Human Activity Recognition with Inertial Sensors using a Deep Learning Approach

Tahmina Zebin
School of EEE
University of Manchester, UK
tahmina.zebin@manchester.ac.uk

Patricia J. Scully
Photon Science Institute
University of Manchester, UK
patricia.scully@manchester.ac.uk

Krikor B. Ozanyan
School of EEE
University of Manchester, UK
k.ozanyan@manchester.ac.uk

***Abstract*—Our focus in this research is on the use of deep learning approaches for human activity recognition (HAR) scenario, in which inputs are multichannel time series signals acquired from a set of body-worn inertial sensors and outputs are predefined human activities. Here, we present a feature learning method that deploys convolutional neural networks (CNN) to automate feature learning from the raw inputs in a systematic way. The influence of various important hyper-parameters such as number of convolutional layers and kernel size on the performance of CNN was monitored. Experimental results indicate that CNNs achieved significant speed-up in computing and deciding the final class and marginal improvement in overall classification accuracy compared to the baseline models such as Support Vector Machines and Multi-layer perceptron networks.**

Keywords— Feature Extraction, Signal Processing, Convolution, Human activity recognition (HAR); Convolutional Neural Networks (CNN)

I. INTRODUCTION

The demands for understanding human activities has grown enormously in recent years for ubiquitous computing, human computer interaction, and healthcare domains such as elder care support, rehabilitation assistance, and cognitive disorder recognition systems [3-5]. To achieve high accuracy in activity recognition with low computational cost is a key challenge. To deal with this challenge, the HAR community is beginning to adopt deep learning to substitute for well-established analysis techniques that rely on hand-crafted feature extraction. Previously, hand-crafted features were mostly limited to statistical features such as mean and variance in time domain, and fast Fourier transform coefficients in the frequency domain. The use of these features required the application prior specific knowledge about the signals, in order to capture essential characteristics between different activities [6]. In contrast, learning based on deep neural networks can automatically extract representative features without any prior knowledge about the signals. Here, instead of exploring hand-crafted features from time-series sensor signals, we aim to show that signal sequences of accelerometers and gyroscopes can be processed by Deep Convolutional Neural Networks (CNN) to automatically learn from the input the optimal features for the activity recognition task.

Tahmina Zebin would like to thank the Presidents Doctoral Scholar award scheme, University of Manchester for funding her PhD studies.

II. RELATED WORK

CNNs, which comprise of one or more convolutional and pooling layers followed by one or more fully-connected layers, have gained popularity due to their ability to learn suitable representations from images or speeches, capturing local dependency and slight-distortion invariance [7]. CNN has recently been applied to the problem of activity recognition in number of research papers [2, 4, 8]. However, most of the research to date has been conducted using datasets that contain either accelerometer data only or data from a single sensor (e.g. smartphone sensor).

In our case, we have collected data from five different sensor locations on the lower body in order to classify activities more accurately. Reference [9] contains a description of our original data collection system based on MPU-9150 sensors which has been used to obtain the reported results. By combining signals from multiple sensors, it was possible to obtain more information compared to the mobile or single sensor [10,11] scenario. Hence, we can expect better accuracy in the activity recognition. In this context we considered the HAR task as a classification problem where time-series data from inertial sensor (consisting of accelerometers and gyroscopes) has been used to extract relevant and discriminative features from them, and finally, to recognize activities by using a classifier.

III. CONVOLUTIONAL NEURAL NETWORK ARCHITECHTURE

The overall CNN structure used in this paper is shown in Fig. 1. To develop more accurate activity recognition algorithm, we adopted techniques such as Pooling, Rectified linear Unit (ReLU) activation function and soft-max classifier popularly used in several deep learning tasks published in [2, 12]. Convolutional neural networks perform convolution instead of matrix multiplication (as with fully-connected neural networks).

For the network shown in Fig. 1, x_t^0 is the sensor data input vector defined as:

$$x_t^0=\left[x_1, x_2, \cdots x_N\right] \quad (1)$$

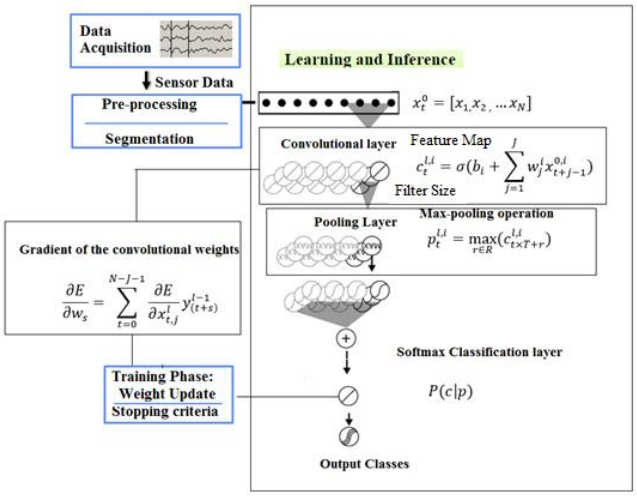

Fig. 1. Deep Convolutional Neural Network Architecture (redrawn from[2])

where N is the number of sensor readings per window after segmentation.

The output of the convolutional layer is

$$c_t^{l,i} = \sigma\left(b_i + \sum_{j=1}^{J} w_j^i x_{t+j-1}^{0,i}\right) \quad (2)$$

where 'l' is the layer index, σ is the activation function, b_i is the bias term for ith feature map, J is the kernel or filter size and w_j^i is the weight for i^{th} feature map and j^{th} filter index. The pooling layer derives a summary statistic of nearby outputs derived from $c_t^{l,i}$.The pooling operation used in this paper, max-pooling, is characterized by outputting the maximum value among a set of nearby inputs, given by

$$f_t^{l,i} = \max_{r \in R} \left(c_{t \times T + r}^{l,i}\right) \quad (3)$$

where R is the pooling size, and T is the pooling stride. A simple softmax classifier is used to recognize activities, which is placed at the final layer. Features from the stacked convolutional and pooling layers are aligned to form feature vectors:

$$f^l = [f_1, f_2, \cdots f_K] \quad (4)$$

where, K is the number of units in the last pooling layer, acting as input to the soft-max classifier:

$$P(c|f) = \text{argmax}_{c \in C}\left(\frac{\exp\left(f^{L-1} w^L + b^L\right)}{\sum_{n=1}^{N_c} \exp\left(f^{L-1} w_n\right)}\right) \quad (5)$$

Here c is the activity class, L is the last layer index, and N_c is the total number of activity classes.

Forward propagation is performed using (2) to (5), which yields the error values of the network. Backpropagation to adjust weights is done by computing the gradient of the convolutional weights:

$$\frac{\partial E}{\partial w_s} = \sum_{t=0}^{N-J-1} \frac{\partial E}{\partial x_{t,j}^l} y_{(t+s)}^{l-1} \quad (6)$$

where E is the error or cost function and $y_{(t+s)}^{l-1} = \sigma\left(x_{(t+s)}^{l-1}\right) + b^{l-1}$ is the nonlinear mapping function. The forward and back propagation procedure is repeated until a stopping criterion is met, e.g., if a maximum number of epochs is reached, among others. Further details on the mathematical derivation can be found in [2].

IV. Experimental Results

For modelling and evaluating physical activity, we collected data from 12 healthy volunteers (age[y]: 24.6±5.2; height [m]: 1.63 ± 0.6, weight [kg]: 64.7 ± 7.1) while undertaking six common daily life activities such as, (1)walking, (2)walking upstairs, (3) downstairs, (4)sitting, (5)standing (6) lying down. Functions and features from the Machine Learning toolbox in MATLAB have been used to implement the CNN network. For processing sensor data from a specific location (pelvis, thigh or shank), we performed a 6-channel 1D convolution on the input, i.e. 3- axis acceleration and 3- axis gyro. The raw accelerometer and gyroscope xyz signals were pre-processed and segmented into 128 values for every activity sample.

Generally, sensor-based activity recognition methods are evaluated in two aspects: recognition accuracy and computational cost [13,14]. To improve the accuracy, in our previous work [1] we extracted effective statistical and frequency domain features (hand-crafted) from 3-axis data from inertial sensors and explored different classifiers including Support Vector Machine(SVM) and Multilayer Perceptron (MLP). The results of the used CNN method and two baseline methods SVM and MLP for the collected data are compared in Table I.

The comparison shows an improvement in accuracy by using deep learning approaches. Additionally, no overtraining is observed without any validation sequence and the training stops when performance on the training set no longer improves.

TABLE I. Quantitative performance Comparison of SVM, MLP based Neural Network and Deep CNN

Learning Method	Classification Accuracy (%)	Computational Load(ms)
SVM [1]	96.4%	10.6
MLP [1]	91.7%	6.7
Deep CNN	97.01%	3.53

TABLE II. Hyper-parameters Range for CNN Implementation

CNN hyper-parameter	Value
Number of convolution layers	3
Learning rate	0.01
Number of feature maps	10 to 100
Filter (kernel) size	1x3 to 1x15
Pooling size	1x2 to 1x15

We also studied the influence of various important hyper-parameters (kernel sizes, network size) of CNN on the overall performance, and showed for almost all the cases that the recognition rates are superior to those of state-of-the-art methods. Experiments show that increasing the number of convolutional layers increases computational load, but the complexity of the derived features decreases with every additional layer. A gradual increase in the performance was observed after adding an extra convolution layer on validation data. On the other hand, on test data there was a noticeable improvement in performance when third layer was added. Tuning the filter and pooling sizes revealed that wider filter (i.e. kernel) size and lower pooling size setting improves the recognition performance of the CNN as well. Table II displays some important CNN hyper-parameters, chosen for yielding the best score on the validation set during the training of the network. In Table II, the number of feature maps can be increased up to 100, depending on the complexity of the activity, and the filter and pooling sizes can be increased up to 1x15.

V. Conclusion and Future Work

Time-series data have inherent local dependency characteristics and daily life physical activities tend to be hierarchical, as well as translation-invariant in nature. Our experimental results demonstrate how these characteristics can be exploited by CNNs.

We used CNN architecture to automate feature learning from the raw inputs for human activity recognition task. We adopted techniques such as pooling, ReLU and soft-max classifier to avoid overfitting of neural network due to small size of the training data. Experimental results indicated that CNNs manifest marginal improvement compared to SVM and MLP in terms of classification accuracy, but it achieved a considerable speed-up in terms of computational load, as shown in Table I. The influence of various important hyper-parameters such as kernel size and number of convolution layers on the performance was also studied, to allow to proceed with the parameter values that displayed best score during the training of the CNN.

Recent research demonstrates increasing interest in classification of more challenging or complex actions and realistic scenarios [15]. Therefore, we plan to verify the effectiveness of our approach by testing it on other challenging activities. This will allow us to confirm the benefit of the deep learning based feature extraction process. Additionally, to take advantage of the availability of multiple sensors, CNN with long short-term memory [8] sequence classifier, involving retro-propagated error, will be implemented and studied in the immediate future.

References

[1] T. Zebin, P. J. Scully, and K. B. Ozanyan, "Inertial Sensor Based Modelling of Human Activity Classes: Feature Extraction and Multi-sensor Data Fusion using Machine Learning Algorithms," in *EAI International Conference on Wearables in Healthcare, budapest, hungary(unpublished)*, ed, 2016.

[2] C. A. Ronao and S. Cho, "Deep Convolutional Neural Networks for Human Activity Recognition with Smartphone Sensors," in *Proceedings from ICONIP 2015, November 9-12, 2015, Part IV*, S. Arik, T. Huang, K. W. Lai, and Q. Liu, Eds., ed: Springer International Publishing, 2015, pp. 46-53.

[3] J. A. Cantoral-Ceballos, N. Nurgiyatna, P. Wright, J. Vaughan, C. Brown-Wilson, P. J. Scully, *et al.*, "Intelligent Carpet System, Based on Photonic Guided-Path Tomography, for Gait and Balance Monitoring in Home Environments," *IEEE Sensors Journal*, vol. 15, pp. 279-289, 2015.

[4] J. B. Yang, M. N. Nguyen, P. P. San, X. L. Li, and S. Krishnaswamy, "Deep convolutional neural networks on multichannel time series for human activity recognition," presented at the Proceedings of the 24th International Conference on Artificial Intelligence, Buenos Aires, Argentina, 2015.

[5] A. Mannini and A. M. Sabatini, "Machine Learning Methods for Classifying Human Physical Activity from On-Body Accelerometers," *Sensors*, vol. 10, pp. 1154-1175, Feb 2010.

[6] S. Ha, J. M. Yun, and S. Choi, "Multi-modal Convolutional Neural Networks for Activity Recognition," in *2015 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2015, pp. 3017-3022.

[7] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436-444, 2015.

[8] F. Ordóñez and D. Roggen, "Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition," *Sensors*, vol. 16, p. 115, 2016.

[9] T. Zebin, P. J. Scully, and K. B. Ozanyan, "Inertial sensing for gait analysis and the scope for sensor fusion," in *SENSORS, 2015 IEEE*, 2015, pp. 1-4.

[10] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine," presented at the Proceedings of the 4th international conference on Ambient Assisted Living and Home Care, Vitoria-Gasteiz, Spain, 2012.

[11] M. Zeng, L. T. Nguyen, B. Yu, O. J. Mengshoel, J. Zhu, P. Wu, *et al.*, "Convolutional Neural Networks for human activity recognition using mobile sensors," in *Mobile Computing, Applications and Services (MobiCASE), 2014 6th International Conference on*, 2014, pp. 197-205.

[12] R. Yeh, M. Hasegawa-Johnson, and M. N. Do, "Stable and symmetric filter convolutional neural network," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 2652-2656.

[13] L. Zhang, X. Wu, and D. Luo, "Recognizing Human Activities from Raw Accelerometer Data Using Deep Neural Networks," in *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, 2015, pp. 865-870.

[14] P. Casale, O. Pujol, and P. Radeva, "Human Activity Recognition from Accelerometer Data Using a Wearable Device," in *Proceeding of 5th Iberian Conference on Pattern Recognition and Image Analysis: IbPRIA 2011*, J. Vitrià, J. M. Sanches, and M. Hernández, Eds., ed: Springer Berlin Heidelberg, 2011, pp. 289-296.

[15] O. D. Lara and M. A. Labrador, "A Survey on Human Activity Recognition using Wearable Sensors," *IEEE Communications Surveys & Tutorials*, vol. 15, pp. 1192-1209, 2013.