

Research Article

Multiscale Bidirectional Input Convolutional and Deep Neural Network for Human Activity Recognition

Yishu Qiu, Lanliang Lin , Lvqing Yang , Dingzhao Li, Runhan Song, Gengchen Xu, and Shaoqin Shen

School of Informatics, Xiamen University, Xiamen, Fujian Province 361005, China

Correspondence should be addressed to Lvqing Yang; lqyang@xmu.edu.cn

Received 15 July 2021; Revised 9 August 2021; Accepted 23 August 2021; Published 18 October 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Yishu Qiu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this paper, we proposed a multiscale and bidirectional input model based on convolutional neural network and deep neural network, named MBCDNN. In order to solve the problem of inconsistent activity segments, a multiscale input module is constructed to make up for the noise caused by filling. In order to solve the problem that single input is not enough to extract features from original data, we propose to manually design aggregation features combined with forward sequence and reverse sequence and use five cross-validation and stratified sampling to enhance the generalization ability of the model. According to the particularity of the task, we design an evaluation index combined with scene and action weight, which enriches the learning ability of the model to a great extent. In the 19 kinds of activity data based on scene+action, the accuracy and robustness are significantly improved, which is better than other mainstream traditional methods.

1. Introduction

The research of using sensors for human activity recognition existed as early as 30 years ago [1, 2]. With the advent of industry 4.0 era, new opportunities and challenges are filled. Human activity recognition research has attracted much attention due to its advantages in intelligent monitoring system, medical care system [3], virtual reality exchange, smart homes [4], anomaly detection [5], and other fields, as well as the ability to provide personalized support and interconnection for different fields. At present, human activity recognition is mainly realized by two ways: one is through indoor and outdoor sensors, and the other is some wearable devices [6–8]. The former is limited by the need to be placed in a fixed location, and the inference of activity completely depends on the user's interaction with these devices. For example, if the user is not within the sensor range or the object moves freely in the scene to introduce varying degrees of occlusion, the activity cannot be recognized. Secondly, the environment is dynamic and complex, such as the weather and sunlight in the background, which also increases the difficulty of recognition. The latter also has many defects, such as high cost and inconvenient carrying.

Smart phones have many advantages in the field of human activity recognition [9]. Due to its small size and convenient portability, the built-in sensors are becoming more and more diverse, and specific types of activities can be effectively classified through the information of multiple sensors. For example, the built-in accelerometers of smart phones [10, 11] can describe human actions, such as standing, walking, and running [12, 13]. Similarly, by collecting audio information from the phone microphone [14], the user's activities can be identified, such as listening to music, speaking, and sleeping [15, 16], running rhythm can be monitored [17], and user respiratory symptoms are related to sound, such as sneezing or coughing [15]. Users' emotions can also be inferred from various sensor data, including Wi-Fi, accelerometer, compass, and GPS [15, 18–20]. The activity identification system based on mobile devices can be perceived from personal perception to group in multiscale [21]. Generally speaking, the quality of data collected by each built-in sensor of a smart phone is also different. For example, the gyroscope of smart phone can sense the change of the movement direction of the person holding the phone, while the acceleration sensor of smart phones can reflect

the speed change of the person holding the phone. Therefore, by fusing the real-time data obtained by the various sensors of the smart phone, the final data obtained has a certain heterogeneity. These data can be widely used in human activity recognition and have broad market and social value in health care, smart home, financial fraud, and other scenarios. At the same time, the user does not need to carry additional equipment with sensors in the process of collecting data. In terms of human activity recognition, using smart phones as research equipment has become the preferred equipment for researchers.

We study the indoor and outdoor human activity recognition, redefine activity as the combination of scene and action, and achieve real-time monitoring of users' indoor and outdoor activity through smart phones.

- (1) From the study of recognition simple activity to recognition complex activity, activity is defined as the combination of scene and action, which has 19 different activities. The aggregation feature is designed manually to help automatic feature extraction and realize more abundant feature information extraction
- (2) A multiscale bidirectional sequence fusion model (MBCDNN) is designed to recognition complex activity. The experiment result shows that the model has higher advantages than the current popular network model

2. Related Work

2.1. Dataset Collected by Smart Phone. In the research of human activity recognition, the data collected by a smart phone through sensors is particularly important. Some of these benchmark datasets have been released to the public. In studies, such as Roggen et al.'s [22], using 72 environmental and body sensors, a set of daily activities were recorded in the sensor rich environment. Similarly, the other researchers have provided dataset, such as Tapia et al. [23] and Hasegawa[24]. In 2012, WISDM Lab released wisdm dataset [25], the device for collecting data is an Android smartphone, and the number of users participating in data collection is 29. Users participating in data collection are required to put the smartphone in the front trouser leg pocket, which is more in line with the real-life scene. Users need to complete a series of actions, including walking, jogging, going up and down the stairs, standing, and sitting. In 2013, the University of Genoa provided UCI-HAR dataset [26] to record the daily activity of each user, including standing, sitting, walking, going up and down the stairs, and lying down. With the help of the acceleration sensor and gyroscope of the smart phone, the data collection device collects 30 users, aged from 19 to 48, and each user is instructed to wear a smart phone at the waist. Each user needs to carry out two experiments. In the first experiment, the smart phone is worn on the left side of the waist, and in the second experiment, the user can freely choose the position. In 2016, Vavoulas and others offered MobiAct dataset [27], Mobiact data is an extension of the MobiFall data set published in

2014 and was initially created with fall detection in mind. This dataset contains four different types of falls and nine daily activities, such as walking, standing, and going up and down the stairs. The number of users collected is 57, including 42 men and 15 women. The users are between 20 and 47 years old, including more than 2500 experiments, all from the collection of smart phone sensors. In the same year, Hnoohom and others provided the UniMiBSHAR dataset [28]. In the authors' opinion, the data collected by smart phone sensors are rarely public, and public data often contain samples of users with too similar characteristics and lack of specific information. They proposed a new dataset, which collected daily activities of users mainly including simple activity such as walking, standing, running, and sitting down and complex activity such as washing dishes, combing hair, and preparing sandwiches. The dataset contains 11771 human activities and 30 users aged between 18 and 60, of which 24 are women and 6 are men. It is worth noting that it includes more elderly people. In 2019, Beijing University of Posts and Telecommunications provided the Sanitation dataset [29]. The user receiving the acquisition needs to wear a smart watch at the wrist. The user selected according to the demand is the sanitation worker. According to the research requirement, sanitation workers were invited as a user, in which 7 kinds of daily life action data that were collected included walking, running, sweeping with a big broom and sweeping with a small broom, cleaning, and taking out garbage.

2.2. HAR Based on AI. Reference [30] collected data from 10 volunteers (4 women and 6 men) through four sensors of smart phones. The volunteers are between 24 and 30 years old. Volunteers are required to carry smart phones to complete the specified actions, standing, walking, walking slowly, go upstairs and downstairs and cycling. In order to eliminate errors, discard the data of the first 2 seconds and the second after the volunteer starts the action, and select 50050 data for each action. They processed data using median filtering, data normalization, sliding window segmentation, feature selection, and optimal feature subset selection are performed on the data. Also, they used the idea of fusion. The traditional random forest (RF) [31], support vector machine (SVM) [32], k -nearest neighbor (KNN) [33], and naive Bayes classification algorithm (NBC) [34] are fused. The accuracy was 99%, which is 6% higher than that of single model

With the development of artificial intelligence, deep learning has become a favorite choice of researchers. Reference [35] used smart phone sensors to collect 6 different activities from 30 volunteers, including walking, going upstairs and downstairs, standing, lying flat, and sitting. The collected data are preprocessed by noise filter, and 2.56 seconds are used as sliding window. A total of 7352 training data and 2947 test data are divided. Three design optimization schemes were proposed to improve the accuracy of human activity recognition: (1) Scheme 1—combining three sensor data; (2) Scheme 2—based on three-dimensional convolution; and (3) Scheme 3—based on difference optimization convolution kernel. By comparing the results of the experimental scheme, it was found that the accuracy of the first scheme is the highest, with an average accuracy of 97.50%.

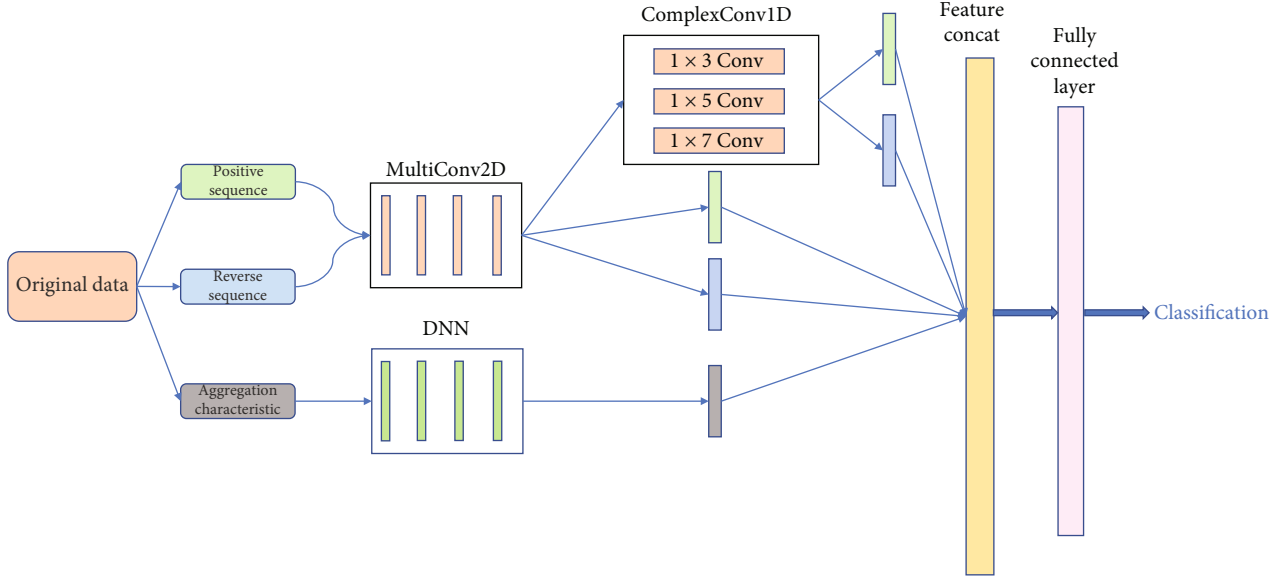


FIGURE 1: MBCDNN model.

Due to the complexity of human activities, there are few researches on excessive activities in the past, such as sitting-standing. Reference [36] proposed a hierarchical recognition model based on support vector machine and random forest, studying 4 kinds of excessive activities and 10 kinds of daily life activities. The data preprocessing adopted filtering denoising and acceleration separation, and the feature extraction adopted the time domain and frequency domain of features, in which the frequency domain uses a total of 118 features, such as Fourier realization, mean, variance, and standard deviation. In the self-collected data, support vector machine and random forest hierarchical recognition model achieved an average accuracy of 98.1%. LSTM is good at processing time series data. The author used bidirectional LSTM network for activity recognition research and achieved 94.1%, 98.5%, and 98.8% accuracy in UCI-HAR, WISDM, and self-collected data, respectively.

3. Method

3.1. MBCDNN Model. Human activity data collected based on scenes+actions are more complex. Since the collected data is time series data, we send the data of the forward sequence and the reverse sequence to the MultiConv2D module for training. Due to the different activity duration, the collected data length is not consistent. Different experiments in the past have segmented the data, trying to find the best segmentation length [15, 37], but in order to ensure the consistency of the sequence fragments in the division process, the filling strategy is adopted. The obvious problem is that the noise interference is increased. Therefore, we introduce the activity convolution module ComplexConv1D, which uses a series of designed one-dimensional convolution kernels to perform convolution operations on forward and reverse sequences to extract diversified features; the interference caused by noise is largely reduced. In addition to the forward and reverse sequences as input to the model, we also added a series of aggregation

features, including three types of features: time domain, frequency domain, and time-frequency domain, which perform feature extraction on the original data, these extracted aggregation features are input into the deep neural network (DNN) module, and this is verified to be effective. Our MBCDNN model is composed of MultiConv2D and ComplexConv1D and DNN modules. Finally, the output results of all modules are connected, and the classification result is obtained through the fully connected layer. The overall structure of the model is shown in Figure 1.

3.1.1. MultiConv2D and ComplexConv1D. The task of the MultiConv2D module is to receive the input of the forward sequence and the reverse sequence, followed by four convolutional layers (Conv2D); the kernel size is 3×3 ; the number of convolution channels is 64, 128, 256, and 512, respectively; the activation function is ReLU. Each layer of convolution passes through the Batch Normalization (BN) layer, then passes through the pooling layer. This module extracts the feature of actions and scenes through a series of convolution operations. The formula for two-dimensional convolution is as follows:

$$\text{conv}_{x,y} = f \left(\sum_i^{k \times k} w_i \times v_i \right), \quad (1)$$

where x, y are the spatial coordinates (x, y) of the input data sample, f is the activation function, the weight of the convolution kernel is w , the size of the convolution kernel is $k \times k$, and the sample data value is v . The convolution process is the sum of the inner product of the weight of the convolution kernel sliding on the value corresponding to the sample data.

The task of the ComplexConv1D module is to extract the features of MultiConv2D from multiple angles when the sequence length is inconsistent, so as to reduce the impact of noise on training. When the input vector dimension is

TABLE 1: MultiConv2D & ComplexConv1D network parameters. Each layer of convolution passes through the BN layer; the selected activation function is ReLU.

Layer name	MultiConv2D	Layer name	ComplexConv1D
	Kernel size, filters/stride, pad		Kernel size, filters/stride, pad
Conv_1 MaxPooling	$3 \times 3, 64/1, \text{same } 2 \times 2/2$	Conv1d_1	$1 \times 7, 128/1, \text{same}$
Conv_2	$3 \times 3, 128/1, \text{same}$	Conv1d_2	$1 \times 5, 256/1, \text{same}$
Conv_3	$3 \times 3, 256/1, \text{same}$	Conv1d_3	$1 \times 3, 128/1, \text{same}$
Conv_4	$3 \times 3, 512/1, \text{same}$		GlobalAveragePooling1D
	GlobalAveragePooling2D		

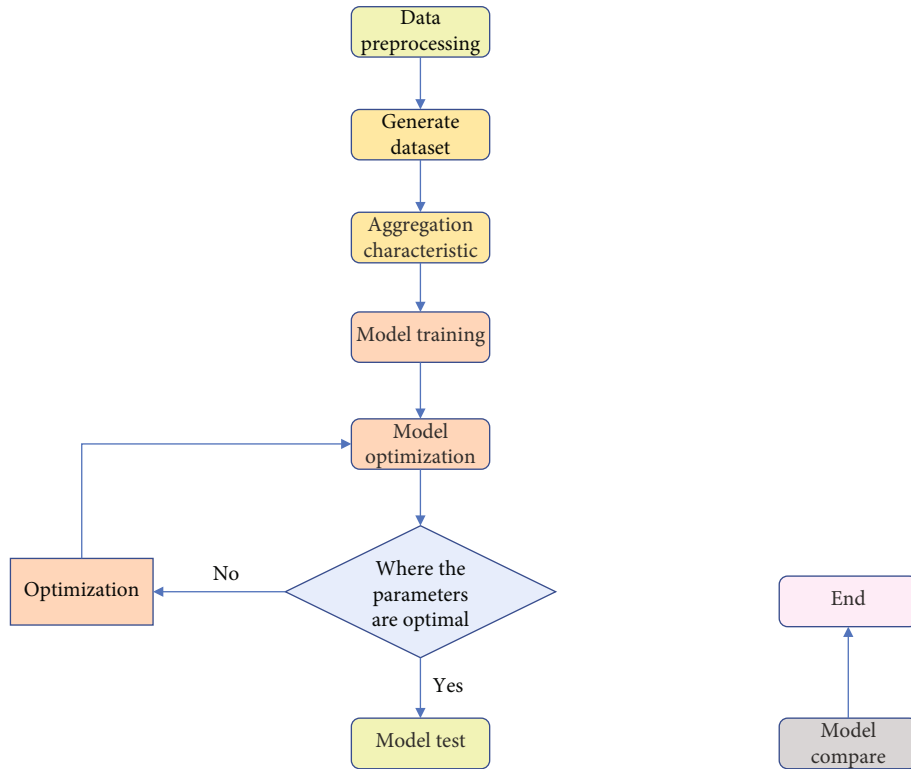


FIGURE 2: Experimental flowchart of MBCDNN.

(N, C_{in}, L_{in}) , the output vector dimension is (N, C_{out}, L_{out}) . As shown in

$$L_{out} = \left\lfloor \frac{L_{in} + 2 \times p - d \times (k - 1) - 1}{s} + 1 \right\rfloor, \quad (2)$$

where L_{in}, L_{out} represent the length of the input and output vectors, respectively, p represents the padding size, d represents the distance between the core points, also called à trous algorithm [38], k represents the size of the convolution kernel, and s represents the convolution stride. To illustrate the relevant parameters of the model, Table 1 lists the parameters of MultiConv2D and ComplexConv1D in detail. The experimental process of the MBCDNN model is shown in Figure 2. The data processing, generate dataset,

and aggregation characteristic part will be described in this section; the rest will be described in the next section.

3.2. Data Processing. The data collected by the smart phone sensor is time series data; for this reason, the research and design collection scenes are divided into 3 categories: walking, standing, sitting, and lying; each scene collects 6 kinds of actions, playing games (mobile game), watching short videos, watching live broadcasts or watching long videos (similar to news broadcasts), browser query or viewing browser content, typing chat or other typing, and other actions (WeChat calls): 6 types of actions. At the same time, an action that is regardless of scene is added, the activity of user A “handing the phone” to user B. We redefine activity as a combination of scene+action, so there are 19 different activities in total. A total of 7292 samples were collected. 80% of the overall data were randomly selected as the training set and 20% as the test set. The sample

TABLE 2: Predefined activity based on smart phone sensors.

Scene	Action	Activity description
Walking/standing/sitting and lying	Playing games	User plays the game in this scene (mobile game)
	Watching short videos	User watches short videos in this scene
	Watching live broadcasts or watching long videos	User watches live broadcast or long videos in this scene (similar to news broadcast)
	Browser query or viewing browser content	User views browser content in this scene
	Typing chat or other typing	User types text in this scene
	Other actions	User makes WeChat calls in this scene, etc.
Regardless of scene	Hand phone	User hands the phone to another user

categories are unbalanced. The proportion of standing playing games is the most, and the proportion of standing brush video is the lowest. Table 2 describes the situation of each activity. Figure 3 shows the proportion of various activities types of collected data.

3.3. Generated Dataset. Use the two built-in sensors of the smart phone: gravity sensor and acceleration sensor to collect data. The volunteers who collected the data were composed of 5 men and 5 women, aged between 20 and 55 years. During the collection process, the volunteers can complete 6 types of actions according to each scene (see Table 2 for details). The collected data includes gravitational acceleration, and nongravitational acceleration data and the action segment are divided by the accelerometer count; every 5 seconds is regarded as a segment. The basic attributes of the collected dataset are described in Table 3. A total of 19 actions were collected based on smart phone sensors. A total of 7292 samples were collected. 80% of the overall data were randomly selected as the training set and 20% as the test set.

3.4. Aggregation Feature. The data is time series data, which needs to be divided into activity segments. Each activity segment data contains 60 pieces of data, which can be used for feature aggregation. Aggregation time series data features generally select three types of features: time domain, frequency domain, and time-frequency domain; they have been verified to be effective. In time domain, the independent variable is time, and the dependent variable is the change of the signal, which describes the value of the signal at different moments. In frequency domain, the independent variable is the frequency, the dependent variable is the amplitude of the signal, and it describes the spectrogram. Among them, the sample features that can be extracted in the time domain include variance, standard deviation, mean, and skewness. The median frequency, average frequency, and energy spectral density of sample features can be extracted in the frequency domain. The specific calculation formula for sample feature is shown in Table 4.

4. Experiment and Analysis

4.1. Evaluation Index. Human activity recognition based on scenes and actions cannot simply use the recognition accuracy to evaluate the quality of the model. Since the activity involves actions under the scene, predicting the scene also

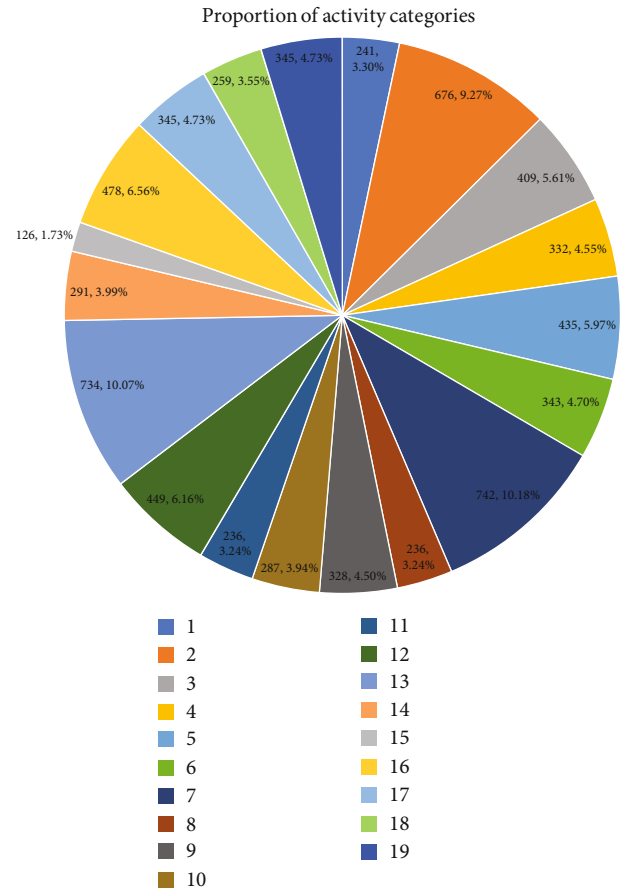


FIGURE 3: Proportion of activity categories. Labels 1 to 6, respectively, represent six types of actions in scene walking: playing games, watching short videos, watching live or long videos, browser query or viewing browser content, typing chat, or other typing and other actions. Labels 7 to 12 represent six types of actions in scene standing. Labels 13 to 18 represent six types of actions in sitting and lying. Label 19 represents the action of delivering the mobile phone alone.

has a certain value. For this reason, we designed an evaluation index that comprehensively considers the weight of the scene and the action: acc_combo, which is more conducive to the recognition efficiency of the real reaction model. The specific rules of the evaluation index are that if the activity of handing over the mobile phone is predicted to get 1

TABLE 3: Data attribute description table.

Num	Attribute name	Data type	Data description
1	fragment_id	int	Activity fragment ID
2	activity_id	int	Activity ID
3	time_point	float	Acquisition time point (ms)
4	acc _x	float	X-axis component without gravitational acceleration (m/s ²)
5	acc _y	float	Y-axis component without gravitational acceleration (m/s ²)
6	acc _z	float	Z-axis component without gravitational acceleration (m/s ²)
7	acc _{xg}	float	X-axis component containing the acceleration of gravity (m/s ²)
8	acc _{yg}	float	Y-axis component containing the acceleration of gravity (m/s ²)
9	acc _{zg}	float	Y-axis component containing the acceleration of gravity (m/s ²)

TABLE 4: Calculation of time and frequency domain features.

Feature	Calculation formula	Meaning
Mean	$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$	The mean of the activity segment data
Variance	$\sigma = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$	The variance of activity segment data
Standard deviation	$\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$	The standard deviation of the activity segment data
Skewness	$\frac{\sum_{i=1}^n (x_i - \bar{x})^3}{(n-1)\sigma^3}$	The skewness of the activity segment data
Quartile deviation	$Q_3 - Q_1$	The difference between the third quartile and the first quartile
Power spectral density	$P_s(w) = \frac{ F_T(w) ^2}{T}$, $F_T(w)$ This is the Fourier transform formula	Power in unit frequency band
Median frequency	$\frac{\int_0^{\infty} fP(f)df}{\int_0^{\infty} P(f)df}$	Median of power spectral density
Average frequency	$\int_0^{\infty} P(f)df$	Mean value of power spectral density
Energy spectral density	$G(f) = F_T(w) ^2$	Signal energy in unit frequency band

point, the prediction error is 0 point. We hope to predict the correct action to get a relatively high positive feedback in the model evaluation. If it is not the activity of handing the phone, the scene+action is completely correct to get 1 point, only the correct scene is predicted to get 1/7 point, and only the correct action is predicted to get 1/3 point; if the scene+action are not correct, the score of acc_combo is 0 points. The acc_combo is expressed by

$$\text{acc}_{\text{combo}} = \frac{\sum_{i=1}^N S(i) + A(i)}{N}, \quad (3)$$

where N represents the total number of prediction samples, S represents the score of the predicted scene, and A represents the score of the predicted action.

4.2. Model Train and Optimization. Through continuous experimentation to explore, the model is upgraded step by step and finally proposed a Activity Bidirectional input Convolution and Deep Neural Network model: MBDCNN, and the effect is the best among all comparative experiments. The data collected by the smart phone sensor contains components of gravitational acceleration (acc_x, acc_y, acc_z) and components that do not contain gravitational acceleration (acc_{xg}, acc_{yg}, acc_{zg}) (see Table 3). In order to increase the effect of the model, as shown in formulas (4), the components are combined into a vector.

$$\begin{aligned} \text{mod} &= \text{acc}_x^2 + \text{acc}_y^2 + \text{acc}_z^2, \\ \text{mod } g &= \text{acc}_{xg}^2 + \text{acc}_{yg}^2 + \text{acc}_{zg}^2. \end{aligned} \quad (4)$$

TABLE 5: Parameter exploration process.

Version	Model	Input	Modify parameter	acc_combo
1	MultiConv2D	Origin data	Adam \ GlobalMaxPooling2D	0.730
2	MultiConv2D	Origin data	Modify the optimization parameter to rmsprop	0.746
3	MultiConv2D	Origin data	Add the BN layer to the convolution layer	0.786
4	MultiConv2D	Origin data	GlobalAveragePooling2D	0.792
5	MultiConv2D	Forward and reverse sequence	—	0.812

TABLE 6: Model optimization process.

Model	Input	Optimized content	acc_combo
MultiConv2D+DNN	Forward, reverse, and aggregation feature	Added reverse sequence and aggregation features	0.846
MultiConv2D+DNN	Forward, reverse, and aggregation feature	Data enhance	0.862
MBCDNN	Forward, reverse, and aggregation feature	Activity module	0.887

TABLE 7: Comparison results of different model.

Model	Input	acc_combo
Lightgbm	Aggregation feature data	0.774
DNN	Aggregation feature data	0.713
MCNN	Forward and reverse sequence	0.834
CNN	Forward and reverse sequence	0.812
ResNet	Forward sequence	0.828
VGG19	Forward sequence	0.435
Bidirectional LSTM	Forward and reverse sequence	0.708
MBCDNN	Forward, reverse and aggregation feature	0.887

Take the component and mod, mod g as the input data of the model. To reduce model deviation and overfitting problems, 5-fold cross-validation and stratified sampling are used to ensure that the proportions of different categories in each compromise are equal. This section takes the MultiConv2D module in the MBCDNN model as a baseline. The optimization methods of the model include RMSProp, adding the BN layer after the convolutional layer, GlobalAveragePooling2D, and combining the input of the forward sequence and the reverse sequence. Among them, adding a reverse sequence to the input part has the best effect. The analysis of the reasons shows that the diversified input facilitates the model to more fully extract the feature of activity. The improvement of the specific experimental model is shown in Table 5. In the preliminary work, it is found that the effect of model fusion is better than the effect of single model. So we optimized the model from the following aspects:

- (1) Combine the input of the MultiConv2D module with the input of the DNN module. The input model data is time series data and aggregation feature data. The time series data is input to the MultiConv2D module, and the feature data is input to the DNN module
- (2) For the input part, due to the existence of various scales of activity sequence fragments, a single padding length

cannot be used, and for too long or too short fragments, interception or padding will bring a lot of noise, so we built a activity module: ComplexConv1D, to make up for the impact of noise and enrich the model learning ability

- (3) Data enhancement (noise enhancement, cubic spline interpolation) increases the generalization ability of the model

Table 6 shows the model optimization process.

According to Tables 5 and 6, the effect of multiple models is obviously better than that of single model. Adding reverse sequence and aggregation feature data to the input data has significantly improved the score. Data enhancement is also an important method for score improvement. The addition of activity modules has further improved the score.

4.3. Model Comparison. We improve our model based on the idea of convolutional neural network and deep neural network and propose the idea of activity and multi-input; after recognizing 19 kinds of activities, the effect of the algorithm model in this paper is fully proved. Comparative experiments include ensemble learning, deep learning single model, single input model, and multiple input model. Table 7 shows the experimental comparison results of different models.

It can be seen from Table 7 that the scores of the multi-input models are all above 0.8. The MBCDNN model we proposed has a maximum score of 0.887, explaining that input of multiple conversion methods of data, data enhancement, and model fusion can obtain better scores, which further shows that our model has a good activity recognition effect.

5. Conclusions

Aiming at the lack of complex action and different scene existing in human activity recognition, we use a smart phone as the carrier equipment and propose complex human activity recognition based on scene+action, by introducing the forward sequence and reverse sequence, as well as aggregation features to help the model with more activity features. However, as the time series data, a truncated or filling strategy will introduce unnecessary noise; for this reason, we propose a ComplexConv1D module to compensate for the impact of unnecessary noise. At the same time, in order to more comprehensively evaluate the performance of the model under a specific activity, we define an evaluation index that combines the weight of the scene and the action. Through experimental comparison and analysis, the performance of our model has indeed been improved, which proves the effectiveness of our method. After all this job we have done, it still needs a lot of things to do on human activity recognition. We believe there will be a more outstanding work in the future.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Research on big data analysis method of beneficiation and metallurgy industry project (project number BGRIMM-KZSKL-2019-03) and the Fujian province industrial field regional development project (project number 2019H4021).

References

- [1] B. Fu, N. Damer, F. Kirchbuchner, and A. Kuijper, "Sensing technology for human activity recognition: a comprehensive survey," *IEEE Access*, vol. 8, pp. 83791–83820, 2020.
- [2] F. Demrozi, G. Pravadelli, A. Bihorac, and P. Rashidi, "Human activity recognition using inertial, physiological and environmental sensors: a comprehensive survey," *IEEE Access*, vol. 8, pp. 210816–210836, 2020.
- [3] C. Huang, Y. Zong, J. Chen, W. Liu, J. Lloret, and M. Mukherjee, "A deep segmentation network of stent struts based on IoT for interventional cardiovascular diagnosis," *IEEE Wireless Communications*, vol. 28, no. 3, pp. 36–43, 2021.
- [4] J. Rafferty, C. D. Nugent, J. Liu, and L. Chen, "From activity recognition to intention recognition for assisted living within smart homes," *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 3, pp. 368–379, 2017.
- [5] C. Kunhui, *Design and Research of Monitoring System for Elderly Abnormal Activity in Smart House*, Huazhong University of Science and Technology, 2019.
- [6] X. Hu, J. Dai, Y. Huang et al., "A weakly supervised framework for abnormal behavior detection and localization in crowded scenes," *Neurocomputing*, vol. 383, pp. 270–281, 2020.
- [7] F. J. Ordóñez and D. Roggen, "Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition," *Sensors*, vol. 16, no. 1, p. 115, 2016.
- [8] J. P. Wolff, F. Grützmacher, A. Wellnitz, and C. Haubelt, "Activity recognition using head worn inertial sensors," in *Proceedings of the 5th international Workshop on Sensor-based Activity Recognition and Interaction*, pp. 1–7, 2018.
- [9] A. Lentzas and D. Vrakas, "Non-intrusive human activity recognition and abnormal activity detection on elderly people: a review," *Artificial Intelligence Review*, vol. 53, pp. 1–47, 2020.
- [10] N. S. Suriani and F. A. N. Rashid, "Smartphone sensor accelerometer data for human activity recognition using spiking neural network," *International Journal of Machine Learning and Computing*, vol. 11, no. 4, pp. 298–303, 2021.
- [11] A. E. Minarno, W. A. Kusuma, H. Wibowo, D. R. Akbi, and N. Jawas, "Single triaxial accelerometer-gyroscope classification for human activity recognition," in *2020 8th international conference on information and communication technology (ICoICT)*, pp. 1–5, Yogyakarta, Indonesia, 2020.
- [12] D. Garcia-Gonzalez, D. Rivero, E. Fernandez-Blanco, and M. R. Luaces, "A public domain dataset for real-life human activity recognition using Smartphone sensors," *Sensors*, vol. 20, no. 8, p. 2200, 2020.
- [13] C. Xu, *Human Activity Recognition Using Smart Phones*, 2020.
- [14] W. A. Kusuma, A. E. Minarno, and M. S. Wibowo, "Triaxial accelerometer-based human activity recognition using 1D convolution neural network," in *2020 international workshop on big data and information security (IW BIS)*, pp. 53–58, Depok, Indonesia, 2020.
- [15] A. Nandy, J. Saha, and C. Chowdhury, "Novel features for intensive human activity recognition based on wearable and smart phones sensors," *Microsystem Technologies*, vol. 26, pp. 1–15, 2020.
- [16] H. du, Z. Yu, F. Yi, Z. Wang, Q. Han, and B. Guo, "Recognition of group mobility level and group structure with mobile devices," *IEEE Transactions on Mobile Computing*, vol. 17, no. 4, pp. 884–897, 2018.
- [17] P. Garabelli, S. Stavarakis, and S. Po, "Smartphone-based arrhythmia monitoring," *Current Opinion in Cardiology*, vol. 32, no. 1, pp. 53–57, 2017.
- [18] A. Mottelsonand and K. Hornbæk, "An affect detection technique using mobile commodity sensors in the wild," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 781–792, 2016.
- [19] B. Cao, L. Zheng, C. Zhang et al., "Deepmood: modeling mobile phone typing dynamics for mood detection," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 747–755, 2017.
- [20] X. Zhang, W. Li, X. Chen, and S. Lu, "MoodExplorer," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 4, pp. 1–30, 2018.

- [21] N. D. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. Campbell, "A survey of mobile phone sensing," *IEEE Communications Magazine*, vol. 48, no. 9, pp. 140–150, 2010.
- [22] D. Roggen, A. Calatroni, M. Rossi et al., "Collecting complex activity datasets in highly rich networked sensor environments," in *2010 seventh international conference on networked sensing systems (INSS)*, pp. 233–240, Kassel, Germany, 2010.
- [23] E. M. Tapia, S. S. Intille, L. Lopez, and K. Larson, "The design of a portable kit of wireless sensors for naturalistic data collection," in *International Conference on Pervasive Computing*, pp. 117–134, Berlin, Heidelberg, 2006.
- [24] T. Hasegawa, "Smart phones sensor-based human activity recognition robust to different sampling rates," *IEEE Sensors Journal*, vol. 21, no. 5, pp. 6930–6941, 2020.
- [25] J. Shi, D. Zuo, and Z. Zhang, "An energy-efficient human activity recognition system based on smart phones," in *2020 7th International Conference on Soft Computing & Machine Intelligence (ISCFMI)*, pp. 177–181, Stockholm, Sweden, 2020.
- [26] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "A public domain dataset for human activity recognition using smart phones," *Esann*, vol. 3, p. 3, 2013.
- [27] G. Vavoulas, C. Chatzaki, T. Malliotakis, M. Padiaditis, and M. Tsiknakis, "The MobiAct dataset: recognition of activities of daily living using smart phones," in *International Conference on Information and Communication Technologies for Ageing Well and e-Health*, pp. 143–151, 2016.
- [28] N. Hnoohom, S. Mekruksavanich, and A. Jitpattanakul, "Human activity recognition using triaxial acceleration data from smart phones and ensemble learning," in *2017 13th international conference on signal-image Technology & Internet-Based Systems (SITIS)*, pp. 408–412, Jaipur, India, 2017.
- [29] Z. Yu, *Research on Human Activity Recognition Algorithm Based on Time Series*, Beijing University of Posts and Telecommunications, 2019.
- [30] H. Jian and Y. Jiaxian, "Human activity recognition technology based on smart phone multi-sensor fusion," *Journal of Beijing University of Technology*, vol. 46, no. 11, pp. 14–21, 2020.
- [31] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [32] J. Platt, "Sequential minimal optimization: a fast algorithm for training support vector machines," 1998.
- [33] R. D. Short and K. Fukunaga, "The optimal distance measure for nearest neighbor classification," *IEEE Transactions on Information Theory*, vol. 27, pp. 622–627, 1981.
- [34] P. Domingos and M. Pazzani, "On the optimality of the simple Bayesian classifier under zero-one loss," *Machine Learning*, vol. 29, no. 2, pp. 103–130, 1997.
- [35] J. Xiangyu, "Using CNN to realize human activity recognition based on smart phone sensors," Lanzhou University, 2019.
- [36] P. Yun, X. Peiwen, Z. Huaiyu, and L. Junjie, *Human Activity Recognition Method Based on Inertial Sensors*, Zhejiang University, 2020.
- [37] P. K. Shukla, A. Vijayvargiya, and R. Kumar, "Human activity recognition using accelerometer and gyroscope data from smartphones," in *2020 International Conference on Emerging Trends in Communication, Control and Computing (ICONC3)*, pp. 1–6, Lakshmanagarh, India, 2020.
- [38] V. Dumoulin and F. Visin, "A guide to convolution arithmetic for deep learning," 2016, <https://arxiv.org/abs/1603.07285>.