

Personnel Identification System Using the Wireless Channel Characteristics and Machine Learning

Yong Li^{1,2,*}, Jie Zhao^{1,2}

¹ College of Engineering and Technology, Chengdu university of technology, Leshan, China

² South Western Institute of Physics, Chengdu, China

*Corresponding Author.

Abstract

In the presence of a Wi-Fi network, movements of individuals lead to changes in the channel state information (CSI) of the Wi-Fi network. Thus, it is possible to use CSI to identify specific individuals based on walking gait and under movements. We have built a hardware system to collect CSI and developed CSI real-time acquisition and visualization software. We further improve results by preprocessing the CSI to remove noise by outlier removal, a low-pass filter, and a discrete wavelet transform. For identity recognition, we propose a parallel learning structure using a convolutional neural network (CNN) and bidirectional long-term and short-term memory network to extract amplitude and time sequence features of human gait at the same time. We then classify the two gait features using Softmax function to identify individuals. Our identity recognition experiment with 30 people yielded a maximum accuracy of 98.7% with our device and software. The system has a value of application in the field of smart home and intrusion detection.

Keywords: CSI, specific individuals, parallel learning structure, CNN, human gait

I. Introduction

Personnel identity recognition is one of the most important technologies in security monitoring systems. Recognition can be accomplished using computer vision [1], but doing so requires sufficient light. It can also be done using biometrics [2], but this approach is easy to forge. It can be done using wearable sensors [3], but this method requires the person to be identified to wear devices, which causes some inconvenience. Micro-Doppler feature identification is also feasible, but the cost of the equipment is high [4]. In the early days of Wi-Fi, the received signal strength indication (RSSI) of the wireless signals was used for indoor positioning [5] and activity recognition [6], but research [7] confirmed that such methods are susceptible to multipath effects, resulting in the poor reliability, low accuracy, and other shortcomings.

In 2011, researchers at the University of Washington developed the CSI-tools package [8], which makes it very easy to obtain channel state information (CSI) using a wireless network card. Recent studies have shown that CSI correlates with movements of a person active within the coverage of the wireless [9, 10]. This correlation enables detection of specific types of activities, including recognizing gestures [11], actions [12], identity [13], sleep patterns [14], and dangerous driving [15]. These studies use the channel state information extracted from Wi-Fi signals to perform these tasks without sensors or contact. When used for personal identification, this method offers user privacy, easy deployment, and low cost.

There are many ways to employ CSI for personal identity recognition. The first to use CSI to identify people is the WiWho model [16], which uses a peak and valley detection algorithm to extract the waveform of each walking step, and uses a decision tree classifier. The FreeSense model [17] uses the k-nearest neighbor (KNN) algorithm to classify and recognize the discrete wavelet features of CSI. The WiID model [18] first uses a Hidden Markov

Model (HMM) to classify DWT features followed by a support vector machine (SVM) for final classification. The Wii model [19] extracts the time-domain and frequency-domain features of a person’s gait and also uses a support vector machine classifier. Wi-Fi-ID model [20] uses the continuous wavelet transform (CWT) algorithm to extract signals of different frequency bands in CSI and uses a classifier based on a sparse approximation classification algorithm (Soft Actor-critic, SAC). The identity recognition model using the Empirical Mode Decomposition (EMD) framework [21] combines time-domain features and frequency-domain features to construct hybrid features and uses the Softmax function as a classifier. The NeuralWave model [22] integrates the amplitude and phase of CSI and uses convolutional neural networks (CNNs) to classify eigenvalues. The CSI-Net model [23] uses transposed convolution networks (TCNs) to perform feature mapping on the CSI and deep neural networks (DNNs) to extract and classify features. The characteristics of various CSI-based identification methods are shown in Table 1.

Table 1 Various personal identification methods using CSI

References	Method	Identifiable number of people	Accuracy
[16]	Decision Tree	2–6	92%–80%
[17]	KNN	2–6	94.5%–88.9%
[18]	HMM, SVM	10	90%
[19]	SVM	2–5	98.7%–91.5%
[20]	SAC	2–6	93%–77%
[21]	Softmax	10	97.5%
[22]	CNN	24	87.7%
[23]	DNN	30	93%

Existing CSI-based identification methods also have some shortcomings. For example, methods based on simple classifiers (SVM, KNN, etc.) require artificial feature selection of samples that do not precisely represent the biological features implicit in the radio frequency signal, limiting the number of people identifiable by these methods to 10 or fewer. Although deep learning enables automatic extraction of the CSI features, those approaches are limited to the extraction of spatial features and do not consider time features. In response to these shortcomings, we combine the amplitude characteristics of CSI with time sequence characteristics and propose an identity recognition model and working device that integrate CNN and BiLSTM. Our main contributions are as follows:

According to the rate of change of the influence of human activities on the channel, we propose a combination of abnormal point removal, low-pass filter, and discrete wavelet transform to remove noise in CSI;

We propose a parallel feature extraction and classification model constructed with CNN and BiLSTM that extracts human gait features from CSI at the same time. The amplitude feature is obtained from the CNN, and the time sequence feature is extracted from the BiLSTM;

We create a Wi-Fi CSI acquisition and processing device using an industrial computer and Wi-Fi network card to identify persons indoors using our CSI training model.

The rest of this paper is organized as follows. In section 2, we analyze and discuss the identification model including the CSI preprocessing method and the feature recognition method using CNN and BiLSTM. In Section

3, we present our hardware and software platform. The experimental results and their analysis and discussion are summarized in Section 4. Finally, we present our conclusions in Section 5.

II. System Model

The estimation of CSI is a prerequisite for demodulating orthogonal frequency-division multiplexing (OFDM) signals. We denote the number of antennas in the transmitter as N_t , the number of antennas in the receiver as N_r , and the number of sub-carriers between antenna links as K . Every time a signal is received, a 3-dimensional CSI matrix of size $N_r \times N_t \times K$ is obtained that describes the changing characteristics of the wireless signal in the channel transmission process. CSI includes the amplitude and phase characteristics of the wireless signal of each antenna pair at the time of transmission. These changes are caused by scattering, fading, power attenuation, shadowing, and other phenomena of channel propagation, which can be expressed as

$$y = Hx + n \tag{1}$$

where y is the received signal, x is the sent signal, H is the channel matrix, and n is the noise in the normal distribution of $N(0,S)$. The channel matrix is then estimated as

$$\hat{H} = y/x \tag{2}$$

\hat{H} is CSI, which can be expressed as

$$\hat{H}(f_k) = \begin{pmatrix} h_{11} & \dots & h_{1N_t} \\ \vdots & \ddots & \vdots \\ h_{N_r1} & \dots & h_{N_rN_t} \end{pmatrix} \tag{3}$$

$$= \|\hat{H}(f_k)\| e^{j\angle \hat{H}(f_k)}$$

where $h_{N_rN_t}$ represents the channel state information between the antenna pair composed of the N_t -th transmitting antenna and the N_r -th receiving antenna, $\hat{H}(f_k)$ represents the channel characteristics of the subcarrier with frequency f_k , $\|\hat{H}(f_k)\|$ represents the amplitude, $e^{j\angle \hat{H}(f_k)}$ represents the phase, and $k \in (1, 30)$ represents the index of the subcarrier.

When there are people moving indoors, the phase and amplitude of the corresponding CSI will change accordingly. Thus, the CSI of the Wi-Fi signal makes it possible to identify the currently active person based on the changing CSI characteristics. Figure 1 presents an overview of the identification method using CSI. The working principle of each part of the figure will be discussed in detail later.

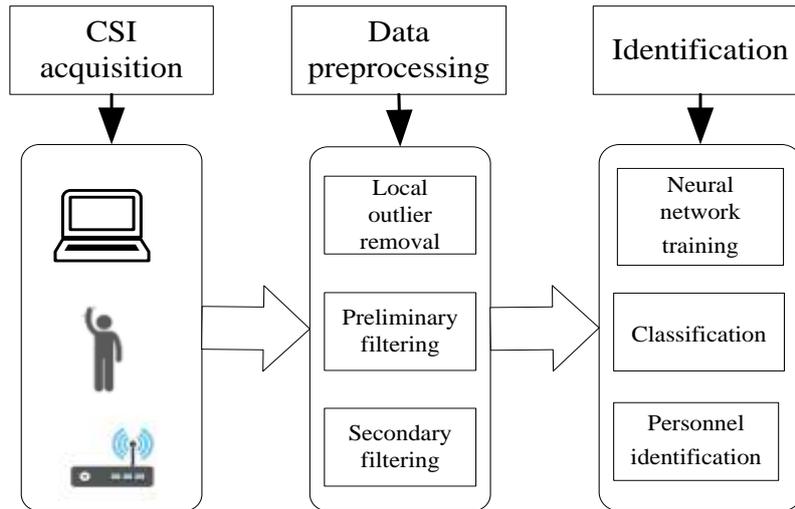


Fig 1: Overview of person identification using CSI

2.1 Data preprocessing

2.1.1 Removal of abnormal points

The original CSI includes interference noise caused by multipath effects. These outliers are not caused by human activities and hinder the extraction of behavior features. Therefore, we use the Hampel algorithm to deal with the outliers in the original CSI data, a low-pass filter to remove the high-frequency components caused by multipath effects, and a wavelet transform to filter out the local features of the subcarrier changes in the OFDM signal.

The Hampel-based abnormal point removal method uses the median (Median, MED) of the data set and the mean absolute deviation (MAD) to discriminate and remove abnormal data [24]. First, for the data set with window width defined in equation (4), the median value z_i is calculated by C_i :

$$C_i = \{c_j | j \in [i, i + w]\}. \quad (4)$$

We then construct a data sequence K_i as

$$K_i = \{k_j | k_j = |c_j - z_i|, c_j \in C_i\}. \quad (5)$$

If $Q = 1.4826K_i$ and $q = |c_m - k_i|$, when $q \geq Q*w$, the current data value is replaced with z_i .

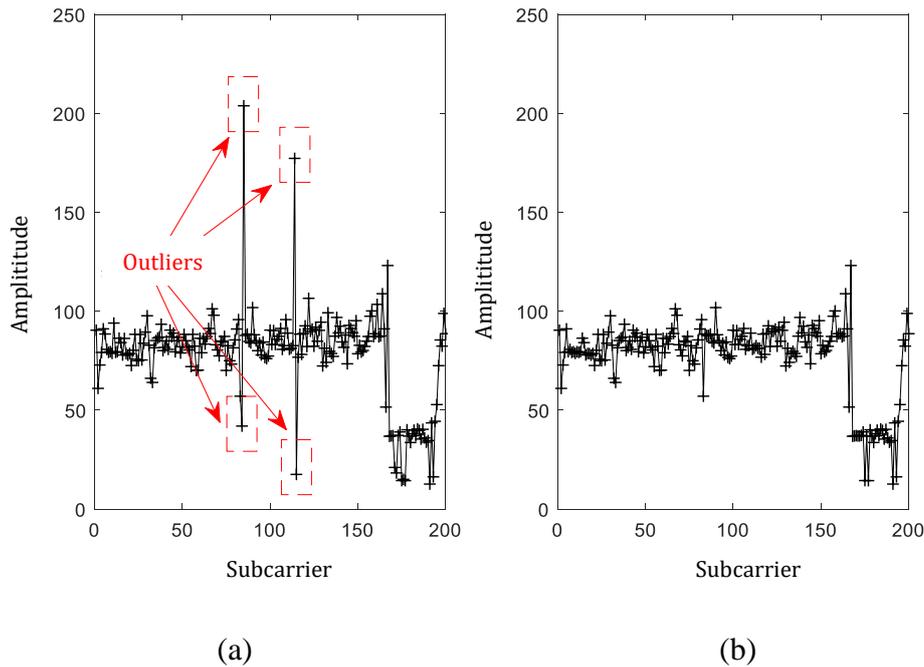


Fig 2: Example of local outlier removal from the original CSI

If the amplitude of a sub-carrier in CSI is not in the effective range $[\mu - \lambda\sigma, \mu + \lambda\sigma]$, it will be regarded as an outlier and eliminated. For the original CSI to be processed, we first calculate the median value of the data in the window and calculate the absolute deviation between each data point and the median value. We then set the threshold according to the calculation results and determine whether the point is an abnormal point. If the data is not in the effective range, it is an abnormal point replaced with the median value. μ in the effective interval is the median value of the original CSI, while λ represents the outlier discrimination parameter (generally equal to 3), which is the absolute median deviation of the subcarrier amplitude. The results of removing local outliers in the original CSI are shown in Figure 2 (b); four outliers have been removed.

2.1.2 Preliminary filtering

In addition to expressing the low-frequency components of human behavior, CSI also contains high-frequency components caused by multipath propagation. Therefore, a low-pass filter is required to perform preliminary filtering. Compared with other filters, the Butterworth filter has the flattest frequency response in the passband with a stopband that eventually tends to zero. Therefore, the Butterworth filter not only ensures the accuracy of the original frequency range signal but also filters out the out-of-band noise. The square expression of its impulse response function is

$$|H(\omega)|^2 = \frac{1}{1 + (\omega/\omega_c)^{2n}} \quad (6)$$

where $H(\omega)$ represents the amplitude of the signal, ω_c represents the cut-off frequency, and n represents the order of the filter. Considering the computational complexity and filtering effect, the order of the filter is 3. As shown in Figure 3(b), the amplitude curves of 3 subcarriers (s_1, s_2, s_3) of CSI have obvious amplitude changes before being filtered. After filtering, their amplitude curves slowed significantly. After filtering the subcarriers of the CSI, the amplitude curves change with time slow noticeably.

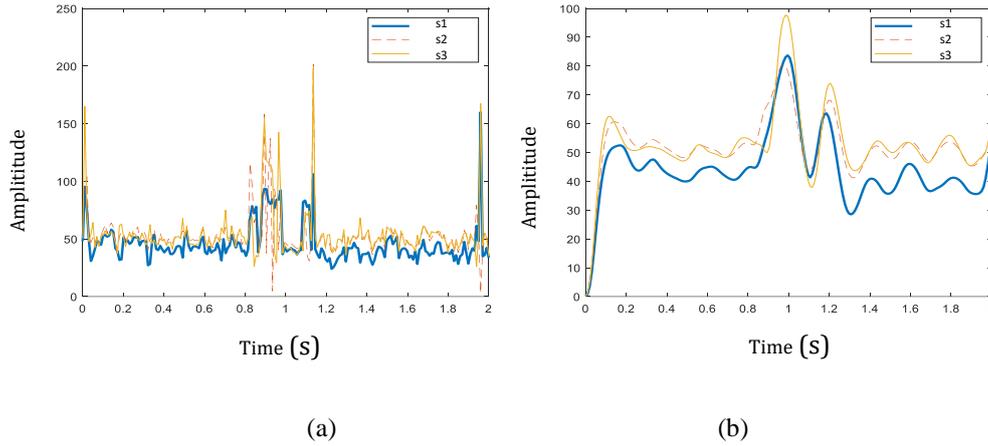


Fig 3: CSI amplitude curves before and after low pass filtering

2.1.3 Secondary filtering

Neural networks learning human behavior features directly from noisy data produce certain deviations that affect recognition accuracy. Wavelet transforms support multiple resolutions good denoising effects with low signal-to-noise ratios (SNRs). In order to more clearly show the local characteristics of human behavior corresponding to the subcarrier changes, we use discrete wavelets for the secondary processing of CSI signals.

Assuming that the CSI sequence is $S(t)$, the wavelet decomposition is defined as

$$S(t) = A_n + D_n + D_{n-1} + \dots + D_1 \tag{7}$$

where A is the low-frequency approximation component, n is the number of layers of wavelet decomposition, and D is the high-frequency detail component. The discrete wavelet transform method implemented by an inner product or convolution is complex and time-consuming to calculate, but Mallat’s algorithm effectively solves this problem [25].

After wavelet transformation, the original CSI can be reconstructed through the set threshold to obtain the CSI with high frequency noise removed. The processing process is shown in Figure 4. The process first decomposes the signal using the selected wavelet function and then determines the threshold, with the decomposed detail coefficients solved by the soft thresholding function. Finally, the processed coefficients are reconstructed by the wavelet transform to obtain the denoised signal. The processing results depend on the wavelet basis and order, the number of decomposition levels, and the chosen detail coefficient.

In order to not only remove high-frequency noise but also ensure that there is enough detailed information to discriminate between gestures, we decompose and reconstruct the CSI using a 4-layer Symlets wavelet. Within the scope value and function mask of the discrete wavelet transform, F_δ and M denote the threshold operator and mask operator of wavelet analysis, respectively; $c_{m,n}$ denotes the decomposition coefficient; δ is the threshold; and the function rule of the threshold operator is as in equation (8). As in equation (9), M preserves the coefficients that meet certain conditions unchanged and sets the ones that are not satisfied to 0.

$$(F_\delta c)_{m,n} = \begin{cases} c_{m,n}, & |c_{m,n}| > \delta \\ 0, & \text{else} \end{cases} \tag{8}$$

$$(Mc)_{m,n} = \begin{cases} c_{m,n}, & |c_{m,n}| \in Q \\ 0, & \text{else} \end{cases} \tag{9}$$

In order to reduce the error and obtain more accurate frequency domain information, we employ the Symlets wavelet function with approximate symmetry and determine the decomposition level N according to the situation. Larger values of N enable stronger local processing ability in the frequency domain at the expense of orthogonality.

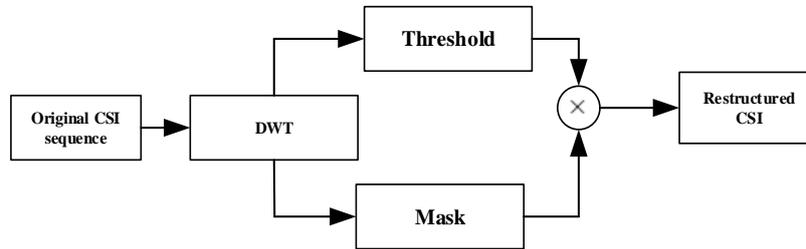


Fig 4: Discrete wavelet transform denoising process

For the wavelet denoising of the CSI, the penalty strategy model is used to calculate the threshold [26]:

$$f(t) = - \sum_{k \leq t} c_k^2 + 2\sigma^2 t (\alpha + \log(\frac{n}{t})) \tag{10}$$

Where c_k^2 represents the k-th largest value among the wavelet decomposition coefficients, n is the total number of wavelet decomposition coefficients, and α is a real number greater than 1 that is generally $\alpha = 2$. The sparseness of the wavelet coefficients after noise reduction is determined by α , with larger values increasing the sparsity and smoothing the reconstructed signal waveform. According to this method, the final threshold is determined to be $\text{threshold} = |t_1|$, where t_1 is the variable that yields the smallest value.

2.2 Personnel identification model

As shown in Figure 5, the personnel identification model composed of deep CNN [27] and BiLSTM networks [28] extracts human gait features from the CSI to perform identity recognition. The deep CNN extracts the gait amplitude features, and the BiLSTM network extracts the gait time features. Finally, the softmax function is used to classify these two features to identify the person.

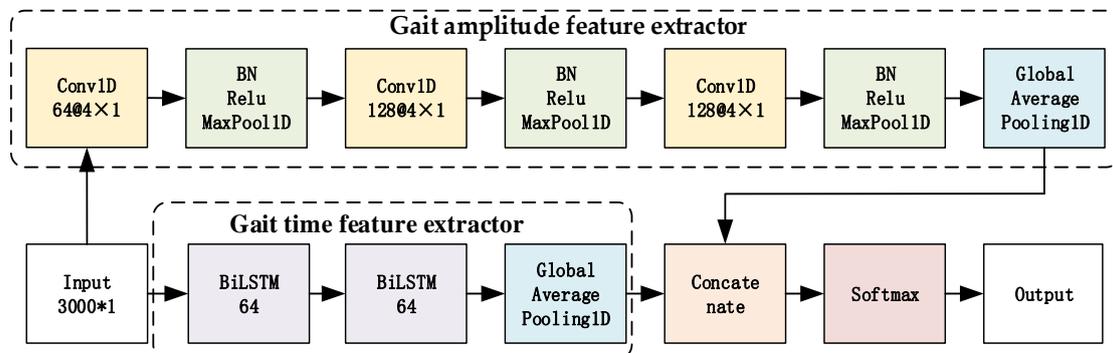


Fig 5: Personnel identification model

2.2.1 Gait amplitude feature extractor

According to our experiments, increasing the number of network layers beyond 3 only slightly improved the recognition accuracy. However, more layers require more hyperparameters to be trained, thereby increasing the time cost and space occupation of the model. Therefore, we use 3 CNN layers. In order to reduce the hyperparameters that the model needs to train, we found it necessary to add a maximum pooling layer (Maxpooling) and a loss layer (Dropout) after each layer of the CNN. Therefore, the structure of the CNN, as shown in Figure 6, consists of 3 convolutional neural network layers (C1, C2, C3), 3 batch normalized network layers (B1, B2, B3), 3 maximum pooling network layers (M1, M2, M3), 3 Dropout layers (D1, D2, D3), and 1 global average pooling layer G1.

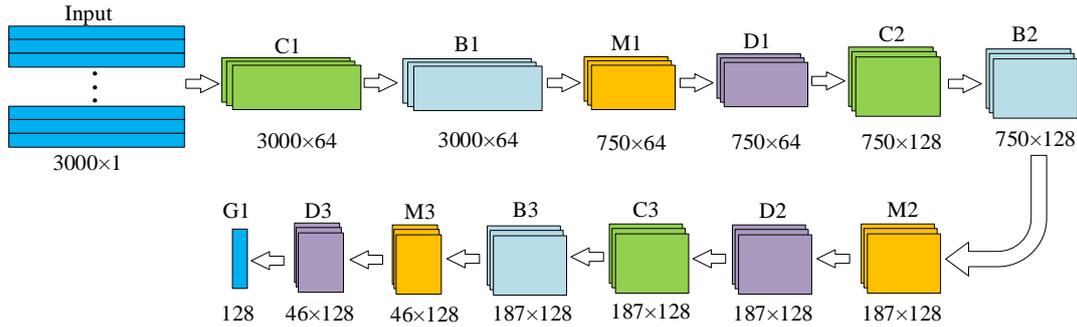


Fig 6: The deep CNN structure used for gait amplitude feature extraction

In the deep CNN, the current convolution layer first uses the convolution kernel and the feature map output by the previous layer to perform the convolution operation, then applies the activation function to the result of the convolution operation, and finally outputs the activated feature map. The output feature map can be expressed as a combination of the convolution kernel and the input feature map:

$$x_j^l = f \left(\sum_{i \in M_j} x_i^{l-1} * k_{ij}^l + b_j^l \right) \quad (11)$$

where M_j represents the selection of the input mapping, x_i^{l-1} represents the previous feature map, b_j^l represents the deviation, and k_{ij}^l represents the convolution kernel. The feature map obtained after the convolution operation represents the gait amplitude characteristics of the person in the CSI. However, when the extractor extracts feature maps from the CSI, each set of convolution kernels generates new feature maps of the same size as the previous layer's feature maps. This may increase the number of channels of the feature maps, which introduces more hyperparameters training in turn.

To alleviate this problem, we add a maximum pooling layer (Maxpool) and a loss layer (Dropout) after each convolutional layer. The maximum pooling layer extracts the largest element in each pooling window to represent the characteristics of this window, and the loss layer randomly discards the connections between some neurons, thus improving the convergence speed of the model and increasing the model's anti-overfitting ability. After completing the convolution, the output of the model is passed to the batch normalization layer (Batch Normalization, BN). The BN layer between the convolutional layer and the activation function normalizes the data by calculating the mean and variance of the training data and likewise normalizes the distribution of neurons in the hidden layer, which helps the activation function learn the data distribution. In this way, we avoid the problem of gradient disappearance in the training process of the model and accelerate the training speed.

A single CNN layer can only extract linear features in CSI samples, so we enable the model to extract non-linear features by overlapping multiple CNN and BN layers and add ReLu activation function after each BN layer. The ReLu activation function introduces nonlinear feature extraction capabilities to the model. Considering the large number of training samples in the experiment, the model compresses the output feature map of the CNN into one-dimensional data through the global average pooling layer.

Table 2 Parameter settings of the deep CNN

Layer order	Number of neurons	Size of kernel	Pooling layer	Accuracy
1	64	4*1	4	92%–80%
2	128	4*1	4	94.5%–88.9%
3	128	4*1	4	90%

The parameters of the structure shown in Figure 6 are given in Table 2. The parameter number of the pooling layers is set to 4, and the parameter of the loss layer is set to 0.6. The CSI data input to the network is one-dimensional data with the dimensions of (3000, 1); the C1 layer extracts a feature map with the dimensions of (3000, 64) from it.

The characteristic map is normalized by the B1 layer as a normal distribution, but the pattern remains unchanged. The feature map of the normal distribution is continuously compressed by the M1 layer and has the dimensions of (750,64) when complete. After the D1 layer randomly discards the connections of 60% of the neurons in the hidden layer, the C2 layer extracts features from the feature map processed by the D1 layer to produce an output feature map with dimensions (750,128).

After the feature map is standardized to a normal distribution by the B2 layer, it is compressed by the M2 layer to obtain a feature map with a dimension of (187,128). The D2 layer randomly discards the connections of 60% of the neurons in the hidden layer. Then, the process extracts the features of the feature map processed by the D2 layer via the C3 layer, to produce a feature map of dimensions (187,128). The feature map continues to be normalized by the B3 layer as a normal distribution, and is then compressed by the M3 layer to produce a map with dimensions (46,128). After 60% of the neurons in the hidden layer are randomly discarded by the D3 layer, the G1 layer performs the global pooling operation on the feature map output by the D3 layer to obtain 128 gait amplitude features, expressed as

$$F = [F_1, \dots, F_i, \dots, F_{128}] \tag{12}$$

where F represents the output vector of the deep CNN, and F_i represents the i-th gait amplitude feature.

2.2.2 Gait time feature extractor

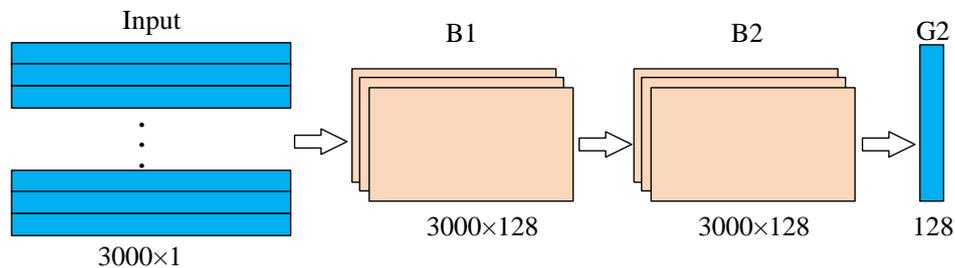


Fig 7: The structure of the deep CNN used for gait amplitude feature extraction

Taking the time sequence of the appearance of different gait amplitudes as biometrics represents the behavioral differences of different people in more detail. Therefore, a deep learning network with time memory is used to

extract the gait time features in the CSI. Because future information is also very important when learning gait characteristics, we use a BiLSTM network to extract gait time characteristics. Our experiments show that increasing number of layers in the BiLSTM network beyond 2 improves the recognition accuracy by less than 1%, even though it significantly increases the training time required. Therefore, the BiLSTM network consists of two bidirectional long- and short-term memory network layers (B1, B2) and an average global pooling layer G2 as shown in Figure 7.

Since the BiLSTM network in Figure 7 includes a forward layer and a backward layer, the network can process the CSI time series from two directions for automatic feature extraction and sequential information coding. According to the contribution of the extracted features to the recognition results, the features are forgotten or retained. Because the two-way operation considers the influence of past and future information on the current hidden state at the same time, a richer time feature is obtained. Because a single-layer BiLSTM network extracts insufficient features, we use a 2-layer BiLSTM network to extract temporal gait features from CSI data. Finally, the extracted feature maps are compressed through the average global pooling layer to reduce the hyperparameters that the model needs.

Both the B1 and B2 layers in Figure 7 contain 64 neurons; the network parameters are shown in Table 3. The dimension of the input data of the CSI sample is (3000, 1). B1 is first used to extract the time feature of the input sample and obtains a feature map with a dimension of (3000,128). Then the B2 layer extracts the features output by the B1 layer, with the resulting feature map dimension of (3000, 128). Finally, the G2 layer pools the output of the B2 layer, so that the output feature map is compressed to obtain 128 temporal features of the gait, expressed as

$$B = [B_1, \dots, B_i, \dots, B_{128}] \tag{13}$$

where B represents the output vector of the gait time feature extractor, and B_i represents the i -th gait time feature information.

2.2.3 fusion and recognition to feature

As shown in Figure 8, the gait amplitude feature F and the time feature B are jointly expressed by the concatenation layer as a fusion gait feature as in equation (14).

$$E = [F_1, \dots, F_i, \dots, F_{128}, B_1, \dots, B_i, \dots, B_{128}] \tag{14}$$

The fusion and classification recognition process of these two features is shown in Figure 8. The Concatenate layer combines the gait amplitude feature and the gait time feature to obtain 256 gait features, and finally classifies the gait features through the Softmax function to identify the person.

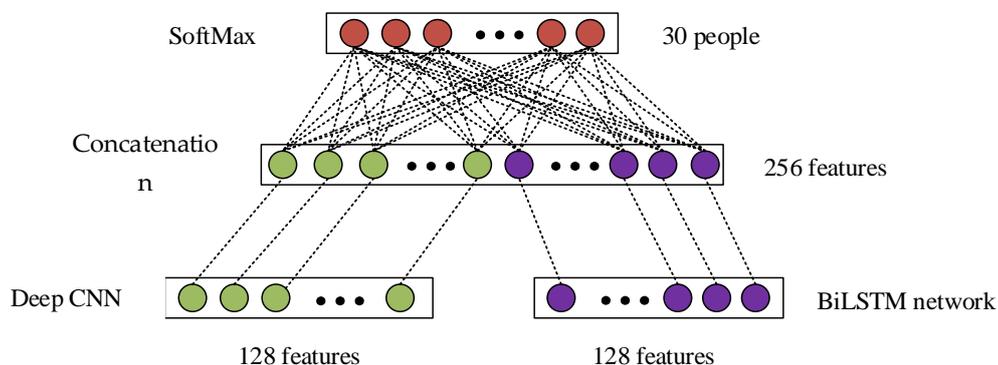


Fig 8: The fusion of gait amplitude and time features

The identity recognition model classifies gait features through the Softmax function. The output layer generates a score vector \vec{Z} by learning the feature mapping of the Concatenate layer, and each element in \vec{Z} represents the score of a specific class. The Softmax function uses exponentiation and normalization to convert the score vector \vec{Z} into a probability vector y_i :

$$y_i = [y_{i1}, y_{i2}, \dots, y_{id}] \quad (15)$$

where y_{id} represents the probability that the i -th CSI sample is predicted to be a class d :

$$y_{id} = P(d|i) = \frac{\exp^{z(d)}}{\sum_{j=1}^D \exp^{z(j)}} \quad (16)$$

Then the vector value with the largest probability value d_{\max} is the predicted class of the i -th CSI sample:

$$d_{\max} = \arg \max_{d \in D} P(d|i) \quad (17)$$

After training the identity recognition model with CSI, the classification accuracy is maximized by minimizing the cross-entropy loss function:

$$L(t, y) = - \sum_{i=1}^N \sum_{d \leq D} t_{id} \log(y_{id}) \quad (18)$$

If and only if the i -th CSI sample belongs to the class d , the value of t_{id} is 1. N indicates the number of samples that need to be trained in each batch of the model.

III. Completion of Personnel Identification System

The hardware for the identity recognition system is relatively simple, as shown in Figure 9. It consists of an Intel 5300 wireless network card connected via the PCMCIA slot of an embedded 32-bit industrial control computer. The network card includes three transceiver antennas that can receive signals through MIMO. The computer has 4 GB of RAM and 128 GB of solid-state mass storage and runs the Linux operating system.



Fig 9: The network identification hardware

The functions implemented in software include both real-time collection of the CSI and our identification software. Although CSI tools can be used to collect CSI, the original CSI gathering software needs the support of Matlab,

which does not run on this computer. For this reason, we have written software to gather CSI data in Python to remove the dependency on Matlab.

Since CSI tools save the collected CSI in a file named “csi.dat”, we have analyzed the data format of the file. The file consists of multiple binary units called Bfee, each of which contains a field called “Field”. Each Field has two parts: a header and a payload. The structure is shown in Figure 10. The header contains the values of the number of transmitting antennas, the number of receiving antennas, RSSI, and noise, and the payload contains CSI data, including a complex matrix composed of data from 30 sub-carriers.

Timestamp_low (32bit)			
Bfee_count (16bit)		Unkonwn (16bit)	
Nrx (8bit)	Ntx (8bit)	Rssi_a (8bit)	Rssi_b (8bit)
Rssi_c (8bit)	Noise (8bit)	Agc (8bit)	Antenna_set (8bit)
Len (16bit)		Fake_rate_n_flags (16bit)	
Payload (LEN × 8 bit)			

Fig 10: The content of each Bfee Field

We defined a Bfee class method in Python language to extract the contents from csi.dat. The Bfee data container is a dictionary type that stores the contents of each Bfee unit in csi.dat in order. The Code index in the Bfee unit is converted and judged in decimal. If the Code value corresponding to the index bit is 187, it means that the Field corresponding to the index bit contains data such as RSSI and CSI. In that case, the Nrx, Ntx, RSSI, Payload, and other data are extracted and stored, including the CSI complex matrix stored from the payload’s fourth bit onward. The program flow chart for reading the CSI is shown in Figure 11.

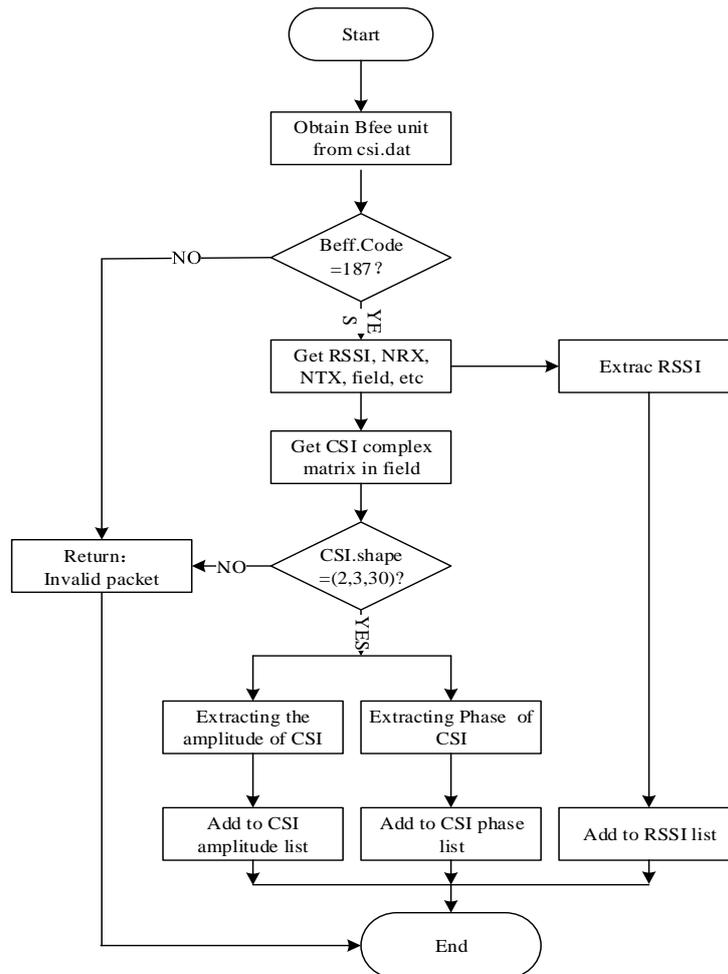


Fig 11: The flow of the Python CSI decoding program

Like the decoding software, we used Python for the identification portion as well, using the tensorflow package to implement the signal preprocessing, identification model, and training discussed in Section 2. We trained the identity recognition model on a workstation Dell Precision T5820 using its graphics operation card NVIDIA Tesla T4. After training, we transferred the needed model parameters to the identification device for indoor human identification test.

IV. Experiments and Results

The experimental environment is shown in Figure 12. The distance between the transmitting antenna and the receiving antenna was 2 meters. The CSI sampling frequency was 1000 Hz, and the sampling duration was 10 s, so the number of samples of a single sub-carrier CSI per acquisition was 10,000. As shown in Figure 12(b), each participant in the experiment walked the same path from A to B.

4.1 CSI changes caused by human activities

Figure 13 shows the varying amplitude of a single sub-carrier CSI for six different participants walking during the experiment. The changing patterns were also different, caused by differences in arm swings, body size, or gait

particulars. To determine whether this relationship exists, we verified the results using our identity recognition model.

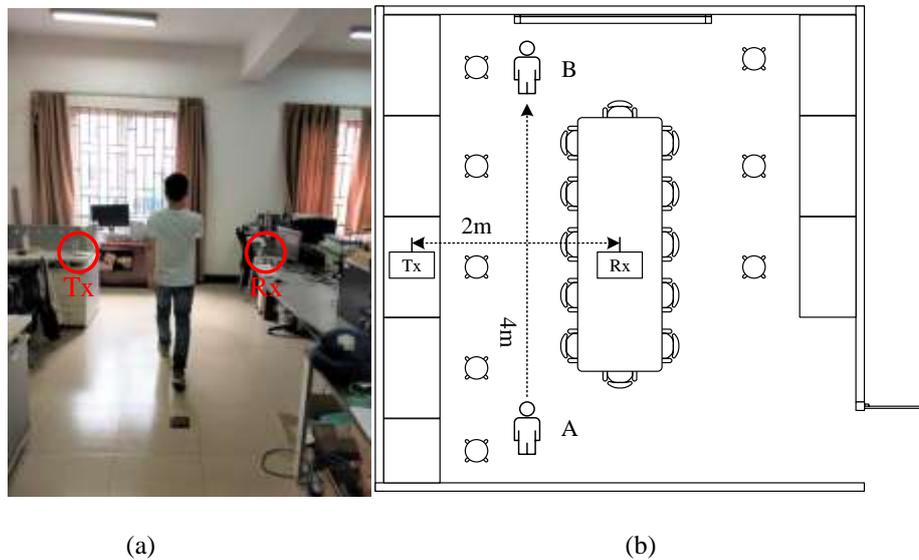


Fig 12: The experimental environment showing

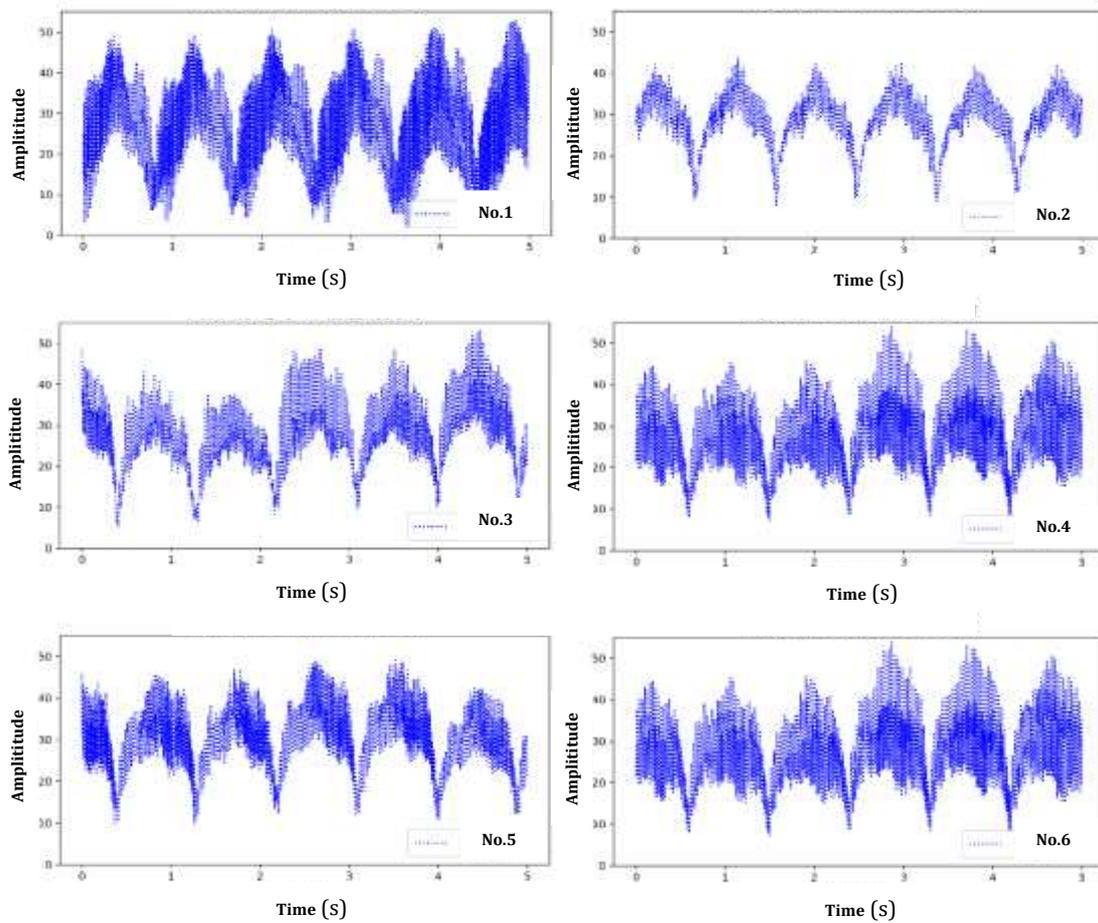


Fig 13: Changes in CSI caused by the activities of different people

4.2 Accuracy of recognition

Training accuracy and test accuracy are two basic evaluation indexes to measure the performance of model recognition. The training accuracy $Acc1$ is the percentage of the number of correctly identified training samples T_A compared to the total number of training set samples T_N . The test accuracy $Acc2$ is the similar percentage for correctly identified test set samples T_a relative to the total number of test samples T_n . However, the test set samples cannot be used to update the model weights, meaning that the test accuracy values measure the performance of the model with unknown data. Training accuracy and test accuracy can be expressed as

$$Acc1 = \frac{T_A}{T_N} \quad (19)$$

$$Acc2 = \frac{T_a}{T_n} \quad (20)$$

We used an existing CNN identification model and our new identification model to carry out experiments, and the results are shown in Figure 14. The accuracy of the new identification model and CNN identification model with training data was close to 100%. However, the accuracy of the CNN identification model for the test samples was 94.2%, while the accuracy of our new model was 98.7%. Because the CNN identification model only extracts amplitude feature information from CSI samples, gait features with the same amplitude at different positions were

considered as the same feature. The new model’s BiLSTM network extracts the time sequence features of different amplitude gaits and overcomes this problem.

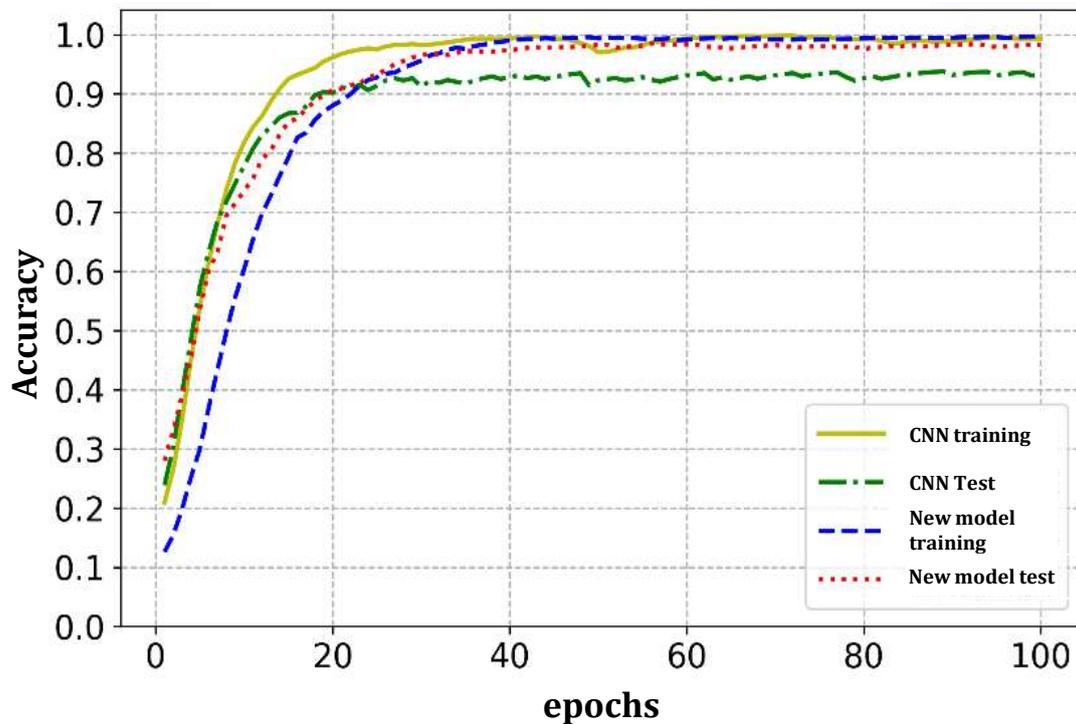


Fig 14: Comparison of recognition accuracy of the two models

4.3 Analysis of loss function

We used a loss function to evaluate the inconsistency between the predicted and actual values of the model, measuring its robustness. The smaller the loss function result, the better the robustness of the model [29]. Figure 15 shows the loss function curve of the new model and the CNN identification model. For the same training batch, the loss function value of the new identification model was lower than that of CNN identification model, indicating that the predicted value of the new model was closer to the real value. In terms of convergence rate, the new model converged with 40 training batches, while the CNN model needed 50 batches to do the same.

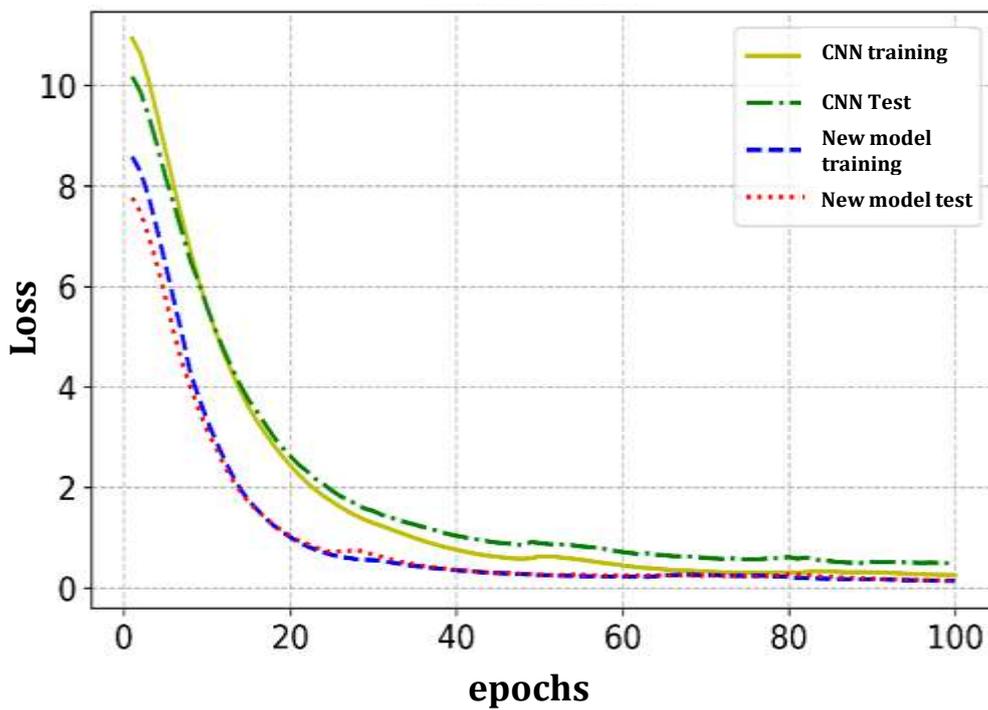


Fig 15: Comparison of recognition accuracy of two models

4.4 Robustness to environmental changes

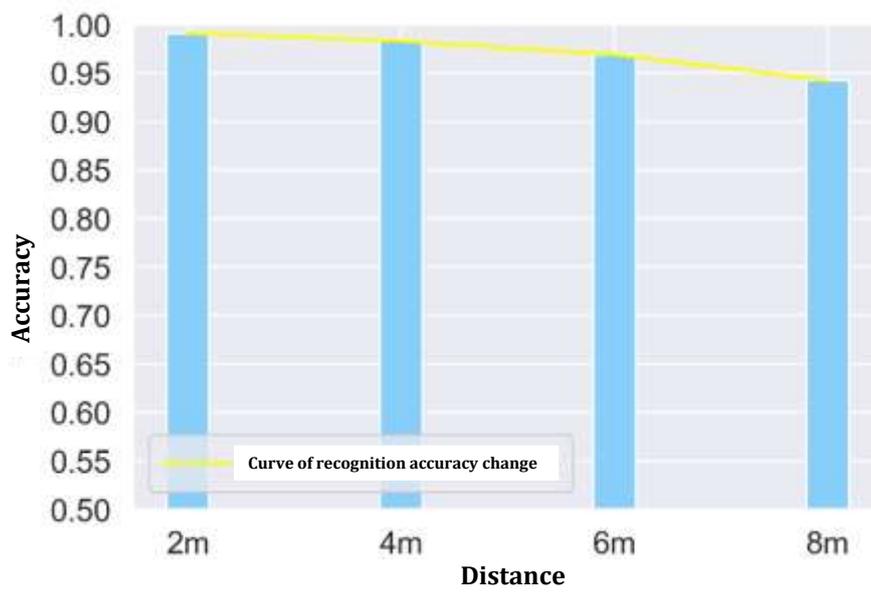


Fig 16: Comparison of recognition accuracy of two models

If the distance between transmitter and receiver is changed, the CSI changes also. If the first test maintains the same recognition accuracy for different CSI, it shows that the model is robust. In the same indoor space, we changed the distance between the receiver and transmitter to be 2 m, 4 m, 6 m, and 8 m in turn while maintaining the height of the receiver and transmitter at 1 m above the ground. We had participants maintain their normal walking speed during the experiments. The experimental results show that when the distance between the transmitter and the receiver was 2 m, 4 m, 6 m, and 8 m, the accuracy of the new model was 99.1%, 98.3%, 96.9%, and 94.2%, respectively. The relationship between the recognition accuracy of the model and the distance between the transmitter and the receiver is shown in Figure 16. This shows that in a specific indoor space, the new identification model still maintained high accuracy.

V. Conclusions

Using the change in CSI caused by indoor human activities, we have implemented a device to identify individuals through the changes in the CSI. The device uses a low-cost embedded computer and Wi-Fi network card, making it portable and concealable. In order to improve the quality of the received CSI, we have also proposed a denoising method combined with outlier removal, a low-pass filter, and a discrete wavelet transform. In order to solve the problems of complexity and insufficient feature extraction of existing identification methods using CSI, we have constructed a new identification model combining a CNN and BiLSTM networks. Our process extracts gait features from the two deep learning models, fuses them into human identity features, and identifies them via classification of the Softmax function. Experimental results show that the accuracy of our new model is 4.5% higher than the existing model using a CNN alone. A test involving 30 people shows that the accuracy can reach 98.7%. In the future, we will further improve our method, increase the ability of transfer learning, make it suitable for different indoor space environment, and reduce the time for repeated training.

VI. Conflict of Interest

The authors declare no conflict of interest.

VII. Author Contributions

Y.L. and J. Z. designed the proposed scheme; Y.L. performed the experiments and analyzed the results.

VIII. Funding

This research was funded by the Innovation Program of SWIP, grant number 201901XWCXRC005.

Acknowledgments

We thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of this manuscript.

References

- [1] Herath S, Harandi M, and Porikli F. Going deeper into action recognition: A survey. *Image and Vision Computing*, 60(2):4–21, 2017.
- [2] Airey P, Verran J. A method for monitoring substratum hygiene using a complex soil: The human fingerprint. *Proceedings of a conference held at Jesus College, Cambridge*, 20–22, 2006.
- [3] Molenberghs G, et al. Review of iris recognition: cameras, systems, and their applications. *Sensor review*, 26(1):66–69, 2006

- [4] Y. Zhang, G. Pan, K. Jia, et al. Accelerometer-based gait recognition by sparse representation of signature points with clusters. *IEEE trans*, 45(9):1864-1875, 2015.
- [5] Orović I, et al. A new approach for classification of human gait based on time-frequency feature representations. *Signal Processing*, 91(6):1448– 1456, 2011.
- [6] Y. Gu, F. Ren, J. Li. PAWS: Passive Human Activity Recognition Based on WiFi Ambient Signals. *Internet of Things Journal, IEEE*. 3(5):796-805, 2016.
- [7] W. Xiao, B. Song, X. Yu, et al. Nonlinear Optimization-Based Device-Free Localization with Outlier Link Rejection. *Sensors*. 15(4):8072-8087, 2015.
- [8] Halperin D, Hu W, Sheth A, et al. Tool Release: Gathering 802.11n Traces with Channel State Information. *Acm Sigcomm Computer Communication Review*, 41(1):53, 2001.
- [9] Kosba A E, Saeed A. RASID:A robust WLAN device-free passive motion detection system. 2012 IEEE International Conference on Pervasive Computing and Communications, IEEE, 180-189, 2012.
- [10] Murakami, T.; Miyazaki, M.; Ishida, S.; Fukuda, A. Wireless LAN-Based CSI Monitoring System for Object Detection. *Electronics* 7, 290,2019. <https://doi.org/10.3390/electronics7110290>
- [11] T. Zhang, T. Song, D. Chen, et al. WiGrus: A Wifi-Based Gesture Recognition System Using Software-Defined Radio. *IEEE Access*, 131102-131113, 2019.
- [12] Z. Chen, L. Zhang, C. Jiang, et al. WiFi CSI Based Passive Human Activity Recognition Using Attention Based BLSTM. *IEEE Transactions on Mobile Computing*, 1-1, 2018.
- [13] Y. Zhang, Y. Lin, Z. Dou, et al. Li. Monitoring and Identification of WiFi Devices for Internet of Things Security. 2019 IEEE Globecom Workshops, 1-5, 2019.
- [14] G. Yu, F. Z. Yi, L. Jie, et al. Sleepy: Wireless Channel Data Driven Sleep Monitoring via Commodity WiFi Devices. *IEEE Transactions on Big Data*, 1-1, 2008.
- [15] Arshad S, Feng C , Elujide I , et al. SafeDrive-Fi: A Multimodal and Device Free Dangerous Driving Recognition System Using WiFi. 2018 IEEE International Conference on Communications(ICC), 1-6, 2018.
- [16] Y. Zeng, Pathak P H, Mohapatra P. WiWho: WiFi-Based Person Identification in Smart Spaces. 15th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN), 1-12, 2016.
- [17] T. Xin, B. Guo, Z. Wang, at al.FreeSense: Indoor Human Identification with Wi-Fi Signals. 2016 IEEE Global Communications Conference (GLOBECOM), 1-7, 2017.
- [18] R. Zheng, Y. Zhao, B. Chen. Device-Free and Robust User Identification in Smart Environment Using WiFi Signal. IEEE International Conference on Ubiquitous Computing and Communications (IUCC), 1039-1046, 2017.
- [19] J. Lv, W. Yang, D. Man, et al. Wii: Device-Free Passive Identity Identification via WiFi Signals. Globecom IEEE Global Communications Conference, 1-6, 2017.
- [20] J. Zhang, B. Wei, W. Hu, et al. WiFi-ID: Human Identification Using WiFi Signal.2016 International Conference on Distributed Computing in Sensor Systems (DCOSS), 75-82, 2016.
- [21] Jie W, Senior M et al. Device-Free Identification Using Intrinsic CSI Features.IEEE Transactions on Vehicular Technology, 2018:8571-8581.
- [22] Pokkunuru A, Jakkala K, Bhuyan A, at al. NeuralWave: Gait-Based User Identification Through Commodity WiFi and Deep Learning. IECON 2018 - 44th Annual Conference of the IEEE Industrial Electronics Society, 758-765, 2018.
- [23] F. Wang, J. Han, S. Zhang, et al. CSI-Net: Unified Human Body Characterization and Action Recognition. Cornell University, 2018.
- [24] Davies L, Gather U. The Identification of Multiple Outliers. *Journal of the American Statistical Association*, 88(423):782-792, 1993.
- [25] Mallat, S. A Theory of Multiresolution Signals Decomposition: The Wavelet Transform. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11, 674-693, 1989.

- [26] P. Li, Y. Zheng, C. Han, et al. Wavelet threshold methods used in lightning transient electrical signals denoising. *Fuzzy Systems and Knowledge Discovery*, 2027-2029. 2012.
- [27] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ImageNet classification with deep convolutional neural networks, in *NIPS*, pp. 11061114, 2012.
- [28] S. Zhang, D. Zheng, X. Hu, M. Yang. Bidirectional long short-term memory networks for relation classification. In: *Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation*; 73–78, 2015.
- [29] Pokkunuru A, Jakkala K, Bhuyan A, et al. NeuralWave: Gait-Based User Identification Through Commodity WiFi and Deep Learning. *IECON 2018 - 44th Annual Conference of the IEEE Industrial Electronics Society*, 758-765, 2018.