Wednesday, November 1, 2023

Grand Challenges Part 1 (10:30-12:30)
15 min per paper, including Q&A;
30 min for organizers presentations, including Q&A

10:30-11:00 – ACM Multimedia Computational Paralinguistics Challenge (ComParE) Organizers presentation, paper mmgc052i, *The ACM Multimedia 2023 Computational Paralinguistics Challenge: Emotion Share & Requests* **(Presenter: Shahin Amiriparian)**
11:00-11:15: Paper mmgc011, *Cascaded Cross-Modal Transformer for Request and Complaint Detection*
11:15-11:30: Paper mmgc013, *Advancing Audio Emotion and Intent Recognition with Large Pre-Trained Models and Bayesian Inference*
11:30-11:45: Paper mmgc 017, *Multi-Layer Acoustic & Linguistic Feature Fusion for ComParE-23 Emotion and Requests Challenge*
11:45-12:00: Paper mmgc 021, *Effect of Attention and Self-Supervised Speech Embeddings on Non-Semantic Speech Tasks*
12:00-12:30: Deep Video Understanding (DVU) 2023 Organizers presentation, paper mmgc044i, *The ACM Multimedia 2023 Deep Video Understanding Grand Challenge* **(Presenter: George Awad)**

Grand Challenges Part 2 (13:30-15:00)
15 min per paper, including Q&A;
30 min for organizers presentations, including Q&A

13:30-14:00: Multi-modal Behaviour Analysis for Artificial Mediation (MultiMediate) Organizers presentation, paper mmgc43i, *MultiMediate '23: Engagement Estimation and Bodily Behaviour Recognition in Social Interactions* **(Presenters Philip Mueller and Michal Balazia)**
14:00-14:15: Paper mmgc025, *MAGIC-TBR: Multiview Attention Fusion for Transformer-based Bodily Behavior Recognition in Group Settings* **(Presenter: Surbhi Madan)**
14:15-14:30: Paper mmgc024, *DCTM: Dilated Convolutional Transformer Model for Multimodal Engagement Estimation in Conversation* **(Presenter: proxy)**
14:30-14:45: Paper mmgc027, *BEAMER: Behavioral Encoder to Generate Multiple Appropriate Facial Reactions* **(Presenter: proxy)**
14:45-15:00: Paper mmgc008, *Gradient Boost Tree Network based on Extensive Feature Analysis for Popularity Prediction of Social Posts* **(Presenter: Chia-Ming Li)**

Grand Challenges Part 3 (15:30-17:30)
15 min per paper, including Q&A;
30 min for organizers presentations, including Q&A

15:30-16:00: Facial Micro-Expression Grand Challenge Organizers presentation, paper mmgc049i, *MEGC2023: ACM Multimedia 2023 ME Grand Challenge* **(Presenter: Adrian Davison)**
16:00-16:15: Paper mmgc028, *Efficient Micro-Expression Spotting Based on Main Directional Mean Optical Flow Feature* **(Presenter: proxy)**
16:15-16:30: Paper mmgc010, *Double-Fine-Tuning Multi-Objective Vision-and-Language Transformer for Social Media Popularity Prediction* **(Presenter: Weilong Chen)**
16:30-17:30: Time for discussion

Videos are available on Conflux System for all Grand Challenges papers, as listed below:

mmgc002	Finetuning Language Models for Multimodal Question Answering
mmgc003	A Hierarchical Deep Video Understanding Method with Large Language Model
mmgc004	Enhanced CatBoost with Stacking Features for Social Media Prediction
mmgc005	Semi-Supervised Multimodal Emotion Recognition with Expression MAE
mmgc006	Towards Realistic Conversational Head Generation: A Comprehensive Framework for Lifelike Video Synthesis
mmgc007	Invisible Video Watermark Method Based on Maximum Voting and Probabilistic Superposition
mmgc008	Gradient Boost Tree Network based on Extensive Feature Analysis for Popularity Prediction of Social Posts
mmgc009	VTQAGen: BART-based Generative Model For Visual Text Question Answering
mmgc010	Double-Fine-Tuning Multi-Objective Vision-and-Language Transformer for Social Media Popularity Prediction
mmgc011	Cascaded Cross-Modal Transformer for Request and Complaint Detection
mmgc012	Multi-scale Conformer Fusion Network for Multi-participant Behavior Analysis
mmgc013	Advancing Audio Emotion and Intent Recognition with Large Pre-Trained Models and Bayesian Inference
mmgc014	Automatic Audio Augmentation for Requests Sub-Challenge
mmgc016	Answer-Based Entity Extraction and Alignment for Visual Text Question Answering
mmgc017	Multi-Layer Acoustic & Linguistic Feature Fusion for ComParE-23 Emotion and Requests Challenge
mmgc018	Sliding Window Seq2seq Modeling for Engagement Estimation
mmgc019	Micro-Expression Spotting with Face Alignment and Optical Flow
mmgc020	UniFaRN: Unified Transformer for Facial Reaction Generation
mmgc021	Effect of Attention and Self-Supervised Speech Embeddings on Non-Semantic Speech Tasks
mmgc023	Data Augmentation for Human Behavior Analysis in Multi-Person Conversations
mmgc024	DCTM: Dilated Convolutional Transformer Model for Multimodal Engagement Estimation in Conversation
mmgc025	MAGIC-TBR: Multiview Attention Fusion for Transformer-based Bodily Behavior Recognition in Group Settings
mmgc026	Hierarchical Audio-Visual Information Fusion with Multi-label Joint Decoding for MER 2023
mmgc027	BEAMER: Behavioral Encoder to Generate Multiple Appropriate Facial Reactions
mmgc028	Efficient Micro-Expression Spotting Based on Main Directional Mean Optical Flow Feature
mmgc030	Mining High-quality Samples from Raw Data and Majority Voting Method for Multimodal Emotion Recognition
mmgc031	Deep Video Understanding with Video-Language Model
mmgc032	Semi-Supervised Multimodal Emotion Recognition with Class-Balanced Pseudo-labeling
mmgc033	Leveraging the Latent Diffusion Models for Offline Facial Multiple Appropriate Reactions Generation
mmgc034	Improvements on SadTalker-based Approach for ViCo Conversational Head Generation Challenge
mmgc035	Multimodal Emotion Recognition in Noisy Environment Based on Progressive Label Revision
mmgc036	Integrating VideoMAE based model and Optical Flow for Micro- and Macro-expression Spotting
mmgc037	Hierarchical Semantic Perceptual Listener Head Video Generation: A High-performance Pipeline
mmgc038	Unveiling Subtle Cues: Backchannel Detection Using Temporal Multimodal Attention Networks
mmgc039	Query-aware Long Video Localization and Relation Discrimination for Deep Video Understanding
mmgc040	Building Robust Multimodal Sentiment Recognition via a Simple yet Effective Multimodal Transformer
mmgc041	MultiMediate 2023: Engagement Level Detection using Audio and Video Features
mmgc044i	The ACM Multimedia 2023 Deep Video Understanding Grand Challenge
mmgc045i	MER 2023: Multi-label Learning, Modality Robustness, and Semi-Supervised Learning
mmgc046i	Learning and Evaluating Human Preferences for Conversational Head Generation
mmgc047i	REACT2023: The First Multiple Appropriate Facial Reaction Generation Challenge
mmgc049i	MEGC2023: ACM Multimedia 2023 ME Grand Challenge
mmgc051i	ACM Multimedia 2023 Grand Challenge Report: Invisible Video Watermark
mmgc052i	The ACM Multimedia 2023 Computational Paralinguistics Challenge: Emotion Share & Requests
mmgc43i	<sc>MultiMediate '23:</sc> Engagement Estimation and Bodily Behaviour Recognition in Social Interactions
mmgc48i	VTQA2023: ACM Multimedia 2023 Visual Text Question Answering Challenge
mmgc50i	SMP Challenge: An Overview and Analysis of Social Media Prediction Challenge