Editorial

Learning-Based Multimedia Analyses and Applications

Chen Gong ⬤,[1] Zechao Li,[1] Xiaojun Chang,[2] and Yong Luo[3]

[1]*Nanjing University of Science and Technology, Nanjing, China*
[2]*Carnegie Mellon University, Pittsburgh, USA*
[3]*Nanyang Technological University, Singapore*

Correspondence should be addressed to Chen Gong; chen.gong@njust.edu.cn

The methodologies on multimedia analysis usually combine different sources/modalities of information such as text, audio, and images to solve various practical tasks related to advertisement, education, art, and so on. In recent years, machine learning has gained much popularity and has been intensively applied to deal with different kinds of multimedia problems. However, massive problems still remain unsolved on both algorithm design and multimedia applications.

This special issue is organized by four guest editors and contains eight research papers. The topics of the accepted papers have a wide coverage and focus on many critical aspects in multimedia analysis, like behavior detection, audio restoration, image retrieval, person reidentification, and so on. These works have shown important discoveries and made a breakthrough in the related research fields.

J. Li et al. study the self-protective behaviors of dairy cows when suffering from dipteral insect infestation for evaluating the breeding environment and cows' selective breeding. Specifically, they develop an automatic monitoring system based on video analysis. By combining the morphological features of head, leg, and tail movements, this method effectively reduces the number of Shi-Tomasi points and eliminates interference from background movement, thus reducing the computational complexity of the algorithm and meanwhile improving the detection accuracy.

C. Jin et al. investigate the distortion problems in the historical audio and video data. Based on the deep learning network, this work designs an objective quality evaluation system for historical audio/video data and evaluates the performance of the system and the audio signal quality from the perspective of feature extraction and network parameter selection.

B. Yang et al. propose an anomaly detection approach by learning a generative model using deep neural network. In their work, a weighted convolutional autoencoder (AE) and long-short-term memory (LSTM) network are proposed to reconstruct raw data and perform anomaly detection based on reconstruction errors to resolve the existing challenges of anomaly detection in complicated background. Convolutional AEs and LSTMs are used to encode spatial and temporal variations of input frames, respectively. A weighted Euclidean loss is proposed to enable the network to concentrate on moving foregrounds to suppress the influence of background.

Y. Ruan et al. study the automatic audio announcement systems for the people with hearing impairment. In this paper, an approach of audio announcement detection and recognition for the hearing-impaired people based on the smart phone is proposed, and a mobile phone application (app) is developed under the bank scenario. For audio announcement detection, a method based on audio segment classification and postprocessing is proposed, which uses a SVM classifier trained on audio announcements and environment noise collected in banks. For announcement speech recognition, an ASR engine is developed by using a GMM-HMM-based acoustic model and a finite state transducer (FST) based grammar.

H. Xu et al. present a novel approach for semantic image retrieval by combining Convolutional Neural Network (CNN) and Markov Random Field (MRF). Unlike previous works that use single-concept classifiers one by one, this paper detects semantic multiconcept by using a multiconcept scene classifier. Specifically, they first train a CNN as a concept classifier, which further includes two types of classifiers:

a single-concept fully connected classifier that is best suited to single-concept detection and a multiconcept scene fully connected classifier that is effective for holistic scene detection. Then they propose an MRF-based late fusion approach that is able to effectively learn the semantic correlation between the single-concept classifier and multiconcept scene classifier.

W. Zhao et al. investigate the singing synthesis in Beijing Opera. The singing of Beijing Opera carries some features of speech but it has its own unique pronunciation rules and rhythms which differ from ordinary speech and singing. Firstly, the speech signals of the source speaker and the target speaker are extracted by using the existing algorithm. After that, through the training of GMM, they complete the voice control model to input the voice to be converted and output the voice after the voice conversion. Finally, by modeling the fundamental frequency, duration, and frequency separately, a melodic control model is constructed using GAN to realize the synthesis of the Beijing Opera fragment.

Q. Leng et al. study the person reidentification, which is a key technique for intelligent video surveillance. To bridge small sample size (SSS) problem and learning model with small labels, a novel semisupervised cometric learning framework is proposed to learn a discriminative Mahalanobis-like distance matrix for label-insufficient person reidentification. Different from typical cotraining task that contains multi-view data originally, single-view person images are firstly decomposed into pseudo two views, and then metric learning models are produced and jointly updated based on both pseudolabels and references iteratively.

F. Kang et al. study the segmentation of piglets in pig farms. In view of the noninteractive and real-time requirements of a video monitoring system, this paper proposes a method of image segmentation based on an improved noninteractive GrabCut algorithm. The functions of preserving edges and noise reduction are realized through bilateral filtering. An adaptive threshold segmentation method is used to calculate the local threshold and to complete the extraction of the foreground target.

In summary, this special issue provides a thorough overview of the present status of learning-based multimedia analyses. The recent progress on techniques and applications is selected and highlighted, which provide valuable references for future research.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

Chen Gong
Zechao Li
Xiaojun Chang
Yong Luo