(Received 4 March 2025, Accepted 14 April 2025, Scheduled 22 April 2025)

Human Action Recognition in Small-Sample Scenarios Based on Improved DCGAN and CNN Models

Cheng Luo[1], Qiusheng Li[2,*], and Yingjie Zhong[2]

[1] *Shenzhen Polytechnic University, Shenzhen 518107, China*
[2] *School of Physics and Electronic Information, Gannan Normal University, Ganzhou 341000, China*

ABSTRACT: Aiming at the prevalent issue of limited dataset scale in radar target micro-Doppler effect-based human action recognition, this study constructs an improved Deep Convolutional Generative Adversarial Network (DCGAN) for radar sample data augmentation and integrates it with a Convolutional Neural Network (CNN) for human action classification. First, a millimeter-wave radar data acquisition system was established to collect human action echo signals. The raw data were preprocessed to extract micro-Doppler features, forming a 2D micro-Doppler time-frequency spectrogram dataset. Second, modifications were made to the original DCGAN by replacing the optimizer of the discriminator network and introducing an L2 regularization term to enhance the quality of micro-Doppler time-frequency spectrogram generation. Finally, a CNN architecture was implemented to classify the augmented human action samples. Experimental results demonstrate that the enhanced DCGAN-CNN framework achieves robust human action classification performance, achieving an accuracy rate of up to 97.5%. This validates the superior capability of generative adversarial networks in few-shot scenarios for radar-based human action recognition.

1. INTRODUCTION

In recent years, human action recognition has found significant applications in telemedicine, autonomous driving, intelligent security systems, and other domains. Current approaches mainly rely on inertial sensors [1] and video data [2]; however, these methods suffer from poor portability, susceptibility to environmental interference, and privacy concerns. Radar technology, with its advantages of all-weather capability, continuous operation, and strong penetration, has driven growing interest in non-contact radar-based human action recognition [3, 4]. Notably, human action recognition leveraging the micro-Doppler effect of radar targets [5] has emerged as a critical research direction.

Early human action recognition approaches primarily employed Support Vector Machines (SVMs) and Decision Trees (DTs), relying on shallow features such as mean values and variances extracted from raw radar echoes, which introduced feature redundancy and compromised recognition accuracy. With the rapid advancement of deep learning, data-driven feature extraction and classification methods have been increasingly applied to micro-motion signal analysis of human targets. Kim and Moon from the University of California pioneered the application of Convolutional Neural Networks (CNNs) for human target detection and gait classification [6]. Subsequent studies [7–9] have prioritized CNN-based implementations for radar human posture recognition. However, deep learning models typically demand substantial training data, while acquiring labeled samples for radar-based human action recognition remains cost-prohibitive, often resulting in insufficient data volume.

To address data scarcity in small-sample scenarios, conventional strategies employ sample augmentation through image processing techniques including random cropping, rotation, and flipping. Nevertheless, such geometric transformations alter the spatiotemporal characteristics of micro-Doppler time-frequency spectrograms, thereby distorting authentic motion signatures. Generative Adversarial Networks (GANs) offer superior solutions by approximating real sample distributions to synthesize augmented datasets [10]. In particular, study [11] implemented GAN for classifying micro-Doppler images of human activities in radar measurements, while [12] utilized Deep Convolutional Generative Adversarial Networks (DCGANs) to enhance micro-Doppler radar spectrograms. Although these GAN-enhanced methods demonstrated improved recognition rates, our proposed improved DCGAN framework advances radar sample augmentation by optimizing both the quality and diversity of generated time-frequency representations compared to baseline DCGAN architectures.

Based on the comprehensive analysis above, this paper proposes a novel method for human action recognition in small-sample scenarios using millimeter-wave radar, which is developed based on improved DCGAN and CNN. First, a millimeter-wave radar acquisition platform is built to collect radar echoes of six different human actions, followed by preprocessing. Micro-Doppler features are extracted through the method of projecting range-Doppler maps onto the velocity dimension and performing frame-wise accumulation, thereby constructing a two-dimensional micro-Doppler time-frequency spectrogram dataset. Subsequently, an improved DCGAN architecture and

* Corresponding author: Qiusheng Li (liqiusheng@gnnu.edu.cn).

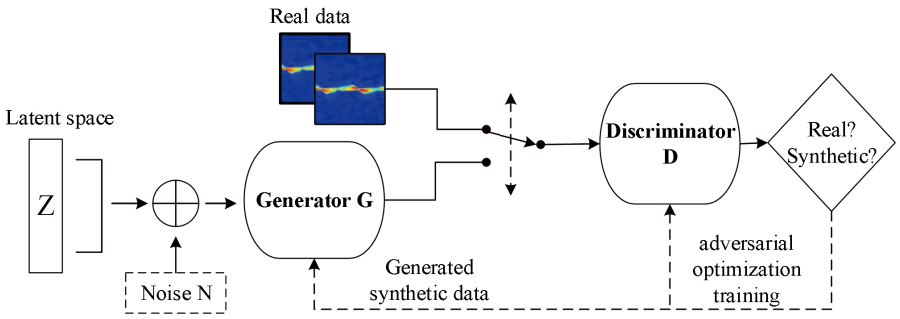

FIGURE 1. Overall architecture of the Generative Adversarial Network (GAN) model.

CNN structure are constructed to achieve data augmentation and classification recognition, with comparative analysis conducted against the original DCGAN and CNN models. Experimental results demonstrate that utilizing the improved DCGAN for radar sample data augmentation in small-sample scenarios represents an effective and viable approach.

2. THEORETICAL FOUNDATION

2.1. Micro-Doppler Effect Analysis

When a radar transmits electromagnetic signals to a moving target, the received signal frequency deviates from the transmitted signal's frequency, a phenomenon known as the "Doppler effect" [13]. The degree of frequency shift is determined by the target's radial velocity relative to the radar. When a moving target undergoes relative motion with respect to the radar system, the distance between the target and radar equipment can be expressed as

$$R(t) = R_0 - vt \tag{1}$$

The time delay of the echo is

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

The high-frequency phase difference is

$$\varphi = -2\pi \frac{2}{\lambda}(R_0 - vt) \tag{3}$$

Therefore, the total Doppler frequency shift of the target is

$$f_D = \frac{1}{2\pi}\frac{\mathrm{d}\varphi}{\mathrm{d}t} = \frac{2}{\lambda}v = f\frac{2v}{c} \tag{4}$$

The primary movement of a target generates a Doppler frequency shift. When any component of the target exhibits micro-motions, these subtle movements induce additional frequency modulation on the radar echoes, creating sideband components in the signal spectrum, a phenomenon termed the "micro-Doppler effect". Human activities constitute typical articulated non-rigid motions, formed by coordinated movements of scattering points across body segments, exhibiting highly complex kinematic patterns. During radar detection of human targets, the received echoes contain hybrid Doppler information modulated by both whole-body translational motion and non-rigid micro-motions, manifesting intricate spectral characteristics. In human action recognition, the micro-Doppler effect serves as a unique signature distinguishing biological targets. Effective extraction of micro-Doppler features from radar echoes provides critical foundations for human action classification and identification.

2.2. Generative Adversarial Networks

Generative Adversarial Networks (GANs), first proposed by Goodfellow et al. [14], constitute a novel deep learning framework comprising a generator (G) and a discriminator (D), with the model architecture illustrated in Figure 1. Radford et al. [15] introduced Deep Convolutional GANs (DCGANs) by replacing multilayer perceptrons in conventional GANs with convolutional neural networks (CNNs). Compared with traditional GANs, DCGANs exhibit four critical architectural innovations: 1) Elimination of pooling layers for up/down-sampling in both generator and discriminator; 2) Systematic integration of Batch Normalization (BN) layers in all network layers except input layers; 3) Adoption of rectified linear unit (ReLU) activations in generator hidden layers with Tanh activation in the output layer; 4) Implementation of LeakyReLU activations throughout the discriminator. The adversarial training process between discriminator D and generator G converges toward Nash equilibrium, formally expressed as the following minimax optimization problem [16]:

$$\min_G \max_D V(D,G) = \underset{x\sim P_{data}(x)}{E}[\log D(x)]$$
$$+E_{z\sim P_z(z)}[\log(1-D(G(z)))] \tag{5}$$

where x denotes the real samples, z the input noise, $D(x)$ the probability of the discriminator classifying real samples as authentic, $G(z)$ the samples generated by the generator, and $D(G(z))$ the probability of the discriminator misclassifying synthetic samples as genuine.

In DCGAN training, the optimization is achieved through alternating execution of the following two gradient update steps [16]:

$$\theta_D^{t+1} = \theta_D^t + \lambda^t \nabla_{\theta_D} V\left(D^t, G^t\right) \tag{6}$$

$$\theta_G^{t+1} = \theta_G^t + \lambda^t \nabla_{\theta_G} V\left(D^{t+1}, G^t\right) \tag{7}$$

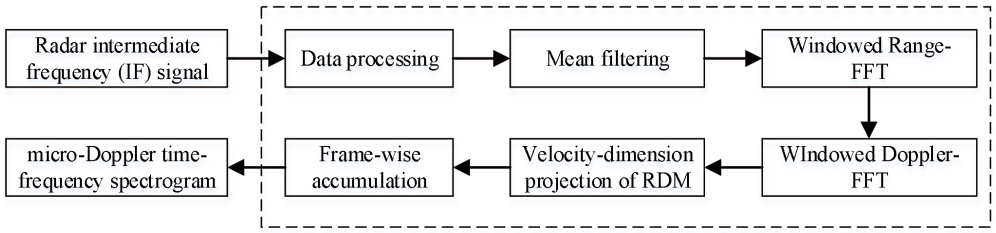

FIGURE 2. Radar data processing pipeline.

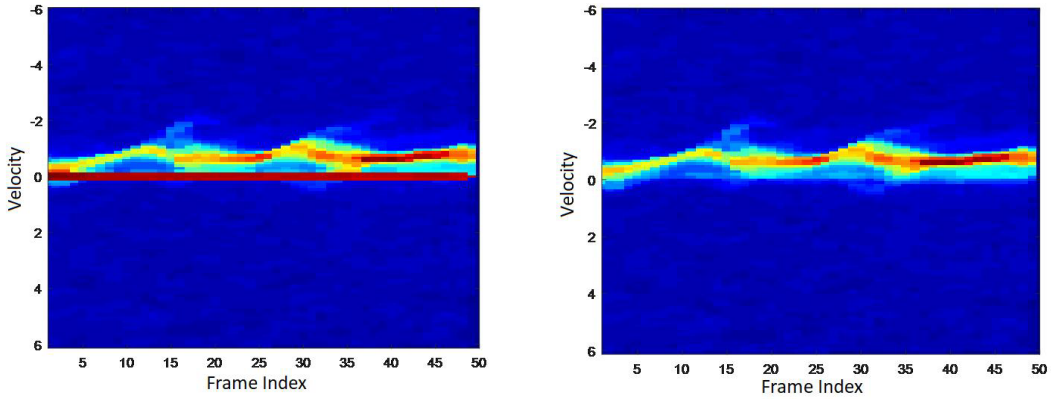

FIGURE 3. Clutter suppression performance comparison.

3. MODEL CONSTRUCTION

3.1. Micro-Doppler Feature Extraction

Radar echoes of different human actions are preprocessed to construct a micro-Doppler time-frequency spectrogram dataset. The processing flow of the radar echo data is shown in Figure 2.

For radar data processing, the mixed intermediate frequency (IF) signal is sampled at the mth sampling point within the nth chirp period, resulting in discrete IF signal samples. These samples are parsed into a two-dimensional data matrix $R[m,n]$, where m denotes the fast-time sampling index, representing the number of sampling points within a single chirp period; n indicates the slow-time sampling index, corresponding to the total number of frames in a complete acquisition cycle.

$$\mathbf{R} = \begin{bmatrix} r_1 & \cdots & r_2 & \cdots & r_n \end{bmatrix} \tag{8}$$

$$\mathbf{r}_n = \begin{bmatrix} r_{1,n} & \cdots & r_{2,n} & \cdots & r_{m,n} \end{bmatrix}^T \tag{9}$$

After successfully parsing the radar echoes, clutter suppression is essential to eliminating interference from walls and environmental clutters. This study employs mean filtering for echo processing — a mature technology with straightforward principles. The comparative time-frequency spectrograms before and after mean filtering are shown in Figure 3. As evidenced in Figure 3, the preprocessing effectively reduces clutter-induced blurring and occlusion of human action-induced micro-Doppler signatures, resulting in enhanced clarity of kinematic features in the spectrograms. The mean filtering operation on the echo matrix R is implemented through two sequential stages:

- Average Reference Calculation: Compute the mean value across all received signal vectors in the slow-time dimension. This establishes a baseline clutter profile derived from stationary environmental reflections.
- Differential Suppression: Subtract the calculated mean reference signal from each individual slow-time frame's received signal vector. This operation effectively isolates dynamic target echoes while suppressing static clutter components.

The mathematical expression can be formulated as

$$\mathbf{R}[m,n] = \mathbf{R}[m,n] - \mathbf{K}[m] \tag{10}$$

where $\mathbf{K}[m]$ denotes the mean reference received signal and expressed as:

$$\mathbf{K}[m] = \frac{1}{N}\sum_{n=1}^{N}\mathbf{R}[m,n] \tag{11}$$

Building upon the acquired data matrix $\mathbf{R}$, the fast-time dimension is processed through Hamming-windowed Range-FFT to obtain high-resolution 1D range profiles. Subsequently, the slow-time dimension undergoes Hamming-windowed Doppler-FFT processing based on the range-resolved data, generating a Range-Doppler Map (RDM). The RDM effectively captures distance and velocity correlation characteristics of all scattering points from human targets within the radar frame. Distinct motion patterns of human actions manifest discriminative RDMs, enabling action-specific feature extraction.

For echo processing of a single human target in this study, let $RD(i,j,t)$ denote the signal power value at the ith range

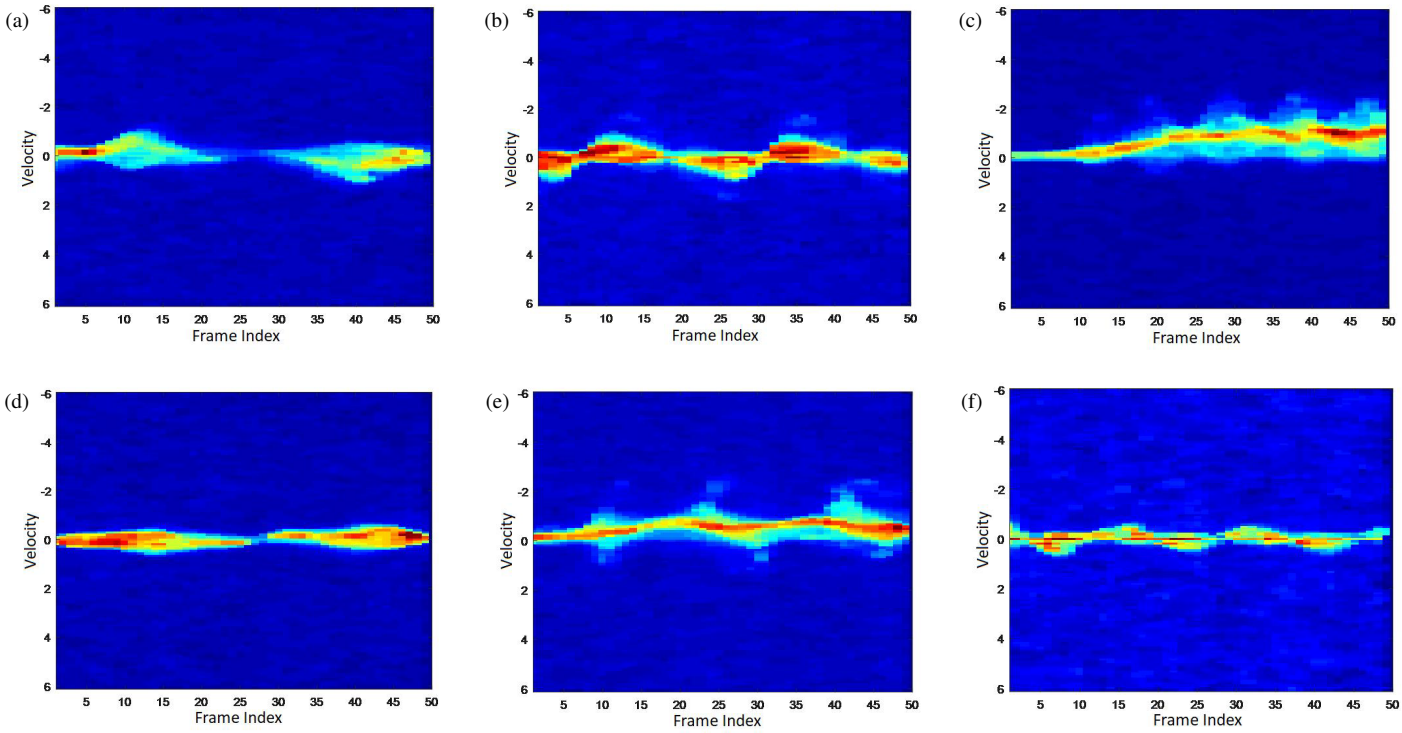

FIGURE 4. Time-frequency spectra of human actions with micro-Doppler signatures, (a) bending, (b) clapping, (c) running, (d) squatting, (e) walking, (f) waving.

bin, jth velocity bin within the Range-Doppler Map (RDM) of the tth frame. The extracted human action RDM undergoes velocity-dimension projection followed by column-wise frame-wise accumulation to generate the micro-Doppler time-frequency spectrogram induced by target motion, as illustrated in Figure 4. The horizontal axis represents the frame index of the echo signal, while the vertical axis corresponds to the velocity magnitude indexed by velocity bins. Consequently, the micro-Doppler frequency can be expressed as:

$$DP(t,j) = \sum_{i} RD(i,j,t) \tag{12}$$

3.2. Improved DCGAN

To address the small-sample scenario in human action recognition, this study introduces the following architectural enhancements to the original DCGAN framework:

1) Optimizer Modification: While the original DCGAN employs the Adam optimizer for both generator and discriminator training, we replace the discriminator's optimization method with the Root Mean Square Propagation (RMSProp) algorithm to mitigate high-noise issues during adversarial training. This adaptation applies an exponentially weighted average of squared gradients to both weight (w) and bias (b) parameters, formulated as:

$$S_{dw} = \beta S_{dw} + (1-\beta)(\mathrm{d}w)^2 \tag{13}$$

$$w = w - lr\frac{\mathrm{d}w}{\sqrt{S_{dw}+\varepsilon}} \tag{14}$$

$$S_{\mathrm{d}b} = \beta S_{\mathrm{d}b} + (1-\beta)(\mathrm{d}b)^2 \tag{15}$$

$$b = b - lr\frac{\mathrm{d}b}{\sqrt{S_{\mathrm{d}b}+\varepsilon}} \tag{16}$$

where lr denotes the learning rate, ε a small constant value for numerical stability, and β the hyperparameter controlling the decay rate.

2) Within the discriminator network, to address the challenge of small-sample datasets, we introduce L2 regularization to enhance the DCGAN's stability and image generation capability. The L2 regularization term is incorporated into the objective function, formulated as:

$$L(w,b) = L_0(w,b) + \frac{\lambda}{2n}\sum_{w} w^2 \tag{17}$$

where L_0 denotes the original loss function, w the weight parameters, λ the regularization coefficient, and n the number of training samples.

3.3. DCGAN-CNN Model Architecture

To address the prevalent challenge of small-scale datasets in radar-based human action recognition leveraging micro-Doppler effects, this work employs a Generative Adversarial Network (GAN) to synthesize radar spectrogram samples. Figure 5 illustrates the operational framework of the DCGAN-based recognition model, which first augments radar spectrograms via DCGAN and subsequently feeds them into a CNN for classification.

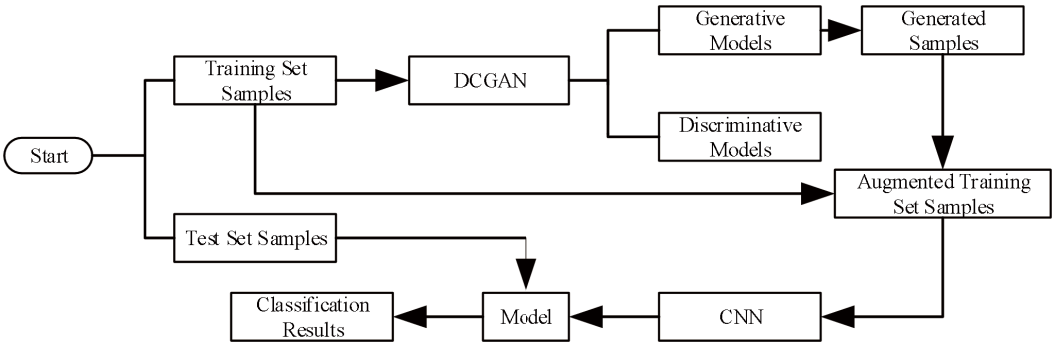

FIGURE 5. DCGAN-CNN architectural framework flowchart.

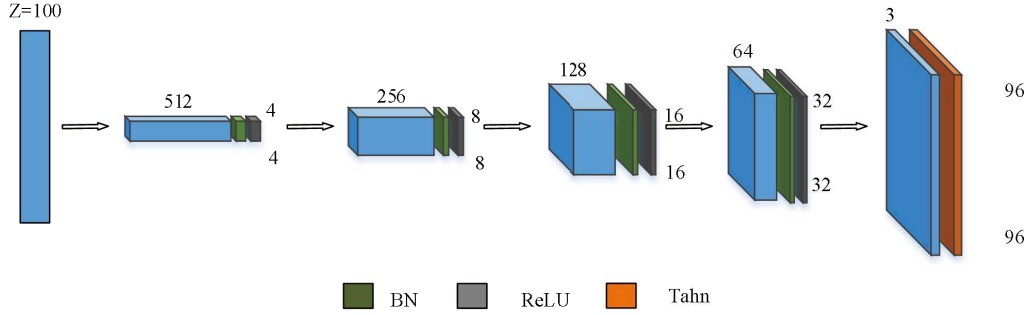

FIGURE 6. Generator architecture diagram.

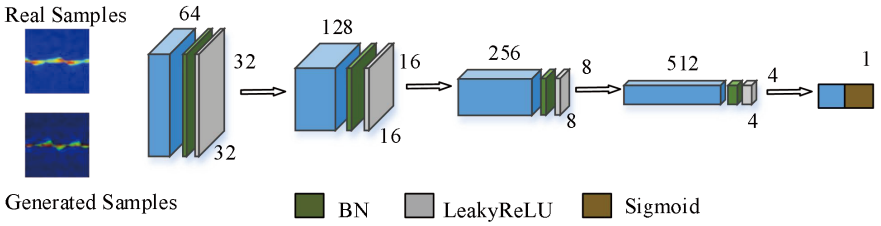

FIGURE 7. Discriminator architecture diagram.

The DCGAN generator architecture in this study, as illustrated in Figure 6, consists of five transposed convolutional layers. The generator network accepts a 100-dimensional random noise vector $\mathbf{Z}$ sampled from a standard normal distribution as input. Transposed convolutional layers are utilized for upsampling, with each layer (except the final one) undergoing Batch Normalization (BN) and ReLU nonlinear activation after the upsampling operation. The output of the final transposed convolutional layer is processed through a Tanh activation function, ultimately generating a $3 \times 96 \times 96$ RGB image.

The DCGAN discriminator architecture is shown in Figure 7. The discriminator network consists of 5 convolutional layers. Each convolutional layer is followed by Batch Normalization (BN) and processed through the nonlinear LeakyReLU activation function. The final layer employs the Sigmoid activation function to ultimately determine the authenticity of the input image.

The human action recognition samples in this study are based on measured radar echo data, constituting a small-sample dataset. After preprocessing, the samples exhibit significant feature differences. Therefore, the constructed convolutional neural network (CNN) adopts a shallow architecture with three layers, as shown in Figure 8. This shallow CNN structure progressively increases channel dimensions and employs 3×3 convolutional kernels. The recognition model utilizes the cross-entropy loss function, with a batch size of 20 and maximum iterations of 100 during training. Stochastic Gradient Descent (SGD) is applied to optimize the network parameters.

4. EXPERIMENTAL RESULTS AND ANALYSIS

4.1. Experimental Setup and Data Acquisition

All experiments were conducted on an NVIDIA RTX 3090 GPU with 24 GB VRAM and Intel Xeon Gold 6226R CPU @ 2.90 GHz. The experiments utilized the IWR1642 industrial millimeter-wave radar sensor and DCA1000 data acquisition board to construct a real-time human action data acquisition platform for capturing radar echoes from human actions. The IWR1642 radar development platform, equipped with the PC-controlled mmWave Studio interface (Texas Instruments), enables real-time echo signal acquisition, transmission, and vi-

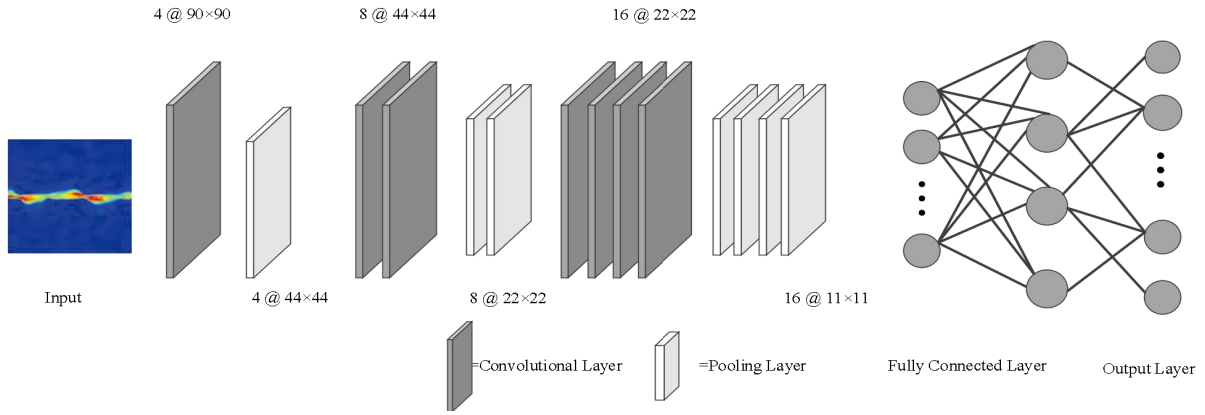

FIGURE 8. CNN-based recognition model.

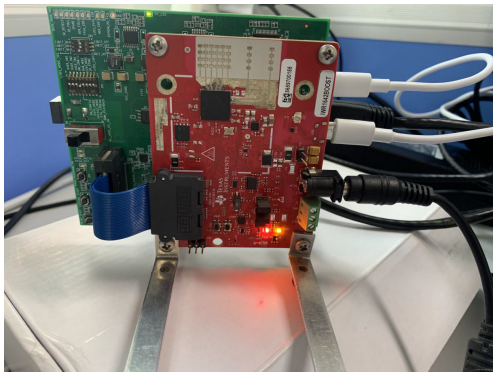

FIGURE 9. Radar hardware configuration.

FIGURE 10. Data acquisition setup.

sualization. The radar data acquisition module is illustrated in Figure 9.

To ensure the validity of the training data, the experimental data acquisition was conducted in a semi-open outdoor environment without interference from other moving targets. The radar antenna was fixed on a 0.9-meter-high tripod, with the target positioned 2.5 meters away from the radar and moving radially relative to the radar, as shown in Figure 10. Figure 11 illustrates the data acquisition software interface. The radar was configured in a monostatic mode (one transmitter, one receiver), with specific parameters listed in Table 1. Five human subjects with heights of 153 cm, 155 cm, 156 cm, 160 cm, and 175 cm participated in the experiment. Six daily human actions were selected: (1) bending, (2) clapping, (3) running, (4) squatting, (5) walking, and (6) waving. Each action was recorded for 2 seconds to ensure motion integrity. These six actions exhibit both similarities and significant differences in their range profiles.

TABLE 1. Radar hardware parameter configuration.

Parameters	Value
Start Frequency (GHz)	77
Frequency Modulation Slope (MHz/μs)	64
Number of ADC Samples	256
Pulses per Frame	128
Sampling Rate (MSPS)	5.12

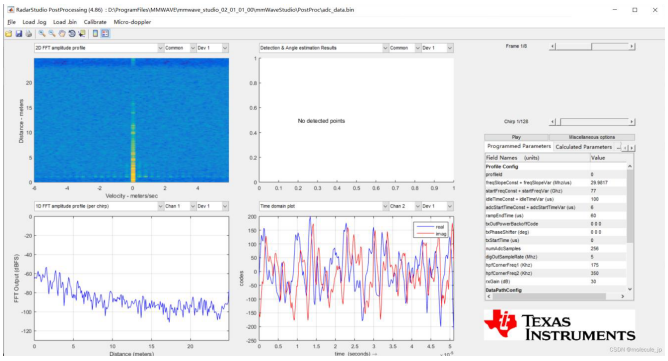

FIGURE 11. Data acquisition software interface.

All motions were performed facing the radar, with each action repeated 50 times.

4.2. Experimental Dataset Construction

Utilizing the developed millimeter-wave radar system, 250 echo samples per action were collected and processed into 2D micro-Doppler time-frequency spectrograms through the proposed preprocessing methodology. For each action category, the samples were randomly partitioned into a training subset D1 and a testing subset D2. The enhanced DCGAN model was then employed to generate an augmented training subset D3 by syn-

TABLE 2. Dataset statistics.

Human Actions	Label	Training Subset D1	Testing Subset D2	Augmented Training Subset D3
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

thesizing samples in a 1 : 1 ratio relative to the original training data. Detailed statistics of the dataset composition are provided in Table 2.

4.3. Analysis of Experimental Results

4.3.1. Evaluation Metrics

In classification tasks, the model's recognition accuracy (accuracy, acc) is defined as:

$$\text{acc} = \frac{P_{\text{true}}}{P_n} \tag{18}$$

where P_{true} denotes the number of correctly classified samples, and P_n represents the total number of samples.

In classification tasks, four fundamental metrics are defined: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). The model is evaluated using precision, recall, and specificity metrics.

1) Precision (precision):

$$\text{precision} = \frac{T_P}{T_P + F_P} \tag{19}$$

2) Recall (recall):

$$\text{recall} = \frac{T_P}{T_P + F_N} \tag{20}$$

3) Specificity (specificity):

$$\text{specificity} = \frac{T_N}{F_P + T_N} \tag{21}$$

Here, T_P and T_N denote the number of human action instances correctly predicted as their true classes. F_P represents the quantity of instances from other human action classes misclassified into a specific target class. F_N indicates the number of target-class human actions erroneously predicted as other classes.

4.3.2. Performance Analysis

The recognition performance of the improved DCGAN-CNN model for micro-Doppler time-frequency spectra of different human actions was analyzed. The neural network was trained using Dataset D3 and evaluated on Test Set D2.

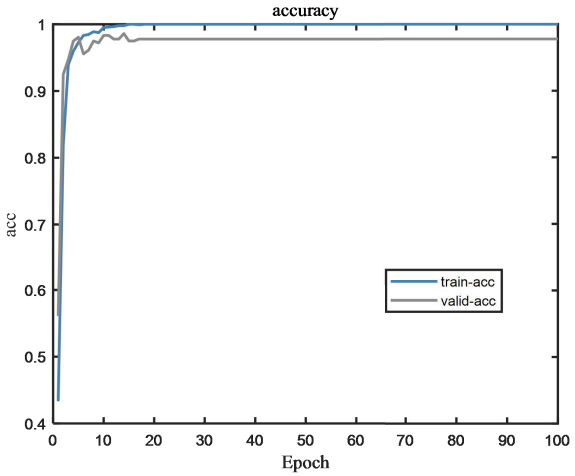

FIGURE 12. Accuracy curve.

During the model training process, the recognition accuracy curves of the training set and validation set are shown in Figure 12. It can be observed that the validation accuracy reaches approximately 97%. The confusion matrix evaluation results are presented in Figure 13. The high values along the diagonal indicate that the proposed network achieves correct classification for most spectrogram categories.

The implemented network attains a human action recognition accuracy of 97.5%. Figure 13 further reveals that hand clapping and hand waving actions are prone to misclassification,

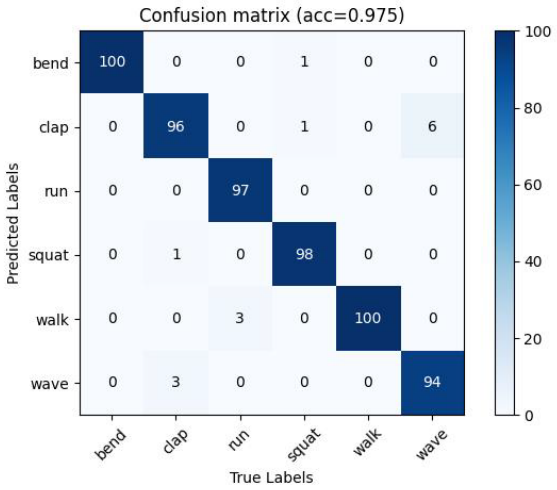

FIGURE 13. Confusion matrix.

followed by running and walking. The model demonstrates superior recognition accuracy for other daily activities. Quantitative evaluations using precision, recall, and specificity metrics are summarized in Table 3.

TABLE 3. Model evaluation metrics.

Class	Precision	Recall	Specificity
bend	0.99	1.0	0.998
clap	0.932	0.96	0.986
run	1.0	0.97	1.0
squat	0.99	0.98	0.998
walk	0.971	1.0	0.994
wave	0.969	0.94	0.994
Macro	0.973	0.968	0.995
Weighted	0.975	0.975	0.995

4.3.3. Model Validation

To validate the data-augmented models, this study compares three approaches: a CNN without augmented samples, the original DCGAN, and the proposed improved DCGAN-augmented model. All experiments uniformly adopt 100 training epochs. Figure 14 illustrates the accuracy curves of different classification methods during training. Table 4 presents the recognition accuracy comparisons across models, while Table 5 provides detailed precision and recall metrics.

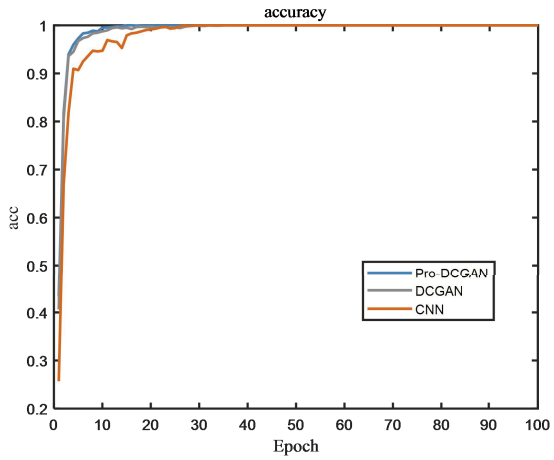

FIGURE 14. Training set accuracy curve.

As shown in Table 4, the recognition accuracy of the baseline CNN model without data augmentation is 94.33%. The DCGAN-augmented model achieves 96.5% accuracy, while the improved DCGAN-augmented model attains 97.5%. Compared to the non-augmented model, the improved DCGAN framework demonstrates a 3.17% accuracy gain, and it outperforms the original DCGAN by 1%. These results confirm that data augmentation significantly enhances recognition accuracy, with the proposed improved DCGAN achieving state-of-the-art performance.

TABLE 4. Recognition accuracy across models/%.

Models	accuracy
Improved DCGAN	97.5
DCGAN	96.5
CNN	94.33

Analysis of Table 5 reveals precision improvements in four action categories: hand clapping, running, walking, and hand waving, with the most substantial precision gain observed for running. For bending and squatting actions, the improved DCGAN-augmented model exhibits marginally lower precision than the original DCGAN, but achieves significantly higher recall rates than both DCGAN and CNN baselines.

4.3.4. Generated Image Evaluation

To evaluate the generation performance of the proposed improved DCGAN model, micro-Doppler time-frequency spectra synthesized by the model are analyzed. Figures 15 and 16 respectively display the real spectrograms and improved DCGAN-generated images. Comparative analysis demonstrates that the improved DCGAN model effectively learns discriminative features of various human actions, with the generated image samples exhibiting high structural similarity to the authentic spectrograms.

The quality and diversity of images generated by Generative Adversarial Networks (GANs) reflect, to some extent, the performance of the network model. Salimans et al. proposed the Inception Score (IS) to evaluate the quality and diversity of images generated by generative models [17]. The IS is calculated as follows:

$$IS(G) = \exp\left(E_{x\sim p_g} D_{KL}(P(y|x)P(y))\right) \tag{22}$$

where x denotes the generated data, y the corresponding labels, $p(y|x)$ the conditional distribution of class assignments given a generated sample, $p(y)$ the marginal distribution of classes, and D_{KL} the Kullback-Leibler divergence between the two distributions.

Based on the IS metric, Table 6 presents a comparative analysis of micro-Doppler-generated images between the improved DCGAN model proposed in this work and the baseline DCGAN model for different human action classes. The results demonstrate that the improved DCGAN achieves higher IS values across all spectrogram categories, confirming its superior capability in generating higher-quality synthetic time-frequency representations.

4.3.5. Comprehensive Evaluation of Improved Model

1. Quantitative Performance Gains

As shown in Table 7, the improved DCGAN demonstrates significant advantages over the baseline model across key metrics.

2. Computational Complexity & Runtime

As shown in Table 8, under NVIDIA RTX 3090 GPU environments, the improved DCGAN incurs marginal computational overhead while achieving superior performance

In summary, the improved DCGAN demonstrated modest yet consistent performance enhancements, achieving a 1% ac-

TABLE 5. Model performance comparison.

Comparison Metrics	Models	bend	clap	run	squat	walk	wave
precision	Improved DCGAN	0.99	**0.932**	**1.0**	0.99	**0.971**	**0.969**
	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	Improved DCGAN	**1.0**	0.96	**0.97**	**0.98**	**1.0**	**0.94**
	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

TABLE 6. Comparison of Inception Score (IS) values across models (↑).

Models	bend	clap	run	squat	walk	wave
DCGAN	1.55	1.38	1.45	1.39	1.16	1.42
Improved DCGAN	1.70	1.45	1.65	1.49	1.30	1.44

TABLE 7. Comparison of improved DCGAN vs. Original DCGAN performance.

Metric	Original DCGAN	Improved DCGAN	Improvement	Statistical Validation
Classification Acc. (%)	96.5	97.5	$+1.0$	t-test ($p = 0.032$)
Inception Score (↑)	1.43 ± 0.12	1.62 ± 0.15	$+13.3\%$	Bootstrap 95% CI
FID (↓)	32.7	28.7	-12.2%	Mann-Whitney U test ($p < 0.05$)
Training Convergence	80 epochs	50 epochs	-37.5%	Loss curve derivative analysis

TABLE 8. Runtime efficiency comparison.

Phase	Original DCGAN	Improved DCGAN	Overhead	Key Factors
Single Iteration (ms)	23.4	25.1	$+7.3\%$	L2 regularization gradients
Full Training (hrs)	4.2	4.5	$+7.1\%$	Faster discriminator convergence
End-to-End Latency (ms)	210	225	$+7.1\%$	Pipeline accumulation

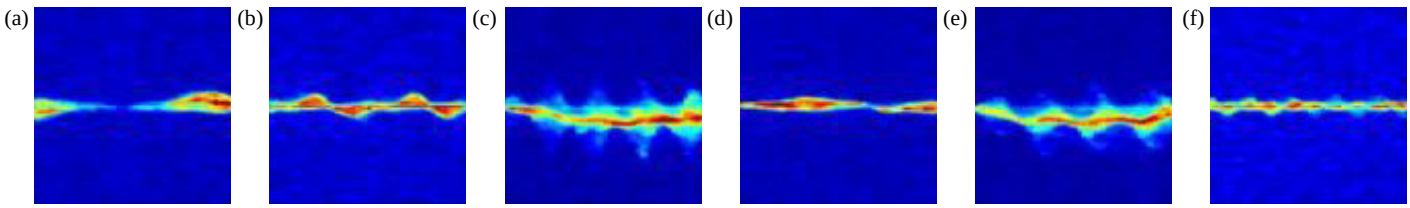

FIGURE 15. Real images, (a) bending, (b) clapping, (c) running, (d) squatting, (e) walking, (f) waving.

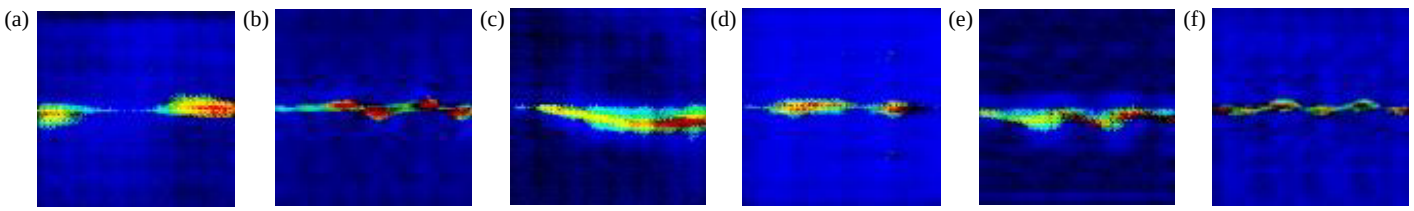

FIGURE 16. Improved DCGAN-generated images, (a) bending, (b) clapping, (c) running, (d) squatting, (e) walking, (f) waving.

curacy improvement, higher IS for generation quality, and enhanced training stability, while maintaining comparable model complexity through optimized adversarial training strategies. Although the integration of L2 regularization and RMSProp optimization introduced minor computational overhead, the framework preserved its computational efficiency without substantially increasing runtime demands. Crucially, this work establishes a cost-effective solution for radar-based motion recognition in small-sample scenarios by systematically addressing data scarcity challenges through refined gradient balance mechanisms and architectural adjustments in the discriminator-generator interplay, thereby advancing practical deployment potential in resource-constrained environments.

5. CONCLUSION

This study proposes a few-shot human action recognition framework integrating an enhanced Generative Adversarial Network (GAN) with Convolutional Neural Networks (CNNs) for millimeter-wave radar systems. The model achieves robust recognition of daily human activities through effective data augmentation. A 77 GHz millimeter-wave radar data acquisition system was deployed to collect human action echo signals, enabling precise extraction of micro-Doppler signatures from complex radar returns. An improved Deep Convolutional GAN (DCGAN) architecture was developed for sample augmentation, coupled with a CNN classifier for motion categorization. Compared to baseline models trained on non-augmented datasets, the enhanced DCGAN achieves a 3.17% accuracy improvement. When benchmarked against the original DCGAN, the proposed model exhibits a 1% accuracy gain, though this marginal enhancement suggests limited architectural superiority. These results validate the framework's practical utility in boosting recognition accuracy and handling complex motion patterns, demonstrating GAN's viability for radar dataset expansion.

However, implementing GANs for motion sample augmentation involves intricate nonlinear adversarial optimization, which may lead to training instability, suboptimal sample quality, and limited diversity during Nash equilibrium pursuit. Future research should focus on developing stabilized training mechanisms to enhance classification performance and generation fidelity. Additionally, current experiments were conducted in ideal interference-free environments; subsequent studies must address motion recognition under real-world noisy conditions.

ACKNOWLEDGEMENT

Firstly, we wish to thank the Jiangxi Provincial Natural Science Foundation (Grant: 20242BAB25052) and the Education Department of Jiangxi province (Grant: GJJ201408) for the support to this research work. Secondly, we also wish to thank the anonymous reviewers for their help in improving this paper.

REFERENCES

[1] Wei, X. X., "Human posture recognition algorithm based on inertial sensors," *Intelligent Computer and Applications*, Vol. 12, No. 6, 97–101, 2022.

[2] Xing, M. M., F. Yang, Z. H. Xin, and G. H. Wei, "Research on human action recognition using depth video," *China Medical Devices*, Vol. 38, No. 1, 36–41, 2023.

[3] Rana, S. P., M. Dey, M. Ghavami, and S. Dudley, "Non-contact human gait identification through IR-UWB edge-based monitoring sensor," *IEEE Sensors Journal*, Vol. 19, No. 20, 9282–9293, 2019.

[4] Ding, C. X., Y. H. Zhang, Z. T. Sun, *et al.*, "Human complex motion recognition based on FMCW radar," *Radar Science and Technology*, Vol. 18, No. 6, 584–590, 2020.

[5] Ding, Y. P., R. J. Liu, and X. M. Xu, "Micro-Doppler frequency estimation method for human targets based on continuous wave radar," *Journal of Central South University (Natural Science Edition)*, Vol. 53, No. 4, 1273–1280, 2022.

[6] 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.

[7] Yao, Z. P., Z. Y. Tang, and Y. J. Sun, "Human action recognition method based on convolutional neural networks," *Journal of Air Force Early Warning Academy*, Vol. 34, No. 5, 360–364, 2020.

[8] Pan, R., "Human action recognition based on graph convolutional neural networks," Ph.D. dissertation, Chongqing University of Posts and Telecommunications, Chongqing, China, 2022.

[9] Gurbuz, S. Z. and M. G. Amin, "Radar-based human-motion recognition with deep learning: Promising applications for indoor monitoring," *IEEE Signal Processing Magazine*, Vol. 36, No. 4, 16–28, 2019.

[10] Chen, F., F. Zhu, Q. X. Wu, J. Zheng, and X. Zhang, "A review of generative adversarial networks and their applications in image generation," *Chinese Journal of Computers*, Vol. 44, No. 2, 347–369, 2021.

[11] 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, 2019.

[12] Qu, B. and H. Li, "Research on gesture recognition based on CNN and DCGAN," *Journal of Heilongjiang Bayi Agricultural University*, Vol. 33, No. 3, 79–84, 2021.

[13] Chen, V. C., D. Tahmoush, and W. J. Miceli, *Radar Micro-Doppler Signatures: Processing and Applications*, IET Digital Library, London, 2014.

[14] Goodfellow, I., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Communications of the ACM*, Vol. 63, No. 11, 139–144, 2020.

[15] Radford, A., L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.

[16] Yu, H. Y., L. Yin, S. F. Li, and S. Lyu, "Few-shot radar modulation signal recognition algorithm based on generative adversarial networks," *Journal of Xidian University*, Vol. 48, No. 6, 96–104, 2021.

[17] Salimans, T., I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training GANs," in *30th Conference on Neural Information Processing Systems (NIPS 2016)*, Vol. 29, 1–9, Barcelona, Spain, 2016.