Editorial: Special Issue on Recent Advances in Cognitive Learning and Data Analysis

Bin Luo[1] · **Jin Tang**[1] · **Cheng-Lin Liu**[2]

Multi-modal big data modeling and analysis have received growing attention recently [1, 2]. In particular, deep learning has achieved great success in a variety of data-intensive applications, especially in computer vision [3, 4] and speech recognition [5, 6]. Continuing advancements in such artificial intelligence (AI) techniques are also promoting interdisciplinary research in cognitive modeling and cognitive systems [7].

To provide a platform for researchers to exchange novel ideas in multidisciplinary research, we successfully organized the 11th International Conference on Brain Inspired Cognitive System (BICS) in Hefei city, China, from Dec 18 to 20, 2020, as a sequel of BICS 2004–2019. The first BICS 2004 was held in Stirling, Scotland, UK, and the last BICS 2019 was held in Guangzhou, China.

BICS 2020 aimed to provide a high-level international forum for scientists, engineers, and educators to present the state of the art of brain-inspired cognitive systems research and applications in diverse fields. The conference featured plenary lectures given by world renowned scholars, regular sessions with broad coverage, and some special sessions focusing on popular and timely topics. Eleven highest quality papers were selected, out of 45 submissions, and significantly extended for further peer review (by at least 3 independent reviewers) and included in this special issue. Papers co-authored by guest editors were peer reviewed by independent editors in line with journal policy.

Among the selected papers, most are related to deep learning, and two aim to advance traditional research paradigms. These reflect emerging research trends in AI and cognitive computation, in particular, introduction of recently developed deep learning techniques, including graph learning, attention networks, and transformer models, into cognitive computation for different applications. The papers in this Special Issue target a variety of application modalities, most devoted to challenging image or video applications.

The first selected paper "Action Recognition with a Multi-View Temporal Attention Network" introduces global temporal attention pooling and feature-level multi-view fusion as part of a novel action recognition model based on a multi-view temporal attention mechanism. The authors found that the temporal attention layer can accurately capture key frames which can improve the performance of action recognition.

In the paper "An Ensemble of Complementary Models for Deep Tracking," the authors introduced an attention mechanism to highlight discriminative features of different convolutional neural network (CNN) architectures and explored complementary properties of different CNNs for visual tracking. The prediction scores of all CNNs were adaptively fused to obtain robust tracking performance. They concluded that a combination of complementary models can better track objects in terms of accuracy and robustness.

By introducing an attention residual refinement module and a feature reuse module, the work "Local Enhancement and Bidirectional Feature Refinement Network for Single-Shot Detector" designed a network which utilizes the inter-channel relationship of higher level visual features and lower level visual features. The experimental results showed that the proposed method outperforms state-of-the-art object detectors.

"Multistage Model for Robust Face Alignment Using Deep Neural Networks" proposed a multistage model based on deep neural networks for face alignment to tackle challenging problems of severe occlusions and large pose

✉ Bin Luo
 luobin@ahu.edu.cn

 Jin Tang
 tj@ahu.edu.cn

 Cheng-Lin Liu
 liucl@nlpr.ia.ac.cn

[1] Anhui University, Hefei, China

[2] Institute of Automation of Chinese Academy of Sciences, Beijing, China

variations. Extensive experiments demonstrate the superior performance of the proposed method over state-of-the-art approaches.

The authors of "Dual Attention with the Self-Attention Alignment for Efficient Video Super-resolution" addressed the lack of suitable attention structures to achieve efficient video super resolution. They proposed a dual position attention and a channel attention module to enhance spatiotemporal features, and a self-attention structure to achieve attention alignment. The long short-time memory (LSTM) network was also used to guarantee coherent consistency of generated video frames both temporally and spatially.

Inspired by the auditory perception principle utilized by humans, the paper "SETransformer: Speech Enhancement Transformer" takes advantage of LSTM and a multi-head attention mechanism to propose a cognitive-inspired speech enhancement model. The proposed model improved the performance in speech quality and speech intelligibility under unseen noise conditions.

To meet the challenge of multi-modal data retrieval, the paper "Unsupervised Multi-modal Hashing for Cross-Modal Retrieval" proposed a novel unsupervised cross-modal hashing method. In particular, the semantic correlation in textual space and locally geometric structure in visual space were reconstructed by unified hashing features. The authors concluded that the proposed framework is effective in learning hash codes and achieves superior retrieval performance compared to state-of-the-art methods.

In "Separable Reversible Data Hiding Based on Integer Mapping and MSB Prediction for Encrypted 3D Mesh Models," the authors proposed a reversible data hiding method for encrypted 3D meshes based on integer mapping and most significant bit (MSB) prediction. Experimental results demonstrated that the proposed method has greater embedding capacity compared to state-of-the-art approaches.

In order to address the social influence prediction problem, the paper "MvInf: Social Influence Prediction with Multi-view Graph Attention Learning" proposes a deep learning framework which combines multi-view learning and graph attention neural network. Experimental results demonstrated the superior performance of the proposed MvInf model compared to previous single view–based approaches.

The paper "A Possible Explanation for the Generation of Habit in Navigation: a Striatal Behavioral Learning Model" proposed a striatal behavioral learning model which is composed of the striosome model and a matrix model. The model highlighted the role of the striatum in reward-based learning, action selection, and exploratory behavior. Comparison results showed that the proposed model was more efficient and robust than the widely used striatal temporal-difference learning model.

The "Joint Adaptive Graph Learning and Discriminative Analysis for Unsupervised Feature Selection" proposed a new unsupervised feature selection method with a predefined graph that was self-adjusted by the original graph and learned subspace. An uncorrelated constraint was also added to enhance the discriminability of the model. Experimental results demonstrated that the proposed adaptive graph learning strategy can learn a high-quality graph with accurate information.

To conclude, this special issue addressed recent interdisciplinary advances in cognitive computation research. The collection of eleven high-quality papers showed the diversity of research topics and applications in order to capture existing and emerging trends.

Finally, we would like to thank the authors who contributed to this special issue, the anonymous reviewers whose invaluable comments and suggestions ensured the high-quality of the contributions, and finally the management and editorial team of *Cognitive Computation*. Our special thanks also go to the speakers of the conference whose papers do not appear in this special issue. Without their great contributions, the conference and this consequent high-quality special issue would not have been possible.

References

1. Bayoudh K, Knani R, Hamdaoui F, Mtibaa A. A survey on deep multimodal learning for computer vision: advances, trends, applications, and datasets. Vis Comput. 2021; online first.
2. Zhang Y, Sidibé D, Morel O, Mériaudeau F. Deep multimodal fusion for semantic image segmentation: a survey. Image Vis Comput. 2020;105:104042.
3. Zhao ZQ, Zheng P, Xu ST, Wu X. Object detection with deep learning: a review. IEEE Trans Neural Netw Learn Syst. 2019;30(11):3212–32.
4. Xie J, Zheng Y, Du R, Xiong W, Guo J. Deep learning-based computer vision for surveillance in ITS: evaluation of state-of-the-art methods. IEEE Trans Veh Technol. 2021;70(4):3027–42.
5. Shahamiri SR. Speech vision: an end-to-end deep learning-based dysarthric automatic speech recognition system. IEEE Trans Neural Syst Rehabil Eng. 2021;29:852–61.
6. Abbaschian BJ, Sierra-Sosa D, Elmaghraby A. Deep learning techniques for speech emotion recognition, from databases to models. Sensors. 2021;21(4):1249.
7. Perconti P, Plebe A. Deep learning and cognitive science. Cognition. 2020;203:104365.