

Received 30 October 2022, accepted 11 November 2022, date of publication 14 November 2022,
date of current version 18 November 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3222381

RESEARCH ARTICLE

A Novel WiFi-Based Personnel Behavior Sensing With a Deep Learning Method

LEI ZHANG¹, YUE ZHANG², RONG BAO¹, YONGHONG ZHU¹, AND XINGUO SHI³

¹School of Information Engineering (School of Big Data), Xuzhou University of Technology, Xuzhou 221000, China

²Hydrology Department (Information Center), The Yi-Shu-Si River Basin Administration, H.R.C., Xuzhou 221000, China

³Information Center, Shandong Energy Zibo Mining Group Company Ltd., Zibo 225100, China

Corresponding author: Rong Bao (baorong_xzit@yeah.net)

This work was supported in part by the Fundamental Science (Natural Science) Research Projects of Jiangsu Higher Education Institutions under Grant 21KJB510025, in part by the Jiangsu Industry University Research Cooperation Project under Grant BY2021160, in part by the Xuzhou Science and Technology Plan Project under Grant KC19208, and in part by the Zibo Mining Group Smart Mine Key Technology Research and Development Open Fund Grant Project under Grant 2019LH05.

ABSTRACT People's activity recognition technology has a large number of applications in indoor monitoring, smart healthcare, and smart homes. Traditional methods require hardware-mounted or wearable sensors, which bring additional costs and impose many limitations on usage, while WiFi devices have the advantages of low cost, wide coverage, and wall penetration, which can compensate for the limitations of traditional means. A WiFi-based behavior sensing system is proposed in this study. As the characteristics of personnel activities would cause path changes, the path decomposition algorithm is designed to use the path information as the feature of the action and to fully exploit the feature information to improve the accuracy of recognition, the system proposes the BI-AT-GRU network in which bi-directional and attention mechanism are added to the gated loop unit to achieve action recognition. The system is evaluated in different environments with different levels of multipath reflections and compared with other systems. The experimental results show that the recognition accuracy of this system in the two scenes is 97.4% and 93.3%, respectively.

INDEX TERMS Personnel behavior identification, channel state information, phase decomposition, deep learning.

I. INTRODUCTION

With the development of Internet of Things (IoT) technology, intelligent applications appear in many life scenarios. Personal activity recognition as an important application of IoT [1], [2], [3] plays an important role in some scenarios such as smart homes [4], [5], intelligent medical care [6], [7], and indoor monitoring [8] where the behavior of the target needs to be effectively monitored to ensure safety.

Computer vision [9] and wearable devices [10] are the two main traditional techniques for personal activity recognition. However, wearable device-based recognition methods require the user to carry the device with him/her to collect information on human activity characteristics, which is not convenient enough.

Computer vision-based activity recognition methods belong to passive detection, which can be divided into two

categories by its sensors. One is a video recognition method aimed at extracting and classifying spatial-temporal features, including three branches of the dual-stream method [11], C3D [12], and CNN-LSTM [13]; the other is a pose estimation method based on human skeletal point information, which requires the use of the Kinect camera to extract skeletal point information [14]. However, the camera can only work properly in scenes with sufficient light. If there are obstacles in the environment to occlude or the person is in the dead-end position of the camera, the camera would not be able to capture the person's action information. In addition, in the bathroom, bedroom, and other indoor environments, the use of cameras for activity recognition would bring the problem of privacy intrusion.

To overcome the shortcomings of the above sensing methods, WiFi signals due to their wide coverage and strong penetration characteristics which have been widely used in the commercial field, have become a hot spot for human activity sensing. As shown in Fig.1, the original indoor environment

The associate editor coordinating the review of this manuscript and approving it for publication was Chan Hwang See.

only has line-of-sight transmission paths and reflection paths of indoor objects. When human activities occur, additional reflection paths caused by human actions are introduced. Since different human activities produce different multipath changes on the WiFi signal, and these multipath changes can be used to distinguish the corresponding activities.

Earlier WiFi-based activity identification methods utilized the Received Signal Strength Indicator(RSSI) as the metric [15], which can be easily obtained from commercial WiFi. In general, a large RSSI variance indicates the possibility of personal activity in the indoor environment, and vice versa, indicating no personal activity. However, RSSI experiences severe degradation in complex environments due to time delay, signal reflection, and diffraction [16], making RSSI-based methods more suitable for describing coarse-grained activities.

Since finer-grained Channel State Information(CSI) can be extracted from commercial WiFi, CSI-based sensing studies have emerged [17], [18]. Compared with RSSI, CSI is a finer-grained signal feature that characterizes multipath effects through OFDM subcarrier in the frequency domain [18]. The E-eyes system is proposed which extracts CSI amplitude information to detect walking and in-situ activities [19]. Based on E-eyes, Wi-Chase employs all 30 CSI subcarriers to discriminate three different person activities [20]. To analyze the relationship between personnel activity and CSI variation, CARM [21] proposed the CSI-speed and CSI-activity models to quantify the relationship between the speed of personnel activity and CSI values. The system was designed for the homebound elderly to classify and identify seven daily activities of the elderly. WiAct [22] used Doppler shift as a feature and classified 10 activities using Extreme Learning Machine (ELM). WiMU first determines the number of executed gestures, generates virtual samples for these gestures based on pretrained working samples, and then compares the real samples with the virtual samples to determine the real executed actions [23].

With a large number of applications of deep learning algorithms, many CSI activity recognition systems based on deep neural networks have emerged. Feng designed a three-layer LSTM network and discussed the impact on accuracy [24]. ABLSTM [25] accomplished the recognition of six-person activities by utilizing a bi-directional LSTM network with an attention mechanism. In addition to LSTM, convolutional neural networks also are used for CSI activity recognition. Alazrai [26] proposed a three-layer CNN network to extract the features and achieves an average recognition accuracy of 86.3% on the collected person interaction dataset [26]. Sheng discussed how to learn rich spatial-temporal information from CSI data automatically and proposes a deep learning framework by putting CNN into a multilayer LSTM as a temporal model [27]. WiDar3.0 [28] proposes a method for constructing cross-domain features that implement the training model of one scene to apply to other different scenes.

In this paper, a system used in an indoor environment is designed to identify daily activities. The system uses the

transmission path information as the metric to describe the person's actions based on the influence of the person's activities on the path and achieves the classification of all person's daily behavioral activities by the designed deep learning network.

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

(1) To be able to use path information to describe the actions of people, a path decomposition algorithm is proposed by using CSI information to realize the decomposition of transmission paths.

(2) The relevant characteristics and physical models of CSI are analyzed, and a series of data processing algorithms have been designed for the specific application of CSI.

(3) A residual network architecture with channel attention and spatial attention mechanisms is designed for action recognition, and the feature information consisting of multiple links is converted into an image matrix as the input to the network using the CSI multiple-input multiple-output technique.

(4) Some impactors on the accuracy of action recognition of the system are investigated, such as the size of sliding windows.

(5) The experiments are carried out in two different scenarios and the performance of the system is evaluated.

The rest of this paper is organized as follows, related work in the field of human activity recognition is briefly presented in Section 2. Section 3 briefly introduces some principles of CSI-based personnel activity discrimination. In Section 4, our proposed activity discrimination system is detailed. The implementation of our system and the evaluation results of our system are discussed in Section 5. The conclusion is drawn in the final section.

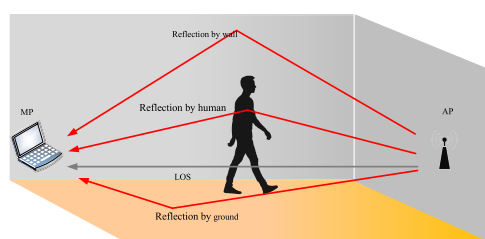


FIGURE 1. WiFi signal propagation in an indoor environment.

II. FUNDAMENTALS

A. CSI

CSI is fine-grained information from the physical layer that can represent the channel properties of the communication link, including information on scattering, environmental fading, and distance fading [29]. In a narrowband flat fading channel, the channel information can be represented in the frequency domain as

$$y = Hx + n \quad (1)$$

TABLE 1. Subcarrier conditions under different bandwidths in 802.11 protocol.

Bandwidth	Grouping number	Number of subcarriers	Subcarrier number
20MHz	1	56	-28,-27,...,-2,-1,1,2,...,27,28
	2	30	-28,-26,...,-2,-1,1,3,5,...,27,28
	4	16	-28,-24,-16,-12,-8,-4,-1,1,5,9,13,17,21,25,28
40MHz	1	114	-58,-57,...,-3,-2,2,3,...,57,58
	2	58	-58,-56,...,-4,-2,2,4,...,56,58
	4	30	-58,-54,...,-6,-2,2,6,...,50,54,58

where y denotes the receiving vector, x denotes the transmitting vector, and n and H denote the noise and channel matrix, respectively.

According to the 802.11n protocol, WiFi can work in 2.4GHz and 5GHz frequency bands, whose bandwidths are 20MHz and 40MHz, respectively. Table 1 shows the distribution of subcarriers under the 802.11n, there are three groups of subcarriers in 20 MHz bandwidth, and the maximum number of subcarriers available to transmit data is 56; in 40 MHz bandwidth, there are also three groups of subcarriers with the maximum number of subcarriers available to transmit data is 114 [30]. Using the CSI-Tool, 30 subcarriers can be acquired, each containing amplitude and phase information, and the acquired subcarriers are represented as follows:

$$H = [H_1, H_2, \dots, H_i, \dots, H_N] \quad (2)$$

Since WiFi uses the multiple-input multiple-output system(MIMO) technique, in which the CSI matrix can be represented as a matrix shown in (3). Where N_t and N_r denote the number of transmitting and receiving antennas, respectively. Each combination of transmitting and receiving antennas can be considered as a separate TX-RX data stream.

$$H_{MIMO} = \begin{bmatrix} H_{11} & H_{12} & \dots & H_{1N_r} \\ H_{21} & H_{22} & \dots & H_{2N_r} \\ \vdots & \vdots & \dots & \vdots \\ H_{N_t1} & H_{N_t2} & \dots & H_{N_tN_r} \end{bmatrix} \quad (3)$$

B. FRESNEL REFLECTION MODEL

The Fresnel Zone is an ellipsoidal region with the transmitting and receiving endpoints as the focal points and the apparent distance path between the two points as the axis. As shown in Fig.2, the Fresnel Zone consists of n ellipses, assuming that the transmitting end is P_1 , the receiving end is P_2 , given the radio wavelength, then for any point Q_n on the n th ellipse, there are the following equations:

$$|P_1Q_n| + |P_2Q_n| - |P_1P_2| = n\lambda/2 \quad (4)$$

where the first Fresnel region is the central region of the geometry, the n th Fresnel region denotes the middle region between the $(n-1)$ th Fresnel boundary and the n th Fresnel boundary, for example, the second Fresnel region denotes the middle range between the first Fresnel boundary and

the second Fresnel boundary. According to the Fresnel Zone principle [31], the important regions for RF transmission are the first 8 to 12 regions, where more than 70% of the energy is transmitted through the first Fresnel region.

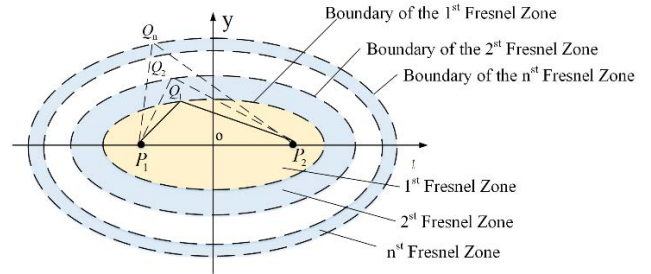


FIGURE 2. Geometric diagram of the Fresnel region.

The paths introduced due to human activity is referred to dynamic paths, and the paths that exist when the scene is unoccupied as static paths. For example, when Q_1 located on the first Fresnel boundary, the difference in length between the reflected path $P_1Q_2P_1$ and the line-of-sight path P_1P_2 is $\lambda/2$, and the phase difference between the static signal and the dynamic signal is π , and when the phase shift π caused by the reflection is added, the phase difference between the two signals is 2π . Thus, the amplitudes of the two signals are superimposed to obtain a larger amplitude. When Q_1 located on the second Fresnel boundary, the phase difference between the two signals is 3π , and the amplitudes are mutually suppressed to obtain a lower amplitude. When moving outward from the first Fresnel region to the n th Fresnel region, the phase difference increases from 2π , 3π , 4π , and so on to $(n+1)\pi$, which results in the superposition of static and dynamic signals, making the CSI signal waveform show periodic fluctuations. Therefore, the use of path information can reflect the activity information of the personnel

III. MOTIVATION AND RECOGNITION SYSTEM DESIGN

In this section, the challenges posed by CSI are investigated to improve the accuracy of action recognition. The architecture of the system is designed with several key constituent blocks, as shown in Fig.3. Each block would be described in detail in the following subsections.

A. OUTLIER REMOVAL

There are always some outliers in the collected CSI data, which are not caused by human activities. To remove these outliers, the Hampel filter [32] is used in this study. Suppose each element in the collected CSI data can be represented as

$$x_k = x_{nom} + e_k \quad (5)$$

where x_{nom} is used to indicate the normal data, e_k represents measurement error, which is the basis for discerning whether the data point is an outlier or not, and the specific identification steps are as follows.

First, all data are arranged in descending order to find the median of the data set, then the median absolute

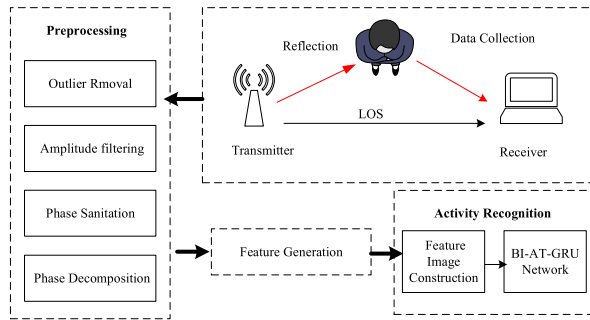


FIGURE 3. System architecture.

deviation (MAD) is obtained according to the median. The formula for calculating the MAD is as follows.

$$MAD = median\{|x_k - x_{median}|\} \quad (6)$$

The discriminant threshold T is shown in (7).

$$T = 1.4826 \frac{|x_k - x_{median}|}{median\{|x_k - x_{median}|\}} \quad (7)$$

The threshold T is generally set empirically, which is set to 3 in this study. The filtering results are shown in Fig.4, where the outlier filtering is performed on the 8th and 10th subcarriers, respectively. It can be seen from the figure that the outlier can be effectively removed using the Hampel filter.

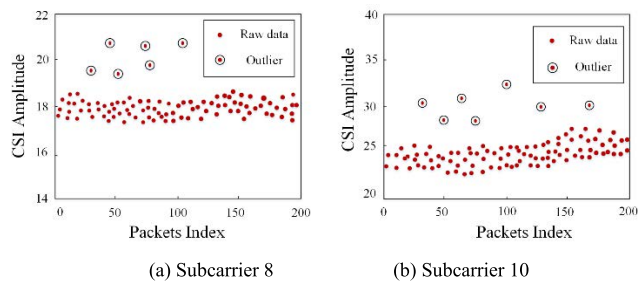


FIGURE 4. An illustration of Hampel filter. (a) Subcarrier 8, (b) Subcarrier 10.

B. AMPLITUDE FILTERING

Although the Hampel filter can effectively remove outliers, the signal after filtering outliers still contains noise generated by the external environment, which makes the obtained CSI data cannot be directly used for action estimation, therefore, considering the frequency of human actions, this study uses the Savitzky-Golay Filter (SGF) for noise removal [33]. The Savitzky-Golay filter retains the main useful components of the signal by fitting the low-frequency band information of the signal, and the specific algorithm is as follows:

Let $x(k)$, $k = -m \dots, 0, \dots m$ be the set of the CSI data, construct a polynomial of order n fit:

$$f(k) = \sum_{i=0}^n b_{ni}k^i + b_{n1}k + b_{n2}k^2 + \dots + b_{nm}k^n \quad (8)$$

The sum of squares of the residuals of the fitted data points and the original data points is:

$$E = \sum_{k=-m}^m (f(k) - x[k])^2 = \sum_{k=-m}^m (\sum_{i=0}^n b_{ni}k^i - x[k])^2 \quad (9)$$

Derivative of the coefficients b_{ni} and making the derivative zero, the m , n , and $x[k]$ are known, which can be used to generate the fitted value of the window center point, and the data can be filtered by moving the sliding window.

C. PHASE SANITIZATION

Due to the clock asynchrony and center frequency deviation between the transmitter and receiver, the collected raw phase information exhibits great randomness, which is not suitable for personnel activity detection. Therefore, in this section, the raw phase data are processed using linear transformation to obtain the CSI available phase information [34], and the acquired CSI phase information is represented as follows:

$$\hat{\phi}_i = \phi_i - 2\pi \frac{k_i}{N} \delta + \beta + Z \quad (10)$$

where $\hat{\phi}_i$ and ϕ_i represent the measured and true CSI phase values of the i th subcarrier, respectively. δ denotes the time difference between transmitter and receiver, β is the phase deviation due to center frequency desynchronization, Z is the Gaussian noise, k_i denotes the index of 30 subcarriers, and N is the size of the Fast Fourier transform. To reduce the effects of β and δ , a linear correction method is applied to remove the phase deviation. The slope and solution of the linear transform equation are denoted by a and b , respectively.

$$a_i = \frac{\hat{\phi}_n - \phi_i}{k_n - k_1} = \frac{\phi_n - \phi_1}{k_n - k_1} - \frac{2\pi}{N} \delta \quad (11)$$

$$b = \frac{1}{n} \sum_{j=1}^n \hat{\phi}_j = \frac{1}{n} \sum_{j=1}^n \phi_j - \frac{2\pi}{nN} \delta \sum_{j=1}^n k_j + \beta \quad (12)$$

Since the frequencies of the CSI subcarriers are symmetric, then $\sum_{j=1}^n k_j = 0$ can be obtained. The noise Z belongs to the measurement error, which can be ignored after several measurements. Finally, the approximate phase $\hat{\phi}_i$ can be obtained by calculating $\hat{\phi}_i - ak_i - b$ which is expressed as follows:

$$\tilde{\phi}_i = \hat{\phi}_i - ak_i - b = \phi_i - \frac{\phi_n - \phi_1}{k_n - k_1} k_i - \frac{1}{N} \sum_{j=1}^n \phi_j \quad (13)$$

The CSI phase after sanitation is shown in Fig. 5(b). Compared with the raw phase which scatters randomly over many different angles, the calibrated phases are less fluctuant and more concentrated.

D. PHASE DECOMPOSITION

From the analysis above, it can be seen that the activities of personnel introduce additional paths, so the information on paths can be used to classify personnel activities. From the definition of CSI, it is clear that the CSI value of each

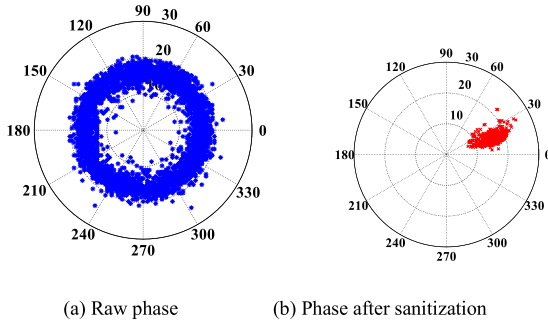


FIGURE 5. Phase sanitization.

subcarrier is the result of the superposition of multiple paths, and thus the CSI of subcarrier k can be written as:

$$H_k = \sum_{n=1}^L \alpha_l e^{-j2\pi(f_0+k\Delta f)\tau_n} \quad (14)$$

where L is the number of paths, α_l and τ_l denote the signal amplitude and the TOF of path n , respectively. f_0 is the central frequency, Δf is the frequency interval of adjacent subcarriers.

For the path l , H_k^l is rewritten as follows:

$$H_k^l = S_k^l \Delta_k^l \quad (15)$$

where $S_k^l = \alpha_l e^{-j2\pi f_0 \tau_l}$ and $\Delta_k^l = e^{-jk\Delta f \tau_l}$.

With the CSITool, the index of subcarriers ranging from -58 to 58 , in which the interval of subcarrier index is 4 . Supposing that there are five paths, a Hankel matrix can be built:

$$\mathbf{X} = \begin{bmatrix} H_{-58} & H_{-54} & H_{-50} & H_{-46} & H_{-42} \\ H_{-54} & H_{-50} & H_{-46} & H_{-42} & H_{-38} \\ H_{-50} & H_{-46} & H_{-42} & H_{-38} & H_{-34} \\ H_{-46} & H_{-42} & H_{-38} & H_{-34} & H_{-30} \\ H_{-42} & H_{-38} & H_{-34} & H_{-30} & H_{-26} \end{bmatrix} \quad (16)$$

Then rewrite \mathbf{X} as:

$$\mathbf{X} = \mathbf{\Delta S \Delta}^T \quad (17)$$

where

$$\mathbf{\Delta} = \begin{bmatrix} \Delta_{-29}^1 & \Delta_{-29}^2 & \Delta_{-29}^3 & \Delta_{-29}^4 & \Delta_{-29}^5 \\ \Delta_{-25}^1 & \Delta_{-25}^2 & \Delta_{-25}^3 & \Delta_{-25}^4 & \Delta_{-25}^5 \\ \Delta_{-21}^1 & \Delta_{-21}^2 & \Delta_{-21}^3 & \Delta_{-21}^4 & \Delta_{-21}^5 \\ \Delta_{-17}^1 & \Delta_{-17}^2 & \Delta_{-17}^3 & \Delta_{-17}^4 & \Delta_{-17}^5 \\ \Delta_{-13}^1 & \Delta_{-13}^2 & \Delta_{-13}^3 & \Delta_{-13}^4 & \Delta_{-13}^5 \end{bmatrix} \quad (18)$$

And

$$\mathbf{S} = \text{diag}(S_0^1, S_0^2, S_0^3, S_0^4, S_0^5) \quad (19)$$

Define $\mathbf{\Delta} = \mathbf{V S}'$

where \mathbf{v} can be transformed into a Vandermonde matrix as follows:

$$\mathbf{V} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ \Delta_4^1 & \Delta_4^2 & \Delta_4^3 & \Delta_4^4 & \Delta_4^5 \\ \Delta_8^1 & \Delta_8^2 & \Delta_8^3 & \Delta_8^4 & \Delta_8^5 \\ \Delta_{12}^1 & \Delta_{12}^2 & \Delta_{12}^3 & \Delta_{12}^4 & \Delta_{12}^5 \\ \Delta_{16}^1 & \Delta_{16}^2 & \Delta_{16}^3 & \Delta_{16}^4 & \Delta_{16}^5 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ (\Delta_4^1)^1 & (\Delta_4^2)^1 & (\Delta_4^3)^1 & (\Delta_4^4)^1 & (\Delta_4^5)^1 \\ (\Delta_4^1)^2 & (\Delta_4^2)^2 & (\Delta_4^3)^2 & (\Delta_4^4)^2 & (\Delta_4^5)^2 \\ (\Delta_4^1)^3 & (\Delta_4^2)^3 & (\Delta_4^3)^3 & (\Delta_4^4)^3 & (\Delta_4^5)^3 \\ (\Delta_4^1)^4 & (\Delta_4^2)^4 & (\Delta_4^3)^4 & (\Delta_4^4)^4 & (\Delta_4^5)^4 \end{bmatrix} \quad (20)$$

And

$$\mathbf{S}' = \text{diag}(S_{-29}^1, S_{-29}^2, S_{-29}^3, S_{-29}^4, S_{-29}^5) \quad (21)$$

Therefore, $\mathbf{x} = (\mathbf{v s})\mathbf{s}(\mathbf{v s})^T$, the problem arises to find a Vandermonde decomposition of \mathbf{X} . With the decomposition algorithm $\left[(\Delta_4^1)^1 (\Delta_4^2)^1 (\Delta_4^3)^1 (\Delta_4^4)^1 (\Delta_4^5)^1 \right]$ can be got.

According to phase decomposition, the phase can be extracted from each path. On the other hand, it cannot be used for the calculation of TOF as it contains some unknown phase offset caused by ADC.

E. FEATURE IMAGE CONSTRUCTION

Since the CSI signal uses MIMO technology, the total number of links is the product of the number of transmit antennas and the number of receive antennas. For a system with 3 receive antennas and 3 transmit antennas, there are generally a minimum of 3 links present. Each link is a CSI matrix that can be considered as a channel of images so that the input data is a multi-channel image matrix fused with the path information and subcarrier data.

A residual network architecture with channel attention and spatial attention mechanisms is designed for feature extraction. In this network, the features extracted by the CNN module can effectively represent the relationship between adjacent subcarriers and suppress invalid subcarriers. While spatial and temporal attention are able to locate and enhance the key links and subcarrier actions, due to the different links' pairs being affected by the actions to different degrees. It is also greatly helpful to further reduce the overhead of the overall parameters by fusing the output of the spatial and channel attention modules by element multiplication. The output of the flatten layer is fused with a bidirectional attention mechanism gated recurrent unit network.

As different actions are both independent and correlated in space-time, to be able to mine the spatial-temporal information of action features, this study uses a bidirectional attention mechanism gated recurrent unit network (BI-AT-GRU) for action classification. This network connects the forward and backward learning vectors as the final structure, which can effectively avoid the problem of missing spatial-temporal

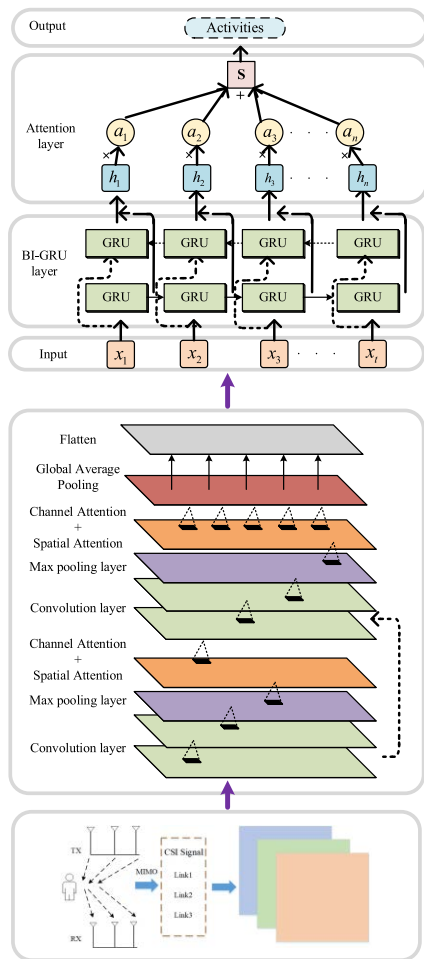


FIGURE 6. Flow chart of activity recognition based on BI-AT-GRU.

information before and after, and at the same time, the attention mechanism that gives different weights to different information pays more attention to the important features of the data and further improves the classification accuracy. The general design of the model is shown in Fig.6.

IV. EXPERIMENTAL ANALYSIS

A. THE EXPERIMENTAL ENVIRONMENT

As shown in Fig.7, two laboratories are chosen as the experimental scene. The environment of Lab1 is simple, which has only a few tables and chairs and electronic devices; while Lab2 occupied with tables, chairs, cabinets, and other furniture is crowded, and there are more multipath reflections.

The hardware equipment of the system consists of a Mini-PC and a wireless access point(D-Link DIR-859), which named MP and AP, respectively. The AP with three antennas is used as the data transmitter, which supports two center frequencies, 2.4 GHz and 5 GHz. The MP is installed with Ubuntu 14.04 and comes with a modified Intel 5300 NIC driver. The AP and MP are shown in Fig.8. As the WiFi network is working in 802.11n protocol, the 5GHz band is used in the experiment to avoid interference from other devices working in the 2.4GHz band in the experimental scenario.

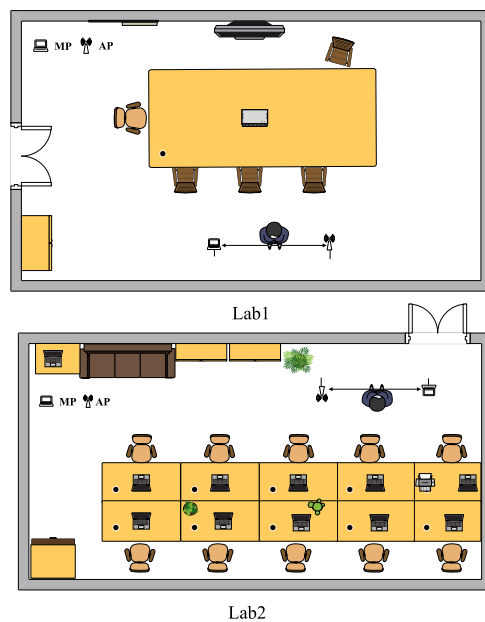


FIGURE 7. Photographs and layouts of the experiment.

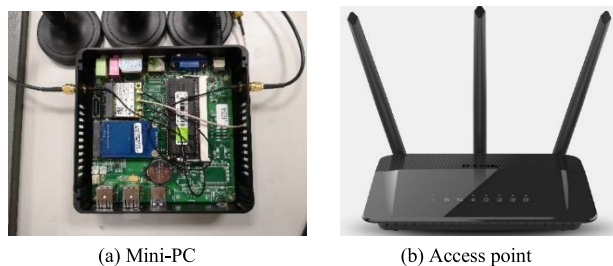


FIGURE 8. Hardware equipment of the system.

Five volunteers with different physical conditions were engaged in data collection for one month in two different test scenarios. The volunteers completed different daily activities according to the experimental setup and the data collected included both upper-limb activities and whole-body activities. In order to ensure that 5 volunteers participated in the data acquisition of different actions in two scenes, before each day’s acquisition, 5 volunteers were divided into two groups according to the pre-designed acquisition schedule in groups of 2-3 people and conducted data acquisition in two scenes simultaneously. In the data collection process, the center point of the distance from the receiving and transmitting terminal was taken as the action completion area to ensure the consistency of daily collection. After the first volunteer completed the action that needed to be collected, the next volunteer entered the action collection area and completed the same action. Since each action lasts for a short period of time, 3500 sets of data were collected for each action in each scene in one month of data collection. 2500 sets of data were randomly selected by us as the training set, and the remaining 1000 sets of data were used as the test set, and the model was



FIGURE 9. Activity data acquisition.

TABLE 2. Summary of activity dataset.

Activity	Expression
Walk	Walk
Put off hat	Put off
Throw	Throw
Sit	Sit
Drink	Drink
Wave	Wave
Run	Run
Sleep	Sleep

trained by a high-performance computer with 2080ti graphics card Partial of activity data acquisition is shown in Fig9.

The final dataset obtained is shown in Table 2, which describes the specific actions and their representation in the text.

B. PERFORMANCE OF BI-AT-GRU NETWORK

Fig.10 shows the curve of the accuracy and loss function of the BI-AT-GRU network. Fig.10(a) shows that the accuracy of the validation set increases as the number of iterations increases, and it gradually stabilizes at 97.3% after the number of iterations reaches 800, which means that the algorithm is able to train the intrinsic features well. Fig.10 shows that as the number of iterations increases, the loss function gradually decreases and eventually converges to 0, which means that the training is effective.

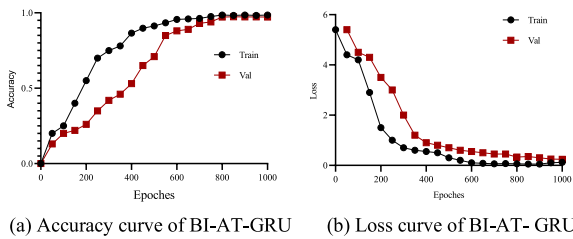


FIGURE 10. BI-AT-GRU training curve.

Adam, SGD, Ada Delta and RMS Pro are selected as the optimizers of our proposed network respectively, and the accuracies with different learning rates are presented in Table 3. Compared with other optimizers, the Adam achieves the highest accuracy with the learning rate of 0.01. And when the learning rate is greater than 0.01, the accuracy of the model starts to decrease. Therefore, in the case of network applications, the learning rate should be less than 0.01.

TABLE 3. Classification accuracy with different optimizers and learning rates.

Learning rate	Ada Delta	SGD	RMS Prop	Adam
0.0001	93.34%	94.75%	97.33%	97.01%
0.001	94.43%	93.27%	95.64%	97.77%
0.01	96.45%	96.18%	96.41%	98.28%
0.05	90.97%	92.64%	90.72%	92.89%
0.1	88.25%	89.76%	91.45%	91.99%

C. OVERALL PERFORMANCE

Our system is evaluated in Lab1 and Lab2 scenes, respectively, and the test results are shown in Fig.11, which shows that the recognition accuracy of all activities in both scenes is above 88%. Lab1 has higher accuracy than Lab2 in classifying human activities, which is mainly due to the fact that Lab2 has a lot of furniture, which generates a lot of extra reflective signals that interfere with the reflective signals from the human body. In contrast, the CSI signals collected in Lab1 are only human reflection and wall reflection, so the accuracy of action recognition is higher.

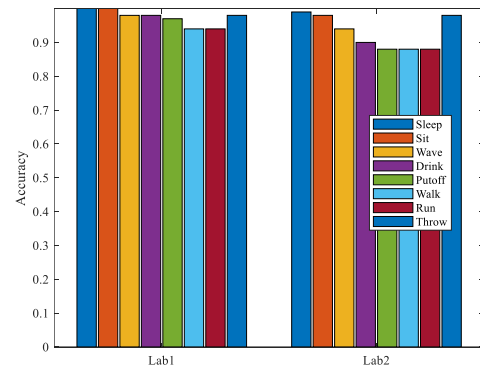


FIGURE 11. The performance of different environments.

D. ACTIVITY CLASSIFICATION RESULTS

Fig.12 shows the confusion matrix obtained after discriminating eight personnel activities. Each row of the matrix represents the true label value, i.e., the activity category, and each column represents the algorithm predicted value. Among them, the actions of sleeping and sitting achieve the highest accuracy because they are relatively simple and coarse-grained actions, which are easier to discriminate compared to other activities. In addition, the three actions of waving, drinking, throwing and removing a hat are all upper limb actions that are easily confused with each other. The remaining are walking and running actions, which are particularly easy to confuse because they perform the same model and sometimes have little difference in speed, and the recognition accuracy of these two activities is lower compared to other activities.

In addition, 8 similar activities of 5 volunteers are also tested, and the results are shown in Fig.13. The average recognition accuracy of the 5 volunteers' actions is 93.4%,

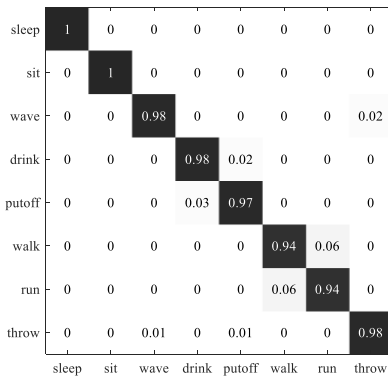


FIGURE 12. Confusion matrix of activity recognition.

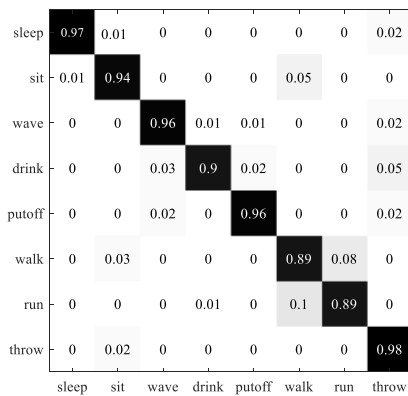


FIGURE 13. Confusion matrix for the classification of similar activities of five people.

which is close to the action recognition accuracy of a single volunteer

To compare the recognition performance of different algorithms, the proposed BI-AT-GRU is compared with LSTM and BI-LSTM where Precision, Recall and F1-score are used as the metrics. The results are shown in Fig.14, from which it can be seen that BI-AT-GRU has the best performance, followed by BI-LSTM, which is mainly due to the fact that BI-AT-GRU turns the three links into a three-channel image matrix when processing the input data, making full use of the information between the links; meanwhile, the residual network with attention can increase the network depth while ensuring the accuracy of feature extraction. In contrast, LSTM and BI-LSTM can only integrate the link information directly into a matrix, which is incomplete for the use of link information. The specific comparison results are shown in Table 4.

TABLE 4. Comparison of different algorithms.

Algorithm	Accuracy	Precision	Recall	F1-Score
LSTM	0.806	0.814	0.815	0.802
BI-LSTM	0.819	0.803	0.818	0.794
BI-AT-GRU	0.973	0.970	0.978	0.971

Our system is also compared with HAR system [35] and WiWave system [36] in two scenarios respectively.

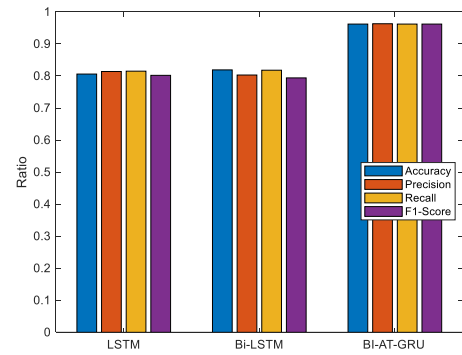


FIGURE 14. Comparison of performance with different algorithms.

The experimental results are shown in Fig.15. In HAR system, the use of KNN as the action recognition model cannot represent the complex relationship between the actions of the features, so the accuracy of recognition in both environments is lower than that of the WiWave system and our system. While WiWave although CNN introduced discrete wavelet transform into the convolutional architectures, which can combine the good time-frequency local characteristics of the wavelet transform with the self-learning ability of the neural network, however, it does not take into account the different characteristics of different actions in continuous time, that results in a lower recognition accuracy than our system.

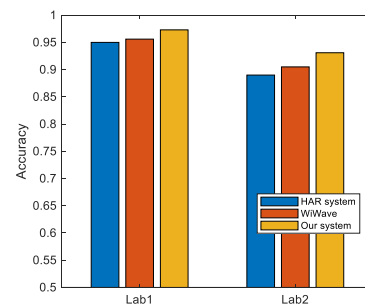


FIGURE 15. Compared with other systems.

E. EFFECT OF SLIDING WINDOW SIZE

Since the data size also affects the recognition accuracy, the effect of different sizes of sliding windows is tested in two different environments, and the results are shown in Fig.16. From the results, it can be seen that the recognition accuracy of person activity in both environments increases with the increase of window size, whose threshold of sizes is 20. This is mainly due to the increase in the sliding window length, the obtained feature samples would also increase, which in turn improves the classification accuracy. However, when the window increases to this length threshold, the recognition accuracy cannot be effectively improved even if the size continues to increase.

F. EFFECT OF DIFFERENT DISTANCE

In addition, the effect of different distances between the transceiver devices on the recognition is also tested in this

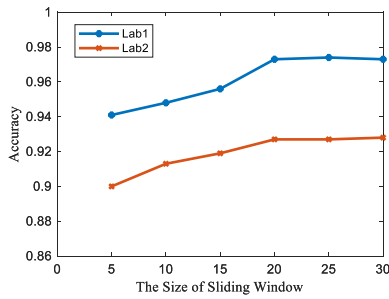


FIGURE 16. The impact of different sizes of sliding window.

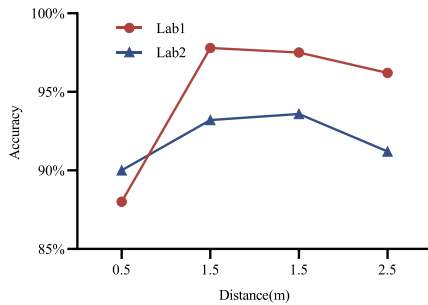


FIGURE 17. Recognition accuracy with different distance.

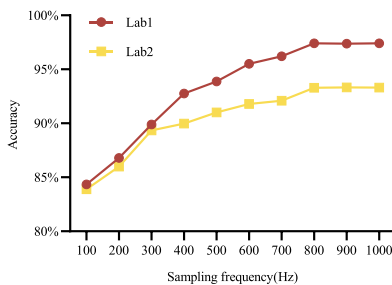


FIGURE 18. Recognition accuracy with different sampling frequencies.

study. The tests are conducted at distances of 0.5m, 1m, 1.5m, and 2m, respectively. Fig.17 shows the action recognition accuracy of different distances under two experimental environments. It can be seen that the accuracy is better than the others at the distance of 1m and 1.5. When the distance is 0.5m, some of the actions with less interference to the signal transmission path would cause the original multi-communication link to degrade into a single communication link due to the closer distance, thus making the obtained feature information less and leading to a decrease in recognition accuracy. And when the distance increases, additional transmission paths would be introduced, leading to the introduction of more noise, which in turn affects the recognition accuracy.

G. EFFECT OF DIFFERENT SAMPLING FREQUENCIES

In order to analyze the effect of different sampling frequencies on the accuracy of action recognition. Different sampling frequency frequencies are tested in two scenes respectively, and the test results are shown in Fig.18, from which it can be seen that as the sampling rate increases, the recognition

accuracy of the system increases. This is mainly due to the fact that as the sampling frequency increases, more CSI data are collected per unit of time, and more CSI samples are used in a fixed time window to more accurately describe the impact of the person's activities on the transmission path. However, it can also be seen from the figure that the accuracy hardly changes when the sampling rate increases to 800 Hz, mainly because there is an upper limit to the impact of each action on the environment, and no more details are obtained at higher sampling rates.

V. CONCLUSION

In this study, a WiFi-based indoor people activity recognition system is presented. A new phase decomposition method is proposed to achieve the features of people activity. Through analyzing the relevant characteristics and physical model of CSI, a series of algorithms is designed to process the data, and the BI-AT-GRU which is a residual network architecture with attention mechanisms is used for action recognition. The performance in two different laboratories is validated. The results reveal a mean accuracy is 97.4% and 93.3%, respectively. We also compare our system with other systems and analyze the impact factors on the recognition accuracy of our system. The results demonstrate that our system can more effectively distinguish personnel activities than the others.

Since the model of the Fresnel zone does not fully explain the complex activity changes of personnel. Therefore, our future work will focus on analyzing the characteristics of different categories of personnel activities and building a unified personnel activity model to improve the recognition accuracy and generalization ability.

REFERENCES

- [1] L. Chen, L. Tian, Z. Xu, and Z. Chen, "A survey of WiFi sensing techniques with channel state information," *ZTE Commun.*, vol. 18, no. 3, pp. 57–63, 2020.
- [2] S. Ren, H. Wang, L. Gong, C. Xiang, X. Wu, and Y. Du, "Intelligent contactless gesture recognition using WLAN physical layer information," *IEEE Access*, vol. 7, pp. 92758–92767, 2019.
- [3] A. S. A. Varshini, G. Bhavani, and R. Thilagavathy, "Real-time hand gesture recognition for robotic arm and home automation," in *Proc. Int. Symp. Electr., Electron. Inf. Eng.*, Feb. 2021, pp. 218–223.
- [4] Y. Ma, G. Zhou, and S. Wang, "WiFi sensing with channel state information: A survey," *ACM Comput. Surv.*, vol. 53, no. 3, pp. 1–36, 2019.
- [5] X. Chu, J. Liu, and S. Shimamoto, "A sensor-based hand gesture recognition system for Japanese sign language," in *Proc. IEEE 3rd Global Conf. Life Sci. Technol. (LifeTech)*, Mar. 2021, pp. 311–312.
- [6] B. Yu, Y. Wang, K. Niu, Y. Zeng, T. Gu, L. Wang, C. Guan, and D. Zhang, "WiFi-sleep: Sleep stage monitoring using commodity Wi-Fi devices," *IEEE Internet Things J.*, vol. 8, no. 18, pp. 13900–13913, Sep. 2021.
- [7] B. B. Atitallah, M. B. Abbasi, R. Barioul, D. Bouchaala, N. Derbel, and O. Kanoun, "Simultaneous pressure sensors monitoring system for hand gestures recognition," in *Proc. IEEE Sensors*, Oct. 2020, pp. 1–4.
- [8] F. S. Botros, A. Phinyomark, and E. J. Scheme, "Electromyography-based gesture recognition: Is it time to change focus from the forearm to the wrist?" *IEEE Trans. Ind. Informat.*, vol. 18, no. 1, pp. 174–184, Jan. 2022.
- [9] I. Rodríguez-Moreno, J. M. Martínez-Otzeta, B. Sierra, I. Rodríguez, and E. Jauregi, "Video activity recognition: State-of-the-art," *Sensors*, vol. 19, no. 14, p. 3160, Jul. 2019.
- [10] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, "Deep learning for sensor-based activity recognition: A survey," *Pattern Recognit. Lett.*, vol. 119, pp. 3–11, Mar. 2019.

- [11] K. Simonyan and A. Zisserman, "Two-stream convolutional networks for action recognition in videos," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 568–576.
- [12] S. Ji, W. Xu, M. Yang, and K. Yu, "3D convolutional neural networks for human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 221–231, Jan. 2013.
- [13] A. Ullah, K. Muhammad, J. D. Ser, S. W. Baik, and V. H. C. de Albuquerque, "Activity recognition using temporal optical flow convolutional features and multilayer LSTM," *IEEE Trans. Ind. Electron.*, vol. 66, no. 12, pp. 9692–9702, Dec. 2019.
- [14] J. M. Paavola, K. E. Oliver, and K. I. Ustinova, "Use of X-box Kinect gaming console for rehabilitation of an individual with traumatic brain injury: A case report," *J. Novel Physiotherapies*, vol. 3, no. 1, pp. 1–6, 2013.
- [15] S. Wang and G. Zhou, "A review on radio based activity recognition," *Digit. Commun. Netw.*, vol. 1, no. 1, pp. 20–29, 2015.
- [16] Z. Yang, Z. Zhou, and Y. Liu, "From RSSI to CSI: Indoor localization via channel response," *ACM Comput. Surv.*, vol. 46, no. 2, pp. 1–32, 2013.
- [17] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "802.11 with multiple antennas for dummies," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 40, no. 1, pp. 19–25, Jan. 2010.
- [18] T. Zhang, T. Song, D. Chen, T. Zhang, and J. Zhuang, "WiGrus: A WiFi-based gesture recognition system using software-defined radio," *IEEE Access*, vol. 7, pp. 31102–31113, 2019.
- [19] Y. Wang, J. Liu, Y. Chen, M. Gruteser, J. Yang, and H. Liu, "E-eyes: Device-free location-oriented activity identification using fine-grained WiFi signatures," in *Proc. 20th Annu. Int. Conf. Mobile Comput. Netw.*, Sep. 2014, pp. 617–628.
- [20] S. Arshad, C. Feng, Y. Liu, Y. Hu, R. Yu, S. Zhou, and H. Li, "Wi-chase: A WiFi based human activity recognition system for sensorless environments," in *Proc. IEEE 18th Int. Symp. World Wireless, Mobile Multimedia Netw. (WoWMoM)*, Jun. 2017, pp. 1–6.
- [21] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Understanding and modeling of WiFi signal based human activity recognition," in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw.*, Sep. 2015, pp. 65–76.
- [22] H. Yan, Y. Zhang, Y. Wang, and K. Xu, "WiAct: A passive WiFi-based human activity recognition system," *IEEE Sensors J.*, vol. 20, no. 1, pp. 296–305, Jan. 2020.
- [23] R. H. Venkatnarayan, G. Page, and M. Shahzad, "Multi-user gesture recognition using WiFi," in *Proc. 16th Annu. Int. Conf. Mobile Syst., Appl., Services*, Jun. 2018, pp. 401–413.
- [24] C. Feng, S. Arshad, R. Yu, and Y. Liu, "Evaluation and improvement of activity detection systems with recurrent neural network," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2018, pp. 1–6.
- [25] Z. Chen, L. Zhang, C. Jiang, Z. Cao, and W. Cui, "WiFi CSI based passive human activity recognition using attention based BLSTM," *IEEE Trans. Mobile Comput.*, vol. 18, no. 11, pp. 2714–2724, Nov. 2018.
- [26] R. Alazrai, M. Hababeh, B. A. Alsaify, M. Z. Ali, and M. I. Daoud, "An end-to-end deep learning framework for recognizing human-to-human interactions using Wi-Fi signals," *IEEE Access*, vol. 8, pp. 197695–197710, 2020.
- [27] B. Sheng, F. Xiao, L. Sha, and L. Sun, "Deep spatial-temporal model based cross-scene action recognition using commodity WiFi," *IEEE Internet Things J.*, vol. 7, no. 4, pp. 3592–3601, Apr. 2020.
- [28] Y. Zhang, Y. Zheng, K. Qian, G. Zhang, Y. Liu, C. Wu, and Z. Yang, "Widar3.0: Zero-effort cross-domain gesture recognition with Wi-Fi," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 11, pp. 8671–8688, Nov. 2021.
- [29] H. Abdelnasser, K. Harras, and M. Youssef, "A ubiquitous WiFi-based fine-grained gesture recognition system," *IEEE Trans. Mobile Comput.*, vol. 18, no. 11, pp. 2474–2487, Nov. 2019.
- [30] *IEEE Standard for Information Technology-Local and Metropolitan Area Networks-Specific Requirements—Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 6: Wireless Access in Vehicular Environments*, IEEE Standard 802, IEEE 802.11 p Working Group, 2010.
- [31] H. D. Hristov, *Fresnel Zones in Wireless Links, Zone Plate Lenses and Antennas*. Norwood, MA, USA: Artech House, 2000.
- [32] R. K. Pearson, "Outliers in process modeling and identification," *IEEE Trans. Control Syst. Technol.*, vol. 10, no. 1, pp. 55–63, Jan. 2002.
- [33] Y. Chen, R. Cao, J. Chen, L. Liu, and B. Matsushita, "A practical approach to reconstruct high-quality Landsat NDVI time-series data by gap filling and the Savitzky-Golay filter," *ISPRS J. Photogramm. Remote Sens.*, vol. 180, pp. 174–184, Oct. 2021.
- [34] K. Qian, C. Wu, Z. Yang, Y. Liu, and Z. Zhou, "PADS: Passive detection of moving targets with dynamic speed using PHY layer information," in *Proc. 20th IEEE Int. Conf. Parallel Distrib. Syst. (ICPADS)*, Dec. 2014, pp. 1–8.
- [35] M. G. Moghaddam, A. A. N. Shirehjini, and S. Shirmohammadi, "A WiFi-based system for recognizing fine-grained multiple-subject human activities," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (I2MTC)*, May 2022, pp. 1–6.
- [36] Y. Mei, T. Jiang, X. Ding, Y. Zhong, S. Zhang, and Y. Liu, "WiWave: WiFi-based human activity recognition using the wavelet integrated CNN," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC Workshops)*, Jul. 2021, pp. 100–105.



LEI ZHANG received the Ph.D. degree from the Perception Mine Research Center. He has been working as a Lecturer with the Xuzhou University of Technology, since 2019. His research interests include machine learning, mobile computing, and personnel sensing.



YUE ZHANG received the B.S. degree in electronic and information engineering from the Changzhou Institute of Technology, Changzhou, Jiangsu, China, in 2018. His research interests include device free wireless recognition, wireless networks, and machine learning.



RONG BAO is currently a Professor with the Xuzhou University of Technology. She is also the Director of Jiangsu Computer Society and the Jiangsu Artificial Intelligence Society. Her research interests include computer networks and embedded development.



YONGHONG ZHU has been an Assistant Professor with the Xuzhou University of Technology, since 2014. She is also the Vice President of Science and Technology under the supervision of Guang Shi Dian Qi. Her research interests include wireless multimedia sensor networks and embedded development.



XINGUO SHI was born in 1972. He received the B.S. degree in control engineering from the Shandong University of Science and Technology, in 1992. He is currently a Senior Engineer with the Information Center of Shandong Energy Zibo Mining Group Company. His research interest includes coal mine information technology.

...