# A Deep Learning Approach to Human Activity Recognition Based on Single Accelerometer

Yuqing Chen, Yang Xue*
School of Elec. & Info. Eng.
South China University of Technology
Guangzhou, China
*E-mail: yxue@scut.edu.cn

***Abstract*—In this paper, we propose an acceleration-based human activity recognition method using popular deep architecture, Convolution Neural Network (CNN). In particular, we construct a CNN model and modify the convolution kernel to adapt the characteristics of tri-axial acceleration signals. Also, for comparison, we use some widely used methods to accomplish the recognition task on the same dataset. The large dataset we constructed consists of 31688 samples from eight typical activities. The experiment results show that the CNN works well, which can reach an average accuracy of 93.8% without any feature extraction methods.**

***Keywords*-tri-axial acceleration signal; human activity recognition; deep architecture; convolution kernel**

## I. INTRODUCTION

Human activity recognition (HAR) is a significant research field full of challenges, most of which focus on accuracy, robustness and real-time capability. It's too difficult for researchers to face all these challenges in the past. Thanks to the great improvements made in sensor and processor technologies during the past decade, now we can acquire more accurate sensors in a smaller size and faster processors with lower power consumption. Recently, with a widely application requirements appearing in health care, physical training and military, HAR has attracted more and more attention from both academia and industry.

As a typical pattern recognition system, a HAR system can be divided into several modules, sensing, segmentation, feature extraction, classification and post preprocessing. Generally, according to sensing method, we can classify HAR system into two types: vision-based and acceleration-based methods. Vision-based method usually uses one or more cameras to collect data, while acceleration-based method asks the users to wear several accelerometers for data collecting. The advantage of vision-based system is that it works without placing any sensors with users, but its recognition performance highly depends on light condition, visual angle and other outer factors. On the contrary, acceleration-based system requires users to wear a device, but almost eliminates all those outer interferences. Today, most of the mobile phones have a built-in accelerometer, and that makes it possible for people to construct an acceleration-based HAR system on the mobile platform, with no extra hardware demand.

In recent years, deep architectures show its strong learning capacity and make a great breakthrough in the fields of image recognition [7], speech recognition [22]. Because of its dramatically encouraging performance, many institutions concentrate on developing a series of applications in deep architecture. Some of the applications based on deep architecture have already been brought into service. With the aim of developing a HAR system with high accuracy, good robustness and quick response, we successfully built up a deep architecture for acceleration-based HAR system. The model has been evaluated on a large dataset (with 31688 samples from 8 typical activities). The result is promising, which reaches a higher accuracy than former methods. And more notably, the deep model directly operates on raw data. In other words, the model doesn't need an extra procedure of feature extraction, which matches our goal of quick response. Fig.1 shows a HAR system without feature extraction.

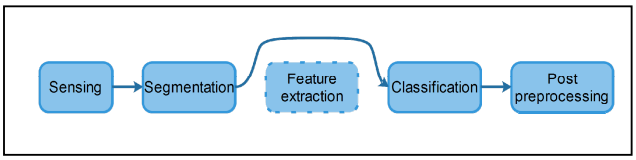


Figure 1. A HAR system without feature extraction

In this paper, we first briefly introduce some related work and widely used methods, then we describe our recognition method using CNN. To address the advantage of using CNN, we set up and implement an experiment about evaluating different methods using the same dataset.

The rest of this paper is organized as follows: Section II gives an overview of related work. Section III demonstrates the proposed HAR method using deep architecture. Section IV is about the evaluation of our method and comparison with other HAR methods. Section V contains some conclusions with some ideas for further work.

978-1-4799-8697-2/15 $31.00 © 2015 IEEE
DOI 10.1109/SMC.2015.263

## II. RELATED WORK

Many researchers have made a lot of attempts on building an accurate and robust acceleration-based HAR system. Those attempts can be summed up in data collection, feature extraction and classifier.

### A. Data collection

The essential module of collecting human acceleration data is acceleration sensor. But the collected data are varied. First of all is the sampling rate. Researchers indicate that the basic frequency of walking is 2 Hz, running is 2.5-3 Hz [12]. So as to collecting more high frequency details about human activity, most researchers set the sampling rate much higher than the Nyquist frequency, range from tens to hundreds Hz. Second is the number of acceleration sensor. Many previous studies used several accelerometers on different body parts like waist, cloth pocket, leg, etc. This kind of HAR system probably requires users to carry complicated devices on body, which is uncomfortable and expensive. Recently, many studies [1] find out that human activity can be accurately recognized by using a single accelerometer.

### B. Feature extraction

It has been proved that many signal processing methods are useful in extracting features for HAR, including time-domain features, frequency-domain features and some others. Time-domain features include mean, variance, short time energy, correlation between axes, zero crossing rate, etc. [1]. In the frequency-domain, many researchers use Fast Fourier Transform (FFT) coefficients and Discrete Fourier Transform (DCT) coefficients [1]. There are also some other features like Principal Component Analysis (PCA) [6], Autoregressive Model (AR) [5] and Haar filters [10], which have been used in experiments and returned good results.

### C. Classifier

There are numbers of classification methods which have been applied for HAR. The popular classifiers include Support Vector Machine (SVM) [1], Decision Tree (DT) [2], Bayesian Methods [2], Neural Networks (NN) [3], Logistic Regression [3] and Hidden Markov Model (HMM) [19]. Moreover, some researchers use the idea of Bagging and Boosting to improve the accuracy.

In this paper, we propose a CNN approach to activity recognition, which can extract features by itself and without any domain specific knowledge about acceleration data. By skipping the procedure of extracting features, the HAR system could become more responsive.

## III. A CNN APPROACH TO ACCELERATION-BASED HAR

Recently, deep neural network architectures have made significant improvements in many fields of pattern recognition. Especially the CNN, one of the most powerful deep architectures, is widely used in computer vision and image recognition.

Our method mainly modifies the CNN architecture from the angle of convolution kernel, we also use a simple method to choose the best number of epochs.

### A. Convolution kernel

The biggest distinction between tri-axial acceleration data and image data is the difference of the size of their data. The length and width of images can be varied, but the width of the tri-axial acceleration data is fixed to 3, which represents the x, y, z component respectively.

Based on this premise, a conventional CNN approach is hardly utilized in tri-axial acceleration data without any changes in the basic model, because the width of the data greatly restricted the construction of a CNN architecture.

There are two alternatives, resizing the data or resizing the convolution kernel. Resizing the data might be easier to implement, but it's not an appropriate approach. On the one hand, it may lose the relevance between adjacent acceleration values, which go against the design idea of CNN. On the other hand, it is inconvenient for generalization, the resized length and width are difficult to be determined when facing different acceleration data with varied length. Oppositely, resizing the convolution kernel is more elegant. The CNN structure can deal with acceleration data with different lengths easily, without losing the information about adjacent acceleration values.

How to determine the width of convolution kernel? Actually, there are only three options: 1, 2 or 3. In order to extract information between different axes, also to improve the flexibility of rotation, we set the convolution kernel width to 2. In fact, Fig.2 shows that the convolution kernel with the width of 2 performs the best.

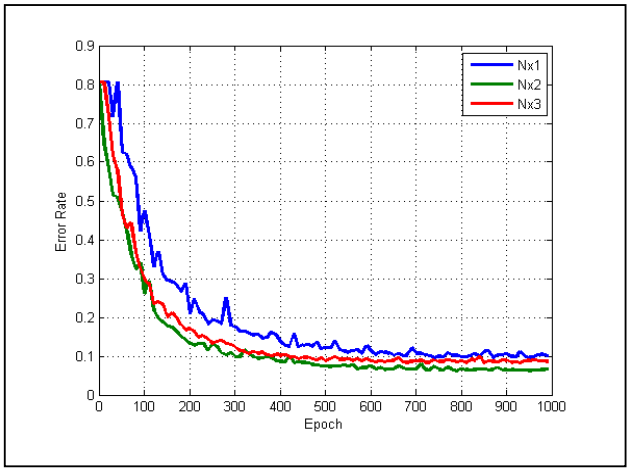


Figure 2. Error rate of CNN with different width of convolution kernels

### B. Locate the best epochs

Once the structure of the CNN is determined, we have to train the network, tuning the best parameters for activity recognition.

Generally, the error rate of training set will gradually reduce as the training proceeds. At the beginning, the error rate on the test set will decrease, after a specific epoch, the error rate will stop decreasing, sometimes even increase. This phenomenon is called over fitting, it's caused by over training

the network. As shown in Fig.3, the best epoch is when the test error starts increasing.

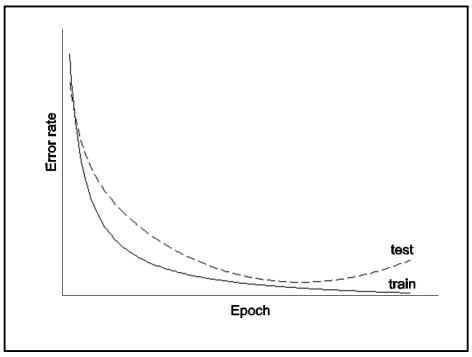


Figure 3. Test and train error in training process

To locate where is the best epochs, we can choose a set of samples for validation, and evaluate the error rates of training set and validation set in every 10 epochs. When the error rate of validation stops decreasing and begins to increase, we stop training and evaluate the network on the test set. The performance of this model will be shown in the next section.

## IV. EXPERIMENT AND ANALYSIS

In this section, we will describe the experiments of evaluating the performance of our model and other popular models.

### A. Our data acquisition

With the aim of train a generalized model, also to evaluate the model from an objective perspective, we need to acquire a great amount of data. Hence, an Android application is developed for data acquisition. Instead of building a sampling device, developing an application is a more appropriate way. First of all, developing an application can offer convenience for volunteers to execute data acquisition. The volunteers who have Android phones can conveniently record and label data by themselves. Besides, building application is cheap and effective. A copy of the application is equal to a new sampling device. That makes it easier for us to get access to a greater amount of data. What's more, it acts as a basic platform for our future work of building a HAR system on intelligent terminals.

We obtain 8 typical activities, including falling (F), running (R), jumping (J), walking (W), walking quickly (WQ), step walking (SW), walking upstairs (U), and walking downstairs (D), in a naturalistic environment from 100 healthy subjects (68 males and 32 females). With the purpose of maintaining diversity of data, all the subjects are required to record data from 3 different body parts: cloth pocket, trouser pocket and waist.

All the data are recorded by a single tri-axial accelerometer in Android phone. Because of the feature of the operating system, the sampling rate of acceleration sensor cannot be fixed to a specific value. Under our estimation, the sampling frequency is 100Hz with a slight fluctuation.

### B. Preprocessing

The recorded raw 3D acceleration signal stream is cropped into the same size with an overlap of 50%. Every sample is a matrix with the size of 256x3. Since the sampling rate is around 100 Hz, one sample records the three dimension acceleration data with a length of about 2.56 seconds.

After the simple preprocessing, we have 31688 labeled samples, 27395 are used for training, the rest 4293 are for testing. It's reasonable to consider the sample in the training set and the test set are independent because the signals of the training set and test set are collected by different volunteers.

### C. Architecture of CNN

The detail of the CNN model is shown in Fig.4. It contains 3 convolution layers and 3 pooling layers. The model works directly on the raw acceleration signal without any other processing.

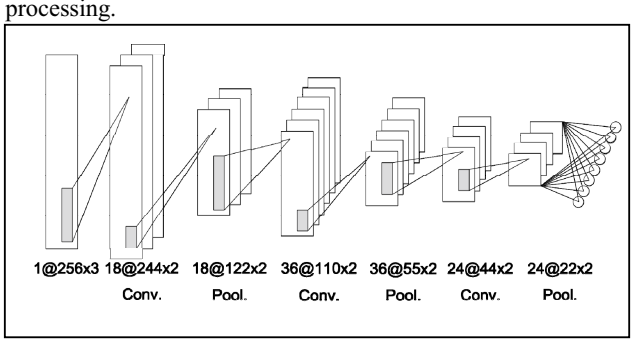


Figure 4. Architecture of Convolution Neural Network

### D. Results

To compare the recognition capability of CNN, we extracted some powerful features and used some popular classifiers to make the evaluation. The recognition results are shown in Fig.5 and Fig.6. The DC component of Fast Fourier Transform (FFT) and Discrete Cosine Transform (DCT) coefficients have been removed. Time-domain Features (TF) include mean, variance, short-time energy and correlation coefficients. Support Vector Machine (SVM) and an 8-layer Deep Belief Network (DBN) are selected as classifiers for performance comparison.

Consistent with previous research, from the figures above, the performance of time domain features is not as good as frequency domain features like FFT or DCT coefficients. What's more, from Fig.5 and Fig.6, we can find out that our proposed CNN model works better than all manually extracted features, and performs up to a high level in every activity, especially in step walking, walking quickly, walking downstairs and walking upstairs which are too difficult to be classified using FFT and DCT coefficients. It implies that CNN must have extracted more effective features than FFT and DCT.

### E. Analysis

TABLE I. shows the confusion matrix of CNN model, which can help us ascertain which activities are more likely to be misunderstood. From TABLE I. , we can see that walking is relatively difficult to be recognized. That's because walking is

often confused with walking quickly, downstairs, and upstairs. These activities are also easily mistaken when using other methods shown in Fig.5 and Fig.6.

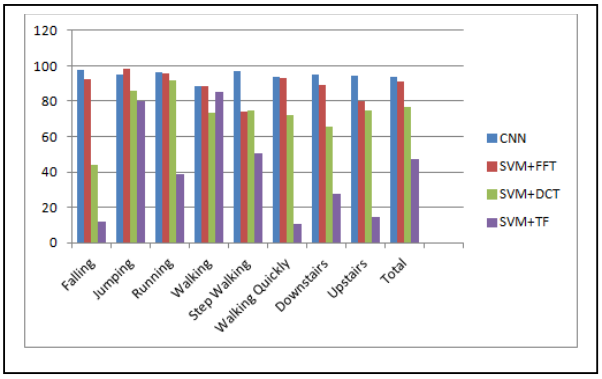


Figure 5. Recognition accuracy of CNN and SVM classifier with different features

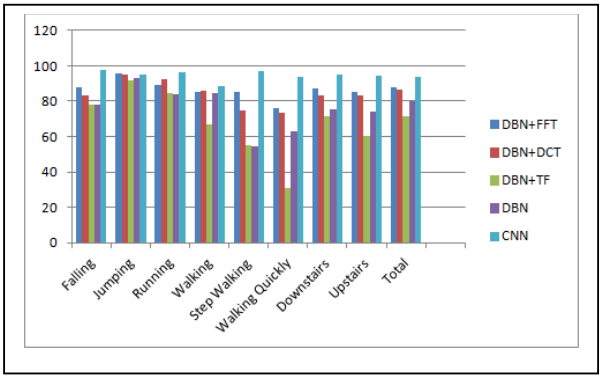


Figure 6. Recognition accuracy CNN and DBN and DBN with different features.

TABLE I. CONFUSION MATRIX OF CNN MODEL, AVERAGE ACCURACY IS 93.8%

| Type | Recognized As | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | F | J | R | W | SW | WQ | D | U |
| F | **0.975** | | | | | | | 0.024 |
| J | | **0.952** | 0.031 | 0.012 | 0.001 | | | 0.003 |
| R | | 0.028 | **0.960** | | | | 0.013 | |
| W | | | 0.006 | **0.883** | 0.001 | 0.026 | 0.034 | 0.050 |
| SW | | | | 0.007 | **0.970** | | 0.003 | 0.020 |
| WQ | | | | 0.058 | | **0.936** | | 0.006 |
| D | | 0.010 | 0.011 | 0.012 | 0.004 | | **0.948** | 0.015 |
| U | | | | 0.037 | 0.002 | 0.006 | 0.008 | **0.946** |

Even though these obstacles may exist, the proposed CNN also achieves a high accuracy, which is 93.8%, on such a huge dataset even without any feature extraction. We wonder what features the model has extracted from the raw acceleration signal. In Fig.7, we visualized the well trained convolution kernels in the first convolution layer. As mention above, the convolution kernels we use are Nx2, we visualize them as 2 curves. It can be seen that the first convolution layer has found out various relationships between axes, some are in the same phase, some with delay or even opposite in phase. The acceleration signals obtained from different kinds of activities are all composed of these basic features.

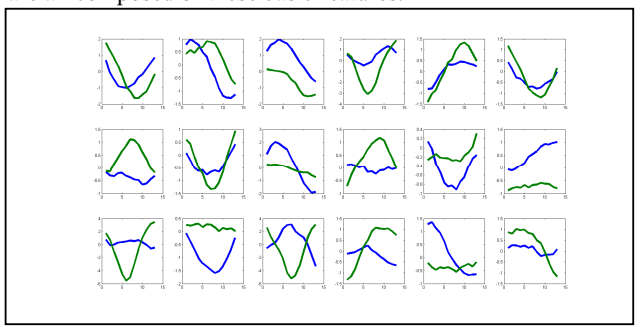

Figure 7. Visualization of the convolution kernels in the first convolution layer

We try to explain why the proposed CNN works better than other human activity recognition systems or fully connected network like DBN. Theoretically, the input units in fully-connected network have the same connection and would not change the performance of network by swapping their positions. That is not a nice property for classification on raw data, because it ignores the dependencies between adjacent units. But CNN, as a sparse neural network, has three important architectural ideas, namely local receptive fields, shared weight and spatial sub-sampling, that ensure the information existing in adjacent units can be extracted fully. As to FFT and DCT coefficients, they can only extract the information located in the same axis, without figuring out the information in different axes.

## V. CONCLUSION

This paper proposes an acceleration-based HAR algorithm using CNN, a popular used deep architecture in image recognition. According to the characteristic of acceleration data, we modified the conventional CNN structure. Then experiments are constructed to evaluate the recognition performance of our proposed CNN and other widely used methods. The experiments are executed on a large dataset of eight kinds of typical activities with 31688 samples from 100 subjects. The results show that the CNN works well, reaches an accuracy of 93.8%. The proposed model is accurate and robust without any feature extraction, which is suitable for building a real-time HAR system on mobile platform.

## ACKNOWLEDGMENT

We would like to thank all those warmly cooperative volunteers. This work was supported in part by the research funding of NSFC (grant no. 61201348) and the Fundamental Research Funds for the Central Universities of China (no. D215106w, 2015ZZ033).

## REFERENCES


[1] Yang Xue, and Lianwen Jin, "A naturalistic 3D acceleration-based activity dataset & benchmark evaluations," Systems, Man and Cybernetics, pp. 4081-4085, 2010.

[2] Ling Bao, and Stephen S. Intille, "Activity recognition from user-annotated acceleration data," Pervasive computing, pp. 1-17, 2004.

[3] Jennifer R. Kwapisz, Gary M. Weiss, and Samuel A. Moore, "Activity recognition using cell phone accelerometers," ACM SigKDD Explorations Newsletter, vol.12, no.2, pp. 74-82, 2011.

[4] Oscar D. Lara, and Miguel A. Labrador, "A survey on human activity recognition using wearable sensors," Communications Surveys & Tutorials, vol.15, no.3, pp. 1192-1209, 2013.

[5] Zhenyu He, and Lianwen Jin, "Activity recognition from acceleration data using AR model representation and SVM," International Conference on Machine Learning and Cybernetics, vol.4, pp. 2245-2250, 2008.

[6] Zhenyu He, and Lianwen Jin, "Activity recognition from acceleration data based on discrete consine transform and SVM," Systems, Man and Cybernetics, pp. 5186-5189, 2009.

[7] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, "Imagenet classification with deep convolutional neural networks," Advances in neural information processing systems, pp. 1097-1105, 2012.

[8] Oscar D. Lara, Perez Alfredo J., Labrador Miguel A., Posada Jose D., "Centinela: A human activity recognition system based on acceleration and vital sign data," Pervasive and mobile computing, vol.8, no.5, pp. 717-729 ,2012.

[9] Zhenyu He, Zhibin Liu, Lianwen Jin, Li-Xin Zhen, Jian-Cheng, Huang, "Weightlessness feature—a novel feature for single tri-axial accelerometer based activity recognition," International Conference on Pattern Recognition, pp. 1-4, 2008.

[10] Yuya Hanai, Jun Nishimura, and Tadahiro Kuroda, "Haar-like filtering for human activity recognition using 3D accelerometer," Digital Signal Processing Workshop and IEEE Signal Processing Education Workshop, pp. 675-678, 2009.

[11] Uwe Maurer, Asim Smailagic, Daniel P. Siewiorek, Michael Deisher, "Activity recognition and monitoring using multiple sensors on different body positions," Wearable and Implantable Body Sensor Networks, pp. 113-116, 2006.

[12] Juha Parkka, et al., "Activity classification using realistic data from wearable sensors," Information Technology in Biomedicine, IEEE Transactions, vol.10, no.1 ,pp. 119-128, 2006.

[13] Akram Bayat, Marc Pomplun, and Duc A. Tran, "A study on human activity recognition using accelerometer data from smartphones," Procedia Computer Science, vol.34, pp. 450-457, 2014.

[14] Stewart G. Trost, Yonglei Zheng, and Weng-Keen Wong, "Machine learning for activity recognition: hip versus wrist data," Physiological measurement, vol.35, no.11, pp. 2183-2189, 2014.

[15] Luciana C. Jatoba, Ulrich Grobmann, Chistophe Kunze, Jorg Ottenbacher, and Wilhelm Stork, "Context-aware mobile health monitoring: Evaluation of different pattern recognition methods for classification of physical activity," Engineering in Medicine and Biology Society, pp. 5250-5253, 2008.

[16] Yen-Ping Chen, Jhun-Ying Yang, Shun-Nan Liou, Gwo-Yun Lee, Jeen-Shing Wang, "Online classifier construction algorithm for human activity detection using a tri-axial accelerometer," Applied Mathematics and Computation, vol.205, no.2, pp. 849-860 ,2008.

[17] Miikka Ermes, Juha Parkka, and Luc Cluitmans, "Advancing from offline to online activity recognition with wearable sensors," Engineering in Medicine and Biology Society, pp. 4451-4454, 2008.

[18] Oscar D. Lara, and Miguel A. Labrador, "A mobile platform for real-time human activity recognition," Consumer Communications and Networking Conference, pp. 667-671, 2012.

[19] Chun Zhu, and Weihua Sheng, "Human daily activity recognition in robot-assisted living using multi-sensor fusion," International Conference on Robotics and Automation, pp. 2154-2159, 2009.

[20] Sungyoung Lee, et al., "Semi-Markov conditional random fields for accelerometer-based activity recognition," Applied Intelligence, vol.35, no.2, pp. 226-241, 2011.

[21] Tâm Huynh, and Bernt Schiele, "Analyzing features for activity recognition," Proceedings of the joint conference on Smart objects and ambient intelligence: innovative context-aware services: usages and technologies, pp. 159-164, 2005.

[22] Geoffrey Hinton, et al, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," Signal Processing Magazine, vol.29, no.6, pp. 82-97 ,2012.

[23] Alex Graves, A-R. Mohamed, and Geoffrey Hinton, "Speech recognition with deep recurrent neural networks," International Conference on Acoustics, Speech and Signal Processing, pp.6645-6649, 2013.