IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 10 September 2023, accepted 1 October 2023, date of publication 9 October 2023, date of current version 16 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3322726

RESEARCH ARTICLE

Residual Neural Network Driven Human Activity Recognition by Exploiting FMCW Radar

CONG LI[1], XIANPENG WANG[1], (Member, IEEE), JINMEI SHI[2], HAN WANG[3], (Senior Member, IEEE), AND LIANGTIAN WAN[4], (Member, IEEE)

[1]School of Information and Communication Engineering, Hainan University, Haikou 570228, China
[2]College of Information Engineering, Hainan Vocational University of Science and Technology, Haikou 571158, China
[3]College of Physical Science and Engineering Technology, Yichun University, Yichun 336000, China
[4]Key Laboratory for Ubiquitous Network and Service Software of Liaoning Province, School of Software, Dalian University of Technology, Dalian 116620, China

Corresponding authors: Xianpeng Wang (wxpeng2016@hainanu.edu.cn) and Jinmei Shi (shijinmei1003@hvust.edu.cn)

This work was supported in part by the Important Science and Technology of Haikou City under Grant 2021-004; in part by the Natural Science Foundation of Hainan Province under Grant 620RC555; in part by the National Natural Science Foundation of China under Grant 61861015, Grant 61961013, and Grant 62101088; in part by the Radar Signal Processing National Defense Science and Technology Key Laboratory Fund under Grant 6142401200101; in part by the Jiangxi Provincial Natural Science Foundation under Grant 20224BAB202005; in part by the Science and Technology Project of Jiangxi Provincial Department of Education under Grant GJJ2201710; in part by the Jiangxi Province 03 Special Project and 5G Project under Grant 20232ABC03W02; and in part by the Innovative Research Projects for Graduate Students in Hainan Province under Grant Qhys2022-149.

ABSTRACT In recent years, radar-based human activity recognition has attracted the interest of a large number of researchers. Many researchers have proposed various effective processing algorithms. However, a good data processing algorithm not only has high recognition accuracy but also should be closer to the real application environment, such as having better detection robustness and detection sensitivity. This paper proposes a residual-bi-LSTM-attention hybrid multi-network, which has high recognition accuracy and the advantages of strong robustness and detection sensitivity. First, we collected data on five different activities of 13 volunteers in a laboratory setting. Then, after processing the collected data through the proposed algorithm, the micro-Doppler characteristics of each action are obtained. Finally, these desired features are fed into the proposed network for classification and recognition. Experimental results confirm the efficiency and accuracy of the proposed algorithm.

INDEX TERMS FMCW radar, data processing, deep learning, micro-Doppler, neural networks.

I. INTRODUCTION

Worldwide, the population is gradually aging, and the phenomenon of older people living alone is becoming increasingly evident. Research shows that for people 65 and older, the incidence of fall accidents increases with age [1]. Falls can cause serious health risks for seniors, ranging from bleeding and fractures to death [2]. Moreover, it also leads to increased time and economic costs [3]. Severe fall consequences are usually caused by not being detected in time [4]. How to effectively detect fall accidents and take timely and effective measures becomes the key [5]. Human activity recognition (HAR) using sensor technology can solve this one problem. Therefore, it is crucial to develop an accurate, easy-to-use, and reliable fall detection technology.

The associate editor coordinating the review of this manuscript and approving it for publication was Chengpeng Hao.

In previous work, researchers have investigated the identification and classification of human activity [6], [7]. Most of these technologies are based on computer vision solutions [8]. Zerrouki et al. investigated the problem of HAR based on body shape changes using a camera sensor solution [9]. In [10], Khan et al. considered the better effect of histogram of oriented gradient (HOG) features for recognition in low-noise environments. In high-noise environments, deep convolutional neural networks (DCNN) have more advantages. They used a method combining HOG and DCNN to identify human activities. Zhang et al. proposed using semantics-guided neural network (SGN) for activity recognition based on human skeleton [11]. They modeled each joint of the human body and analyzed the correlation between the joint points of the human body in the same frame of data. This end-to-end SGN has improved the performance of graph convolutional networks (GCN) and convolutional

neural networks (CNN). The experimental results show that it achieves better results in gesture recognition. However, when using this solution for elderly fall detection, users are concerned about their privacy being compromised. The solution does not guarantee its operating performance in poor lighting conditions. In order to solve this problem, non-invasive sensor solutions were considered by Researchers. Researchers used wearable sensors for HAR, such as acceleration, gyroscopes, and pressure sensors [12], [13]. In [14], Xugang et al. investigated surface electromyography (sEMG) features for pattern recognition and classification. These sensors can detect the signals generated during human activity accurately. The method has a very high recognition accuracy by combining these features with support vector machines. When a person is active, these sensors record various parameters of human movement. Moreover, the posture of the human body can be accurately detected and recognized [15], [16]. In [17], Mim et al. considered the characteristics of both temporal and spatial information of the feature, and proposed a model of convolutional block attention module (CBAM) based on gated recurrent unit (GRU). Zhang et al. also use accelerometer sensors to complete HAR [18]. Different from [17] and [18] uses acceleration sensors with three channels of x, y, and z to collect data, and then performs recognition based on the attention mechanism and CNN recognition network. Finally, the recognition accuracy rate of 96.4% is achieved, which solves the problem that the case-based method is difficult to recognize normally in the real environment. In [19], Khatun et al. proposed to combine CNN and long short-term memory network (LSTM) based on the self-attention mechanism. Multiple sets of data verified the feasibility of the model in HAR. However, it does not detect whether an elderly person has fallen when he or she is not wearing a sensor. This solution is not suitable as a full-time fall monitoring solution for the elderly. In addition, Chu et al. also conducted research on HAR using wifi channel state information (CSI) [20]. This is a contactless and non-intrusive solution for HAR. But this method can only have a good detection effect on the signal transmission channel, which makes the system less robust.

With the continuous development of radar technology, it has been widely used in target locating [21], [22]. In [23], Guo et al. proposed a method based on tensor decomposition for localization of traffic targets. At the same time, radar non-contact monitoring methods are being considered to be used in smart healthcare by researchers [24]. In [25], Qi et al. proposed a HAR algorithm based on ultra-wideband (UWB) radar. The algorithm first classifies the targets roughly based on their radial features. Then, power spectrum and micro-Doppler characteristics are extracted to refine the HAR further. Finally, these two features are fed into CNN for HAR. The final recognition accuracy can reach 98%. However, the UWB radar is susceptible to external interference. In order to reduce the impact of the radar regime, frequency-modulated continuous wave (FMCW) millimeter wave radar is of interest to researchers due to its high accuracy, low power consumption, and non-invasiveness [26], [27]. In [27], Ahmed et al. used FMCW radar to collect data on seven different activities. The algorithm uses distance variation as a feature, which is generated during human activity. Then, this feature is fed into a CNN and performs softmax classification. Since the algorithm only considers the distance feature, the final recognition accuracy is 91%. However, during the motion, the swing of the arms affects the distance feature. In order to fully show the characteristics of the target during its movement, micro-Doppler features are considered for use in HAR [28]. In [29], Kim et al. propose the application of combining micro-Doppler features with DCNN in human detection and classification. The results show that the algorithm's accuracy for HAR is 90.9%. In [30], Helen et al. started from the RGB three-channel data of Doppler spectrum characteristics and sent them into the network in three data channels for identification. However, this parallel input architecture will make the network more complex and will lead to longer network training time. Aman et al. proposed a recognition algorithm based on a combination of micro-Doppler features and a bi-directional long short-term memory network (Bi-LSTM) [31]. Six movements of 15 participants were used as experimental radar data to validate the proposed method. The average accuracy of the experiments was above 90%. Most recognition work is done using neural networks [32], [33]. In [34], Jiang et al. used the DenseNet neural network to identify human activities in order to extract as much effective feature information as possible from the deep network. Different from the above studies that use Doppler spectrum features, [34] and [35] use the point cloud information of the target as input features. In [35], Yu et al. proposed a dual-view neural network, which comprehensively analyzes from different observation perspectives, so as to judge the activity of the target. Although this method can display the target's activity changes in the form of a point cloud, a more complex antenna is required to obtain the three-dimensional information of the target. The powerful performance of these neural networks has been demonstrated in various fields. Hence, some researchers have tried to optimize the network structure to improve recognition accuracy. Huan et al. proposed an attention-based network for feature recognition [36]. With optimization from a feature perspective, the network part of the optimization can be expanded. In [37], Wang et al. introduced an attention-based vision transformer for HAR by millimeter wave radar. They proposed a vision transformer network with slicing, allowing the system to focus on classifying effective features.

Inspired by the research described above, a good feature extraction algorithm can reduce the pressure of subsequent feature recognition work. At the same time, it is also very important to have a good feature recognition network.

This paper proposes a data processing algorithm, including a data preprocessing module and a recognition network module. In the preprocessing module of the proposed

algorithm, the data adjustment module and distance window interception are introduced. By preprocessing the data, it makes the characteristics of the target movement more obvious, and the detection sensitivity is better, which is closer to the application scenario of the algorithm. In the recognition network part, a residual-Bi-LSTM hybrid multi-network structure based on the attention mechanism is proposed. The network structure combines the advantages of a residual network, long short-term memory network (LSTM), and attention mechanism. For the recognition work, a data-driven approach is used to conduct experiments. The data for the datasets used in the experiments were collected in a laboratory environment. The experimental results show that for the same dataset, the proposed algorithm can better describe the data characteristics, which have better recognition accuracy and better performance than traditional data processing methods.

The main work and contributions of this paper are summarized as follows:

(1) A data processing algorithm based on micro-Doppler features is proposed. The algorithm includes a data preprocessing module and a recognition network module.
(2) The data in the dataset used in this article was collected in a laboratory environment. These data are used for subsequent feature extraction and training for different recognition networks.
(3) The algorithm proposed in this paper is verified experimentally, and the experimental results show that the system has certain stability and robustness.

The remaining parts are organized in the following manner. The data model is formulated and analyzed in Section II. Section III introduces a data processing algorithm for HAR. The various parts of the identification system are described in detail. Data acquisition and experimental setup are described in section IV. The experimental results and discussion are placed in Section V. In section VI, we conclude the paper.

II. DATA MODEL

Our follow-up work is based on Texas Instruments' AWR 1642 single-chip FMCW radar system. As shown in Figure 1, the system has two transmitting antennas and four receiving antennas. The compact system can measure the distance, Doppler, and angle information.

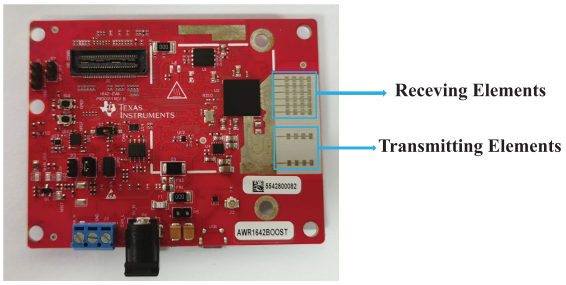

FIGURE 1. Texas Instruments' single-chip FMCW radar system AWR 1642.

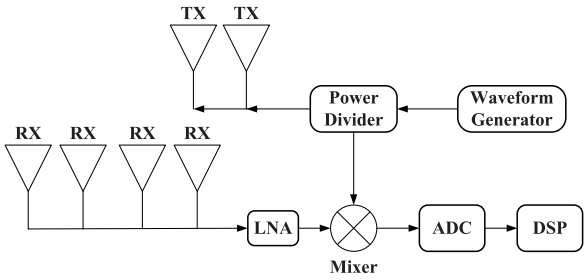

FIGURE 2. The structure diagram of the FMCW radar system.

The structure diagram of the FMCW radar system is shown in Figure 2. The system consists of a waveform generator (WG), power divider (PD), transmitting antenna (TX), low noise amplifier (LNA), receiving antenna (RX), mixer, analog to digital converter (ADC), and digital signal processor (DSP). The WG generates the FMCW radar signal, which varies linearly with time. The data model of the transmit signal can be expressed as:

$$s(t) = e^{j2\pi\left(f_c t + 0.5\mu_s t^2\right)} \tag{1}$$

where f_c is the center frequency of the baseband signal, μ_s is the slope of the FMCW wave. For the cases where there are K targets present, the received signal can be expressed as follow:

$$r(t) = \sum_{j=1}^{K} \gamma_j s(t - \tau_j) \tag{2}$$

where $j(j = 1, 2, 3, \ldots, K)$ is the target index, γ_j is the reflection coefficient of the j-th target, and τ_j is the time delay of the j-th target, which is closely related to the target range information. The relationship between them is:

$$\tau_j = \frac{2d_j}{c} \tag{3}$$

where c is the speed of light and d_j is the range of the j-th target. For HAR, the angle information is not important. Thus, the delay caused by the internal array elements of the transmitter and receiver units is ignored. Only the effect of the target range on the delay is considered. After the received data passes through the LNA, it is multiplied with the transmit signal for mixing. The signal is then passed through low-pass filtering and is sampled by ADC. After that, the signal can be expressed as:

$$x[n] = x[nT_s] = \sum_{j=1}^{K} \gamma_j \exp\left[j2\pi\left\{V_j n T_s + \frac{2f_c}{c} d_j\right\}\right] \tag{4}$$

$$V_j = \frac{2\mu_s}{c} d_j + \frac{2f_c}{c} v_j \tag{5}$$

where V_j is induced with the micro-movement of the target at moment j, and v_j represents the relative velocity of the j-th target and T_s is the ADC sampling period. Finally, the single baseband chirp signal is modeled n times.

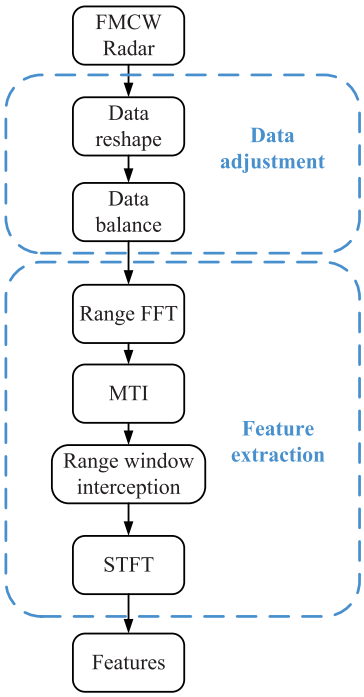

FIGURE 3. Micro-Doppler signal preprocessing framework.

III. PROPOSED DATA PROCESSING ALGORITHM

A. DATA ADJUSTMENT

In this section, the micro-Doppler features of the data are first extracted using data and processing algorithms. Then, it is fed into the proposed network for training. Figure 3 shows the preprocessing framework of the complete algorithm. The data features are extracted from such a data framework.

Users want the system to respond as quickly as possible when a fall accident occurs. However, some current algorithm research blindly pursues the recognition accuracy of the algorithm without considering the application of the algorithm in the natural environment. Therefore, the first step is to adjust the radar data after acquiring the radar data.

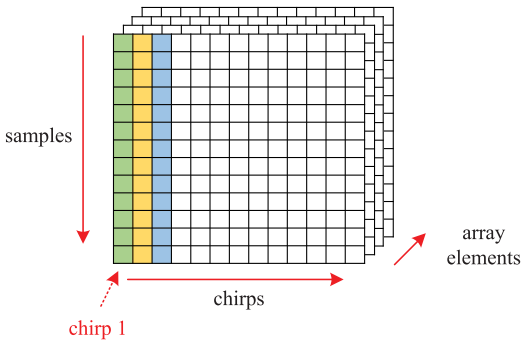

FIGURE 4. 3D data block format.

Note that the data sampled by the ADC is transferred to the computer via the serial data port. The initial data transfer to the computer does not facilitate our data processing. Firstly, the data is rearranged into a 3D data block. Figure 4 shows its format, where each column corresponds to all samples in a chirp signal. For each frame of data, different columns represent different chirps, and the target data is sampled multiple times. Data received by different receive array elements are stored on different pages of the data block. The data in each data cell of this data block is the intermediate frequency (IF) signal after mixing and filtering. Then, the data from different array elements are balanced. The advantage is that interest signals are strengthened. Moreover, signals that are not of interest are weakened. The specific step is to sum the data with the same row vertical coordinates but different pages in a 3D data block and then average.

In order to make the proposed algorithm closer to the actual application scene, the algorithm takes ten frames of data each time for processing, and the action duration is 1 second. When processing the next data, if the data is fetched following the end of the previous data, the data will be artificially cut off. If the old man has a fall accident when the data is cut off, the algorithm will fail. In order to ensure that the algorithm can work normally at all times, the starting point of data processing at the next moment is the middle moment of data processing at the current moment. In this way, a piece of data is detected and processed multiple times, which can effectively avoid detection blind spots.

B. FEATURE EXTRACTION

To obtain target range information, in the frequency domain, we can perform the Fourier transform of Equation 4 to obtain the baseband signal, which can be expressed as:

$$X[k] = \sum_{n=0}^{N-1} x[n] \exp\left(-j2\pi \frac{n}{N} k\right) \tag{6}$$

where $k(k = 0, 1, 2, \ldots, K-1)$ denotes the frequency of the index. The spectrogram can be obtained by acquiring the baseband signal in the frequency domain. Since the transformed frequency information is proportional to the distance of the target, the result in the frequency domain can be considered as the result of the distance dimension. The spectrogram of the collected baseband signal over the

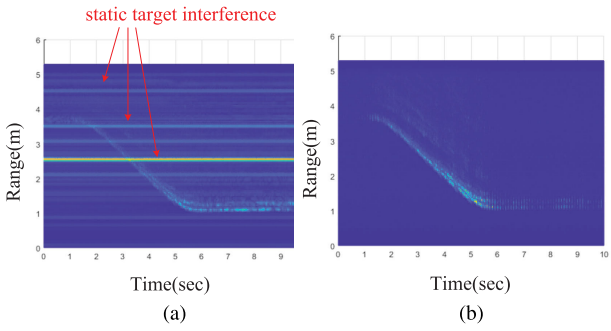

FIGURE 5. Range changing with time. (a) Processed without MTI. (b) Processed with MTI.

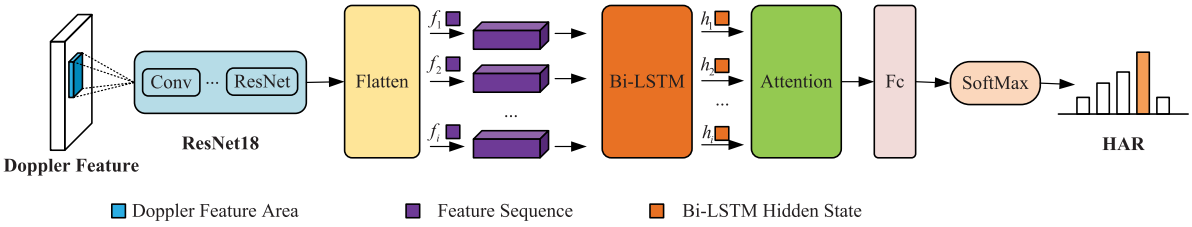

FIGURE 6. Proposed RBLA network structure.

N_q interval can be expressed as:

$$\mathbb{X}^{(N_q)} = \left[\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \ldots, \mathbf{X}^{(N_q)}\right] \tag{7}$$

where $\mathbf{X}^{(i)} = \left[X^{(i)}(0), X^{(i)}(1), \ldots, X^{(i)}(K-1)\right]^T$ denotes the frequency domain of the baseband signal as a vector and i denotes the index of the interval.

However, in reality, much static target interference will affect the results. Therefore, the moving target indication (MTI) algorithm is used to indicate dynamic target [38]. The MTI algorithm subtracts the previous moment's signal from the original signal at different distance units respectively. The enhanced micro-Doppler feature signals were subsequently obtained. In this paper, it can be expressed as:

$$X_{RT\_MTI}(m, n) = X_{RT}(m, n) - X_{RT}(m, n-1) \tag{8}$$

where m is the index of the distance dimension, n is the index of the slow time dimension. The variation in the target range over time is shown in Figure 5. Figure 5(a) and Figure 5(b) show the results without MTI and with MTI, respectively.

The horizontal lines in Figure 5(a) are static target interference, which MTI can eliminate. Figure 5(b) shows the display results. For the HAR and fall detection system, the signal of interest occurs only in the range of human motion. Therefore, the range window interception (RWI) cropped out the signal that exceeds the range.

After that, the target's frequency domain characteristics can be analyzed using Fourier transform. However, this method does not apply to non-stationary signals that change over time. This paper uses short-time Fourier transform (STFT) to extract the combined time and frequency domain micro-Doppler signals after RWI. STFT can be specifically described as:

$$\mathrm{STFT}(t, \omega) = \int_{-\infty}^{\infty} s(\tau) g^*(\tau - t) e^{-j\omega\tau} d\tau \tag{9}$$

where $g(t)$ is a window function that is used to select signals within a particular window area. After STFT, the characteristic pattern of the micro-Doppler spectrogram can be obtained during collection.

Figure 7 shows the micro-Doppler spectrum feature maps of five kinds of motion data, which are obtained after the above data preprocessing algorithm. The duration of these feature example images shown is very short, only 1 second. When intercepting data, the features in the dataset are not always so perfect and easy to distinguish. This is just an example of processing some of the features.

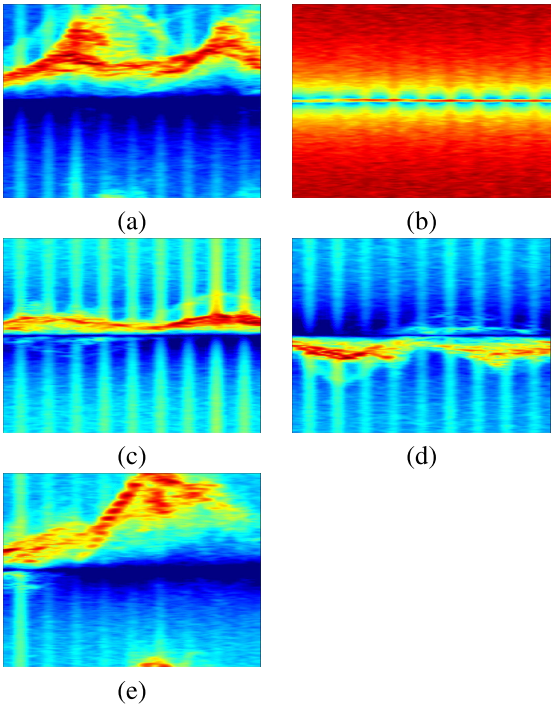

FIGURE 7. Five examples of motion signatures after micro-Doppler signal processing (a) Walking. (b) Standing still. (c) Standing up. (d) Sitting down. (e) Falling down.

C. RESIDUAL-BI-LSTM-ATTENTION HYBRID MULTI-NETWORK

In the study of HAR, the micro-Doppler feature is the preferred feature by most researchers. In some studies, researchers extended the time window to 4 seconds or longer in order to obtain micro-Doppler features that are more easily distinguished. This will be reflected in the actual situation that when the old man falls, the fall cannot be detected immediately, and it may be detected after waiting for a period of time. However, in reality, we hope that the system can detect it immediately when a fall occurs.

This research fully considers the real-time performance of the system and ensures a relatively good recognition accuracy. The advantages of the three networks are combined to form a new network. In this paper, we proposed a Residual-Bi-LSTM-Attention Hybrid Multi-Network (RBLA) to class

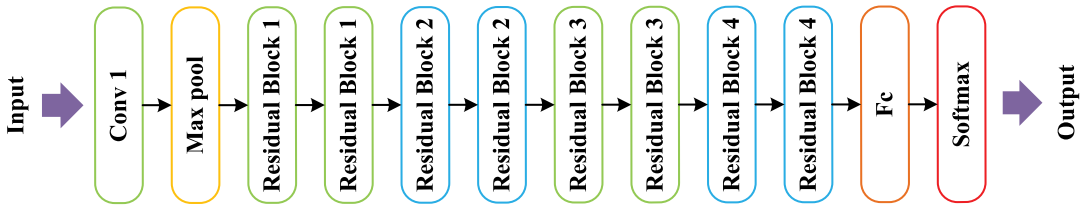

FIGURE 8. The architecture of ResNet-18.

different kinds of postures. The structure of the RBLA network is shown in Figure 6. The residual network extracts effective features from micro-Doppler spectral features through its special feedback link. The obtained data features are flattened through the flattened layer. Then, the flattened features are fed into the Bi-LSTM network in batches to obtain the correlation features between before and after the time window. Finally, the weight assignment mechanism in the attention mechanism is exploited to highlight features in the hidden layers of Bi-LSTM.

In previous studies, the excellent performance of CNN in HAR has been verified. In the network of this study, CNN was not simply used, but a residual module was used instead. The CNN network can extract some features of the data, but it can't do anything about some deep features. The residual network is optimized on the basis of CNN, which retains the advantages of the CNN network. For the insufficiency of CNN in extracting deep data features, the residual module adds a feedback link, which effectively solves this problem.

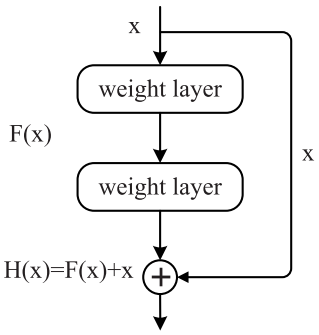

FIGURE 9. The structure of the residual block.

Residual network was proposed by He in 2016 [39]. Generally, deeper layers of networks will have powerful performance and can extract deeper features. Nevertheless, when the number of network layers rises to a certain level, the performance of neural networks will not improve through increasing layers of the network. On the contrary, it may cause attenuation. In order to solve this problem, the residual block structure is proposed to increase the layers. At the same time, hidden information can be learned. According to the residual block, gradient disappearance, gradient explosion, and network generalization can be avoided. Figure 9 shows the structure of the residual block. Residual block increases a shortcut connection compared to the general neural network, weakening the connection between each layer to learn more

TABLE 1. Parameters of ResNet-18 network.

Residual convolution module	Filter configurations
Conv 1	7 x 7, 64, stride = 2
Residual Block 1	1 x 1, 64, stride = 1 3 x 3, 64, stride = 1
Residual Block 2	3 x 3, 128, stride = 1 3 x 3, 128, stride = 1
Residual Block 3	3 x 3, 256, stride = 1 3 x 3, 256, stride = 1
Residual Block 4	3 x 3, 512, stride = 1 3 x 3, 512, stride = 1

characteristics in deeper layers. Two paths can lead the data to deeper layers. One needs to go through each layer of the network. The other flows to the deep network via the shortcut connection. In addition, the performance of Residual network-18 (ResNet-18) surpassed the human level for the first time in ImageNet's image classification.

ResNet-18 consists of 17 convolutional layers (Conv) and a fully connected layer (Fc) not used in all 17 convolutional layers. In order to reduce network parameters, global average pooling was introduced. The architecture of ResNet-18 is shown in Figure 8, where Conv represents the residual convolution module, and the specific information on the parameters of each layer in the module structure is given in TABLE 1. Conv 1 has only one convolutional layer, while the remaining four convolutional modules contain two convolutional operations. The 17 Conv consist of one Conv 1 unit, two Conv 2 units, two Conv 3 units, two Conv 4 units, and two Conv 5 units. The structure of this residual network can extract deep data features, while ensuring the real-time requirements of the system at the same time.

Bi-LSTM is a special kind of LSTM. And understanding LSTM needs to be traced back to the recurrent neural network (RNN) because LSTM is extended from RNN. Because of the particular structural unit of the RNN, it can analyze and model the dynamic behavior of the data sequence. The LSTM's structural unit controls the flow and status of current data through three different gates. A gate is used to determine whether the current state of the data should be forgotten. The next gate is used to control whether the data should be read. The last gate is used to control whether the status should be updated. LSTM only extracts data features from a single direction, while Bi-LSTM processes data from both directions in order to obtain more hidden features.

Figure 10 shows the structure of the LSTM operation unit. When the input time is the data sequence x at t, x_t is jointly

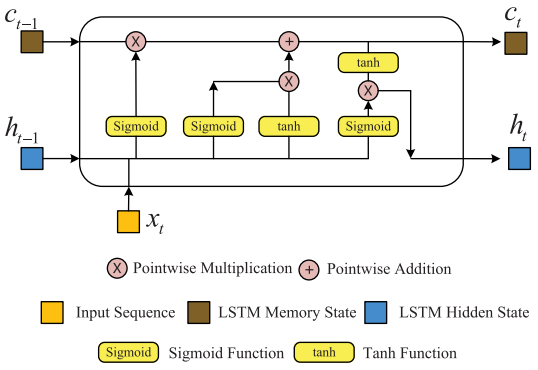

FIGURE 10. The network structure of the LSTM operation unit.

affected by the parameters c_{t-1} and h_{t-1}, where t represents the moment of the data. c and h are two transmission state parameters of LSTM, representing the memory unit and hidden unit of LSTM, respectively. The parameters c_t and h_t generated at the current t time will affect the output at $t+1$ time.

Compared to the LSTM algorithm, Bi-LSTM operates on sequences in the forward and backward directions. According to the past state, the forward LSTM operation determines whether the current data is saved or forgotten. The backward LSTM operation judges the current data by observing the data of the future state. Finally, the features extracted from both directions are merged to obtain the final feature output. Figure 11 shows the structural model of Bi-LSTM. It is composed of multiple LSTM structural units and extracts features from the data from two directions of the data sequence. Then, the features extracted from these two directions are combined and output as a first-level feature, shown as h_1, h_2, and h_3 in Figure 11.

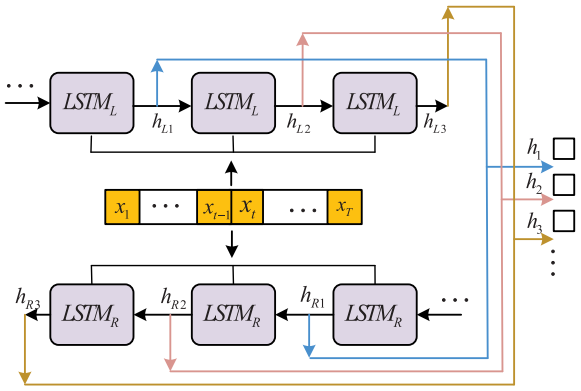

FIGURE 11. The network structure of Bi-LSTM.

The attention mechanism can adjust the weight of its features according to the different characteristics of the input data. In the proposed network, after processing the data through the previous residual network and Bi-LSTM, multiple segmentation features are obtained. The input of the attention mechanism network is such features as h_1, h_2, and so on obtained by Bi-LSTM. After weighing the input

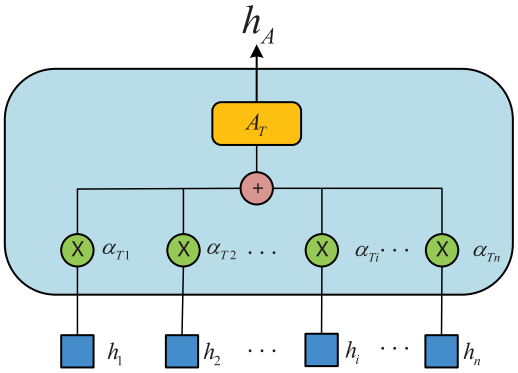

FIGURE 12. The network structure of the attention mechanism module.

parameters, the weight of each part is obtained to highlight the characteristics of important parts. After weighing the input parameters, the weight of each part is obtained to highlight the characteristics of important parts. Then, weigh and sum the data features of each part to obtain the final data features. Figure 12 shows the network structure of the attention mechanism module, where α_{Ti} represents the $i-th$ input feature value h_i corresponding to the time T.

IV. DATA ACQUISITION AND EXPERIMENTAL SETUP

A. EXPERIMENTAL DATA ACQUISITION

This data was collected in a laboratory environment with a Texas Instruments AWR 1642. Its specific configuration parameters are shown in Table 2. Because the angle has little impact on human activities, this article adopts the mode of one transmitting antenna and four receiving antennas. This greatly reduces the data size of human activities. The radar

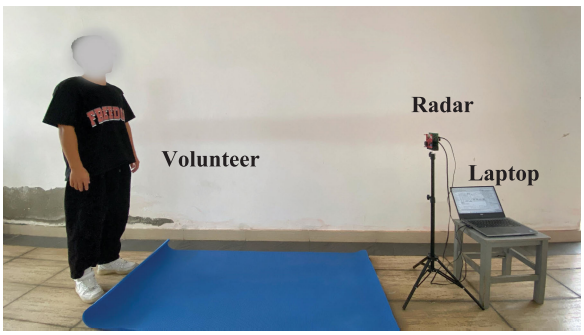

FIGURE 13. Laboratory collection environment.

TABLE 2. The configuration parameters of the radar evaluation.

Parameter name	Value
Carrier Frequency	77GHz
Frequency Slope	53GHz/s
Sample Rate	2.5M
Period of data per frame	100ms
ADC Sample	128
Chirp Loops per frame	255
Number of frames	100
Number of transmit channels	1
Number of receiving channels	4

TABLE 3. Volunteer Details.

Volunteers	Gender	Age(years)	Weight(kg)	Height(m)	BMI(kg/m^2)
1	M	25	55	1.63	20.70
2	M	24	75	1.77	23.94
3	M	27	70	1.70	24.22
4	F	26	45	1.70	15.57
5	F	25	67	1.60	26.17
6	M	25	75	1.72	25.35
7	M	25	70	1.70	24.22
8	F	25	56	1.67	20.08
9	M	26	59	1.73	19.71
10	M	31	71	1.73	23.72
11	M	25	62	1.66	22.50
12	M	25	66	1.72	22.31
13	M	23	70	1.77	22.34
Total	M/F (10/3)	25.54±1.82	64.69±8.51	1.70±0.05	22.37±2.72

is placed on a tripod with a vertical distance of 1.2 meters from the ground, and the side with the radar antenna is placed facing the volunteers. Volunteers completed various actions at a horizontal distance of 2 meters from the radar. Figure 13 shows the laboratory collection environment.

Five common postures in daily life are collected, including walking, standing still, standing up, sitting down, and falling. Thirteen volunteers were selected to generate experimental data, including ten males and three females. Table 3 shows the details of the volunteers. They are 23 to 31 years old, weigh 45 to 75 kilograms, and are 1.60 to 1.77 meters tall. For each type of motion data, volunteers were repeated six times. Taking into account various human movement patterns, walking and falling motions were captured in two cases: motion toward the radar and motion away from the radar, respectively. During repeated walks and falls six times, the activities of the first three times moved towards the radar, and the last three times were away from the radar. Altogether, 384 data on human activity can be obtained.

B. DATASET

In this paper, the acquisition time of each action is 10 seconds. When the proposed algorithm processes the data, the data length of each feature window is 1 second. It is necessary to truncate the collected long data, which is more in line with the processing of data in real application environments. To this end, the collected 384 sets of data were processed according to the proposed algorithm and then truncated. Then, according to the difference of each micro-Doppler characteristic segment picture, manually select its motion segment as the data of the dataset. In the process of collecting motion data from volunteers, they are not always in motion during the whole process. For example, a fall accident occurs only for a moment, and other moments may be static. In the end, a data set corresponding to the five actions of walking, standing, standing, sitting, and falling was obtained. In the dataset, the number of each action is shown in Table 4.

This is also consistent with the actual situation. Because each group of actions has idle time and the target is still, it has the largest amount of data in the dataset. In the process of motion, the duration of walking is the longest, so more data will be generated in the same time window. The action sequence of standing up and sitting down is exactly the opposite, so the amount of data is roughly the same. The fall event always happens in an instant, which also leads to the least amount of data among the five actions.

TABLE 4. Detail quantitative information of six different postures.

Postures	Number of samples
Walking	808
Standing still	5868
Standing up	366
Sitting down	349
Falling	289
Total	7680

The five types of human activities of 13 volunteers recorded by the self-built dataset: (1) walking, (2) standing still, (3) standing up, (4) sitting down, (5) falling down are shown in Figure 14. Compared to the dataset from the University of Glasgow, UK [40]. Their dataset records six kinds of older people's daily activities: walking, sitting down, standing up, picking up an object, drinking water, and falling. They used the FMCW radar to acquire data at an operating frequency of 5.8GHz with a chirp bandwidth of 400MHz. A total of 1589 micro-Doppler signature samples were generated. The duration of each of their features is 10 seconds, and we further divide the dataset on the basis of 10 seconds. The data set used in this article has a larger data set, and the time window of the data is divided into 1 second, which is more in line with the real situation.

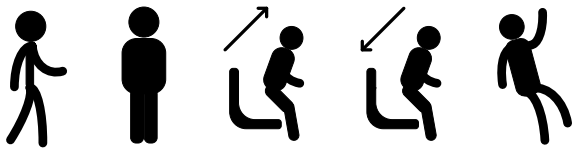

FIGURE 14. The five types of human activities.

V. RESULT AND DISCUSSION

In this section, the control variable method will be used to conduct comparative experiments on the optimized modules to compare the recognition effects of each module in different networks. The first subsection V-A introduces the evaluation indicators of the detection results. The following subsections are control experiments for adjusting the optimization modules. Among them, subsection V-C and subsection V-D are adjusted on the basis of subsection V-B, and subsection V-E is the overall control experiment of the proposed algorithm. It is worth noting that the results presented in this section are obtained after repeated verification of the data in the dataset. Before identifying the network, we divided the data set into ten equal parts according to the activity labels. Then, these ten pieces of data are combined into training sets and data sets in turn to train and verify the network. The results shown are the final results after averaging these multiple experiments.

A. EVALUATION METRICS

To evaluate the performance of different methods for HAR, four metrics are used in this paper to evaluate system performance: accuracy, precision, recall, and F1-score. These metrics are respectively defined as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{11}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{12}$$

$$\text{F1-Score} = \frac{2(\text{Precison} \times \text{Recall})}{\text{Precison} + \text{Recall}} \tag{13}$$

where TP is truly positive, which means correct classification. TN is truly negative, which means correct rejection. And FP is the abbreviation of false positive, which means false alarm. FN is the abbreviation of false negative, which means missing detection.

B. GENERAL DATA PROCESSING

In this subsection, the data are used for various machine learning methods to recognize human activities after general data processing as a benchmark for comparison. The recognition accuracy of the general algorithm is shown in Figure 15. In the basic comparison experiment, the algorithm RBLA proposed in this article has the best recognition accuracy rate of 95.79% in all machine learning. LSTM has the lowest recognition accuracy. Although the recognition accuracy of LSTM is the worst among the five machine learning algorithms, its accuracy can still reach 90.18%, which means that it can correctly recognize most poses. The network combined with CNN and LSTM ranked second with 92.98% recognition performance, followed by ResNet-18 and CNN.

Table 5 shows the scores of the four evaluation indicators, including accuracy, precision, recall, and F1-score. These four evaluation indicators are used to comprehensively evaluate the performance of the network. Generally, higher accuracy algorithms have higher precision, recall, and F1-score.

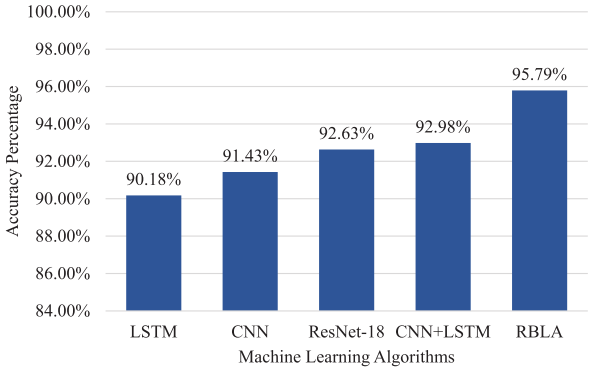

FIGURE 15. Recognition accuracy of different machine learning methods after general data processing.

TABLE 5. Results of different metrics of different machine learning methods after general data processing.

Algorithm	Accuracy	Precision	Recall	F1-Score
LSTM	90.18%	0.74	0.76	0.74
CNN	91.43%	0.88	0.79	0.77
ResNet-18	92.63%	0.83	0.81	0.82
CNN+LSTM	92.98%	0.84	0.82	0.83
RBLA	95.79%	0.90	0.89	0.89

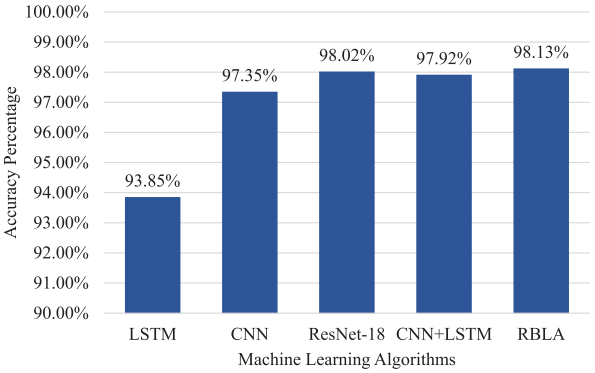

FIGURE 16. Recognition accuracy of different machine learning methods after data adjustment.

C. DATA ADJUSTMENT

After data adjustment, the performance of the five selected machine learning networks has been significantly improved. Among them, RBLA has the best accuracy rate of 98.13%, the combined network of CNN and LSTM is 97.92%, and the LSTM with the worst recognition accuracy rate is 93.85%. Figure 16 shows the recognition accuracy of different machine learning algorithms after data adjustment. Table 6 shows specific scores for each machine learning algorithm.

It can be seen from Table 6 that after data manipulation, the values of the four evaluation indicators have all increased. Compared with the recognition accuracy in Table 5. The improvement effect of CNN is the most significant, with

TABLE 6. Results of different metrics of different machine learning methods after data adjustment.

Algorithm	Accuracy	Precision	Recall	F1-Score
LSTM	93.85%	0.58	0.43	0.41
CNN	97.35%	0.84	0.84	0.84
ResNet-18	98.02%	0.87	0.90	0.88
CNN+LSTM	97.92%	0.88	0.85	0.86
RBLA	98.13%	0.90	0.93	0.91

an improvement of 5.92%. Secondly, ResNet-18 has an improvement of 5.39%, followed by the network combined with CNN and LSTM and LSTM, which have an improvement of 4.93% and 3.68%, respectively. After adjusting the data, the characteristics of each action data set are more clear. Experimental results show that data adjustment has a positive impact on HAR.

D. RANGE WINDOW INTERCEPTION

Figure17 shows the accuracy of different methods after range window interception. It can be seen that RBLA has the best recognition accuracy, reaching 99.06%. The recognition effect of LSTM is the lowest, which is 94.15%. CNN, ResNet-18, and CNN combined with LSTM have similar performance. After replacing the data adjustment step with a distance window interception, there is only a small fraction of the improvement. However, it can still be seen from Table 7 that the distance window operation improves the network performance.

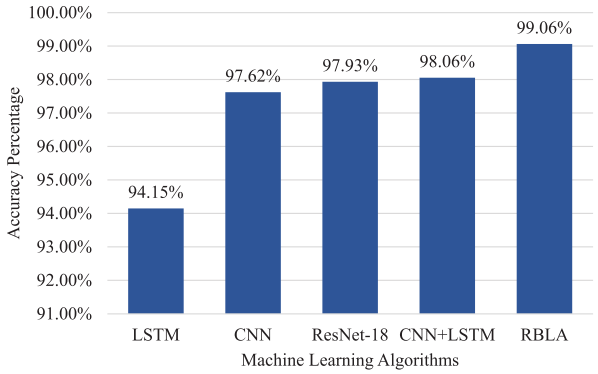

FIGURE 17. Recognition accuracy of different machine learning methods after range window interception.

As can be seen from Table 7, CNN, ResNet-18, CNN, and LSTM combined network performance recognition accuracy have similar values. However, the other three evaluation indicators are different. The remaining three parameters of CNN are relatively stable, but overall, its performance is worse than that of ResNet-18 and CNN combined with LSTM. Compared with precision and f1-score, the Recall of ResNet-18 has a larger value, indicating that the network is more sensitive and more suitable for detection tasks that require high security. The precision, recall, and f1-score parameters of the network combined with CNN and LSTM are close, indicating that it has relatively stable performance.

TABLE 7. Results of different metrics of different machine learning methods after range window interception.

Algorithm	Accuracy	Precision	Recall	F1-Score
LSTM	94.15%	0.72	0.43	0.41
CNN	97.62%	0.86	0.86	0.86
ResNet-18	97.93%	0.86	0.90	0.88
CNN+LSTM	98.06%	0.89	0.88	0.89
RBLA	99.06%	0.94	0.98	0.96

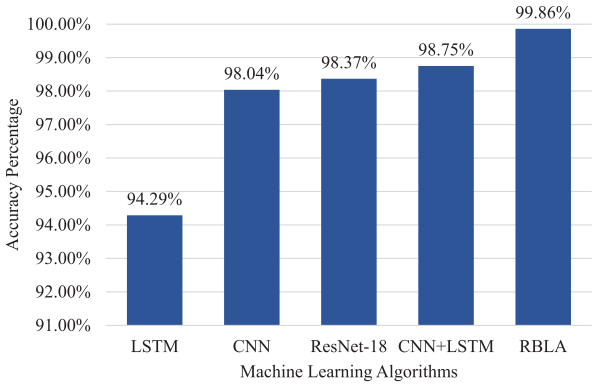

FIGURE 18. Recognition accuracy of different machine learning methods after using proposed data processing algorithm.

TABLE 8. Results of different metrics of different machine learning methods after using proposed data processing algorithm.

Algorithm	Accuracy	Precision	Recall	F1-Score
LSTM	94.29%	0.75	0.44	0.43
CNN	98.04%	0.88	0.89	0.88
ResNet-18	98.37%	0.91	0.92	0.92
CNN+LSTM	98.75%	0.93	0.89	0.91
RBLA	99.86%	0.99	0.99	0.99

E. PROPOSED DATA PROCESSING ALGORITHM

After putting the data adjustment module and the range window interception module back into the overall algorithm framework, the data is processed to obtain the final human activities features. Five machine-learning algorithms were used to identify these extracted features. Figure 18 shows the recognition accuracy of different machine learning algorithms. RBLA has the highest recognition accuracy, reaching 99.86%, which is 4.07% higher than the accuracy after conventional data adjustment in Table 5. Because the initial recognition accuracy of the network is relatively high, the improvement in network performance is not obvious. The second best method is to use a network that combines CNN and LSTM, which improves by 5.77%. Among the five recognition algorithms, LSTM has the worst recognition accuracy. However, after the optimization of the two modules, some performance is still improved, with a high recognition accuracy of 94.29%. The recognition performance of CNN and ResNet-18 also increased by 6.61% and 5.74% from the initial 91.43% and 92.63, respectively, reaching the final 98.04% and 98.37%. Table 8 shows the specific situation of its four evaluation indicators. It can be seen from Table 8

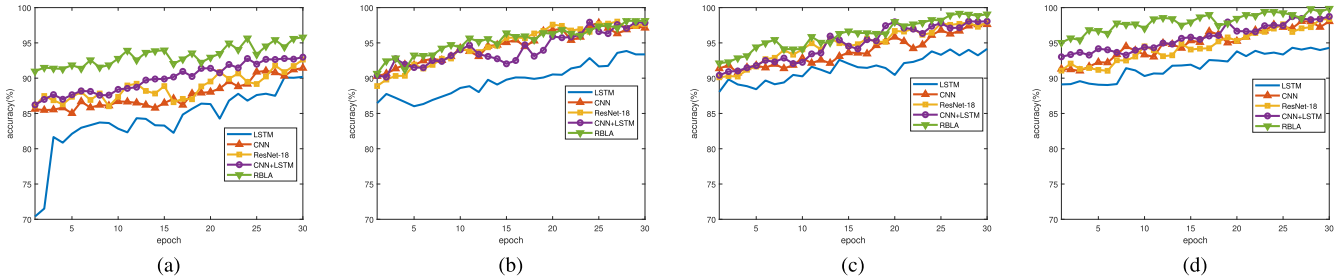

FIGURE 19. The recognition accuracy rate of five neural network algorithms at four different stages changes with epoch. (a) General data processing. (b) Data adjustment. (c) Range window interception. (d) Proposed data processing algorithm.

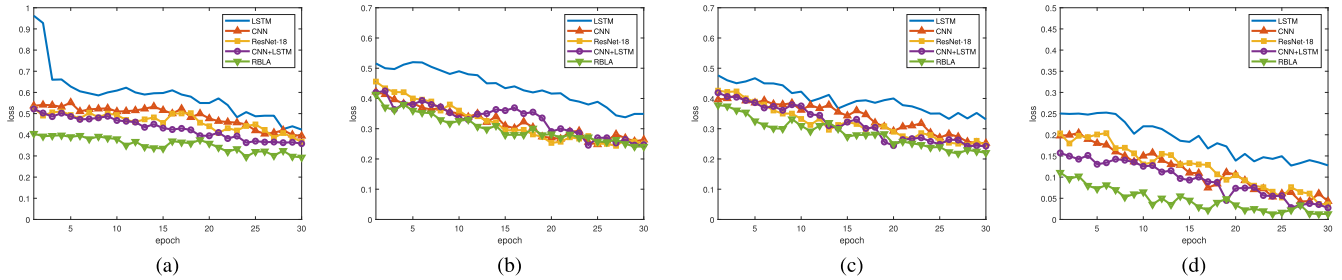

FIGURE 20. The change curve of the loss of five neural network algorithms in four different stages with epoch. (a) General data processing. (b) Data adjustment. (c) Range window interception. (d) Proposed data processing algorithm.

TABLE 9. Comparison of proposed algorithm with existing architectures.

Deep neural network type	Type of activity	Input type	Accuracy(%)
DCNN [27]	7-Indoor	Range-time map	91.00
Tower CNN [30]	6-(Indoor and outdoor)	RGB three-channel spectrogram	97.58
Dual-view CNN(DVCNN) [35]	7-Indoor	Radar point clouds	98.00
Vision transformer-slice [37]	5-(Indoor and outdoor)	Micro-Doppler map	99.12
Proposed RBLA	5-Indoor	Doppler spectrogram	99.86

that the proposed algorithm is superior to the other four comparison algorithms in four evaluation indicators.

F. ANALYSIS AND COMPARISON OF THE PROPOSED MODEL AT DIFFERENT STAGES

In the section V-B to V-E, the experimental results show the four performance indicators of the proposed algorithm in different stages. In order to make the results more sufficient and specific, this section will compare and analyze the accuracy and loss results in each stage. As shown in Figure 19, the four sub-figures respectively show the curves of the accuracy of the five neural network algorithms at different stages with the number of iterations. The accuracy of the five algorithms in the four stages as a whole continues to improve with the increase in the number of iterations. In the first stage, the accuracy of the proposed algorithm is mostly higher than the four comparison algorithms. As the data processing stage continues, the gap in the accuracy of these five algorithms continues to narrow. In the second and third stages of data processing, the four algorithms, except LSTM, ended up with similar accuracy. As can be seen from Figure 19(a), most of the accuracy rates of the proposed algorithm are better than those of the four comparison algorithms. As the processing of the proposed data preprocessing algorithm continues to deepen, the accuracy of the five algorithms has been improved to a certain extent. At the same time, the change curves of the loss of the five neural networks with the number of iterations are shown in Figure 20. Corresponding to the accuracy results in Figure 19, higher accuracy has a lower loss.

G. COMPARISON OF PROPOSED ALGORITHM WITH EXISTING ARCHITECTURES

Table 9 summarizes the results of existing algorithms for HAR. The data used in the five methods in the table are collected based on FMCW radar. Although the data sets and input types used by these algorithms are different, they are still representative to a certain extent. Among these four existing algorithms, Wang et al. obtained the highest recognition accuracy by using the visual converter slice algorithm based on the attention mechanism, reaching 99.12% [37]. The method proposed in this article is also a deep learning algorithm based on the attention mechanism. The difference is that the proposed algorithm further considers the correlation between the data before and after the time dimension and also attempts to use a residual neural network to replace the traditional CNN. The finally proposed algorithm achieved an accuracy of 99.86%, which was 0.74%

higher than the algorithm in [37]. It is worth noting that the dataset used in this paper has a total of 7680 sets of data, while the dataset used in [37] has only about 1500 sets of data. To a certain extent, it reflects the superiority of the algorithm proposed in this paper.

VI. CONCLUSION

This paper proposes a residual-Bi-LSTM hybrid multi-network data processing algorithm based on the attention mechanism, optimizes two sub-modules in data preprocessing, and combines the advantages of multi-network. The algorithm can effectively extract the deep features in the data and enhance the signal strength of the features of interest in the data. After the data is processed by the proposed algorithm, the feature data is transmitted to the network as an input signal. RBLA achieves the best performance of 99.86%. Although the algorithm has achieved good results, this experiment still has its limitations. Even for the same human activity, there may be some subtle changes in the movements of different people. This is related to various factors such as the volunteer's age, gender, and whether the body is slim or obese. In subsequent research, we will also strengthen communication and cooperation with hospitals and other nursing institutions to build coverage for people of all ages. Second, matching different features with different recognition networks will also have performance differences. Therefore, we plan to optimize the recognition network in follow-up work to improve the matching degree of features and network. Multi-level feature extraction and improved deep learning algorithms are used to further improve recognition performance. For this system, although the implementability in real environments has been improved, there is still a certain distance to achieve complete real-time performance. The lightweight network structure will make it possible to realize the algorithm in real-time.

REFERENCES

[1] R. Igual, C. Medrano, and I. Plaza, "Challenges, issues and trends in fall detection systems," *Biomed. Eng. OnLine*, vol. 12, no. 1, p. 66, 2013, doi: 10.1186/1475-925X-12-66.

[2] X. Li, Y. He, and X. Jing, "A survey of deep learning-based human activity recognition in radar," *Remote Sens.*, vol. 11, no. 9, p. 1068, May 2019, doi: 10.3390/rs11091068.

[3] Y. Seo, K. Kim, and J.-S. Kim, "Trends of nursing research on accidental falls: A topic modeling analysis," *Int. J. Environ. Res. Public Health*, vol. 18, no. 8, p. 3963, Apr. 2021, doi: 10.3390/ijerph18083963.

[4] J. R. Ehrlich, S. E. Hassan, and B. C. Stagg, "Prevalence of falls and fall-related outcomes in older adults with self-reported vision impairment," *J. Amer. Geriatrics Soc.*, vol. 67, no. 2, pp. 239–245, Feb. 2019, doi: 10.1111/jgs.15628.

[5] M. G. Amin, Y. D. Zhang, F. Ahmad, and K. C. D. Ho, "Radar signal processing for elderly fall detection: The future for in-home monitoring," *IEEE Signal Process. Mag.*, vol. 33, no. 2, pp. 71–80, Mar. 2016, doi: 10.1109/MSP.2015.2502784.

[6] S. Sehgal, "Human activity recognition using BPNN classifier on HOG features," in *Proc. Int. Conf. Intell. Circuits Syst. (ICICS)*, Apr. 2018, pp. 286–289, doi: 10.1109/ICICS.2018.00065.

[7] H. Ramirez, S. A. Velastin, I. Meza, E. Fabregas, D. Makris, and G. Farias, "Fall detection and activity recognition using human skeleton features," *IEEE Access*, vol. 9, pp. 33532–33542, 2021, doi: 10.1109/ACCESS.2021.3061626.

[8] E. Alam, A. Sufian, P. Dutta, and M. Leo, "Vision-based human fall detection systems using deep learning: A review," *Comput. Biol. Med.*, vol. 146, Jul. 2022, Art. no. 105626, doi: 10.1016/j.compbiomed.2022.105626.

[9] N. Zerrouki, F. Harrou, Y. Sun, and A. Houacine, "Vision-based human action classification using adaptive boosting algorithm," *IEEE Sensors J.*, vol. 18, no. 12, pp. 5115–5121, Jun. 2018, doi: 10.1109/JSEN.2018.2830743.

[10] M. A. Khan, M. Sharif, T. Akram, M. Raza, T. Saba, and A. Rehman, "Hand-crafted and deep convolutional neural network features fusion and selection strategy: An application to intelligent human action recognition," *Appl. Soft Comput.*, vol. 87, Feb. 2020, Art. no. 105986, doi: 10.1016/j.asoc.2019.105986.

[11] P. Zhang, C. Lan, W. Zeng, J. Xing, J. Xue, and N. Zheng, "Semantics-guided neural networks for efficient skeleton-based human action recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1112–1121.

[12] H. Chander, R. F. Burch, P. Talegaonkar, D. Saucier, T. Luczak, J. E. Ball, A. Turner, S. N. K. K. Arachchige, W. Carroll, B. K. Smith, A. Knight, and R. K. Prabhu, "Wearable stretch sensors for human movement monitoring and fall detection in ergonomics," *Int. J. Environ. Res. Public Health*, vol. 17, no. 10, p. 3554, May 2020, doi: 10.3390/ijerph17103554.

[13] H. Li, A. Shrestha, H. Heidari, J. Le Kernec, and F. Fioranelli, "Bi-LSTM network for multimodal continuous human activity recognition and fall detection," *IEEE Sensors J.*, vol. 20, no. 3, pp. 1191–1201, Feb. 2020, doi: 10.1109/JSEN.2019.2946095.

[14] X. Xi, M. Tang, S. M. Miran, and Z. Luo, "Evaluation of feature extraction and recognition for activity monitoring and fall detection based on wearable sEMG sensors," *Sensors*, vol. 17, no. 6, p. 1229, May 2017, doi: 10.3390/s17061229.

[15] A. S. Syed, D. Sierra-Sosa, A. Kumar, and A. Elmaghraby, "A deep convolutional neural network-XGB for direction and severity aware fall detection and activity recognition," *Sensors*, vol. 22, no. 7, p. 2547, Mar. 2022, doi: 10.3390/s22072547.

[16] L. Chen, X. Liu, L. Peng, and M. Wu, "Deep learning based multimodal complex human activity recognition using wearable devices," *Appl. Intell.*, vol. 51, no. 6, pp. 4029–4042, Jun. 2021, doi: 10.1007/s10489-020-02005-7.

[17] T. R. Mim, M. Amatullah, S. Afreen, M. A. Yousuf, S. Uddin, S. A. Alyami, K. F. Hasan, and M. A. Moni, "GRU-INC: An inception-attention based approach using GRU for human activity recognition," *Expert Syst. Appl.*, vol. 216, Apr. 2023, Art. no. 119419, doi: 10.1016/j.eswa.2022.119419.

[18] H. Zhang, Z. Xiao, J. Wang, F. Li, and E. Szczerbicki, "A novel IoT-perceptive human activity recognition (HAR) approach using multihead convolutional attention," *IEEE Internet Things J.*, vol. 7, no. 2, pp. 1072–1080, Feb. 2020, doi: 10.1109/JIOT.2019.2949715.

[19] M. A. Khatun, M. A. Yousuf, S. Ahmed, M. Z. Uddin, S. A. Alyami, S. Al-Ashhab, H. F. Akhdar, A. Khan, A. Azad, and M. A. Moni, "Deep CNN-LSTM with self-attention model for human activity recognition using wearable sensor," *IEEE J. Transl. Eng. Health Med.*, vol. 10, pp. 1–16, 2022, doi: 10.1109/JTEHM.2022.3177710.

[20] Y. Chu, K. Cumanan, S. K. Sankarpandi, S. Smith, and O. A. Dobre, "Deep learning-based fall detection using WiFi channel state information," *IEEE Access*, vol. 11, pp. 83763–83780, 2023, doi: 10.1109/ACCESS.2023.3300726.

[21] X. Wang, Y. Guo, F. Wen, J. He, and T.-K. Truong, "EMVS-MIMO radar with sparse Rx geometry: Tensor modeling and 2D direction finding," *IEEE Trans. Aerosp. Electron. Syst.*, early access, Jul. 21, 2023, doi: 10.1109/TAES.2023.3297570.

[22] Z. Zhang, F. Wen, J. Shi, J. He, and T.-K. Truong, "2D-DOA estimation for coherent signals via a polarized uniform rectangular array," *IEEE Signal Process. Lett.*, vol. 30, pp. 893–897, 2023.

[23] Y. Guo, X. Wang, X. Lan, and T. Su, "Traffic target location estimation based on tensor decomposition in intelligent transportation system," *IEEE Trans. Intell. Transp. Syst.*, early access, Apr. 18, 2022, doi: 10.1109/TITS.2022.3165584.

[24] H. Cui and N. Dahnoun, "High precision human detection and tracking using millimeter-wave radars," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 36, no. 1, pp. 22–32, Jan. 2021, doi: 10.1109/MAES.2020.3021322.

[25] R. Qi, X. Li, Y. Zhang, and Y. Li, "Multi-classification algorithm for human motion recognition based on IR-UWB radar," *IEEE Sensors J.*, vol. 20, no. 21, pp. 12848–12858, Nov. 2020, doi: 10.1109/JSEN.2020.3000498.

[26] Y. Jin, B. Kim, S. Kim, and J. Lee, "Design and implementation of FMCW surveillance radar based on dual chirps," *Elektronika Elektrotechnika*, vol. 24, no. 6, pp. 60–66, Dec. 2018, doi: 10.5755/j01.eie.24.6.22292.
[27] S. Ahmed, J. Park, and S. H. Cho, "FMCW radar sensor based human activity recognition using deep learning," in *Proc. Int. Conf. Electron., Inf., Commun. (ICEIC)*, Feb. 2022, pp. 1–5.
[28] Y. Kim and H. Ling, "Human activity classification based on micro-Doppler signatures using a support vector machine," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 5, pp. 1328–1337, May 2009, doi: 10.1109/TGRS.2009.2012849.
[29] Y. Kim and T. Moon, "Human detection and activity classification based on micro-Doppler signatures using deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 1, pp. 8–12, Jan. 2016, doi: 10.1109/LGRS.2015.2491329.
[30] A. H. Victoria and G. Maragatham, "Activity recognition of FMCW radar human signatures using tower convolutional neural networks," *Wireless Netw.*, pp. 1–17, 2021. [Online]. Available: https://link.springer.com/article/10.1007/s11276-021-02670-7
[31] A. Shrestha, H. Li, J. L. Kernec, and F. Fioranelli, "Continuous human activity classification from FMCW radar with bi-LSTM networks," *IEEE Sensors J.*, vol. 20, no. 22, pp. 13607–13619, Nov. 2020, doi: 10.1109/JSEN.2020.3006386.
[32] Y. He, X. Li, and X. Jing, "A mutiscale residual attention network for multitask learning of human activity using radar micro-Doppler signatures," *Remote Sens.*, vol. 11, no. 21, p. 2584, Nov. 2019, doi: 10.3390/rs11212584.
[33] H. Sadreazami, M. Bolic, and S. Rajan, "Contactless fall detection using time-frequency analysis and convolutional neural networks," *IEEE Trans. Ind. Informat.*, vol. 17, no. 10, pp. 6842–6851, Oct. 2021, doi: 10.1109/TII.2021.3049342.
[34] W. Jiang, Y. Ma, W. Zhuang, Z. Wu, Y. Hua, M. Li, and Z. Wang, "Human activity recognition based on frequency-modulated continuous wave and DenseNet," *J. Comput. Commun.*, vol. 11, no. 7, pp. 15–28, 2023, doi: 10.4236/jcc.2023.117002.
[35] C. Yu, Z. Xu, K. Yan, Y.-R. Chien, S.-H. Fang, and H.-C. Wu, "Noninvasive human activity recognition using millimeter-wave radar," *IEEE Syst. J.*, vol. 16, no. 2, pp. 3036–3047, Jun. 2022, doi: 10.1109/JSYST.2022.3140546.
[36] S. Huan, L. Wu, M. Zhang, Z. Wang, and C. Yang, "Radar human activity recognition with an attention-based deep learning network," *Sensors*, vol. 23, no. 6, p. 3185, Mar. 2023, doi: 10.3390/s23063185.
[37] Z. Wang, S. Huan, L. Wu, Q. Wang, J. Liu, and Z. Hu, "Attention-based vision transformer for human activity classification using mmWave radar," in *Proc. 4th Int. Conf. Video, Signal Image Process.*, Nov. 2022, pp. 128–134, doi: 10.1145/3577164.3577184.
[38] M. Ash, M. Ritchie, and K. Chetty, "On the application of digital moving target indication techniques to short-range FMCW radar data," *IEEE Sensors J.*, vol. 18, no. 10, pp. 4167–4175, May 2018, doi: 10.1109/JSEN.2018.2823588.
[39] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[40] F. Fioranelli and J. Le Kernec, "Radar sensing for human healthcare: Challenges and results," in *Proc. IEEE Sensors*, Oct. 2021, pp. 1–4.

CONG LI was born in 1999. He received the B.S. degree from the College of Physical Science and Engineering Technology, Yichun University, China, in 2021. He is currently pursuing the M.S. degree in information and communication engineering with Hainan University, Haikou, China. His research interests include radar signal processing and HAR.

XIANPENG WANG (Member, IEEE) was born in 1986. He received the M.S. and Ph.D. degrees from the College of Automation, Harbin Engineering University, Harbin, China, in 2012 and 2015, respectively. He was a full-time Research Fellow with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, from 2015 to 2016. He is currently a Professor with the School of Information and Communication Engineering, Hainan University, Haikou, China. He is the author of more than 100 papers published in related journals and international conference proceedings. His major research interests include communication systems, array signal processing, radar signal processing, and compressed sensing and its applications. He was a reviewer of more than 30 journals.

JINMEI SHI received the B.Sc. degree from Hainan Normal University, in 2009, the M.Sc. degree from Jiangxi Normal University, in 2015, and the Ph.D. degree from Universiti Sabah Malaysia, in 2022. She is a Professor with the School of Information Engineering, Hainan Vocational University of Science and Technology. Her current research works focus on network traffic prediction, intelligent optimal control, and machine learning and their application. In recent years, she has published 16 papers about swarm intelligence algorithm and network traffic prediction, five sponsored projects, three books, and 21 awards above the provincial level.

HAN WANG (Senior Member, IEEE) received the B.S. degree in electrical engineering from the Hubei University of Nationalities, China, in 2009, and the M.S. and Ph.D. degrees in information and communication system from Hainan University, Haikou, China, in 2013 and 2017, respectively. He was with China Mobile, Jiangxi Branch, as a Network Engineer, in 2014. He was a Postdoctoral Research Fellow with the City University of Macau, from June 2020 to May 2022. Currently, he is an Associate Professor with Yichun University. He is the author of over 50 papers published in related international conference proceedings and journals. His current research interests include intelligent communication, MIMO communications, and information theory.

LIANGTIAN WAN (Member, IEEE) received the B.S. and Ph.D. degrees from the College of Information and Communication Engineering, Harbin Engineering University, Harbin, China, in 2011 and 2015, respectively. From October 2015 to April 2017, he was a Research Fellow with the School of Electrical Engineering, Nanyang Technological University, Singapore. He is currently an Associate Professor with the School of Software, Dalian University of Technology, China. His current research interests include graph learning, data science, and array signal processing. He has been serving as an Associate Editor for IEEE Access and *Journal of Information Processing Systems*.