Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 7821241, 12 pages
https://doi.org/10.1155/2021/7821241

Research Article

Recognition for Human Gestures Based on Convolutional Neural Network Using the Off-the-Shelf Wi-Fi Routers

Haixia Yang,[1] Zhaohui Ji,[2] Jun Sun,[3] Fanan Xing,[2] Yixian Shen,[2] Wei Zhuang,[2] and Weigong Zhang[1]

[1]*School of Instrument Science and Engineering, Southeast University, Nanjing 211189, China*
[2]*School of Computer, Nanjing University of Information Science and Technology, Nanjing 210044, China*
[3]*College of Computer and Information, Hohai University, Nanjing 210098, China*

Correspondence should be addressed to Weigong Zhang; zhangwg@seu.edu.cn

Received 21 August 2021; Revised 4 October 2021; Accepted 9 October 2021; Published 3 November 2021

Academic Editor: Pengfei Wang

Human gestures have been considered as one of the important human-computer interaction modes. With the fast development of wireless technology in urban Internet of Things (IoT) environment, Wi-Fi can not only provide the function of high-speed network communication but also has great development potential in the field of environmental perception. This paper proposes a gesture recognition system based on the channel state information (CSI) within the physical layer of Wi-Fi transmission. To solve the problems of noise interference and phase offset in the CSI, we adopt a model based on CSI quotient. Then, the amplitude and phase curves of CSI are smoothed using Savitzky-Golay filter, and the one-dimensional convolutional neural network (1D-CNN) is used to extract the gesture features. Then, the support vector machine (SVM) classifier is adopted to recognize the gestures. The experimental results have shown that our system can achieve a recognition rate of about 90% for three common gestures, including pushing forward, left stroke, and waving. Meanwhile, the effects of different human orientation and model parameters on the recognition results are analyzed as well.

1. Introduction

Gesture is a common visual body activity that contains a rich content of indications. Unlike physical device-based control methods, gesture control does not require specific devices for smart home applications such as keyboards, mice, and remote controller. Nowadays, human gesture is becoming a promising indication for human-computer interaction due to the contactless and device-free characteristics for many Internet of Things applications [1].

In traditional gesture recognition systems, a variety of sensing devices such as high-resolution cameras and various sensors have been proved to be able to recognize gesture activities [2]. But these sensing devices have great limitations in installation difficulty, privacy security, and installation cost [3]. In contrast, wireless signal gradually becomes a very attractive carrier for gesture recognition. Because with the advent of the urban Internet of Things, Wi-Fi devices will become ubiquitous, and Wi-Fi signals also spread throughout the home environment [4]. Wireless signals do not need to be received in the range of sight and can easily penetrate walls, so Wi-Fi has the advantages of easy deployment and low cost in activity recognition [5].

Therefore, Wi-Fi is easy and inexpensive to deploy for motion recognition. At the same time, the gesture recognition based on Wi-Fi signal can also realize the gesture recognition in a more private way without strong user perception, and the security of privacy is also guaranteed.

In the nineteenth century, the famous physicist Fresnel proposed the Fresnel zone model, which is widely used in the field of optical wave dynamics to explain the interference and diffraction of light. With the development of wireless sensing, researchers found that the propagation of radio frequency signal can also be explained by the Fresnel zone model. Different gestures will crossdifferent Fresnel zone at different angles and speeds, so the interference is also

different, and the final CSI sequence is also different. This provides a theoretical basis for identifying different gestures through the CSI sequence of Wi-Fi devices.

2. Related Works

The gesture recognition based on the wireless signal of commercial Wi-Fi devices can be roughly divided into two categories, one is the early RSS-based system, and the second is the more popular CSI-based system at present.

In 2015, Abdelnasser et al. proposed a gesture recognition system WiGest [6]. The system identified five gestures by analyzing received signal strength (RSS) changes. The recognition rate reaches 87.5% when using a single wireless access point and more than 96% when there are multiple wireless access points. However, RSS can only get a value of overall signal change in measurement, so the RSS-based perceptual recognition system can only identify some coarse-grained human activities. In 2010, Halperin et al. from the University of Washington, in collaboration with Intel, released a tool called CSITOOL based on the Intel 5300 network card [7]. This tool provides a method to extract CSI values from commercial Wi-Fi devices on Linux systems. Since then, passive sensing technology based on CSI of commercial Wi-Fi devices soon became a research hotspot in wireless sensing. In 2018, Yu et al. from Nanjing University proposed a QGesture gesture recognition system [8]. This system separates gesture movements from the daily environment by modeling the amplitude and phase changes in CSI with the direction and distance of gesture movements. The system finally achieves more than 90% recognition rate in the presence of multiple interfering daily activities. Jiang et al. at the State University of New York at Buffalo, USA, proposed an EI framework for device-free activity recognition based on deep learning techniques [9]. The core module of the EI framework is an adversarial network, which consists of three main components: feature extractor, human activity recognizer, and environment discriminator. The EI framework removes the environmental information contained in the activity data in the process of integrated learning, and the recognition rate reached more than 90% [10]. Zheng et al. of Tsinghua University proposed a gesture recognition system called WIDAR3.0 in 2019 [11]. This system adopts an environment-independent signal feature—velocity spectrum BVP in human body coordinate system. It applies time-frequency analysis to CSI sequence to estimate Doppler frequency shift profile, restores BVP sequence, and realizes crossdomain gesture recognition. Its intradomain recognition rate for pushing, waving, and other activities reaches 92.7%, and the crossdomain recognition rate reaches between 82.6% and 92.4%.

It is not difficult to see from the above; using channel state information for behavior recognition has become the mainstream scheme in the field of wireless perception. Moreover, the behaviors that can be identified through CSI gradually shift from coarse-grained daily activities such as falls and walking to fine-grained actions such as gestures and gaits. In this paper, we proposed a gesture recognition system based on the CSI of Wi-Fi devices. The main content of our paper is structured as follows: in Section 2, we outlined the basic content of the system design. In Section 3, we introduced the Fresnel zone model, discussed the influence of gesture activity on CSI signal, and finally explained the characteristics and principles of CSI quotient model. In Section 4, the principles and steps of the four basic modules, namely, system data acquisition and extraction, CSI data preprocessing, gesture feature extraction, and gesture classification and recognition, were detailed. In Section 5, we designed the experimental scheme and made a comprehensive analysis and comparison of the experimental results. In Section 6, we summarized the system design process.

3. The Analysis of Wi-Fi Signals of Gesture Activities

3.1. Channel State Information of Wi-Fi. CSI estimates the channel information by representing the channel attribute of the communication link, which can reflect the influence of the environment on the signal transmitted from the transmitter equipment to the receiver [12]. CSI can be extracted from commercial Wi-Fi devices based on 802.11.n protocol. Wi-Fi equipment adopts multiple-input multiple-output (MIMO) technology. A communication system consists of multiple transmitting antennas and receiving antennas. Therefore, CSI data collected from commercial Wi-Fi devices include multiple subcarriers on different receiving and transmitting antenna channels. Each pair of amplitude and phase of CSI signal describes the state information of subcarriers.

In a narrowband smooth fading channel, the orthogonal frequency division multiplexing system in the frequency domain is modeled as

$$Y = HX + N. \tag{1}$$

In the formula, Y and X represent the received signal vector and the transmitted signal vector, respectively, and H and N represent the channel matrix and the random Gaussian white noise vector, respectively. The channel state information of the subcarrier can be estimated from the above equation as

$$\widehat{H} = \frac{y}{x}. \tag{2}$$

It describes the channel gain between the transmitting antenna of a commercial Wi-Fi device and the receiving antenna of a network card, where the CSI of a single subcarrier is mathematically represented as

$$h = |h| e^{j \bullet \sin(\angle h)}, \tag{3}$$

where $|h|$ is the amplitude and $\angle h$ is the phase of each subcarrier. CSI-based sensing systems usually use the amplitude and phase values of CSI to achieve the sensing of action behavior.

3.2. The Influence of Gesture Activities on CSI. In this paper, we design a preexperiment to verify the interference of hand gestures to Wi-Fi signal CSI sequences. In the case of no change in the position of the receiving end and the sending end, the experimenters made two different gestures of push and clap. The Savitzky-Golay filter is used to smooth the CSI amplitude, and the final CSI amplitude sequence is shown in Figures 1 and 2, respectively. Through comparative analysis, we can draw the following conclusions. Gesture movements will have an impact on the CSI sequence waveform, and different gesture movements have different impacts on the CSI sequence waveform. These conclusions provide a theoretical basis for the recognition of different gestures through CSI sequence.

3.3. CSI Quotient Model. The CSI amplitude extracted in the actual typical indoor environment has noise due to hardware defects and environmental interference, which greatly limits the clarity of CSI data. However, in the recognition of fine-grained gestures, there is a high requirement for the clarity of CSI sequence. The amplitude of CSI can be filtered to reduce some ambient noise so as to restore a relatively regular CSI amplitude sequence. However, there is a random offset noise in the CSI phase sequence that is difficult to compensate, which completely limits the use of CSI phase information in monitoring the direction and displacement of hand movement.

In 2019, Zeng et al. proposed the CSI quotient model to push the range limit of Wi-Fi-based respiration sensing [13]. This model abandons the CSI value collected by the original single antenna and takes the quotient of the CSI values received by the two adjacent antennas at the receiver as a new metric. Mathematically, the magnitude of the CSI quotient is the quotient of the original CSI magnitude, while the phase of the CSI quotient is the phase difference of the original CSI data.

$$\mathrm{CSI}_{\mathrm{quotient}} = \frac{\mathrm{CSI}_{\mathrm{antenna1}}}{\mathrm{CSI}_{\mathrm{antenna2}}}. \tag{4}$$

The CSI quotient model is based on the following two conclusions. (1) For commercial wireless NETWORK CARDs, the same set of crystal clocks and RF processing circuits are shared between different antennas on the NETWORK CARD, and their amplitude noise is equally proportional. At the same time, the time-varying phase offsets of different antennas are the same. (2) When the observation target moves for a small distance, the difference of the reflected path length between two adjacent antennas can be considered as a constant.

In order to further compare the difference between original CSI and CSI quotient, we conducted a comparative experiment. During the experiment, the volunteer made a pushing forward gesture within 1 m of the receiving and transmitting equipment of the receiver and then extracted the CSI values and CSI quotient amplitude sequences collected on the original two receiving antennas for comparison. The amplitude sequence of CSI value is shown in Figures 3 and 4, and the amplitude sequence of CSI quotient is shown in Figure 5. It can be seen that the amplitude waveform of CSI quotient is obviously better than that of the other two original CSI values.

From the above analysis and comparison experiments, it is easy to find that the CSI quotient has more excellent characteristics than the original CSI value. CSI quotient can also provide better performance addition to the system when used for gesture recognition. Therefore, the CSI quotient model is used for gesture recognition in this paper.

4. Design of Gesture Recognition System

4.1. Software Environment. The system is built on personal laptop. This study uses Python to parse, preprocess, extract features, and classify CSI gesture data and uses Pycharm as an integrated development environment for system development. Anaconda3 is selected as the interface. Anaconda integrates a large number of Python tool libraries, which is one of the common development environments for developers. In terms of data preprocessing, scipy.signal library is used, which contains common filter functions. In the selection of neural network framework, the system uses Keras to train the neural network model. Keras is a commonly used highly modular deep learning framework. The Scikit-Learn library in Python is used for machine learning, which integrates various machine learning algorithms, and is also very friendly in data preprocessing and classification results visualization.

4.2. Raw Data Acquisition and CSI Extraction. In order to further verify the stability of the gesture recognition system, the system also uses raw CSI gesture data from some public datasets on the basis of self-collected CSI gesture data. The experimental data source in this system design includes the gesture recognition data set adopted by the Zheng et al. team of Tsinghua University in WiDar3.0 in 2019, which is the original CSI gesture package in multiple scenarios [14]. By studying the original CSI data of gestures in different scenarios, it provides great help to study the mechanism behind gestures. At the same time, the WiAr data set released by Guo of Dalian University of Technology is also used in the system [15]. The data set includes 30 groups of reference cases of 16 different activities of 10 volunteers, including gestures such as waving, sliding, and throwing. These data provide a large number of reliable experimental data reference for the study of gesture data preprocessing and gesture feature extraction in this paper.

In the process of actual experiments, phase and amplitude information were used as the comparison for several times. These experimental results showed that the phase information of CSI quotient was identified with high recognition rate and stability in the real environments. Therefore, we decided to use CSI quotient phase information as the original data for identification of human gestures.

4.3. Data Preprocessing

4.3.1. Filtering and Denoising. The CSI data parsed from the original CSI packet contains a large amount of noise, which seriously interferes with the clarity of the CSI sequence waveform. Although CSI quotient model is adopted in this

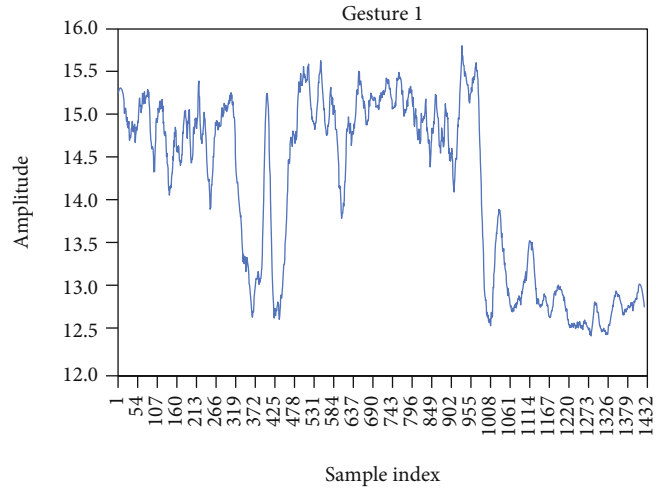

Figure 1: Amplitude sequence of gesture 1.

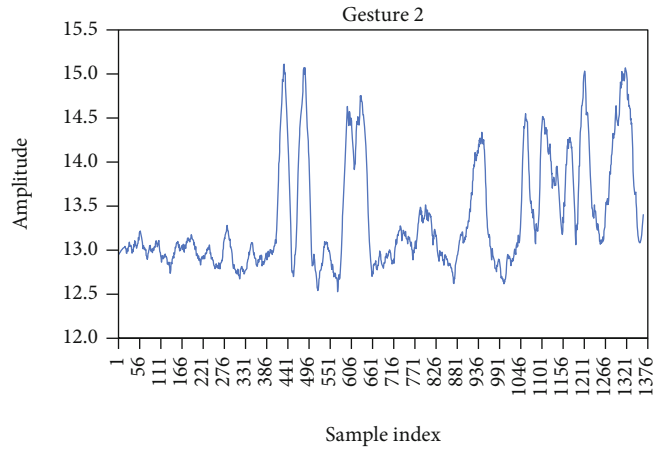

Figure 2: Amplitude sequence of gesture 2.

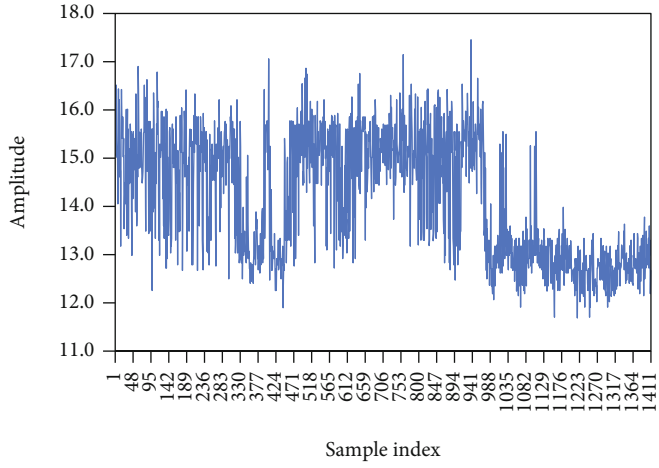

Figure 3: The first receiving antenna amplitude sequence.

system and a large number of noises have been filtered out, the amplitude waveform is not smooth enough. Therefore, the Savitzky-Golay filter is used to further smooth the waveform for subsequent feature extraction and classification.

Savitzky-Golay filtering is a filtering method based on local area polynomial least square fitting for time series signals. One of the characteristics of the filter is that it can ensure the basic stability of the signal width while filtering

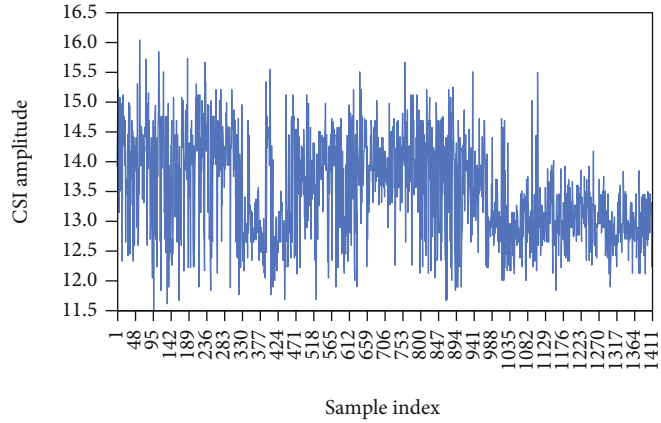

Figure 4: The second receiving antenna amplitude sequence.

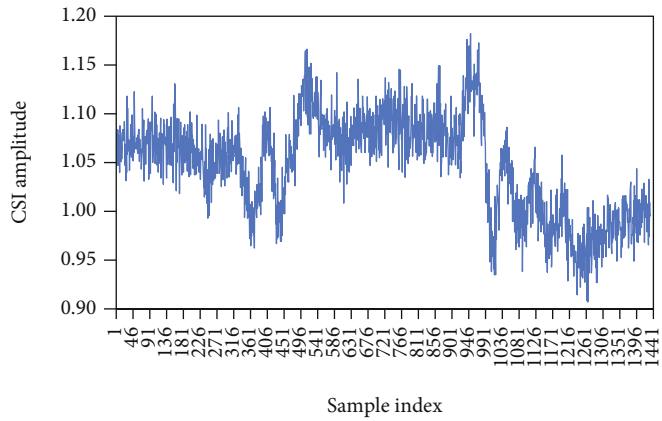

Figure 5: Amplitude sequence of CSI quotient.

out the high-frequency noise in the original signal. Savitzky-Golay method is used to smooth the CSI data, which can improve the smoothness of the CSI gesture data sequence and reduce the interference of environmental noise, so as to facilitate the subsequent gesture feature extraction and classification. When the typical peak of the signal is narrow, the filter has good filtering effect.

The method requires first setting a sliding window of size n and a fitting order k, where $n = 2m + 1$, to perform a left-to-right filtering of the curve before filtering with this window size. Suppose the original signal is $x[r]$, where $r = 0$, 1, -1, 2, -2… Firstly, the filtering center $x[0]$ is selected, and 2 $m + 1$ points of each m point around the center are selected as the primary filtering object. The polynomial fitting of $k - 1$ order is used for n data points in the selected window, as shown in the following formula.

$$y = \sum_{k=0}^{N} a_k n^k. \tag{5}$$

The fitting equation is expressed as a matrix, as the following.

$$Y = X \bullet A + \varepsilon. \tag{6}$$

The least square solution $\widehat{A}$ of A is

$$\widehat{A} = \left(X^T \bullet X\right)^{-1} \bullet X^T \bullet Y. \tag{7}$$

The final predicted value obtained $\widehat{Y}$ is

$$\widehat{Y} = X \bullet \widehat{A} = X \bullet \left(X^T \bullet X\right)^{-1} \bullet X^T \bullet Y. \tag{8}$$

The window slides from left to right on the original signal, and the data points are fitted sequentially until all data points are fitted to the end. The fitted CSI data are finally used for subsequent gesture feature extraction and classification recognition.

A comparison between the original CSI data amplitude sequence and the CSI amplitude sequence filtered by Savitzky-Golay filter is shown in Figures 6 and 7. Through image comparison, it can be obviously observed that most of the noise in the original CSI sequence is removed by two filtering, and most of the change information of the original CSI amplitude curve is retained. After filtering, the CSI amplitude sequence fluctuation caused by gesture can be clearly identified.

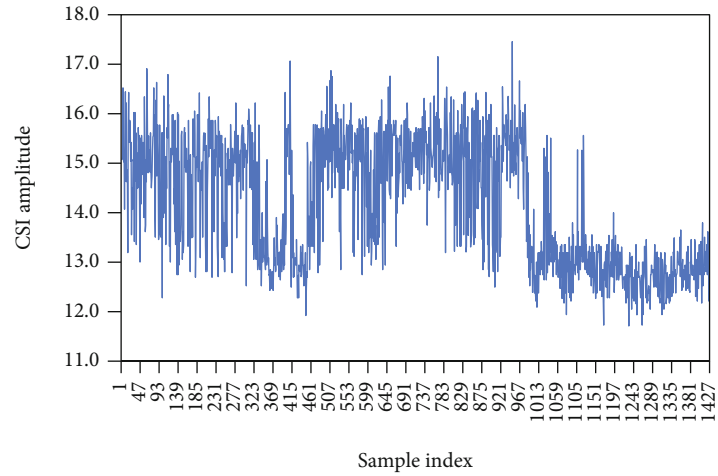

Figure 6: CSI amplitude sequence before denoising.

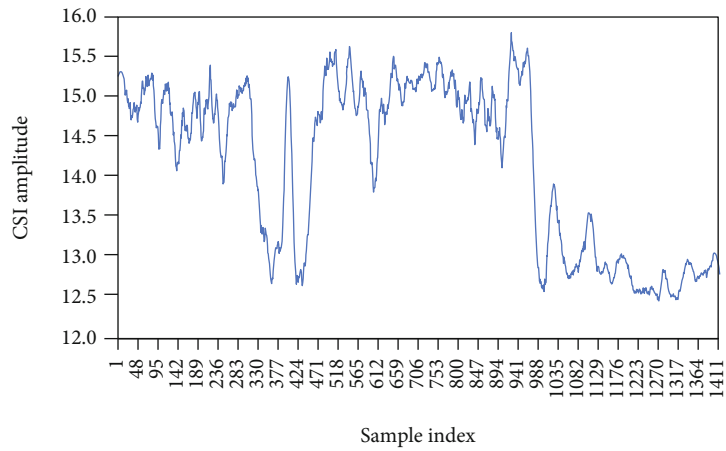

Figure 7: CSI amplitude sequence after denoising.

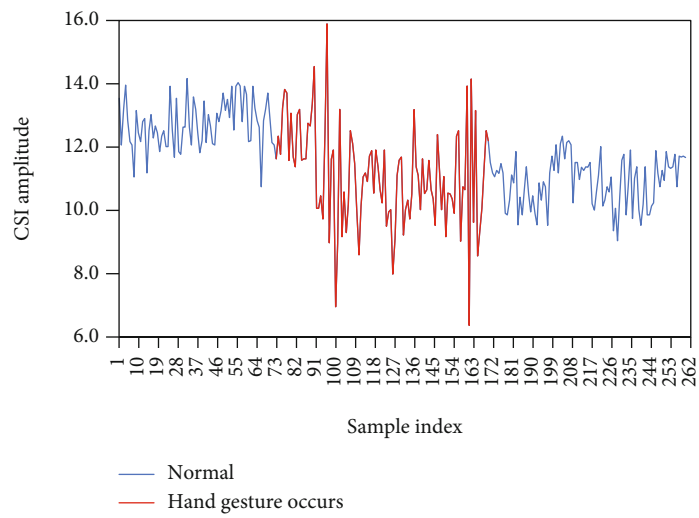

Figure 8: Gesture occurrence interval.

4.3.2. Segmentation of CSI Stream Data. As shown in Figure 8, the starting and ending points of gestures can be determined by detecting the fluctuation of this mutation. In this system, the sliding window variance comparison method is used to detect the occurrence of gesture activity, and the specific implementation process is as follows. Firstly,

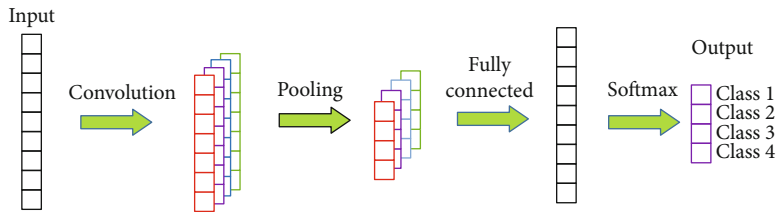

Figure 9: Structure of a one-dimensional CNN.

the window size is set as 10 CSI packets, and the sliding distance is set as 5 CSI packets. By calculating the variance of CSI amplitude in each window, a variance sequence can be obtained. By comparing the variance sequence, the position where the amplitude of CSI fluctuates greatly due to the influence of gesture can be detected. It can be used to roughly identify the beginning and end of each gesture from the data sequence. It is stipulated that when the variance of the latter window is three times that of the former window, gesture activity can be considered to occur between the two windows. The end of the former window is recorded as the occurrence point of the activity. Similarly, the end point of the activity can be found from the back forward.

4.4. Gesture Feature Extraction

4.4.1. 1D-CNN-Based Gesture Feature Extraction. Convolutional neural network (CNN) is a deep learning structure. It consists of multiple convolutional layers and pooling layers combined with each other and is capable of feature extraction and classification recognition of images, sounds, and texts [16].

Since CNNs are widely used in the field of image recognition, researchers usually prefer to convert CSI data into the form of images and use two-dimensional CNNs for image recognition to achieve the purpose of classification [17]. However, CSI values are a set of one-dimensional time series data that can be classified and recognized by a one-dimensional CNN. The one-dimensional CNN has a simpler structure and fewer parameter settings, so it can save computer resources and time, and is conducive to rapid recognition of gestures. The basic structure of the 1D-CNN is shown in Figure 9.

In the convolution layer, the neural network convolutions the segments of the input signal to generate the corresponding one-dimensional feature map. Various convolution kernels extract different features from the input signal, and each convolution kernel detects the specific features of all positions on the input feature map to achieve the weight distribution on the same input feature map. After the convolution layer, the number of feature maps and the dimension of data features increases greatly, which makes the next work difficult. Therefore, the pooling layer is used here to process the features, reduce the number of parameters, and extract the main features. Through multilayer convolution and pooling, the classification results are finally output through the dense layer, and the classification activation function is generally set as softmax.

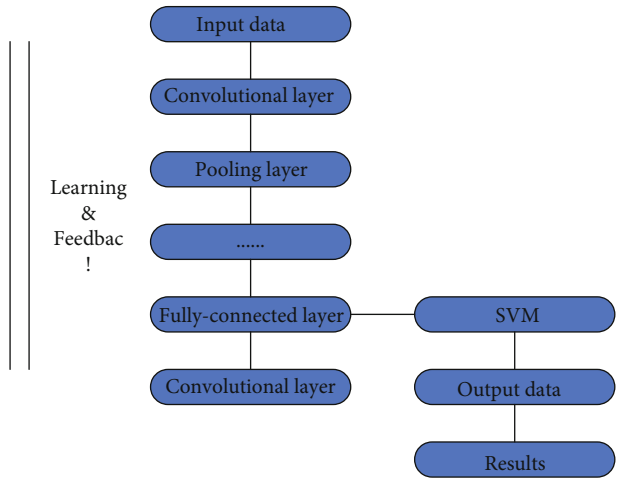

Figure 10: CNN-SVM model diagram.

4.4.2. 1DCNN-SVM Model. Based on the RCNN model proposed by Lei [18], this paper does not use the softmax function to output the results at the end of CNN, but only uses CNN as a feature extractor and inputs the extracted feature vectors into the SVM classifier for training and classification. Girshick and others from Harbin Institute of Technology applied the model to Chinese text recognition [19], and Liu et al. applied the CNN-SVM model to classification of liquids [20], both of which achieved good results. Based on the above research, this paper uses neural network as a feature extractor and SVM as a classifier. The CNN-SVM model is shown in Figure 10.

This paper is always faced with the problems of small amount of data and difficulty to selecting gesture features. 1DCNN-SVM model is undoubtedly very suitable to solve the above two problems. SVM classifier can solve the model overfitting problem caused by the small amount of data to some extent. CNN is a powerful feature extractor, and 1DCNN can automatically and effectively extract the features of CSI gestures [21]. A comparative discussion of the recognition rate of the 1DCNN-SVM model and the CNN model alone will be presented in chapter 4 below.

Through the above analysis, this paper finally adopts the hierarchical structure as shown in Table 1. Regardless of the additional layers, such as batch normalization and dropout layers, the network mainly consisted of five convolutional layers, three pooling layers, one flat layer, and two fully connected layers.

In this paper, the output of full connection layer of one-dimensional convolutional neural network is put into SVM

Table 1: Hierarchical structure of proposed 1D-CNN.

Number	Layer name	Function	Parameter setting
1	Convolutional layer	Extracting local features	Number of convolutional kernels: 16 Size of convolutional kernel: 3
2	Pooling layer	Spatially dimension reduction	Number of pooling kernels: 16 Size of pooling kernel: 3
3	Convolutional layer	Extracting local features	Number of convolutional kernels: 32 Size of convolutional kernel: 3
4	Pooling layer	Spatially dimension reduction	Number of pooling kernels: 16 Size of pooling kernel: 3
5~7	Convolutional layer	Extracting local features	Number of convolutional kernels: 64 Size of convolutional kernel: 3
8	Pooling layer	Spatially dimension reduction	Number of pooling kernels: 16 Size of pooling kernel: 3
9	Flat layer	Realizing the transition from convolution layer to full connection layer	
10	Fully connected layer		Output size: 64
11	Fully connected layer	Classification and recognition using SVM classifier	Output size: 3

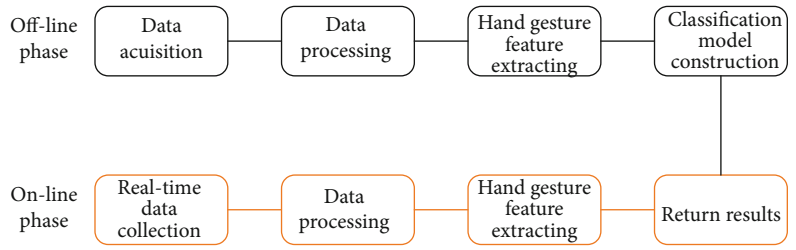

Figure 11: Real-time processing flow chart.

classifier as feature vector to classify and recognize. Therefore, the number of neurons in the full connection layer in the model is the characteristic number of SVM training. By comparing the system recognition accuracy and training time under different feature number, the most suitable value of feature number is 64.

4.5. Design of the Real-Time Recognition System. A set of real-time acquisition system is designed for the data collected by the self-collected Atheros network card. The real-time system is divided into the offline stage and the online stage. In the off-line stage, the basic process of data acquisition and extraction, data preprocessing, and feature extraction process classification and recognition are completed, and the classification model is trained to detect and classify the collected data. The flow chart is shown in Figure 11.

To meet the requirements of data volume and real-time, an unconventional gesture acquisition method is used in this project for the actual data acquisition process. In the process of gesture acquisition, the single gesture activity is no longer stored separately, but the same gesture activity is repeated many times in a fixed time. The same gesture activity is collected in the same CSI original packet. The gestures are sliced, and the models are built using these fragments. In the process of detection, the gestures collected on the scene are sliced with the same size. Then, each fragment is classified, and the classification results are determined to return the final recognition results.

5. Experimental Result and Analysis

5.1. Platform and Data Extraction. In this design, gesture data samples are collected in two indoor environments. The first indoor environment is the home environment outside the school, and 20 gesture samples of hand waving, push forward, and stroke are collected. The second indoor acquisition environment is the school dormitory. Here, 20 gesture samples of hand waving, push forward, and left stroke in four different directions of the human body are collected for comparative experiments.

Two TP-Link TL-WDR4310 Wi-Fi routers were used as signal transceiver devices for data acquisition. The routers were equipped with Atheros AR9344 network cards, and a computer was used as the console to control the two routers for interactive communication for the acquisition of the original CSI data. The two routers are located at the same height, 1.5 m apart, and the collector is located on the vertical line of the two routers, 0.5 m apart from the transceiver equipment. Human face toward two routers for gesture data collection.

In this paper, the data collected by two kinds of network cards are tested simultaneously. The CSI data of Atheros network card are collected by ourselves, and the data collected by Intel5300AGN network card are from the WiAr data set published by Guo, Dalian University of Technology [22–24]. The data of real-time system is collected by Atheros network card.

5.2. Analysis of Experimental Results

5.2.1. Classification Results Based on CSI Data. In order to improve the efficiency of data acquisition and data volume, the data collected by the Atheros network card is classified by window sliding slice. The window size is set to 2 seconds; sliding step size is set to 0.2 seconds. Figure 12 shows the change curves of accuracy and loss during the training process, and it can be seen that the model accuracy and loss tend to smooth out after 10 iterations. Finally, the recognition rate of three different hand gestures, pushing forward, waving, and left stroke, can reach up to 92% or more. After 50 times of random division of the data set, the comprehensive accuracy of the recognition results is between 85% and 92%.

The recognition rate for each gesture is shown in Figure 13, where the recognition rate is 90.1% for the left stroke, 88.6% for the waving motion, and 97.6% for the forward pushing motion. Among the three different gestures, the recognition rate of forward pushing activity is the highest, and the recognition rate of left stroke and waving activity is relatively lower. Left strokes and hand wavings are both right-to-left in physical space, crossing the Fresnel zone similarly and having similar effects on CSI. Therefore, it is easy to confuse the recognition, and the recognition rate is lower than that of the former.

5.2.2. Classification Results. Due to the limited type and quantity of gesture data collected by the Atheros network card, it is difficult to fully detect the stability of the system. Therefore, the system readapts the CSI data of Intel 5300 network card. The gesture data in WiAr dataset are classified as a supplement to the correctness of the detection system [25–27]. Figure 14 shows the curves of accuracy and loss during the training process, and it can be seen that the model accuracy as well as the loss tends to be smooth after 10 iterations. The final combined accuracy for the recognition of the four different gestures can reach up to 94% or more. After 50 random divisions of the data set, the combined accuracy is between 85% and 94%.

5.3. Experimental Analysis

5.3.1. The Influence of Different Human Orientations on the Results. In order to explore the influence of different human orientations on the accuracy of gesture recognition, a comparative experiment is set up in this paper. Three different gesture data are collected in four different directions of the human body. The final recognition results for the three gestures are shown in Figure 15. Direction 1 is the direction facing the transmitter and the receiver connection, direction 2 facing the receiver, direction 3 back to the transmitter and the receiver connection, and direction 4 back to the receiver. From the results, we can see that the recognition rate is lower when on the direction that is facing the transmitter and the receiver connection, and the recognition results of the other three directions are basically similar [28–31].

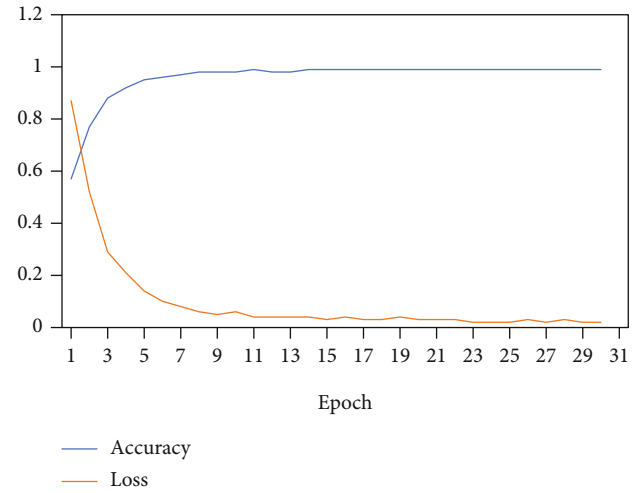

Figure 12: Model training results based on CSI data of the Atheros Network Interface Card.

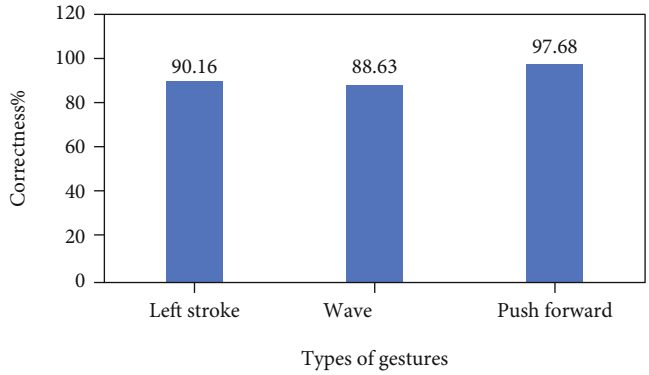

Figure 13: The results of recognition rate of different gestures.

5.3.2. Comparison of Experimental Results of 1DCNN-SVM Model. At the same time, the experiment also compared the performance of separate CNN and CNN-SVM under the same gesture data set [32]. After 50 comparative experiments, the accuracy of CNN and CNN-SVM is shown in Figure 16. It can be seen in 50 comparative experiments that CNN-SVM model is better than the softmax function-activated CNN model in 43 times, accounting for 86%. Among them, the average recognition rate of CNN model using softmax activation classification is 89.8%, and the average recognition rate of CNN-SVM can reach 91.4%. It can be seen that CNN-SVM is suitable for CSI gesture data [33], which can slightly improve the recognition rate of gesture recognition system [34].

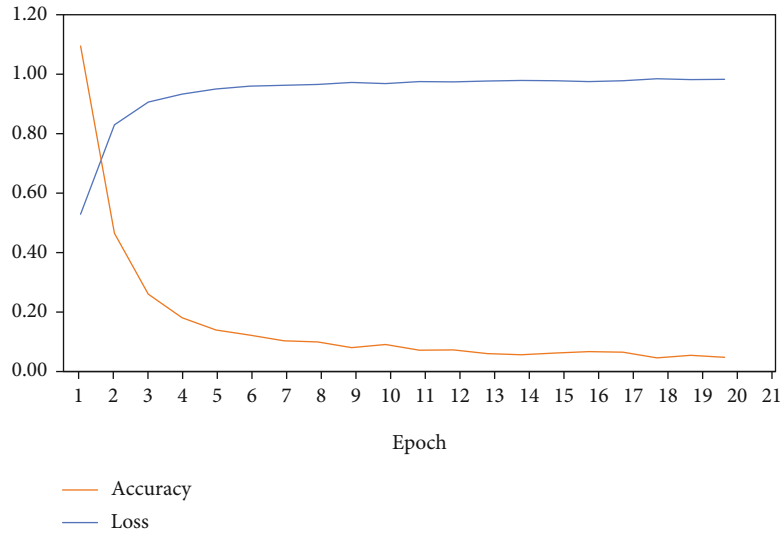

Figure 14: Model training results based on CSI data of Intel 5300 Network Interface Card.

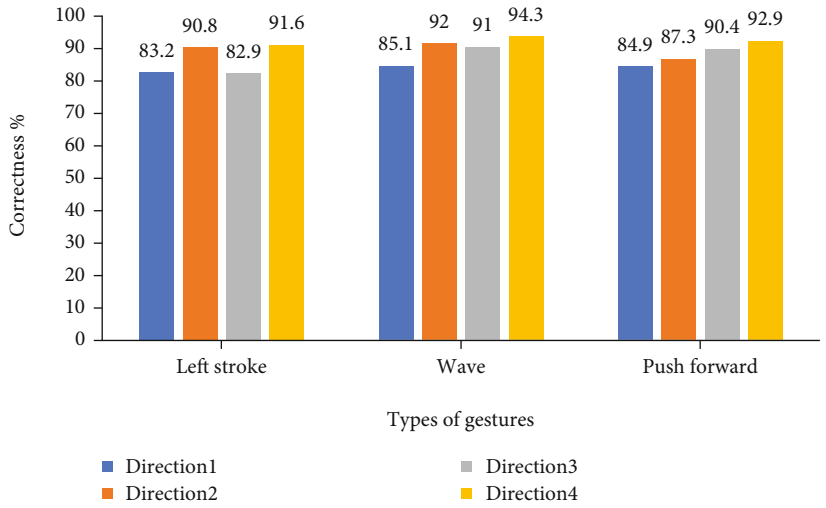

Figure 15: Recognition results under different orientations.

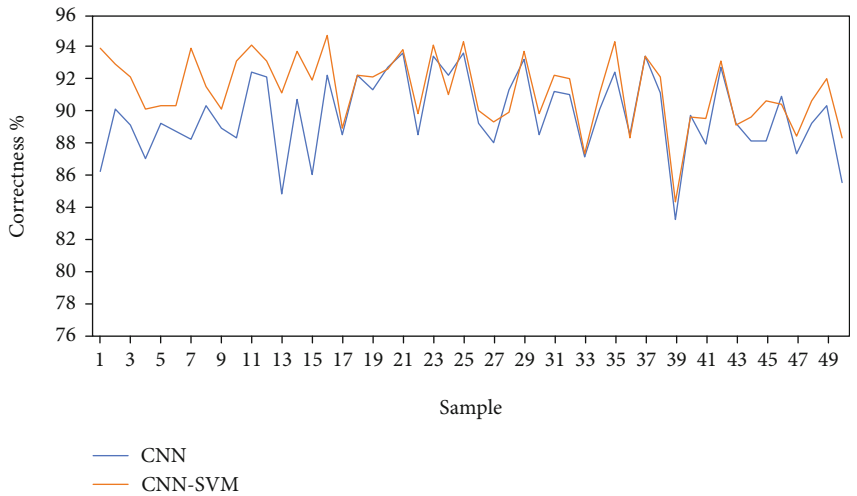

Figure 16: Comparison of the recognition rate.

6. Conclusion

Gesture recognition technology based on the CSI of Wi-Fi device is an emerging gesture recognition scheme. Compared with the traditional gesture recognition technologies based on video images and sensors, it has unique advantages of high privacy and convenient deployment. As a result, the scheme has great potential for use in smart homes as well as smart healthcare. In this paper, a gesture recognition system based on CSI is designed based on Wi-Fi equipment. The system adapts the CSI data collected on two different Network Interface Cards and also compares the correctness of system recognition when the human body makes gestures in different orientations. The Savitzky-Golay filter is used to reduce the noise and smooth the curve of CSI sequence, and the sliding window is used to separate the process of static and motion from CSI, so as to extract the segment of gesture activity. Then, we build appropriate 1D-CNN model to achieve gesture feature extraction. Furthermore, we use the SVM classifier to recognize the gestures and compare the effect of different kernel functions on the gesture recognition results, then select the appropriate kernel function.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest to report regarding the present study.

Acknowledgments

This research was funded by the National Natural Science Foundation of China, grant number 61802196; Jiangsu Provincial Government Scholarship for Studying Abroad; the Priority Academic Program Development of Jiangsu; the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD); and NUIST Students' Platform for Innovation and Entrepreneurship Training Program, grant number 202010300080Y.

References

[1] W. Zhuang, Y. Shen, L. Li, C. Gao, and D. Dai, "Develop an adaptive real-time indoor intrusion detection system based on empirical analysis of OFDM subcarriers," *Sensors*, vol. 21, no. 7, p. 2287, 2021.

[2] J. Su, Z. Sheng, A. Liu, Z. Fu, and Y. Chen, "A time and energy saving based frame adjustment strategy (TES-FAS) tag identification algorithm for UHF RFID systems," *IEEE Transactions on Wireless Communications*, vol. 19, no. 5, pp. 2974–2986, 2020.

[3] J. Su, Z. Sheng, A. X. Liu, Y. Han, and Y. Chen, "Capture-aware identification of mobile RFID tags with unreliable channels," *IEEE Transactions on Mobile Computing*, vol. 20, no. 1, pp. 1–14, 2020.

[4] J. Su, Z. Sheng, V. C. M. Leung, and Y. Chen, "Energy efficient tag identification algorithms for RFID: survey, motivation and new design," *IEEE Wireless Communications*, vol. 26, no. 3, pp. 118–124, 2019.

[5] H. Yang, Y. Shen, W. Zhuang, C. Gao, D. Dai, and W. Zhang, "A smart wearable ring device for sensing hand tremor of Parkinson's patients," *Computer Modeling in Engineering and Sciences*, vol. 126, no. 3, pp. 1217–1238, 2021.

[6] H. Abdelnasser, K. A. Harras, and M. Youssef, "WiGest demo: an ubiquitous WiFi-based gesture recognition system," in *2015 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pp. 1472–1480, Hong Kong, Hong Kong, 2015.

[7] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release," *ACM SIGCOMM Computer Communication Review*, vol. 41, no. 1, p. 53, 2011.

[8] N. Yu, W. Wang, A. X. Liu, and L. Kong, "QGesture," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 1, pp. 1–23, 2018.

[9] W. Jiang, C. Miao, F. Ma et al., "Towards environment independent device free human activity recognition," in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking Mobi Com*, pp. 289–304, New York: ACM, 2018.

[10] Y. Zheng, Y. Zhang, K. Qian et al., "Zero-effort cross-domain gesture recognition with Wi-Fi," in *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services*, pp. 313–325, New York: ACM, 2019.

[11] Y. Zhang, Y. Zheng, G. Zhang, K. Qian, C. Qian, and Z. Yang, "GaitID: robust Wi-Fi based gait recognition," *Wireless Algorithms, Systems, and Applications: 15th International Conference, WASA 2020, Qingdao, China, September 13–15, 2020, Proceedings, Part I*, , pp. 730–742, Springer Nature, 2020.

[12] W. Xi, H. Dong, K. Zhao, Y. Yan, and D. Chen, "Device-free human activity recognition using CSI," in *In Proceedings of the 1st Workshop on Context Sensing and Activity Recognition CSAR*, pp. 31–36, new York: ACM, 2015.

[13] Y. Zeng, D. Wu, J. Xiong, E. Yi, R. Gao, and D. Zhang, "FarSense," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, no. 3, pp. 1–26, 2019.

[14] H. Ge, Z. Yan, W. Yu, and L. Sun, "An attention mechanism based convolutional LSTM network for video action recognition," *Multimedia Tools and Applications*, vol. 78, no. 14, pp. 20533–20556, 2019.

[15] L. Guo, L. Wang, J. Liu, and W. Zhou, "A survey on motion detection using Wi-Fi signals," in *12th International Conference on Mobile Ad-Hoc and Sensor Networks (MSN)*, pp. 202–206, Hefei, China, 2016.

[16] W. Zhuang, Y. Shen, L. Li, C. Gao, and D. Dai, "A brain-computer interface system for smart home control based on single trial motor imagery EEG," *International Journal of Sensor Networks*, vol. 34, no. 4, p. 214, 2020.

[17] L. Guo, L. Wang, J. Liu, W. Zhou, and B. Lu, "HuAc: human activity recognition using crowdsourced Wi-Fi signals and skeleton data," *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 6163475, 2018.

[18] L. P. Lei, "Curve smoothing denoising based on Savitzky-Golay algorithm," *Computer and Information Technology*, vol. 22, no. 5, pp. 30-31, 2014.

[19] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic

segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580–587, Columbus, OH, USA, 2014.

[20] L. Lu, P. Yang, S. Weiwei, and J. Ma, "Similar handwritten Chinese character recognition based on CNN-SVM," in *In Proceedings of the International Conference on Graphics and Signal Processing*, pp. 16–20, new York: ACM, 2017.

[21] L. Hu, *Research on Liquid Classification Detection of WI-FI Channel State Information Based on CNN-SVM*, Hunan University, Changsha, 2018.

[22] W. He, *Research on WI-FI-Based Gesture Recognition Technology*, Shenzhen University, Shenzhen, 2015.

[23] T. Zhang, *Research on Hand Gesture Recognition Based on WI-FI Channel State Information*, University of Electronic Science and Technology, Chengdu, 2019.

[24] Y. Lu, S. Lv, X. Wang, and X. M. Zhou, "A review of research on human behavior sensing technology based on Wi-Fi signal," *Journal of Computer Science*, vol. 2, pp. 1–21, 2019.

[25] L. Jiahui, W. Yujie, and L. Yi, "LSTM-based gesture recognition method for CSI," *Computer Science*, vol. 46, Supplement 2, pp. 283–288, 2019.

[26] X. Pan, *Research on the Key Technology of Dynamic Gesture Recognition Based on 1D-CNN in Wi-Fi Environment*, Beijing University of Posts and Telecommunications, Beijing, 2019.

[27] W. Zhuang, Y. Shen, J. Zhang, C. Gao, Y. Chen, and D. Dai, "A contactless sensing system for indoor fall recognition based on channel state information," *International Journal of Sensor Networks*, vol. 34, no. 3, p. 188, 2020.

[28] G. Yang, *Research on Wi-Fi-Based Hand Action Recognition Algorithm*, Northwestern University, Xi'an, 2019.

[29] K. Niu, Z. Fusang, D. Wu, and D. Q. Zhang, "Exploring the stability of Wi-Fi-aware system with Fresnel zone model," *Computer Science and Exploration*, vol. 15, no. 1, pp. 60–72, 2021.

[30] K. Oyedotun and A. Khashman, "Deep learning in vision-based static hand gesture recognition," *Neural Computing and Applications*, vol. 28, no. 12, pp. 3941–3951, 2017.

[31] Z. Yang, Z. Zhou, and Y. Liu, "From RSSI to CSI: indoor localization via channel response," *ACM Computing Surveys*, vol. 46, no. 2, p. 25, 2013.

[32] F. Youbing, Y. Lu, and Z. Weibo, "Design and implementation of face recognition system based on CNN and SVM," *Computer and Digital Engineering*, vol. 49, no. 2, pp. 378–420, 2021.

[33] Y. Di, *Research on Location-Unknown Gesture Recognition Method Based on Wi-Fi Signal*, Henan University, Kaifeng, 2020.

[34] X. Li, D. Q. Zhang, Q. Lv et al., "IndoTrack," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 3, pp. 1–22, 2017.