TAF

Journal of ICT, Design, Engineering and Technological Science

JITDETS

VOL. 1, NO. 1, pp. 1-5, 2017
DOI: 10.20474/JITDETS-1.1.1

ORIGINAL CONTRIBUTION

Human Activity Recognition Using a Single Wrist IMU Sensor via Deep Learning Convolutional and Recurrent Neural Nets

E. Valarezo [1], P. Rivera [2], J. M. Park [3], G. Gi [4], T. Y. Kim [5], M. A. Al-Antari [6], M. Al-Masni [7], T.-S. Kim [8*]

[1, 2, 3, 4, 5, 6, 7, 8] Department of Biomedical Engineering, College of Electronics and Information, Kyung Hee University, Yongin, South Korea

[1] Escuela Superior Politecnica del Litoral, ESPOL, FIEC, Campus Gustavo Galindo, Guayaquil, Ecuador

Abstract— **Recognizing and recording human activities using a smart sensor device is an essential technology for smart living. The recorded activities (i.e., life logs) could be used as valuable information to support smart life, lifecare, and healthcare services. For sensing human activities, smart sensors are required and most smart devices such as smart phones, smart bands, and smart watches incorporate Inertial Measurement Units (IMUs) which could be utilized for this purpose. However, implementing a robust Human Activity Recognition (HAR) system with high recognition accuracy using only a single sensor (i.e., no multiple sensors) is still a technical challenge. In this paper, we propose novel deep learning-based HAR systems with a single wrist IMU sensor. We used time series activity data from only one IMU sensor at a wrist to build two deep learning algorithm-based HAR systems: one is based on Convolutional Neural Nets (CNN) and the other Recurrent Neural Nets (RNN). Our two HAR systems are evaluated by 5-fold cross-validation tests to compare the performance of both systems. Five primary daily activities including standing, walking, running, walking downstairs, and walking upstairs were recognized. Our results show that the CNN-based HAR system achieved an average accuracy of 95.43% and the RNN-based HAR system an accuracy of 96.95%**.

Index Terms— **Human Activities, Inertial Measurement Units (IMUs), Convolutional Neural Nets (CNN), Recurrent Neural Nets (RNN), HAR System**

Received: 01 April 2017; **Accepted:** 02 May 2017; **Published:** 18 June 2017

I. INTRODUCTION

HAR opens new opportunities to personalized life care and healthcare services, since daily activity logs (i.e., lifelogs) of a person can provide information of personal behaviors and patterns. A HAR system requires two key components: smart sensors and pattern recognition techniques. Recently, IMUs are readily available inside smart phones, smart bands, and smart watches. With IMUs, human movements or activities can be translated into information of acceleration via accelerometer and angles via gyroscope. Recent works utilizing IMUs for HAR can be found in [1], [2], [3]. In these studies, multiple IMUs, positioned in different body parts (especially waist, lower limbs, and upper limbs), are used for HAR. Although deploying multiple sensors increases the accuracy of HAR, this approach is impractical and inconvenient to users due to multiple sensor deployment. For making HAR practical, a single IMU is preferred in some common body areas (i.e., wrist, waist, or ankle). This constitutes a need to develop a HAR system using a single IMU embedded in a smart watch or smart band to be accepted by general users.

In [4], performed HAR with a pair of wrist sensors performing an evaluation of different classification methods for activity recognition and fall detection. They made a comparison between performances of recognition using information from the right and left wrist. They applied traditional classifiers such as decision tree and Support Vector Machines over some selected features as input. They concluded that the left wrist sensor is more informative than the dominant right one: accuracy of 72% from the left wrist vs. accuracy of 68% from the right. Till now, most studies use multiple sensors [5]. Regarding, pattern recognition techniques, most studies utilize traditional classifiers with handcrafted features [6, 7]. There is still a lack of robust system to discern similar human activities.

This work focuses on HAR using a single IMU sensor at one wrist via deep learning CNN and RNN approaches. We have developed two HAR systems: one is using CNN and the other RNN. With our developed HAR systems, five primary daily activities such as walking, standing, running, walking upstairs and walking downstairs are recognized in this study. In addition, we evaluate our systems with continuous time series data to test the feasibility of real-time continuous HAR. The structure of this paper is organized as follows: Section 2 describes some related works. Section 3 our proposed CNN- and RNN-based HAR systems. In section 4, our experimental results are presented. Section 5 corresponds to discussion section. Finally, we present the conclusion in Section 6.

II. RELATED WORKS

Deep learning is getting a major attention lately because of its applicability to various machine learning problems. In fact, it becomes a new trend in the fields of pattern recognition and machine learning. Deep learn-

[*] Corresponding author: T.-S. Kim

[†] Email: tskim@khu.ac.kr

ing approaches are proven to overcome traditional classifiers in many applications. Also, deep learning performs an unsupervised feature extraction, saving time and computation resources used in feature extraction and selection. In the work of [7], they compared traditional machine learning techniques such as Naïve Bayes, K-Nearest Neighbors (KNN), Decision Tree, and Support Vector Machines (SVMs) to CNN for HAR with two IMUs at both wrists. Their results showed that CNN outperformed in activity recognition the traditional machine learning algorithms. In the work of [8], they tested a CNN-based approach in two different data sets: one compound of multiple body-worn sensors including accelerometer and gyroscopes located in some parts of the body such as waist and upper limbs for HAR and the other database for hand gestures. Again, CNN was tested against some conventional approaches of SVM, KNN, and Deep Belief Network (DBN) and CNN achieved better HAR.

There are some works in which recurrent neural nets are used in the HAR systems. We can mention [5] in their work report the performance using RNN. They proposed a generic deep framework for HAR based on Long Short Term Memory (LSTM) using a full set of body-worn sensors. Their results showed RNN produced better recognition than CNN and better classification decisions between similar activities such as open and close door. In the work of Shin et al. [9], they developed two dynamic hand gesture recognition techniques: one using QVGA images from the Cambridge-Gesture database [10], and the other using accelerometer signals from the smart watch gestures database [11]. Both models were implemented with simple RNN.

Most deep learning approaches show that they can outperform traditional classifiers for better HAR. However, most of these studies employed multiple sensors till now. Therefore, it remains a challenge to develop a robust HAR system using a single sensor via deep learning approaches.

III. METHODS

This section presents our CNN- and RNN-based HAR systems. As aforementioned, only one IMU including one tri-axial accelerometer and one tri-axial gyroscope was attached at dominant wrist.

A. Our Proposed CNN-Based HAR System

CNN is a feed forward neural network that involves the use of a convolutional operation in at least one of their layers [12]. In general, a basic CNN architecture involves a combination of different layers like convolutional, subsampling, and fully connected. In the convolutional layer, a mathematical operation (i.e., convolution) applies a set of local filters (or kernels) to obtain the most representative features. After the convolution operation, a bias is added to the result and an activation function generates the output for this layer. Subsampling or pooling operations reduce data size by reducing the dimension of the input using an average or maxing filter. The fully connected layer is applied at the end, combining all features' maps obtained by the previous steps and using it as input for a classification layer. This hierarchical organization generates classification results, where the lower layers obtain local dependencies of the input and the higher layers obtain high level representation of the data. Input for the proposed CNN-based HAR system is one-dimensional blockwise segments (i.e., epochs) extracted from accelerometer and gyroscope signals. Our proposed architecture captures epochs from multi-channel time series: The CNN architecture was changed to work with multichannel time series as input and the architecture of convolutional kernels and pooling kernels were modified to work in one dimension. Our proposed architecture for the CNN-based HAR uses five convolutional layers. The corresponding filter sizes include 1x13, 1x13, 1x11, 1x3, and 1x3. The pooling layers are used after the first and second convolutional layers, both are of the same dimension of 1x2. The activation function used in each layer is Rectified Linear Unit (ReLU). At the end, a fully connected (F1) and a softmax (F2) layers are added as shown in Fig. 1. Once the architecture was specified, the networks parameters for training were defined: mini-batch size of 264, learning rate of 6.5e-6, and dropout of 0.5.

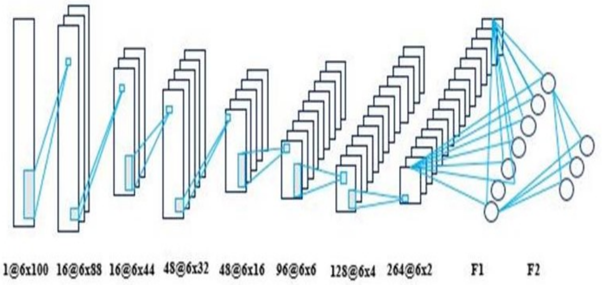

Fig. 1. Architecture of our proposed CNN-based HAR system

B. Our Proposed RNN-Based HAR System

Considering temporal variations in the natural movements of humans, we utilized RNN to encode these temporal changes. This deep learning approach has recurrent connections between hidden units: those connections generate a temporal memory, in which recurrent nets store the value of previous stage. Giving the opportunity to decide not only depending on the input (as in the case of CNN or other neural network), also considering the previous stage. In order to avoid the vanishing gradient problem (i.e., turning slow learning process by vanishing the magnitude of gradient in time) or exploding gradient, we have used LSTM proposed by Hochreiter and Schmidhuber [13]. LSTM produces internal paths where the gradient can flow for long durations: those paths are the structure of the gates to LSTM. This flavor of LSTM has the following gates: input gates determine which value is an update; forget gates determine what information is throwing away; output gates control what information is going to be the output of the cell. Fig. 2 shows a single LSTM cell with its internal connections with Equations (1)~(5) describing operations.

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \quad (1)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \quad (2)$$

$$c_t = f_t c_{t-1} + i_t tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (3)$$

$$o_t = \sigma(W_{xo}x_t) + W_{ho}h_{t-i} + b_c \quad (4)$$

$$h_t = o_t tanh(c_t) \quad (5)$$

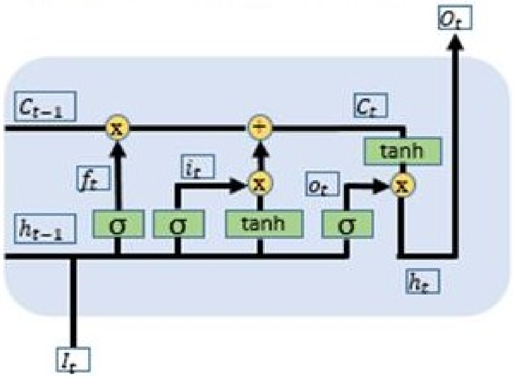

Fig. 2. LSTM structure

Backpropagation Through Time (BPTT) was used to train the model, as described [14]. For tuning network parameters, an algorithm for the first-order gradient-based optimization Adam Optimizer was used as described in [15]. In Fig. 3, our RNN architecture is shown with 100 LSTM cells, reflecting the length of data in our epochs (i.e., 2 seconds' data with 50 Hz sampling frequency).

The output comes after the last LSTM cell. We used the learning rate of 1.5xe-3 and 105 hidden units as the network structure.

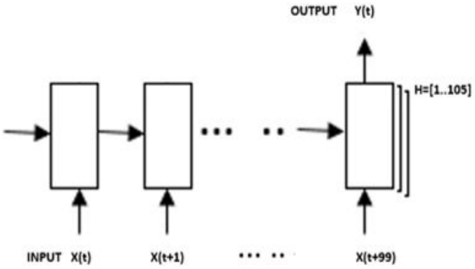

Fig. 3. Architecture of our proposed RNN-based HAR system

IV. EXPERIMENT AND RESULTS

We worked with a public database. Details and general information are given in Section A. Training and testing epochs were obtained by segmenting accelerometer and gyroscope time series signals. Two methods of validation were used to test the proposed algorithms. First validation was done by splitting the data 80% for training and 20% for testing avoiding overlap between them and running five-fold tests. The second validation was done using the time-series continuous data sets from one subject. More details are described in the following section.

A. Activity Data Set

We used the Physical Activity Monitoring for Aging People (PAMAP2) database [16], [17]. This database includes activity information of 9 subjects (8 males and 1 female) using three IMUs located at chest, wrist, and ankle with the sampling frequency of 100Hz. In this work, we used data from the 16g-scale accelerometer to avoid saturation problems in activities. There are 12 lifestyle activities including sports activities such as running and Nordic walking; household activities such as ironing, vacuum cleaning, walking, etc., and 6 optional activities such as watching TV, computer work, car driving, folding laundry, house cleaning, and playing soccer. This data set has a total of 8 hours of data collection.

In this study, we used time series data from Subjects 2, 5, 6, 7, and 8 from the PAMAP2 database to build training and testing data sets. We selected five daily activities, which are commonly used in human activity recognition: namely, standing, walking, running, walking upstairs, and walking downstairs. We restricted the information to only a single IMU (each of tri-axial accelerometer and tri-axial gyroscope) at one wrist. Time series signals were down sampled to 50 Hz following the suggestions mentioned in [18] to remove the gravity effect using a high-pass Butterworth filter with cutoff frequency of 0.25 Hz.

We took the following actions to obtain training and testing data epochs: First, a sliding window was used to generate epochs with duration of 2 seconds. We used 50% overlap in epochs. Then put all channels together and made a matrix dimension of 6 by 100. Each matrix corresponds to a single epoch to train or test the systems. From the total number of epochs, we split 80% for training and 20% for testing. Table I shows the number of epochs for each activity.

TABLE I
TOTAL NUMBER OF TRAINING AND TESTING DATA SETS FROM EPOCH DATA SETS FROM THE PAMAP2 DATABASE

Type of Activity	Input Information	
	Training	Testing
Standing	488	122
Walking	1176	294
Running	560	140
Walking Upstairs	512	128
Walking Downstairs	400	100

B. HAR Results with Epoch Activity Data

The recognition results of this test are presented in Tables II and III. The average accuracy of the CNN-based HAR system is 95.43% with a standard deviation of 0.02. There is some confusion found between walking upstairs and walking downstairs due to the similarity of the activities.

Table III presents the results of the RNN-based HAR system. Average accuracy of 96.95% and the standard deviation of 0.76. In this case, the system has produced similar performance for standing, walking, and running activities as the CNN-based system. Nevertheless, in the remaining activities, RNN outperforms convolutional nets.

TABLE II
CONFUSION MATRIX OF HAR VIA OUR CNN-BASED HAR SYSTEM

Type of Activity (%)	Model Classification				
	Stand	Walk	Run	Walking Upstairs	Walking Downstairs
Standing	100	0.00	0.00	0.00	0.00
Walking	0.00	97.29	0.00	1.69	1.02
Running	0.00	0.71	99.29	0.00	0.00
Walking Upstairs	0.00	8.59	0.00	89.84	2.32
Walking Downstairs	0.00	2.97	0.00	10.89	86.14

TABLE III
CONFUSION MATRIX OF HAR VIA OUR RNN-BASED HAR SYSTEM

Type of Activity (%)	Model Classification				
	Stand	Walk	Run	Walking Upstairs	Walking Downstairs
Standing	100	0.00	0.00	0.00	0.00
Walking	0.00	97.08	0.00	2.37	0.54
Running	0.00	0.41	99.43	0.00	0.00
Walking Upstairs	0.00	4.18	0.15	93.33	2.32
Walking Downstairs	0.20	2.57	0.20	3.55	93.46

ISSN: xxxx-xxxx
DOI: 10.20474/JITDETS-1.1.1

C. HAR Results with Continuous Activity Data

This test was done using continuous epochs from Subject 7 got following the same procedure explained earlier in this section. The collected epochs were inputs to the algorithms in the same order as they were obtained. Label for each epoch was defined by the mode of the ground truth labels.

The results of our proposed CNN- and RNN-based HAR systems are shown in Fig. 4. The recognition accuracy was 79.97% by the CNN-based system and 88.96% by the RNN-based system.

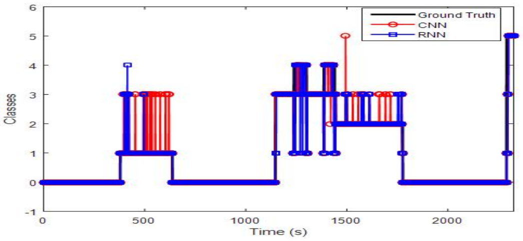

Fig. 4. HAR results from the CNN-based and RNN-based HAR systems: The ground-truth activities (black), by the CNN-based HAR system (red), and by the RNN-based HAR system (blue). Class 1 represents stand Class 2 walk, Class 3 up-stairs, Class 4 down-stairs and, Class 5 Run

V. DISCUSSION

Based on the results with the continuous activity data, RNN outperforms 9% better than 1-D CNN. We consider that this is due to the capability of RNN better handling the time sequential information of alike activities such as walking upstairs and downstairs: the recall values (main diagonal values) for alike activities are 89.84% for walking upstairs and 86.14% for walking downstairs by the CNN-based HAR system whereas 93.33% for walking upstairs and 93.46% for walking downstairs by the RNN-based HAR system. Lower recall values produce confusion between alike activities and the remaining activities as shown in Fig. 3.

Finally, Table IV shows the comparisons against other deep learning approaches. Considering the different sizes of data per classes in PAMAP, F1-score [3] analysis shows that our CNN- and RNN-based HAR systems outperform some previous deep learning approaches even with the use of a single IMU. This result presents a feasibility of HAR for some macro human activities with only a single wearable IMU device.

TABLE IV
BASELINE COMPARISON

Model		F1-Score (%)
DNN	[3]	90.40
LSTM-F	[3]	92.9
CNN	[3]	93.7
CNN	Our Based HAR system	94.43
RNN	Our Based HAR system	96.68

VI. CONCLUSION

In this study, we present two deep learning algorithm-based HAR systems: one is based on CNN, the other RNN. We only used a tri-axial accelerometer and a tri-axial gyroscope from one IMU positioned at the dominant wrist. With our HAR systems, daily activities such as standing, walking, running, walking upstairs, and walking downstairs are recognized with an overall recognition rate of 95.43% by the CNN system and about 96.95% by the RNN system.

ACKNOWLEDGMENT

This work was supported by International Collaborative Research and Development Programme (funded by the Ministry of Trade, Industry and Energy (MOTIE, Korea) (N0002252).

REFERENCES

[1] M. Zeng, L. T. Nguyen, B. Yu, O. J. Mengshoel, J. Zhu, P. Wu and J. Zhang, “Convolutional neural networks for human activity recognition using mobile sensors,” in *Mobile Computing, Applications and Services (MobiCASE), 6th International Conference* IEEE, New York, NY, 2014, pp. 197-205. **DOI:** https://doi.org/10.4108/icst.mobicase.2014.257786

[2] Y. Chen and Y. Xue, “A deep learning approach to human activity recognition based on single accelerometer,” in *Systems, Man, and Cybernetics (SMC), IEEE International Conference*. IEEE, New, NY, 2015 pp. 1488-1492. **DOI:** https://doi.org/10.1109/SMC.2015.263

[3] N. Y. Hammerla, S. Halloran and T. Ploetz. (2016). *Deep, convolutional, and recurrent models for human activity recognition using wearables* [Online]. Available: https://arxiv.org/abs/1604.08880

[4] M. Gjoreski, H. Gjoreski, M. Luštrek and M. Gams, "How accurately can your wrist device recognize daily activities and detect falls?," *Sensors*, vol. 16, no. 6, p. 800, 2016.

[5] F. J. Ordóñez and D. Roggen, “Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition,” *Sensors*, vol. 16, no. 1, p. 115, 2016.

[6] S. Ha, J. M. Yun and S. Choi, “Multi-modal convolutional neural networks for activity recognition,” in *Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference*, New York, NY, 2015, pp. 3017-3022. **DOI:** https://doi.org/10.1109/SMC.2015.525

[7] H. H. Gjoreski, J. Bizjak, M. Gjoreski and M. Gams, "Comparing deep and classical machine learning methods for human activity recognition using wrist accelerometer," in *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI)*, New York, NY, 2016, pp. 1-7.

[8] J. B. Yang, M. N. Nguyen, P. P. San, X. Li and S. Krishnaswamy, “Deep convolutional neural networks on multichannel time series for human activity recognition,” in *International Joint Conference on Artificial Intelligence (IJCAI)*, California, CA, 2015, pp. 3995-4001.

[9] S. Shin and W. Sung, “Dynamic hand gesture recognition for wearable devices with low complexity recurrent neural networks,” in *Circuits and Systems (ISCAS), IEEE International Symposium* IEEE, New York, NY, 2016, pp. 2274-2277.
DOI: https://doi.org/10.1109/ISCAS.2016.7539037

[10] T. K. Kim and R. Cipolla, “Canonical correlation analysis of video volume tensors for action categorization and detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 8, pp. 1415-1428, 2009. **DOI:** https://doi.org/10.1109/TPAMI.2008.167

[11] G. Costante, L. Porzi, O. Lanz, P. Valigi and E. Ricci, “Personalizing a smartwatch-based gesture interface with transfer learning,” in *Proceedings of the 22nd European Signal Processing Conference (EUSIPCO)*, New, York, NY, 2014, pp. 2530-2534.

[12] S. Ha and S. Choi, “Convolutional neural networks for human activity recognition using multiple accelerometer and gyroscope sensors,” in *IEEE International Joint Conference Neural Networks (IJCNN)*, New York, NY, 2016, pp. 381-388. **DOI:** https://doi.org/10.1109/IJCNN.2016.7727224

[13] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 1997.
DOI: https://doi.org/10.1162/neco.1997.9.8.1735

[14] F. A. Gers, N. N. Schraudolph and J. Schmidhuber, “Learning precise timing with LSTM recurrent networks,” *Journal of Machine Learning Research*, vol. 3, pp. 115-143, 2002.

[15] N. Neverova, C. Wolf, G. Lacey, L. Fridman, D. Chandra, B. Barbello and G. Taylor, “Learning human identity from motion patterns,” *IEEE Access*, vol. 4, pp. 1810-1820, 2016.
DOI: https://doi.org/10.1109/ACCESS.2016.2557846

[16] A. Reiss and D. Stricker, “Creating and benchmarking a new dataset for physical activity monitoring,” in *Proceedings of the 5th International Conference on Pervasive Technologies Related to Assistive Environments*, New York, NY, 2012.
DOI: https://doi.org/10.1145/2413097.2413148

[17] A. Reiss and D. Stricker, “Introducing a new benchmarked dataset for activity monitoring,” in *IEEE 16th International Symposium Wearable Computers (ISWC)*, New York, NY, 2012, pp. 108-109.
DOI: https://doi.org/10.1109/ISWC.2012.13

[18] A. Khan, N. Hammerla, S. Mellor and T. Plötz, "Optimising sampling rates for accelerometer-based human activity," *Pattern Recognition Letters*, vol. 73, pp. 33-35, 2016.
DOI: https://doi.org/10.1016/j.patrec.2016.01.001

— This article does not have any appendix. —