Progress In Electromagnetics Research M, Vol. 113, 101–113, 2022

Human Motion Recognition in Small Sample Scenarios Based on GAN and CNN Models

Ying-Jie Zhong and Qiu-Sheng Li*

Abstract—In the research of radar-based human motion classification and recognition, the traditional manual feature extraction is more complicated, and the echo data set is generally smaller. In view of this problem, a method of human motion recognition in small sample scenarios based on Generative Adversarial Network (GAN) and Convolutional Neural Network (CNN) models is proposed. First, a 77 GHz millimeter wave radar data acquisition system is built to obtain echo data. Secondly, the collected human motion echo data is preprocessed, the micro-Doppler features are extracted, and the range Doppler map (RDM) is used to project the velocity dimension and accumulate the two-dimensional micro-Doppler time-frequency map data set of the human motion frame by frame. Finally, the deep convolution generative adversarial network (DCGAN) is constructed to achieve data augmentation of the sample set, and the CNN is constructed to realize automatic feature extraction to complete the classification recognition of different human motions. Experimental studies have shown that the combination of GAN and CNN can achieve effective recognition of daily human motions, and the recognition accuracy can reach 96.5%. Compared with the manual feature extraction, the recognition accuracy of CNNs is improved by 7.3%. Compared with the original data set, the system recognition accuracy based on the sample augmentation data set is improved by 2.17%, which shows that the GAN has an excellent performance in human motion recognition in small sample scenarios.

1. INTRODUCTION

Human motion recognition research based on sensors [1] and video data [2] has been developed rapidly and widely used in safety intelligent monitoring, somatosensory games, automatic driving, etc. However, this passive imaging method is greatly affected by the environment, and the recognition rate fluctuates greatly under camouflage conditions. Radar technology has the advantages of all-day, all-weather, and strong penetration ability, and its application in human motion recognition has received more and more attention. Among them, human motion recognition based on radar target micro-Doppler effect [3] is an important research direction. However, in practice, using radar to collect target echoes is time-consuming and labor-intensive, and radar data samples are generally small in scale. Therefore, further research on the expansion algorithm of radar sample data sets is undoubtedly of great significance to the research on human motion recognition.

The performance of radar-based human action recognition schemes largely depends on the effective extraction of micro-motion features. At present, most of them use traditional machine learning methods, rely on statistical theory, and rely on shallow features extracted from raw echo data for recognition. Reference [4] uses Principal Component Analysis (PCA) to extract features from the raw echo signal and combines it with support vector machine (SVM) for motion recognition and classification. Reference [5] extracts the weighted distance-time-frequency spectra, uses PCA for feature selection and reduction, and

Received 2 July 2022, Accepted 25 August 2022, Scheduled 13 September 2022
* Corresponding author: Qiu-Sheng Li (bjliqiusheng@163.com).
The authors are with the Research Center of Intelligent Control Engineering Technology, School of Physics and Electronic Information, Gannan Normal University, Ganzhou, Jiangxi 341000, China.

performs performance verification through SVM. Reference [6] uses PCA to analyze the characteristics of the raw echo signal and uses the improved SVM to classify and recognize human motions.

Deep learning can learn the essence of data from a large amount of data. It has made breakthrough progress in image classification [7], speech processing [8], text classification [9], etc. At present, deep learning has also begun to be applied to the classification of micro-motion signals. Kim and Moon of the University of California first used CNNs for human object detection and human gait classification [10], as well as the recognition of different swimming styles and ships underwater [11]. Their research work shows that deep learning has important potential in solving human micro-motion feature extraction and target classification and recognition.

However, although deep learning has the advantages of strong learning ability and high accuracy of target classification and recognition, it requires a large amount of training data. In addition, it is difficult and expensive to obtain labeled samples for human motion recognition, and overfitting is prone to occur when training in scenarios with a lack of data. Therefore, in recent years, many studies have focused on expanding datasets, mainly including random cropping, rotation, flipping, etc. However, for the micro-Doppler time-frequency map with different human motion features, the traditional extension method will change the feature semantics, making the motion change in time and space, and cannot describe the real motion information. This requires the study of new learning strategies to expand the sample size, effectively utilize prior knowledge, and achieve data augmentation with fewer labeled samples. Generative Adversarial Networks (GANs) have been used in dataset generation in recent years [12–14]. Therefore, in order to construct a reasonable network model, it is necessary to design effective learning strategies and make better use of prior knowledge to achieve effective recognition of human motions. Using GAN to construct a network model that describes the small differences of human micro-motion signals and characterizes the time series characteristics is an effective way to use. This paper intends to explore a new method for human motion recognition based on deep convolutional generative adversarial networks (DCGANs) by using millimeter-wave radar.

2. THEORETICAL BASIS

2.1. Millimeter Wave Radar Echo Model

Different from the traditional pulse radar system that periodically transmits short pulses, this paper adopts a 77 GHz frequency-modulated continuous wave (FMCW) radar to measure the distance and speed of the target by continuously transmitting a frequency-modulated signal (Chirp signal) whose signal frequency increases linearly with time [15]. At the same time, the Short Range Radio (SRR) band in the 77 GHz frequency band can provide a scanning bandwidth up to 4 GHz, and the range resolution and accuracy are significantly improved, which is conducive to the capture of subtle human movements. The radar uses sine wave to modulate the carrier frequency. At this time, the radar transmission frequency can be expressed as

$$f_t = f_0 + \frac{\Delta f}{2}\cos 2\pi f m_t \tag{1}$$

The transmitted signal can be expressed as

$$u_t = U_t \sin\left(2\pi f_0 + \frac{\Delta f}{2f_m}\sin 2\pi f_m t\right) \tag{2}$$

At this time, the echo signal reflected by the target is expressed as

$$u_r = U_r \sin\left[2\pi f_0(t-T) + \frac{\Delta f}{2f_m}\sin 2\pi f_m(t-T)\right] \tag{3}$$

where f_m is the modulation frequency, Δf the frequency offset, and $T = 2R/c$ the residence time. The intermediate frequency (IF) signal is obtained by subtracting the transmitted signal from the received signal in the mixer. Therefore, the differential frequency voltage is expressed as follows:

$$u_b = kU_tU_r \sin\left\{\frac{\Delta f}{f_m}\sin \pi f_m T * \cos[2\pi f_m(t-T) + 2\pi f_0 T]\right\} \tag{4}$$

2.2. Micro-Doppler Effect Analysis

The radar transmits electromagnetic signals to the target through the transmitting antenna, while the receiving antenna receives the target echo. If the target is in motion, the frequency of the received signal will change and deviate from the frequency of its transmitted signal, which is called "Doppler effect" [16]. The degree of Doppler shift depends on the radial velocity of the moving radar target. Generally speaking, the radial velocity v of the target relative to the radar is much less than the electromagnetic wave propagation velocity c. When there is movement between the moving target and the observation radar, the distance between them at time t is

$$R(t) = R_0 - vt \tag{5}$$

The time delay can be expressed as

$$T = \frac{2R(t)}{c} = \frac{2}{c}(R_0 - vt) \tag{6}$$

High frequency phase difference is

$$\varphi = -2\pi\frac{2}{\lambda}(R_0 - vt) \tag{7}$$

Therefore, the total Doppler shift of the target is

$$f_D = \frac{1}{2\pi}\frac{\mathrm{d}\varphi}{\mathrm{d}t} = \frac{2}{\lambda}v = f\frac{2v}{c} \tag{8}$$

The movement of the main body of the target generates Doppler offset frequency. If any part of the target has small movement, the micro-motion will induce additional frequency modulation on the radar echo, thus resulting in signal side frequencies. This additional Doppler modulation is called "micro-Doppler effect". Therefore, micro-Doppler effect refers to the physical phenomenon that the vibration, rotation, and other small movements in the radar target or target structure produce Doppler frequency modulation in the radar echo signal.

2.3. GAN

GAN is a new deep network framework. It abandons the traditional way of improving learning ability through the simple stacking of network layers. It has two parts, the generator and discriminator. It uses their "countermeasure game", that is, the discriminator feeds back the difference between the identified real samples and the generated samples to the generator, improves the generator to continuously improve its ability to "forge" data, and finally realizes that it is difficult for the discriminator to distinguish the real and generated data. The structure of the GAN model is shown in Figure 1. Random noise and latent variables are input into generator G in the GAN model to generate realistic samples. The distribution of the samples is as consistent as possible with the distribution of the real micro-motion signals. The real data and generated data are input into the discriminator D at the same time to discriminate the authenticity; the discriminant result is fed back to the generator; and the training is repeated in a loop

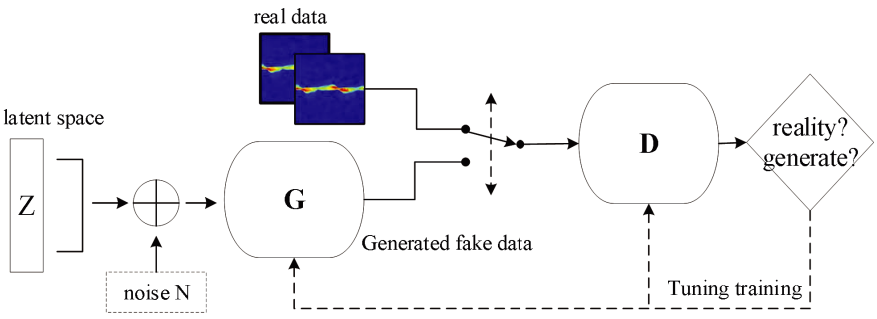

Figure 1. The overall framework of the GAN model.

until the Nash equilibrium is reached. The training of GAN can be regarded as a minimax optimization process, as shown in Equation (9).

$$\min_G \max_D V(D,G) = E_{x\sim P_{data}(x)}[\log D(x)] + E_{z\sim P_z(z)}[\log(1 - D(G(z)))] \tag{9}$$

where x represents the real data, z the random noise, $P_{data}(x)$ the distribution of real data, and $P_z(z)$ the prior distribution of random noise.

$V(D,G)$ is a binary cross entropy function. The discriminator D attempts to maximize $V(D,G)$ while the generator G attempts to minimize it, thus forming a minimax relationship, which is solved by alternately performing the following two gradient updates.

$$\theta_D^{t+1} = \theta_D^t + \lambda^t \nabla_{\theta_D} V(D^t, G^t) \tag{10}$$

$$\theta_G^{t+1} = \theta_G^t + \lambda^t \nabla_{\theta_G} V(D^{t+1}, G^t) \tag{11}$$

3. MODEL BUILDING

3.1. Extraction of Human Micro-Doppler Features

Human walking is formed by the association of scattering points of various parts of the body, and its motion forms are extremely rich, which is a typical articulated non-rigid body motion. When the radar detects the human target, its echo signal contains the mixed Doppler information modulated by the human body's translation and non-rigid micro-motion, and the expression is also extremely complex. In the research of human motion recognition, the micro-Doppler effect is considered to be the unique reflection of human target movement characteristics in radar echoes and is a unique feature that marks human targets. Nowadays, the effective extraction of micro-motion features of radar targets and the analysis of micro-Doppler effects have become a research hotspot in the field of radar technology [17]. Aiming at the problem of human motion recognition, this paper preprocessed the echo of human behavior to extract the unique features that reflect different human motions, and then constructs a micro-Doppler time-spectrogram dataset. Figure 2 shows the processing flow of radar echo data.

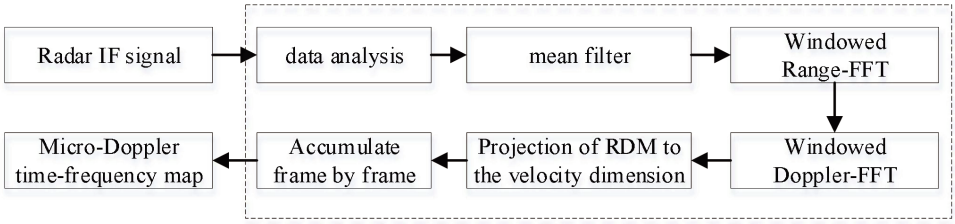

Figure 2. Data processing flow.

For radar data, the IF signal after mixing is sampled at the mth sampling point of the nth cycle to obtain a discrete IF signal, which is parsed into a two-dimensional data matrix $\mathbf{R}[m,n]$, where m is the sampling points in a frequency modulation cycle, that is, m is the fast time sampling point, and n is the total number of frames in a sampling period, i.e., n is the slow time sampling point.

$$\mathbf{R} = [\ \mathbf{r}_1 \quad \dots \quad \mathbf{r}_2 \quad \dots \quad \mathbf{r}_n\ ] \tag{12}$$

$$\mathbf{r}_n = [\ r_{1,n} \quad \dots \quad r_{2,n} \quad \dots \quad r_{m,n}\ ]^T \tag{13}$$

After the radar echo is effectively analyzed, in order to effectively eliminate the influence of the other clutter in the detection environment, it is necessary to suppress the echo data. In this paper, the mean value filtering is used to process the echo. This method is mature in technology and simple in principle. The time-frequency diagram comparison after mean filtering is shown in Figure 3. It can be seen from Figure 3 that the preprocessing reduces the blurring and covering, and the motion features in the time-frequency map are clearer.

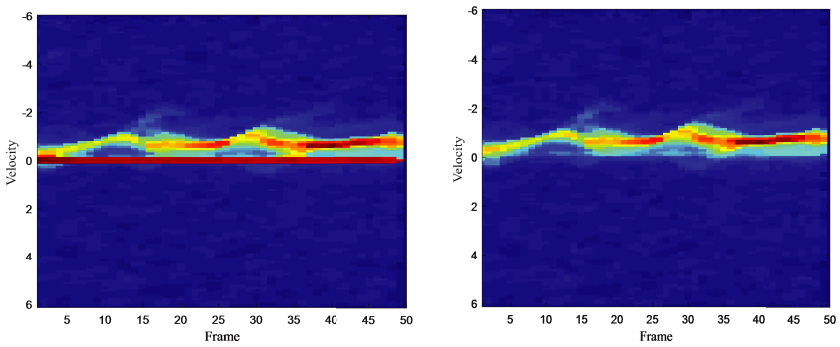

Figure 3. Comparison of clutter suppression.

Using the mean filtering processing matrix $\mathbf{R}$, firstly average all received signals $\mathbf{r}_n$, and then subtract the mean reference received signal from each received signal to obtain the target echo signal after clutter suppression. The mathematical expression can be expressed as

$$\mathbf{R}[m,n] = \mathbf{R}[m,n] - \mathbf{K}[m] \tag{14}$$

where $\mathbf{K}[m]$ is the mean value reference received signal, expressed as follows:

$$\mathbf{K}[m] = \frac{1}{N}\sum_{n=1}^{N}\mathbf{R}[m,n] \tag{15}$$

Based on the obtained data matrix $\mathbf{R}$, the windowed range Fast Fourier Transform (FFT) is performed after adding Hamming window to the fast time dimension to obtain high-resolution one-dimensional range information. And then the Doppler FFT is performed in the slow time dimension with a Hamming window to obtain a Range-Doppler Map (RDM). RDM effectively extracts the range and velocity related features of all scattering points of human targets in the radar signal frame. Different RDMs can be obtained based on the differences between different human movements.

For the echo processing of a single human target in this paper, it is assumed that $RD(i,j,t)$ is used to represent the signal power value of the tth frame of the RDM at the ith distance gate and the jth speed gate. Then, the extracted RDM of human motion is further projected to the velocity dimension and accumulated column by column to obtain the micro-Doppler time-spectrogram generated by the target movement, as shown in Figure 4. The horizontal axis represents the frame number corresponding to the echo signal, and the vertical axis represents the corresponding speed under the speed gate label. Therefore, the micro-Doppler frequency can be expressed as

$$DP(t,j) = \sum_{i} RD(i,j,t) \tag{16}$$

3.2. Data Augmentation Recognition Model Based on DCGAN

Human motion recognition based on Radar generally has the problem that the echo dataset is too small. Therefore, it is necessary to expand the number of training samples and improve the recognition accuracy. At present, the commonly used data augmentation methods include rotation, mirroring, local cropping, etc. The rotation of the time-spectrogram will cause the direction of the motion feature to be opposite to the real motion, which cannot be applied to the recognition experiment. Also, mirroring will cause the time-varying micro-Doppler frequency to reverse the time sequence, so it is not suitable for the human motion recognition. Local cropping can easily lead to the lack of complete motion information. In the cropping process, it is necessary to strictly control the region to obtain complete information of a motion, and the operation is complicated. Therefore, this paper uses a GAN to expand the sample dataset. Figure 5 shows the specific flowchart of the recognition model, that is, the dataset is first expanded by using DCGAN and then input to the CNN for classification and recognition.

DCGAN is an improvement on the basis of GAN. It uses CNN to replace the multi-layer perceptron in the GAN discriminator and generator to improve the stability of the model. In the discriminator, a

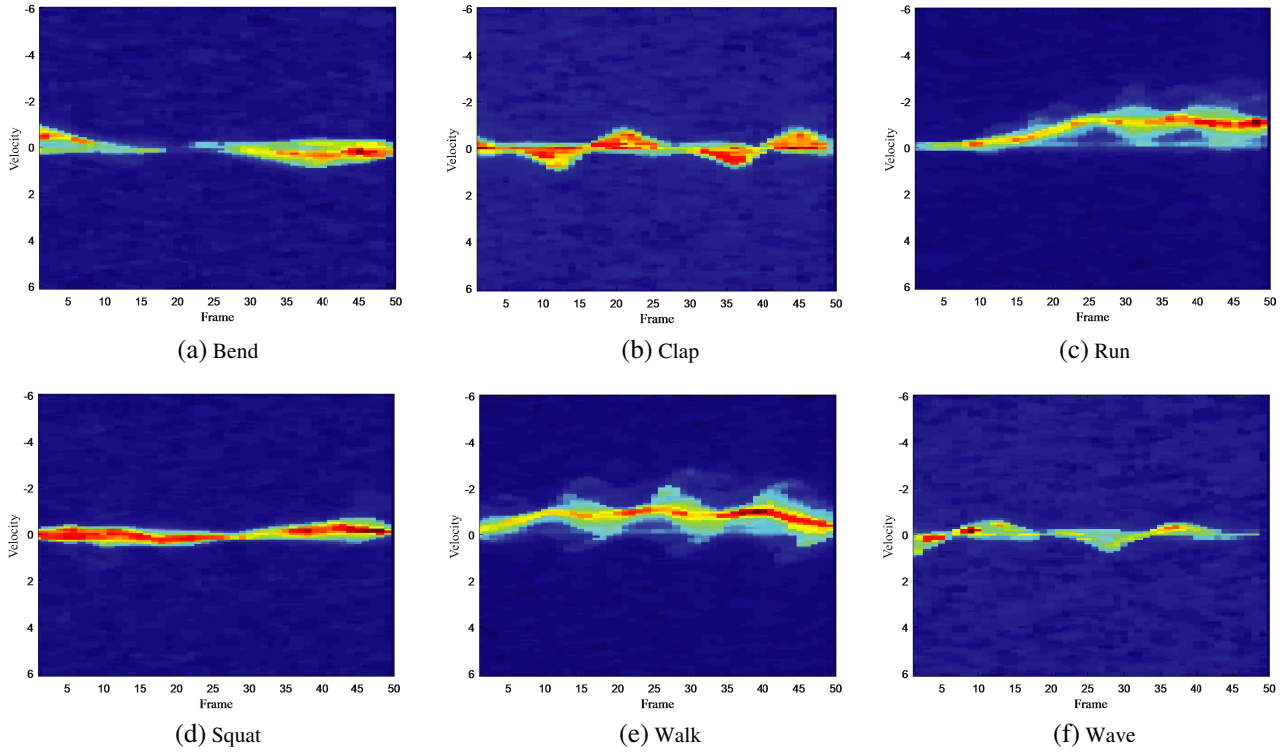

Figure 4. Micro-Doppler time-frequency map of different human motions.

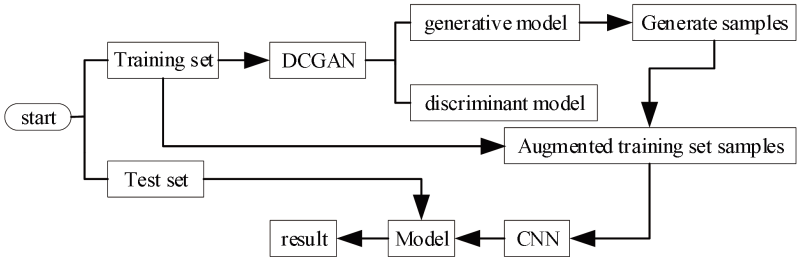

Figure 5. Flow chart of DCGAN and CNN model construction.

convolutional layer is used to extract features; in the generator, a transposed convolutional layer is used to restore the information in the image. The structures of the DCGAN generator and discriminator are shown in Figure 6 and Figure 7.

As can be seen from Figure 6, the generator network structure consists of a total of 5 layers of transposed convolutions. The convolution kernel size of the first four layers in the generator network is set to 4×4, and the last layer is set to 5×5. The convolution stride is [1, 2, 2, 2, 3]. The input to the generator network is random normally distributed noise with dimension $Z = 100$. The transposed convolutional layer is used for upsampling. Except for the last layer, the rest of the layers are sampled and then processed by Batch Normalization (BN), and processed by the nonlinear Rectified Linear Unit (ReLU) activation function. The Tanh activation function is used behind the last transposed convolutional layer to output a $3 \times 96 \times 96$ RGB image. As can be seen from Figure 7, the discriminator network structure consists of 5 layers. The size of the convolution kernel of the first convolutional layer is 5×5; the remaining convolutional layers is 4×4; and their convolution strides are [3, 2, 2, 2, 1], respectively. BN is performed after each convolution layer, and then processed by the nonlinear LeakyReLU activation function. The last layer uses the Sigmoid activation function, and finally judges the authenticity of the input image. The optimizer selected by DCGAN in this paper is Adam. The

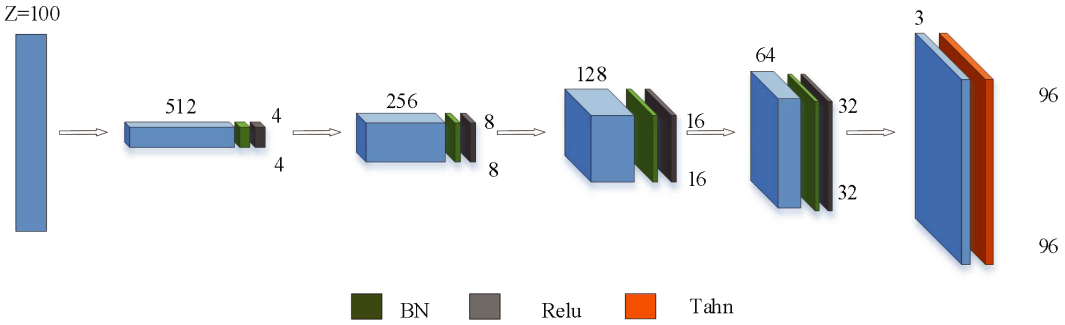

Figure 6. Schematic diagram of the structure of the generator.

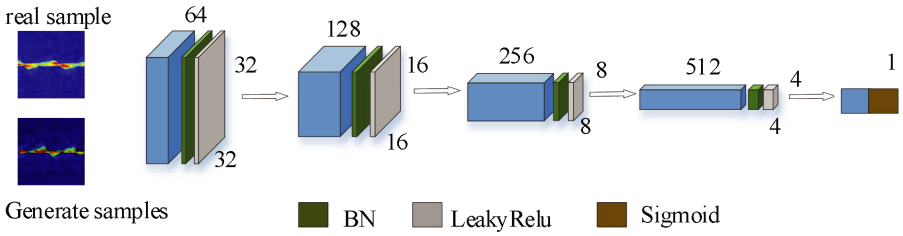

Figure 7. Schematic diagram of the structure of the discriminator.

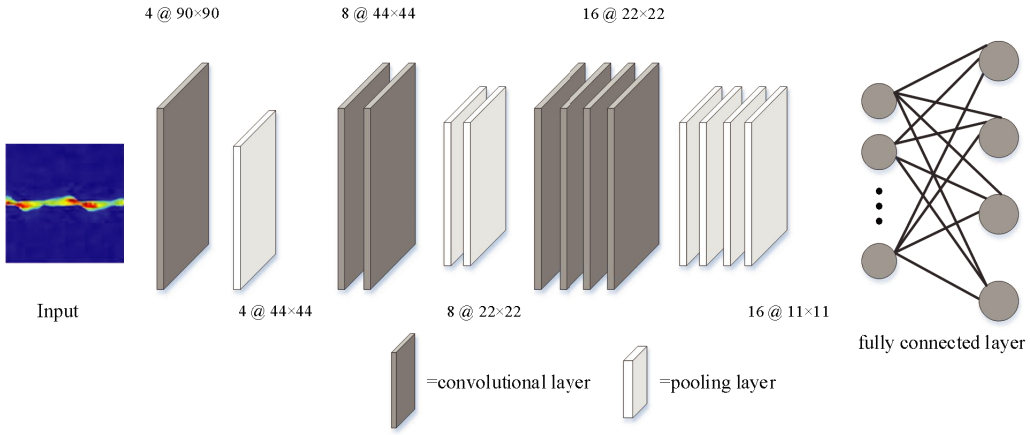

Figure 8. Recognition model based on convolutional neural network.

batch size of the training process is set to 10, the learning rate of the generator and discriminator set to 0.0002, and the total number of training times is 100.

There are various convolutional neural network structures for different classification and recognition applications. The samples of human motion recognition studied in this paper are based on measured radar echo data, which is a small sample dataset, and the feature differences between samples after preprocessing are obvious. Therefore, the constructed convolutional neural network is a small shallow network as shown in Figure 8. The shallow CNN structure is composed of three layers, each of which contains a convolution layer and a maximum pooling layer, and the number of channels in each layer is $[4, 8, 16]$. In this paper, the CNN recognition model adopts the cross entropy loss function; the iteration batch is set to 20; and the total number of training times is 100. Using stochastic gradient descent (SGD) optimizes the model parameters. In the experiment, to achieve a better training effect, the learning rate is adjusted at equal intervals, and the learning rate is set to decrease every epoch. At this time, the learning rate decay factor is set to 0.98.

4. EXPERIMENTAL RESULTS AND ANALYSIS

4.1. Experimental Setup and Data Acquisition

The experiment uses the industrial millimeter wave radar sensor IWR1642 and DCA1000 data acquisition board to build a real-time data acquisition platform to obtain the radar echo data of human motion. IWR1642 radar equipment development platform is equipped with PC (Personal Computer) control terminal mmWave Studio, which can realize real-time acquisition, transmission processing, and visualization of target echoes. The acquisition module of this radar equipment is shown in Figure 9. The mmWave sensor IWR1642 has a continuous bandwidth of up to 4 GHz covering 76 to 81 GHz, uses a low-power 45 nm Radio-Frequency Complementary Metal-Oxide Semi-Conductor (RFCMOS) process, and its TX Power (transmission power) is 12.5 dBm, making it an ideal solution for low-power radar systems. The IWR1642 continuously emits FMCW forward to capture radar echoes of moving humans in its path. The echo contains the speed and distance information of the human target. After preprocessing the echo, the two-dimensional micro-Doppler time spectrum of human action can be obtained. The performance and application scenarios of the radar system depend on the different parameters of the radar. The main indicators that determine its performance are the Maximum Unambiguous Range ($R_{\max}$), the Maximum Unambiguous Speed ($V_{\max}$), the Range Resolution (ΔR), and the Velocity Resolution (ΔV). Their expression is as follows

$$R_{\max} = \frac{F_s c}{2S} \tag{17}$$

$$V_{\max} = \frac{\lambda}{4T_c} \tag{18}$$

$$\Delta R = \frac{c}{2B} \tag{19}$$

$$\Delta V = \frac{\lambda}{2NT_c} \tag{20}$$

where F_s is the Analog to Digital Converter (ADC) sampling rate, S the frequency modulation (FM) slope, T_c the frequency modulation period, B the bandwidth, and N the number of pulses per frame. In order to ensure the validity of the training data, the data collection work in this experiment is selected to be carried out in an outdoor semi-open open environment, and there is no other moving target interference in the environment. The radar antenna is fixed on a tripod with a height of 0.9 m. The measured target is 2.5 m away from the radar and moves radially relative to the radar. Figure 10 shows the data acquisition site. In the experiment, the radar is set as a transmitter-receiver type, and the radar parameter settings are shown in Table 1. The number of tested human targets is five people. Choose six kinds of human motions, which are: (1) bending; (2) clapping; (3) running; (4) squatting; (5) walking; (6) waving. To ensure the integrity of the acquisition actions, the acquisition duration is set to 2 s. The six human motions belong to daily behaviors, and these motions have both similarities and great differences in range profiles. The target moves towards the radar, and each motion is repeated 50 times.

Table 1. Radar equipment parameter settings.

Parameter	Value
FM start frequency (GHz)	77
FM slope (MHz/μs)	64
ADC sampling points	256
Pulses per frame	128
Sampling Rate (msps)	5.12

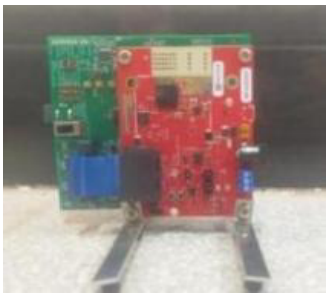

Figure 9. Radar equipment.

Figure 10. Data collection site.

4.2. Construction of Experimental Sample Set

In the experiment, millimeter wave radar echoes sample data were acquired for each human motion. In the experiment, 250 sets of radar echo data have been collected for each different human motion. After the micro-Doppler feature extraction is performed on the raw echo data, a two-dimensional micro-Doppler spectrogram containing motion features is obtained; 150 spectrograms are randomly selected as the training sample set $\mathbf{D}_1$; and the remaining 100 spectrograms are used as the test sample set $\mathbf{D}_2$. On the basis of $\mathbf{D}_1$, the constructed DCGAN is used to generate the time-spectrograms of 6 different types of human motions at a ratio of 1 : 1 to obtain the extended sample training set $\mathbf{D}_3$. The specific information statistics of the sample sets are shown in Table 2.

Table 2. Statistics of sample set information.

Human motions	Label	Training sample set $\mathbf{D}_1$	Test sample set $\mathbf{D}_2$	Augmented training sample set $\mathbf{D}_3$
bending	bend	150	100	300
clapping	clap	150	100	300
running	run	150	100	300
squatting	squat	150	100	300
walking	walk	150	100	300
waving	wave	150	100	300

The sample sets in Table 2 are randomly scrambled. From the statistical information of training samples, test samples, and extended samples in Table 2, it can be seen that the total number of samples in $\mathbf{D}_1$ is 900; the total number of samples in $\mathbf{D}_2$ is 600; and the number of samples in $\mathbf{D}_3$ is 1800.

4.3. Analysis of Experimental Results

4.3.1. Model Evaluation

The DCGAN and CNN models in this paper are trained using the processed sample set. Use the augmented training sample set $\mathbf{D}_3$ to train the neural network, save the training model, and test the performance of the designed model using the test sample set $\mathbf{D}_2$ that is not involved in the training. Define Accuracy *acc* as:

$$acc = \frac{P_{\text{true}}}{P_n} \tag{21}$$

where P_{true} is the number of all correctly classified samples, and P_n is the total number of samples. Figure 11 shows the change curves of the recognition accuracy of the training set and the test set in the experiment. The prediction accuracy of the test set is about 96%.

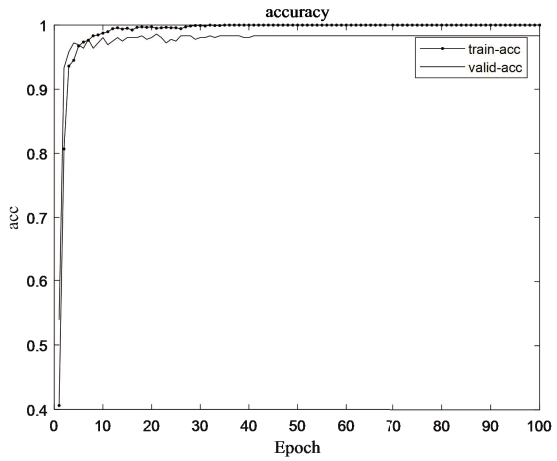

Figure 11. Accuracy change curves.

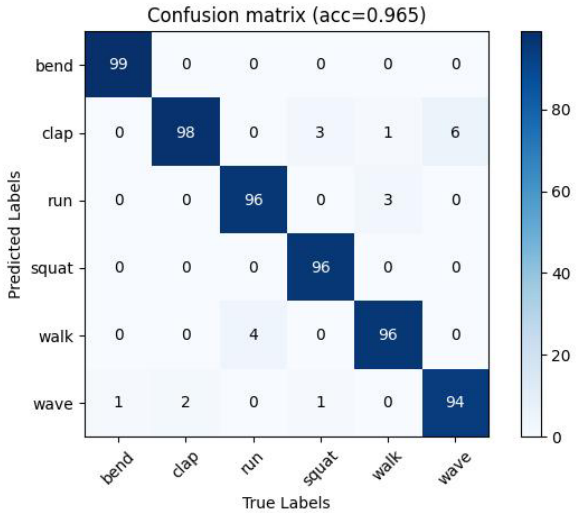

Figure 12. Confusion matrix.

The confusion matrix is a visual representation of what is correctly identified and what is incorrectly identified by the network. Each column element of the matrix represents the actual type of the input time-spectrogram; each row element represents the model's predicted type; and the diagonal elements represent correctly identified cases. In this paper, the confusion matrix evaluation model is used to obtain the results shown in Figure 12. Observing Figure 12, it can be seen that the values on the diagonal are generally higher, indicating that the proposed network model can correctly identify most of the time-spectrograms in various motions. The accuracy of the model recognition rate is 96.5%. Figure 12 also shows that two motions, clapping and waving, are prone to misrecognition, followed by the two motions of running and walking, while the proposed model has higher recognition accuracy for other daily motions.

There are four indicators in the classification task, namely: True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN). Indicators such as precision, recall, and specificity are commonly used to evaluate the performance of the model. These performance indicators are defined as follows:

(1) Precision: represents the probability that the predicted positive samples are actually positive samples, and its expression is:

$$precision = \frac{T_P}{T_P + F_P} \tag{22}$$

(2) Recall: represents the probability that the actual positive sample is predicted to be a positive sample, and its expression is:

$$recall = \frac{T_P}{T_P + F_N} \tag{23}$$

(3) Specificity: represents the proportion of predicted negative samples to actual negative samples, and its expression is:

$$specificity = \frac{T_N}{F_P + T_N} \tag{24}$$

where T_P and T_N represent the number of each type's human actions correctly predicted as this type of action. F_P is the number of other actions incorrectly identified as this type of action, and F_N is the number of this type's actions incorrectly predicted as other actions. Table 3 shows the evaluation results of the model using the above three indicators. It shows that the model presented in this paper shows good performance.

Table 3. Model evaluation indicators.

Tag category	Precision	Recall	Specificity
bend	1.0	0.99	1.0
clap	0.907	0.98	0.98
run	0.97	0.96	0.994
squat	1.0	0.96	1.0
walk	0.96	0.96	0.992
wave	0.959	0.94	0.992

4.3.2. Comparison with Manually Extracted Features

In order to make an objective and fair evaluation of the system that uses CNN to solve human motion recognition in this paper, this system is compared with different human posture recognition schemes, such as the most commonly used and classic SVM recognition methods. In the following experiment, we will analyze the performance of the CNN-based classification method proposed in this paper with the PCA+SVM-based human motion recognition method adopted in [5] as a contrast. Table 4 shows the comparison of accuracies of the two methods. It can be seen from the table that the accuracy of the deep model based on CNN is much higher than that of SVM, and the accuracy of different motions by the CNN-based method is nearly 92%. The PCA+SVM-based method has slightly better classification performance for waving than the CNN method in this paper, and the accuracies for the other five motions (bending, clapping, running, squatting, and walking) does not exceed the CNN method, especially for bending and squatting. Therefore, the overall performance of the CNN model in this paper is significantly better than the SVM method.

Table 4. Comparison of the Accuracies of PCA+SVM and CNN/%.

Compare items	PCA+SVM	CNN
bend	80	99
clap	83	90
run	91	97
squat	82	92
walk	93	96
wave	93	92
Average Accuracy	87	94.3

4.3.3. Model Validation

For the model test of data augmentation, this paper chooses to compare and analyze the CNN network without data augmentation and the DCGAN-CNN model after data augmentation using DCGAN. The confusion matrix for classification recognition using only the CNN model is shown in Figure 13. In the contrast experiment, the uniform iteration number is set to 100, and the two different classification methods are discussed, respectively. Figure 14 shows the accuracy change curves of the training set and test set of the two classification methods. The comparison of precision and recall of the models is shown in Table 5.

It can be seen from Figure 12 to Figure 14 that the accuracy of using only the CNN model is 94.33%, while the accuracy of the model after data augmentation reaches 96.5%, which is 2.17% higher than that of the model without data augmentation. It can be seen from Table 5 that the precision of bending, running, squatting, and waving has been improved, and better classification performance

Table 5. Model performance comparison.

Comparative indicators	Model	bend	clap	run	squat	walk	wave
precision	CNN-DCGAN	1.0	0.907	0.97	1.0	0.96	0.959
	CNN	0.99	0.909	0.96	0.979	0.97	0.86
recall	CNN-DCGAN	0.99	0.98	0.96	0.96	0.96	0.94
	CNN	0.99	0.90	0.97	0.92	0.96	0.92

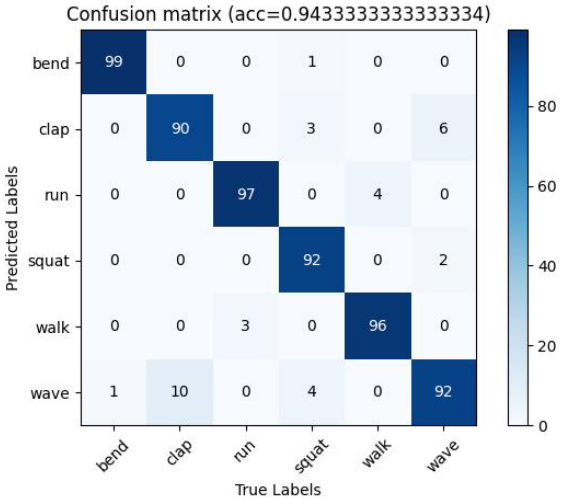

Figure 13. CNN identification confusion matrix.

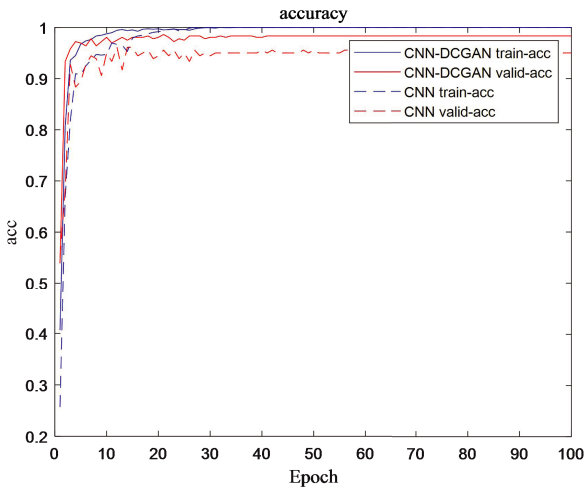

Figure 14. Comparison of training results.

has been obtained, especially the precision of waving. For the two types of motions, clapping and walking, the precision of the data-augmented model is slightly lower than that of the model without data augmentation, but the recall rate of the two types of actions is significantly higher than or equal to the model without data augmentation. In general, the recognition experiments after data augmentation using GAN have achieved better results. Experiments show that the CNN-DCGAN model can effectively handle the classification tasks of different daily motions of the human body, so it is an effective way to use GAN to realize the effective expansion of the radar human motion echo dataset.

5. CONCLUSION

In this paper, a human motion recognition model of millimeter wave radar based on the GAN and CNN in small sample scenes is proposed, which can recognize human motions on the basis of effective data augmentation. By building a 77 GHz millimeter wave radar data acquisition system, the human motion radar echo data set is obtained to effectively extract the micro-Doppler information of complex human motion echo. At the same time, DCGAN is constructed to enhance the sample data, and the CNN is combined to recognize different human motions. Compared with artificial feature extraction, the recognition accuracy of CNN is improved by 7.3%. Compared with the original data set without sample augmentation, the recognition accuracy of the system based on sample augmentation data set is improved by 2.17%. The above results prove that the recognition model proposed in this paper has high practical value in improving the recognition accuracy and complexity. However, the experimental simulation is set up in an ideal environment without other target interference. In the follow-up research and practical application, it is also necessary to consider the recognition and detection of human motions in complex environments.

REFERENCES

1. Deng, P. and M. Wu, "Human motion and gesture recognition method based on machine learning," *Chinese Journal of Inertial Technology*, Vol. 30, No. 1, 37–43, 2022.
2. Luo, H., K. Tong, and F. Kong, "Review of human action recognition in video based on deep learning," *Journal of Electronic Arts*, Vol. 47, No. 5, 1162–1173, 2019.
3. Ding, Y., R. Liu, and X. Xu, "Micro-Doppler frequency estimation method for human target based on continuous wave radar," *Journal of Central South University (Natural Science Edition)*, Vol. 53, No. 4, 1273–1280, 2022.
4. Bryan, J. D., J. Kwon, N. Lee, et al., "Application of ultra-wide band radar for classification of human activities," *IET Radar, Sonar & Navigation*, Vol. 6, No. 3, 172–179, 2012.
5. Ding, C., L. Zhang, C. Gu, et al., "Non-contact human motion recognition based on UWB radar," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, Vol. 8, No. 2, 306–315, 2018.
6. Jiang, L., X. Zhou, and L. Che, "Small-sample human action recognition based on carrier-free ultra-wideband radar," *Journal of Electronic Engineering*, Vol. 48, No. 3, 602–615, 2020.
7. Zheng, Y., G. Li, and Y. Li, "A review of the application of deep learning in image recognition," *Computer Engineering and Applications*, Vol. 55, No. 12, 20–36, 2019.
8. Li, A., M. Yuan, C. Zheng, et al., "Speech enhancement using progressive learning-based convolutional recurrent neural network," *Applied Acoustics*, Vol. 166, 107347, 2020.
9. Prabhakar, S. K., D.-O. Won, and Y. Maleh, "Medical text classification using hybrid deep learning models with multihead attention," *Computational Intelligence and Neuroscience*, Vol. 2021, 9425655, 2021.
10. Kim, Y. and T. Moon, "Human detection and activity classification based on micro-Doppler signatures using deep convolutional neural networks," *IEEE Geoscience and Remote Sensing Letters*, Vol. 13, No. 1, 8–12, 2016.
11. Park, J., R. J. Javier, and Y. Kim, "Micro-Doppler based classification of human aquatic activities via transfer learning of convolutional neural networks," *Sensors*, Vol. 16, No. 12, 1990, 2016.
12. Sun, X., K. Zhou, S. Shi, et al., "A new cyclical generative adversarial network based data augmentation method for multiaxial fatigue life prediction," *International Journal of Fatigue*, 162, 2022.
13. Jin, H., Y. Li, J. Qi, et al., "GrapeGAN: Unsupervised image enhancement for improved grape leaf disease recognition," *Computers and Electronics in Agriculture*, 198, 2022.
14. Alnujaim, I., D. Oh, and Y. Kim, "Generative adversarial networks for classification of micro-Doppler signatures of human activity," *IEEE Geoscience and Remote Sensing Letters*, Vol. 17, No. 3, 396–400, 2020.
15. Cha, D., S. Jeong, M. Yoo, et al., "Multi-input deep learning based FMCW radar signal classification," *Electronics*, Vol. 10, 1144, 2021.
16. Chen, V. C., D. Tahmoush, and W. J. Miceli, "Radar micro-doppler signatures: Processing and applications," *IET Digital Library*, 406, 2014.
17. Jin, T., Y. He, X. Li, et al., "Research progress on human behavior perception by ultra-wideband radar," *Journal of Electronics and Information*, Vol. 44, No. 4, 1147–1155, 2022.