IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 12, 2019, accepted November 26, 2019, date of publication December 2, 2019,
date of current version December 16, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2956952

WiFi CSI-Based Human Activity Recognition Using Deep Recurrent Neural Network

JIANYANG DING AND YONG WANG, (Member, IEEE)
School of Telecommunications Engineering, Xidian University, Xi'an 710071, China

Corresponding author: Jianyang Ding (jyding@stu.xidian.edu.cn)

This work was supported in part by the Natural Science Foundation of China under Grant 61671346, and in part by the 111 Project under Grant B08038.

ABSTRACT Human activity recognition based on channel state information (CSI) using commercial WiFi devices plays an increasingly important role in many applications, such as smart home and interactive games. In this paper, we propose a WiFi CSI based human activity recognition approach using deep recurrent neural network (HARNN). HARNN mainly exploits four key techniques to recognize different human activities. HARNN firstly constructs a novel two-level decision tree for using two environment variation statistics efficiently. Meanwhile, a linear regression method is also introduced to seek for the optimal parameter for the designed decision tree. Depending on this, the decision tree is used to sense indoor environment variation, and then detect whether there is any human activity occurring in a target area. In addition, a noise removal mechanism is devised to eliminate the influence of random noise derived from indoor environments. Then, to characterize various human activities, two representative features are extracted from different statistical profiles, including channel power variation (CPV) and time-frequency analysis (TFA). Finally, a recurrent neural network (RNN) model is utilized to recognize different human activities by leveraging the extracted representative features above. According to the above steps, the proposed HARNN could establish a robust relationship between human activities and WiFi CSI compared with most of the existing WiFi CSI based approaches. The proof-of-concept prototype of HARNN is implemented on a set of commercial WiFi devices, and its overall performance is evaluated in several typical indoor environments. The experimental results demonstrate that HARNN can achieve better recognition performance compared with some benchmark approaches.

INDEX TERMS Human activity recognition, commercial WiFi, CSI, CPV, TFA, RNN.

I. INTRODUCTION

Human activity recognition is increasingly popular in pervasive computing and human-computer interaction, and can support a wide variety of applications, such as health care [1], smart home [2], augmented reality, and building surveillance [3], etc. In order to recognize various human activities, a number of sensors have been employed in the previous works, such as cameras based approaches [4], wearable sensors based approaches [5], ambient devices based approaches [6], [7], and smartphones based approaches [8]. Cameras based approaches are intended to recognize different human activities by capturing image and then performing matching. The merit of cameras based approaches is the capability of detecting and recognizing different human activities effectively. Meanwhile, these approaches suffer from some issues especially with regard to the illumination condition, dead angle, personal privacy concern, and high-energy consumption. In addition, wearable sensors are also popular for human activity recognition with high accuracy. However, these approaches require the sensors to be attached to the human body, which is very inconvenient and obstructive. Additionally, ambient devices based approaches generally include pressure sensors and vibration sensors, and measure changes in the ambience to detect human activity. However, other sources of pressure or vibration in indoor environments can also cause false alarms in these approaches. Another widely used device for human activity recognition is the modern smartphones. Due to many sensors, such as accelerator, gyroscope, and barometer, embedded in smartphones, they can be view as human activities sensing platforms for recognition. Certainly, if users forget to take their phones or

The associate editor coordinating the review of this manuscript and approving it for publication was Zhen Ren.

lose their smartphones anywhere, the human activity recognition would terminate naturally. Meanwhile, their personal information security would be threatened seriously in smartphones.

Considering that human activities between a pair WiFi transmitter and receiver could affect the characteristics of WiFi signal, human activity recognition can be feasible by using WiFi signal. Besides, due to low cost, no invasion, and wide deployment of commercial WiFi devices in indoor environments, human activity recognition using WiFi signal can become reality. More importantly, these passive human activity recognition approaches using WiFi signal do not require users to take any dedicated sensor or device. Additionally, the commonly used signal for WiFi is the received signal strength indication (RSSI) which has been widely used for indoor localization [9] and human activity recognition [10]–[12]. In [10], S. Sigg *et al*. designed a RSSI-based activity recognition system for the mobile phones. In general, it collected WiFi RSSI data at mobile phones to monitor single-handed gestures. In [11], Y. Gu *et al*. presented an online activity recognition system, which explored WiFi ambient signals for RSSI fingerprint of different human activities. In [12], Y. Gu *et al.* used data mining techniques to abstract footprints of different human activities on the RSSI data, and designed a novel fusion algorithm to distinguish different human activities. Although these approaches have the advantages of passive recognition and easy deployment, the performance of activity recognition is limited because the RSSI information could be influenced heavily by the severe multipath and random noise in indoor environments.

Instead of RSSI, there are increasing applications using channel state information (CSI) [13]–[15] of WiFi, which reveals a set of channel measurements depicting the amplitude and phase information for multiple orthogonal frequency division multiplexing (OFDM) subcarriers. For instance, IEEE 802.11 a/g/n WiFi framework is based on OFDM and the relatively wideband 20MHz channel is partitioned into 52 subcarriers. Generally, WiFi CSI information can be obtained from commercial WiFi devices, because it contains significant information about channel impulse response (CIR). Comparing with RSSI, which only provides a single measurement of the power over the whole channel bandwidth, WiFi CSI is a more informative and fine-grained characteristic. Up to now, WiFi CSI based human activity recognition approaches have been widely studied [16]–[21], such as E-eyes [16], CARM [17], WiGest [18], RT-Fall [19], ABLSTM [20], and others. In [16], E-eyes used the commercial WiFi devices to monitor human activity with one transmitter and multiple receivers, which could achieve in-place and walking activity recognition. In [17], CARM exploited a CSI-speed model and a CSI-activity model to recognize different human activities. In [18], WiGest introduced a WiFi signal propagation model in physical space under the existence of human activity, and then leveraged the effect of the in-air hand motion on the wireless signal strength received by the device from an access point to recognize the performed gesture. In [19], to promote performance, RT-Fall exploited both the amplitude and phase information of WiFi CSI data to detect human falls simultaneously. In [20], ABLSTM proposed a new deep learning based approach for passive human activity recognition by learning representative features in two directions from raw sequential CSI measurements. Although these human activity recognition approaches can ensure certain recognition accuracies, they suffer from some issues at the same time. Firstly, it was necessary to establish an efficient mechanism for conducting human activity detecting. Secondly, so far, it was very hard to establish a robust corresponding relationship between human activities and WiFi CSI.

Motivated by the observations above, in this paper, we propose a human activity recognition approach using deep recurrent neural network (HARNN). HARNN can recognize six basic human activities in daily life, including running, walking, standing, sitting, crouching, and lying. To realize HARNN approach, we mainly face four technical challenges. The first one is how to know that a person is performing an activity in a target area. It is well known that when humans perform any activity, their motion can induce fluctuation of WiFi signal in a unique manner. Therefore, we can extract useful variation statistics from raw WiFi CSI as indications of activities occurring. Specially, the variance and correlation coefficient have been widely used in many statistics applications, which can be found in [22] and [23]. Most of the existing WiFi CSI based approaches only leveraged the variance or the correlation coefficient directly to detect the human activities. However, their detection performances were influenced heavily by the random noise or channel variation, which may cause false recognition result. Now, a natural question to ask: which strategy is better for detecting human activities in a target area? The second challenge is to obtain effective WiFi CSI data. The reason is that the collected raw WiFi CSI data could be interfered by the random noise derived from indoor environments. For that, it is a challenging problem that how to preserve WiFi signal details while filtering out the noise in the WiFi signal. The third challenge is to obtain multiple representative statistical features. The existing works display that human activities can be depicted by a number of features in time domain, such as maximum value, minimum value, mean value, and so forth. It is conceivable that so many features would reduce the efficiency of activity recognition. As a result, it is necessary to seek for representative features that can characterize various human activities profoundly. The fourth challenge is how to leverage the extracted representative features for performing human activity recognition. Considering feature extraction and recognition part are not jointly optimized, it is not enough to rely on representative statistical features alone. Thus, we need to utilize a suitable method to recognize different human activities.

To overcome the first challenge, we design a novel two-level decision tree for using the variance and correlation coefficient of raw WiFi CSI data efficiently.

Meanwhile, a linear regression method is introduced, i.e. least median of squares (LMS) algorithm, to seek for the optimal parameter for the designed decision tree. Based on this observation, the decision tree is utilized to sense the indoor environment variation, and then detect whether there is any human activity occurring in a target area. To overcome the second challenge, the discrete wavelet transform (DWT) strategy (e.g., [24]) is employed to eliminate the influence of the random noise in indoor environments by signal decomposition. It is helpful to reduce the interference of both the complicated background environment and random noise effectively. To overcome the third challenge, we only extract two representative features from denoised WiFi CSI data in time and frequency domains, including channel power variation (CPV) and time-frequency analysis (TFA), which can characterize various human activities profoundly. To overcome the fourth challenge, a recurrent neural network (RNN) (e.g., [25]–[29]) model with long short-term memory (LSTM) block is used to recognize different human activities by leveraging the extracted representative features above. As the term recurrent implies, the RNN model takes not only the current input data but also several previous input data. In other words, it has a memory that can obtain the variation in input data. Thus, the selected method has the capacity for efficiently capturing the complicated non-linear relationship between input and output data through training process, then conducts human activity recognition. According to the above steps, the proposed HARNN can establish a robust relationship between human activities and WiFi CSI. Real experiments have been conducted to verify the high performance of the proposed HARNN for human activity recognition using WiFi CSI. Then, the results are compared with some existing works.

In summary, the main contributions of this work are as follows:

1) To detect human activities in a target area, we firstly design a novel two-level decision tree. Unlike the existing CSI based approaches, the designed decision tree not only uses the variance but also the correlation coefficient of raw WiFi CSI data. Meanwhile, a linear regression method is also introduced to seek for the optimal parameter for the designed decision tree. This helps avoiding false activity generated by other noise in indoor environments from nearby users.
2) In order to characterize various human activities, the DWT strategy is employed to eliminate the random noise of indoor environments presented in raw WiFi CSI data. After that, only two representative features are extracted from denoised WiFi CSI data, including channel power variation (CPV) in time domain and time-frequency analysis (TFA) in frequency domain. Therefore, this helps improving the efficiency of activity recognition.
3) To realize HARNN, the proposed RNN model is used to recognize different human activities. More importantly, it can perform activity recognition from the extracted representative features, instead of raw WiFi CSI data. This reduces the influence of the random noise derived from indoor environments significantly.

The remaining paper is organized as follows. The preliminaries of WiFi CSI are discussed in Section II, and Section III illustrates the architecture and design of the proposed HARNN approach. Section IV describes the data collection for experiments and presents the experimental setups. Next, the experimental results are presented and discussed in this section. Finally, our conclusion can be obtained in Section V.

II. PRELIMINARY

In this section, a short overview of WiFi CSI is presented. The most of modern off-the-shelf WiFi devices operate on both the 2.4GHz and 5GHz bands and employ OFDM at the physical (PHY) layer. The OFDM channel is a wide channel divided into subcarriers where each transmitted signal on every subcarrier has a different signal strength and phase. Unlike RSSI that represents the total received signal strength at the receiver, WiFi CSI contains more significant information of individual subcarrier between each pair of transmitter and receiver antenna. In view of this, compared with RSSI, the WiFi CSI is regarded as a more fine-grained and informative measurement for wireless channel, especially, it can distinguish the multipath components at the subcarrier level. Furthermore, as discussed above, WiFi CSI can also provide the channel response, including the magnitude and phase.

A group of subcarriers channel measurements can be revealed to upper layer users in the format of WiFi CSI in frequency domain, which is commonly used as follows:

$$\mathbf{H}_{sub} = \left[H_1 e^{j\angle H_1}, H_2 e^{j\angle H_2}, \ldots H_i e^{j\angle H_i}\right]^T \quad i \in [1, N_{sub}], \tag{1}$$

where $\mathbf{H}_{sub}$ has a dimension of $N_{sub} \times 1$, H_i and $\angle H_i$ are the amplitude and phase of the $i-th$ subcarrier, and N_{sub} is the total number of subcarriers.

Generally, the continuous raw WiFi CSI data of the $i-$ th subcarrier with a predefined sliding time window of length m is collected, which is able to be expressed by

$$\mathbf{H}_i = \left[H_{1,i} e^{j\angle H_{1,i}}, H_{2,i} e^{j\angle H_{2,i}}, \ldots H_{\kappa,i} e^{j\angle H_{\kappa,i}}, \ldots H_{m,i} e^{j\angle H_{m,i}}\right] \quad \kappa \in [1, m], \tag{2}$$

where $\mathbf{H}_i$ has a dimension of $1 \times m$, $H_{\kappa,i}$ and $\angle H_{\kappa,i}$ are the amplitude and phase of the $\kappa - th$ data point of the $i - th$ subcarrier.

III. SCHEME DESIGN

A. OVERVIEW

The basic idea of our proposed approach is to extract representative features for RNN training, and then perform activity recognition. Fig. 1 illustrates the overall architecture and block diagram of the proposed HARNN approach. One important question for the correct operation of HARNN is when to know that a person is generating an activity.

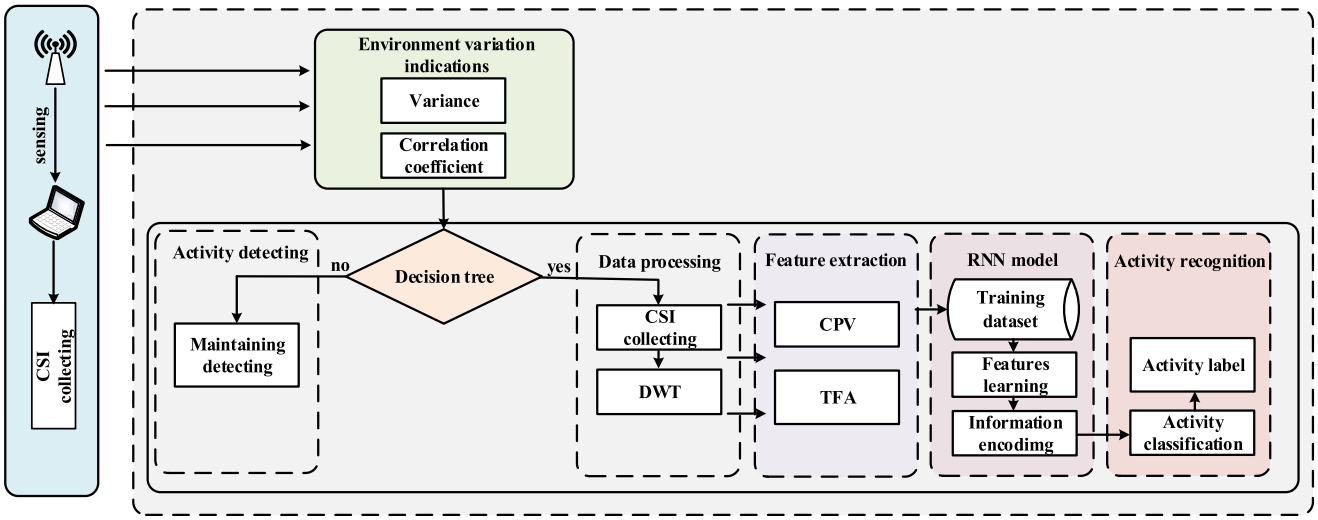


FIGURE 1. Overall architecture of the proposed HARNN approach.

This helps avoiding false activity generated by other noise in indoor environments from nearby users and also leads to energy-efficiency. In view of this, it is assumed that a person without wearing any specified sensor or device performs random activity in a target area, and the raw WiFi CSI data is continuously collected. At the same time, it is analyzed to sense whether there is environment variation. If sensed that indoor environments generate variation caused by human activities, the collected raw WiFi CSI data needs to be processed for human activity recognition. To solve the above technical issues, in our study, we can divide the proposed HARNN approach into four main modules: human activity detecting, data processing, feature extraction, and human activity recognition.

Human Activity Detecting: This module is responsible for detecting human activities. We can leverage two sensitive statistics, including the variance and correlation coefficient of raw WiFi CSI data, as environment variation indications. Then, the corresponding two-level decision tree can be constructed. Based on this, if there is a response 'yes', the proposed HARNN approach would start the data processing module, otherwise continue maintaining detecting.

Data Processing: This module is responsible for implementing denoising. Since the collected raw WiFi CSI data is obtained from commercial WiFi Network Interface Cards (NICs), it contains the random noise from various sources, such as nearby electronic devices. This noise can add false edges and affect HARNN accuracy and robustness. Therefore, in order to increase HARNN quality, we can leverage a wavelet based denoising method to eliminate the influence derived from indoor environments.

Feature Extraction: This module is responsible for extracting representative features from different statistical profiles. It is well known that when humans perform random activity in a target area, their motion could affect indoor environments in a unique manner, which can be changed into the impact on WiFi signal. After that, the significant impact is in turn manifested as distinguishing perturbation in WiFi CSI data. Considering that each human activity in daily life is highly different, it is very possible to recognize different activities by closely examining representative features exhibited in WiFi CSI data. In this study, we exploit two representative features with regard to the same human activity, including the CPV in time domain and the TFA in frequency domain.

Human Activity Recognition: This module is responsible for distinguishing different human activities. The RNN model with LSTM block is introduced to recognize different activities by automatically learning representative features and encoding the temporal information during feature learning. For that, a large amount of raw WiFi CSI data needs to be collected for extracting representative features of different human activities. Meanwhile, the extracted effective features and the corresponding activities labels form the dataset, which is leveraged for RNN training. When every parameter of the RNN model is adjusted appropriately, it can perform activity recognition. However, if the experimental setting changes, the RNN model needs to be trained again.

B. ENVIRONMENT VARIATION INDICATIONS

Fig. 2(a) and (d) illustrate the indoor environment variation when a person is running between a pair transmitter and receiver. We can observe that although the amplitude value of raw WiFi CSI performs variation, the fluctuation is very small. This reason is that the amplitude of raw WiFi CSI is insensitive to environment variation. In view of this, we need to utilize some typical environment variation statistics for human activity detecting in our study. Based on the previous research [22], [23], when humans perform random activity in a target area, both the variance and correlation coefficient of raw WiFi CSI data would have corresponding variation. Thus, we can present the variance and correlation

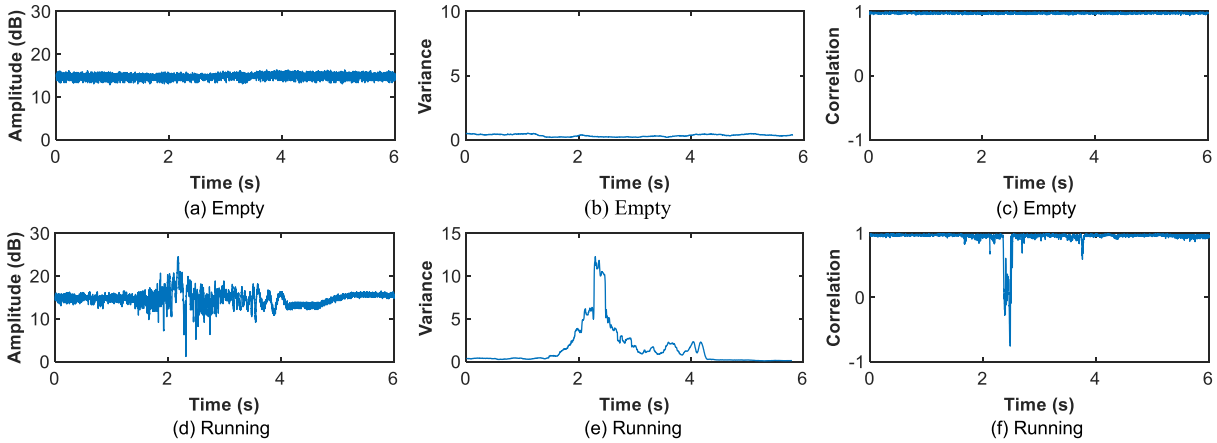


FIGURE 2. The variation of the amplitude, variances, and correlation coefficients of raw WiFi CSI data when a person is running in the same target area.

coefficient as environment variation indications for human activity detecting.

1) VARIANCE

When humans perform any activity in a target area, the variance of each subcarrier with a predefined sliding time window would generate variation. Then, the variances of all subcarriers with a predefined sliding time window can constitute a $N_{sub} \times 1$ vector, which can be calculated with following equation:

$$\mathbf{V} = [v_1, v_2, \ldots v_i]^T \; i \in [1, N_{sub}], \tag{3}$$

and the variance of each subcarrier can be given by

$$v_i = \mathrm{var}\left(\|\mathbf{H}_i\|\right) i \in [1, N_{sub}], \tag{4}$$

where v_i is the variance of the $i - th$ subcarrier, $var(\cdot)$ is the variance operation, and $\|\cdot\|$ is the magnitude operation, respectively.

In ideal condition, human activities should have the same influence on each subcarrier, that is to say, the variances of all subcarriers should be the same. However, due to the influences of the severe multipath and random noise in indoor environments, the variances of different subcarriers perform differently. In order to take full advantage of each subcarrier, we can leverage the variance of the vector $\mathbf{V}$, which is given as follows:

$$V_m = \mathrm{var}\left(\mathbf{V}\right), \tag{5}$$

where V_m represents the fluctuation degree of WiFi signal for environment variation. It is well known that in the case when no human activity occurs in a target area, we can get $V_m = 0$. In contrast, we can have $V_m \succ 0$.

Fig. 2(b) and (e) illustrate the variances variation when a person is running in a target area. As we can see, in the case when there is no human activity, as analyzed above, the variances are close to 0. In contrast, the values of the variances become large drastically.

2) CORRELATION COEFFICIENT

Considering that human activities can induce fluctuation of WiFi signal, the authors in [30] and [31] have studied that the correlation coefficient between successive raw WiFi CSI data is sensitive to the indoor environment variation. As a result, the correlation coefficient can be leveraged for detecting human activities. In addition, we calculate all correlation coefficients between subcarriers. After that, we can obtain a $N_{sub} \times N_{sub}$ correlation coefficients matrix, which can be expressed mathematically by

$$\mathbf{C} = \begin{bmatrix} \rho_{(1,1)} & \cdots & \rho_{(1,N_{sub})} \\ \vdots & \ddots & \vdots \\ \rho_{(N_{sub},1)} & \cdots & \rho_{(N_{sub},N_{sub})} \end{bmatrix}, \tag{6}$$

and we can have

$$\rho_{(i,j)} = \mathrm{corr}\left(\|\mathbf{H}_i\|, \left\|\mathbf{H}_j\right\|\right) \; i, j \in [1, N_{sub}], \tag{7}$$

where $\rho_{(i,j)}$ denotes the correlation coefficient between the $i - th$ and $j - th$ subcarriers, and corr$(\cdot)$ represents the correlation operation, respectively.

Ideally, it is noted that human activity can attach an identical influence on each subcarrier principally, thus all correlation coefficients should be the same. However, the severe multipath and random noise on each subcarrier differ from each other, which can lead to inconsistent parameter. In view of this, the challenge we face now is how to obtain a reliable and effective parameter from these primary correlation coefficients.

Motivated by this observation, the incorrect correlation coefficients can be sifted out by conducting a least median of squares (LMS) [30] outlier detection. First of all, the objective function minimizing the median of the squared residuals can be defined as follows:

$$E = \min\left(\mathrm{med}\left(r^2_{(i,j)}\right)\right), \tag{8}$$

and we can have

$$\begin{aligned} r_{(i,j)} &= \left\{\rho_{(i,j)} - \tilde{\rho}\right\}, \\ \rho_{(i,j)} &= \left\{\rho_{(i,j)}, i \prec j \in [1, N_{sub}]\right\}, \\ \Gamma &= N_{sub}\left(N_{sub} - 1\right)/2, \end{aligned} \tag{9}$$

where min(·) and med(·) are the minimum and median operation, $r_{(i,j)}$ is residuals, $\tilde{\rho}$ is the LMS estimation, and Γ is the number of correlation coefficients between different subcarriers.

Mathematically, we determine whether $\rho_{(i,j)}$ is an outlier or not following a typical LMS regression, which can be given as follows:

$$I\left(\rho_{(i,j)}\right) = \begin{cases} 1 & if \left|r_{(i,j)}/\sigma^*\right| \leq 2.5 \\ 0 & otherwise \end{cases}, \tag{10}$$

where we can have

$$\sigma^* = \sqrt{\frac{\sum_{j=1}^{N_{sub}} \sum_{i=1}^{N_{sub}} q_{(i,j)} r_{(i,j)}^2}{\sum_{j=1}^{N_{sub}} \sum_{i=1}^{N_{sub}} q_{(i,j)}}}, \tag{11}$$

$$q_(i,j = \begin{cases} 1 & if \left|r_{(i,j)}/s^0\right| \leq 2.5 \\ 0 & otherwise \end{cases},$$

$$s^0 = 1.4826 * \left(1 + \frac{5}{\Gamma^2}\right) \sqrt{\underset{(i,j)}{\text{med}}\left(r_{(i,j)}^2\right)}. \tag{12}$$

After diagnosing the biased and incorrect correlation coefficients, they can be sifted out simply, and then average the remained ones as the ultimate estimation for a group of correlation coefficients, it is denoted as

$$\rho = \frac{1}{\sum_{j=1}^{N_{sub}} \sum_{i=1}^{N_{sub}} I\left(\rho_{(i,j)}\right)} \sum_{j=1}^{N_{sub}} \sum_{i=1}^{N_{sub}} \rho_{(i,j)} I\left(\rho_{(i,j)}\right). \tag{13}$$

According to the above analysis, the ultimate correlation coefficient ρ can be obtained. Generally, in the case when there is no human activity in a target area, the correlation coefficient of successive raw WiFi CSI data would tend to be larger. In contrast, the correlation coefficient would tend to be smaller if there is any human activity occurring.

Fig. 2(c) and (f) illustrate the correlation coefficients variation when there is a person running. As we can see, in the case when there is no human activity, the correlation coefficients are close to 1. In contrast, they would decrease dramatically.

C. HUMAN ACTIVITY DETECTING

The first step towards recognizing different human activities using raw WiFi CSI data is to detect whether a person performs any activity or not in a target area. In this section, we discuss how we can detect human activities by using the variance and correlation coefficient effectively.

In our study, in order to make full use of the above two environment variation indications, we design a two-level decision tree to conduct the activity detecting. Besides, the detailed detecting process is shown in Fig. 3. There are four key parameters of the two-level decision tree, the variance V_m and the correlation coefficient ρ, and the corresponding thresholds ν and α. In the first level, if the variance V_m is over ν, it would move to the second level, otherwise the result is activity absence. In the second level, if the correlation coefficient ρ is over α, the result is absence, otherwise it

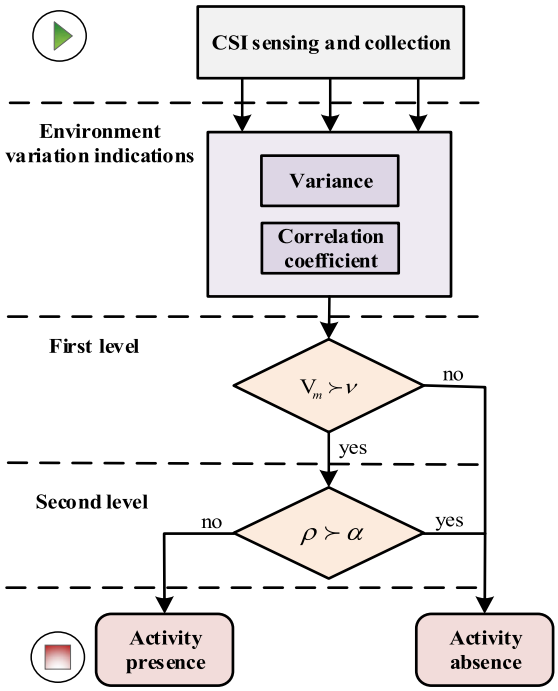


FIGURE 3. The architecture of human activity detecting.

is presence. The activities detecting results can be described with following equation:

$$Y = \begin{cases} 1 & V_m \succ \nu and \rho \prec \alpha \\ 0 & otherwise \end{cases}, \tag{14}$$

where Y is the label: positive label 'Y=1' for human activity presence, and negative label 'Y=0' for human activity absence. The corresponding thresholds can be determined by some preliminary measurements. In addition, we do not need to calibrate for each different case because these thresholds can gracefully apply to various indoor environments. Besides, the design of human activity detecting is lightweight because the calculation of the variance and correlation coefficient of raw WiFi CSI data is fast, convenient, and small complexity. Furthermore, the usage of variance and correlation coefficient is sufficient and efficient for recognition since both of them are sensitive to the indoor environment variation.

D. DATA PROCESSING

After data processing module is triggered, raw WiFi CSI data needs to be processed to eliminate the random noise. The first row in Fig. 4 shows raw WiFi CSI data of different human activities collected in the same experimental environment. We can observe that the amplitude has been interfered over time, which may result in inaccurate WiFi CSI profiles construction for activity recognition. In view of this, our strategy to deal with the issue is to utilize wavelet based denoising method [24]. The selected strategy has the advantage of being computationally efficient. In addition, it does not make any particular consumption with regard to the nature of the WiFi signal, and permits discontinuities in the WiFi signal. Here DWT can decompose raw WiFi CSI data into two terms:

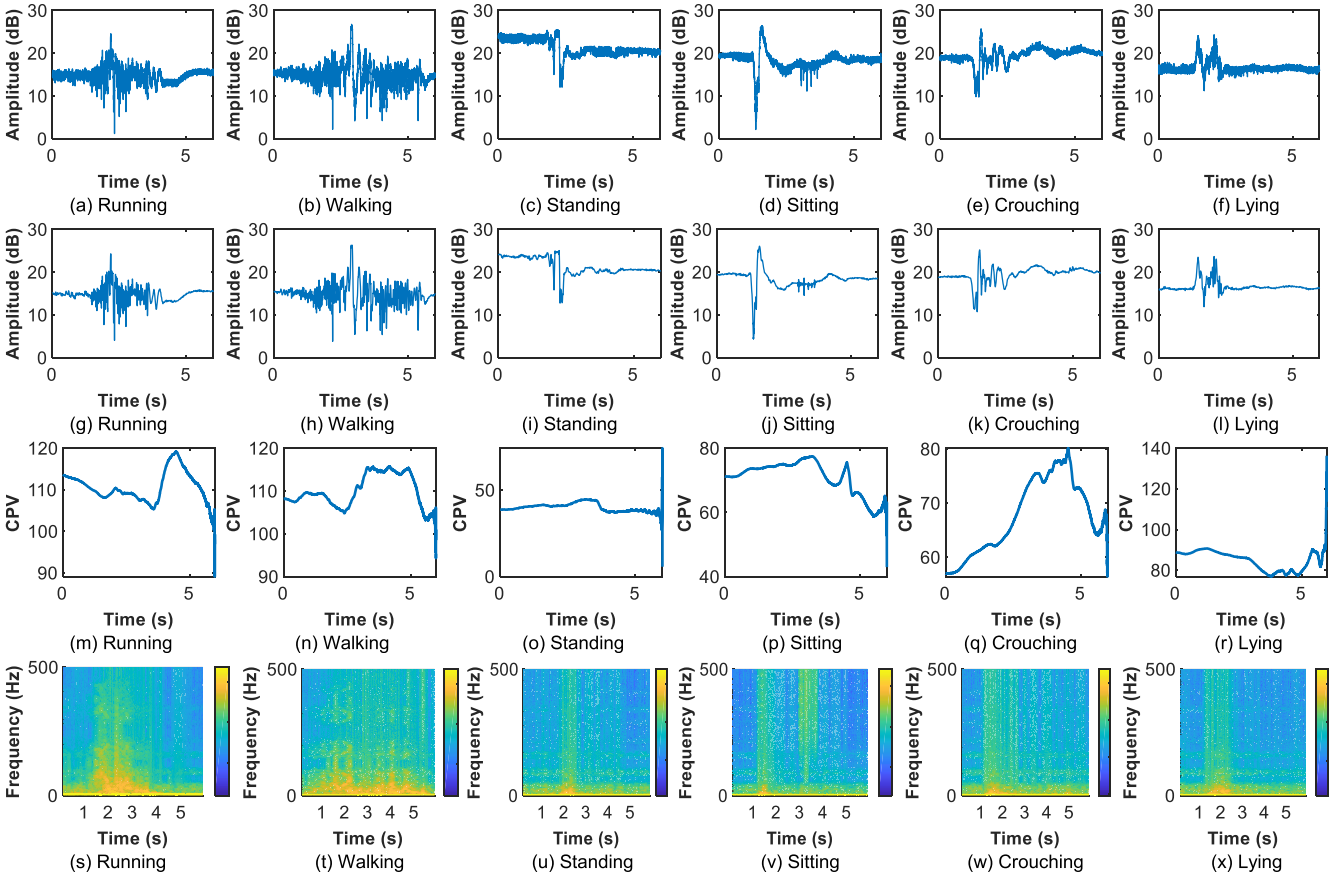


FIGURE 4. The raw WiFi CSI, denoised WiFi CSI, CPV, and TFA results when human performs different activities in the same target area.

the approximation coefficients and detail coefficients. Among them, the approximation coefficients can describe the raw WiFi CSI data shape, and the detail coefficients can capture both the random noise and fine details of raw WiFi CSI data. Considering the random noise is mainly located in the detail coefficients parts after the decomposition of each layer, the denoising can be achieved by handling these detail coefficients of each layer (i.e. j =1,2,...J). Finally, we reconstruct the WiFi CSI data by inversing transformation of the processed wavelet coefficients.

If the raw WiFi CSI data with a predefined sliding time window is interfered by the random noise, the observed value can be expressed as follows:

$$\mathbf{H}_i=\left[H_{1,i}e^{j\angle H_{1,i}}, H_{2,i}e^{j\angle H_{2,i}}, \ldots H_{\kappa,i}e^{j\angle H_{\kappa,i}}, \ldots H_{m,i}e^{j\angle H_{m,i}}\right]. \tag{15}$$

The expansion of the raw WiFi CSI data in wavelet space can be denoted as follows:

$$H_{\kappa,i} = \sum_{j\in Z}\sum_{k\in Z} <H_{\kappa,i}, \psi_{j,k}(\kappa)> \psi_{j,k}(\kappa)\, \kappa \in [1, m], \tag{16}$$

where $\langle\cdot\rangle$ is the dot product operation, and $\psi_{j,k}(\kappa)$ is the wavelet function, which is expressed by

$$\psi_{j,k}(\kappa) = \left\{2^{j/2}\psi\left(2^j\kappa - k\right)\right\} j, k \in Z, \tag{17}$$

Next, the pyramid algorithm of multiresolution analysis can be used for wavelet decomposition in this study. It is able to be expressed mathematically by

$$H_{\kappa,i} = \sum_{k\in Z} < H_{\kappa,i}, \varphi_{j,k}(\kappa) > \varphi_{j,k}(\kappa) + \sum_{j\le j'\le J}\sum_{k\in Z} < H_{\kappa,i}, \psi_{j',k}(\kappa) > \psi_{j',k}(\kappa), \tag{18}$$

where $\varphi_{j,k}(\kappa)$ is the scale function, we can have the value

$$\varphi_{j,k}(\kappa) = \left\{2^{j/2}\varphi\left(2^j\kappa - k\right)\right\} j, k \in Z. \tag{19}$$

The approximation coefficients and detail coefficients can be given by

$$\begin{aligned} c_{j,k} &= < H_{\kappa,i}, \varphi_{j,k}(\kappa) >, \\ d_{j,k} &= < H_{\kappa,i}, \psi_{j,k}(\kappa) >, \end{aligned} \tag{20}$$

where the value of $\varphi_{j,k}(\kappa)$ is able to be given mathematically with following equations:

$$\begin{aligned} \varphi_{j,k}(\kappa) = &\sum_{n\in Z} < \varphi_{j,k}(\kappa), \varphi_{j+1,n}(\kappa) > \varphi_{j+1,n}(\kappa), \\ &< \varphi_{j,k}(\kappa), \varphi_{j+1,n}(\kappa) \\ &> = \frac{2^j}{\sqrt{2}}\sum_{\kappa=1}^{m}\varphi\left(2^j\kappa - k\right)\overline{\varphi\left(2^{j+1}\kappa - n\right)}, \end{aligned} \tag{21}$$

next, the h_k can also be obtained by

$$\begin{aligned} h_k &= <\varphi(\kappa), \sqrt{2}\varphi(2\kappa - k) \\ &> = \sqrt{2}\sum_{\kappa=1}^{m}\varphi(\kappa)\overline{\varphi(2\kappa - k)}, \\ h_{n-2k} &= <\varphi_{j,k}(\kappa), \varphi_{j+1,n}(\kappa)>, \end{aligned} \tag{22}$$

finally, we can have the value of $\varphi_{j,k}(\kappa)$

$$\varphi_{j,k}(\kappa) = \sum_{n\in Z} h_{n-2k}\varphi_{j+1,n}(\kappa) = \sum_{n'\in Z} h_{n'}\varphi_{j+1,n'+2k}(\kappa). \tag{23}$$

The expected value $c_{j,k}$ can be given through the above calculating. We let

$$c_{j,k} =< H_{\kappa,i}, \varphi_{j,k}(\kappa) >= \sum_{n\in Z} h_n c_{j+1,n+2k}. \tag{24}$$

Once the approximation coefficients are obtained, and then the corresponding detail coefficients can be given by

$$d_{j,k} =< H_{\kappa,i}, \psi_{j,k}(\kappa) >= \sum_{n\in Z} g_n c_{j+1,n+2k}. \tag{25}$$

To remove the random noise component while retaining sufficient details for activity recognition, a soft threshold method should be applied to the detail coefficients. The soft threshold function can be given as follows:

$$S = \begin{cases} d_{j,k} - T, & d_{j,k} > T \\ d_{j,k} + T, & d_{j,k} < -T \\ 0, & |d_{j,k}| \le T \end{cases}, \tag{26}$$

where T is the dynamic threshold.

For the obtained detail coefficients, we can set the detail coefficients with a small absolute value to zero, and decrease the coefficients with a large absolute value. According to (26), the estimated detail coefficients are used to reconstruct the WiFi CSI data directly, which can be denoted as

$$\hat{H}_{\kappa,i} = \sum_{k\in Z} c_{J,k}\varphi_{J,k} + \sum_{j=-\infty}^{J}\sum_{k\in Z}\hat{d}_{j,k}\psi_{j,k}. \tag{27}$$

After that, we can obtain the value of denoised WiFi CSI data, which can be given by

$$\hat{\mathbf{H}}_i = \left[\hat{H}_{1,i}e^{j\angle H_{1,i}}, \hat{H}_{2,i}e^{j\angle H_{2,i}}, \ldots \hat{H}_{\kappa,i}e^{j\angle H_{\kappa,i}}, \ldots \hat{H}_{m,i}e^{j\angle H_{m,i}}\right], \tag{28}$$

where $\hat{\mathbf{H}}_i$ is the WiFi CSI data output after wavelet based denoising. The second row in Fig. 4 shows the transformation results of noisy subcarrier amplitude of different human activities to a noiseless signal through the selected strategy. We can observe that the denoised WiFi CSI data is smoother, and it is obvious that the wavelet based denoisng can eliminate the random noise presented in raw WiFi CSI data effectively.

E. FEATURES EXTRACTION

After raw WiFi CSI data collection and processing, we concern about how to extract representative features for charactering different human activities. It is well known that human activities could influence indoor environment propagation in a unique manner, which can be changed into impact on WiFi signal, and then WiFi CSI can capture the impact. That is to say, for different activities, the ways of expression are different, and the influence on the indoor environment differs from each other. In order to describe the above influence, we can apply fast Fourier transform (FFT) to the denoised WiFi CSI data. After transforming into time domain, we can extract CPV as a representative feature in time domain.

The third row in Fig. 4 shows the CPVs of different human activities. As we can see, the standing activity has less power gain compared with other activities. In contrast, the lying activity has more power gain. It is possible that the lying activity has larger influence on the WiFi signal reflected from human body.

Additionally, Y. Kim *et al*. [32] have studied that there is a function relationship between the velocities of movement of human activities and the frequency components of the WiFi CSI data. Besides, the authors in [33] have studied that the velocities of different human activities are highly distinguishing. This suggests that the influence of different human activities is likely to be most pronounced in a specific frequency range. In order to obtain the TFA, a transformation to the time-frequency domain is necessary to conduct features extraction. Specially, among various time-frequency feature extraction methods, short-time Fourier transform (STFT) has better spectrogram characteristics compared with the others. Therefore, we can select STFT for TFA. After that, we can perform TFA of denoised WiFi CSI data to separate the desired components in frequency domain. Based on this observation, human activities can be modeled by profiling the each frequency component by using TFA tool.

The fourth row in Fig. 4 shows the TFAs of different human activities. It is evident that the running activity performs higher frequency components compared with the other human activities, followed by the walking activity. In addition, the sitting down activity has lower frequency components. The possible reason is that the sitting activity has a very low movement velocity, which could generate minute influence on WiFi signal.

F. RNN MODEL

In this section, the RNN model is utilized to recognize different human activities by leveraging the extracted representative features. For activity recognition, the first step needs to create an appropriate input vector from the extracted representative features. Furthermore, in order not to lose important information contained in the WiFi CSI data and improve the accuracy of activity recognition, we include denoised WiFi CSI data in the input vector because it also provides very important information. Then, the input vector can be

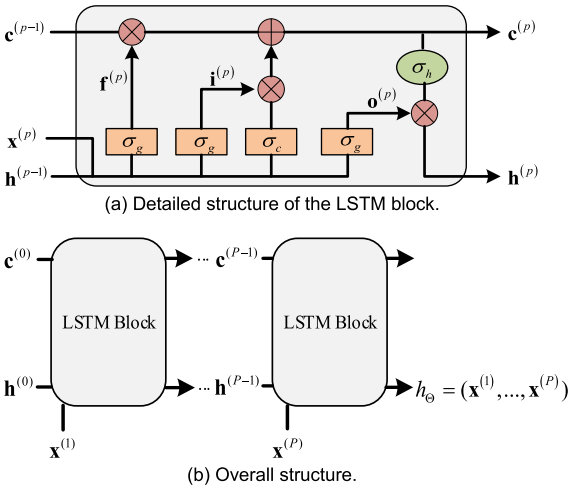


FIGURE 5. RNN structure with the LSTM block.

expressed by

$$\mathbf{x} = [\text{CPV, TFA, denoised WiFi CSI}]^T, \tag{29}$$

where the dimension of the input vector is D_x.

In our study, in order to obtain better performance for activity recognition, we consider an RNN model taking P sets of representative features by extending CPV, TFA, and denoised WiFi CSI to make a decision. Fig. 5 illustrates the structure of RNN model with the LSTM block, where $\mathbf{x}^{(p)}$ and $\mathbf{h}^{(p)}$ represent input vector and hidden layer at the $p-th$ time step, respectively. It is assumed that the dimension of the hidden layer is D_h. Except for the hidden layer, the LSTM block in RNN model has one more unit, called cell state $\mathbf{c}^{(p)}$, which can control the flow of information by using three important gates, including forget, input, and output gates, respectively.

For the input vector, each gate vector needs to be calculated. They are can be calculated mathematically by

$$\begin{aligned}
\mathbf{f}^{(p)} &= \sigma_g\left(\mathbf{W}_f\mathbf{x}^{(p)} + \mathbf{U}_f\mathbf{h}^{(p-1)} + \mathbf{b}_f\right),\\
\mathbf{i}^{(p)} &= \sigma_g\left(\mathbf{W}_i\mathbf{x}^{(p)} + \mathbf{U}_i\mathbf{h}^{(p-1)} + \mathbf{b}_i\right),\\
\mathbf{o}^{(p)} &= \sigma_g\left(\mathbf{W}_o\mathbf{x}^{(p)} + \mathbf{U}_o\mathbf{h}^{(p-1)} + \mathbf{b}_o\right),
\end{aligned} \tag{30}$$

where every matrix represented by $\mathbf{W}$ has a dimension of $D_h \times D_x$, those represented by $\mathbf{U}$ have a dimension of $D_h \times D_h$, and the vectors denoted by $\mathbf{b}$ have a dimension of $D_h \times 1$. In addition, $\sigma_g(\cdot)$ denotes an element-wise activation function for all gates, which is able to be given by a sigmoid function in our study, i.e. $\sigma_g(z) = 1/(1+e^{-z})$. In addition, both previous cell state and hidden layer of the LSTM block at the first time step need to be initialized as zero vectors. Because of these gate vectors, the cell state and hidden layer in the LSTM block can be updated in real time, which can be expressed as follows:

$$\mathbf{c}^{(p)} = \mathbf{f}^{(p)} \circ \mathbf{c}^{(p-1)} + \mathbf{i}^{(p)} \circ \sigma_c\left(\mathbf{W}_c\mathbf{x}^{(p)} + \mathbf{U}_c\mathbf{h}^{(p-1)} + \mathbf{b}_c\right), \tag{31}$$

and we can have

$$\mathbf{h}^{(p)} = \mathbf{o}^{(p)} \circ \sigma_h\left(\mathbf{c}^{(p)}\right), \tag{32}$$

where the $\circ$ represents the Hadamard product operation, $\sigma_c(\cdot)$ and $\sigma_h(\cdot)$ are element-wise activation functions for the cell state and hidden layer, respectively. In our study, the hyperbolic tangent is leveraged for these activation functions. Finally, the output of the proposed RNN model can be obtained directly from the hidden layer at the last time step, which can be given by

$$h_\Theta\left(\mathbf{x}^{(1)}, \ldots \mathbf{x}^{(P)}\right) = \sigma\left(\mathbf{V}\mathbf{h}^{(P)} + b\right), \tag{33}$$

where $\mathbf{V}$ is the row vector of the D_h dimension, and b is a bias constant. The set Θ contains every parameter in the RNN model, for example, elements in the matrices $\mathbf{U}$ and $\mathbf{W}$, vectors $\mathbf{b}$ and $\mathbf{V}$, and the constant b.

Every parameter in the model can be adjusted by training representative features dataset. The sequence of P input data vectors can be denoted as $\mathbf{X} = (\mathbf{x}^{(1)}, \ldots \mathbf{x}^{(P)})$ and the corresponding label as y, where y can be represented as activity A, activity B, etc. From the representative features and corresponding activities labels, a set of input and output pairs $(\mathbf{X}, y)$ is generated to train and verify the used RNN model. For that, ς is introduced as such a activities dataset, and every parameter in the model is adjusted in the direction of minimizing the cost function, which can be expressed mathematically by

$$J(\Theta) = -\frac{1}{|\varsigma|}\sum_{g\in\varsigma} C(g), \tag{34}$$

where $C(g)$ is the cost of the $g-th$ input and output pair and it can measure how accuracy the RNN model generates the output compared with the ground truth data, and $|\cdot|$ is the number of elements in a set, respectively.

Among many choices of evaluating the cost performance, cross-entropy is used in this study, we have

$$C(g) = y^{(g)} \log h_\Theta\left(\mathbf{X}^{(g)}\right) + \left(1 - y^{(g)}\right)\log\left(1 - h_\Theta\left(\mathbf{X}^{(g)}\right)\right), \tag{35}$$

where the superscript indicates the index of the input and output pairs. Then, every parameter in the RNN model can be updated in an iterative manner by taking use of the gradient descent method. It can update each parameter at each iteration in the direction of the steepest descent at the same time, which is able to given by

$$\Theta_{n+1} = \Theta_n - \eta\nabla_\Theta J(\Theta), \tag{36}$$

where ∇_Θ is the gradient operation with regard to Θ, and η is the learning rate, which can determine the step size. Many variations of the gradient descent methods have been studied widely, such as Adam optimizer. It is well known that these optimizers can change adaptively the learning rate to achieve the minimum cost efficiently and precisely. Regardless of which optimizer is used in the RNN model, we need to

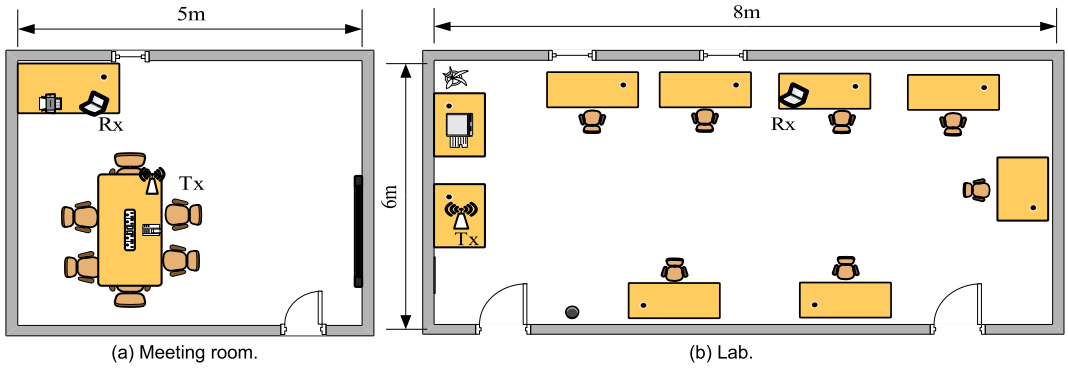


FIGURE 6. The experimental environments for the proposed HARNN approach.

take the gradient of the cost function at each iteration, and the back propagation algorithm can do that by using matrix multiplications.

Due to containing many significant parameters, the recurrent neural network can efficiently capture non-linearity relationship between the input and output layers, but it has a high risk of over-fitting. In view of this, an early stopping is very necessary to avoid that. In addition, the activities dataset of input and output pairs, ς, are separated into three non-overlapping datasets: training dataset ς_{tr}, validation dataset ς_v, and testing dataset ς_{te}. Then, we train the proposed RNN model to determine all parameters of the RNN model based on the training dataset with true labels. In our study, we set the number of hidden layers as $D_h = 200$. We leverage the Adam optimizer that can solve effectively the adaptive learning rates for each parameter during optimization. η is the learning rate which can be set to be 1×10^{-4}. Additionally, here we can adopt the Adam optimizer to reduce the risk of over-fitting. Besides, at each iteration, these values of the cost function (34) with respect to the validation dataset are evaluated and tracked, and we select the optimal parameters when the cost function is minimized. Finally, the performance of the proposed RNN model is verified with the testing dataset.

IV. PERFORMANCE AND EVALUATION

In this section, we describe the raw WiFi CSI data collection for experiments first. After that, the experimental results are shown and discussed. Then, we evaluate the accuracies of human activity recognition with different approaches in the lab scenario. Furthermore, we also evaluate performance of the proposed HARNN approach with different realistic settings in typical indoor environments. According to the experimental results, we can understand intuitively the theory of this paper.

A. IMPLEMENTATION

In our study, we employ a commercial WiFi device as a transmitter and a Lenovo laptop with Inter 5300 NIC as a receiver with a sampling rate of 1000Hz. In addition, a sliding time window with a window size of 6 seconds is used to segment raw WiFi CSI data collected. All the experiments are performed at 5GHz band with the channel bandwidth of 20 MHz. The proposed HARNN can acquire WiFi CSI data by using WiFi CSI tool [34] and then, perform human activity detecting, WiFi CSI data processing, feature extraction, and activity recognition using MATLAB.

B. DATA COLLECTION

In the study, we collected activities dataset in two different indoor environments, i.e. a meeting room and a lab. The layouts of the two indoor environments are shown in Fig. 6(a) and Fig. 6(b). Among them, the meeting room has a size of $5m \times 6m$, which contains various furniture and electronic devices. In this scenario, there is a pair trans-receiver device that is placed on the top of the table. In addition, the lab has a size of $8m \times 6m$, which contains nine tables and the trans-receiver device is placed on the top of the table as well. During raw WiFi CSI data collection, each person performs each activity in the same target area for a period of 15 seconds. At the beginning of each activity, the person needs to remain stationary, and then performs the corresponding activity as usual until the time is over. Totally, ten persons are involved for raw WiFi CSI data collection with six common daily activities of ''running, ''walking'', ''standing'', ''sitting'', ''crouching'', lying''. Each person needs to perform each activity 30 times. To make the activities dataset more diverse, we consider multiple different experimental settings in this study. Therefore, we need to repeat the above procedure for the collection of raw WiFi CSI data when experimental settings change.

C. RESULTS WITH DIFFERENT APPROACHES

There are WiFi CSI based device-free human activity recognition approaches, among them, E-eyes and CARM are device-free passive approaches to recognize different human activities in typical indoor environments. To verify the effectiveness of the proposed HARNN approach, we conduct a comparison with the existing approaches. Fig. 7 shows the result in the lab scenario. Three recognition approaches can achieve average accuracies of 95%, 96%, and 98% for all the human activities, respectively. It can be found that E-eyes approach performs the worst. Compared with E-eyes,

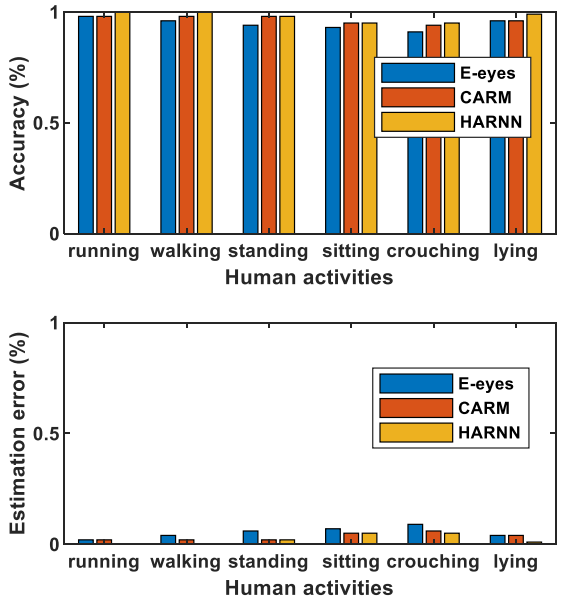


FIGURE 7. The recognition accuracies for different human activities of the three approaches in the lab scenario.

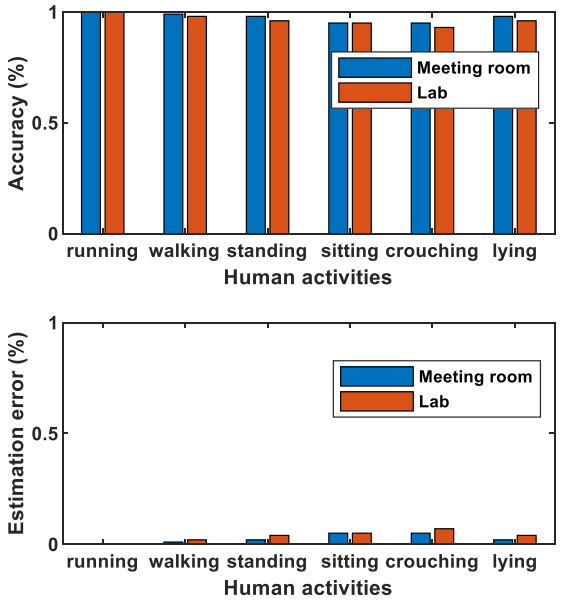


FIGURE 8. The recognition accuracies for different human activities in two typical indoor environments.

CARM has better recognition performance. Due to establishing a robust corresponding relationship between various human activities and WiFi CSI, the proposed HARNN approach is able to achieve the best performance for recognition of all the six activities, and the recognition accuracy of each activity can be more than 95%.

D. IMPACT OF DIFFERENT ENVIRONMENTS

The testing environment for experiments is an important factor for WiFi CSI data. In the additional experiment, two typical indoor environments are considered, i.e. a meeting room and a lab. The recognition accuracies in the meeting room and lab scenarios are shown in Fig. 8. As it is indicated, the proposed HARNN approach can achieve average recognition accuracies of 96% and 95% for all the activities, respectively. Besides, the lab area is much larger than that of the meeting room, which means that the lab has a more complicated multipath environment. Therefore, the recognition accuracies have better performance in the meeting room scenario compared with that in the lab scenario.

E. IMPACT OF THE NUMBER OF HIDDEN NODES

The number of hidden nodes is a significant parameter for RNN training for activity recognition. In view of this, we need to conduct an additional experiment to investigate the influence of this parameter on the performance of activity recognition in the lab scenario. The experimental result is shown in Fig. 9. As we can see, in the case when the number of hidden nodes is only 50, the recognition accuracies for all the activities are very low. With the increasing of the number of hidden nodes from 50 to 200, the recognition accuracies increase with hidden nodes. Then, when further increasing the number of hidden nodes from 200 to 300, the recognition accuracies remain relatively stable. This reason is that when

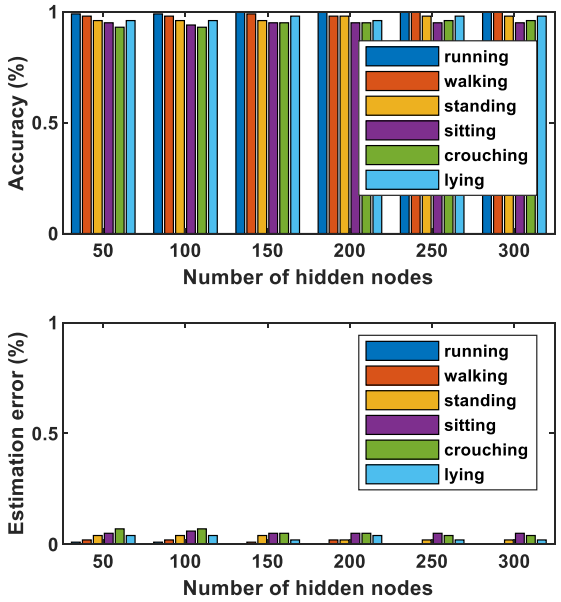


FIGURE 9. The recognition accuracies for different human activities with different number of hidden nodes in the lab scenario.

the number of hidden nodes is few, every parameter of the proposed RNN model is not adjusted appropriately. It should be noted that more hidden nodes would lead to longer training time. Therefore, 200 hidden nodes for the proposed RNN model are chosen.

F. IMPACT OF MULTIPLE RECEIVERS

The receiver's densities are a key parameter for evaluating system performance, which has been studied in [16] and [35]. In view of this, an additional experiment needs to be conducted to investigate the influence of this parameter on the performance of human activity recognition. We can analyze it through changing the number of receivers from one to five.

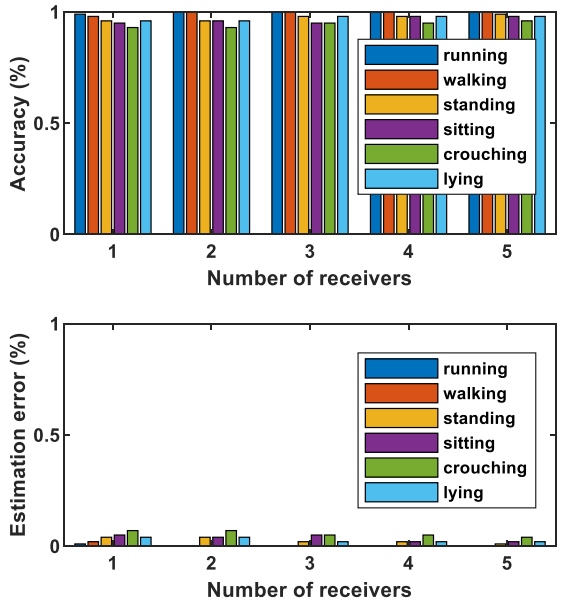


FIGURE 10. The recognition accuracies for different human activities with different number of receivers in the lab scenario.

The experiment is performed in the lab scenario where the receivers are deployed at the corner and the transmitter is placed at the center of the scenario. The experimental result is shown in Fig. 10. With the increasing of the number of receivers from one to five, the recognition accuracies increase with that. Concretely, the recognition accuracy for all the activities increases to 96% on average when there are two receivers, and reaches 98% on average when deploying five receivers in the lab scenario. This reason is that there are more direct paths as the number of receivers increases. As a result, with a higher density deployment, we believe our proposed HARNN can achieve a robust performance.

V. CONCLUSION

In this paper, we propose a WiFi CSI based passive human activity recognition approach using deep recurrent neural network. The proposed HARNN approach can establish a robust relationship between human activities and WiFi CSI. Then, we conduct some real experiments to verify the high recognition performance of the proposed HARNN approach and compare it with the existing approaches. The experimental result shows HARNN can achieve the best recognition performance compared with the other approaches in the same indoor environment. Besides, we evaluate the recognition performance of HARNN approach in two typical indoor environments. As it is indicated, the recognition accuracies have better performance in the meeting room scenario compared with that in the lab scenario. Moreover, we also verify its high recognition performance with different settings in the lab scenario. The experimental results show that the proposed HARNN approach can achieve a satisfactory performance.

REFERENCES

[1] Y. Wang, K. Wu, and L. M. Ni, "WiFall: Device-free fall detection by wireless networks," *IEEE Trans. Mobile Comput.*, vol. 16, no. 2, pp. 581–594, Feb. 2017.

[2] C. Tran and M. M. Trivedi, "3-D posture and gesture recognition for interactivity in smart spaces," *IEEE Trans. Ind. Informat.*, vol. 8, no. 1, pp. 178–187, Feb. 2012.

[3] X. Li, D. Zhang, Q. Lv, J. Xiong, S. Li, Y. Zhang, and H. Mei, "IndoTrack: Device-free indoor human tracking with commodity Wi-Fi," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 1, no. 3, p. 72, Sep. 2017.

[4] X. Yang and Y. Tian, "Super normal vector for human activity recognition with depth cameras," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 5, pp. 1028–1039, May 2017.

[5] O. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 1192–1209, Jul. 2013.

[6] Q. Wu, Y. D. Zhang, W. Tao, and M. G. Amin, "Radar-based fall detection based on Doppler time–frequency signatures for assisted living," *IET Radar, Sonar Navigat.*, vol. 9, no. 2, pp. 164–172, Feb. 2015.

[7] Y. Li, K. C. Ho, and M. Popescu, "A microphone array system for automatic fall detection," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 2, pp. 1291–1301, May 2012.

[8] Z. Chen, Q. Zhu, S. Y. Chai, and L. Zhang, "Robust human activity recognition using smartphone sensors via CT-PCA and online SVM," *IEEE Trans. Ind. Informat.*, vol. 13, no. 6, pp. 3070–3080, Dec. 2017.

[9] M. T. Hoang, B. Yuen, X. Dong, T. Lu, R. Westendorp, and K. Reddy, "Recurrent neural networks for accurate RSSI indoor localization," *Electr. Eng. Syst. Sci.*, Mar. 2019, pp. 1–13. [Online]. Available: http://arxiv.org/abs/1903.11703

[10] S. Sigg, U. Blanke, and G. Tröster, "The telepathic phone: Frictionless activity recognition from WiFi-RSSI," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun.*, Budapest, Hungary, Mar. 2014, pp. 148–155.

[11] Y. Gu, F. Ren, and J. Li, "PAWS: Passive human activity recognition based on WiFi ambient signals," *IEEE Internet Things J.*, vol. 3, no. 5, pp. 796–805, Oct. 2016.

[12] Y. Gu, L. Quan, and F. Ren, "WiFi-assisted human activity recognition," in *Proc. IEEE Asia–Pacific Conf. Wireless Mobile*, Bali, Indonesia, Aug. 2014, pp. 60–65.

[13] D. Halperin, W. J. Hu, A. Sheth, and D. Wetherall, "Predictable 802.11 packet delivery from wireless channel measurements," in *Proc. ACM Special Interest Group Data Commun.*, New Delhi, India, Sep. 2010, pp. 159–170.

[14] Z. Yang, Z. Zhou, and Y. Liu, "From RSSI to CSI: Indoor localization via channel response," *ACM Comput. Surv.*, vol. 46, no. 2, pp. 25:1–25:32, Dec. 2013.

[15] Y. Ma, G. Zhou, and S. Wang, "WiFi sensing with channel state information: A survey," *ACM Comput. Surv.*, vol. 52, no. 3, pp. 1–36, Jul. 2019.

[16] Y. Wang, J. Liu, Y. Chen, M. Gruteser, J. Yang, and H. Liu, "E-eyes: Device-free location-oriented activity identification using fine-grained WiFi signatures," in *Proc. 20th Annu. Int. Conf. Mobile Comput. Netw.*, Maui, HI, USA, 2014, pp. 617–628.

[17] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Understanding and modeling of WiFi signal based human activity recognition," in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw.*, Paris, France, 2015, pp. 65–76.

[18] H. Abdelnasser, M. Youssef, and K. A. Harras, "WiGest: A ubiquitous WiFi-based gesture recognition system," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Hong Kong, Apr./May 2015, pp. 1472–1480.

[19] H. Wang, D. Zhang, Y. Wang, J. Ma, Y. Wang, and S. Li, "RT-Fall: A real-time and contactless fall detection system with commodity WiFi devices," *IEEE Trans. Mobile Comput.*, vol. 16, no. 2, pp. 511–526, Feb. 2017.

[20] Z. Chen, L. Zhang, C. Jiang, Z. Cao, and W. Cui, "WiFi CSI based passive human activity recognition using attention based BLSTM," *IEEE Trans. Mobile Comput.*, vol. 18, no. 11, pp. 2714–2724, Nov. 2019.

[21] S. Yousefi, H. Narui, S. Dayal, S. Ermon, and S. Valaee, "A survey on behavior recognition using WiFi channel state information," *IEEE Commun. Mag.*, vol. 55, no. 10, pp. 98–104, Oct. 2017.

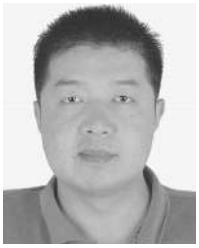
[22] J. Xiao, K. Wu, Y. Yi, L. Wang, and L. M. Ni, "Pilot: Passive device-free indoor localization using channel state information," in *Proc. IEEE 33rd Int. Conf. Distrib. Comput. Syst.*, Philadelphia, PA, USA, Jul. 2013, pp. 236–245.

[23] J. Lv, W. Yang, L. Gong, D. Man, and X. Du, "Robust WLAN-based indoor fine-grained intrusion detection," in *Proc. IEEE Global Commun. Conf.*, Washington, DC, USA, Dec. 2016, pp. 1–6.

[24] S. Sardy, P. Tseng, and A. Bruce, "Robust wavelet denoising," *IEEE Trans. Signal Process.*, vol. 49, no. 6, pp. 1146–1152, Jun. 2001.

[25] M. Inoue, S. Inoue, and T. Nishida, "Deep recurrent neural network for mobile human activity recognition with high throughput," *Artif. Life Robot.*, vol. 23, no. 2, pp. 173–185, Jun. 2018.
[26] D. Singh, E. Merdivan, I. Psychoula, J. Kropf, S. Hanke, M. Geist, and A. Holzinger, "Human activity recognition using recurrent neural networks," in *Proc. Int. Cross-Domain Conf. Mach. Learn. Knowl. Extraction*, Reggio Calabria, Italy, 2017, pp. 267–274.
[27] M. Z. Uddin, W. Khaksar, and J. Torresen, "Activity recognition using deep recurrent neural network on translation and scale-invariant features," in *Proc. 25th IEEE Int. Conf. Image Process.*, Athens, Greece, Oct. 2018, pp. 475–479.
[28] W. Y. Kwon, Y. Park, S. H. Lee, and I. H. Suh, "Human activity recognition using deep recurrent neural networks and complexity-based motion features," Georgia Inst. Technol., Atlanta, GA, USA, Tech. Rep., 2015, pp. 1–7.
[29] A. Murad and J.-Y. Pyun, "Deep recurrent neural networks for human activity recognition," *Sensors*, vol. 17, no. 11, p. 2556, Nov. 2017.
[30] J. Xiao, K. Wu, Y. Yi, L. Wang, and L. M. Ni, "FIMD: Fine-grained device-free motion detection," in *Proc. IEEE 18th Int. Conf. Parallel Distrib. Syst.*, Singapore, Dec. 2012, pp. 229–235.
[31] C. Wu, Z. Yang, Z. Zhou, X. Liu, Y. Liu, and J. Cao, "Non-invasive detection of moving and stationary human with WiFi," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 11, pp. 2329–2342, Nov. 2015.
[32] Y. Kim and H. Ling, "Human activity classification based on micro-Doppler signatures using a support vector machine," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 5, pp. 1328–1337, May 2009.
[33] P. Van Dorp and F. C. A. Groen, "Feature-based human motion parameter estimation with radar," *IET Radar, Sonar Navigat.*, vol. 2, no. 2, pp. 135–145, May 2008.
[34] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: Gathering 802.11n traces with channel state information," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 41, no. 1, p. 53, Jan. 2011.
[35] X. Li, S. Li, D. Zhang, J. Xiong, Y. Wang, and H. Mei, "Dynamic-MUSIC: Accurate device-free indoor localization," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, Heidelberg, Germany, 2016, pp. 196–207.

JIANYANG DING received the B.E. degree from Xidian University, Xi'an, China, in 2018, where he is currently pursuing the Ph.D. degree in indoor positioning. His current research interests include wireless communication and signal processing.



YONG WANG was born in Shaanxi, China. He received the B.S., M.E., and Ph.D. degrees from Xidian University, Xi'an, China, in 1997, 2002, and 2005, respectively. He is currently an Associate Professor with the State Key Laboratory of Integrated Services Networks, Xidian University. His research interests include full-duplex wireless communications and nonlinear predistorter technique.