

Editorial

Complex Deep Learning and Evolutionary Computing Models in Computer Vision

Li Zhang ¹, **Chee Peng Lim**,² and **Jungong Han**³

¹Northumbria University, Newcastle, UK

²Deakin University, Geelong, Australia

³Lancaster University, Lancaster, UK

Correspondence should be addressed to Li Zhang; li.zhang@northumbria.ac.uk

Received 15 April 2019; Accepted 15 April 2019; Published 1 July 2019

Copyright © 2019 Li Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Computer vision, which deals with how computers can be used to gain high-level understanding pertaining to information contained in digital images or videos, is an important, yet challenging, technology. The advent of deep learning and associated paradigms such as evolutionary computing models has propelled computer vision to the next level, solving a variety of complex problems in diverse applications, such as object detection, motion tracking, semantic segmentation, and emotion recognition. In this special issue, recent advances with respect to mathematical modeling, simulation, and/or analysis of deep learning and evolutionary computing models for undertaking complex phenomena in computer vision, are presented. A variety of problems, which include automatic object tracking, understanding image content, optical character recognition, personalized recommendation and movie summarization systems, and underwater video streaming, are covered. A description of each article is presented, as follows.

The paper titled “Multimodal Deep Feature Fusion (MMDF) for RGB-D Tracking” addresses the limitation of existing tracking methods in processing geometrical features extracted from depth images by using a multimodal deep feature fusion model. The proposed model consists of four deep convolutional neural networks (CNNs). It extracts RGB (red, green, blue) and depth features from images using RGB-specific and depth-specific CNN models and exploits their correlated relationship using an RGB-depth correlated CNN model. In addition, a motion-specific CNN is used to provide high-level motion information for object tracking. Empirical evaluation with two RGB-depth datasets demonstrates that the proposed method achieves better performances,

especially in situations where occlusion occurs, the movement is fast, and the target size is small, as compared with those from other state-of-the-art trackers.

Semantic image segmentation is useful for understanding the semantic information contained in an image. To improve the time-consuming pixel-level annotation process, a weakly supervised semantic segmentation method using the CNN and extreme learning machine is proposed in the paper titled “Weakly Supervised Deep Semantic Segmentation Using CNN and ELM with Semantic Candidate Regions”. Semantic inference of candidate regions is realized based on the relationship and neighborhood rough set associated with semantic labels. After completing the inference step of all semantic labels, the extreme learning machine is used to learn the inferred candidate regions. The experimental results of two benchmark datasets show the proposed method is able to outperform several state-of-the-art alternatives for deep semantic segmentation.

To undertake imbalanced datasets, a deep learning model for unbalanced distribution character recognition based on a focal connectionist temporal classification (CTC) loss function is proposed in the paper titled “Focal CTC Loss for Chinese Optical Character Recognition on Unbalanced Datasets”. The proposed model consists of three main components: convolutional, recurrent, and transcription layers. The residual network is used as the convolutional layers, which extract a feature sequence from the input image. The bidirectional long short-term memory is used as the recurrent layers, which predicts a label distribution for each frame. The focal CTC function is used in the transcription layer, which translates the per-frame predictions into the final

label sequence. A series of experimental studies using both synthetic and real image sequence datasets indicates that the proposed model is able to achieve better performance as compared with those from traditional CTC on imbalanced datasets.

The paper titled “A New Type of Eye Movement Model Based on Recurrent Neural Networks for Simulating the Gaze Behavior of Human Reading” tackles gaze behavior of human reading as a word-based sequence labeling task in natural language processing. The eye movement data are used to train a model that can predict the eye movements of the same reader reading a previously unseen text. Based on a combination of CNN models, bidirectional long short-term memory networks, and conditional random fields, a recurrent neural network is used to generate a gaze point prediction sequence. The empirical results indicate that the recurrent neural network-based model is able to achieve similar accuracy in predicting a user’s fixation points during reading, with the advantages of less reliance on data features and less preprocessing than some existing machine learning models.

In e-commerce services, it is important for a personalized recommendation system to learn the latent user and item representations from implicit interactions between users and items. A neural personalized ranking model for collaborative filtering with implicit frequency feedback is introduced in the paper titled “Neural Personalized Ranking via Poisson Factor Model for Item Recommendation”. A ranking-based Poisson factor model is developed, which adopts a pair-wise learning method to learn the rankings of preferences between items. A multilayer perceptron is used to learn the nonlinear user-item interaction relationships. The personalized ranking model is able to capture the complex structure of user-item interactions. The empirical results indicate the superiority of the proposed method over state-of-the-art recommendation algorithms.

To allow viewers to have an idea about the semantics of a movie in a short time, a movie summarization system produces a short video sequence that contains the most important scenes from the movie. In the paper titled “Personalized Movie Summarization using Deep CNN-Assisted Facial Expression Recognition”, a user preference-based movie summarization technique is developed using a deep CNN model to analyze the emotional state of the characters through facial expression recognition. Segmentation of movie shots with faces using an entropy method is conducted. Then, a summary with respect to user preference from seven basic emotion classes is produced. A subjective evaluation using five Hollywood movies shows the effectiveness of the proposed scheme in terms of user satisfaction, while an objective evaluation indicates its superiority over state-of-the-art movie summarization methods.

Traffic loads and congestion management are important issues in wireless sensor networks. A proactive caching strategy based on a stacked sparse autoencoder to predict data package content popularity is devised in the paper titled “Deep Learning Based Proactive Caching for Effective WSN-Enabled Vision Applications”. A simple frame structure of software defined network and network function virtualization technologies, coupled with the autoencoder in

the sink and control nodes of the wireless sensor network, is constructed. The structure and model parameters associated with the stacked sparse autoencoder are optimized through training. A series of simulation studies to compare the proposed method with traditional classical caching strategies indicate that the stacked sparse autoencoder is able to improve the prediction accuracy for enhancing performance of wireless sensor networks.

Most of the classical monocular visual simultaneous localization and mapping (SLAM) methods do not consider the motion characteristics of the platform during the initialization phase. As such, a motion hypothesis to transform the solution of camera motion into an error elimination problem during the initialization process is introduced in the paper titled “Passive Initialization Method Based on Motion Characteristics for Monocular SLAM”. The error is reduced by using a multiframe optimization method based on Bundle Adjustment, thus improving the accuracy of the initialization process. A hardware-in-the-loop simulation system on a fixed-wing aircraft is established as the test platform. The results indicate that the success rate of monocular SLAM initialization can be greatly improved, as compared with that of existing methods. However, this method cannot be used indiscriminately on platforms characterized by randomized motions.

In the paper titled “Design and Implementation of an Assistive Real-time Red Lionfish Detection System for AUV/ROVs”, a remotely operated underwater vehicle with a robotic system for divers to locate red lionfish through real-time object recognition with a CNN-based model is designed and implemented. The assistive robot is able to identify red lionfish such that the divers can maximize their catch before each dive. The underwater vehicle is equipped with a camera to collect live videos underwater, and the video streams are processed in real-time to detect red lionfish. The developed system has been evaluated in areas currently invaded by red lionfish in the Gulf of Mexico. The outcome indicates the usefulness of the system for detecting red lionfish with high confidence in real-time.

It is hoped that this special issue serves as a cornerstone to further stimulate and promote research studies related to theory and application of deep learning and evolutionary computing models for advancing computer vision technology and delivering benefits to our society.

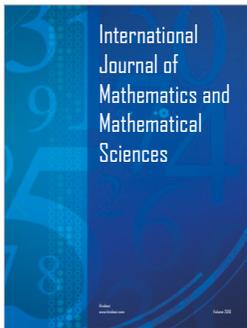
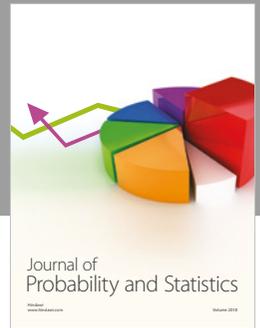
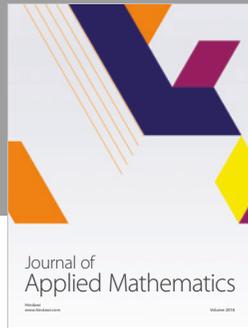
Conflicts of Interest

The editors declare no conflicts of interest.

Acknowledgments

The guest editors would like to thank the authors for contributing their articles and the reviewers for improving the quality of the articles through constructive comments and suggestions.

*Li Zhang
Chee Peng Lim
Jungong Han*



Hindawi

Submit your manuscripts at
www.hindawi.com

