Subject Independent Human Activity Recognition with Foot IMU Data

Xiaotong Zhang
Department of Computer Science and Engineering
Southern University of Science and Technology
Shenzhen, China
zhangxt@mail.sustech.edu.cn

Jin Zhang
Department of Computer Science and Engineering
Southern University of Science and Technology
Shenzhen, China
zhangj4@sustc.edu.cn

***Abstract*—Human activity recognition is a very active research area on pervasive computing and mobile health application. Many human activity systems based on inertial measurement unit (IMU) sensor data were proposed in the past few years. These systems mainly use IMU sensor placed on torsos and limbs to collect data and utilize supervised machine learning algorithms on sensor data. One main issue of these systems is that wearing multiple on-body IMU sensors may bring inconvenience to users' daily life. The other issue of these exiting methods is that an activity recognition model that is trained on a specific subject does not work well when being applied to predict another subject's activities since IMU sensor data always carry information that is specific to the human subject who conducts the activities. In our work, inspired by the principle of domain adaption, we proposed a new deep-learning activity recognition model based on an adversarial network which can remove the subject-specific information within the IMU activity data and extract subject-independent features shared by the data collected on different subjects. We also for the first time use data collected from insole based IMU sensors on 8 participants for 5 common activities to build a new real world human activity dataset which can minimize the inconvenience for users to wear. We conducted experiments with our new real-world dataset. Results show that our subject independent activity recognition model outperforms state-of-art supervised learning techniques and eliminates the effects of individual differences between subjects successfully. The average recognition accuracy under the leave-one-out (L1O) condition achieves 99.0% which is higher than the performance of traditional human activity recognition system based on CNNs.**

***Index Terms*—Human Activity Recognition, Foot-based IMU, Subject Independent**

I. Introduction

Human Activity Recognition (HAR) has become a very attracted research field of pervasive computing and wearable devices. It has a wide range of applications on mobile health in recent years due to the rapid updating of smart sensing devices and deep learning algorithms. It uses specific sensor data generated by different sensors placed on body of users or in the environment to predict which activity is being taken by users. Human activity recognition plays an important role in a wide range of real world application scenarios, such as health care [15], smart wearable device and fitness monitoring. For instance, recognition of users' activity can help to analyze users' exercise habits and infer users' health state.

Various human activity recognition systems have been proposed to extract useful features from raw sensor data and perform activity classification in the past decade [9] [11] [4]. Most of existing works using IMU sensors embedded in smart mobile devices such as wristbands, watches and other wearable devices to perform activity recognition. Many human activity recognition systems use multiple IMU sensors placed on different parts of subjects' body such as arm, waist, ankle and leg to collect data. This can capture the posture of various parts of the user's body which can improve the recognition accuracy effectively [2]. However, it is very inconvenient and uncomfortable for users to carry several on-body sensors in their daily life, thus, such human activity recognition system based on multiple on-body IMU sensor has an very limited real world application context. There are also some human activity recognition systems use single inertial sensor embedded in smart watch or smart phone to collect data [11] [9]. Though this kind of human activity recognition systems can solve the inconvenient problem in users' daily live, a single IMU sensor can only provide limited activity data of a single part of users' body that may reduce the final recognition accuracy. For instance, a human activity recognition system based on smart phone sensors may fail to work when users operate the phone.

To address this challenge, we design a human activity recognition system that uses insole-based IMU sensor to collect data, two IMU sensors are embedded into a pair of insoles which bring zero inconvenience to users. Users wear shoes with these insoles to perform daily activities and the IMU sensor in the insole can provide better performance, as the difference of most activities can be reflected in the foot motion.

In addition, our proposed system is a subject independent one which can eliminate the subject-specific information within the activity data and extract subject independent features from raw sensor data. Most of the existing methods are subject-specific activity recognition which requires labeled data for training but satisfactory results are hard to get when we use the subject-specific activity recognition model trained with specific subject data to predict another subject's activities. Since different subjects with different ages, genders, heights, weights, and body shapes affect the signals in different ways

978-1-7281-5212-7/19/$31.00 ©2019 IEEE

DOI 10.1109/MSN48538.2019.00054

even if they are performing the same activity. However, in the real world application scenario, it is almost impossible to label the activity data of every user before the device is distributed to the users. Our proposed subject independent human activity recognition system which uses labeled data of several other people to train and adapt the resulting trained model to new subjects can solve this problem. Our system can eliminate the subject-specific information within the activity data and extract subject independent features from raw sensor data by the inside deep learning model. In our system, we first use insole-based IMU sensor to collect activity data and all data are normalized, segmented and manually labeled. Then all processed data are input to our deep learning model to perform activity recognition. The kernel of our deep learning model is an adversarial network, which consists of three main parts: feature extractor, activity recognizer and subject discriminator. The feature extractor is a convolutional neural network which works with activity recognizer to classify human activities, meanwhile, tries to fool the subject discriminator to build a subject independent model.

Our main contribution in this article are as follows:

- We are the first to use insole-based IMU data to present a new real world human daily activity dataset and perform human activity recognition with this dataset.
- We proposed a new subject independent human activity recognition system based on an adversarial neural network and the principle of domain adaption.
- The experimental results show that our system can eliminate the effects of individual difference between subjects successfully achieving an 99.0% recognition accuracy under leave-one-out (L1O) condition.

The rest of this paper is organized as follows. We first discuss the related work in Section II. Then an overview of our human activity recognition system is provided in Section III. In Section IV, details of our methodology are introduced. We conduct a series of experiments on our new real world human activity dataset with our proposed system in Section V and conclude the article in Section VI.

II. Related Work

Human activity recognition is a widely studied research area of ubiquitous computing, human-computer interaction and human behavior analysis. Human activity recognition systems use signals generated by different kind of sensors such as body-worn inertial sensors, video, WiFi and etc. to recognize human behaviors. In this section, we review the existing works from the following three aspects.

A. Device free human activity recognition systems

Human activity recognition (HAR) has been widely studied in recent years. some researchers use wireless signals (e.g., ultrasound, WiFi, mmWave, visible light, etc.) to perform device-free human activity recognition. Many researchers propose human activity recognition system based on acoustic signals [3] [17]. Due to the Doppler effect, acoustic signals can get frequency shift when they reflect by human bodies. In addition, Acoustic signals in these systems can achieve frequency higher than 17 kHz, which is inaudible to most people [16], will not bring inconvenience to users' daily life. Some recent work [12] [13] propose to recognize human activities or gestures by analyzing unique continuous shadow maps produced by human activity. The mmWave with short wavelength can create stronger reflection from small objects since wireless signals cannot easily bypass objects larger than wavelength [20]. However, the major challenge of device free human activity recognition is that the wireless signals arriving at the receiving devices usually carry substantial information that is specific to the environment where the activities are recorded [8]. Therefore, the recognition accuracy of HAR using wireless signals are usually not satisfied. Compared with device free HAR systems, using body-worn inertial sensors can be a better choice.

B. Device based human activity recognition systems

Body-worn inertial measurement unit (IMU) sensors are one of the most common modalities in human activity recognition [19], such as accelerometer, magnetometer, and gyroscope. Since IMU is widely embedded in smart devices such as smart phones, watches, bands and glasses, it has a wide real world application scenario. In 1990s, pioneering researchers started using accelerometer data to perform human activity recognition [5]. However, the most cited work on HAR which can produce satisfactory results using acceleration data collected from five small biaxial accelerometers placed on different parts of subjects' body and various data mining algorithms was made in 2004 [2]. In this work, the results showed that the best performance recognizing daily activities with an overall accuracy rate of 84% which is a good performance in those years. Although the use of multiple inertial sensors that are simultaneously placed a human body can improve the recognition accuracy, carrying multiple on-body sensors like this work can bring inconvenience to users undoubtedly. Magnetometer sensors are used together in [1] to perform HAR. However, the price of magnetometers are always expensive, so very few smart wearable devices have this sensor which means human activity recognition systems rely on magnetometers have restricted application scenario.

Motivated by the inconvenience caused by multiple on-body sensor system, some human activity recognition systems using single sensor are proposed. [11] uses cell phone accelerometers to perform human activity recognition. With the deep learning model, [4] uses a single accelerometer to perform HAR and reach an average accuracy of 93.8% without any feature extraction methods. Though human activity recognition system using single sensor can reduce the wearing burden of users, these systems may be unable to discriminant similar activities, such as downstairs and upstairs [14] and the behavior of operating mobile phone in users daily life can also affect the recognition performance.

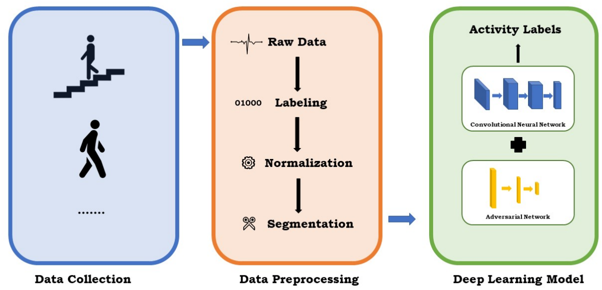

Fig. 1. Human Activity Recognition System

C. Domain adaption on human activity recognition

Technically, our work is based on the idea of domain adaption to solve the subject-specific problem in human activity systems. The main challenge of domain adaption HAR system is to transfer the available labeled data from a set of existing activities in one domain to help recognize the activities in another different but related domain [23]. With the development of deep learning algorithm, many domain adaption techniques are proposed [6] [7] [18], [8] develops an environment independent device free human activity recognition system that can remove the environment and subject-specific information contained in the activity data and extract environment/subject-independent features shared by the data collected on different subjects under different environments based on these approaches. Inspired by these applications of domain adaption in human activity recognition, we design a novel subject independent human activity system based on an adversarial network.

III. System Overview

In this section, we provide an overview of our proposed human activity recognition system which consists of three main parts: data collection, data preprocessing and deep learning model. The structure of our system is shown in Fig. 1.

- **Data Collection**. In our work, we use insole-based IMU sensors to collect human activity data since foot movement is closely related to human activities and wearing shoes with a pair of insoles embedded with IMU sensor will not disturb subjects' daily life. Our system first collects human activity data conducted by different subjects to build a new real world human activity dataset for the further experiments. All data are collected in non-laboratory environment which means subjects have no restrictions on the places and ways for them to take activities.
- **Data Preprocessing**. Raw IMU sensor data are divided into two parts, one part of data are manually labeled with activity labels while the other part of data are unlabeled in the data preprocessing part. However, all collected data are manually labeled with subject labels since the subject discriminator in the adversarial network needs subject labels to train. There are two kinds of data generated by IMU sensors, acceleration data and angular velocity data, they have different variation ranges. Hence, we must normalize the raw sensor data in addition to labeling them to prevent the input of deep learning model from having different weights in the neural network. Besides labeling and normalization, sensor data will be split to short segments with overlaps as the input of activity recognition model.
- **Deep Learning Model**. After the data preprocessing step, the collected IMU sensor data is still too complicated to be analyzed directly. Therefore, it is difficult for traditional machine learning algorithms to classify activities with such complex data. However, deep learning technologies which have been proved effective in mining useful information from complex data can be used to solve this problem. In particular, we proposed a deep learning model based on an adversarial neural network and the idea of domain adaption, to predict users' activity. This model consists of three main parts: feature extractor, activity recognizer and subject discriminator. We used a convolutional neural network which can effectively mine useful features from data as the feature extractor. Then the fully connected layer in the activity recognizer uses these features to predict activity labels. Meanwhile, the subject discriminator is responsible for subject classification which can help to remove the subject-specific information in the sensor data and extract common features shared with data from different subjects. Our deep learning model not only use labeled data to train but also can take advantage of unlabeled data in the training process which can improve the recognition performance.

IV. Methodology

An overview of our proposed deep learning model is shown in Fig. 2. In this paper, our work focuses on the practical application context: the labeled data and unlabeled data are come from different subjects but the human activity recognition system can work well on those subjects who contributed unlabeled data. Therefore, this real world application context requires that our proposed deep learning model must be able to learn transferable features for different subjects and eliminate subject-specific information with data.

A. Model Input

The proposed deep learning model can recognize human activities with preprocessed insole-based IMU sensor data. We provide a general description of the model inputs in the subsection.

All raw sensor data $\boldsymbol{x_{raw}}$, are first normalized as follows:

$$\boldsymbol{x_{Norm}} = \frac{\boldsymbol{x_{raw}}}{x_{Max} - x_{Min}}, \tag{1}$$

where $\boldsymbol{x_{Norm}}$ is the normalization result, x_{Max} is the maximum value of the raw data and x_{Min} is the minimum value. After normalization, $\boldsymbol{x_{Norm}}$ are split to short segments with

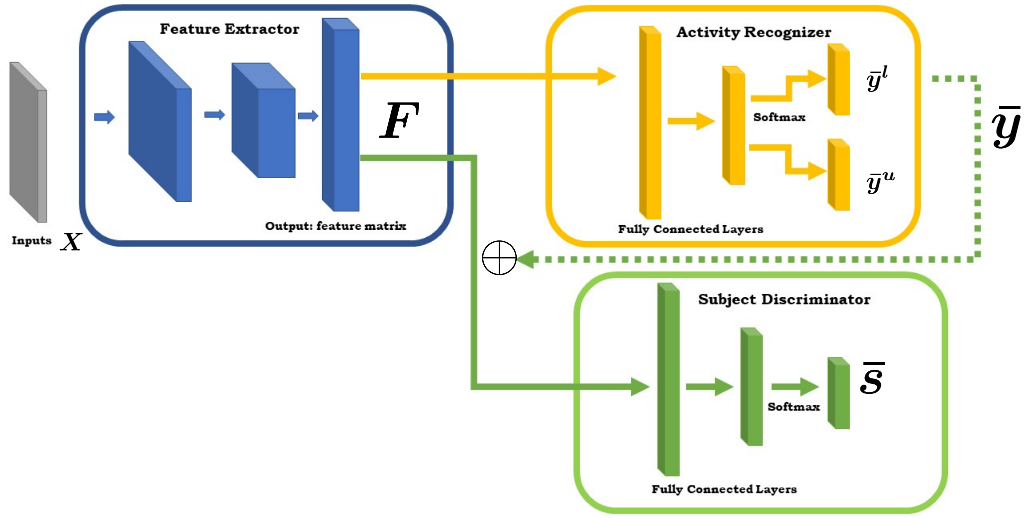

Fig. 2. Deep Learning Model

overlaps, the segment results are $\boldsymbol{x_{Segment}}$. Then, $\boldsymbol{x_{Segment}}$ are labeled manually to create the final input matrix $\boldsymbol{X}$.

According to the principle of domain adaption, we divide all subjects into two parts: source subjects and target subjects. Source subjects are those who provide labeled activity data while target subjects are those who provide unlabeled data. The goal of our deep learning model is to predict activities conducted by target subjects. The input to the proposed model has two parts, labeled activity data and unlabeled ones, represented by $\boldsymbol{X^l}$ and $\boldsymbol{X^u}$ respectively and the total input data are represented by $\boldsymbol{X}$. Each labeled data X_i^l of $\boldsymbol{X^l}$ has a true activity label $y_i^l \in \boldsymbol{y}$, where $\boldsymbol{y}$ is the set of all activities. Besides, each data X_i of $\boldsymbol{X}$ has a true subject label $S_i \in \boldsymbol{S}$, where $\boldsymbol{S}$ is the set of all the source and target subjects. Therefore, the inputs of our proposed model are all activity data $\boldsymbol{X}$, subject label matrix $\boldsymbol{s}$ of $\boldsymbol{X}$ and activity label matrix $\boldsymbol{y^l}$ of $\boldsymbol{X^l}$ (ground truth).

B. Feature Extractor

Due to its strong ability on data mining, we choose convolutional neural networks which are widely used in human activity recognition systems as the feature extractor [21]. There are three stacks in our CNNs, each stack has a convolutional layer, a 2×2 max pooling layer to reduce the size of representation, a rectified linear unit (ReLU) layer to create nonlinearity, and a normalization layer. In the convolutional layer of the first stack, there are 32 filters with the size of 5×5, the second one has 64 filters with the size of 5×5 and 1024 filters with the size of 5×5 in the third one. Let $\boldsymbol{\theta_f}$ be the set of parameters in CNNs, and the output feature matrix $\boldsymbol{F}$ of the feature extractor can be represented as follows:

$$\boldsymbol{F} = CNN(\boldsymbol{X}, \boldsymbol{\theta_f}). \tag{2}$$

C. Activity Recognizer

The activity recognizer consists of two fully connected layer followed by a softplus activation function. The output vector of the feature extractor $\boldsymbol{F}$ is input to the activity recognizer to predict the label of human activities. The output $\boldsymbol{Z}$ of the first fully connected layer is as follows:

$$\boldsymbol{Z} = Softplus(\boldsymbol{W_{a1}}\boldsymbol{F} + \boldsymbol{b_{a1}}), \tag{3}$$

and output $\bar{\boldsymbol{y}}$ of the activity recognizer is as follows:

$$\bar{\boldsymbol{y}} = Softmax(\boldsymbol{W_{a2}}\boldsymbol{Z} + \boldsymbol{b_{a2}}), \tag{4}$$

where $\boldsymbol{W_{a1}}$,$\boldsymbol{W_{a2}}$,$\boldsymbol{b_{a1}}$,$\boldsymbol{b_{a2}}$ are parameters to be learned, $\boldsymbol{W_{a1}}$ and $\boldsymbol{b_{a1}}$ are the weight matrix and bias matrix in the first fully connected layer while $\boldsymbol{W_{a2}}$ and $\boldsymbol{b_{a2}}$ are those in the second layer, let $\boldsymbol{\theta_y}$ be the set of these parameters. The output $\bar{\boldsymbol{y}}$ including $\bar{\boldsymbol{y}}^{\boldsymbol{l}}$ and $\bar{\boldsymbol{y}}^{\boldsymbol{u}}$ represents the predicted probabilities of all input data $\boldsymbol{X}$ where $\bar{\boldsymbol{y}}^{\boldsymbol{l}}$ denotes the output of labeled data and $\bar{\boldsymbol{y}}^{\boldsymbol{u}}$ denotes the output of unlabeled data since the input $\boldsymbol{X}$ also includes both labeled data $\boldsymbol{X^l}$ and unlabeled data $\boldsymbol{X^u}$.

Cross entropy can be used to calculate the loss between the prediction result $\bar{\boldsymbol{y}}^{\boldsymbol{l}}$ and the ground truth $\boldsymbol{y^l}$ for the labeled data:

$$L_a = -\frac{1}{|\boldsymbol{X^l}|}\sum_{i=1}^{|\boldsymbol{X^l}|}\sum_{n=1}^{N} \boldsymbol{y_{in}^l} log(\bar{\boldsymbol{y}}_{\boldsymbol{in}}^{\boldsymbol{l}}), \tag{5}$$

where $|\boldsymbol{X^l}|$ is the number of input labeled data and N is the number of activities. The traditional human activity system using CNNs only optimize (3) to learn model parameters and conducted activity label prediction. However unlabeled data can also help to improve recognition performance in our system. The cross entropy loss function for unlabeled data can also be calculated as follows [8]:

$$L_u = -\frac{1}{|\boldsymbol{X^u}|}\sum_{i=1}^{|\boldsymbol{X^u}|}\sum_{n=1}^{N}\boldsymbol{\bar{y}^u_{in}}log(\boldsymbol{\bar{y}^u_{in}}), \tag{6}$$

where $|\boldsymbol{X^l}|$ is the number of input unlabeled data. We can increase the recognition accuracy by minimizing (3) and (4).

D. Subject Discriminator

In the real world application context, it is impossible to get labeled data from the users first use this system, thus, a subject independent human activity recognition system must have the ability to extract features shared by all subjects without subject-specific information. In our work, domain adaption technique is used to achieve the goal of extract subject independent features from all activity data. When the target domains (target subjects in our work) are fully unlabeled, the technique is called unsupervised domain adaptation [6]. In this paper, we utilize the technique of unsupervised domain adversarial training [6] and the idea of domain independent human activity system [8] to remove the subject-specific information within activity labels by using unlabeled data. of activities.

The input of the subject discriminator $\boldsymbol{C}$ is the concatenation of output matrix $\boldsymbol{F}$ of feature extractor and the prediction matrix $\boldsymbol{\bar{y}}$ as follows [22]:

$$\boldsymbol{C} = \boldsymbol{F} \oplus \boldsymbol{\bar{y}}, \tag{7}$$

where $\oplus$ is the concatenation operation. $\boldsymbol{F}$ contains both subject independent and subject-specific features which are necessary for the task of extracting common features shared with all subjects, Therefore, we need to put the whole feature matrix $\boldsymbol{F}$ into the concatenation in stead of using subject independent features only.

The subject discriminator consists of two fully connected layer followed by a softplus activation function which is exactly the same structure as activity recognition. The output $\boldsymbol{K}$ of the first fully connected layer in subject discriminator is as follows:

$$\boldsymbol{K} = Softplus(\boldsymbol{W_{s1}}\boldsymbol{C} + \boldsymbol{b_{s1}}), \tag{8}$$

and the output $\boldsymbol{\bar{S}}$ of the subject discriminator is as follows:

$$\boldsymbol{\bar{S}} = Softmax(\boldsymbol{W_{s2}}\boldsymbol{K} + \boldsymbol{b_{s2}}), \tag{9}$$

where $\boldsymbol{W_{s1}}$,$\boldsymbol{W_{s2}}$,$\boldsymbol{b_{s1}}$,$\boldsymbol{b_{s2}}$ are parameters to be learned, $\boldsymbol{W_{s1}}$ and $\boldsymbol{b_{s1}}$ are the weight matrix and bias matrix in the first fully connected layer while $\boldsymbol{W_{s2}}$ and $\boldsymbol{b_{s2}}$ are those in the second layer, let $\boldsymbol{\theta_s}$ be the set of these parameters. Cross entropy is also used to calculate the loss between the prediction subject label $\boldsymbol{\bar{S}}$ and the true subject label matrix $\boldsymbol{S}$ for the labeled data:

$$L_s = -\frac{1}{|\boldsymbol{X}|}\sum_{i=1}^{|\boldsymbol{X}|}\sum_{j=1}^{|\boldsymbol{S}|}\boldsymbol{S_{ij}}log(\boldsymbol{\bar{S}_{ij}}), \tag{10}$$

where $|\boldsymbol{X}|$ represents the number of all subjects and $\boldsymbol{S_{ij}}$ is the true one-hot vector of subject labels inside $\boldsymbol{S}$.

E. Training Step

We have already got three loss functions so far: the prediction loss function of both labeled and unlabeled data in the activity recognition $\boldsymbol{L_a}$ and $\boldsymbol{L_u}$ and the prediction loss of subject labels in the subject discriminator $\boldsymbol{L_s}$. The goal of the activity recognizer is to maximize the activity label prediction accuracy by minimizing $\boldsymbol{L_a}$ and $\boldsymbol{L_u}$. Meanwhile, the goal of the subject discriminator is to maximize the subject label prediction accuracy by minimizing $\boldsymbol{L_s}$, but our ultimate goal is learning subject independent features of activities which means the subject discriminator should fail to predict subject labels. To address this contradiction, we use the opposite number of $\boldsymbol{L_s}$ in the final loss function:

$$L = L_a + \alpha L_u - \beta L_s, \tag{11}$$

where α and β are predefined hyper-parameters.

In our work, the target of the training process is to seek parameters $\boldsymbol{\theta_f}$, $\boldsymbol{\theta_y}$, $\boldsymbol{\theta_s}$ that deliver a saddle point of $\boldsymbol{L}$. At this saddle point, the parameters $\boldsymbol{\theta_f}$ minimize the activity label prediction loss while maximize the subject label prediction loss, the parameters $\boldsymbol{\theta_y}$ minimize the activity label prediction loss and parameters $\boldsymbol{\theta_s}$ minimize the subject label prediction loss [6]. This saddle point can be found by the following stochastic steps:

$$\boldsymbol{\theta_f} \leftarrow \boldsymbol{\theta_f} - \mu\frac{\partial \boldsymbol{L}}{\partial \boldsymbol{\theta_f}}, \tag{12}$$

$$\boldsymbol{\theta_y} \leftarrow \boldsymbol{\theta_y} - \mu\frac{\partial \boldsymbol{L_a} + \partial \alpha \boldsymbol{L_u}}{\partial \boldsymbol{\theta_y}}, \tag{13}$$

$$\boldsymbol{\theta_s} \leftarrow \boldsymbol{\theta_s} - \mu\frac{\partial \boldsymbol{L_s}}{\partial \boldsymbol{\theta_s}}, \tag{14}$$

where μ is the learning rate.

V. Evaluation

In this section, we conducted experiments on our proposed new real world insole-based IMU dataset to evaluate the performance of our proposed human activity recognition system.

A. Baseline Model

To evaluate the performance of our proposed human activity recognition system, we use a traditional subject-specific activity recognition system based on CNNs to compare with our system. The CNNs are exactly the same as the feature extractor in our deep learning model which means we want to compare the performance of the system with and without the subject discriminator.

B. Dataset

In our proposed new real world human daily activity dataset, we employ 8 volunteers to conducted 5 basic daily activities: standing, walking, running, laying and walking downstairs. Each subject performs five minutes of each activity. Since several seconds are enough for people to complete a whole body motion such as a pace in walking, we can collect hundreds of complete body motion data in five minutes, which

TABLE I
BASIC INFORMATION OF 8 SUBJECTS

Subject	Gender	Height (cm)	Weight (kg)	Age
P1	Male	183	90	22
P2	Male	180	87	24
P3	Male	178	75	20
P4	Male	182	65	26
P5	Male	168	60	32
P6	Male	172	67	22
P7	Male	175	73	25
P8	Male	173	59	22

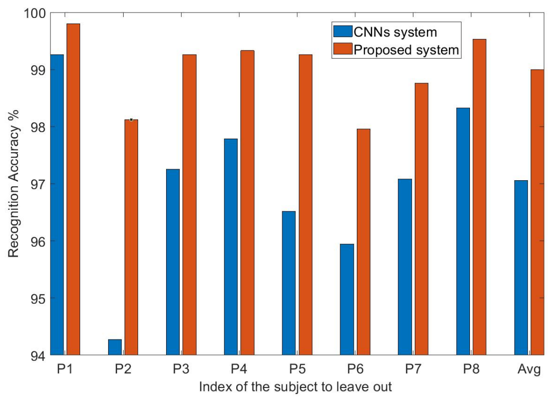

Fig. 3. Recognition Accuracy of the Proposed Model under L1O Condition

is sufficient for our training step. [1] also use five minutes as the data collection duration for each activity. The basic information of 8 subjects are shown in Table I.

The sampling rate of our IMU sensor is 100 Hz. Therefore, total $(100Hz \times 60seconds \times 5minutes \times 5activities \times 8subjects =)$ 1200000 samples activity data are collected. In particular, all activity data are collected under non-laboratory environment which means subjects can perform activities in any places under none constrained condition. For instance, subjects can conduct running activity anywhere at any speed they want. This is to simulate the real world application context as much as possible.

Retrieving important and useful information from continuous stream of sensor data is a difficult issue for continuous activity and motion recognition [10]. Therefore, we segment the data every 200 samples ($N = 200$) with 50 samples overlap which represents the human activity of 2 seconds in the real world. After segmentation and manually labeling process, the insole-based IMU activity dataset is prepared already.

C. Performance

In our work, we use activity recognition accuracy as the evaluation of the proposed system. We first use leave-one-out (L1O) approach to conduct experiments which the number of target subject is one and the number of source subject is seven. Since there are 8 subjects in total, the L1O experiments are conducted for 8 times. The results of recognition accuracy

TABLE II
CONFUSION MATRIX OF SUBJECT 2 USING CNNS SYSTEM

Activity Label	Standing	Laying	Walking	Running	Walking downstairs
Standing	597	0	0	0	0
Laying	0	597	0	0	0
Walking	0	0	545	2	50
Running	0	0	2	595	0
Walking downstairs	0	0	105	12	480

TABLE III
CONFUSION MATRIX OF SUBJECT 2 USING PROPOSED SYSTEM

Activity Label	Standing	Laying	Walking	Running	Walking downstairs
Standing	597	0	0	0	0
Laying	0	597	0	0	0
Walking	0	0	572	0	25
Running	0	0	0	597	0
Walking downstairs	0	0	30	1	566

under L1O condition is shown in Fig. 3. From Fig. 3, we can observe that for all eight subjects, the recognition accuracy is higher than the CNNs system, and the average accuracy of our proposed system is 98.92% which is a very satisfied performance.

To observe the details of recognition results, the confusion matrixs of subject 2 shown in Table II and Table III. The reason why we choose subject 2 is that the recognition results of this subjects are relative bad inside all 8 people. From the confusion matrix, we can conclude that standing, laying and running activity have almost 100 % recognition accuracy. Walking and walking downstairs are sometimes confuse with each other. However, we consider this a “correct” error, since there are many platforms between two set of staircases, and the data subjects collect during these platforms are exactly the same as the walking activity data. The difference between the confusion matrix of the CNNs system and the one of our proposed system shows that our system can greatly alleviate the confusion problem.

To further demonstrate the performance of our system, We gradually increase the number of target subjects from 1 to 4 (choose fewer subjects’ data to train gradually). The recognition accuracy under this condition is shown in Fig. 4. We can observe from Fig. 4 that, the recognition accuracy of the CNNs model and our proposed model are both going down with the decreasing number of source subjects. However, our proposed system can utilize unlabeled data of the target subjects while the CNNs system only input labeled data. This fundamental difference can make our proposed system a adaptive one, so our system can keep higher recognition accuracy with the decreasing number of source subjects but the CNNs will lose its performance gradually. This again validates that our proposed human activity recognition system is able to remove subject-specific information and extract environment independent features from unlabeled data.

VI. CONCLUSION

In this paper, we proposed a new effective subject independent human activity recognition system. Our proposed

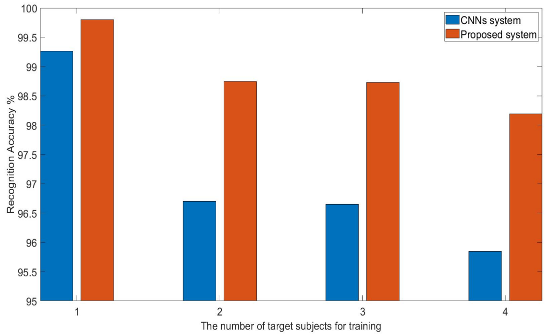

Fig. 4. Recognition Accuracy of the Proposed Model with Different Number of Source Subjects

system consists of three main parts: data collection, data preprocessing and deep learning model which has three main components (feature extractor, activity recognizer and subject discriminator). We also for the first time use insole-based IMU sensors to present a new human activity dataset. The results show that our proposed system can remove the subject-specific information within data and learn common features of activities achieving an 99.0% average recognition accuracy under the leave-one-out (L1O) condition. It is worth mentioning that more sensors can be integrated into the insole in the future such as pressure sensors, temperature sensors, which can bring more functionality.

ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China under Grant No. 61701216 and Shenzhen Science, Technology and Innovation Commission Basic Research Project under Grant No. JCYJ201805071 81527806, Grant No. JCYJ 20160531190935987 and Grant No. JCYJ 20160531191011045.

REFERENCES

[1] Kerem Altun, Billur Barshan, and Orkun Tunçel. Comparative study on classifying human activities with miniature inertial and magnetic sensors. *Pattern Recognition*, 43(10):3605–3620, 2010.

[2] Ling Bao and Stephen S Intille. Activity recognition from user-annotated acceleration data. In *International conference on pervasive computing*, pages 1–17. Springer, 2004.

[3] Ke-Yu Chen, Daniel Ashbrook, Mayank Goel, Sung-Hyuck Lee, and Shwetak Patel. Airlink: sharing files between multiple devices using in-air gestures. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 565–569. ACM, 2014.

[4] Yuqing Chen and Yang Xue. A deep learning approach to human activity recognition based on single accelerometer. In *2015 IEEE International Conference on Systems, Man, and Cybernetics*, pages 1488–1492. IEEE, 2015.

[5] F Foerster, M Smeja, and J Fahrenberg. Detection of posture and motion by accelerometry: a validation study in ambulatory monitoring. *Computers in Human Behavior*, 15(5):571–583, 1999.

[6] Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. *arXiv preprint arXiv:1409.7495*, 2014.

[7] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. *The Journal of Machine Learning Research*, 17(1):2096–2030, 2016.

[8] Wenjun Jiang, Chenglin Miao, Fenglong Ma, Shuochao Yao, Yaqing Wang, Ye Yuan, Hongfei Xue, Chen Song, Xin Ma, Dimitrios Koutsonikolas, et al. Towards environment independent device free human activity recognition. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*, pages 289–304. ACM, 2018.

[9] Adil Mehmood Khan, Young-Koo Lee, Sungyoung Y Lee, and Tae-Seong Kim. A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer. *IEEE transactions on information technology in biomedicine*, 14(5):1166–1172, 2010.

[10] Narayanan C Krishnan, Colin Juillard, Dirk Colbry, and Sethuraman Panchanathan. Recognition of hand movements using wearable accelerometers. *Journal of Ambient Intelligence and Smart Environments*, 1(2):143–155, 2009.

[11] Jennifer R Kwapisz, Gary M Weiss, and Samuel A Moore. Activity recognition using cell phone accelerometers. *ACM SigKDD Explorations Newsletter*, 12(2):74–82, 2011.

[12] Tianxing Li, Qiang Liu, and Xia Zhou. Practical human sensing in the light. In *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services*, pages 71–84. ACM, 2016.

[13] Tianxing Li, Xi Xiong, Yifei Xie, George Hito, Xing-Dong Yang, and Xia Zhou. Reconstructing hand poses using visible light. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(3):71, 2017.

[14] Uwe Maurer, Asim Smailagic, Daniel P Siewiorek, and Michael Deisher. Activity recognition and monitoring using multiple sensors on different body positions. Technical report, CARNEGIE-MELLON UNIV PITTSBURGH PA SCHOOL OF COMPUTER SCIENCE, 2006.

[15] Kirsten KB Peetoom, Monique AS Lexis, Manuela Joore, Carmen D Dirksen, and Luc P De Witte. Literature review on monitoring technologies and their outcomes in independently living elderly people. *Disability and Rehabilitation: Assistive Technology*, 10(4):271–294, 2015.

[16] A Rodríguez Valiente, A Trinidad, JR García Berrocal, C Górriz, and R Ramírez Camacho. Extended high-frequency (9–20 khz) audiometry reference thresholds in 645 healthy subjects. *International journal of audiology*, 53(8):531–545, 2014.

[17] Wenjie Ruan, Quan Z Sheng, Lei Yang, Tao Gu, Peipei Xu, and Longfei Shangguan. Audiogest: enabling fine-grained hand gesture detection by decoding echo signal. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 474–485. ACM, 2016.

[18] Rui Shu, Hung H Bui, Hirokazu Narui, and Stefano Ermon. A dirt-t approach to unsupervised domain adaptation. *arXiv preprint arXiv:1802.08735*, 2018.

[19] Jindong Wang, Yiqiang Chen, Shuji Hao, Xiaohui Peng, and Lisha Hu. Deep learning for sensor-based activity recognition: A survey. *Pattern Recognition Letters*, 119:3–11, 2019.

[20] Teng Wei and Xinyu Zhang. mtrack: High-precision passive tracking using millimeter wave radios. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, pages 117–129. ACM, 2015.

[21] Ming Zeng, Le T Nguyen, Bo Yu, Ole J Mengshoel, Jiang Zhu, Pang Wu, and Joy Zhang. Convolutional neural networks for human activity recognition using mobile sensors. In *6th International Conference on Mobile Computing, Applications and Services*, pages 197–205. IEEE, 2014.

[22] Mingmin Zhao, Shichao Yue, Dina Katabi, Tommi S Jaakkola, and Matt T Bianchi. Learning sleep stages from radio signals: A conditional adversarial architecture. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 4100–4109. JMLR. org, 2017.

[23] Vincent Wenchen Zheng, Derek Hao Hu, and Qiang Yang. Cross-domain activity recognition. In *Proceedings of the 11th international conference on Ubiquitous computing*, pages 61–70. ACM, 2009.