Neurocomputing 444 (2021) 260–273

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Towards CSI-based diversity activity recognition via LSTM-CNN encoder-decoder neural network

Linlin Guo [a,*], Hang Zhang [b,*,1], Chao Wang [b], Weiyu Guo [b], Guangqiang Diao [c], Bingxian Lu [a], Chuang Lin [b,*], Lei Wang [a,**]

[a] School of Software Technology, Dalian University of Technology, China
[b] Shenzhen Institutes of Advanced Technology Chinese Academcy of Sciences, University of Chinese Academy of Science, China
[c] Shandong Youth University of Political Science, China

ARTICLE INFO

Article history:
Received 1 October 2019
Revised 17 December 2019
Accepted 3 February 2020
Available online 24 November 2020

Keywords:
WiFi signals
Human activity recognition
Channel state information
Long short term memory
Convolutional neural network

ABSTRACT

Human activity recognition using WiFi signals is widespread for smart-environment sensing domain in recent years. Existing researches use learning-based methods to obtain several features of activity data and then recognize human activities. As we know, propagation characteristics of WiFi signals are different for individuals under different place conditions even in the same environment. In this paper, we focus on how to weaken the accuracy differences among individuals on activity recognition and improve the robustness in one indoor environment. Based on this, we design a novel deep learning model called LCED which consists of one LSTM-based Encoder, features image presentation, and one CNN-based Decoder to weaken the accuracy differences among individuals on activity recognition. We first use a low-pass filter to remove high-frequency noise data in time-sequence signal data and design variance-based window method to determine the start and the end of time-sequence signal data corresponding to an activity. After that, we utilize the proposed LCED model to learn informative features space of activity data and improve the accuracy of sixteen activities. Experimental results show that the average accuracy of sixteen activities is high 95% and the accuracy differences among individuals on activity recognition averagely decreases by 3%.

1. Introduction

With the rapid development of wireless devices and the ubiquitous deployment of WiFi infrastructure, WiFi-based sensing has attracted increasing academia and industry attention. WiFi sensing provides non-intrusive and continuous sensing over the air, finding a new way to sense human activities without attaching any device. The early researches always utilize received signal strength indicator (RSSI) to roughly locate targets [30] and sense human behaviour like daily activity or gestures [1]. During the process of signals propagation, existing multipath effect leads to the performance instability of RSSI-based human sensing. Daniel et al. [10] firstly extracted channel state information (CSI) from Intel 5300 wireless card for indoor localization. Subsequent works proposed to use CSI for fall detection [11], gait recognition [29,5,31], gesture classification [23,18,21] and fine-grained vital signs [2,33]. The principle behind WiFi-based human activity recognition is that human activity affects the wireless signal propagation to cause different signal variation patterns. In other words, each signal variation pattern corresponds to an activity. Therefore, the learning-based method is typically applied to train a model for each signal variation pattern and infer the corresponding activity.

Recent learning-based works are divided into two classes: one class [28,21,14] explores how to extract effective features from raw activity data to recognize activities with high recognition accuracy; another class [25,38,4,36,24,13] explores how to construct a deep model from raw activity data to improve the robustness of human activity recognition in different scenarios. In the same indoor environment, these works [28,21,14,25] can achieve satisfying results. However, as we know, an activity recognition model that is trained on a specific individual in a specific indoor environment typically does not work well when being applied to predict another individual's activity recorded in another indoor

* Corresponding author.
** Principal corresponding author.
E-mail addresses: linlin.teresa.guo@gmail.com (L. Guo), hang.zhang1@siat.ac.cn (H. Zhang), chao.wang@siat.ac.cn (C. Wang), guoweiyu96@gmail.com (W. Guo), dgq@sdyu.edu.cn (G. Diao), bingxian.lu@dlut.edu.cn (B. Lu), chuang.lin@siat.ac.cn (C. Lin), lei.wang@dlut.edu.cn (L. Wang).
1 Equal contribution-first author.

https://doi.org/10.1016/j.neucom.2020.02.137
0925-2312/

environment. The interesting research [13] aims to address the above problem by designing a deep-learning model. The deep-learning model can remove the environment and individual of specific information contained in the activity data and extract environment/individual-independent features shared by the collected activity data on different individuals under different environments. The above solution needs precision devices to collect high-quality data for good performance. Based on the limit, Zhang et al. [35] firstly utilizes the theories of signal propagation (Fresnel Zone Model) to propose a diffraction-based sensing model which can quantitatively determine the signal change with respect to an individual's movements.

Based on the above analysis, we want to explore one problem that is to weaken the accuracy differences among individuals on human activity recognition. Compared with the work [13], we focus on individuals while it focuses on environments. On the technical side, the difficulty of our problem is easier but this problem is very important for the home-level indoor environment. We do not remove interference factors as well as the previous work did but focus on the shared features for various activities using the temporal and spatial models to improve recognition accuracy in different individuals. The challenge to solve the problem is how to extract the problem corresponding to the structure required in spatial–temporal features learning. In this paper, we propose a deep learning model named LSTM-CNN Encoder-Decoder model (LCED) to weaken the accuracy differences among individuals on human activity recognition. First, we try to encode the input WiFi time-series signal into a fixed-size vector by the Long Short-Term Memory (LSTM) and convert the vector into a two-bit grayscale image. Next, we use the idea of image classification method, taking the 2D grey map as input through the Convolutional Neural Network (CNN), extracting high-dimensional feature information and recognizing it. In this way, we solve the inconsistent problem of input and output lengths and time learning dependencies. The LSTM-based encoder encodes a time-sequence of WiFi signals into a fixed dimension of grey representation, and the CNN-based decoder decodes the different grayscale representation into the class of the activity. The encoder and the decoder of the LCED model are jointly trained to maximize the conditional probability of a vector of target labels given source labels.

2. Related work

We propose a novel deep learning model (LCED) to weaken the accuracy differences among individuals on human activity recognition. According to the goal, the related works are divided into learning-based activity recognition and model-based activity recognition. The details are shown as follows.

2.1. Learning-based activity recognition

Learning-based activity recognition learns effective features from large of activity data and leverages the extracted features to construct a learning model to recognize activities. In this paper, we mainly introduce several classic machine learning algorithms and deep learning models according to the learning character and the requirement of data scale. The learning algorithms used in WiFi-based activity recognition have Decision Tree (DT), Naive Bayes (NB), Dynamic Time Wrapping (DTW), k-Nearest Neighbor (kNN), Support Vector Machine (SVM), Hidden Markov Models (HMM), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), and Long Short-Term Memory (LSTM). The early researches use classic classification algorithms to recognize a single activity, gestures, simple behaviour, and daily activities [11,28,1,21,16,37,14,6,22]. WiFall [11] system extracts several statistical features to detect single fall using one-class SVM algorithm. WiSee [21] leverages the combination of Doppler Frequency Shift (DFS) and classification algorithms to remotely sense gestures for controlling appliance switch.

Due to the propagation characteristics of WiFi signals, WiFi-based activity recognition using machine learning cannot transfer the designed system to a new environment. Researchers begin to explore deep learning models to improve the robustness regarding WiFi-based activity recognition domain. The recent research works [25,39,38,4,36,24,13] explore how to leverage current neural network structures to construct a novel network model for the robust sensing system. These novel deep models help researchers to solve simultaneous localization and activity recognition [25], increasing the sample scale [38], constructing body skeleton using wireless signals to sense through-wall objects [36], improving the ability to sense granularity from sample level to data point [24] and weakening differences of indoor environments [13].

In summary, classification algorithms only can achieve simple effective information to satisfy the basic demand. However, deep learning model can excavate more effective information to satisfy the high demand for WiFi sensing in practical applications.

2.2. Model-based activity recognition

Model-based activity recognition is to model the signal pattern corresponding to an activity by theoretical models based on physical theories [35,27,19,32] or statistical models based on empirical measurements [26,27]. Theoretical models of WiFi signals mainly utilize fresnel zone model [35], multipath fading [19], and phase shift [32] to analyze the attributes of human behaviour and then recognize human behaviour. For example, the latest work [35] leverages the signal patterns reflected off moving person or human activity in the Fresnel zone to quantify the principles of signal propagation in terms of distance, multipath, and dynamic level. For statistical models based on empirical measurements, it mainly explores and determines laws of data change by statistical metrics. For example, CARM [26] explores the relationship between signal patterns and activities to build a CSI-Activity model for human activity recognition. The relationship depends on the quantitative analysis between body parts' speed and signal change. As we know, quantitative analysis is very difficult for dynamic signals.

3. Preliminary

3.1. WiFi signals

With the rapid development of related wireless techniques, wireless devices and smart devices, wireless signals increase the coverage range from research institutes, colleges and universities to public places and even home environments. Wireless signals not only can provide communication function but also have sensing function which has been attracting attentions inspired by the development of the artificial intelligence domain. Sensing function of wireless signals denotes that signals data can record dynamic change occurred in the indoor environment during the propagation process. Wireless signals sent by one transmitter propagate to one receiver through multiple paths, which consist of direct paths (Line-of-Sight, LoS) and reflected paths as shown in Fig. 1(a). LoS path denotes the shortest distance between the receiver and the transmitter and Time-of-Flight (ToF) in LoS path is also least compared to other reflected paths. The reflected paths are divided into static-based reflected paths and dynamic-based reflected paths. Static-based reflected paths caused by static objects like wall, ceiling and floor have stable distance and quantity within a short period in one indoor environment. Dynamic objectors change the

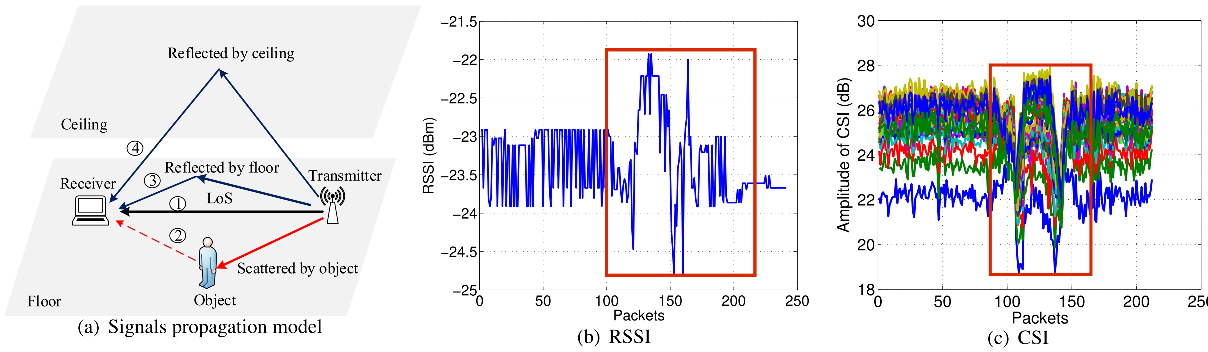


Fig. 1. WiFi signals. (a) During signal propagation process, WiFi signals face the signal attenuation with increasing distance and time-varying property due to the multipath effect. (b) RSSI represents the signal strength of all signal propagation paths. The time-sequence signal data covered by the red rectangular box corresponds to an activity. (c) CSI is a physical layer measurement with thirty subcarriers provided by the wireless network card. The red rectangular box covers CSI change corresponding to an activity. Note that the trends of CSI changes for thirty subcarriers are stable.

length of propagating paths named dynamic-based reflected paths which can precisely describe objects' behaviour. Research institutions exploit WiFi measurements in terms of coarse-grained RSSI and fine-grained CSI. In this paper, we briefly introduce WiFi RSSI and pay more attention to the research of CSI. The details are shown as follows.

3.1.1. RSSI (Received Signal Strength Indicator)

Commodity wireless devices can record RSSI without additional hardware like a smartphone. RSSI is a sum of signal energy from multiple paths which include a direct path (LoS path) between the transmitter and the receiver, and multiple reflected paths caused by walls, furniture, and people in the macro-view as shown in Fig. 1(b). RSSI denotes the received signal power in decibels (dBm) [30]:

$$RSSI = 10log_2\left(\|V\|^2\right) \tag{1}$$

where V denotes signal voltage. Due to the multipath effect, it is hard to distinguish signal components produced by different paths in an indoor environment. Therefore, we consider RSSI as coarse-grained information of WiFi signals to roughly sense dynamic change such as human movement. In this paper, we use RSSI to evaluate the change of a meeting room. Once the variance of RSSI change exceeds the threshold value, we need to consider the effect of the indoor environment on human activity recognition.

3.1.2. CSI (Channel State Information)

CSI describes how signals propagate in the wireless channel combining the effect of time delay, energy attenuation and phase shift [11]. Compared with RSSI, CSI represents fine-grained information of WiFi signals with thirty subcarriers as well as rainbow reflected by sunlight. Leveraging the off-the-shelf Intel 5300 NIC with a modified driver, a group of sampled versions of channel frequency response (CFR) within the WiFi bandwidth is revealed to upper layers in the format of channel state information [10]. CSI of a single subcarrier is in the following mathematical formula:

$$H(k) = \|H(k)\|e^{j\angle H(k)} \tag{2}$$

where $H(k)$ is a CSI of the kth subcarrier. $\|H(k)\|$ and $\angle H(k)$ are CSI amplitude and CSI phase respectively. CSI can capture more fine-grained changes like gestures, breaths and heartbeats. Thirty subcarriers have various sensitivities to the same activity due to existing frequency selective fading. This characteristic is utilized to explore the relationship between signal patterns and activities. As shown in Fig. 1(c), CSI has a unique signal pattern produced by an activity, and we use different colours to represent 30 subcarriers.

3.2. CNN and LSTM

In this paper, we combine the advantages of CNN and LSTM to design an encoder-decoder model which can explore the stability of human activity recognition regarding accuracy using CSI. In the following part, we introduce the related details of CNN and LSTM to better explain why we select both network structures.

3.2.1. CNNs (Convolutional Neural Networks)

CNN is a classic neural network structure which can extract spatial features to represent objects and compress the raw data. Compared with manual-based features extraction, CNNs can select a kernel function to obtain enrich features maps with an optimal way. CNNs have three basic blocks including convolution operation, activation function and pooling as shown in Fig. 2. The convolution operation can select one filter to compute one feature map. For the activation function in CNN, *ReLu* function is commonly used to keep the raw values for features with greater than 0 and 0 instead of the raw values of features with less than 0. This operation can transform linear results into non-linear results. The pooling block only keeps optimal features in the feature map to realize data compression. The number of basic blocks in the dotted can be increased or reduced according to the actual demand of the research. Moreover, several works add one full connected layer or several full connected layers into CNNs to achieve the corresponding probability of each class.

3.2.2. LSTM-based RNN (Long Short-term Memory)

LSTMs can avoid the long-term dependency problems including gradient explode, gradient vanish by designing three gates, which contain forget gate layer, input gate layer and output gate layer [15]. The key to LSTMs is the cell state corresponding to one rectangular box as shown in Fig. 3. The details of each layer of LSTM are supported by the related techblog [20][2].

Forget gate layer determines what information to be given up from the cell state. Fig. 4 demonstrates that the process of how to decide what information to be neglected by the output h_{t-1} of the previous cell state and current input data x_t. The forget gate layer outputs f_t which is between 0 and 1 for each number in the previous cell state C_{t-1} where 1 represents completely keeping this while 0 represents completely getting rid of this.

Input gate layer decides what new information we are going to store in the cell state as shown in Fig. 5. There are two parts. First, a sigmoid layer called the input gate layer decides which values we

[2] In this paper, we directly use the related details of LSTM provided by the related techblog [20] and combine with our activity data to analysis.

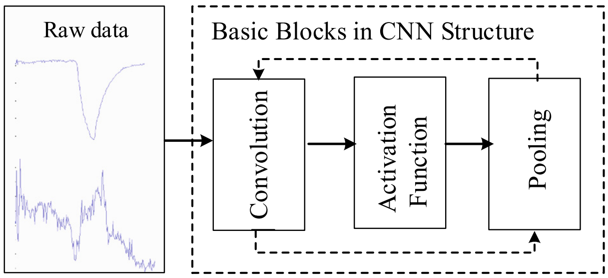


Fig. 2. Structure of CNN.

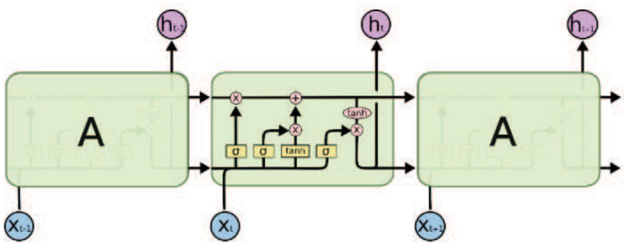


Fig. 3. Structure diagram of LSTM provided by [20].

update. Next, a *tanh* layer creates a vector of new candidate values, C_t, that could be added to the cell state. And then we combine the two parts to create an update to the cell state. Meanwhile the old cell state, C_{t-1}, is updated into the new cell state C_t. The previous steps already decided what to do, we just need to follow it. We multiply the old state C_{t-1} by f_t and forget the things we decided to forget earlier. Then we add $i_t * \widetilde{C}_t$. This is the new candidate values, scaled by how much we decided to update each value of cell state.

Output gate layer decides what we are going to output as shown in Fig. 6. First, we run a sigmoid layer that decides which parts of the cell state we are going to output. Then, we put the cell state through *tanh* and multiply it by the output of the sigmoid gate. Finally, we only output the parts we decided to.

3.3. Motivation

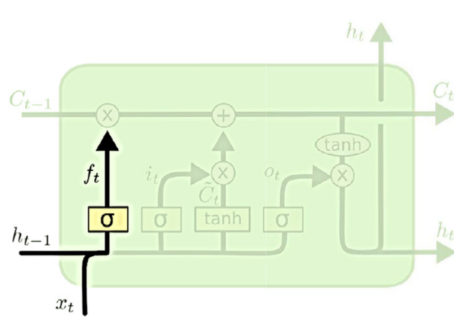
The previous works utilize machine learning algorithms and deep learning algorithms to improve the accuracy of human activity recognition. The accuracy of human activity recognition exists instability for different individuals and even the same individual in one indoor environment. Based on this challenge, we design a model based on CNN and LSTM to improve the accuracy stability of human activity recognition on individuals. Note that we emphasize the differences among persons regarding human activity recognition in this paper. The differences between indoor environments are beyond the range in this paper. **Existing problems and the corresponding solutions are shown as follows**.

The accuracy difference of the same individual. In other words, the collected activity data with respect to the same activity at different times have a slight change in signal patterns due to signal propagation. Therefore, the change leads to the instability of the activity recognition accuracy. For example, the accuracy difference of the same person's activity recognition using machine learning algorithms is between 5% and 10%, the early version [8] between 3% and 7%. Although the early version can reduce the accuracy difference, the result still cannot be up to the demand in daily life. To solve the problem, we design the LSTM-CNN Encoder-Decoder model to reduce generalization error about the accuracy of activity recognition. Compared with the early version, we increase three convolution layers and two full-connection layers to learn more effective information from the raw activity data.

The accuracy differences among different individuals. Each individual has a unique behaviour habit during the process of operating an activity. The behaviour habit consists of speed, strength and the characteristics of the human body. The speed and strength factors affect signal pattern in terms of frequency, the speed of signal change and signal fluctuation except for the trend of signal change. In other words, the key we can recognize the same activity operated by different individuals is to depend on the trend of signal change. Therefore, we design a novel network structure to achieve global characteristics in time-sequence data, not local characteristics.

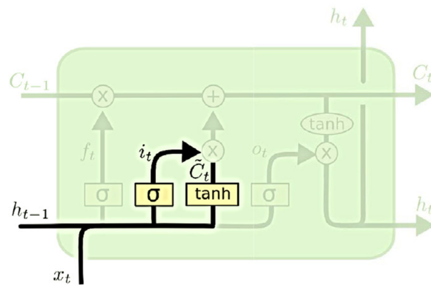
3.4. System overview

In this paper, we mainly design a novel neural network model named LCED for learning time-sequence signals data. The LCED model consists of LSTM-based encoder and CNN-based decoder to enhance the learning ability and weaken accuracy differences among individuals on human activity recognition. The overview of this system is shown in Fig. 7. In the collected phase, we aim to collect activity data of individuals under different distances and heights in one indoor environment. The collected signals data is time-sequence, interval and instability in the receiving process for human activity recognition. The signal pattern reflected on the same activity is different from each other over time. Moreover, dynamic environment, moving behaviour produce a large fluctuation on the collected signals data. In the pre-processing phase, we use the low-pass filter to remove the high-frequency noise and design a variance-based window method to determine the start and the end of time-sequence data corresponding to an activity. Principal component analysis (PCA) helps us to roughly analyze the quality of collected activity data and extract the informative principal components from the collected sequence signals. To better make a comparison between machine learning algorithms and



$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \ + \ b_f\right)$$

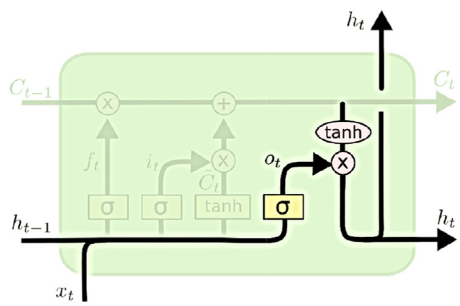
Fig. 4. Forget gate layer of LSTM provided by [20].



$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \ + \ b_i\right)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \ + \ b_C)$$

Fig. 5. Input gate layer of LSTM provided by [20].



$$o_t = \sigma\left(W_o \ [h_{t-1}, x_t] \ + \ b_o\right)$$

$$h_t = o_t * \tanh\left(C_t\right)$$

Fig. 6. Output gate layer of LSTM provided by [20].

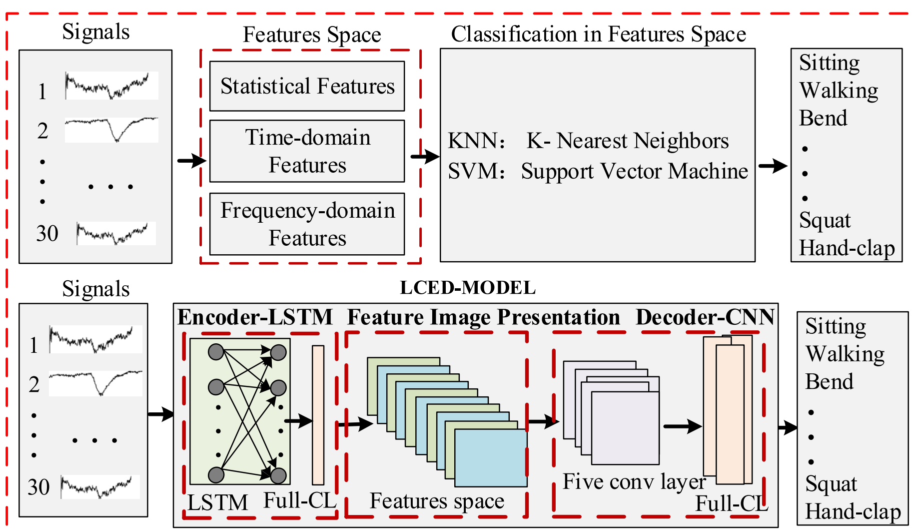


Fig. 7. Schematic illustration of human activity recognition using machine learning and deep learning algorithms.

deep learning algorithms regarding CSI-based human activity recognition, we design two schemes respectively. For the machine learning algorithms, we extract several statistical features as inputs to evaluate the performance of individuals on activity recognition. For the proposed LCED model, we directly learn informative time-features through LSTM-based encoder and use a CNN-based decoder to obtain features map and complete activity recognition.

4. LCED neural network

To improve the robustness and universality of the multi-class activities recognition under the diverse individuals, we propose a novel encoder-decoder model called LCED. The architecture of the LCED model is described in Fig. 8, which has twelve layers including the LSTM layer, convolutional layer and fully connected layer. First, the raw signals data are treated as inputs of the encoder to extract the fixed featured vector. Then, the vector is transformed into a featured image. Finally, the spatial features of the image are extracted and classified by CNN. The idea to design the LCED model is based on the following principles: (i) The diversity activities have the distinct clusters in the space according to the visualization of PCA technique, and the featured image presentation produces diverse intensity distribution among the activities. (ii) The encoder part of the LCED is capable of learning the time-series representation and can encode it to a fixed-length vector using nonlinear transformation. What is more, the fixed-length feature vector is automatically extracted by the proposed model from the raw signals compared with the original methods based on the manual way, which can adaptively extract features and

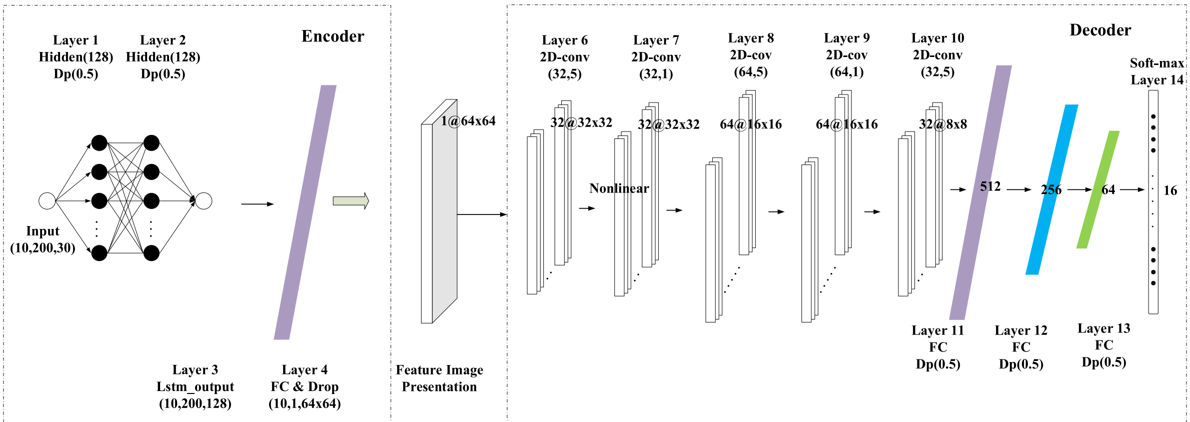


Fig. 8. LCED model: LSTM-based Encoder and CNN-based Decoder.

compress the key information corresponding to the differences among the individuals and activities. (iii) The decoder part is based on CNN which can effectively capture the local features of the image based on the spatial distribution and has better classification accuracy. In general, the proposed model has better robustness and can be more generalized due to the adaptive spatiotemporal feature extraction.

4.1. The encoder-LSTM

Our encoder has three layers, and it is designed by the LSTM to extract the fixed featured vector. The input to the encoder consists of a 200-length sequence of the 10 batch size with 30 channel subcarriers. The first two hidden layers are LSTM block layers, each of which consists of 128 hidden units with dropout 0.5. The last hidden layer is a fully connected layer that can compress the output of the LSTM into a fixed feature vector of size $64 * 64$. In this encoder, the time t is from 1 to 200, and the last output is used to feature extraction and reduction by the FC (Full Connected) layers. The output of the featured vector output is shown as follows:

$$h_t = \psi(x_t, c_{t-1}, h_{t-1}) \tag{3}$$

$$y_t = \sigma\left(W_y h_t + b_y\right) \tag{4}$$

where x_t is channel state information (CSI) after filtering and y_t denotes the output of LSTM respectively. The c_t represents the status of a cell unit and h_t is the hidden vector for each layer. The σ is an activation function and *logsoftmax* is regarded as the activation function. The W_y is the matrix weight needed to be optimized. The parameter b_y is a bias of in the activation function. After passing through the LSTM layers, the LSTM output is a fixed vector and then feeds into a FC layer for features extraction and reduction. Finally, we can get the output with a $64 * 64$-dimensional featured vector.

4.2. The decoder-CNN

For the designed decoder, there are nine layers. The output of the encoder is transformed to feature image presentation of size $(64 * 64)$ regarded as the input of the decoder. The first five hidden layers are combined with convolutional layers and locally connected layers. The convolutional layers consist of 32 filters of $5 * 5$ with stride 1 and 64 filters of $5 * 5$ with stride 1, which can extract the spatial feature distribution corresponding to the image intensity and the local feature representation. The two locally connected layers consist of $32, 64$ non-overlapping filters with $1 * 1$ are added after each convolutional layer. The convolution layers operation can be equal to the downsampling, which can learn more local features and extract high-level abstract feature representation. The locally connected layers (1x1) can be equivalent to the nonlinear transformer using fewer parameters to augment the ability of the model to learn more complex features [7]. The next three layers are FC and composed of $512, 256, 64$ units with dropout 0.5, respectively. The decoder ends up with an FC with a *LogSoftmax* function to classify sixteen action categories. The detail calculation can be expressed as follows:

$$\sigma^l_m(j) = b^l(j) + \sum_{i=1}^{i \leqslant N^l_{kernel}} \left(I^{l-1}_{i,j} k_1(l,m,i) + I^{l-1}_{i,j+1} k_2(l,m,i) \right) \tag{5}$$

$$I^l = \left[\sigma^l_1(j), \sigma^l_2(j), \sigma^l_3(j), \cdots, \sigma^l_m(j)\right]^T \tag{6}$$

$$out\_class = max(FC(\sigma * W + b)) \tag{7}$$

where $b^l(j)$ refers to the bias in the l layer for the neuron. The N^l_{kernel} is the number of the kernel used in layer. The $k_1(l, m, i)$ is the filter used in neural network. The scalar of l, m, i and w represent the layer, the map, the *ith* kernel and the position in the kernel respectively. $I^l_{i,j}$ represents the *Ith* element which the *jth* neuron in *ith* kernel of the features extraction maps after convolution from the previous layers or the initial elements. Then, *Logsoftmax* function is applied to maximize the probability distribution of the last output layer $FC(\sigma * W + b)$ to get the classification category corresponding to the given signals.

In the proposed model, the batch normalization (BN) is adopted after each convolutional layer, which can effectively avoid the phenomena of the internal covariate shift and optimize the parameter settings of the model [12]. Similarly, the *ReLu* function is added after each decoder layer to improve the learning ability of the model and it enhances the robustness among the diverse individuals.

5. Experiments and implement

The section mainly explains the details of the experimental setting, pre-processing of activity data and the evaluation metrics of the proposed LCED model.

5.1. Experimental setting

This paper aims to design a novel deep learning model to reduce the accuracy differences among different individuals on human activity recognition. Based on this thought, we select one meeting room within several furniture as the only experimental scenario. Fig. 9 shows the details of the experimental environment in terms of furniture layout, devices layout and the size of the meeting room. We collect activity data using two T400 laptops within Intel 5300 wireless network card supported by the CSI-TOOL [9] which can be widely used for WiFi-based applications. One T400 laptop with one antenna as a transmitter transmits signals data, and another one with three antennae as a receiver receives the propagated signals data from different reflected paths. The packet delivery fraction is $30pkt/s$ which is widely used in WiF-based applications. The packet delivery fraction can keep sensory ability and does not affect normal network applications. The collect platform can not only real-time receive signals data but also accomplish real-time signals pattern display in the process of doing activities. It can help us to discover the low-quality signal data without delay and improve the quality of the collected signals. We use activity data of 10 volunteers to analyze the accuracy differences among individuals on activity recognition. Ten volunteers contain four females and six males whose ages are from 18 to 30 and the heights are from 160cm to $190cm$. To explore the impact of different distances between the transmitter and the receiver on human activity recognition, we design three levels regarding the distances. Note that volunteer locates in the middle of the two laptops. The three distances are $1m, 3m$ and $6m$ respectively. For the impact of different heights on human activity recognition, we also design three levels regarding the height between the receiver and the floor. The heights are $0.6m, 0.9m$ and $1.2m$ respectively.

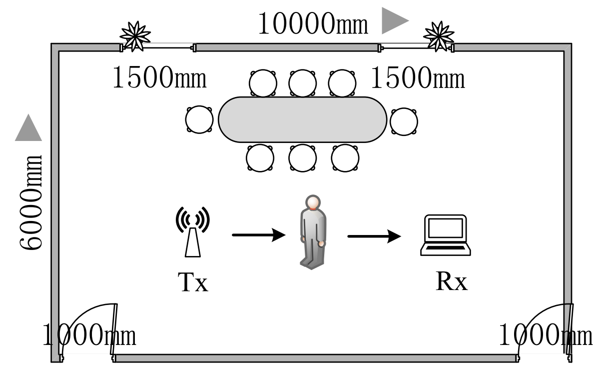


Fig. 9. Experimental environment. The length and width of the meeting room are $10m$ and $6m$ respectively.

5.2. Activity data

The section mainly introduces the details of activity data and leverages PCA technique to validate the quality of collected WiFi activity data. We collect sixteen activities for ten volunteers in a meeting room. The related definitions of sixteen activities: Horizontal arm wave (0), High arm wave (1), Two hands wave (2), High throw (3), Draw x (4), Draw tick (5), Toss paper (6), Forward kick (7), Side kick (8), Bend (9), Hand clap (10), Walking (11), Phone call (12), Drink water (13), Sitting (14), and Squat (15). Each activity has thirty samples and each sample is time-sequence signal data[3] which is a $30 * packets * 3$ matrix. The format of the matrix describes thirty subcarriers[4], the length of time-sequence signal data and three antennas. As we know, the data error brought by the collected activity data offsets the differences among individuals' activities. To explore the accuracy difference of single person and multiple people regarding activity recognition, we first ensure the quality of collected activity data and reduce the influence of activity data self on human activity recognition. We use eigendecomposition-based PCA to analyze the distribution of each activity in activity dataset and roughly evaluate the quality of activity data. The Fig. 10 demonstrates the activity distribution of each antenna for randomly selecting two volunteers. The activity data of the volunteer2 is superior to the volunteer3's data because the former activity data has a distinct boundary of each class. However, the latter cannot distinguish some activities from sixteen activity data due to the overlap of some activities. Once occurring the bad/low quality of activity data, we analyze caused reasons and design two ways to deal with them.

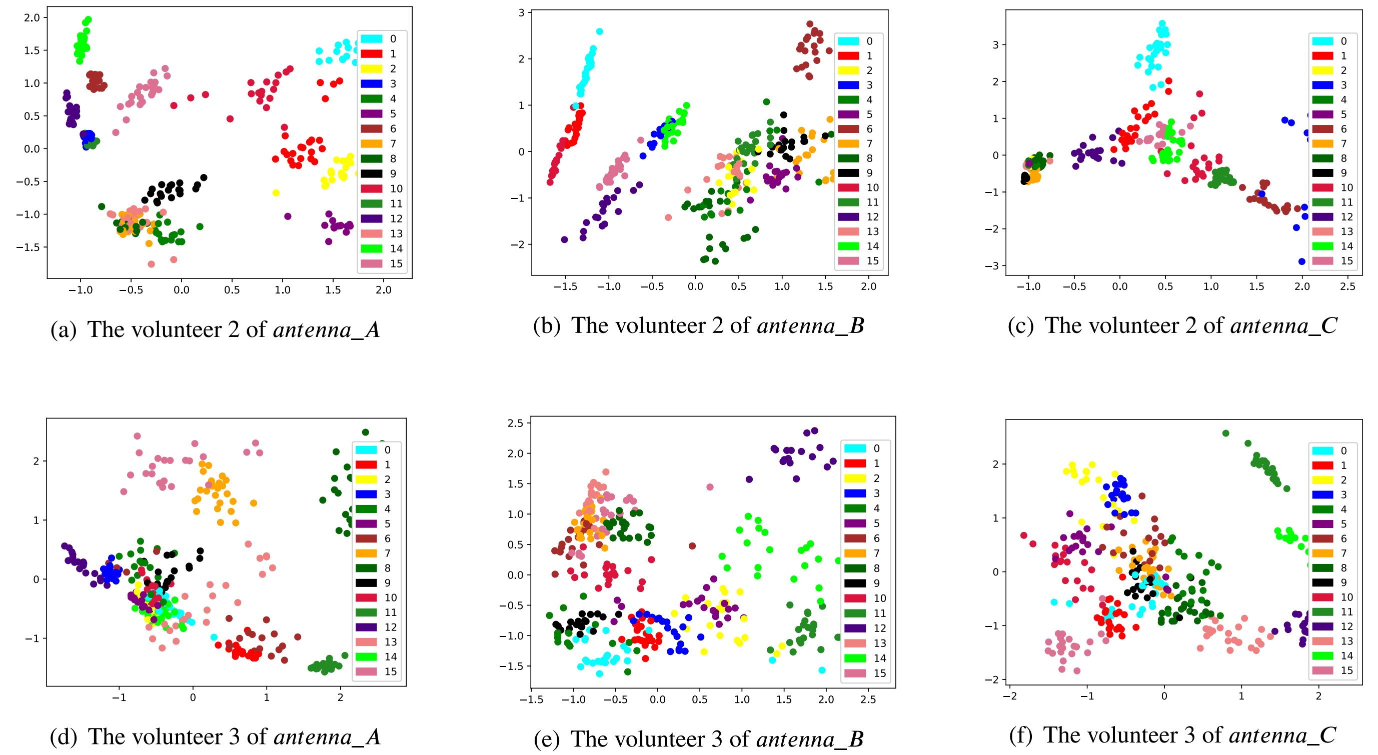
- Bad-quality data caused by device faults. The device faults contain the poor contact between antennas and the mainboard, reading an issue of the wireless network card and high drop rates caused by a bad network environment. Device faults decrease the analyzability of the collected activity data due to large of empty packets or irregular change of time-sequence data corresponding to an activity. For the bad quality activity data caused by these reasons, we directly ignore these activity data.
- Low-quality data caused by individual behaviour. Individual behaviour issues stem from non-standard activity, body shaking and velocity change of human activity. The non-standard activity and body shaking can change the corresponding signal pattern and the corresponding frequency. Human activity with high/slow speed can change the corresponding frequency but cannot change the trend of the signal pattern. Compared with the bad-quality activity data, the low-quality activity data have a chance to be recovered using the data analysis technique.

5.3. Pre-processing

The pre-processing is an important step which aims to reduce noise data and removes outlier for achieving high-quality activity data. During our work, we design an LSTM-CNN based deep learning model to reduce the accuracy error of human activity recognition caused by individuals. To make a comparison between machine learning algorithms and the proposed deep learning model regarding human activity recognition, we need to deal with raw activity data to meet the demands of the data format for the above algorithms respectively. Note that the emphasis of this paper still focuses on the proposed deep learning model. The common step of the two ways is to utilize a low-pass filter technique to remove high-frequency signals corresponding to noises and remain with low-frequency signals corresponding to an activity. After that, we design a variance-based moving window to determine the start and the end of time-sequence activity data corresponding to an activity. The remaining data is treated as the goal of the pre-processing phase.

- ML-based pre-processing needs to extract effective feature in a manual-based way. In this paper, we extract statistical features of the time-sequence signals corresponding to an activity as inputs of machine learning algorithms such as KNN, SVM. Before extracting statistical features, we use an iteration method to determine the optimal length of the sliding window to smooth the time-sequence signals. The statistical features include average value, variance and standard deviation. Moreover, we also explore time-domain features like signal waveform, correlation coefficient with time as evaluation metrics to analyze each activity sample.

[3] Here only introduces CSI because RSSI is not an emphasis in this paper.
[4] Each subcarrier value in frequency-domain is a format of CSI



(a) The volunteer 2 of *antenna_A* (b) The volunteer 2 of *antenna_B* (c) The volunteer 2 of *antenna_C*

(d) The volunteer 3 of *antenna_A* (e) The volunteer 3 of *antenna_B* (f) The volunteer 3 of *antenna_C*

Fig. 10. The visualization of the two volunteers' spatial category distribution in 2-dimensional based PCA.

- DL-based pre-processing has two goals to be completed. One is to determine the start and the end of the time-sequence signal corresponding to an activity. We leverage variance-based moving window function and adaptive threshold to determine the start and end of each time-sequence signals regarding an activity. Another one is to cut out the rest of more than 200 packets and be filled with less than 200 packets for given time-sequence signals. Due to the difference of activities at the time cost, we set a size of 200 packets for each activity sample to meet the need of data format for the proposed deep learning model.

5.4. FIP (Featured Image Presentation)

In this experiment, we design a novel classification method based featured image presentation. We convert the 4096-dimensional vector which is the output of the encoder into an image ($64 * 64$) by linear transforming. Once time-sequence signal data is transformed into an image, we can direct use several image-based techniques to analyze and extract key information for high-stability performance. Fig. 11 describes featured image presentation of five activities on the *antenna_A* for two volunteers. The featured image records the spatial distribution of the WiFi signals. In this way, it can integrate the spatial with temporal characteristics of the raw signals. Besides, the max–min is applied to normalize the pixel of the image for improving the training speed.

5.5. Training of LCED

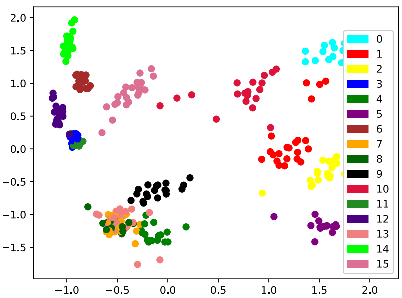
We evaluate our model using activity data of seven individuals. In the training of LCED experiments, we divide all dataset by a ratio of 3 to 1. For the hyper-parameters, the batch size is set as 10, and the input was randomly chosen samples from the training set. We apply 1000 trials for model training for learning the deeper features. Regraded as the optimizer, the *Adam* is initialized to 0.001 to update the parameters with less memory requirement, and the *CorssEntropyLoss* function is applied to calculate the error loss between the predicted value and the real one while training, the random initialization is used to initialize the model weights, it can effectively break the symmetry problem. Moreover, the dropout with 50% is adopted after each fully connected layers. After completing each training epoch, the LCED model is set as the test mode, and then we experiment with the entire testing set to evaluate the performance of the proposed model.

5.6. Reference models

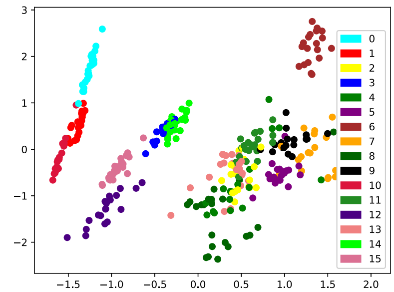
To verify the stability and accuracy, we compare the proposed LCED model with the diverse reference models in the time-series classification problem. These methods are described as follows.

KNN is universally used in multi-class classification. We implement the KNN program using the scikit-learn library. Firstly, we manually extract the mean value of 30 subcarriers of root antenna as the feature extraction of WiFi signals and normalize the processed signals. Next, the PCA is applied to reduce the feature dimension to 3*D* space. Lastly, the KNN model is built with neighbours 3 to train and test the all set. For the evaluation of the performance, we calculate three analysis measures: *Accuracy*, *F1Score* and *Recall* which these are served as a general evaluating certificate.

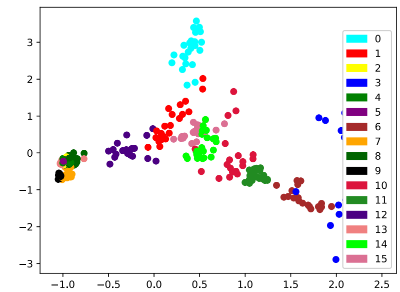
SVM algorithm is often used on nonlinear binary classification tasks. In the same way, the SVM is built for data preprocessing and model training as KNN. For the evaluation of the performance, we calculate the accuracy and plot the matrix for each action. sRNN is a variant of RNN, which has a simple structure and is consisted of a single hidden layer 128. The same machine learning-based preprocessing method is applied to divide the datasets. For hyper-parameters, the sRNN is initialized in the same way as the proposed LCED. Besides, the K-Average operation is utilized to estimate the accuracy of the LCED performance.



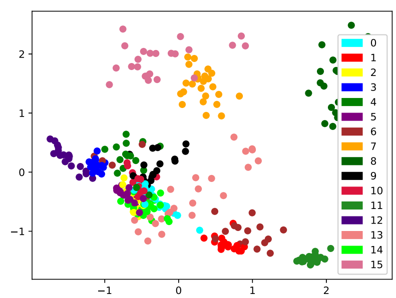
(a) The volunteer 2 of *antenna_A*



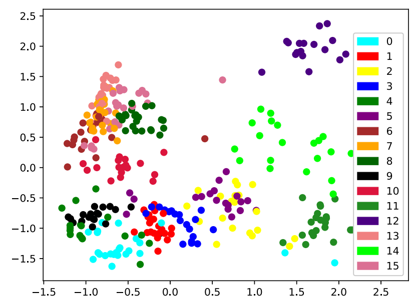
(b) The volunteer 2 of *antenna_B*



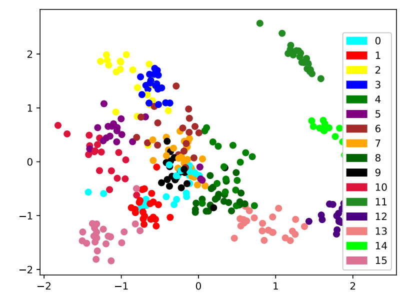
(c) The volunteer 2 of *antenna_C*



(d) The volunteer 3 of *antenna_A*



(e) The volunteer 3 of *antenna_B*



(f) The volunteer 3 of *antenna_C*

6. Evaluation

We analyze the performance of human activity recognition using machine learning algorithms and the proposed LCED model in terms different individuals, different distances between one transmitter and one receiver, and different heights between wireless devices and the floor. The details of these analyses are shown as follows.

6.1. Testing performance measures

We use three evaluation indicators including *Recall*, *F1-score*, and *K-Average accuracy* to evaluate the performance of human activity recognition using ML algorithms and the proposed LCED model.

- *Recall*: the number of detecting true samples from the whole samples.
  $Recall=\frac{TP}{TP+FN}$.
- $F1-score$: to evaluate and balance the contradiction between *Precision* [5] and *Recall* to achieve the science results.
  $F1-score=\frac{Precision* \ Recall}{Precsion+Recall}$
- K-average accuracy[6]: taking an average of k times as the final result.$Accuracy = \frac{1}{K} * \sum_{k=0}^{K} acc_k$

[5] *Precision*: the number of detecting true samples from the number of used samples. $Precision=\frac{TP}{TP+FP}$

[6] Usually, the entire testing sets are once directly calculated to attain the accuracy of the model. In this case, the method has a limitation because the distribution of data in the testing set is generally uneven. Hence, evaluating the performance at different distribution scales is necessary. The $K-fold$ is widely used in model verification. Owing to this idea, we propose a k-average accuracy calculation using *kth* average accuracy as a result. During the testing process, we perform four times randomly shuffle testing data selection, and calculate the average of the four results.

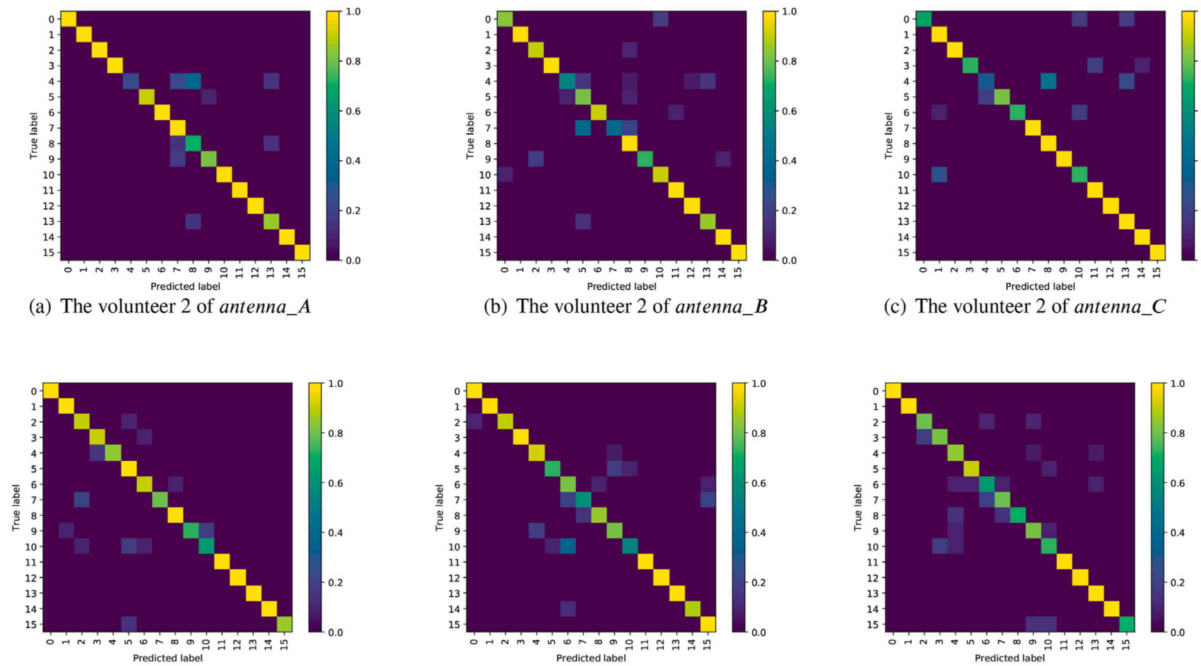
6.2. Performance analysis

6.2.1. Performance of individuals

The research goal is to explore the impact of individuals on the accuracy of human activity recognition. We utilize KNN, SVM, RNN, and the proposed LCED model to evaluate the performance of individuals regarding human activity recognition using CSI.

As shown in the Fig. 12, the top row shows the accuracy of the first volunteer on human activity recognition using the SVM algorithm. The *antenna_A*'s performance is superior to the *antenna_B* and the *antenna_C*. There are some activities with low accuracy in three antennae. According to the experimental analysis, we infer two reasons: one is the low-quality data; another one is that the activity is hard to be recognized by using SVM. The second row is the second volunteer's accuracy for human activity recognition. Compared with the first volunteer, we observe that existing the same activity achieves low accuracy. The Table 1 shows the performance of sixteen activities using KNN and Fig. 13 shows the accuracy of an individual on two antennas using the proposed LCED model.

We randomly select activity data of three volunteers as analysis objects and the results are shown in Fig. 14. According to the comparison of three volunteers regarding recognition accuracy using KNN, the second volunteer achieves an average accuracy of 90%, high 80% accuracy for the third volunteer, and the worst accuracy is the first volunteer with average 77%. Although the first volunteer's accuracy still is lower than the other two individuals using SVM, RNN, the stability of recognition accuracy on three antennas is higher than the other two individuals. Compared with KNN, SVM, and RNN, the accuracy of activity recognition increases 10% on average using the proposed LCED model. The important performance is to not only reduce the difference of recognition accuracy on three antennas but also enhances the robustness of activity recognition on individuals.



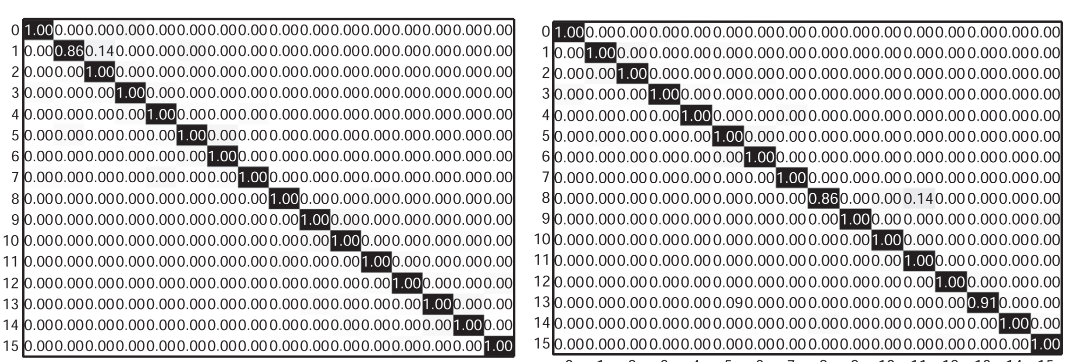
(a) The volunteer 2 of *antenna_A* (b) The volunteer 2 of *antenna_B* (c) The volunteer 2 of *antenna_C*

(d) The volunteer 3 of *antenna_A* (e) The volunteer 3 of *antenna_B* (f) The volunteer 3 of *antenna_C*

Fig. 12. Two volunteers and three antennae: the accuracy of sixteen activities using SVM.

Table 1
KNN performance for one individual with two antennae.

	Antenna_A			*Antenna_B*		
	Precision	Recall	F1-score	Precision	Recall	F1-score
0	1.00	0.83	0.91	0.62	0.83	0.71
1	0.89	1.00	0.94	0.88	0.88	0.88
2	1.00	1.00	1.00	0.56	0.90	0.69
3	1.00	1.00	1.00	0.79	1.00	0.88
4	0.50	0.62	0.55	0.62	0.38	0.48
5	0.91	0.91	0.91	0.57	0.73	0.64
6	1.00	1.00	1.00	1.00	0.73	0.84
7	0.57	0.80	0.67	-	-	-
8	0.43	0.43	0.43	1.00	0.71	0.83
9	0.77	0.91	0.83	1.00	0.27	0.43
10	1.00	0.91	0.95	0.91	0.91	0.91
11	1.00	1.00	1.00	0.88	0.88	0.88
12	1.00	1.00	1.00	0.83	1.00	0.91
13	1.00	0.43	0.60	0.45	0.71	0.56
14	1.00	1.00	1.00	1.00	1.00	1.00
15	1.00	0.71	0.83	0.70	1.00	0.82



(a) The confusion matrix of the volunteer 4 about *antenna_A* (b) The confusion matrix of the volunteer 4 about *antenna_B*

Fig. 13. One individual and two antennas: the accuracy of sixteen activities using LCED model.

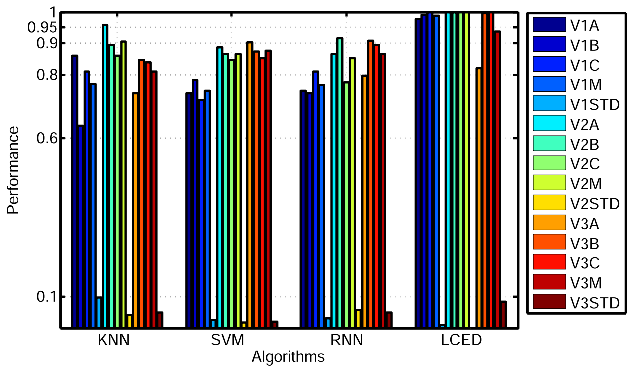


Fig. 14. Performance analysis of different algorithms. *V1A* is the accuracy of the first volunteer's activity data on *antenna_A*. *V1M* denotes the average accuracy of the first volunteer's activity data on three antennas' (*antenna_A*, *antenna_B*, *antenna_C*). *V1STD* is the standard deviation of the first volunteer's activity data on three antennas. The definition of the following is the same.

In the following parts, we analyze the accuracy difference of three antennas on human activity recognition. Fig. 15(a) shows the accuracy of the *antenna_As* of different individuals as the analysis object. The three antennas have stability accuracy with high 80% on average accuracy respectively and up to 95% accuracy by using LCED model. The standard deviation of the three antennas is less than 0.1 for KNN, SVM, and RNN. Compared with the above, the standard deviation of the *antenna_B* and *antenna_C* using LCED model is less than 0.01. Fig. 15(b) is to show the accuracy of three antennas of different individuals as an analysis object by using LCED model. The average accuracy of three antennas regarding activity recognition is higher than 93% and the standard deviation is less than 0.02 except for the performance of the third individual.

6.2.2. Performance of different distances

We set three levels of distances to evaluate the impact of different distances on human activity recognition using the proposed LCED model. First, we discuss why we select $1m, 3m, 6m$ as examining distance. As we know, the communication distance of WiFi signals is less than $15m$ and the sensing distance with satisfying accuracy is less than $5m$. Once occurring interference or fine-grained behaviour sensing, the sensing distance is less than $3m$. Fig. 16(a) shows the accuracy differences of different distances on activity recognition. We observe that three of four algorithms achieve similar accuracy of activity recognition except for KNN with less than 90%. Compared with ML algorithms, although the proposed LCED model has no best performance, the accuracy stability of three antennas on activity recognition is the best. This performance indicator is an important factor to evaluate the applicability of the proposed LCED model in a practical indoor environment.

Compared with the accuracy of the other two distances, the accuracy of activity recognition on $6m$ distance is the best and this result contradicts the previous common sense. We discuss the interesting issue occurred in experimental results. The issue is why is a small accuracy difference of three distance on activity recognition. In other words, what is the limit range of sensing distance using commodity wireless devices to sense human behaviour. From the results shown in Fig. 16(a), the $6m$ distance cannot up to the limit of sensing distance [34,3]. Therefore, the following research will explore the issue.

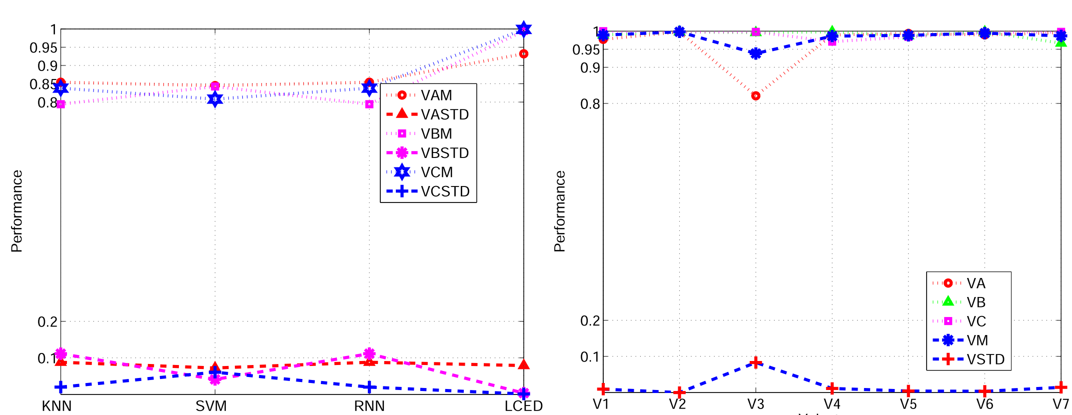
6.2.3. Performance of different heights

We set three levels of heights to evaluate the impact of different heights on human activity recognition using the proposed LCED model. First, we give reasons why we select $60cm, 90cm, 120cm$ of heights as evaluated standards respectively. The $60cm$ of height corresponds to the lower-body to capture the lower-body activities like a side-kick. The $90cm$ corresponds to the torso part to sense whole-body activities like running. The $120cm$ of height corresponds to upper-body activities to recognize gestures like hand wave. Fig. 16(b) shows the accuracy of three heights on human activity recognition using four algorithms. The results of KNN, SVM, RNN are less than normal level with 80% due to the low-quality activity data. The reason stems from additional interference caused by non-standard activities. The proposed LCED model can offset the passive effect of additional interference on activity recognition and achieve an average accuracy of 98.5%.

Compared with the accuracy of the other two heights, the accuracy of activity recognition on $120cm$ is best. According to the accuracy analysis of three heights on human activity recognition, we observe that the proposed LCED model can only offset the passive effect of low-quality activity data but also keep the high stability of the accuracy on different heights.

6.3. Discussion

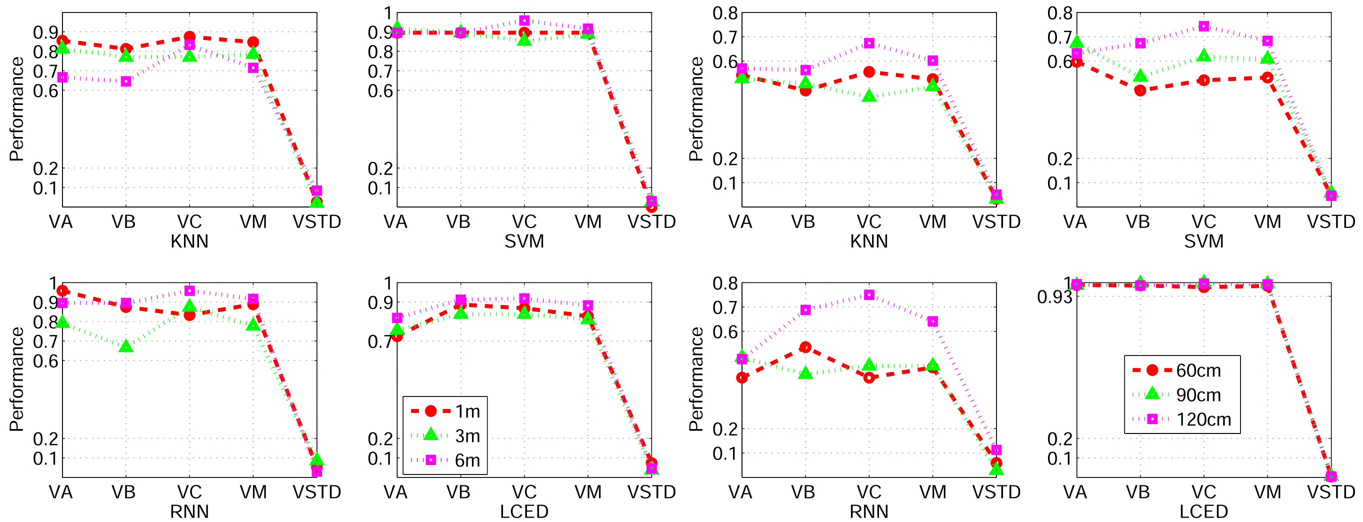
Existing works using machine learning algorithms usually achieve excellent performance in human activity recognition but lose effective in practical life due to the narrow bandwidth, insta-



(a) Performance analysis of algorithms

(b) Performance analysis of individuals

Fig. 15. Performance of antennas and individuals. (a) *VAM* is the average accuracy of all *antenna_As*; *VASTD* is the standard deviation of all *antenna_As*. (b) *VA* is the accuracy of a *antenna_A*; *VM* is the average accuracy of three antennas; *VSTD* is the standard deviation of three antennas.



(a) Effect of different distances on human activity recognition. (b) Effect of different heights on human activity recogniton.

Fig. 16. Performance analysis of distances and heights. *VA* represent the *antenna_A*; *VB* represent the *antenna_B*; *VC* represent the *antenna_C*; *VM* represent the average accuracy of all the antennas; *VSTD* represent the standard deviation of all the antennas.

bility, limited propagation distance and multipath effect. The above factors [19,17] hinder the development of WiFi-based application in practical applications. Therefore, the general scheme will play an important role between the experimental environments and practical applications. The two key issues need to be discussed as follows.

- **The principles of signal propagation in different indoor environments**. Currently, the researches mainly focus on the data-driven way to explore and quantify the rough relationship between signal patterns and human activities. As we all know, the way with high performance does not keep stability in another environment. The following work about the thought will be an important research topic.
- **Environment and activity-self**. We cannot neglect the impact of the environment on the reflected signal corresponding to human activity. The method to quantify the impact of the environment is also an interesting research topic for smart-environment sensing domain.

7. Conclusion

In this paper, we propose a novel deep learning model named LCED to explore the accuracy differences among individuals on human activity recognition using WiFi signals. To be specific, we design a series of the pre-processing methods to obtain high-quality activity data and use KNN, SVM, RNN as comparative objects to analyze the performance of LCED model on human activity recognition. The results show that the comparative objects achieve the accuracy of 75%–85%, and LCED model achieves the average accuracy is high 95% with less than cost. In the following research, we will focus on the accuracy difference of indoor environments on human activity recognition and attempt to design a deep mutual learning model to improve the stability of activity recognition under different indoor environments.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The work was supported by “Dalian Science and Technology Innovation Fund” with No. 2019J11CY004, by “National Key Research and Development Program of China” with 2017YFC0821003-2, by “the Shenzhen Basic Research Grants” with No. JCYJ20170413152804728 and JCYJ20180507182508857, by “National Natural Science Foundation of China” with No. 61842601 and 61902052, by “Fundamental Research Funds for the Central Universities” with No. DUT19RC(3)003.

References

[1] H. Abdelnasser, M. Youssef, K.A. Harras, WiGest: a ubiquitous WiFi-based gesture recognition system, Proc. of IEEE INFOCOM. (2015).
[2] F. Adib, H. Mao, Z. Kebelac, D. Katabi, C. Miller, Smart Homes That Monitor Breathing and Heart Rate, in: Proc. of ACM CHI, 2015.
[3] L. Chen, X. Chen, S.I. Lee, K. Chen, D. Han, D. Fang, Z. Tang, Z. Wang, WIDESEE: Towards Wide-Area Contactless Wireless Sensing, in: Proc. of ACM SenSys, 2019.
[4] Z. Chen, L. Zhang, C. Jiang, W. Cui, WiFi CSI based passive human activity recognition using attention based BLSTM, IEEE Trans. Mob. Comput. (2018).
[5] H. Chen-Yu, L. Yuchen, K. Zach, H. Rumen, K. Dina, L. Christine, Extracting Gait Velocity and Stride Length from Surrounding Radio Signals, in: Proc. of ACM CHI, 2017.
[6] B. Fang, N.D. Lane, M. Zhang, A. Boran, F. Kawsar, BodyScan: Enabling Radio-based Sensing on Wearable Devices for Contactless Activity and Vital Sign Monitoring, in: Proc. of IEEE MobiSys, 2016.
[7] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, S. Bing, T. Liu, X. Wang, W. Gang, Recent Advances in Convolutional Neural Networks, Computer Science, 2015.
[8] L. Guo, H. Zhang, C. Wang, W. Guo, L. Wang, C. Lin, Towards Diversity Activity Recognition Via LSTM-CNN Encoder-Decoder Neural Network, Proc. of ACM IJCAI Workshop (2019) 1–8.
[9] D. Halperin, W. Hu, A. Sheth, D. Wetherall, Linux 802.11n csi tool, 2010.http://dhalperi.github.io/linux-80211n-csitool/.
[10] D. Halperin, W. Hu, A. Sheth, D. Wetherall, Tool Release: Gathering 802.11n Traces with Channel State Information, Proc. of ACM SIGCOMM CCR 41 (2011).
[11] C. Han, K. Wu, Y. Wang, L.M. Ni, WiFall: Device-free fall detection by wireless networks, in: Proc. of IEEE INFOCOM, 2014.
[12] S. Ioffe, C. Szegedy, Batch normalization: accelerating deep network training by reducing internal covariate shift, in: International Conference on International Conference on Machine Learning, 2015.
[13] W. Jiang, C. Miao, S.Y. Fenglong Ma, Y. Wang, Y. Yuan, H. Xue, C. Song, D.K. Xin Ma, W. Xu, L. Su, Towards Environment Independent Device Free Human Activity Recognition, in: Proc. of ACM MobiCom, 2018, pp. 1–16.
[14] Linlin Guo, J.L.W.Z. Lei Wang, B. Lu, HuAc: Human Activity Recognition Using Crowdsourced WiFi Signals and Skeleton Data, in: Wireless Communications and Mobile Computing, 2018.
[15] Z.C. Lipton, J. Berkowitz, A Critical Review of Recurrent Neural Networks for Sequence Learning, 2015. CoRR abs/1506.00019. URL: http://arxiv.org/abs/1506.00019, arXiv:1506.00019.

[16] X. Liu, J. Cao, S. Tang, J. Wen, Wi-Sleep: Contactless Sleep Monitoring via WiFi Signals, in: Pro of IEEE RTSS, 2014.
[17] Y. Ma, G. Zhou, S. Wang, WiFi Sensing with Channel State Information: A Survey, ACM Comput. Surv., 2019. abs/1506.00019, 46:1–46:36. URL: https://doi.org/10.1145/3310194.
[18] P. Melgarejo, X. Zhang, P. Ramanathan, D. Chu, Leveraging Directional Antenna Capabilities for Fine-grained Gesture Recognition, in: Proc. of ACM UbiComp, 2014.
[19] K. Niu, F. Zhang, X.L. Jie Xiong, E. Yi, D. Zhang, Boosting fine-grained activity sensing by embracing wireless multipath effects, Proc. of ACM CoNEXT (2018) 1–13.
[20] C. Olah, Understanding LSTM Networks, 2015. URL: https://colah.github.io/posts/2015-08-Understanding-LSTMs/.
[21] Q. Pu, S. Gupta, S. Gollakota, S. Patel, Whole-home Gesture Recognition Using Wireless Signals, in: Proc. of ACM MobiCom, 2013.
[22] K. Qian, Z. Zhou, Y. Zheng, Z. Yang, Y. Liu, Inferring Motion Direction using Commodity WiFi for Interactive Exergames, in: Proc. of ACM CHI, 2017.
[23] S. Sigg, S. Shi, Y. Ji, Teach your WiFi device: recognize simultaneous activities and gestures from time-domain RF-features, Int. J. Ambient Comput. Intell. 6 (2014) 20–34.
[24] F. Wang, Y. Song, J. Zhang, J. Han, D. Huang, Temporal Unet: sample level human action recognition using WiFi, Proc. of arXiv (2019) 1–14.
[25] J. Wang, X. Zhang, Q. Gao, H. Yue, H. Wang, Device-free simultaneous wireless localization and activity recognition with wavelet feature, IEEE Trans. Veh. Technol. 66 (2017) 1659–1669.
[26] W. Wang, A.X. Liu, M. Shahzad, K. Ling, S. Lu, Understanding and Modeling of WiFi Signal Based Human Activity Recognition, in: Proc. of ACM MobiCom, 2015.
[27] X. Wang, X. Wang, S. Mao, RF Sensing in the internet of things: a general deep learning framework, IEEE Commun. Mag. (2018) 62–67.
[28] Y. Wang, J. Liu, Y. Chen, M. Gruteser, J. Yang, H. Liu, E-eyes: Device-free Location-oriented Activity Identification Using Fine-grained WiFi Signatures, in: Proc. of ACM Mobicom, 2014.
[29] W. Wei, X.Liu, A., M, S., Gait Recognition Using Wi-Fi Signals, in: Proc. of ACM UbiComp, 2016.
[30] Z. Yang, Z. Zhou, Y. Liu, From RSSI to CSI: Indoor Localization via Channel Response, ACM Comput. Surv. 46 (2013) 25:1–25:32.
[31] N. Yu, W. Wang, A.X. Liu, L. Kong, QGesture: Quantifying Gesture Distance and Direction with WiFi Signals, in: Proc. of ACM Interact. Mob. Wearable Ubiquitous Technol, 2018, pp. 1–23.
[32] Y. Zheng, Y. Zhang, K. Qian, G. Zhang, Y. Liu, C. Wu, Z. Yang, Zero-Effort Cross-Domain Gesture Recognition with Wi-Fi, in: Proc. of ACM Mobisys, 2019, pp. 1–13.
[33] Y. Zeng, D. Wu, R. Gao, T. Gu, D. Zhang, FullBreathe: Full Human Respiration Detection Exploiting Complementarity of CSI Phase and Amplitude of WiFi Signals, in: Proc. of ACM Interact. Mob. Wearable Ubiquitous Technol, 2018, pp. 148:1–148:19.
[34] Y. Zeng, D. Wu, J. Xiong, E. Yi, R. Gao, D. Zhang, FarSense: Pushing the Range Limit of WiFi-based Respiration Sensing with CSI Ratio of Two Antennas, in: Proc. of ACM UbiComp, 2019.
[35] F. Zhang, K. Niu, J. Xiong, B. Jin, T. Gu, Y. Jiang, D. Zhang, Towards a Diffraction-based Sensing Approach on Human Activity Recognition, in: Proc. of ACM Interact. Mob. Wearable Ubiquitous Technol, 2019, pp. 33–57.
[36] M. Zhao, T. Li, M.A. Alsheikh, Y. Tian, H. Zhao, A. Torralba, D. Katabi, Through-Wall Human Pose Estimation Using Radio Signals, in: Proc. of ACM CVPR, ACM, 2018, pp. 7356–7365.
[37] X. Zheng, J. Wang, L. Shangguan, Z. Zhou, Y. Liu, Smokey: Ubiquitous Smoking Detection with Commerical WiFi Infrastructures, in: Proc. of IEEE INFOCOM, 2016.
[38] H. Zou, J. Yang, Y. Zhou, L. Xie, C.J. Spanos, Robust WiFi-enabled Device-free Gesture Recognition via Unsupervised Adversarial Domain Adaption, in: Proc. of IEEE ICCCN, 2018, pp. 1–8.
[39] H. Zou, Y. Zhou, J. Yang, H. Jiang, L. Xie, C.J. Spanos, DeepSense: Device-Free Human Activity Recognition via Autoencoder Long-Term Recurrent Convolutional Network, IEEE International Conference on Communications (ICC) (2018) 1–6, https://doi.org/10.1109/ICC.2018.8422895.

Linlin Guo received the B.S. degree in computer and technology department from Taishan University in 2011, the M.S. degree from Qufu Normal University in 2014. She has been working towards the PhD degree in the School of Software technology at Dalian University of Technology. Her research interests include device-free wireless localization and human activity recognition, wireless network, and machine learning. Email: linlin.teresa.guo@gmail.com

Hang Zhang received the B.E. degree in computer science and engineering department from Changchun University of Technology in 2018. He has been working towards the M.E. degree in Computer Science at Shenzhen Institutes of Advanced Technology, University of Chinese Academy of Sciences. His research interests include digital signal processing, deep learning, machine learning, natural language processing. Email: hang.zhang1@siat.ac.cn

Chao Wang, a postgraduate majoring in Biomedical Engineering from Northeastern University, is visiting Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences. Her main research is the application of Deep Learning in human–computer interaction, including motion recognition and continuous motion estimation. Email: chao.wang@siat.ac.cn

Weiyu Guo received the B.S. degree in School of Software technology at Dalian University of Technology in 2019. He has been working towards the M.S. degree in the School of Artificial Intelligence at University of Chinese Academy of Sciences. His research interests include artificial intelligence and human activity recognition, neuroengineering and deep learning. Email: guoweiyu96@gmail.com

Guangqiang Diao received the B.S. degree in computer and technology department from Taishan University in 2010, the M.S. degree from Zhejiang Sci-Tech University in 2014. He is engaged in teaching at Shandong youth university of political science. His research interests include machine learning and artificial intelligence. Email: dgq@sdyu.edu.cn

Bingxian Lu is currently an assistant professor in the School of Software, Dalian University of Technology, China. He received his B.S., M.E., and PhD from Dalian University of Technology, in 2012, 2014, and 2019, respectively. His research interests include the wireless network, mobile computing, and pervasive computing applications. He is a member of the IEEE and ACM.

Chuang Lin received the M.Sc. and the PhD degrees in signal processing from the Harbin Institute of Technology, Harbin, China, in 2004 and 2008, respectively. He was a Research Scientist of European Research Council Project DEMOVE in the Department of Neurorehabilitation Engineering, University Medical Center Göttingen, Georg-August University, Germany, within the Bernstein Center for Computational Neuroscience and the Bernstein Focus Neurotechnology Göttingen. Now he is an Associate Professor of Chinese Academy of Sciences, Shenzhen Institutes of Advanced Technology, Shenzhen, China. He presided and participated many grants, including the National Natural Science Foundation of China under grants, "Hundred Talents Program" of the Chinese Academy of Sciences, and the Shenzhen Governmental Basic Research Grant, etc. He has (co)-authored more than 20 papers in peer-reviewed journals and over 30 conference papers/abstracts, and over 10 patents. His research interests include biomedical signal processing, neural rehabilitation engineering, pattern recognition, and machine learning.

Lei Wang is currently a full professor of the School of Software, Dalian University of Technology, China. He received his B.S., M.S. and Ph.D. from Tianjin University, China, in 1995, 1998, and 2001, respectively. He was a Member of Technical Staff with Bell Labs Research China, a senior researcher with Samsung, South Korea, a research scientist with Seoul National University, and a research associate with Washington State University, Vancouver, WA, USA. His research interests involve wireless ad hoc network, sensor network, social network and network security. He has published 180 + papers and the papers have 2000 + citations. Email: lei.wang@dlut.edu.cn