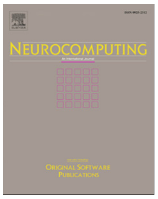
Neurocomputing 444 (2021) 214–216

Contents lists available at ScienceDirect

Neurocomputing

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



Editorial

Deep learning for human activity recognition

1. Introduction

In the past two decades, a large number of artificial intelligence (AI) systems have been developed in many areas, including computer vision, data mining, machine learning, natural language processing, business intelligence, robotics, and others. Among these AI systems, deep neural networks have become one of the most popular machine learning techniques, achieving great success especially in speech recognition and image processing fields [1]. Many novel deep learning techniques have emerged, driven by the boost in parallel computing from graphical processing unit (GPU) acceleration and the growing availability of large datasets.

Human activity recognition (HAR) can benefit various real-world applications, such as health-care services and smart home applications [2]. Different types of sensors have been utilized for human activity recognition, such as wearable sensors, smartphone sensors, cameras, radio frequency (RF) sensors (e.g., WiFi and RFID), LED light sensors, etc. Owing to the rapid development of wireless sensor network, a large amount of data has been collected for the recognition of human activities using different kinds of sensors. Different machine learning algorithms have been applied to classify them into different activities. Nevertheless, conventional shallow machine learning algorithms, such as Support Vector Machines and Random Forest, require human beings to manually extract some representative features from large and noisy sensory data [3,6]. However, manual feature engineering requires expert knowledge and will inevitably miss implicit yet useful features. Meanwhile, deep learning mentioned above has achieved great success in many challenging research areas, such as image recognition and natural language processing. The key merit of deep learning is to automatically learn representative features from massive amount of data. This technology can thus be a good candidate for human activity recognitions [4,5]. While some initial attempts can be found in the literature, many challenging research problems in terms of detection accuracy, device heterogeneities, environment changes, etc. remain unsolved.

This special issue collection intends to prompt both state-of-the-art approaches and latest innovative methods on deep learning for human activity recognition. We have received 43 submissions. Considering the high quality of these submissions and very positive feedbacks from peer-reviewers, a total of 18 submissions have been accepted for publication on this special issue. Below, we will group them into 4 different categories based on sensor types and activity types, where the first 3 categories focus on the wearable sensors, Radio Frequency (RF)-based sensors, camera sensors respectively, and last category focuses on special activities.

1. Due to the low-cost and easy-to-use properties of wearable sensors, four works explored wearable sensor based HAR with deep learning. The first paper, *Human Activity Recognition by Manifold Regularization Based Dynamic Graph Convolutional Networks*, by *Liu et al.*, proposed a novel semi-supervised learning method, i.e., manifold regularized dynamic graph convolutional network (MRDGCN), for human activity recognition. It is able to automatically update the structure information via manifold regularization during model training. When evaluating on human activity recognition datasets, the MRDGCN outperforms the conventional GCN and other semi-supervised learning algorithms. The second paper, *Human Activity Recognition based on Smartphone and Wearable Sensors Using Multiscale DCNN Ensemble*, by *Souza et al.*, firstly processed each sensor separately, learning their features and performing the classification before fusing with the other sensors. Then, an ensemble of Deep Convolution Neural Networks (DCNN) was developed to extract patterns in multiple temporal scales of the data. The third paper, *Domain Models for Data Sources Integration in HAR*, by *Hamidi and Osmani*, considered data source integration for the trade-off between the quantities of sampled data and recognition performance for human activity recognition. Three domain models were developed to integrate information source efficiently. The experimental results show that their method can significantly improve the performance of baseline setting with only one-half of the data. The last paper, *Wearables-based multi-task gait and activity segmentation using recurrent neural networks*, by *Martindale et al.*, developed a multi-task recurrent neural network (RNN) framework that uses inertial sensor data to both segment and recognizes activities and cycles. In particular, the framework is a combination of a convolutional neural network (CNN) for detecting edges and a RNN for modelling the temporal dependency of the data. Experiments on three public datasets show that it outperforms several state-of-the-arts for activity recognition and cycle analysis.
2. Another popular sensor for HAR is Radio Frequency (RF)-based sensors, such as WiFi and radar. We have accepted two papers. The first paper, *Towards CSI-based Diversity Activity Recognition Via LSTM-CNN Encoder-Decoder Neural Network*, by *Guo et al.*, deals with HAR based on WiFi Channel State Information (CSI). It designed a novel deep learning model called LCED which consists of one LSTM-based Encoder and one CNN-based Decoder to weaken the accu-

https://doi.org/10.1016/j.neucom.2020.11.020

racy differences among individuals on activity recognition. Experimental results show that the average accuracy of the proposed deep learning method is higher than 95% for the recognition of sixteen activities. The second paper, *Multi-frequency and Multi-domain Human Activity Recognition Based on SFCW Radar Using Deep Learning*, by *Jia et al.*, proposed a specific deep learning network consisting of multiple parallel deep convolutional neural networks (DCNNs) and a sparse autoencoder for human activity recognition with stepped frequency continues wave (SFCW) radar signals which provide two types of features, i.e., spectrograms and range maps. Specifically, each DCNN was to extract the detailed micro-Doppler features from a spectrogram, while sparse autoencoder learnt prime range distribution features by compressing each range map to reduce complexity and improve robustness.

3. There are six papers focusing on camera sensors or vision-based HAR. The first paper, *A Fast Human Action Recognition Network Based on Spatio-Temporal Features*, by *Xu et al.*, proposed an end-to-end fast network for human action recognition. It combines spatial and temporal features into fusion features. Besides, a CNN with OFF network was proposed to obtain abundant features based on the fused optical flow features. With only RGB inputs, this method can achieve state-of-the-art performance in four datasets. The second paper, *Conflux LSTMs Network: A Novel Approach for Multi-View Action Recognition*, by *Ullah et al.*, presented a conflux long short-term memory (LSTM) network to recognize actions from multi-view cameras. Firstly, deep features from a sequence of frames were extracted by using a pre-trained VGG19 for each view. Then, the extracted features were forwarded to the conflux to learn the view self-reliant patterns. After that, the inter-view correlations were computed for learning the view inter-reliant patterns. Finally, the flatten layers followed by SoftMax classifier were employed for action recognition. The third paper, *Adaptive Multi-View Graph Convolutional Networks for Skeleton-based Action Recognition*, by *Liu et al.*, proposed an adaptive multi-view graph convolutional networks (AMV-GCNs) for skeleton based action recognition. It firstly constructed a novel skeleton graph with two kinds of graph nodes which are defined to model the spatial configuration and temporal dynamics respectively. Then, the generated graphs were fed into the AMV-GCNs. Furthermore, multiple GCNs based streams were utilized to learn action information from multiple viewpoints. Finally, the classification scores from multiple streams are fused to provide the recognition results. The fourth paper, *Action Anticipation for Collaborative Environments: The Impact of Contextual Information and Uncertainty-Based Prediction*, by *Santos et al.*, presented a LSTM model for the prediction of action anticipation with contextual information. It also used the uncertainty about each prediction as an online decision-making criterion for action anticipation. The uncertainty was modeled as a stochastic process applied to a time-based neural network architecture, which improves the conventional class-likelihood criterion. The fifth paper, *DB-LSTM: Densely-connected Bi-directional LSTM for Human Action Recognition*, by *He et al.*, proposed a novel deep learning model to capture the spatial and temporal patterns from videos for human action recognition. Firstly, a sample representation learner was designed to extract the video-level temporal feature. To boost the effectiveness and robustness of modeling long-range action recognition, a Densely-connected Bi-directional LSTM (DB-LSTM) network was developed. Two modalities from appearance and motion are integrated with a fusion module to further improve the performance. The last paper, *Normal Graph: Spatial Temporal Graph Convolutional Networks based Prediction Network for Skeleton based Video Anomaly Detection*, by *Luo et al.*, proposed a spatial temporal graph convolutional network for abnormal behavior detection by analyzing graph connections of human joints in video. In other words, it built a normal graph describing graph connection of joints in normal data, where joints of abnormal events will be outliers of this graph. Most of vision-based solutions consider spatial and temporal features from vision. The powerful GCN has been widely adopted for HAR.
4. The rest of six papers handle some special human activities, such as two-person interaction, human parsing and pose, 3D hand gesture, human-object interaction, facial expression and eye-movement. The first paper, *Knowledge Embedded GCN for Skeleton-based Two-person Interaction Recognition*, by *Li et al.*, dealed with two-person interaction recognition by proposing a knowledge embedded graph convolutional network. Two graphs were designed by exploiting the knowledge for two-person interaction recognition. Specifically, a knowledge-given graph was constructed to build the direct connection between two persons. Besides, a knowledge-learned graph was developed to build the adaptive correlations. Moreover, a multi-level scheme was proposed to model the joint-level and part-level information simultaneously, which is able to further enhance the performance for interaction recognition. The second paper, *A Novel Framework for Simultaneous Human Parsing and Pose Estimation*, by *Xu et al.*, proposed a Separation-and-UnioN Network (SUNNet) for simultaneous human parsing and pose estimation. Task-specific features and common features were extracted for both tasks. An initial prediction was achieved with these features. Then, they refined the initial prediction by explicitly leveraging the features from parallel task to predict the kernels' receptive fields in a convolutional neural network. Extensive experiments show the effectiveness of the SUNNet model for human body configuration analysis. The third paper, *Deformation Representation based Convolutional Mesh Autoencoder for 3D Hand Generation*, by *Zheng et al.*, proposed a convolutional mesh autoencoder for 3D hand generation with complex gestures. It firstly built a large-scale high-quality hand mesh dataset based on MANO with a novel mesh deformation method. Then, a VAE was trained based on this dataset, which is able to obtain the low-dimensional representation of hand meshes for hand generation. The fourth paper, *GID-Net: Detecting Human-object Interaction with Global and Instance Dependency*, by *Yang et al.*, handled the interesting problem of human-object interactions. It proposed a two-stage trainable reasoning mechanism, named GID block, which breaks through the local neighborhoods and captures long-range dependency of pixels both in global-level and instance-level for capturing interactions between instances. Moreover, a multi-steam network, named GID-Net, that consists of a human branch, an object branch and an interaction branch, was proposed for human-object interaction recognition. The fifth paper, *Multi-attention based Deep Neural Network with hybrid features for Dynamic Sequential Facial Expression Recognition*, by *Sun et al.*, explored the recognition of a special type of human activity, i.e., facial expression. Specifically, it firstly proposed an attention shallow model to describe the action units of the facial action coding system. Then, an attention deep model was designed to extract deep features on sequence facial images. Finally, a multi-attention shallow

and deep model was developed to achieve dynamic sequence facial expression recognition. The last paper, *Deep-learning-based reading eye-movement analysis for aiding biometric recognition*, by *Wang et al.*, considers a special type of HAR, i.e., reading eye-movement recognition, by proposing a deep learning model. The model takes the text, fixation, and text-based linguistic feature sequences as inputs and identifies a human subject by measuring the similarity distance between the predicted fixation sequence and the actual one.

The guest editors would like to take this opportunity to thank all the authors for their valuable contributions to this special issue, and all of the reviewers, who provided constructive suggestions and thorough reviews during the paper selection process. The encouragements and support from the Editor-in-Chief and the editorial staff of the Elsevier Neurocomputing Journal throughout the preparation of this issue are also greatly appreciated.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.
[2] O.D. Lara, M.A. Labrador, A survey on human activity recognition using wearable sensors, IEEE Commun. Surv. Tutorials 15 (3) (2012) 1192–1209.
[3] Z. Chen, Q. Zhu, Y.C. Soh, L.e. Zhang, Robust human activity recognition using smartphone sensors via CT-PCA and online SVM, IEEE Trans. Ind. Inf. 13 (6) (2017) 3070–3080.
[4] J.B. Yang, M.N. Nguyen, P.P. San, X.-L. Li, P.S. Krishnaswamy, Deep convolutional neural networks on multichannel time series for human activity recognition, IJCAI (2015).
[5] Z. Chen, C. Jiang, S. Xiang, J. Ding, W.u. Min, X.-L. Li, Smartphone sensor based human activity recognition using feature fusion and maximum full a posteriori, IEEE Trans. Instrum. Meas. (2020).
[6] H. Cao, M.N. Nguyen, C.W. Phua, S.P. Krishnaswamy, X.-L. Li, An integrated framework for human activity classification, ACM International Conference on ubiquitous computing (2012).

Xiaoli Li is currently a principal scientist and department head (Machine Intellection) at the Institute for Infocomm Research, A*STAR, Singapore. He also holds adjunct professor position at Nanyang Technological University. His research interests include data mining, machine learning, AI, and bioinformatics. He has been serving as area chairs/senior PC members/workshop chairs/session chairs in leading data mining and AI related conferences (including IJCAI, AAAI, KDD, ICDM, SDM, PKDD/ECML, WWW, ACL and CIKM). Xiaoli has published more than 200 high quality papers, and won 6 best paper awards and 2 international benchmark competitions.

Peilin Zhao is currently a Principal Researcher at Tencent AI Lab, China. Previously, he has worked at Rutgers University, Institute for Infocomm Research (I2R), Ant Group. His research interests include: Online Learning, Recommendation System, Automatic Machine Learning, Deep Graph Learning, and Reinforcement Learning etc. He has published over 100 papers in top venues, including JMLR, ICML, KDD, etc. He has been invited as a PC member, reviewer or editor for many international conferences and journals, such as ICML, JMLR, etc. He received his Ph.D. degree from Nanyang Technological University, Singapore.

Min Wu is currently a senior scientist in Data Analytics Department, Institute for Infocomm Research, Agency for Science, Technology and Research (A*STAR), Singapore. He received his Ph.D. degree in Computer Science from Nanyang Technological University (NTU), Singapore, in 2011 and B.S. degree in Computer Science from University of Science and Technology of China (USTC) in 2006. He received the best paper awards in InCoB 2016 and DASFAA 2015. He also won the IJCAI competition on repeated buyers prediction in 2015. His current research interests include machine learning, data mining and bioinformatics.

Zhenghua Chen received the B.Eng. degree in mechatronics engineering from University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2011, and Ph.D. degree in electrical and electronic engineering from Nanyang Technological University (NTU), Singapore, in 2017. He has been working at NTU as a research fellow. Currently, he is a scientist at Institute for Infocomm Research, Agency for Science, Technology and Research (A*STAR), Singapore. His research interests include sensory data analytics, machine learning, deep learning, transfer learning and related applications.

Le Zhang received the B.E. degree in communication engineering from the University of Electronic Science and Technology of China, Chengdu, China, and the M.Sc. degree in signal processing and Ph.D. degree in machine learning and computer vision from the Nanyang Technological University, Singapore, in 2011, 2012, and 2016, respectively. He has been working as a Researcher with the Advanced Digital Sciences Center, Singapore, the Singapore-based research center of the University of Illinois at Urbana-Champaign, from 2016 to 2018. Currently, he is a scientist at Institute for Infocomm Research (I2R) A*STAR, Singapore. His current research interests include machine learning and computer vision.

Xiaoli Li [a,*]
Peilin Zhao [b]
Min Wu [a]
Zhenghua Chen [a]
Le Zhang [a]
[a] *Institute for Infocomm Research (I2R) A*STAR, Singapore*

[b] *Tencent AI Lab, China*
* Corresponding author.
E-mail addresses: xlli@i2r.a-star.edu.sg (X. Li), wumin@i2r.a-star.edu.sg (M. Wu), chen0832@e.ntu.edu.sg (Z. Chen)

Received 23 October 2020
Accepted 8 November 2020

Available online 28 November 2020