International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

HUMAN ACTIVITY RECOGNITION USING GYROSCOPE, ACCELERATION SENSORS AND DEEP LEARNING

Binh A. Nguyen[1], Hanh T. Pham[1], Ngoc T. Le[1], Truong V. Le[1], and Giao N. Pham[2] (corresponding author)

[1]ICT Department, FPT University, Hanoi, Vietnam
[2]Dept. of Computing Fundamentals, FPT University, Hanoi, Vietnam
E-mail: binhnase04865@fpt.edu.vn, hanhpthe130014@fpt.edu.vn, ngoclthe131028@fpt.edu.vn, truongtvhe130302@fpt.edu.vn, giaopn@fe.edu.vn

ABSTRACT

Nowadays, mobile phones are often integrated smart sensors for purpose to develop healthcare applications or applications on wearable devices. This is popularly become with the development of smart phones. This paper presents a solution to recognize human activities using mobile sensor that is integrated on smart phone and deep learning. The solution used a deep learning method that is based on Convolutional Neural Networks (CNNs) and Center Loss to recognize daily living activities. **Experimental results showed that the proposed solution achieves achieves the accuracy rate of 94.2% F1-Score on the smartphone dataset. This result is better than previous works.**

Keywords: *Huma Activity Recognition, CNNs, Smart Sensor, Deep Learning, Accelerometer, Gyroscope*

1. INTRODUCTION

Recently, due to the fact that human activity recognition brings many benefits, there are many researches to recognize human activities based on smart sensors that is integrated in smart phone or wearable [1, 2]. These researches focus on how to use the traditional techniques or machine learning/deep learning methods to recognize human activities [3, 4]. However, almost traditional methods or machine learning methods have a serious drawback, which is that input features have to be discriminative. It requires domain knowledge to design features. To solve this problem, researchers applied the power of deep learning technology for human activity recognition (HAR). Deep learning approaches have achieved significant benefit to directly extract discriminative features from raw data without the feature-engineering step**. In order to obtain discriminative features, the well-known loss function used to solve the problem of classification is Cross-Entropy loss. If just using Cross-Entropy loss, embedded features extracted by a deep learning model are not strong enough for recognizing classes that have similar patterns.**

In order to extract discriminative features for human activity recognition above, this paper presents a deep learning method to recognize human activities such as walking, standing, sitting, walking up, walking down, and laying based on gyroscope sensor and acceleration sensor that is integrated on smartphone. The proposed solution is based on CNNs and Center Loss. By combining Cross-Entropy Loss with Center-Loss, the proposed solution could enhance the quality of the prediction of similar activity classes. Dataset is collected from a smart phone placed on human. **To clarify the proposed solution, our paper is organized as follow: The related works are shown in Sec. 2; Sec. 3 describes the proposed solution in detail; The settings and experimental results of the proposed solution will be illustrated in Sec. 4; Finally, the conclusion of this work is explained in Sec. 6.**

2. RELATED WORKS

Cross-Entropy loss is used to calculate the similar between two probability distributions. It is well-known for the problem of classification. Let's consider the problems of k-class classification. Let $X \in R^d$ be the feature space and $Y = \{1, 2, \dots, c\}$ be the label space. In an idea scenario, a classifier is a function that maps input feature space to the label space$f: X \to R^c$. Cross-Entropy Loss's mathematics formulation [14] is defined as follows:

$$L(f(x_i, \theta), y_i) = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{c} y_{i,j} \log\left(f_j(x_i, \theta)\right) \qquad (1)$$

Where, θ is the parameters of the classfier, c is the number of classes.

To improve the discriminate of the embedded features, Center Loss minimizes the intra-class variations while keeping the features of different classes separable [5]. The mathematics formulation of Center Loss is defined as follows.

$$L_c = \frac{1}{2}\sum_{i=1}^{m}\left\|x_i - c_{y_i}\right\|_2^2 \qquad (2)$$

Where $x_i \in \mathrm{R}^{\mathrm{d}}$ is learned features of ith sample, $c_{y_i} \in \mathrm{R}^{\mathrm{d}}$ denotes the $\in \mathrm{R}^{\mathrm{d}}$ class center of deep features and m is the number of classes. The formulation effectively characterizes the intra-class variations. c_{y_i} is updated as the deep features changed. Updating the centers is performed based on mini-batch. In each iteration, the centers are calculated by averaging the features of the corresponding classes. To avoid large perturbations caused by mislabeled samples, α factor is used to control the learning rate of the centers [6]. The gradients of $L_c$with repect to x_i be computed as:

$$c_j = c_j - \alpha \times \frac{\partial L_c}{c_j} \quad (3) \qquad (4)$$

$$\frac{\partial L_c}{c_j} = \frac{\sum_{i=1}^{m} \delta(y_i = j) \times \left(c_{y_i} - x_i\right)}{1 + \sum_{i}^{m} \delta(y_i = j)} \quad (4) \qquad (5)$$

Where $\delta(y_i = j) = 1$ if the condition is satisfied and $\delta(y_i = j) = 0$ if not.

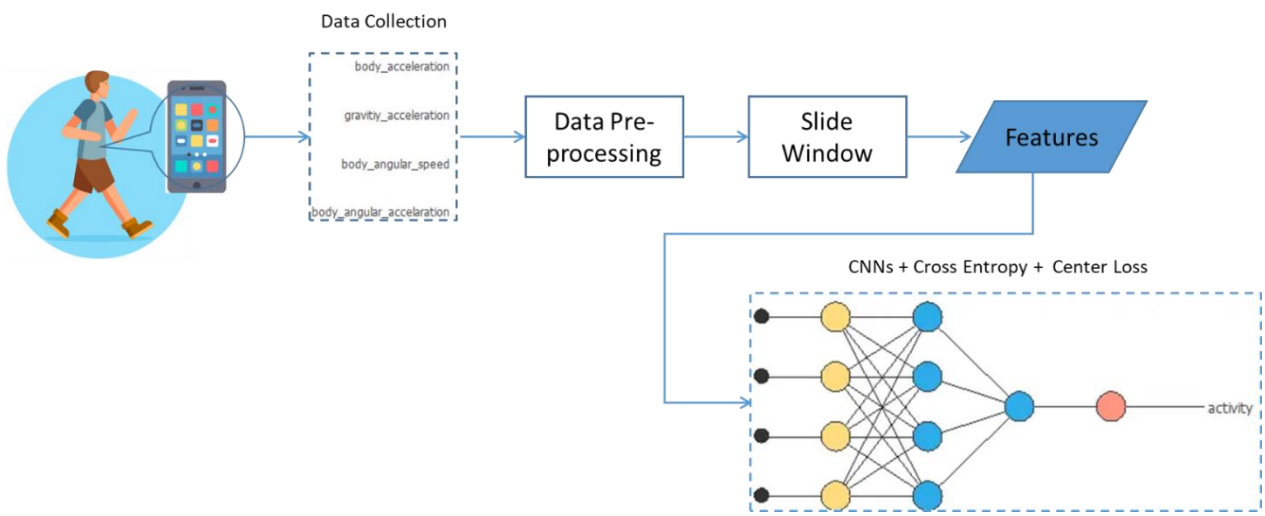
3. THE PROPOSED SOLUTION

The proposed solution is described in Fig. 1. Dataset is collected from gyroscope and acceleration sensors on Smartphone, and then performed the pre-processing step before applying Slide Window to extract features**. In the pre-processing step, a low-pass filter is used to de-noise the raw signals. In order to get samples for training and testing phases, we use a 2.56-second sliding window with 50% overlap to extract samples from the pre-processed signals.** Finally, features are used as the input of CNNs combined with Cross Entropy and Center Loss. By stacking multiple CNN layers, we have a novel deep learning architecture that has a good performance on Human Activity Recognition. One disadvantage of deep learning is over fitting, to prevent this phenomenon, we drop some CNN layers rapidly. By doing this, our architecture is simple and still enough capability to perform the task of classifying. The proposed network architecture is showed in Tab. 1.

TABLE I. OUR NETWORK

Input: 128× 6
Conv1d: 3×64, stride 1, padding same, LeakReLu
Conv1d: 3×64, stride 1, padding same, LeakReLu
MaxPooling: 2, stride: 2
Conv1d: 3×128, stride 1, padding same, LeakReLu
Conv1d: 3×128, stride 1, padding same, LeakReLu
Conv1d: 3×128, stride 1, padding same, LeakReLu
MaxPooling: 2, stride: 2
FC: 512, LeakReLu
FC: 256, LeakReLu
FC: 64, LeakReLu
FC: 6

FIGURE I. THE PROPOSED SOLUTION



4. EXPERIMENTAL RESULTS

The dataset consists of accelerometer and gyroscope tri-axis sensor data were collected from 30 volunteers within an age bracket of 19-48 years. Each person performed six activities (walking, walking up stair, walking down stair, sitting, standing, and laying) wearing a smartphone (Samsung Galaxy S V) on the waist. Using its embedded accelerometer and gyroscope sensors, 3-axial linear acceleration and 3-axial angular velocity were sampled at a constant rate of 50Hz as shown in Fig. 2, Tab. II, and Tab. III.

FIGURE II. DATASET

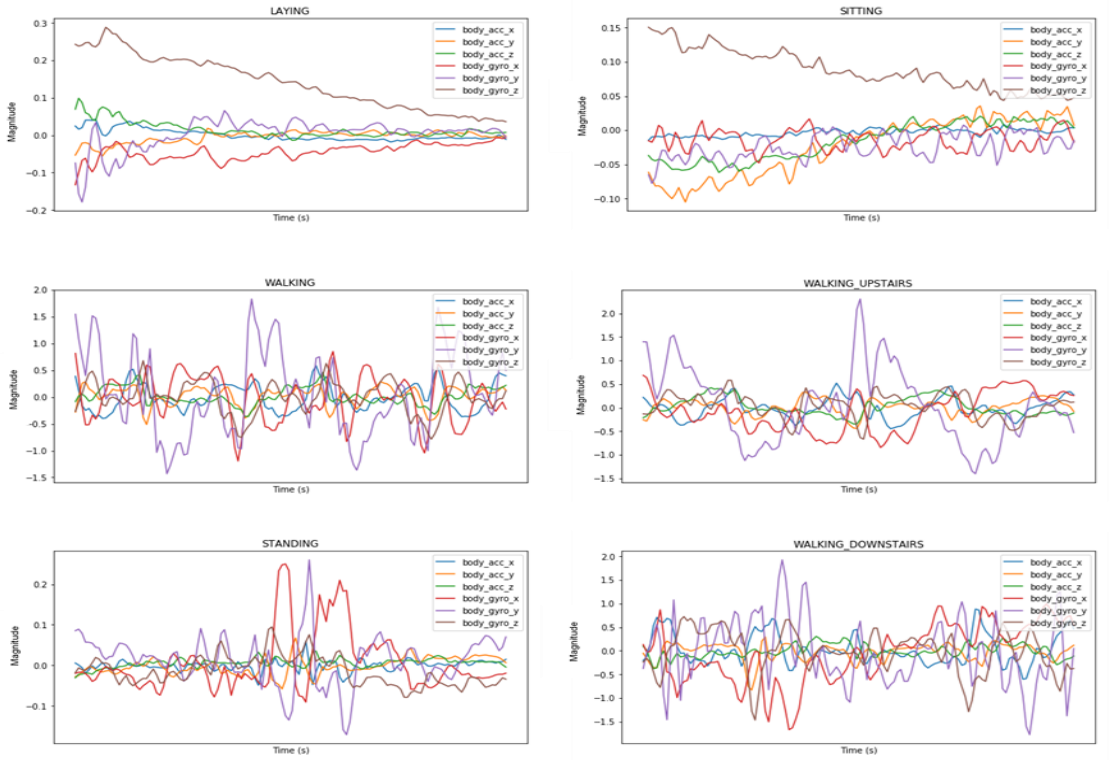


TABLE II: Dataset Information

Set	Number Samples
Training Set	5147
Validation Set	2205
Test Set	2947

TABLE III: Class distribution of the validation set

Activity	Number Samples	Proportion (%)
Sitting	904	0.18
Standing	960	0.19
Laying	976	0.19
Walking	866	0.17
Walking up stair	745	0.14
Walking down stair	696	0.14

TABLE IV: Class distribution of the training set

Activity	Number Samples	Proportion (%)
SITTING	904	0.18
STANDING	960	0.19
LAYING	976	0.19
WALIKING	866	0.17
WALKING_UPSTAIRS	745	0.14
WALKING_DOWNSTAIRS	696	0.14

To train our model, we set parameters as follow: Learning rate = 0.001, Batch Size =32, Number of epochs = 300, λ1=0.01, λ2=0.02, Center-Loss's learning rate=0.1, Optimizer =Adam Optimizer. From Tab. 5, Tab.6 and Tab. 7, we concluded that the proposed solution get better results than previous methods.

TABLE V: Comparison with recent studies in terms of F1-Score

Model	F1-Score (macro)
SVM [7]	0.893
CNN [12]	0.915
DeepConvLSTM [13]	0.910
Our Achitecture + Cross-Entropy Loss	**0.920**
Our Achitecture + Cross-Entropy Loss + Center Loss	**0.942**

TABLE VI: Comparison between two proposed methods on accuracy, precision and recall metrics

Model	Accuracy	Precision	Recall
Our Achitecture + Cross-Entropy Loss	0.916	0.922	0.919
Our Achitecture + Cross-Entropy Loss + Center Loss	**0.940**	**0.942**	**0.942**

TABLE VII: Comparison between two proposed methods on the recall of each class

Class	**Recall**	
	Our Achitecture + Cross-Entropy Loss	Our Achitecture + Cross-Entropy Loss + Center Loss
SITTING	**0.892**	0.884
STANDING	0.897	**0.930**
LAYING	0.840	**0.898**
WALIKING	0.988	**0.996**
WALKING_UPSTAIRS	0.910	**0.957**
WALKING_DOWNSTAIRS	**0.988**	0.986

5. CONSLUSION

The proposed approach can achieve high performance for the human activity recognition task. Specially, the method obtains 94.2% F1-Score on Smartphone Dataset, which outperforms some recent studies results. High performance is the result of using our CNN architecture with joint loss functions. The effects of incorporating the Center Loss which is expected to make embedded features of the similar activities become more discriminative is demonstrated in this paper. In the future, we still focus on increasing the quality of prediction by considering difference backbone architectures to produce good embedding features. Moreover, more efforts will be put on scaling up our method to tackle the problem of daily living activity classification based on multi-sensor data. Finally, to apply the method to the real world, there are some classes that are rarely occurred, therefore some techniques related to few-shot learning will be considered.

ACKNOWLEDGEMENT:

This work was supported by FPT University, Hanoi, Vietnam.

REFERENCES

[1] F. Ordóñez and D. Roggen, "Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition," Sensors, vol. 16, no. 1, p. 115, Jan. 2016.

[2] Du, Y., Lim, Y., and Tan, Y., “A Novel Human Activity Recognition and Prediction in Smart Home Based on Interaction.” Sensors (Basel, Switzerland) vol. 19,20 4474. 15 Oct. 2019.

[3] Russell Stewart and Stefano Ermon, “Label-free supervision of neural networks with physics and domain knowledge,” In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI’17). AAAI Press, pp. 2576–2582, 2017.

[4] C. Xu, D. Chai, J. He, X. Zhang and S. Duan, "InnoHAR: A Deep Neural Network for Complex Human Activity Recognition," IEEE Access, vol. 7, pp. 9893-9902, 2019.

[5] Wen Y., Zhang K., Li Z., Qiao Y., “A Discriminative Feature Learning Approach for Deep Face Recognition,” ECCV, Lecture Notes in Computer Science, vol. 9911, 2016.

[6] Karen Simonyan and Andrew Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” International Conference on Learning Representations, 2015.