Signal Processing: Image Communication 47 (2016) 463–464

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

journal homepage: www.elsevier.com/locate/image

Guest Editors' Introduction: Special issue on deep learning with applications to visual representation and analysis

The last decade has witnessed an exponential growth of visual information and an unprecedentedly broad range of applications of image and video. To efficiently access, utilize, and transmit visual information, visual representation and analysis have attracted intensive attention and become one of the most active research topics in the fields of signal processing, computer vision, and artificial intelligence (AI). Recently, deep learning has advanced as an AI approach that can automatically discover good representations and model high-level abstractions from data. It has achieved record-breaking performance on a spectrum of visual analysis tasks. At the same time, the full potential of deep learning for visual representation and analysis has yet to be explored and many theoretical and practical issues remain unsolved. This special issue consists of nine papers selected from the submission, which report new research explorations in employing, improving, and designing deep learning algorithms for visual representation and analysis. As guest editors, we received strong support from a team of committed reviewers and the journal management team, especially EIC Frédéric Dufaux.

Two papers of this special issue are about facial image analysis. The paper "Landmark Perturbation-Based Data Augmentation for Unconstrained Face Recognition" written by Jiang jing Lv, Cheng Cheng, Guodong Tian, Xiangdong Zhou, and Xi Zhou aims to improve the robustness of deep convolutional neural networks with respect to facial landmark misalignment. A data augmentation method is proposed to achieve this goal. Given a facial image, this method automatically perturbs landmark positions to generate a large number of misaligned copies to train the networks. Experimental study on multiple benchmark data sets shows that this augmentation method can attain better recognition performance in both face verification and identification cases. The paper "FaceHunter: A Multi-task Convolutional Neural Network Based Face Detector" by Dong Wang, Jing Yang, Jiankang Deng, and Qingshan Liu proposes a new multi-task-based convolutional neural network for face detection. Its objective function considers both the discrimination between face and non-face images and the accuracy of bounding box regression, which allows region proposal classification and bounding box refinement to be jointly performed. To improve computational efficiency, regional proposal network is employed to directly generate proposals from convolutional feature map. Experimental result on five face data sets demonstrates the state-of-the-art performance of this detection framework.

Pedestrian detection plays a key role in a wide range of practical applications on automotive, surveillance and robotics. A deep learning based pedestrian detection system is developed in the paper "Deep Convolutional Neural Networks for Pedestrian Detection" by Denis Tomè, Federico Monti, Luca Baroffio, Luca Bondi, Marco Tagliasacchi, and Stefano Tubaro. This system thoroughly analyzes and optimizes the key steps such as region proposals, fine-tuning and data augmentation in the pedestrian detection pipeline based on convolutional neural networks. Extensive experiments are conducted to demonstrate the improved detection accuracy achieved by this system. In addition, a lightweight version of this system is developed and tested. The paper "Robust Object Representation by Boosting-like Deep Learning Architecture" by Lei Wang, Baochang Zhang, Jungong Han, Linlin Shen, and Cheng-shan Qian develops a framework to fuse hand-crafted features and learning-based features. A multi-stage feature structure is proposed to concatenate the output at different stages of the neural networks, and a boosting-like deep learning algorithm is invented to adaptively weight samples in the training process. Experimental study on pedestrian detection and action recognition demonstrates the competitive performance of this framework.

Two papers of this special issue develop algorithms to assess esthetic image quality, which has recently attracted much research attention in the field of computer vision and image analysis. Considering that the assessment criterion could be dependent on the category of images, the paper "Hierarchical Esthetic Quality Assessment using Deep Convolutional Neural Networks" written by Yueying Kao, Kaiqi Huang, and Steve Maybank trains convolutional neural networks for scene, object and texture separately to learn the image representation for esthetic quality assessment. In addition, a single convolutional neural network is proposed to jointly perform image categorization and the assessment. The other paper "A Multi-Scene Deep Learning Model for Image Esthetic Evaluation" by WeiNing Wang, Mingquan Zhao, Li Wang, Jiexiong Huang, Chengjia Cai, and Xiangmin Xu also considers the dependence of assessment criterion on image type. In that work, a special layer is embedded into convolutional neural networks to discriminatively learn representation for seven types of images. In addition, that paper takes advantage of pre-trained networks and a pre-training stage to handle the scarcity of training samples, the unbalance of image classes, and noisy label information. Both papers demonstrate the effectiveness of the proposed methods on the benchmark data sets of image esthetic assessment.

Inspired by the visual attention mechanism in human visual system, the paper "Ada-Sal Network: Emulate the Human Visual System" by Yunong Wang, Nenghai Yu, and Taifeng Wang proposes to adaptively learn a visual saliency mask with the feature map of

http://dx.doi.org/10.1016/j.image.2016.09.003

an image by modifying the lower end of a deep learning network. This mask is multiplied with the feature map to realize a mechanism of region selection and resource allocation, and the masked feature map is used as the input to go through the following layers to perform classification. Experimental result on two benchmark data sets shows the benefit brought by this saliency mask.

Finally, this special issue includes two papers that develop deep neural networks for interesting applications, Indian classical dance understanding and driver's cognitive performance prediction. The paper "Nrityabodha: Towards Understanding Indian Classical Dance using a Deep Learning Approach" by Aparna Mohanty, Pratik Vaishnavi, Prerana Jana, Anubhab Majumdar, Alfaz Ahmed, Trishita Goswami, and Rajiv R Sahay aims to understand the semantic meaning of the dance performance by recognizing hand gestures and body postures with convolutional neural networks. This work creates a set of data sets and demonstrates the superiority of the learned feature representation to the hand-crafted ones. Also, the help of transfer learning to faster convergence and better generalization performance is shown in this application. The other paper "EEG-based Prediction of Driver's Cognitive Performance by Deep Convolutional Neural Network" by Mehdi Hajinoroozi, Zijing Mao, and Yufei Huang predicts driver's cognitive states through EEG signals. To take the characteristics of EEG data into account, it develops a novel channel-wise convolutional neural network to learn feature representation for classification. In addition, it considers using Restricted Boltzmann Machine to replace the convolutional filter for feature extraction. The advantage of the proposed framework is demonstrated on a large EEG data set collected by this paper.

This special issue demonstrates the variety of image and signal analysis tasks that deep learning technique has been applied to and the necessity of developing advanced learning frameworks and algorithms to fully release the power of this fascinating technique. It is hoped that this special issue will provide new insights and spark fresh ideas for the research on deep learning for visual representation and analysis.

Guest editors
Lei Wang, Ce Zhu, Jieping Ye, Juergen Gall