wei-Hao Huang (Institute of Data Science, National Cheng Kung University, Taiwan), Wen-Hai Tseng (Institute of Data Science, National Cheng Kung University, Taiwan), Ming-Hsuan Wu (Institute of Data Science, National Cheng Kung University, Taiwan), Ren-Jung Xu (Institute of Data Science, National Cheng Kung University, Taiwan), and Chia-Ming Lee (Institute of Data Science, National Cheng Kung University, Taiwan)

GESA: Exploring Loss-Based Adversarial Attacks in Volumetric Media Streaming 342
Most Husne Jahan (Concordia University) and Abdelhak Bentaleb (Concordia University)

LivePics-24: A Multi-Person, Multi-Camera, Multi-Settings Live Photos Dataset 349
Omkar Kulkarni (University at Albany, USA), Aryan Mishra (University of California, USA), Shashank Arora (University at Albany, USA), Vivek K. Singh (Rutgers University, USA), and Pradeep K. Atrey (University at Albany, USA)

Zero-Shot Retrieval of User Intent in Human-Robot Interaction with Large Language Models 355
Jesse Atuhurra (Nara Institute of Science and Technology, Japan)

TextSleuth: A New Dataset and Baseline for Scene Text Manipulation Detection 362
Abhineet Kumar Pandey (University At Albany, State University of New York, USA), Ming-Ching Chang (University At Albany, State University of New York, USA), and Xin Li (University At Albany, State University of New York, USA)

Session 11: Deep Learning in Multimedia Data and Fusion

MambaTab: A Plug-and-Play Model for Learning Tabular Data 369
Md Atik Ahamed (University of Kentucky, USA) and Qiang Cheng (University of Kentucky, USA)

Deep Learning-Based Text-in-Image Watermarking 376
Bishwa Karki (University of Nebraska Omaha, USA), Chun-Hua Tsai (University of Nebraska Omaha, USA), Pei-Chi Huang (University of Nebraska Omaha, USA), and Xin Zhong (University of Nebraska Omaha, USA)

Single-Frame Supervised Action Temporal Localization Based on Multi-View Contrastive Learning 383
Haoran Tong (University of Chinese Academy of Sciences, China), Xu Cui (University of Chinese Academy of Sciences, China), and Laiyun Qing (University of Chinese Academy of Sciences, China)

Mutual Information Analysis in Multimodal Learning Systems 390
Hadi Hadizadeh (Simon Fraser University, Canada), Seyedeh Faegheh Yeganli (Simon Fraser University, Canada), Bahador Rashidi (Huawei Canada, Canada), and Ivan Bajic (Simon Fraser University, Canada)

Mathematics-Inspired Learning: A Green Learning Model with Interpretable Properties 396
Ling Guan (Toronto Metropolitan University, Canada), Lei Gao (Toronto Metropolitan University, Canada), Kai Liu (Toronto Metropolitan University, Canada), and Zheng Guo (Toronto Metropolitan University, Canada)

Learning to Switch off, Switch on, and Integrate Modalities in Large Pre-Trained Transformers 403
Haoran Tong (University of Chinese Academy of Sciences, China), Xu Cui (University of Chinese Academy of Sciences, China), and Laiyun Qing (University of Chinese Academy of Sciences, China)

Session 12: Novel Applications

Cultural Relevance Index: Measuring Cultural Relevance in AI-Generated Images 410
Wala ELsharif (Hamad Bin Khalifa University, Qatar), Marco Agus (Hamad Bin Khalifa University, Qatar), Mahmood Alzubaidi (Hamad Bin Khalifa University, Qatar), and James She (The Hong Kong University of Science and Technology, Hong Kong)

Parameter-Efficient Adaptation of Foundation Models for Damaged Building Assessment 417
Fei Zhao (The University of Alabama at Birmingham, USA) and Chengcui Zhang (The University of Alabama at Birmingham, USA)

Exploring the Impact of Hand Pose and Shadow on Hand-Washing Action Recognition 423
Shengtai Ju (Purdue University, USA) and Amy Reibman (Purdue University, USA)

Enabling Paper-Based Surface Authentication via Digital Twin and Experimental Verification..... 430
Prasun Datta (NC State University), Chau-Wai Wong (NC State University), and Min Wu (University of Maryland, College Park)

DeepFake-o-Meter v2.0: An Open Platform for DeepFake Detection 439
Yan Ju (State University of New York at Buffalo, USA), Chengzhe Sun (State University of New York at Buffalo, USA), Shan Jia (State University of New York at Buffalo, USA), Shuwei Hou (State University of New York at Buffalo, USA), Zhaofeng Si (State University of New York at Buffalo, USA), Soumyya Kanti Datta (State University of New York at Buffalo, USA), Lipeng Ke (Amazon Lab126, USA), Riky Zhou (State University of New York at Buffalo, USA), Anita Nikolich (University of Illinois Urbana-Champaign, USA), and Siwei Lyu (State University of New York at Buffalo, USA)

GeoSecure-B: A Method for Secure Bearing Calculation 446
Vikram Patil (University at Albany, USA; GoDaddy LLC, USA), Sharmilee Rajkumar Rajan (GoDaddy LLC, USA), and Pradeep K. Atrey (GoDaddy LLC, USA)

Session 13: Multimedia Data Modeling, Indexing, Mining

Clearing Text Images: A Non-Blind Deblurring with Convex Total Variation Regularization Model 452
Narendra Kumar (Indian Institute of Technology Jodhpur, India) and Gaurav Bhatnagar (Indian Institute of Technology Jodhpur, India)

Algorithmic Stock Trading Strategies 458
Craig Rainey (University of Washington Bothell, USA) and Min Chen (University of Washington Bothell, USA)

Benchmarking the Robustness of UAV Tracking Against Common Corruptions 465
Xiaoqiong Liu (University of North Texas, USA), Yunhe Feng (University of North Texas, USA), Shu Hu (Purdue University, USA), Xiaohui Yuan (University of North Texas, USA), and Heng Fan (University of North Texas, USA)

LitAI: Enhancing Multimodal Literature Understanding and Mining with Generative AI 471
Gowtham Medisetti (University of North Texas, USA), Zacchaeus Compson (University of North Texas, USA), Heng Fan (University of North Texas, USA), Huaxiao Yang (University of North Texas, USA), and Yunhe Feng (University of North Texas, USA)

GaugeTracker: AI-Powered Cost-Effective Analog Gauge Monitoring System 477
Beitong Tian (University of Illinois Urbana-Champaign, USA), Mingyuan Wu (University of Illinois Urbana-Champaign, USA), Ruixiao Zhang (University of Illinois Urbana-Champaign, USA), Haozhen Zheng (University of Illinois Urbana-Champaign, USA), Bo Chen (University of Illinois Urbana-Champaign, USA), Yaohui Wang (University of Illinois Urbana-Champaign, USA), Shiv Trivedi (University of Illinois Urbana-Champaign, USA), Shanbo Zhang (University of Illinois Urbana-Champaign, USA), Robert Bruce Kaufman (University of Illinois Urbana-Champaign, USA), Leah Espenhahn (University of Illinois Urbana-Champaign, USA), Gianni Pezzarossi (University of Illinois Urbana-Champaign, USA), Mauro Sardela (University of Illinois Urbana-Champaign, USA), John Dallesasse (University of Illinois Urbana-Champaign, USA), and Klara Nahrstedt (University of Illinois Urbana-Champaign, USA)

Balancing Explanations and Adaptation in Offline Continual Learning Systems using Active Augmented Reply 484
Md Abdullah Al Forhad (University of North Texas, USA) and Weishi Shi (University of North Texas, USA)

Session 14: Short Paper Session (2)

Controllable Universal Edge-Preserving Image Filtering 491
Shijun Liang (Michigan State University) and Dongdong Fu (Dolby Laboratories, Inc.)

Attenuation-Aware Weighted Optical Flow with Medium Transmission Map for Learning-Based Visual Odometry in Underwater Terrain 495
Bach Nguyen Gia (Shibaura Institute of Technology, Japan), Chanh Tran Minh (Shibaura Institute of Technology, Japan), Kamioka Eiji (Shibaura Institute of Technology, Japan), and Tan Phan Xuan (Shibaura Institute of Technology, Japan)

Harmful Brain Activity Classification of Spectrograms with Transfer Deep Learning 499
Shanker Ram (Lynbrook High School, USA), Sambhu Ganesan (Lynbrook High School, USA), and Yajat Nagaraj Kiran (Lynbrook High School, USA)

TQCompressor: Improving Tensor Decomposition Methods in Neural Networks via Permutations 503
Vadim Abronin (Terra Quantum AG, Switzerland), Aleksei Naumov (Terra Quantum AG, Switzerland), Denis Mazur (Terra Quantum AG, Switzerland), Dmitriy Bystrov (Terra Quantum AG, Switzerland), Katerina Tsarova (Terra Quantum AG, Switzerland), Artem Melnikov (Terra Quantum AG, Switzerland), Sergey Dolgov (University of Bath, United Kingdom), Reuben Brasher (Terra Quantum AG, Switzerland), and Michael Perelshtein (Terra Quantum AG, Switzerland)

BubbleSig: Same-Hand Ballot Stuffing Detection 507
Fei Zhao (The University of Alabama at Birmingham, USA), Chengcui Zhang (The University of Alabama at Birmingham, USA), Maya Shah (Texas A&M University, USA), and Nitesh Saxena (Texas A&M University, USA)

Towards a Novel Blob Detection Approach for Concealed Object Detection in Passive Terahertz Imaging 511
Sushmita Chandel (Indian Institute of Technology Jodhpur, India), Preeti Dwivedi (Indian Institute of Technology Jodhpur, India), Gaurav Bhatnagar (Indian Institute of Technology Jodhpur, India), and Marcin Kowalski (Military University of Technology, Poland)

Session 15: Trustworthy Multimedia & 3D Modeling

A VR 360°-Video Encoding Framework with Differentiated Tile Compression Based on Digital-Twin Technology 515
Andrea Caruso (University of Catania - CNIT, Italy) and Giovanni Schembra (University of Catania - CNIT, Italy)

Trustworthy and Robust Machine Learning for Multimedia: Challenges and Perspectives 522
Katsuaki Nakano (Rochester Institute of Technology, USA), Michael Zuzak (Rochester Institute of Technology, USA), Cory Merkel (Rochester Institute of Technology, USA), and Alexander Loui (Rochester Institute of Technology, USA)

Counterfactual Gradients-Based Quantification of Prediction Trust in Neural Networks 529
Mohit Prabhushankar (Georgia Institute of Technology, USA) and Ghassan AlRegib (Georgia Institute of Technology, USA)

A Framework for Single-View Multi-Plane Image Inpainting 536
Zachary McBride Lazri (Dolby Laboratories, Inc., USA; University of Maryland, USA), Dae Yeol Lee (Dolby Laboratories, Inc., USA), and Guan-Ming Su (Dolby Laboratories, Inc., USA)

GPSR: A Green Point Cloud Surface Reconstruction Method 542
Qingyang Zhou (University of Southern California), Jiawei Yu (University of Southern California), Shan Liu (Tencent Media Lab, USA), and C.-C. Jay Kuo (University of Southern California)

Behavioral Emotion Analysis Model for Large Language Models 549
Edward Y. Chang (Stanford University)

Session 16: Multimedia Applications for Health and Sports

MISS: Memory-Efficient Instance Segmentation for Sport-Scenes with Visual Inductive Priors..... 557
Chih-Chung Hsu (National Cheng Kung University, Taiwan) and Chia-Ming Lee (National Cheng Kung University, Taiwan)

Simultaneous Classification and Segmentation of Subretinal Lesions on ICGA Images 562
Ming-Wen Kuan (National Chung Cheng University, Taiwan), Wei-Yang Lin (National Chung Cheng University, Taiwan), Chia-Ling Tsai (City University of New York, USA), Shih-Jen Chen (Taipei Veterans General Hospital, Taiwan), Paisan Ruamviboonsuk (Rajavithi Hospital, Thailand), and Dong-Jie Jiang (National Chung Cheng University, Taiwan)

Automatic Clipping and Text Logging for Baseball Game Videos using Deep Learning 566
Chen-Wei Wang (Feng Chia University, Taiwan) and Hwai-Jung Hsu (Feng Chia University, Taiwan)

Advancing Retinal Image Segmentation: A Denoising Diffusion Probabilistic Model Perspective 572
Alnur Alimanov (Bahcesehir University, Turkey) and Md Baharul Islam (Florida Gulf Coast University, USA; Bahcesehir University, Turkey)

Automated Recognition of Optic Disc and Blood Vessels in Diabetic Fundoscopy Images Using Real-Time Image Analysis 579
Kaixuan Li (Virginia State University, USA), Wei-Bang Chen (Virginia State University, USA), Yongjin Lu (Oakland University, USA), Xiaoliang Wang (Department of Computer Science, USA), and He Gao (Virginia State University, USA)

Robust Covid-19 Detection in CT Images with CLIP 586
Li Lin (Purdue University), Yamini Sri Krubha (Purdue University), Zhenhuan Yang (Etsy, USA), Cheng Ren (University at Albany, State University of New York), Thuc Duy Le (University of South Australia), Irene Amerini (Sapienza University of Rome), Xin Wang (University at Albany, State University of New York), and Shu Hu (Purdue University)

Session 17: Remote Presentation Session

Enhancing Video Stability with Object-Centric Stabilization 593
Aparna Tiwari (Center for Advance Studies, A.K.T.U. Lucknow, India), Hitika Tiwari (Indian Institute of Technology Kanpur, India), K.S. Venkatesh (Indian Institute of Technology Kanpur, India), and Anuj Kumar Sharma (Center for Advance Studies, A.K.T.U., India)

CIA: Controllable Image Augmentation Framework Based on Stable Diffusion 600
Mohamed Benkedadra (University of Mons (UMONS), Belgium), Dany Rimez (Université Catholique de Louvain (UCLouvain), Belgium), Tiffanie Godelaine (Université Catholique de Louvain (UCLouvain), Belgium), Natarajan Chidambaram (University of Mons (UMONS), Belgium), Hamed Razavi Khosroshahi (Université Libre de Bruxelles (ULB), Belgium), Horacio Tellez (Multitel, Belgium), Matei Mancas (University of Mons (UMONS), Belgium), Benoit Macq (Université Catholique de Louvain (UCLouvain), Belgium), and Sidi Ahmed Mahmoudi (University of Mons (UMONS), Belgium)

Robust Light-Weight Facial Affective Behavior Recognition with CLIP 607
Li Lin (Purdue University), Sarah Papabathini (Purdue University), Xin Wang (University at Albany, State University of New York), and Shu Hu (Purdue University)

Guarding Against ChatGPT Threats: Identifying and Addressing Vulnerabilities 612
Dingzong Zhang (The University of Queensland, Australia), Khushi Jain (The University of Queensland, Australia), and Priyanka Singh (The University of Queensland, Australia)

Advection-Diffusion for Feature-Based Cancer Diagnosis 616
Fayadh Alenezi (Jouf University, Saudi Arabia)

Exploiting Correlation Between Facial Action Units for Detecting Deepfake Videos 622
Quoc Hoan Vu (The University of Queensland, Australia) and Priyanka Singh (The University of Queensland, Australia)

Perceptual Image Compression via Stable Diffusion at Low Bitrate 626
Luoxu Jin (Waseda University, Japan) and Hiroshi Watanabe (Waseda University, Japan)

Session 18: Demo Session

Building a Generative AI Showroom for Foundation Models with Different Modalities 630
Benny Stein (Fraunhofer Institute for Intelligent Analysis and Information Systems IAIS, Germany), Niklas Beck (Fraunhofer Institute for Intelligent Analysis and Information Systems IAIS, Germany), Daniel Becker (Fraunhofer Institute for Intelligent Analysis and Information Systems IAIS, Germany), and Dennis Wegener (Fraunhofer Institute for Intelligent Analysis and Information Systems IAIS, Germany)

Where You Look Matters in Group Photos: A Demo of GARGI iOS App 634
Omkar Kulkarni (University at Albany, SUNY, USA), Thomas Lloyd-Jones (University at Albany, SUNY, USA), My Tran (University at Albany, SUNY, USA), Gregory Vincent (University at Albany, SUNY, USA), Vivek K. Singh (Rutgers University, USA), and Pradeep K. Atrey (University at Albany, SUNY, USA)

Early Alzheimer's Detection: The Promise of AI-Powered MRI Analysis 638
Dominic Baker (Virginia State University), Wei-Bang Chen (Virginia State University), and He Gao (Virginia State University)

ProSchedule: A Comprehensive Mobile Solution for Seamless Academic Scheduling 642
He Gao (Virginia State University) and Wei-Bang Chen (Virginia State University)

Workshop Session: Workshop BDH: Big Data in Healthcare

Improving Machine Learning-Based Activity Type Prediction from Time-Series EEG Data 646
Syed Muhammad Raza Abidi (Dublin City University, Ireland), Tomas Emmanuel Ward (Dublin City University, Ireland), David Henshall (Royal College of Surgeons, Ireland), and Gabriel-Miro Muntean (Dublin City University, Ireland)

A Clustering-Based Sequence Variants Analysis Method for Electronic Medical Records of Multimedical Institutions 653
Hieu Hanh Le (Ochanomizu University, Japan), Yuki Yasumitsu (Tokyo Institute of Technology, Japan), Ryosuke Matsuo (Ochanomizu University, Japan), Tomoyoshi Yamazaki (Ochanomizu University, Japan), and Haruo Yokota (Josai University, Japan)

Privacy-Preserving Disease Prediction with Secure Data Deduplication on Untrusted Cloud Servers 660
Khushi Jain (The University of Queensland, Australia), Priyanka Singh (The University of Queensland, Australia), and Xue Li (The University of Queensland, Australia)

Understanding eGFR Trajectories and Kidney Function Decline via Large Multimodal Models 667
Chih-Yuan Li (National Sun Yat-sen University, Taiwan), Jun-Ting Wu (National Sun Yat-sen University, Taiwan), Chan Hsu (National Sun Yat-sen University, Taiwan), Ming-Yen Lin (Kaohsiung Medical University Hospital, Kaohsiung Medical University, Taiwan), and Yihuang Kang (National Sun Yat-sen University, Taiwan)

Self-Monitoring the Mental-Health State of a Focused Population with Multiple Self-Questionnaires and Sentiment Descriptions 674
Sukhan Lee (Sungkyunkwan University), Soojin Lee (Sungkyunkwan University), and Yaejin Lee (Sungkyunkwan University)

Big Data and Bigger Dilemmas: Ethical Concerns of Data in Healthcare 681
Nisha Daga (Indian Institute of Technology Jodhpur, India) and George Kodimattam Joseph (Indian Institute of Technology Jodhpur, India)

Patient 3D Data Visualisation with AR-Based Interactive Technology for Brain MRI 685
Vishakha Pareek (Technology Innovation Hub - iHub Drishti, Indian Institute of Technology Jodhpur, India), Shreyansh Sharma (Indian Institute of Technology Jodhpur, India), Vibhor Singh (JECRC University, India), and Shaswat Singh (IIT Jodhpur)

Author Index **691**