Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 4555136, 16 pages
https://doi.org/10.1155/2022/4555136

Research Article

Gait Recognition Using Spatio-Temporal Information of 3D Point Cloud via Millimeter Wave Radar

Tao Li,[1,2] Zhichao Zhao,[1] Yi Luo,[1] Benkun Ruan,[1] Dawei Peng,[1] Lei Cheng,[1] and Chenqi Shi[1]

[1] *School of Computer Science, China University of Mining and Technology, Xuzhou, 221116 Xuzhou, China*
[2] *Mine Digitization Engineering Research Center of Ministry of Education of the People's Republic of China, China University of Mining and Technology, Xuzhou, 221116 Xuzhou, China*

Correspondence should be addressed to Tao Li; li_t@cumt.edu.cn and Chenqi Shi; shichenqi@cumt.edu.cn

Received 9 January 2022; Accepted 25 July 2022; Published 16 August 2022

Academic Editor: Rajesh Kaluri

Gait recognition is one of the crucial methods in identity recognition, which has a wide range of applications in many fields, such as smart home, smart office, and health monitoring. The camera is the most mainstream traditional solution. But the camera is difficult to maintain stable performance in the dark, low light, and bad weather conditions. In addition, privacy leakage is also one of the important issues that people worry about. In contrast, as the latest research progress in gait recognition, millimeter wave radar can not only protect people's privacy, but also maintain normal perception performance in dark conditions. In this paper, we propose a system for gait recognition named MTPGait using spatio-temporal information via millimeter wave radar. We specially design a neural network that can extract multiscale spatio-temporal features along space and time dimensions of 3D point cloud concisely and efficiently. We use LSTM to design the context flow of local and global time and space, fusing local and global spatio-temporal features. In addition, we construct and release a millimeter wave radar 3D point cloud data set, which consists of 960-minute gait data of 40 volunteers. Using the data set, we evaluate the system and compare it with four state-of-the-art algorithms. The experimental results show that MTPGait is able to achieve 96.7% recognition accuracy in a single-person scene on fixed route and 90.2% recognition accuracy when two people coexist, while none of the existing methods is more than 90% recognition accuracy in either scenario.

1. Introduction

As an important link in human-computer interaction, identity recognition plays an important role in many fields. For example, personalized control of room temperature, selection of background music, and adjustment of light brightness are all dependent on accurate identity information. At present, traditional recognition methods are based on wearable devices [1, 2]. The user's identity is recognized through the smart phone, ID card, token, and other devices carried by the user. However, carrying extra equipment may cause inconvenience to users in some cases. Therefore, no wearable recognition methods based on biometric features, such as sclera [3], fingerprints [4], and iris [5], have been rapidly developed to adapt to more general scenarios. The visual recognition technology is widely used in various scenes [6–9]. However, the camera is difficult to obtain clear images in low light or even dark scenes. In addition, as people pay more and more attention to privacy, especially considering the factors such as the hijacking of cameras by malicious users, people increasingly feel the serious privacy threat of cameras in the home and office environment.

To eliminate issues such as privacy and weak light, the researchers propose using WiFi for identity recognition, based on the principle that each person's unique body features and gait characteristics lead to different channel state information (CSI) patterns to recognize subtle differences between people [10]. [11] combines the convolutional layer with the LSTM layer and proposes a simple and effective deep learning method to realize automatic identification of people by WiFi. [12] proposes a gait recognition method based on WiFi and LSTM and designs an effective six-layer recursive neural network to extract human gait biometric features. However, the WiFi-based approach requires the deployment of a transmitter and a receiver, and the user needs to walk between the transceivers, which severely limits the recognition area and the universality of the application. In addition, the existing WiFi identification technology is not able to identify multiple people in the same scene well due to limited factors such as the bandwidth of the device. Recently, millimeter wave radar technology has made a prominent appearance in the field of autonomous driving, and meanwhile, research in the field of indoor behavior perception has also increased gradually.

The millimeter wave radar does not have the privacy problem of the camera and can still work normally in low light or dark environment. Compared with WiFi identification technology, millimeter wave radar has a bandwidth of up to 4 GHz, which can provide more accurate spatial resolution. In addition, millimeter wave radar can provide relatively fine-grained velocity information. Importantly, the current WiFi recognition technology relies heavily on the information in the environment. Although some studies use transfer learning and adversarial learning to solve this problem, it still needs to spend a large deal of time and energy to retrain or improve the model after the environment changes. Gait recognition based on millimeter wave radar can filter the static point cloud directly and has high robustness to the environment.

The 3D point cloud of pedestrian gait contains rich timing and spatial information, such as walking speed, spatial trajectory, local limb swing amplitude, stride size, and frequency. However, the existing methods for gait recognition using 3D point clouds do not have a good fusion of time sequence and spatial information in the gait point clouds at the same time. For example, mID [13] uses LSTM for feature extraction of point clouds and only focuses on the timing information in the point cloud. [14] directly inputs the 3D point cloud into CNN for feature extraction. Although this method uses the spatio-temporal convolution kernel to extract the spatio-temporal information of point cloud, it pays little attention to the fused features because the five attributes are input independently for feature extraction. The fused features can better represent the coordination between body and limb, the correspondence between speed and spatial position, and the matching between stride frequency and stride length.

However, there are still some problems in gait recognition based on millimeter wave radar. In [15], the gait motion features were transformed into micro-Doppler images, and the CNN model was trained and evaluated to achieve classification. CNN alone can extract local and global spatial characteristics, but in the process of walking, the spatial and velocity information of the whole human body and the local body are closely related in time series, while CNN alone cannot well extract the time series features and cannot effectively use millimeter wave radar speed information. [13] uses millimeter wave radar data to generate 3D point cloud gait data sets and trains and evaluates LSTM models to realize 12 person gait recognition. LSTM can well extract the temporal features of point cloud data, but the morphology of point cloud itself also contains a large amount of feature information, which cannot be effectively extracted only by using LSTM model.

In order to solve the above challenges, this paper proposes a new gait recognition method based on 3D point cloud. This method is based on the concise and efficient CNN + LSTM network structure for multiscale spatio-temporal features extraction. In addition, since there are very few public millimeter wave radar gait data sets, this paper collects the single and double persons gait data of 40 volunteers in three scenarios to form a large-scale 3D point cloud data set of gait (https://github.com/caoxu907/MMWAVE_gait). Then, we conduct model training and evaluation on this data set and achieve the highest recognition accuracy of 96.7%.

The main contributions of this paper are as follows:

(1) We specially design a neural network that can extract multiscale spatio-temporal features along space and time dimensions of 3D point cloud concisely and efficiently. And LSTM is used to design the context flow of local and global time and space, fusing local and global spatio-temporal features

(2) We use the DBSCAN algorithm to cluster the point clouds and use the Hungarian algorithm to perform interframe multitarget matching

(3) We have built and published a large-scale 3D point cloud gait data set of 40 volunteers in three scenarios, with a total length of 960 minutes

(4) We evaluate the accuracy of single and double persons gait recognition in multiple environments, and the results show that single and double persons recognition accuracies are 96.7% and 90.2%

2. Related Work

2.1. Gait Recognition Based on Wireless Signals. There has been a large amount of work in the research of gait recognition based on wireless signals. Among them, WiFi-based gait recognition has the most research work. Using WiFi for identification, the principle is that each person's unique physical characteristics and gait characteristics lead to different CSI patterns to identify the subtle differences between people [21–23]. WiWho [24] uses step analysis and walking analysis to extract the step features and overall walking features of each detected target from the CSI data and then matches these features with the walking signatures pretrained by machine learning to achieve classification. WiFi-ID [25] first uses continuous wavelet transform to separate signals, then uses RelieF feature selection algorithm to extract time and frequency domain features from the separated signals, and finally uses sparse approximation algorithm to achieve classification. Because the WiFi-based sensing method has a strong dependence on the environment, it is difficult to achieve better results after the environment changes. In addition, it is difficult for WiFi signals to separate targets like millimeter wave radars, so it is impossible to realize the perception of multiple people.

Researches on millimeter wave radar-based identification are also gaining attention. Table 1 reviews the recent research methods and recognition accuracy of identification using millimeter wave radar. [15] uses CNN to extract the micro-Doppler features of gait to realize gait recognition. Experimental results show that the error rate of the test set is 21.54% in gait recognition of five targets. mID [13] is the first to use millimeter wave radar point cloud for gait recognition. This method uses LSTM to extract the time sequence characteristics of gait point cloud data. Finally, it can achieve 89% accuracy in single-person gait recognition and does not introduce the accuracy of multiperson recognition. Meng et al. establish the first public millimeter wave radar gait point cloud data set mmGait and propose mmGaitNet [14]. This method can achieve 90% accuracy in single-person recognition, but the pedestrian route is a fixed route. In random routes, this method can only achieve 45% accuracy.

2.2. Other Perception Based on Millimeter Wave. In addition to using millimeter wave for gait recognition, researchers have also conducted many other perceptual studies [26, 27]. mmTrack [28] improves spatial resolution by applying digital beamforming to the receiving antenna and proposes a new target detection method to solve near-far effect and measurement noise. In addition, mmTrack designs a robust clustering method to estimate the location of multiple targets and finally achieves continuous tracking of multiple trajectories, with a median tracking error of 9.9 cm for dynamic targets. mSense [29] proposes a novel method for material identification using millimeter wave radar. This method uses CIR interpolation, direct path-based synchronization, and background and noise elimination techniques to characterize the intrinsic reflectivity of the target and then achieves the target category's association. Finally, the method achieves an average of 93% recognition accuracy for five materials such as aluminum, ceramics, plastics, wood, and water. mmVib [30] proposes a noninvasive micron-level vibration measurement method using millimeter wave radar. This method introduces the Multi-Signal Consolidation (MSC) model to capture the multifrequency and multiantenna characteristics of the reflected signal of the vibrating object and achieves a comprehensive description of the reflected millimeter wave signal model description of the vibrating object. In the end, the method achieves a relative amplitude error of 8.2% and a relative frequency error of 0.5%, and the median error in the amplitude of 100 micrometers was only 3.4 micrometers. VIMO [31] proposes a method for detecting breathing and heartbeat using millimeter wave radio. The autocorrelation function of this method calculates the CIR phase to estimate the respiratory frequency and uses the cubic spline interpolation method to estimate the heart rate. Finally, the respiratory frequency estimation can be achieved. The median accuracy is 0.19 BPM, and the median accuracy of the heart rate estimate is 1 BPM. mHomeGes [32] proposes a concentrated position-Doppler profile (CPDP) based on point cloud and a lightweight neural network mGesNet to extract gesture features, proposes a new user discovery method to eliminate multipath effect, and finally designs a hidden-Markov model-based voting mechanism (HMM-VM) to realize continuous gesture recognition. The recognition accuracy of mHomeGes can reach 95.3% in the case of ambient interference.

3. Principle of Millimeter Wave Radar

The millimeter wave radar determines the distance, speed, and angle of the object by capturing the signal reflected by the obstruction of the object on the transmission path. The signal emitted by the millimeter wave radar is a continuous frequency modulation wave, which is essentially a sine wave signal whose frequency increases linearly with time. This kind of continuous frequency modulation wave signal is also called chirp, which is transmitted and received by two antenna arrays equipped with millimeter wave radar. The transmitted signal of millimeter wave radar can be expressed as

$$s_T(t) = A_T \cos 2\pi \left[f_c t + \int_0^t f_T(\tau) d\tau \right]. \tag{1}$$

The received signal is

$$s_R(t) = A_R \cos 2\pi \left[f_c(t - \Delta t_d) + \int_0^t f_R(\tau) d\tau \right]. \tag{2}$$

Among them, $s_T(t)$ and $s_R(t)$ represent the transmitted signal and the received signal, respectively. A_T and A_R are the amplitudes of the transmitted signal and the received signal, respectively. f_c is the center frequency of the carrier. $f_T(\tau)$ and $f_R(\tau)$ are the frequencies of the transmitted signal and the received signal, respectively. Δt_d indicates the flight delay from the transmitted signal back to the receiving end after reflection.

The frequency of the transmitted signal and the received signal can be obtained by the following formula:

$$f_T(\tau) = \tau \frac{B}{T}, \tag{3}$$

$$f_R(\tau) = (\tau - \Delta t_d) \frac{B}{T} + \Delta f_d. \tag{4}$$

Among them, B is the signal bandwidth, T is the sweep period of the chirp signal, and Δf_d represents the Doppler frequency shift. After the receiving antenna captures the reflected signal, the mixer combines the transmitted signal and the echo signal reflected by the target scene to generate an intermediate frequency signal, which can be expressed as

$$\begin{aligned} s_{\mathrm{IF}}(t) = f_{\mathrm{LPF}}\{s_T(t) s_R(t)\} = \frac{1}{2} A_T A_R \\ \cdot \cos 2\pi \left[f_c \Delta t_d + \left(\frac{B}{T} \Delta t_d - \Delta f_d \right) t \right]. \end{aligned} \tag{5}$$

Table 1: A review of various conventional studies on radar-based personal identification.

References	Data set	Method	Number and type of radar	Recognition accuracy
[16]	20 volunteers 2,880 radar signals each lasting 10 seconds	A deep network constructed from several individually sparse AEs The input data are μ D signals	Using a single radar A continuous wave Doppler radar, the ST200 system by RFbeam	The recognition accuracy of single person on random routes can reach 96.20%.
[17]	8 volunteers Total 40,000 frames	Multichannel 3D convolutional neural network Gait point cloud data as input	Using a single radar IWR1443BOOST	The recognition accuracy of single person on random routes can reach 93%
[18]	10 volunteers Total 1500 samples 2.3 seconds per sample	Input the gait spectrum map into the convolutional neural network	Using a single radar AWR1443	The recognition accuracy of single person on fixed and random routes can reach 91% and 90%
[19]	4 volunteers 100000 frames of data	Input point cloud data into multibranch CNN networks	Using a single radar CAL60S244 IBM AiP	The recognition accuracy of single person on random routes can reach 85.8%
[20]	6 volunteers Three environments Total 75 minutes of data	Input point cloud data into spatio-temporal graph convolutional network	Using a single radar IWR6843ISK	The recognition accuracy of single person and two persons on random routes can reach 68.88% and 75.51%
[14]	95 volunteers Total 30 hours of data	Input the five attributes of the point cloud into the CNN network	Using two radars IWR6843 and IWR1443	The recognition accuracy of single person and two persons on fixed route can reach 90% and 86%
[15]	5 volunteers 150 minutes of gait data	Input the micro-Doppler signature into the DCNN model	Using a single radar An FMCW radar device produced by industrial radar systems GmbH	The method achieved an error rate of 24.70% on the validation set and an error rate of 21.54% on the test set
[13]	12 volunteers Total 120 minutes of data	Each frame of data is flattened into a 16000 dimensional vector and fed into the bidirectional LSTM	Using a single radar IWR1443	The recognition accuracy of single person on random routes can reach 89%

According to the delay of each chirp pulse, the distance formulas for different multitargets can be derived as follows:

$$\Delta d = \frac{\Delta f c}{2S}, \tag{6}$$

where Δf is the frequency of the intermediate frequency signal and S is the slope of the chirp.

The millimeter wave radar measures the speed by emitting two chirp pulses separated by Tc and obtains the distance of the object by performing FFT processing on each reflected echo. According to the different phases of the same peak position, the speed of the object movement is derived:

$$v = \frac{\lambda \Delta \Phi}{4\pi T_c}. \tag{7}$$

Among them, λ is the wavelength, and $\Delta \Phi$ is the phase difference between the two chirped signals.

The millimeter wave radar estimates the angle by measuring the phase change caused by the small change in the distance of the object. The estimated angle can be expressed as

$$\theta = \sin^{-1}\left(\frac{\lambda \Delta \Phi}{2\pi l}\right), \tag{8}$$

where I is the distance between the antennas and $\Delta \Phi$ is the phase difference between the two chirp signals.

4. System Design

MTPGait uses millimeter wave radar to obtain biological characteristics of human gait when walking for identity recognition. The continuous frequency modulation wave emitted by the millimeter wave radar is used to obtain the reflected signal of the walking person and then generate 3D point cloud. The system realizes the classification by analyzing point cloud data. Figure 1 is the flow chart of MTPGait. There are six steps in total:

(i) Point cloud generation: the millimeter wave radar transmits the chirp signal and records the reflected

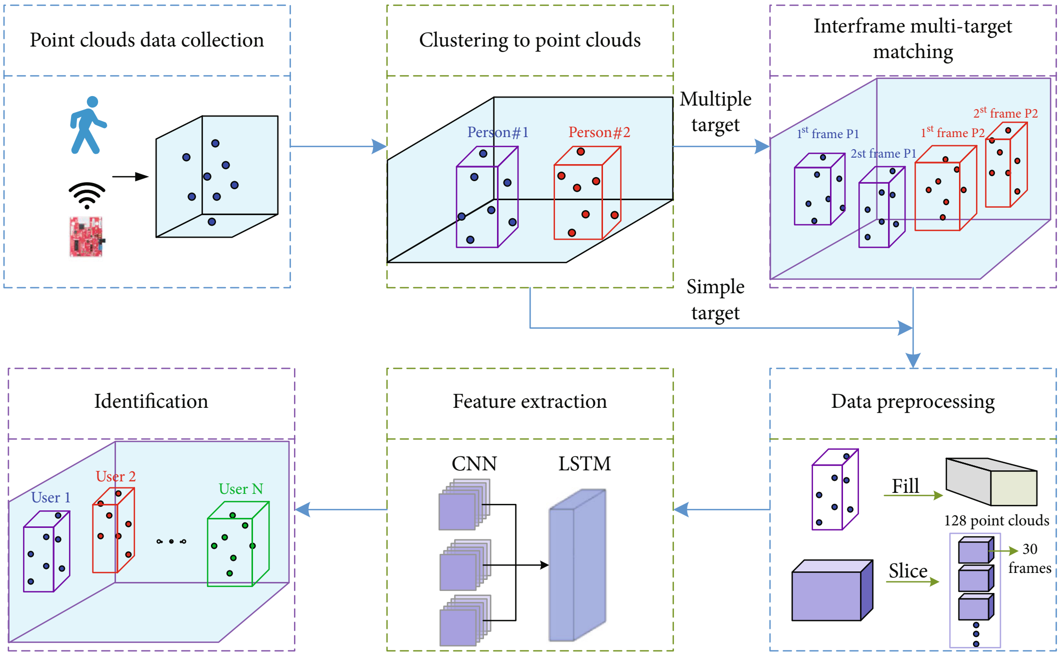


Figure 1: The flow chart of MTPGait.

signal. Then it calculates and generates point clouds based on the reflected signal. In this process, static clutter filtering and Constant False-Alarm Rate (CFAR) algorithms are applied to remove interference point clouds [33]. Interference point clouds include static objects, such as tables and walls, as well as dynamic objects other than pedestrians, such as rotating fans and curtains stirred by the wind

(ii) Point cloud clustering analysis: in this module, a clustering algorithm is applied to gather sparse point clouds into one or more point cloud clusters to form single or double persons point cloud features. If the clustering result is a single person point cloud, go directly to the fourth step; otherwise, go to the third step

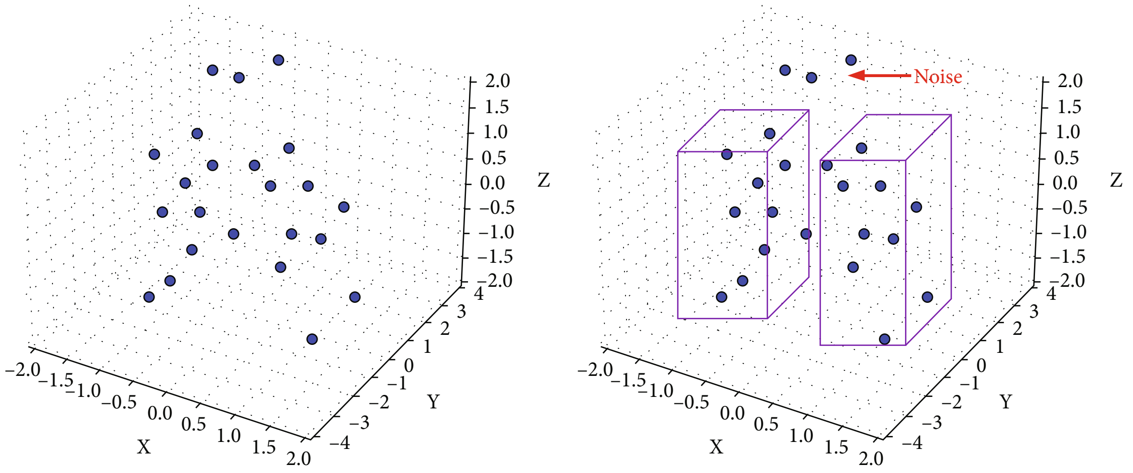
(iii) Multitarget matching between frames: the Hungary algorithm is applied to match multiple point cloud clusters between the upper and lower frames and then obtain multiple continuous point cloud features of the gait trajectory

(iv) Data preprocessing: in this module, we preprocessed the data before training. Frames with less than 128 point clouds are filled to 128 point clouds to ensure consistent length of input data. Then, we slice the point cloud data with the length of 30 frames as a window. In addition, we divide the data into training set and test set according to the ratio of 11 : 1

(v) Feature extraction: we input five attributes of 3D point cloud gait data as five independent channels into spatio-temporal convolution kernel for local feature extraction and then fuse the features of five attributes through fusion network. The global feature is extracted by the long- and short-term memory network

(vi) Target classification: the fully connected layer deduces the score of each classification and finally realizes the classification of the targets

4.1. Point Cloud Clustering and Multitarget Matching between Frames. The gait point cloud collected by millimeter wave radar is sparse and scattered. After applying the static clutter filtering algorithm, most of the static interference point cloud is filtered out. By adjusting the CFAR threshold, most nontarget dynamic interference can also be removed. However, there will still be some dynamic interference that is difficult to filter out. This is because the movement speed of some distractors is high or low, and some are close to the speed of the human body, such as swinging air-conditioning fan blades. This leads to too much dynamic interference when the threshold is too small, and when the threshold is too large, part of the human body point cloud is missing. Fortunately, most of the dynamic interference is at a certain distance from the user. Therefore, we use the DBSCAN clustering algorithm to gather the point cloud data reflected by the human body and remove the remaining noise points. In addition, for double persons point cloud data, it needs to be divided into single-person point cloud data for processing. Therefore, the DBSCAN clustering algorithm is used to cluster and segment double persons point clouds. Figure 2 shows the point cloud data before and after clustering. The DBSCAN algorithm is a classic clustering algorithm based on density perception that does not need to set the number of clusters in advance. It divides the clusters by



(a) The original 3D point cloud of two people

(b) The point cloud of two people after clustering

Figure 2: Point clouds of two people before and after clustering.

calculating the closeness between sample points. When processing point cloud data, the algorithm divides the point cloud according to the Euclidean distance between the point clouds in the three-dimensional space. For point clouds between different people who are far away, such as maintaining a social distance of 0.3-0.6 m or more, the algorithm has better performance. But when the distance is close, the segmentation accuracy of this algorithm decreases.

4.2. Neural Network Structure. The point cloud data is composed of five attributes: distance, speed, pitch angle, horizontal angle, and signal-to-noise ratio. After data analysis and data transformation, we can get the three-dimensional coordinates (X, Y, and Z) of the point cloud. Correspondingly, we can get the five attributes X, Y, Z, V, and N of the point cloud, where X, Y, and Z are the three-dimensional coordinates of the point cloud, V is the speed, and N is the signal-to-noise ratio. Due to the sparse and scattered characteristics of the point cloud itself, the point cloud mapping will generate a large amount of redundant data on the picture, resulting in a rapid increase in network consumption. Therefore, in order to reduce network consumption and improve training speed, we directly use the five attributes of the point cloud as the input data of the network instead of mapping the point cloud to the picture.

In order to effectively extract the spatio-temporal features of point cloud data, we specially design a neural network that can extract multiscale spatio-temporal features along space and time dimensions of 3D point cloud concisely and efficiently. And LSTM is used to design the context flow of local and global time and space, fusing local and global spatio-temporal features. We take the five attributes of the point cloud independently as input and use the spatio-temporal convolution kernel to extract the local spatio-temporal features of the point cloud. However, it is difficult for five independent gait attributes to fully represent the inherent characteristics of gait, so we need to fuse these five attributes. Therefore, in the third layer, we design a fusion network F to fuse the five attributes. After fusion, the features are more comprehensive and can more comprehensively represent the coordination of a person's torso and limbs, the correspondence of speed and spatial position, and the matching of stride frequency and stride length.

$$GF = Cat(F(X), F(Y), F(Z), F(V), F(N)). \tag{9}$$

GF represents the output after the fusion network F, and Cat represents the feature fusion by concat. At the fourth layer, we extract the global temporal features through LSTM. The temporal sequence of the fused features can well represent the global spatio-temporal trajectory and global speed of walking. Finally, a full connection layer network FC is set at the end to derive and calculate the classification score scr. The loss function of the network is

$$L(scr, t) = -\log\left(\frac{\exp\,(scr[t])}{\sum_j \exp(scr[j])}\right) = -scr[t] + \log\left(\sum_j \exp(scr[j])\right). \tag{10}$$

As shown in Figure 3, the network contains a total of six layers. Among them, the layer 0 is the input layer, and the five attributes of X, Y, Z, V, and N are directly input to the layer 1; the layer1 is composed of five independent modules, each of which is composed of a 7×7 spatio-temporal convolution sum 3×3 maximum pooling composition. We use batch normalization followed by the ReLU activation functions after the layer. We use the 3×3 max pooling with 2×2 strides. Next is layer 2 composed of the first layer of five independent ResNet50. Layer 3 uses a 3×3 spatio-temporal convolution and 2×2 average pooling to fuse the five attributes features. The next layer is an LSTM layer with an input size of 256 and a hidden unit of 128 and a dropout of 0.5. After this,

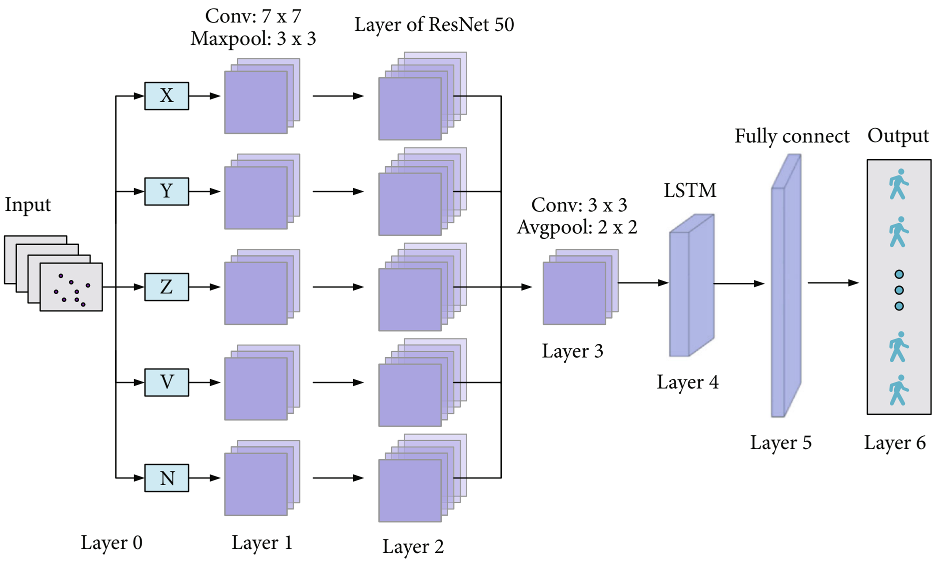


Figure 3: The neural network structure of MTPGait.

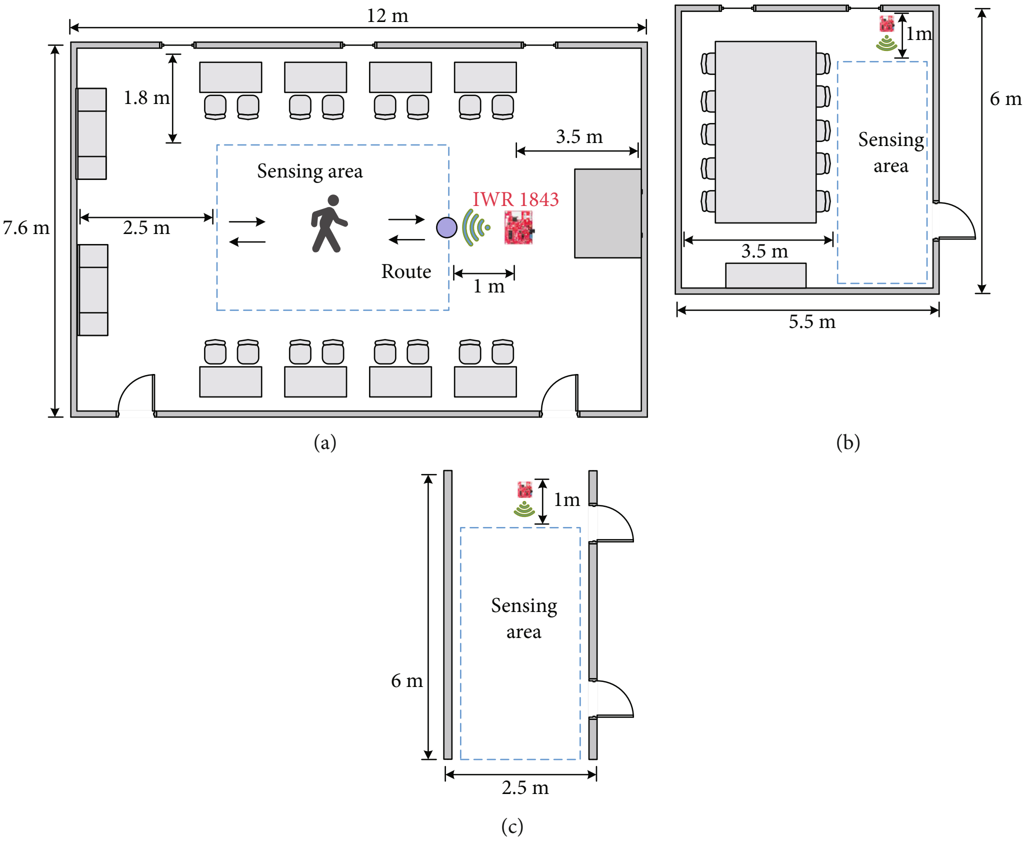


Figure 4: The diagram of the three experimental scenarios. (a), (b), and (c) represent laboratory, conference room, and corridor scenarios, respectively.

it is layer 5 composed of two fully connected layers and a ReLU activation function. The last one layer 6 is the output layer. The initial value of the learning rate lr is 0.05. For each 4 epoch, we set $lr = lr \times 0.005$. The batch size is 256. The optimization function of the network is Adam. We implement our network in PyTorch.

5. Experiment

5.1. Device Parameter Setting. TI's millimeter wave radar IWR1843BOOST is used for 3D point cloud acquisition of human gait. The radar is configured with three transmitting antennas and four receiving antennas, as well as

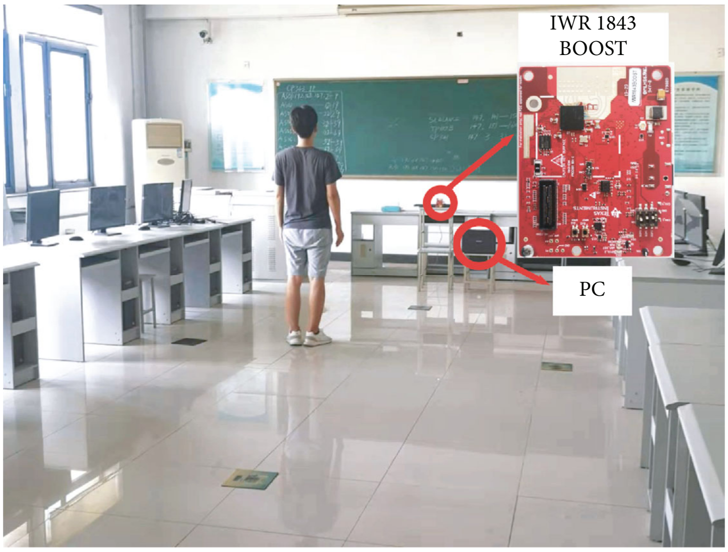


Figure 5: The experimental scene and equipment.

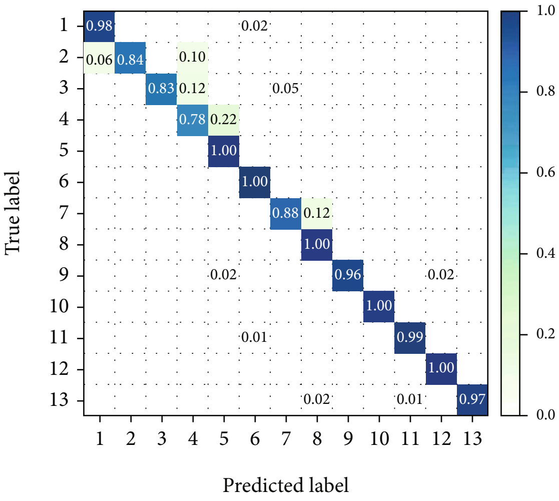


Figure 6: Confusion matrix of 13 people on fixed route.

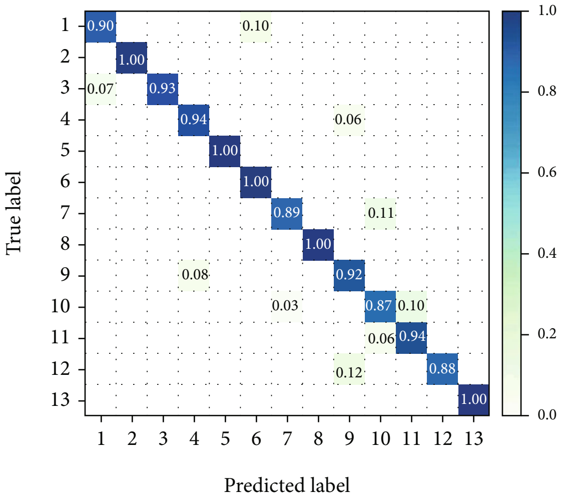


Figure 7: Confusion matrix of 13 people on random route.

a built-in phase-locked loop (PLL) and analog-to-digital converter (ADC). The radar equipment integrates TI's high-performance C674x DSP for radar signal processing. The frequency range is 77GHZ to 81GHZ, with a bandwidth of 4GHZ. We set the frame rate to transmit 10 frames per second and send 128 chirps per frame. The chirp transmission period *tc* is 162.14 microseconds. The frequency slope is set to 70GHZ/ms. The range of azimuth and elevation angle of arrival is between -60 degrees and 60 degrees. Under this parameter configuration, the maximum detection speed of the radar is 2.35 m/s, and the speed resolution is 0.15 m/s. The maximum detection distance is 8 m, and the distance resolution is set to 0.044 m.

5.2. Data Collection. We implemented data collection in three scenarios as shown in Figures 4(a)–4(c) in Figure 4 which are laboratory, conference room, and corridor scenarios, respectively. Figure 5 is a real picture of data collection in the laboratory scene. The size of the collection area in the three scenes is 4 m × 5 m, 2 m × 5 m, and 2.5 m × 5 m, respectively. In the experiment, we used a single commercial radar device to collect single and double persons 3D point cloud gait data on 40 volunteers. The age of the volunteers is between 19 and 29 years old, the height is between 158 and 182 cm, and the weight is between 45 kg and 105 kg. As the millimeter wave radar has certain restrictions on the elevation and horizontal angles, we place the radar at a height of 1 m and keep the data collection area at a minimum distance of 1 m from the radar equipment to improve the coverage of the human body by the radar beam.

The total duration of the data we finally collected was 960 minutes. We completed the data collection in a period of three months. The data collection process is divided into fixed route and random route, and the duration is 200 s each time. Since a single device may cover each other when

Table 2: A comparison of various radar-based personal identification studies with our method.

Method	[17]	[18]	[19]	[20]	[14]	[15]	[13]	Our method
Single-fixed	NA	91%	NA	NA	90%	NA	NA	96.7%
Single-random	93%	90%	85.8%	68.88%	45%	78.46%	89%	94.9%
Two-fixed	NA	NA	NA	NA	86%	NA	NA	90.2%
Two-random	NA	NA	NA	75.51%	NA	NA	NA	87.4%
Number of people identified	8	10	4	7	10	5	12	13
Number of coexisting persons	1	1	1	2	5	1	1	2
Type of input data	Point cloud	Gait spectrum	Point cloud	Point cloud	Point cloud	Micro-Doppler	Point cloud	Point cloud
Number of training data required	40,000 frames	NA	100,000 frames	75 minutes	900 MB	100 minutes	NA	127 MB
Number of experimental scenes	3	1	1	3	2	2	1	3

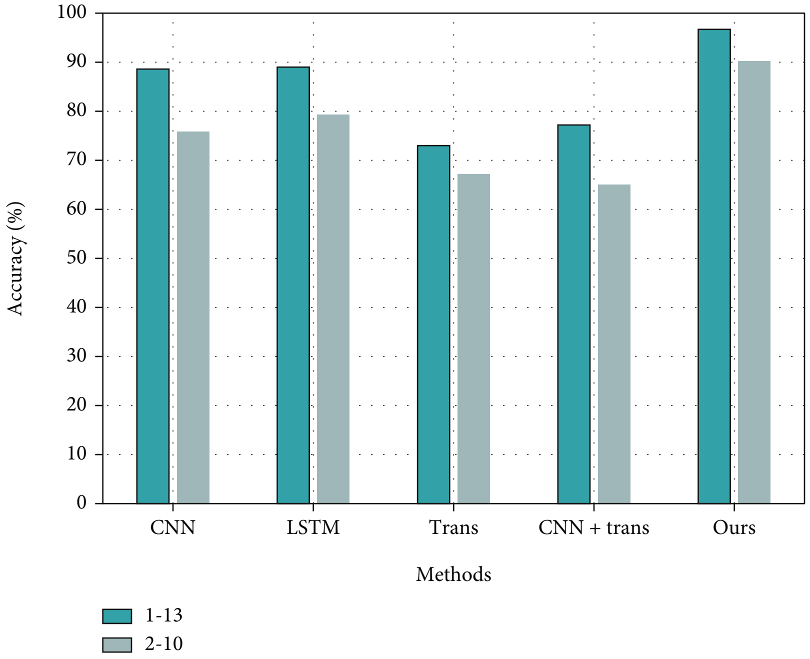


Figure 8: The accuracy of five network structures on fixed route.

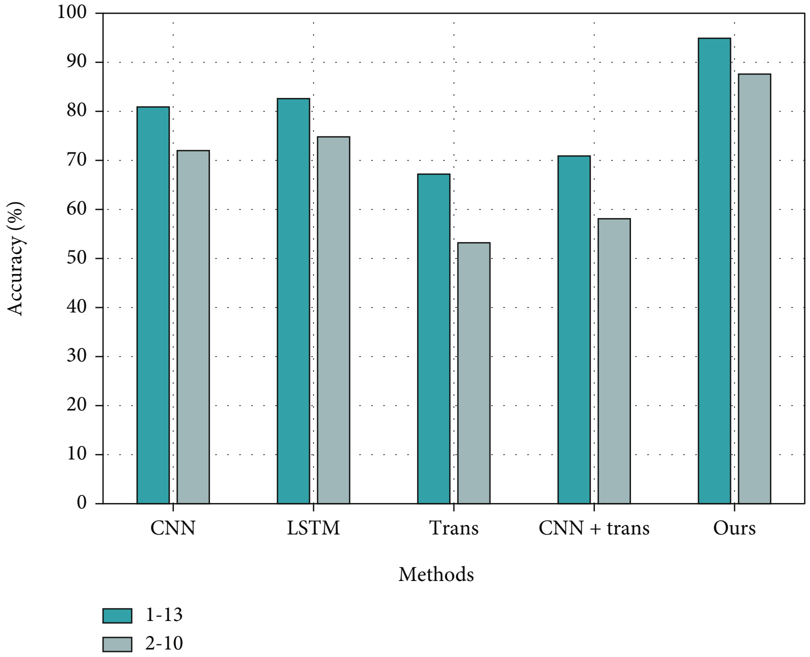


Figure 9: The accuracy of five network structures on random route.

collecting data from double persons, we only consider the short-term coverage between persons in most cases. In this case, when we use DBSCAN clustering, we will skip the multiframe data in such cases. Since the coverage time is short and the number of data frames generated is also less, it will not have a significant impact on the overall accuracy. For the long-term coverage of multiple people, we will adopt a multidevice collection solution in future work.

5.3. Data Preprocessing. Since the number of point clouds in each frame will be different, this will cause the length of the input data to be inconsistent. Therefore, we have filled in each frame of data. If the number of point clouds does not meet 128, the copy is notified that the existing point clouds are filled to 128. When inputting the attribute network for training, we use a 3-second data frame, that is, 30 frames of data as the input of each attribute network, and each

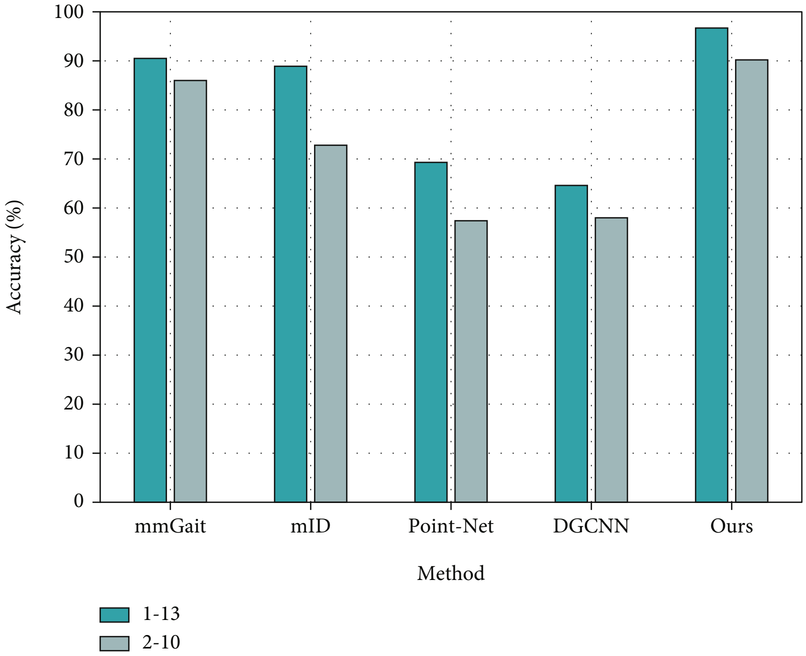


Figure 10: The accuracy of five methods on fixed route.

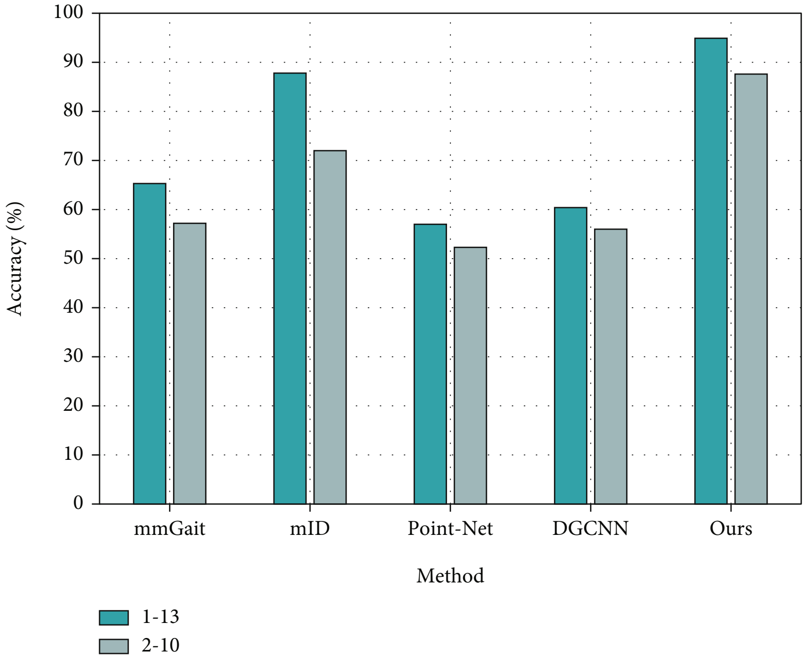


Figure 11: The accuracy of five methods on random route.

frame of data has 128 point clouds. We divide the data into training set and test set according to the ratio of 11 : 1.

6. Evaluation

After the data set was constructed, we evaluated the system performance. Because the 3D point cloud data generated by millimeter wave radar is too sparse and scattered, it is difficult to directly use traditional vision-based methods to recognize various parts of the human body, and the point cloud mapping to the picture will generate a large deal of redundant data resulting in a rapid increase in network consumption. Therefore, we chose to directly consume the point cloud for training. However, the network structure suitable

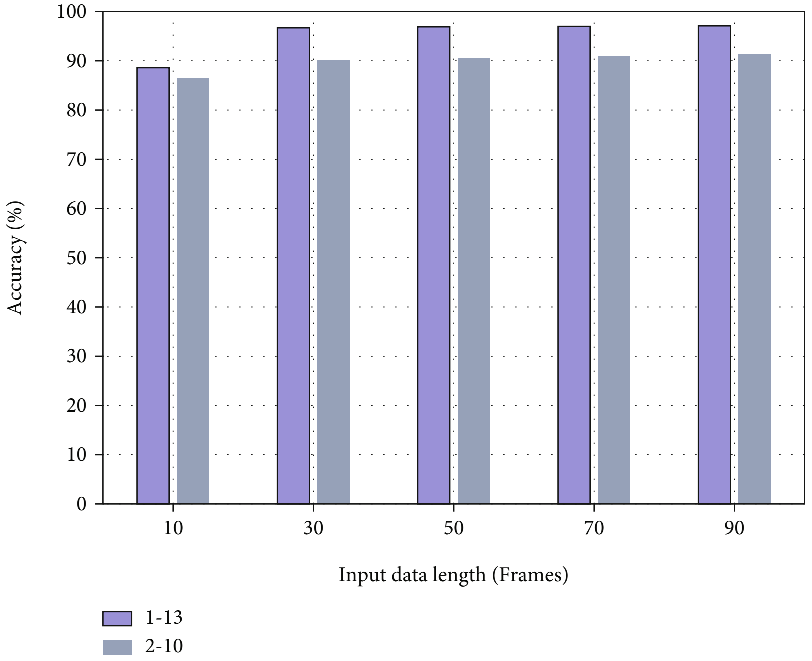


Figure 12: The accuracy of different lengths of input data on fixed route.

for 3D point cloud is not so easy to determine. Therefore, we further analyze the characteristics of 3D point cloud and found that although it is sparse and scattered, it still contains the spatial distribution characteristics of the human body. In addition, the point cloud has the speed attribute, and the person is walking. The movement characteristics of the overall torso and each body part are reflected in the speed attribute, so the point cloud of consecutive frames contains the time sequence law of human walking. Therefore, we guess that a suitable network must have both the ability to extract spatial and temporal features and to merge them. Based on this, we design the network structure of CNN + LSTM. In order to verify our guess, we compare a variety of different types of network structures and different network sizes by conducting ablation research evaluations, which will be discussed below.

6.1. Overall Performance. In general, the accuracy of the system can reach 96.7% and 94.9% in the fixed route and random route with 13 participants. The confusion matrix of the two cases is shown in Figures 6 and 7. In the case of 10 people participating and two people walking at the same time, the system accuracy of the fixed route and the random route reached 90.2% and 87.4%, respectively. This is because when a single device collects data, the greater the number of people in the case of random routes, the more serious the mutual obstruction, the higher the probability of obstruction, and the longer the duration of occurrence. We compare our approach with various conventional studies of radar-based personal identification in Table 2.

6.2. Comparison of Neural Network Structure. In order to verify our conjecture, we compare four different types of network CNN, LSTM, transformer, and CNN + transformer. Among them, CNN and LSTM are the first half and the second half of our network, respectively. Comparing with them, we can effectively verify whether the combination of CNN + LSTM is better than CNN and LSTM alone through ablation studies. In addition, once transformer appeared, it has taken the lead in the field of NLP. In the past two years, it has slowly migrated to the field of vision and achieved good results. There are two main ways to apply transformer in the field of vision. One is the pure transformer structure. For example, Vision Transformer directly applies the pure transformer architecture to a series of image blocks for classification and achieves good results on several larger image data sets. The other is a hybrid structure combining CNNs/backbone network and transformer. For example, DETR [34] is a target detection framework that combines CNN and transformer's pipeline. Therefore, we compare the two structures with transformer and CNN + transformer. Among them, transformer uses ViT-Base in Vision Transformer [35], and CNN + transformer uses DETR.

The five neural network methods are evaluated with accuracy as their evaluation index. The results of the five methods are shown in Figures 8 and 9. Figures 8 and 9, respectively, report the performance of the five algorithms under the fixed route and the random route. Among them, 1-13 means there are 13 volunteers in this experiment, and one volunteer is walking at the same time. 2-10 means there are 10 volunteers in this experiment, and two volunteers are walking at the same time. The experimental results show that, first, the combined effect of CNN + LSTM is significantly better than that of CNN or LSTM alone. This verifies our previous conjecture that a network with both spatial and temporal feature extraction capabilities can better extract gait point cloud features. Second, the application effect of transformer and CNN + transformer on the point cloud data

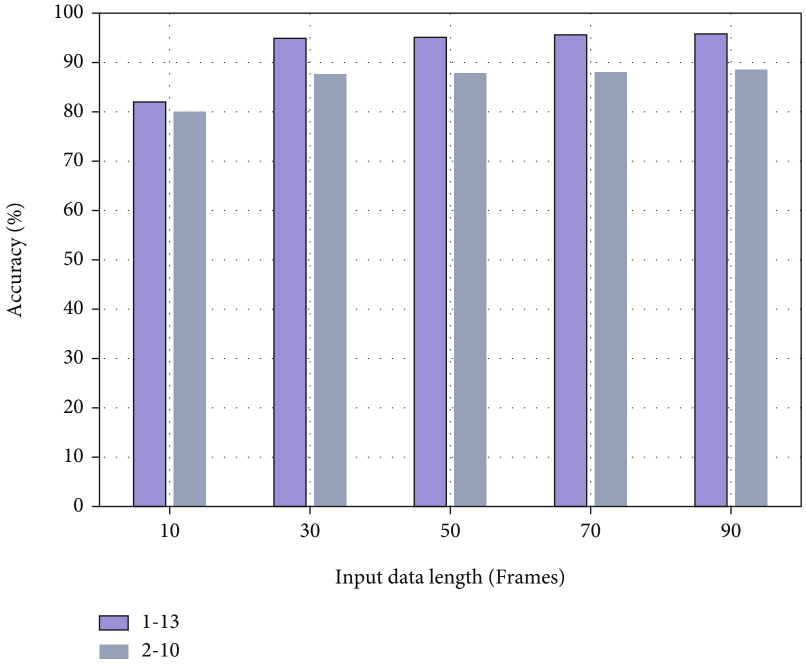


Figure 13: The accuracy of different lengths of input data on random route.

set is not very good. We guess that this is because the mechanism of the transformer limits its ability to converge better on large-scale data sets. Although the total size of our data set has reached 16 hours, it is still a far cry from ImageNet, which has more than 14 million images, and JFT, which has more than 350 million images. Therefore, the transformer is not suitable for radar gait point cloud data sets that do not yet have large-scale public data sets.

6.3. Comparison with Other Point Cloud Processing Algorithms. Four existing point cloud-based classification algorithms mmGait [14], mID [13], Point-Net [36], and DGCNN [37] are used to compare with our algorithm. mmGait and mID classify volunteers based on the point cloud sequence generated by volunteers walking. However, both Point-Net and DGCNN classify based on static object point clouds. Therefore, in order to ensure the objectivity of the comparison, we add LSTM to Point-Net and DGCNN to extract timing features.

We evaluate five algorithms with accuracy as their evaluation index. Figures 10 and 11, respectively, report the single and double persons gait recognition performance of the five algorithms in the fixed route and the random route. Experimental results show that our algorithm performance is better than the other four algorithms overall. Both mmGait and mID have achieved good results in fixed routes, while Point-Net and DGCNN are not very effective. This is because the first two algorithms are designed for gait point clouds and have been tested by corresponding data sets. The latter two are completely designed for static object point clouds. Although LSTM is added, it is difficult to achieve higher accuracy. In addition, by further analyzing the results in Figure 10, it can be seen that the accuracy of mmGait in the random route is much lower than that of the fixed route. This is because mmGait can only achieve 45% accuracy in the case of random routes in the original paper. This shows that the mmGait algorithm itself is difficult to adapt to gait point cloud data under random routes. In addition, the results show that the performance of mID in multiplayer scenes is also very unsatisfactory. This is because mID itself is only for single-player scenes. In the single-player scene in mID, when 12 people participate, a good accuracy of 89% is achieved, but it has not been trained and tested in a multiplayer scene. Therefore, compared with several existing point cloud processing algorithms, our method can achieve high accuracy under both fixed and random routes and can still maintain good performance in double persons situations.

6.4. Impact of Input Data Length. The length of the input data represents the length of a continuous walking trajectory for the volunteers. When the length of input data is short, it cannot fully characterize the characteristics of gait, such as walking speed, spatial trajectory, local limb swing amplitude, stride size, and frequency. When the input data is too long, the training duration will be greatly increased, and the number of total training samples will be reduced. Therefore, proper input data length will greatly improve the training effect of the model. We train the model when the length of input data is 10, 30, 50, 70, and 90 frames, and the accuracy results are shown in Figures 12 and 13. By observing Figures 12 and 13, it can be concluded that the accuracy is significantly improved when the length of input data increases from 10 to 30 frames. After that, with the increase of the length of input data, the accuracy continues to increase, but slowly. However, the increase of input data length will continue to increase the training cost. Therefore, in order to ensure high precision and avoid unnecessary

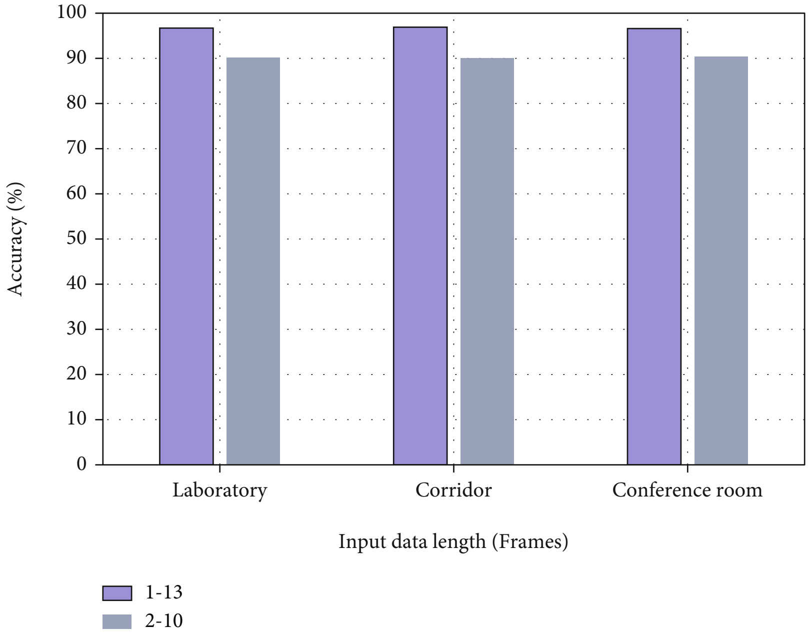


Figure 14: The accuracy of different environments on fixed route.

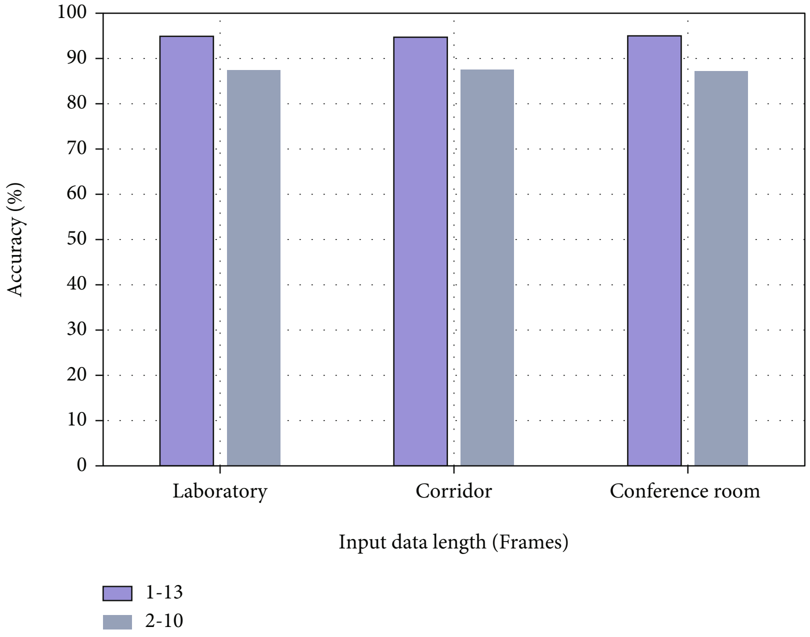


Figure 15: The accuracy of different environments on random route.

training cost, we choose to set the input data length as 30 frames.

6.5. Impact of Different Environments. In order to verify the robustness of the system in different environments, we conduct experiments in three classic environments, including laboratories, corridors, and conference rooms. In three scenarios, we collect a total of 40 volunteers' gait point cloud data. The recognition accuracy of the three scenarios is shown in Figures 14 and 15. Experimental results show that changes in the environment will hardly affect the recognition accuracy. This is because when using millimeter wave radar to collect data, CFAR and static clutter filtering algorithms can be used to remove static point clouds, so that the point clouds of static objects in the environment such as walls, ceilings, tables, and chairs can be removed. Therefore, compared with WiFi-based gait recognition, millimeter wave radar-based gait recognition has greater advantages in environmental independence and robustness.

7. Conclusion

In this paper, we propose a gait recognition system using millimeter wave radar 3D point cloud. Therefore, we design a neural network that can efficiently extract temporal and spatial features of gait. This network greatly improves the recognition accuracy. We collect and publish online a gait data set with a duration of 960 minutes for 40 volunteers. As far as we know, this is the second large-scale public millimeter wave point cloud gait data set besides mmGait. A large number of experiments conducted in laboratory, corridor, and conference room scenarios show that the accuracy of MTPGait can reach 96.7% and 94.9% in the case of fixed route and random routes. Compared with several existing methods, this method can achieve higher accuracy in fixed route and random route, single, and double persons situations.

Data Availability

The data is available at https://github.com/caoxu907/MMWAVE_gait.

Conflicts of Interest

We declare that we have no conflicts of interest.

Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities (No. 2020ZDPY0305).

References

[1] C. Nickel, C. Busch, S. Rangarajan, and M. Möbius, “Using hidden markov models for accelerometer-based biometric gait recognition,” in *IEEE 7th international colloquium on signal processing and its applications*, pp. 58–63, Penang, Malaysia, March 2011.

[2] T. T. Ngo, Y. Makihara, H. Nagahara, Y. Mukaigawa, and Y. Yagi, “The largest inertial sensor-based gait database and performance evaluation of gait-based personal authentication,” *Pattern Recognition*, vol. 47, no. 1, pp. 228–237, 2014.

[3] Z. Zhou, E. Y. Du, N. L. Thomas, and E. J. Delp, “A new human identification method: sclera recognition,” *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 42, no. 3, pp. 571–583, 2011.

[4] P. Airey and J. Verran, “A method for monitoring substratum hygiene using a complex soil: The human fingerprint,” in *Fouling, Cleaning and Disinfection in Food Processing*, Paul Airey and Joanna Verran, Cambridge, UK, 2006.

[5] Y. E. Du, “Review of iris recognition: cameras, systems, and their applications,” *Sensor Review*, vol. 26, no. 1, pp. 66–69, 2006.

[6] S. Li, W. Liu, H. Ma, and S. Zhu, “Beyond view transformation: cycle-consistent global and partial perception gan for view-invariant gait recognition,” in *2018 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6, San Diego, CA, USA, July 2018.

[7] Y. Makihara, A. Suzuki, D. Muramatsu, X. Li, and Y. Yagi, “Joint intensity and spatial metric learning for robust gait recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5705–5715, Honolulu, HI, United states, 2017.

[8] H. Chao, Y. He, J. Zhang, and J. Feng, “Gaitset: regarding gait as a set for cross-view gait recognition,” in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, pp. 8126–8133, Honolulu, HI, United states, 2019.

[9] S. Li, W. Liu, and H. Ma, “Attentive spatial–temporal summary networks for feature learning in irregular gait recognition,” *IEEE Transactions on Multimedia*, vol. 21, no. 9, pp. 2361–2375, 2019.

[10] H. Zou, Y. Zhou, J. Yang, W. Gu, L. Xie, and C. Spanos, “Wifi-based human identification via convex tensor shapelet learning,” in *Thirty-Second AAAI Conference on Artificial Intelligence*, vol. 32, New Orleans, LA, United states, 2018no. 1.

[11] D. Wang, Z. Zhou, X. Yu, and Y. Cao, “CSIID: WiFi-based human identification via deep learning,” in *2019 14th International Conference on Computer Science & Education (ICCSE)*, pp. 326–330, Toronto, ON, Canada, August 2019.

[12] X. Ming, H. Feng, Q. Bu, J. Zhang, G. Yang, and T. Zhang, “Humanfi: Wifi-based human identification using recurrent neural network,” in *2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, pp. 640–647, Leicester, UK, August 2019.

[13] P. Zhao, C. X. Lu, J. Wang et al., “Mid: tracking and identifying people with millimeter wave radar,” in *2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS)*, pp. 33–40, Santorini, Greece, May 2019.

[14] Z. Meng, S. Fu, J. Yan et al., “Gait recognition for co-existing multiple people using millimeter wave sensing,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, pp. 849–856, New York, NY, United states, 2020.

[15] B. Vandersmissen, N. Knudde, A. Jalalvand et al., “Indoor person identification using a low-power FMCW radar,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 7, pp. 3941–3952, 2018.

[16] H. T. Le, S. L. Phung, and A. Bouzerdoum, “Human gait recognition with micro-Doppler radar and deep autoencoder,” in *2018 24th International Conference on Pattern Recognition (ICPR)*, pp. 3347–3352, Beijing, China, August 2018.

[17] X. Jiang, Y. Zhang, Q. Yang, B. Deng, and H. Wang, “Millimeter-wave array radar-based human gait recognition using multi-channel three-dimensional convolutional neural network,” *Sensors*, vol. 20, no. 19, p. 5466, 2020.

[18] J. Wu, J. Wang, Q. Gao, M. Pan, and H. Zhang, “Path-independent device-free gait recognition using mmwave signals,” *IEEE Transactions on Vehicular Technology*, vol. 70, no. 11, pp. 11582–11592, 2021.

[19] Z. Xia, G. Ding, H. Wang, and F. Xu, “Person identification with millimeter-wave radar in realistic smart home scenarios,” *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2021.

[20] C. Wang, P. Gong, and L. Zhang, “Stpointgcn: spatial temporal graph convolutional network for multiple people recognition

using millimeter-wave radar," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3433–3437, Singapore, Singapore, May 2022.

[21] W. Wang, A. X. Liu, and M. Shahzad, "Gait recognition using wifi signals," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 363–373, Washington, DC, United states, September 2016.

[22] T. Xin, B. Guo, Z. Wang, M. Li, Z. Yu, and X. Zhou, "Freesense: indoor human identification with Wi-Fi signals," in *IEEE Global Communications Conference (GLOBECOM)*, pp. 1–7, Washington, DC, USA, December 2016.

[23] A. Pokkunuru, K. Jakkala, A. Bhuyan, P. Wang, and Z. Sun, "Neuralwave: gait-based user identification through commodity WiFi and deep learning," in *IECON 2018-44th Annual Conference of the IEEE Industrial Electronics Society*, pp. 758–765, Washington, DC, USA, October 2018.

[24] Y. Zeng, P. H. Pathak, and P. Mohapatra, "Wiwho: WiFi-based person identification in smart spaces," in *15th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, pp. 1–12, Vienna, Austria, April 2016.

[25] J. Zhang, B. Wei, W. Hu, and S. S. Kanhere, "Wifi-id: human identification using wifi signal," in *2016 International Conference on Distributed Computing in Sensor Systems (DCOSS)*, pp. 75–82, Washington, DC, USA, May 2016.

[26] M. Ritchie, F. Fioranelli, H. Borrion, and H. Griffiths, "Multistatic micro-Doppler radar feature extraction for classification of unloaded/loaded micro-drones," *IET Radar, Sonar & Navigation*, vol. 11, no. 1, pp. 116–124, 2017.

[27] F. Guidi, A. Guerra, and D. Dardari, "Personal mobile radars with millimeter-wave massive arrays for indoor mapping," *IEEE Transactions on Mobile Computing*, vol. 15, no. 6, pp. 1471–1484, 2016.

[28] C. Wu, F. Zhang, B. Wang, and K. R. Liu, "mmtrack: passive multi-person localization using commodity millimeter wave radio," in *IEEE INFOCOM 2020-IEEE Conference on Computer Communications*, pp. 2400–2409, Toronto, ON, Canada, July 2020.

[29] C. Wu, F. Zhang, B. Wang, and K. R. Liu, "mSense: towards mobile material sensing with a single millimeter-wave radio," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 3, pp. 1–20, 2020.

[30] C. Jiang, J. Guo, Y. He, M. Jin, S. Li, and Y. Liu, "mmvib: micrometer-level vibration measurement with mmwave radar," in *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, pp. 1–13, London, United Kingdom, April 2020.

[31] F. Wang, F. Zhang, C. Wu, B. Wang, and K. R. Liu, "Vimo: multiperson vital sign monitoring using commodity millimeter-wave radio," *IEEE Internet of Things Journal*, vol. 8, no. 3, pp. 1294–1307, 2020.

[32] H. Liu, Y. Wang, A. Zhou et al., "Real-time arm gesture recognition in smart home scenarios via millimeter wave sensing," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 4, pp. 1–28, 2020.

[33] M. A. Richards, J. Scheer, and W. A. Holm, *Principles of Modern Radar*, M. A. Richards and W. L. Melvin, Eds., Raleigh, NC, USA: SciTech Pub, 2010.

[34] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *European Conference on Computer Vision*, pp. 213–229, 2020.

[35] A. Dosovitskiy, L. Beyer, A. Kolesnikov et al., "An image is worth 16x16 words: transformers for image recognition at scale," 2020, http://arxiv.org/abs/2010.11929.

[36] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: deep learning on point sets for 3d classification and segmentation," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, United states, 2017.

[37] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, "Dynamic graph cnn for learning on point clouds," *Acm Transactions On Graphics (tog)*, vol. 38, no. 5, pp. 1–12, 2019.