Research Article

Wi-Exercise: An Indoor Human Movement Detection Method Based on Bidirectional LSTM Attention

ZhanJun Hao,1,2 YunZhu Kang,1 XiaoChao Dang,1,2 and Ying An1

1College of Computer Science and Engineering, Northwest Normal University, Lanzhou, Gansu 730070, China
2Gansu Province Internet of Things Engineering Research Center, Northwest Normal University, Lanzhou, Gansu 730070, China

Correspondence should be addressed to ZhanJun Hao; zhanjunhao@126.com

Received 9 November 2021; Revised 19 March 2022; Accepted 22 March 2022; Published 11 April 2022

Academic Editor: Mario Muñoz-Organero

Copyright © 2022 ZhanJun Hao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Human motion recognition has become a hot topic in the field of human-computer interaction. Due to the traditional method of detecting human movement in indoor environments, there are still problems such as high hardware costs, influenced by the environment and the need for experimenters to be equipped with relevant equipment. With the large-scale deployment of wireless devices and the establishment of wireless network infrastructure, indoor personnel movement recognition systems based on wireless signals are gradually proposed. In this paper, we propose a device-independent motion-detection method Wi-exercise based on temporal recurrent neural networks. The acquired channel state information (CSI) is preprocessed by wavelet transform, and the output CSI matrix is filtered by the principal component analysis (PCA) to extract the feature values. The wavelet function is processed to further remove the noise also to make the action data more distinctive, and then PCA is used to track the entire CSI time series to improve and maintain its dimensional energy. The preprocessed data are subjected to feature extraction to obtain CSI of human motion and establish feature sequences, and then the action model is trained using the bidirectional long short-term memory attention (ABLSTM) mechanism to finally realize human motion recognition. Finally, the robustness of Wi-exercise is tested experimentally, and its performance is compared with that of existing recognition methods. The experimental results show that the average recognition rate of the method is 94.43% in the complex indoor environment, and the method is better than the traditional classifier method for action recognition, with lower and higher recognition accuracy.

1. Introduction

In recent years, the growth of mobile devices has stimulated people’s exploration of new forms of human-computer interaction. As a basic solution to human-computer interaction, human motion recognition has become a hot spot in the field of human-computer interaction [1]. Human motion recognition system has been widely used in electronic products and mobile devices, such as smart phones [2], notebook computers [3], navigation devices [4], and game control systems [5]. These systems usually use various sensors available on the equipment to realize the recognition of human motion, such as inertial sensors [6], acoustics [7], and light sensors [8]. However, there are still many limitations when implementing these technologies, such as equipment placement for specific environments, light sensitivity, high installation cost or equipment cost, and requiring handheld devices or other sensors.

Based on the above problems, researchers began to use wireless signals to realize the identification of human activities without devices. With the large-scale deployment of wireless equipment and the establishment of wireless network infrastructure, an indoor human movement recognition system based on wireless signals is gradually proposed. WiFi signal can be used not only for indoor human movement recognition but also for positioning, gesture recognition, and so on.

Literature [9] proposed an indoor positioning algorithm based on channel state information (CSI) by using phase difference correction.

Literature [10] proposed a gesture recognition method to efficiently recognize digital dynamic gestures in the air.
These systems operate by analysing changes in the characteristics of the wireless signal to correlate, for example, the CSI caused by human movement [11] or changes in the received signal strength indication.

In the existing human behavior recognition, the correlation matrix of CSI amplitude and phase is extracted in literature [12], and the maximum and submaximum eigenvalues are taken to form four-dimensional eigenvectors. However, the recognition rate will increase with the increase of eigenvectors.

Wan et al. [13] proposed a deep learning framework based on a smartphone inertial accelerometer using three layers of convolutional pooling to extract local features for human action classification; the authors also used methods such as LSTM and BiLSTM for comparison; Maitre et al. [14] access the raw data and classical features of the sensors to the convolutional, dense, and connected layers for feature-level fusion; Mahmud et al. [15] used a two-layer LSTM network to extract temporal features from different sensors and fed these features aggregated into a global feature optimizer network consisting of three consecutive LSTM layers to optimize the aggregated feature vectors. Chen et al. [16] proposed a complex human action recognition method based on CNN and LSTM, which designed specific convolutional subnetwork structures for different sensor data and extracted the output of all subnetworks through fully connected and convolutional layers to fuse features and then used a two-layer LSTM network to extract potential action features; Xia et al. [17] proposed a deep neural network combining LSTM and CNN. The original data of this model is fed to the two-layer LSTM and then to the convolutional layer, which enables the LSTM to learn the temporal dynamics on different time scales based on the learned parameters, resulting in better accuracy. And the use of a global average pooling layer instead of the convolutional fully connected layer greatly reduces the model parameters while maintaining high recognition rate; Zhenyu and Lei [18] proposed a DeepConvGRU model, which performs automatic feature extraction and action classification using convolutional neural networks and gated recursive units as learning models.

In this paper, we propose a highly accurate CSI-based Wi-exercise, which uses a WiFi device with an Atheros 9580 NIC chip for CSI data collection. Wavelet transform is used to preprocess the signal, and the output CSI matrix is filtered and downsampled by principal component analysis (PCA), and then the classification model is trained using a bidirectional long short-term memory network-attention mechanism to identify human actions. The method is experimentally proven to improve the accuracy of person movement recognition.

The main work of this paper is as follows:

(1) How to achieve low-cost, high-precision, and fine-grained action recognition

There are various ways of action-based human-computer interaction, but in terms of the technologies implemented, they can be divided into three main categories, sensor-based, image recognition, and wireless signals. In recent years, there has been an explosion in the variety and number of smart devices, and the most representative wearable smart devices are Google’s Google Glass and Apple’s iWatch, which use cutting-edge artificial intelligence technology to integrate motion-detection and health-monitoring functions and are very popular among consumers. However, the rapid popularity also revealed some problems with existing devices: users either need to wear a wide variety of sensors or need to deploy additional sensor nodes in the environment, which undoubtedly affects the user experience. Although vision-based identification or surveillance technologies can overcome the above-mentioned drawbacks, they inevitably violate the user’s privacy because video surveillance needs to be conducted in a visual environment. In contrast, recognition techniques implemented using wireless signals overcome these problems very well. Secondly, a wireless signal identification monitoring system is simple to deploy, now based on the IEEE802.11 series of wireless devices in ordinary households, and has been popular, using WiFi signals to achieve monitoring and identification costs that are very low, because WiFi signals can be seen everywhere. With the rapid development of WiFi technology, research related to human motion recognition can be conducted using common commercial WiFi devices, greatly reducing the cost of hardware. Youssef obtained the received signal strength indication (RSSI) to detect the location of the human body by the prevalence of commercial WiFi devices [19]. Compared with RSSI, channel state information (CSI), as a physical layer, can measure the amplitude and phase information of each subcarrier and has a certain multipath discrimination capability. In a complex environment, the structural characteristics of CSI can remain relatively stable. In this general context, with the popular use of the 802.11n protocol, we can extract channel state information from WiFi devices. CSI is present when the signal is emitted from the transmitter, after a series of refraction and reflection to reach the receiver attenuation situation, so we can link the CSI with the pedestrian passing between the transmitter and receiver, so as to achieve the purpose of action recognition.

(2) How to extract CSI that characterizes motion and map CSI to human motion

The WiFi signal propagation model is used to establish a corresponding connection between wireless signals and human motion in physical space. When a person moves indoors, the received radio signal changes significantly, and the CSI amplitude information changes sensitively. Not all signals in the transmission process are useful for practical applications. Distortions can occur at every step of the signal generation, transformation, and transmission process due to environmental and other interfering signals. In many cases, the distortions are so severe that they can seriously interfere with the basic information carried by the signal, ultimately leaving useful information hidden in the noise. After rejecting the bad links by wavelet transform, the noise with less variation is filtered out by PCA to increase the signal-to-noise ratio of the data. The pulse coefficients and discrete
coefficients are used as features reflecting the CSI waveform transformation in the feature extraction stage.

(3) How to efficiently train deep learning models

Under the method of pattern recognition of human action, the features need to be classified after extraction, and many domestic and foreign researches have applied machine learning, neural network, and deep learning methods to classification. Gao et al. proposed an enhanced KNN method to determine the location of a smartphone [20]. A backpropagation (BP) neural network model was developed by Yang et al. to identify apnea states [21]. Sheng et al. proposed a new deep learning framework (called the dual-stream network) to mine spatial-temporal cues in CSI [11]. In this paper, we use ABLSTM to train the action recognition model by inputting 6000 feature values into the ABLSTM network, using the activation function in the output layer and finally using the softmax layer to identify different activities and train the action recognition model.

(4) How to select classification models for different movements

The conventional LSTM can only handle continuous CSI measurements in one direction [22], which means that the current hidden state only considers past CSI, and future information is also essential for activity determination. The traditional CNN (convolutional neural networks) structure for the input image can maintain a certain degree of translational rotation invariance in space for a specific location. This spatial invariance only works in the local range of the input image, and the whole map cannot achieve the overall spatial rotation invariance in the local area of the stack. The pooling layer in the CNN structure has many constraints, such as many useful information is lost when extracting features, the input data is only a local operation, and the feature map in the middle of the CNN framework produces large distortion, which makes it difficult for the CNN to realize spatial transformations such as rotation and scaling of images. The feature map generated during feature extraction by CNN is not an overall transformation of the input data [23], which is more restrictive. In this paper, the proposed ABLSTM network includes both forward and backward layers. Both past and future CSI can be considered in the feature learning.

2. Preliminaries Work

2.1. Channel Status Information. CSI uses signal delay, attenuation, and phase shift that can describe the propagation of the signal. In the frequency domain, the frequency of the signal received by the orthogonal frequency division multiplexing technique is represented as follows:

\[ Q = H \times X + Y, \]  

where \( X \) represents the transmit signal vector, \( Q \) represents the receive signal vector, \( H \) is the channel gain matrix, and \( Y \) is the noise vector. According to Equation (1), \( H \) can be expressed as

\[ \hat{H} = Q/X, \]  

where \( \hat{H} \) is the channel frequency response in the frequency domain.

A set of CSI \( \{H_k(i)\}_{i=1}^{30} \) is a discrete sampling of the channel frequency response within the WiFi bandwidth to represent the amplitude and phase of the subcarriers:

\[ H_k(i) = \|H_k(i)\|e^{j\angle H_k(i)}, \]

where \( H_k(i) \) is the CSI of the \( k \) subcarrier \( t_i(i \in [1,N]) \) in each link at the sampling moment and \( \|H_k(i)\| \) and \( \angle H_k(i) \) are its amplitude and phase, respectively.

2.2. Human Motion Recognition Model under WiFi. In this paper, the WiFi signal propagation model is used to establish a corresponding connection between wireless signals and human motion in physical space. The propagation path of WiFi signals in an indoor environment is shown in Figure 1. When a person moves indoors, the received radio signal changes significantly and the CSI amplitude information changes sensitively [24]. Based on this principle, this paper proposes Wi-exercise, a personnel behavior detection method based on WiFi signals. The CSI data is acquired using a low-cost wireless line, and the tester does not need to carry any additional special sensors. It works in both line-of-sight (LOS) and non-line-of-sight (NLOS) situations.

When a person is present in an indoor environment, the human body generates a number of dispersive paths. Adding these scattered powers to the final received power [25].

\[ P_r(d) = \frac{P_tG_tG_r\lambda^2}{(4\pi)^2(d^2 + 4h^2 + \Delta^2)}, \]

where \( P_t \) is the transmit power, \( d \) denotes the distance from the transmitter to the receiver, \( P_r(d) \) denotes the receive power, \( d \) is a function of distance, \( G_t \) is the transmitter antenna gain, \( \lambda \) denotes the wavelength, \( h \) is the distance from the reflection point on the ceiling or floor to the LOS path, and \( \Delta \) is the path length caused by the human body.

2.3. Wavelet Transform. Wavelet transform (WT) is also a transform analysis method, which is based on the short-time Fourier transform improvement [26], and solves the window size that does not change with frequency; it can provide a frequency change with the “time-frequency” window and is the ideal tool for signal time-frequency analysis and processing. The formula is

\[ WT(a,\tau) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \times \psi \left( \frac{t-\tau}{a} \right) dt. \]  

After rejecting the bad links, human motion causes changes in amplitude and phase on the subcarriers, and
the correlation between the subcarriers obtained by sampling in consecutive sampling times can be used to show the changes in CSI, highlighting the characteristics of different human behaviors. And for the CSI on all $n$ subcarriers remaining after removing the bad links, there are

$$H(i) = [H_1(i), H_2(i), \ldots, H_n(i)]. \quad (6)$$

The CSI $H = [H(i)]_{N \times N}$ during the continuous sampling time $N$ and the correlation matrices $A$ and $P$ of their amplitude phases are calculated separately:

$$A = [a(i,j)]_{N \times N},$$
$$P = [p(i,j)]_{N \times N}, \quad (7)$$

where $a(i,j)$ and $p(i,j)$ denote the correlation coefficients between their $H(i)$ and $H(j)$:

$$a(i,j) = \text{corr}(\|H(i)\|, \|H(j)\|),$$
$$p(i,j) = \text{corr}(\angle H(i), \angle H(j)). \quad (8)$$

The maximum 3 eigenvalues of the magnitude and phase correlation matrices are then extracted separately as features characterizing the behavior of each human body:

$$\alpha = \max (\text{eigen}(A)),$$
$$\rho = \max (\text{eigen}(P)). \quad (9)$$

2.4. PCA Dimensional Reduction. PCA is a data analysis method that increases the signal-to-noise ratio of data by filtering out noise with small changes in the process of dimensionality reduction [27]. It is implemented by transforming the original features with a set of orthogonal vectors to obtain a linear combination of new features.

$$F = Ax,$$
$$\text{cor}(F_i, F_j) = 0,$$
$$\text{var}(F_i) > \text{var}(F_j), i < j, \quad (10)$$

where $\text{cor}(F_i, F_j)$ and $\text{var}(F_i)$ denote the correlation coefficient and variance of the transformed features, respectively, and the larger the amount of information, the larger the variance.

2.5. LSTM Model. The long short-term memory (LSTM) network [28], which is capable of processing sequence-changing data, is a time-based recurrent neural network that has applications in many technical fields. Since LSTM solves the gradient disappearance problem caused by the gradient backpropagation process gradually decreasing, LSTM is chosen as the learning model for action classification in this paper.

The LSTM structure is the unit state, which runs directly along the entire chain, and information can be added to or removed from the unit state. Gate is an optional way of inputting information, and the unit state can be carefully adjusted. The LSTM consists of an S-type neural network layer and a point-by-point product operation. The LSTM structure is shown in Figure 2.

2.6. Channel Attention Mechanism. The attention mechanism consists of a multilayer perceptron (MLP) and a softmax layer [29]. The MLP is divided into three layers, and the number of neurons in the hidden layer of the MLP is reduced to $C/r$ in order to reduce the number of parameters and improve the computational efficiency of the system, with $r$ being the decay rate.

The signal sequence is passed through the MLP and softmax layers to generate attention maps, and finally, the input signal sequence is multiplied with the attention maps by element to obtain the weighted signal sequence. The structure of the attention mechanism is shown in Figure 3.

3. System Design

3.1. System Flow. Wi-exercise supports user motion counting and user motion recognition in scenes in a given environment. Compared with traditional WiFi-based people behavior recognition systems, it supports motion recognition and counting in typical indoor scenes without interference. The Wi-exercise system is divided into three stages; the first stage is CSI data preprocessing: the CSI data set is collected from the hardware device and stored in the database, due to the presence of factors such as external and internal system network card noise, which interferes with the original CSI sequence, so signal preprocessing is performed to reduce the interference of these noisy signals; the second stage is action feature extraction: after the preprocessing is completed, the feature extraction stage is carried out, and labels are attached to the sample sequences; the third stage is the classification and recognition of human behavioral actions. After the training of the model is completed, the collected action data are classified and recognized.

In this paper, a subcarrier with the highest energy is selected as the sample information of CSI [30], and the data of six motor movements of the experimenter are collected in real time. The human motion system model is shown in Figure 4.
3.2. Data Processing

3.2.1. Data Noise Reduction. The received CSI raw data has more noise information unrelated to human action, so this paper introduces wavelet transform and dimensionality reduction by PCA.

In this paper, the db3 wavelet and sym8 wavelet are used for further processing. Firstly, we input the data into the first wavelet function, the wavelet function of the first wavelet function is set to db3, the detail coefficient is sure, and the number of decomposition layers is 8. We input the data after db3 wavelet processing into the second wavelet function, the wavelet function of the second wavelet function is set to sym8, and the number of decomposition layers is 5. The wavelet function further removes the noise and makes the characteristics of the action data more obvious.

Bursts and impulse noise generated by transmit power versus rate change unpredictably between transmitter and receiver. In this paper, we use PCA to track the entire CSI time series to improve and maintain its dimensional energy.

The noise reduction steps are as follows:

1. Normalization: for continuous CSI stream, \(N \times S_p\)-dimensional CSI matrix \(H_{t,r}\) is obtained, which is expressed as follows:

\[
H_{t,r} = \begin{bmatrix} H_{t,r}(1) \\ H_{t,r}(2) \\ \vdots \\ H_{t,r}(N) \end{bmatrix}^T,
\]

where \(N\) is the number of CSI packets and \(S_p\) is the total number of subcarriers. \(H_{t,r}(m)\) is the \(S_p \times 1\) dimensional vector, and \(m\) is the CSI grouping at the \((t, r)\) antenna. Normalize \(H_{t,r}(m)(t, r)\) to get the mean value and unit variance of zero \(Z_{t,r}(m)\):

\[
H_{m,k} = \frac{H_{m,k} - \bar{H}_k}{S_k},
\]

where \(H_{m,k}\) is the array element in \(H_{t,r}\), \(\bar{H}_k\) is the mean of the \(k\)-dimension, and \(S_k\) is the standard deviation of the \(j\)-dimension

2. Calculate the covariance matrix: calculate the normalized covariance \(Z_{t,r}(i)\), and construct the covariance \(n \times n\) matrix of size, where is the number of \(n\) subcarriers

3. Eigendecomposition: the eigenvalue of the covariance matrix \(e\) and its corresponding eigenvector \(q\) are obtained

4. Principal component calculation: principal component \(H \times q\) can be decomposed to reconstruct the information in the original CSI packet

PCA maps the \(m\)-dimensional principal components of sample data to \(n\)-dimensional \((n < m)\) principal components, and \(n\) features are \(m\) linearly independent of each other, as the linear representation. After denoising processing, denoising data of dimension \(N \times 10\) is obtained. The computational complexity is greatly reduced.

In an empty and unmanned environment, Figure 5(a) is the original CSI data after sampling. It can be seen from the figure that the original CSI data has high noise, and the CSI data waveform after Kalman and PCA processing is shown in Figure 5(b).

3.2.2. Feature Extraction. The preprocessed data are subjected to feature extraction to improve the system performance while reducing the computational complexity. As shown in Figure 6, different human actions cause different CSI frequency domain transformations, which are manifested in different CSI data waveforms. In this paper, pulse coefficients and discrete coefficients are selected as features reflecting the CSI waveform transformations. Pulse coefficients reflect the CSI peak-to-peak and trough transformations, while discrete coefficients reflect the fluctuation degree of CSI signals.

The denoised CSI subcarrier time series \(X_i = \{X_1, X_2, \ldots, X_N\}\) is calculated, and its pulse coefficient \(X_p\) is the ratio of the root mean square to the mean value in a sliding window:

\[
X_p = \frac{X_p}{X_r}.
\]
The root mean square value $x_p$ is

$$x_p = \sqrt{\frac{\sum_{i=1}^{N} X_i^2}{N}} = \sqrt{\frac{(X_1^2 + X_2^2 + \cdots + X_N^2)}{N}}. \quad (14)$$

Dispersion coefficient $X_v$ is the ratio of standard deviation to mean value, which can be obtained by the following formula:

$$X_v = \frac{H_v}{X_r}. \quad (15)$$

Four people were assigned to do 6 movements of squats, lunges, push-ups, sit-ups, jumping jacks, and high knee lift, respectively. Each person did 50 groups of each movement, and each group had 10 times. Each movement contained $10 \times 50 \times 6 \times 2 = 6000$ characteristic values, and the characteristic sequence after rearrangement was $R_v = \{ r_{v1}, r_{v2}, \ldots, r_{v30} \}$.

### 3.2.3. Feature Extraction

Traditional LSTM can only deal with continuous CSI measurements in one direction, that is, forward, which means that only past CSI is considered in the current hidden state, and future information is also necessary to determine activity. For example, the activities of deep squats and lunges both require the body to be lowered first, but the final body position of the two activities is different. The BLSTM is used to learn valid features from the original CSI measurements. The BLSTM network consists of two layers, forward and backward, and the ABLSTM network structure is shown in Figure 7. Both past and future CSI can be considered in feature learning. The forward layer encodes the past time step information into the current hidden state, that is, the past information of the CSI sequence is considered. The backward layer encodes the future time step information into the current hidden state, that is, the future information of the CSI sequence is considered. The complete context of the sequences is learned considering the past and future dependency information of the CSI sequences, which are used to identify human activities.
The input of the LSTM time step $t$ is $X_t$, namely, the normalized data. The hidden state $S_{t-1}$ of the previous time step is introduced into the LSTM block, and the calculated hidden state $S_t$ is as follows:

Step 1: from the forgetting gate $f_t$, determine the information to be discarded from the cell state:

$$f_t = \sigma(X_t U^f + S_{t-1} W^f + b_f)$$  \hspace{1cm} (16)$$

Step 2: update new information in the unit status. First of all, the input gate $i_t$ layer is confirmed to be updated, and the tan $h$ layer is introduced to create a candidate value $\tilde{C}$:

$$i_t = \sigma(X_t U^i + S_{t-1} W^i + b_i),$$

$$\tilde{C} = \tanh(X_t U^c + S_{t-1} W^c + b_c)$$  \hspace{1cm} (17)$$

Step 3: update the old cell status, $C_{t-1}$ enters the new state $C_t$, and the output gate $o_t$ outputs cell status:

$$C_t = C_{t-1} \otimes f_t \otimes i_t \otimes \overline{C_t},$$

$$o_t = \sigma(X_t U^o + S_{t-1} W^o + b_o),$$  \hspace{1cm} (18)$$

$$S_t = o_t \otimes \tanh(C_t),$$

where $U^f, U^i, U^c$ is the input weight, $W^f, W^i, W^c$ is cycle weight, $b^f, b^i, b^c$ is the deviation, and $S_t$ will be the input for the next hidden layer.

Step 4: the hidden states of the forward layer and the backward layer can be represented as $S_t$ and $\overline{S_t}$, respectively, where “$-$” and “$-$” represent the forward and reverse processes, respectively. The fully hidden state of BLSTM in

Figure 6: The CSI waveform of the four movements. (a) Jumping jack. (b) Press-up. (c) High knee lift. (d) Sit-up. (e) Lunge. (f) Squat-up. 

Mobile Information Systems
the time step $t$, i.e., $S_t$, is the series of the hidden state of the forward layer and the backward layer:

$$S_t = \overrightarrow{S}_t \oplus \overleftarrow{S}_t$$ (19)

Step 5: 6000 eigenvalues are entered into the ABLSTM network. The features learned in BLSTM are taken as input

Step 6: there are 256 neurons in the input layer, and the discard rate is set as 20%, that is, a certain proportion of nodes are randomly selected to be discarded in each round of weight update

Step 7: use activation functions at the output layer. After the input layer, there is a BLSTM layer with 128 neurons, a fully connected dense layer with 64 neurons, a fully connected dense layer with 32 neurons, and a one-dimensional output layer. The leakage rate between two hidden layers and between hidden layers and output layers is 10%.

Step 8: a penalty item is added to the activation parameter or layer activation value to apply the rule item to the weights and outputs and activate the model. Using the mean square error as the loss function, the model achieves the best effect by adjusting the learning rate of the important parameter of stochastic gradient descent

Step 9: the softmax layer is used to identify different activities and train the action recognition model.

4. Experimental Results and Analysis

4.1. Experimental Device. The experiment was conducted with two TP-Link WDR4310 routers loaded with Atheros 9580 network card chips and two laptop computers. Connect two laptops to two routers, respectively, one as the transmitter and the other as the receiver. The transmitter and receiver form a local area network for data transmission and reception, which is transmitted to the processing computer through the TCP/IP protocol. WiFi packets are broadcast at the rate of 1000 packets per second, and CSI in WiFi signal is extracted by the CSI tool. Three transmitting antennas and two receiving antennas are set up in the experiment, and a total of six communication links are set up. The operation information of personnel is collected at 5.7 GHz center frequency.

Experiments are set in empty classrooms and conference rooms to collect experimental data. More computers and desks are placed in the conference room, which is strongly affected by multipath interference. Compared with the conference room, the empty classroom has less interference. Under the line-of-sight condition, the distance between the transmitting end and the receiving end is set as 2.0 m, and the height is 0.5 m (this setting is to obtain a large signal-to-noise ratio and facilitate the perception of human movement). The size of the empty classroom is $7 \text{m} \times 5 \text{m}$, and the size of the laboratory is $10 \text{m} \times 5 \text{m}$. The perceived real experimental scene diagram is shown in Figure 8. The plane structure is shown in Figure 9.

The experimenter was arranged to carry out indoor sports containing 6 movements in two scene areas, respectively, and the computer collected CSI signals of each movement in real time and saved them. The Wi-exercise preprocesses the signals collected in real time and outputs the recognition results. In this paper, four groups of comparative experiments were set up to explore the influence of user diversity, movement direction, visual distance, and influence on human motion recognition under different experimental scenarios. The experimenter was given a five-second period to begin an exercise and remained stationary at the beginning and end of that period. The exercises were squats, lunges, push-ups, sit-ups, leg lifts, and jumping jacks, with 10 experiments for each group and 50 experiments for each movement. In the test of each human movement, a movement was repeated 500 times. In order to evaluate the robustness of the Wi-exercise, error rate and recognition accuracy were used as measures to evaluate the accuracy of action recognition.

4.2. Experimental Verification

4.2.1. Effect of Different Distances on Action Recognition in Different Situations. In order to study the influence of the distance between humans and link on action recognition rate, the researchers performed squats at a range of five distances. The intervals of the experiment were set between 0 cm and 200 cm, 0 cm, 50 cm, 100 cm, 150 cm, and 200 cm, respectively. Planar structures of different distances are shown in Figure 10(a). The error rates of the experimenter at different distances are shown in Figure 10(b).

It can be seen from Figure 10 that when the distance is greater than or equal to 50 cm, the average recognition rate of the action decreases with the increase of the distance. Since the strength of the signal decreases with the increase
of the interval, the recognition rate decreases. When the distance is \( \leq 50 \) cm, the recognition rate of the average movement increases. When we test at a spacing of 0 cm, the average movement recognition rate decreases. Because the multipath effect is less when the distance is reduced, the disturbance caused by the irregular movement of the limb will also increase. A distance of about 15 cm is ideal for Wi-exercise applications.

In order to study the influence of the distance between the transmitter and the receiver on the movement recognition rate, the experimenter performed squatting in three distance ranges, with an interval of 1 m to 3 m (1 m, 2 m, and 3 m, respectively). Experiment scenes of different distances are shown in Figure 11(a), and planar structures of different distances are shown in Figure 11(b).

We detect motion at different distances between the transmitter and the receiver. It is found that when the distance is greater than 2 m, the recognition rate of the average movement decreases with the increase of the distance. In general, the closer the AP is to the computer, the more accurate the results will be. This is because as the communication distance is shortened, the receiving effect of the received WiFi signal is enhanced, thus providing a more reliable CSI feature extraction to capture the different movements of the human body. Although the multipath effect is small when the distance is reduced, the disturbance caused by the irregular movement of the limb is also increased. As shown in the experimental results, the experimental results are better when the distance between transmitter and receiver is 2 m than other distances. Therefore, the following experimental distances in this paper are set as the distance between the transmitter and receiver is 2 m, as shown in Figure 11(c).

4.2.2. The Influence of User Diversity on Movement Recognition. In order to test the influence of user diversity on the error rate of calculated action recognition, the influence of one experimenter performing different actions and the influence of different experimenters performing the same action were selected to verify.

One of the experimenters was selected to do six movements such as squats, sit-ups, push-ups, lunges, jumping jacks, and leg lifts, and the error rate of each movement was calculated to verify the influence of different movements. The error rate of different actions made by the same person is shown in Figure 12(b). It was observed that the error rate in counting the number of movements varied between movements, ranging from 0% to 4%. Push-ups had a relatively low rate of motion identification errors, while squats, lunges, and sit-ups were higher. The reason is that when a person performs such a movement, part of the limb’s range of motion is relatively large, resulting in a large difference in the signal received.

To study the change in the error rate of counting the number of movements when different subjects performed the same exercise, four researchers performed the same exercise separately. Taking squat as an example, four experimenters did a squat one by one to detect the error rate of movement recognition when different people did the same movement; the results are shown in Figure 12(a). The average error rate is less than 5%. There was no significant difference in the number of counted movements among the
4.2.3. The Influence of Different Motion Directions on Motion Recognition.

Different direction of movement is a prerequisite to exploring the robustness of Wi-exercise. One experimenter was asked to do squat exercises in four directions, respectively, with the door as the orientation reference point, facing the door as the front, and the door as the back. The subjects mainly did squat exercises in four directions, namely, forward, backward, left, and right. Planar structures in different directions are shown in Figure 13(a); the identification error rate graph is shown in Figure 13(b).
As can be seen from Figure 13, the error rate of anterior and posterior motion recognition is higher than that of left and right motion recognition. The error rate of the forward motion calculation was slightly lower than that of the other three directions, and the error rate of backward motion calculation was slightly lower than that of left and right motion. This is because the cross-section of the body is larger when people are moving forward and backward than when they are moving left and right. Therefore, the inconsistency of the direction of movement of the experimenter will cause different signal fluctuations, resulting in different error rates for the number of motions. But even with a complete change of direction, the average error rate remained around 4%. In short, the Wi-exercise recognition of motor movements in different directions was highly robust.

4.3. Performance Analysis

4.3.1. The Influence of Different Experimental Scenes on Movement Recognition. The lab is full of desks and computers, and the multipath environment is complicated. In the two scenarios, the average detection rates of the six movements are as follows: squats, sit-ups, push-ups, lunges, jumping jacks, and high leg lifts were 96.57, 95.41, 92.975, 94.847, 92.73, and 94.015%, respectively. The average detection rates of empty classrooms were 97.16, 96.96, 94.24, 95.41, 93.45, and 95.01%, respectively. The detection rates in the laboratory were 95.98, 93.86, 91.71, 94.34, 92.01, and 93.02%, respectively, as shown in Figure 14.

4.3.2. Comparison of the Recognition Rate of Different Classification Methods. In order to verify the overall performance of Wi-exercise, this paper compares it with KNN [20], SVM [31], and BP, and the recognition rate of the six actions is shown in Table 1. The experimental results show that the recognition accuracy of Wi-exercise method is more accurate. It can be seen that the accuracy of KNN model is less than 80%, and that of the SVM and BP models is more than 80%, but SVM and BP are less accurate for the motion recognition with less obvious activity. The accuracy of ABLSTM is 92%, and the overall performance is better than the other two traditional methods.
KNN needs to store all the training samples and is computationally intensive because for each sample to be classified, the distance from it to all known samples is computed to find its $K$-nearest neighbors. Since SVM solves support vectors with the help of quadratic programming, and solving quadratic programming will involve the computation of a matrix of order $m$ ($m$ is the number of samples), the storage and computation of this matrix will consume a lot of machine memory and computing time when the number of $m$ is large. The BP algorithm is a fast gradient descent algorithm that can easily fall into the problem of local minima because of the large number of parameters in the BP neural network and the need to update a large number of thresholds and weights each time. In this paper, the features are extracted by the BLSTM network structure, then the weights are assigned to the features extracted by BLSTM using the attention mechanism, and finally, the classification results are output by softmax. The recognition accuracy of ABLSTM is more than 92%, and the overall performance is better than the other two traditional methods.

### 4.3.3. Comparison of Recognition Results of Different Models

In order to verify the superiority of the model proposed in this paper, a comparison is made with the models given in other papers. Among the models are the CNN and LSTM models proposed by Wan et al. mentioned above; the DeepConvLSTM model is proposed by Francisco and Daniel [32]; the DeepConvGRU model is proposed by Wang et al.; and the LSTM-CNN model is proposed by Xia et al. The models are compared in terms of recognition rates in two experimental scenarios, respectively. The comparison results are given as shown in Table 2. As can be seen in Table 2, the evaluation metrics of single deep learning methods such as CNN and LSTM are not very high. Among the two methods using a combination of CNN and RNN, the recognition rate of DeepConvLSTM is 3.11% and 3.45% lower than that of this paper for empty classrooms and laboratories, respectively. The recognition rate of DeepConvGRU is 2.63% and 3.31% lower than that of this paper for empty classrooms and laboratories, respectively. The recognition rates of LSTM-CNN in an empty classroom and laboratory are 1.76% and 1.45% lower than the method in this paper,
respectively. The performance of the ABLSTM network architecture is improved compared with other models, and the proposed Wi-exercise is better than other models.

CNN are not entirely suitable for learning time series and thus will require various auxiliary processing and are not always effective. When faced with time series sensitive problems and tasks, RNN (e.g., LSTM) are usually more appropriate. RNN is used for sequential data and has got some memory effect. Although the gradient problem of RNN has been solved to some extent inside LSTM and its variants, it is still not enough. It can handle sequences of 100 magnitudes, while it will still be tricky for sequences of 1000 magnitudes, or longer. DeepConvLSTM ignores the convergence layer in the convolutional network and replaces it with a single LSTM layer to reduce the loss of local information. DeepConvGRU is easier to converge than DeepConvLSTM, but the performance of DeepConvLSTM expression is better for large data sets. CNN-LSTM models need more epochs to learn and reduce overfitting quickly. Compared with other models, Wi-exercise training results will be better.

5. Conclusions

In this paper, an indoor person motion-detection method based on WiFi is proposed, which uses low-cost wireless routing to obtain CSI signals and later uses the CSI signals to extract features and detect the motion changes of indoor people. To achieve this goal, the raw CSI data are processed by the db3 wavelet and sym8 wavelet, PCA is used to track the whole CSI time series, the eigenvalues are extracted after data noise reduction, and the impulse coefficients and discrete coefficients are selected as the features reflecting the CSI waveform transformation. The 6000 feature values are input to the ABLSTM network for training, using the mean square error as the loss function, and the softmax layer is used to identify different activities and train the action recognition model to finally achieve the detection of indoor human movement actions. The average detection rate of Wi-exercise is 93.48% and 95.37% in two different experimental environments, laboratory and open classroom, with high recognition accuracy, as derived from the experimental results. This method has high recognition accuracy, as derived from the experimental environments, laboratory and open classroom, and the proposed Wi-exercise is 93.48% and 95.37% in two different real environments.

Data Availability

Data are available on request to the authors. The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work is supported by the Science and Technology Innovation Funding Project of Gansu Province (4066-004) and Lanzhou Virtual Reality Joint Laboratory Project (4066-005).

References

[1] Y. Lu, S. Lu, and X. Wang, “Research review of human behavior sensing technology based on WiFi signal,” Chinese Journal of Computer Science, vol. 42, no. 2, pp. 3–23, 2019, (in Chinese).
[2] W. Zhang, S. Zhou, L. Yang, L. Ou, and Z. Xiao, “WiFiMap+: high-level indoor semantic inference with WiFi human activity and environment,” IEEE Transactions on Vehicular Technology, vol. PP (99), pp. 1–1, 2019.
[3] Z. Hao, Y. Duan, X. Dang, Y. Liu, and D. Zhang, “Wi-SL: contactless fine-grained gesture recognition uses channel state information,” Sensors, vol. 20, no. 14, p. 4025, 2020.
[4] M. Ali, T. El Batt, and M. Youssef, “SenseIo: Realistic Ubiquitous Indoor Outdoor Detection System Using Smartphones,” IEEE Sensors Journal, vol. 18, no. 9, pp. 3684–3693, 2018.
[5] J. Shotton, T. Sharp, A. Kipman et al., “Real-time human pose recognition in parts from single depth images,” Communications of the ACM, vol. 56, no. 1, pp. 116–124, 2013.
[6] C. Chen, R. Jafari, and N. Kehtarnavaz, “Improving human action recognition using fusion of depth camera and inertial sensors,” IEEE Transactions on Human-Machine Systems, vol. 45, no. 1, pp. 51–61, 2015.
[7] L. Zhang, K. Liu, Y. Jiang et al., “Montage: combine frames with movement continuity for real time multi-user tracking,” IEEE Transactions on Mobile Computing, vol. 16, no. 4, pp. 1019–1031, 2017.
[8] T. Li, C. An, Z. Tian, A. T. Campbell, and X. Zhou, “Human Sensing Using Visible Light Communication,” in International Conference on Mobile Computing & Networking, Paris France, 2015.
[9] Y. Zhang, D. Li, and Y. Wang, “An indoor passive positioning method using CSI fingerprint based on Adaboost,” IEEE Sensors Journal, vol. 19, no. 14, pp. 5792–5800, 2019.
[10] H. Abdelnasser, K. A. Harras, and M. Youssef, “A ubiquitous WiFi-based fine-grained gesture recognition system,” IEEE Transactions on Mobile Computing, vol. 18, no. 11, pp. 2474–2487, 2019.
[11] B. Sheng, Y. Fang, F. Xiao, and L. Sun, “An accurate device-free action recognition system using two-stream network,” IEEE Transactions on Vehicular Technology, vol. 99, pp. 7930–7939, 2020.
[12] K. Qian, C. Wu, Z. Yang, Y. Liu, F. He, and T. Xing, “Enabling contactless detection of moving humans with dynamic speeds using CSI,” ACM Transactions on Embedded Computing Systems, vol. 17, no. 2, pp. 1–18, 2018.
[13] S. Wan, L. Qi, X. Xu, C. Tong, and Z. Gu, “Deep learning models for real-time human activity recognition with smartphones,” Mobile Networks & Applications, vol. 25, no. 2, pp. 743–755, 2020.
[14] J. Maitre, K. Bouchard, and S. Gaboury, “Alternative deep learning architectures for feature-level fusion in human activity recognition,” Mobile Networks and Applications, vol. 26, no. 5, pp. 2076–2086, 2021.

[15] T. Mahmud, S. S. Akash, S. A. Fattah, W. P. Zhu, and M. O. Ahmad, “Human activity recognition from multi-modal wearable sensor data using deep multi-stage LSTM architecture based on temporal feature aggregation,” in 2020 IEEE 63rd International Midwest Symposium on Circuits and Systems (MWSCAS), pp. 249–252, Springfield, MA, USA, 2020.

[16] L. Chen, X. Liu, L. Peng, and M. Wu, “Deep learning based multimodal complex human activity recognition using wearable devices,” Applied Intelligence, vol. 51, no. 6, pp. 4029–4042, 2021.

[17] K. Xia, J. Huang, and H. Wang, “LSTM-CNN architecture for human activity recognition,” IEEE Access, vol. 8, pp. 56855–56866, 2020.

[18] W. Zhenyu and Z. Lei, “Sensor motion recognition based on deep convolution and gated recurrent neural networks,” Journal of Electronic Measurement and Instrumentation, vol. 32, no. 1, p. 9, 2020.

[19] M. Youssef, “Challenges: device-free passive localization for wireless environments,” in Proceedings of the 13th annual ACM international conference on Mobile computing and networking, Montréal Quebec Canada, 2007.

[20] Z. Gao, Y. Gao, S. Wang, D. Li, and Y. Xu, “CRISLoc: reconstructable CSI fingerprinting for indoor smartphone localization,” IEEE Internet of Things Journal, vol. 8, pp. 3422–3437, 2020.

[21] X. Yang, X. Yu, L. Xie, M. Zhou, and Q. Jiang, “Sleep apnea monitoring system based on Channel State Information,” in International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness, Springer, Cham, 2021.

[22] S. Yousefi, H. Narui, S. Dayal, S. Ermon, and S. Valaee, “A survey on behavior recognition using WiFi channel state information,” IEEE Communications Magazine, vol. 55, no. 10, pp. 98–104, 2017.

[23] M. Jaderberg, K. Simonyan, A. Zisserman et al., “Spatial transformer networks computer,” Science, vol. 4, pp. 2017–2025, 2015.

[24] C. Han, K. Wu, Y. Wang, and L. M. Ni, “WiFi Fall: device-free fall detection by wireless networks,” in IEEE INFOCOM 2014-IEEE Conference on Computer Communications, Toronto, ON, Canada, 2014.

[25] C. G. Keller and D. M. Gavrila, “Will the pedestrian cross? A study on pedestrian path prediction,” IEEE Transactions on Intelligent Transportation Systems, vol. 15, no. 2, pp. 494–506, 2014.

[26] T. Xin, B. Guo, Z. Wang, P. Wang, and Z. Yu, “FreeSense: human-behavior understanding using Wi-Fi signals,” Journal of Ambient Intelligence and Humanized Computing, vol. 9, no. 5, pp. 1611–1622, 2018.

[27] J. Lv, D. Man, W. Yang, L. Gong, X. du, and M. Yu, “Robust device-free intrusion detection using physical layer information of WiFi signals,” Applied Sciences, vol. 9, no. 1, p. 175, 2019.

[28] X. Wang, L. Gao, P. Wang, X. Sun, and X. Liu, “Two-stream 3-D convNet fusion for action recognition in videos with arbitrary size and length,” IEEE Transactions on Multimedia, vol. 20, no. 3, pp. 634–644, 2018.