# A GAN-Based Data Augmentation Approach for Sensor-Based Human Activity Recognition

## Wen-Hui Chen\*, Po-Chuan Cho

Graduate Institute of Automation Technology, National Taipei University of Technology, Taipei, Taiwan.

\* Corresponding author. Tel.: +886-2-27712171-4323; email: whchen@ntut.edu.tw Manuscript submitted March 21, 2021; accepted May 5, 2021. doi: 10.17706/ijcce.2021.10.4.75-84

**Abstract:** Recently, deep learning has emerged as a powerful technique and been successfully employed for various tasks. It has also been applied to human activity recognition and showed better performance than traditional machine learning algorithms. However, the success of deep learning always comes with large labeled datasets when the learning model goes deeper. If the training data is limited, the performance of the classification model may not generally perform well due to overfitting of the networks to the training data, which can be alleviated through data augmentation. Generative adversarial networks (GANs) can be used as a technique to produce data artificially. GAN-based approaches have made rapid progress in generating synthetic data, but they are mostly studied for image data. Comparatively little research has been conducted to examine the effectiveness of generating sensor data using GANs. This study aims to investigate the data scarcity problem by using conditional generative adversarial networks (CGANs) as a data augmentation method. The proposed approach was experimentally evaluated on a benchmark sensor dataset for activity recognition. The experimental results showed that the proposed approach can boost the model accuracy and has better performance when compared with existing approaches.

Key words: Human activity recognition, inertial measurement units, generative adversarial networks.

# 1. Introduction

Human activity recognition (HAR) has many important potential applications such as driving behavior analysis [1], video surveillance, and healthcare [2]. In terms of sensor types used for classifying activities, the approach to HAR can be generally categorized into vision-based and sensor-based directions [3]. Vision-based approaches recognize activities by analyzing image sequences captured by image sensors such as video cameras, while sensor-based approaches use inertial measurement units (IMUs) such as accelerometers and gyroscopes or environmental sensors such as status sensors and pressure sensors to identify activities based on sensor signal variations. Although image sensors can capture rich information that helps recognize complex activities, visual monitoring is considered intrusive and could raise privacy and ethical issues. Therefore, sensor-based approaches are more popular when it comes to privacy concerns. This study focused on sensor-based HAR, specifically, IMU-based wearable sensors.

Sensor-based HAR requires sensor deployment to acquire information. Sensors can be attached to the subject being observed in a wearable fashion to gather activity data. However, sensor data generally contains noise and the data variation could be irregular. As such, it is difficult to identify activities accurately through observation of signal variations. Hence, feature extraction from sensor data plays an important role and is the key to activity recognition applications in earlier studies [4]. Previous studies heavily rely on experts'

experience and knowledge to perform feature extraction before feeding raw sensor data into the activity classifier [5]. As the HAR can be considered as a classification problem, the algorithms used for classification in machine learning have been widely explored, such as decision trees, random forests, support vector machines, and Naive Bayes [6].

Traditional machine learning approaches for HAR require feature extraction and need much labor work to label the training data, which is time-consuming. In addition, the performance of the recognition system developed by such approaches is easily affected by humans' subjective judgments on feature selection. Deep learning has been proven effective to learn and extract features automatically from data without human intervention. Hence, current research for HAR tends to circumvent the aforementioned issue of handcrafted features by exploiting deep learning-based approaches [7].

When we train a machine-learning model, we adjust its model parameters such that the trained model can map the input into the desired output. State-of-the-art deep neural networks typically contain parameters that need to be tuned in the order of millions or more. Limited data only provides a small amount of data, so it can not cover the input space well and will fail to construct a good model to capture the underlying pattern of the dataset.

In practice, sufficient data for training the learning model is not always available. Data augmentation is a common way to deal with the lack of data problems. In image recognition tasks, data augmentation can be implemented by artificially increasing the number of training samples via performing scaling, shifting, flipping, translating, or rotating on the original images [8]. Those operations provide more data and increase a variety of training samples leading to better model performance.

Sensor data is a time series of data points sampled from raw sensor signals in a one-dimension format, which is different from image data. Therefore, augmented techniques successfully employed to image data are not necessarily applicable to wearable sensor data. In addition, unlike the image data, the physical meaning of sensor data can hardly be explained by visual observation. In general, augmentation for sensor data yet to be well studied. In this study, we proposed an effective approach to data augmentation for sensor data based on conditional generative adversarial networks (CGANs).

The remainder of this paper is organized as follows. Section 2 provides an overview of HAR and major existing methods for deep learning-based HAR. Section 3 introduces the concept of generative adversarial networks and sensor data augmentation in the application of HAR. Section 4 demonstrates the experimental results to verify the effectiveness of the proposed approach. Conclusions are drawn in Section 5.

#### 2. Overview of Human Activity Recognition

Activity recognition is a well-known problem and has been studied for decades. Sensor-based HAR can be considered as the problem of classifying inertial sensor data such as tri-axial signals of accelerometers and gyroscopes into well-defined human activities. Traditional approaches to the HAR problem involve a two-stage design process. The first stage is to extract features from sensor signals based on human knowledge. Examples of features are the mean, variance, frequency, and amplitude of the sensor signals. Then, the extracted features are used to train a classifier in the second stage. Therefore, previous studies on HAR mainly work on discovering effective features for sensor signals in the first stage and/or exploring better algorithms for training classifiers in the second stage to increase the classification accuracy.

Although past studies have made much progress on HAR, the feature engineering requires strong domain knowledge and the exploration of good features is still a challenge as they heavily affect the performance of classification models.

Recently, deep learning methods have shown state-of-the-art results in many tasks. Taking advantage of deep learning on feature learning, deep learning-based HAR has become the research trend. A brief review of

major HAR-based deep learning is described as follows.

## 2.1. HAR Based on Dense Neural Networks (DNNs)

In the earlier research, fully-connected dense neural networks were used to serve as a classification model after feature extraction based on human experience and domain knowledge. Although some authors used multi-layer DNNs as a learning model, those networks were rather shallow compared to the current deep neural structure. In [9], the authors used human selected features to feed into a multi-layer feedforward neural network with two hidden layers to train the classifier. As fully connected neural networks are not able to capture local dependencies of sensor signals, they may not always perform well. Deeper networks provide better representation as suggested in [10], which can help increase the model performance in identifying complex activities, but it needs more data to train the deep network.

Today, it is widely recognized that a deep network structure has a direct impact on model performance, but how to find the best network structure remains an open question. In addition, many fine-tuning works are required to obtain a good generalization model.

## 2.2. HAR Based on Convolutional Neural Networks (CNNs)

Convolutional neural networks are a type of neural networks with convolutional layers, which were originally designed for image classification tasks. As such, the operation of CNNs is well suited for the twodimensional (2-D) data format. A typical architecture of CNNs contains a mix of convolution layers, pooling layers, and fully-connected layers [11].

CNN has the ability to extract features from data in a deeper fashion and can achieve promising results in many tasks such as image classification and speech recognition. When applied to HAR, CNN has its merits to capture local dependency by convolution oprations in convolutional layers and scale invariance by pooling operations in pooling layers. Local dependency in HAR means that the nearby sensor signals are likely to be correlated, while scale invariance refers to being invariant for different movements.

As the raw sensor signal is a sequence of one-dimension data, when applying CNNs to HAR, the input sensor data has to be pre-processed. A common way to do this is to formulate sensor data into a two-dimension format to adapt itself for performing convolution and pooling operations. Otherwise, a strategy has to be made to perform 1-D convolution and pooling operations.

## 2.3. HAR Based on Recurrent Neural Networks (RNNs)

DNNs and CNNs are feedforward networks. There is no difference in the order of each input for a feedforward network and the only input is the current observation it receives, while recurrent neural networks are different from feedforward networks. An RNN has the feedback loop connected to past states. An RNN take the current observation and what it has perceived previously as inputs. The ability of RNNs in discovering the temporal relationship of input data make RNNs well-suited for time series analysis. Variations of RNNs with gating mechanisms such as long short-term memory (LSTM) or gated recurrent units (GRUs) provide a solution to overcome the gradient vanishing or exploding problem, which brings LSTMs and GRUs models to the research of HAR.

LSTMs have become a popular HAR model in past few years till the rise and success of attention model. The authors in [12] explored deep, convolutional, and recurrent approaches across three sensor datasets and provided guidelines for applying deep leaning to HAR.

## 2.4. HAR Based on Hybrid Models

Each network structure of neural networks has its pros and cons. Combining CNNs and RNNs to train a hybrid model can take advantages from both CNN and RNN models and achieve better results in recognizing

various activities [13]. The use of attention mechanisms in deep learning models has recently attracted much interest for its success in language processing [14] and computer vision tasks. A hybrid model includes attention mechanism is an emerging research direction. With the inclusion of attention component in HAR models, a new state-of-the-art performance has been reported on four diverse HAR benchmarks [15].

## 3. Data Augmentation Using Generative Adversarial Networks (GANs)

Data augmentation is a technique to avoid overfitting by generating synthetic data to improve the model generalization. In [16], the authors used CNNs to augment data by transforming existing labelled samples for time-series classification and presented a window slicing with warping technique to increase the number of training samples. They showed that the mixing dataset could improve the classification accuracy. However, window slicing is similar to image cropping for reducing the dependency on event locations, which may cause the problem of data label change. In addition, the CNN-based architecture will not be able to capture the temporal dependency. As such, it may not be applicable to sensor data in the application of HAR.

An effective way to increase the data size is to make use of the invariant properties of the data with certain operations. In [17], several operations on original data for data augmentation are addressed. Those operations aim to add synthetic data to enrich training sets and help the model learn the range of intra-class invariances. In image recognition, minor changes in an image such as jittering, scaling, permutation, cropping, and rotating are proven to be effective as they do not change the data labels. Although those transformations have been shown effective in the computer vision community, they are not always applicable to sensor data.

GANs are a type of generative models comprising two different networks, a generator network (G) and a discriminator network (D) [18]. Both G and D networks are trained as a two-player minimax zero-sum game. The generator's task is to learn how to generate samples that are similar to the real data, while the discriminator is responsible for distinguishing the generated data from the real data. The task is expected to end up at an equilibrium point where the generator can generate samples that are not able to be distinguished by the discriminator.

In general, the generator in a typical GAN is trained to learn the synthetic data that is likely to come from the sample space by taking a latent variable drawn from some distribution such as normal distribution or uniform distribution. The discriminator then evaluates the generated data and gives a scale value representing the probability that the generated sample comes from the real dataset. When the training is complete, the generator can be used to produce data to increase the amount of the required training data as the purpose of data augmentation.

There are many GAN variants and applications since its debut in 2014 [18]. In this study, we investigate data augmentation based on conditional generative adversarial networks [19] to generate synthetic data for wearable sensors. Various methods for using GANs to expand training datasets have been proposed. In [20], the authors employed GAN-based data augmentation techniques to increase the emotion classification accuracy and showed their success on three benchmark datasets. Data augmentation using GANs has also been applied to medical image classification with promising results [21].

The results reported in previous research suggest that GANs could have a significant benefit when used for data augmentation in image classification tasks. However, limited studies associated with GAN-based data augmentation have been conducted on wearable sensor data. In this work, we adopted a GAN-based framework that can effectively generate available sensor data for HAR.

#### 4. Experimental Results

## 4.1. Benchmark Dataset

The objective of this experiment was to observe the impact of training data size on recognition accuracy.

The dataset used in this study was derived from [22]. It was collected by attaching a smartphone on the waist, with 30 participants at the age from 19 to 48, conducting six activities (walking, walking upstairs, walking downstairs, sitting, standing, and lying) in a laboratory environment. The sensor readings were sampled at 50 Hz on tri-axial accelerometers and tri-axial gyroscopes. We set the time step for the LSTM parameter to 100 and preprocessed the original data by removing unfilled data points within a time step for each activity, as listed in Table 1.

Table 1. Experimental Dataset				
Activities	Instances	Training set(ratio)	Test set(ratio)	
Walking	1,162	924 (16.36%)	238 (16.84%)	
Upstairs	1,079	879 (15.55%)	200 (14.16%)	
Downstairs	982	776 (13.73%)	206 (14.58%)	
Sitting	1,212	972 (17.21%)	240 (16.98%)	
Standing	1,317	1,049 (18.56%)	268 (18.96%)	
Lying	1,312	1,051 (18.59%)	261 (18.48%)	
Total	7,064	5,651 (100%)	1,413 (100%)	

## 4.2. Performance Evaluation

There are several performance metrics used in the study of recognition problems, such as the confusion matrix, accuracy, precision, and recall. As the instances of the dataset used in this experiment were an imbalance, a simple accuracy metric is unable to fairly evaluate recognition performance. As such, we used the F1-measure to calculate the number of misclassified activity instances and evaluated the recognition performance. Defined in (1), the F1-score is a measure of the test accuracy by combining the measure that assesses the precision and recall scores in (2) and (3), respectively.

$$F1-score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$
(1)

Precision 
$$= \frac{1}{N} \sum_{i=1}^{N} \frac{TP_i}{TP_i + FP_i}$$
(2)

$$\operatorname{Recall} = \frac{1}{N} \sum_{i=1}^{N} \frac{TP_i}{TP_i + FN_i}$$
(3)

where

N is the total number of class instances TP is true positive FN is false negative

FP is false positive

## 4.3. Classification Model

In this study, a CNN-LSTM architecture, as illustrated in Fig. 1, was adopted as the base classification model for evaluating the performance of data augmentation. CNNs are used to extract features as a sequence of sensor input for the LSTMs. To adapt the sensor inputs to form a virtual image or 2-D format, we treat each axis of the accelerometer and gyroscope as a channel and then perform convolution and pooling operations separately. The parameters for this experiment are shown in Fig. 1 with the stride size set to one in convolution operations.



Fig. 1. A CNN-LSTM architecture as a base model for activity classification.

We use different portions of the training set listed in Table 1 to train the base model. To investigate the impact of augmented data on model performance, we trained a CGAN to generate synthetic data using 20%, 50%, and 80% of the training set conditioned on their activity labels, encoded as one-hot vectors and generated the required amount of data equal to 100% in total. The results are listed in Table 2 and Fig. 2. From Table 2, we can observe that the amount of generated training data has an impact on model performance.

Ta	<u>able 2. Test Performance with Various Sizes of Tr</u>	aining Data
	Percentage of training data without Augmented data	F1 <sub>score</sub>
	20% Training set	84.80%
	50% Training set	88.21%
	80% Training set	91.53%
	100% Training set	94.27%





Fig. 2. Test performance with various percentages of the training set.

Table 3	. Test	Perforn	nance	with	Augme	ented	Data	Using	CGAN	٧s
				-	. 0			0		-

0	0
The mixture of the training set and augmented data	F1 <sub>score</sub>
20% Training set and 80% Augmented data	85.01%
50% Training set and 50% Augmented data	90.1%
80% Training set and 20% Augmented data	92.91%

In the generator network of the CGAN, a noise prior *z* was drawn from a normal distribution. Both *z* and the activity label y are mapped into hidden layers with the rectified linear unit (ReLU) activation, and the layer size was set to 600. The number of units of the LSTM was set to 64. The dimension of latent vector was set to 100. In the discriminator network of the CGAN, the discriminator maps input x to a layer with 240 units, and y to a layer with 50 units. The model was trained using stochastic gradient descent with the learning rate, batch size, and time step set to 0.0002, 64, and 100, respectively. We then have a final layer as our output for generating the 6-dimensional (tri-axial accelerometers and tri-axial gyroscopes) sensor samples.



Fig. 3. Test Performance with various amounts of augmented data.

The results indicated that data augmentation is effective for boosting model performance with limited training data. Table 3 and Fig. 3 show the test performance with various amounts of training data and augmented data.

Table 4. Comparison with Other Approaches				
Sources	Methods	Performance		
Zhao, Y. et al. [23]	Deep residual bidirectional LSTM	93.57%		
Ronao, C. A., et al. [24]	Deep CNN with temporal FFT	95.75%		
Our Approach	GAN-based data augmentation	96.39%		

Table 5. The Confusion Matrix of the Proposed Approach Activities Walking Lying Upstairs Downstairs Sitting Standing Walking Upstairs Downstairs Sitting Standing Lying



Fig. 4. A visual representation of the confusion matrix derived from the proposed approach.

To make a comparison with other deep learning approaches for sensor-based HAR, we used the entire training dataset listed in Table 1 to train the CGAN model and generate additional 1,000 samples for each activity class. By combing the original training data and the data generated by CGAN, we can increase the number of training samples and obtain better performance at 96.39% accuracy on the test set. The results also outperform deep residual bidirectional LSTM [23] and deep CNN with temporal fast Fourier transform [24] as listed in Table 4. The corresponding confusion matrix derived by our approach was listed in Table 5 and Fig. 4.

## 5. Conclusion

In this study, we have presented a CGAN-based framework for synthetic sensor data generation to improve the model performance for sensor-based HAR applications. As a complex learning model will generally not perform well with limited data, the proposed CGAN-based approach allows us to increase the amount of the training data, resulting in better performance. In addition, we have experimentally demonstrated the proposed approach on a benchmark dataset and made a comparison with other approaches to validate the effectiveness of the proposed approach.

## **Conflict of Interest**

The authors declare no conflicts of interest.

## Author Contributions

Dr. Wen-Hui Chen provided methodology, conceptualization, and supervised the research project. Mr. Po-Chuan Cho was involved in model design, coding, and reporting the results.

## Acknowledgment

The authors would like to thank Mr. Chieh-Yuan Huang for the experimental setup. This work was supported by the Taiwan Ministry of Science and Technology (MOST) under Grant No. MOST 108-2221-E-027-104 & 109-2221-E-027 -049.

# References

- [1] Lu, D., Nguyen, D., Nguyen, T., & Nguyen, H. (2018). Vehicle mode and driving activity detection based on analyzing sensor data of smartphones. *Sensors*, *18* (*4*), 1-25.
- [2] Kim, E., Helal, S., & Cook, D. (2010). Human activity recognition and pattern discovery. *IEEE Pervasive Computing*, *9*(1), 48-53.
- [3] Aggarwal, J. K., & Ryoo, M. S. (2011). Human activity analysis: a review. *ACM Computing Surveys*, *43(3)*, 1-47.
- [4] Plötz, T., Hammerla, N. Y., & Olivier, P. (2011). Feature learning for activity recognition in ubiquitous computing. *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence* (pp. 1729-1734). Barcelona, Spain.
- [5] Huynh, T., & Schiele, B. (2005). Analyzing features for activity recognition. *Proceedings of the 2005 Joint Conference on Smart Objects and Ambient Intelligence* (pp. 159–163). New York, USA.
- [6] Lara, O., & Labrador, M. (2013). A survey on human activity recognition using wearable sensors. *IEEE Communications Surveys & Tutorials*, *15(3)*, 1192-1209.
- [7] Wang, J., Chen, Y., Hao, S., Peng, X., & Hu, L. (Dec. 2017). Deep learning for sensor-based activity recognition: A survey. *Pattern Recognition Letters.* arXiv:1707.03502v2.
- [8] Hauberg, S., Freifeld, O., Larsen, A., Fisher, J., & Hansen, L. (2015). Dreaming more data: Class-dependent

distributions over diffeomorphisms for learned data augmentation. arXiv:1510.02795.

- [9] Vepakomma, P., De, D., Das, S. K., & Bhansali, S. (2015, June). A-wristocracy: Deep learning on wrist-worn sensing for recognition of user complex activities. *Proceedings of the IEEE 12<sup>th</sup> International Conference on Wearable and Implantable Body Sensor Networks*. Cambridge, USA.
- [10] Bengio, Y. (2013). Deep learning of representations: Looking forward. *Proceedings of the International Conference on Statistical Language and Speech Processing* (pp. 1–37).
- [11] Zeng, M., Nguyen, L. T., Yu, B., Mengshoel, O. J., Zhu, J., Wu, P., & Zhang, J. (2014). Convolutional neural networks for human activity recognition using mobile sensors. *Proceedings of the 6th International Conference on Mobile Computing, Applications and Services* (pp. 197-205). Austin, USA.
- [12] Hammerla, N. Y., Halloran, S., & Ploetz, T. (2016). Deep, convolutional, and recurrent models for human activity recognition using wearables. Retrieved from https://arxiv.org/abs/1604.08880
- [13] Ordonez, F. J., & Roggen, D. (2016). Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition. *Sensors*, *16(1)*, 1-25.
- [14] Vaswani, A., et al. (2017). Attention is all you need. Proceedings of the Advances In Neural Information Processing Systems (pp. 5998–6008). Long Beach, USA.
- [15] Abedin, A., Ehsanpour, M., Shi, Q., Rezatofighi, H., & Ranasinghe, D. C. (2020). Attend and discriminate: Beyond the state-of-the-art for human activity recognition using wearable sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies.*
- [16] Guennec, A. L., Malinowski, S., & Tavenard, R. (2016). Data augmentation for time series classification using convolutional neural networks. *Proceedings of the Workshop on Advanced Analytics and Learning on Temporal Data*. Riva Del Garda, Italy.
- [17] Um, T. T., Pfister, F. M. J., Pichler, D., Muriel Lang, S. E., Fietzek, U., & Kulic, D. (2017, Nov.). Data augmentation of wearable sensor data for Parkinson's disease monitoring using convolutional neural networks. *Proceedings of the 19th ACM International Conference on Multimodal Interaction*. New York, USA.
- [18] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Proceeding of the Advances in Neural Information Processing Systems*. Montreal, Canada.
- [19] Mirza, M., & Osindero, S. (2014). Conditional generative adversarial networks. arXiv:1411.1784.
- [20] Zhu, X., Liu, Y., Qin, Z., & Li, J. (2017). Data augmentation in emotion classification using generative adversarial networks. arXiv:1711.00648, 2017.
- [21] Frid-Adar, M., Klang, E., Amitai, M., Goldberger, J. & Greenspan, H. (2018). Synthetic data augmentation using GAN for improved liver lesion classification. arXiv:1801.02385.
- [22] Anguita, D., Ghio, A., Oneto, L., Parra, X., & Reyes-Ortiz, J. L. (2013, April). A public domain dataset for human activity recognition using smartphones. *Proceedings of the 21st European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning* (pp. 24-26). Bruges, Belgium.
- [23] Zhao, Y., Yang, R., Chevalier, G. & Gong, M. (2017). Deep residual Bidir-LSTM for human activity recognition using wearable sensors. arXiv: 1708.08989.
- [24] Ronao, C. A., & Cho, S. B. (2016). Deep convolutional neural networks for human activity recognition with smartphone sensors. *Expert Systems with Applications*, *59(15)*, pp. 235-244.

Copyright © 2021 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (<u>CC BY 4.0</u>).



**Wen-Hui Chen** received the M.S. and Ph.D. degrees from the Department of Electrical Engineering, National Taiwan University in 1993 and 2000, respectively. He is currently a professor with the Graduate Institute of Automation Technology, National Taipei University of Technology, Taipei, Taiwan. His research interests are in the areas of artificial intelligence, machine learning, and smart living technology.



**Po-Chuan Cho** received a bachelor's degree from the Department of Mechanical Engineering, Tatung University, Taiwan in 2007 and a master's degree from the Graduate Institute of Automation Technology, National Taipei University of Technology, Taipei, Taiwan in 2009. He is currently working at Foxconn Technology Group and pursuing his Ph.D. degree at the Graduate Institute of Automation Technology, National Taipei University of Technology, Taipei, University of Technology, Taipei, Taiwan.