EDITORIAL

Special issue on human-centric intelligent multimedia understanding

Zhenguang Liu[1] · Roger Zimmermann[2] · Li Cheng[3]

The basic objective of artificial intelligence techniques is to meet the demands of people, which is inseparable from a proper understanding of humans. Human-centric intelligent multimedia understanding, as an essential part of machine intelligence, relies on a wide spectrum of human-centric multimedia understanding tasks such as human pose estimation, action recognition, person re-identification, transportation data mining, medical image processing, and social media data understanding. From another perspective, human behavior understanding from multimedia contents is a fundamental driving force for related artificial intelligence applications such as smart city, medical image understanding, and human behavior analysis. However, there are still various open questions in improving the robustness, explainability, and accuracy of human-centric multimedia understanding approaches, while large-scale benchmark datasets and effective methods on these tasks are yet to be explored.

This special issue aims to bring together the latest advancements in human-centric intelligent multimedia understanding, diving deeper into human-centered behavior understanding tasks. Overall, we have accepted 11 high quality papers. We expect the papers to contribute innovative techniques, including methodologies and algorithmic approaches to solve theoretical and practical multimedia problems. We hope these high-quality works could push forward the boundary of human-centric multimedia understanding.

The paper entitled "Position Constrained Network for 3D Human Pose Estimation" proposes a human pose estimation

framework which exploits 3D root coordinates as subordinate input to 2D joint coordinates to provide positional reference to the recovered 3D joint coordinates, and employs inner camera parameters to construct additional projection constraints for recovering 3D joint coordinates.

The paper entitled "Pedestrian Attribute Recognition based on Attribute Correlation" presents a Cross Attribute and Feature Network (CAFN) that fully exploits the correlations between any pair of attributes for pedestrian attribute recognition.

The paper entitled "Attribute-aware Style Adaptation for Person Re-identification" designs an attribute-aware style adaptation based on CamStyle to combine fine-grained style adaptation and discriminative person re-identification.

The paper entitled "Multi-head attention based two-stream EfficientNet for action recognition" develops a multi-head attention-based two-stream EfficientNet for action recognition.

The paper entitled "Context-aware and Ethics-first Crowd Mobility Portraits over Massive Smart Card Data" introduces an efficient ranking algorithm to figure out trajectories mostly contributing to the city-wide mobility pattern and constructs time-varying land functions to textualize trajectories.

The paper entitled "Perturbation Consistency and Mutual Information Regularization for Semi-Supervised Semantic Segmentation" tries to reduce the number of required labels in the field of human-centric understanding by proposing a semi-supervised model for semantic segmentation.

The paper entitled "FPF-Net: Feature Propagation and Fusion based on Attention Mechanism for Pancreas Segmentation" advocates to extract features in a propagation and fusion manner.

The paper entitled "Asymmetric Exponential Loss Function for Crack Segmentation" analyzes the loss functions during the training process of crack segmentation tasks and proposes an asymmetric exponential loss function that addresses sample biases and dataset biases.

The paper entitled "Evil vs Evil: Using Adversarial Examples to Against Backdoor Attack in Federated

✉ Zhenguang Liu
liuzhenguang2008@gmail.com

Roger Zimmermann
rogerz@comp.nus.edu.sg

Li Cheng
lcheng5@ualberta.ca

[1] Zhejiang University, Hangzhou, China

[2] National University of Singapore, Singapore, Singapore

[3] University of Alberta, Edmonton, AB, Canada

Learning" investigates adversarial attacks on training models in the presence of multiple human clients.

The paper entitled "Improving Text-Image Cross-Modal Retrieval with Contrastive Loss" attempts to improve text-image cross-modal retrieval for human users.

The paper entitled "SMPC: Boosting Social Media Popularity Prediction with Caption" introduces word-level, sentence-level and length-level processing into the method and designs seven variants of architectures for exploring which level of caption and which combination of capture features wins best performance.