International Journal for Housing Science and Its Applications
Publish August 4, 2025. Volume 46, Issue 3 Pages 3171-3180

https://doi.org/10.70517/ijhsa463266

Research on Human Action Recognition and Behavior Analysis Based on Long and Short-Term Memory Networks

Heng Zhang[1,*] and Fa Wang[1]
[1] College of Electronic Information and Engineering, Huaibei Institute of Technology, Huaibei, Anhui, 235000, China
Corresponding authors: (e-mail: zh2568610041@163.com).

Abstract Effective recognition of human actions is a must for the further development of artificial intelligence. In this paper, wearable sensors are utilized to collect human action data based on time series. Combined with median filtering technique to process the action data and keep the edge information, instantaneous and local features are highlighted by wavelet transform. Long and short-term memory network (LSTM) is added into the convolutional neural network to solve the limitation of insufficient dependency of the convolutional neural network in learning human actions and improve the model classification accuracy. A large human action database is selected for model training and testing to compare the performance advantages of this paper's model. The results show that the actual number of misclassifications of this paper's model in the 2 datasets is only 16 and 5. During the training process, the loss value is always no more than 0.1, and it shows a stable decreasing trend. The average recognition accuracy is steadily improved from 0.03891 to 0.76787. In the multi-method comparison, the accuracy of CS and CV metrics of this paper's model is 87.39% and 90.87%, and that of Top-1 and Top-5 metrics is 41.76% and 60.63%, which is higher than that of the comparison methods. The model in this paper can effectively realize the smooth recognition of human body movements with high accuracy.

Index Terms wearable sensor, median filter, wavelet transform, LSTM, convolutional neural network, human action recognition

I. Introduction

People's daily activities constitute an important part of social production and life, and human behavior recognition and behavior analysis play an important role in daily life, and are widely used in the fields of medical rehabilitation, intelligent care, motion monitoring, and human-computer interaction [1]-[4]. According to different data sources, human behavior recognition is divided into video image-based human behavior recognition and wearable sensor-based human behavior recognition [5], [6]. With the rapid development of microelectromechanical systems and wireless communication technologies, these sensors can be integrated into wearable devices, smartphones or smartwatches, which greatly facilitates people's daily carrying [7]-[9]. And the actual use is not restricted by the place and surrounding environment, and does not bring the threat of violating personal privacy to the user, which promotes the application of human behavior recognition based on wearable sensors in people's daily life [10]-[12].

Traditional pattern recognition analysis mainly uses machine learning algorithms such as artificial neural networks, support vector machines, decision trees, plain Bayes, K-nearest neighbors, and hidden Markov models [13], [14]. In the past decade or so, these machine learning algorithms have made great progress in the problem of human behavior recognition, but there are some inescapable drawbacks [15]-[17]. For example, when using traditional machine learning algorithms for human behavior recognition, human behavior data features need to be manually extracted in advance, and manual feature extraction is limited by domain-specific knowledge and people's existing knowledge and experience [18]-[20]. Some shallow features can only be used to recognize low-level activities of human behavior (e.g., standing, walking, running, etc.), and it is difficult to recognize more complex and high-level human behaviors, especially in the current situation of multimodal and high-dimensional sensor data emergence, and these features are unable to effectively deal with the complex activities and realize the accurate analysis of human activities [21]-[24]. To further improve the classification performance of human behavior recognition without relying on manual feature extraction, this paper proposes a human action recognition and behavior analysis method based on long and short-term memory networks [25], [26].

Aiming at the problem of complex human action recognition scene and high recognition difficulty, this paper first analyzes the process of collecting human action data based on wearable sensors. The prediction error is reduced by multi-classification cross entropy loss function calculation and so on. Combining median filter and wavelet transform methods, the collected human action data are preprocessed to maintain the edge information while improving the clarity of local features. The LSTM network is introduced into the convolutional neural network for time series modeling and to solve the problem of the lack of long-term dependence of the original learning of the convolutional neural network, so as to enhance the model's recognition and classification accuracy of the human body movements. The KTH database and UCF101 database are chosen as the sources of training and testing sets, and the classification effect and stability of the model are analyzed through comparative experiments.

II. Analysis of human action recognition technology based on long and short-term memory network

Based on the human action recognition task, this chapter analyzes the specific application process of human action recognition technology from the aspects of action data preprocessing and hybrid neural model construction.

II. A.Human Movement Recognition Task

Sensor-based human action recognition can be considered as a typical pattern recognition problem, where samples are classified based on their nature through computational methods. The behavior itself is characterized by uncertainty and ambiguity, which makes this recognition method very difficult. To describe it in mathematical language, it is assumed that the user performs some actions belonging to a set S, and the probability of occurrence of these events is represented by a matrix P of time series collected by multiple sensors, the

$$S=\{S_1,S_2,\cdots,S_m\} \tag{1}$$

$$P=\left(P_1,P_2,\cdots,P_t,\cdots,P_n\right) \tag{2}$$

where m denotes the number of categories of human actions, P_t denotes the column vector composed of data collected by wearable sensors at the moment t, and n represents the length of the sequence. Then the time series P is inputted into the model $\mathrm{F}\left(P\right)$, and the sequence of predicted behavioral categories of human actions $\hat{A}$ can be obtained after the computation, however, the exact sequence of behavioral categories is A^*, the

$$\hat{A}=\left\{\hat{A}_j\right\}_{j=1}^{n}=\mathrm{F}\left(S\right),\hat{A}_j\in A \tag{3}$$

$$A^*=\left\{A_j^*\right\}_{j=1}^{n}\ A_j^*\in A \tag{4}$$

The function of the system based on wearable sensor human action recognition is to reduce the gap between the predicted category $\hat{A}$ and the accurate category A^* by training the model F, which can be calculated by a loss function $L\left(F\left(S\right),A^*\right)$ to calculate the two The gap, in this paper, we use the multiclassification cross entropy loss function, the calculation formula is shown in equation (5):

$$loss=-\frac{1}{m}\sum_{i=1}^{m}\tilde{y}_i\log y_i \tag{5}$$

where $\tilde{y}_i$ denotes the true value of the i th action and y_i denotes the predicted value of the i th action of the model.

The smaller the loss $loss$ value indicates that the predicted category is closer to the true value, so the problem is converted into how to make the loss function as small as possible through the training of the model.

Generally, the raw data collected by the sensors will not be directly passed into the training model, but first go through several processes such as data preprocessing, feature extraction and feature selection, and finally the result of the model training is to minimize the loss function $L\left(F\left(S\right),A^*\right)$.

II. B.Data noise reduction preprocessing

In the process of data acquisition by sensors, the collected data are often subject to noise and other interferences due to factors such as the environment and equipment. These interferences will affect the quality and accuracy of the data, so it is necessary to calibrate the collected data to make up for the defects in the process of sensor acquisition. Effective calculations can compensate for the errors generated during the sensor data acquisition process, so as to obtain data closer to the true value. For accelerometers, gyroscopes and magnetometers and other sensor equipment, the need for different calibration operations to solve their respective problems. With the continuous development of sensor technology, many sensors now have the function of automatic calibration. Noise reduction is an important step when performing data processing. Common noise reduction methods include median filter, wavelet transform, mean filter, Gaussian filter and Butterworth filter to name a few.

II. B. 1) Median Filter

The median filter is a nonlinear filtering technique that does not require weighted averaging of pixel values, but rather filters by taking the median value. This characteristic makes median filtering more suitable for removing non-Gaussian noise such as impulse noise, while preserving image details and not affecting image brightness and contrast.

Another good feature of median filtering technique is that it preserves edge information. Since median filtering only considers pixel values within a local region and does not take into account the distance between pixels and the difference in gray values, it can effectively remove noise and preserve edge information. Median filtering is suitable for the removal of small noise, and can remove the noise that is difficult to handle by linear filtering such as mean filtering and Gaussian filtering. In the field of digital image processing, median filtering is a basic filtering technique with a wide range of applications.

For a median filter template of size $n*n$, assuming that the pixels to be filtered are $p(i,j)$, and the pixel points in the template are $p(k,l)(0<=k<n,0<=l<n)$, the output of median filtering $g(i,j)$ can be expressed as:

$$g(i,j)=median\left(p(k,l)\right)(0<=k<n,0<=l<n) \tag{6}$$

where $median$ means to find the middle value after sorting all the pixel values within the template. If there are an even number of pixels in the template of $n*n$, the average of the middle two pixels is taken as the output.

II. B. 2) Wavelet transform

Wavelet transform is a signal analysis method based on orthogonal functions. Different from other traditional transform methods such as Fourier transform and discrete cosine transform, wavelet transform is a time-domain transform method with localization, which can better reflect the local characteristics of the signal and have better effect on non-smooth signal and non-linear signal processing.

Wavelet transform decomposes the signal into two directions, scale and frequency, which makes wavelet transform able to express both instantaneous and local features of the signal effectively. In wavelet transform, decomposition at different resolutions is achieved by choosing different wavelet basis functions, thus realizing multi-resolution analysis. Commonly used wavelet basis functions include Haar wavelet, Daubechies wavelet, Symlet wavelet and so on.

Wavelet transform is widely used in signal processing, image processing, pattern recognition, data compression and other fields. Among them, in image processing, wavelet transform can use two-dimensional wavelet transform for image noise reduction, edge detection, texture analysis and other operations, while in signal processing, wavelet transform can use one-dimensional wavelet transform for filtering, smoothing and other operations.

Wavelet transform is a relatively complex transformation method, which requires reasonable analysis and design of signal characteristics and selection of wavelet basis functions. At present, with the continuous development of computer technology, wavelet transform has become a very important signal processing method.

The wavelet transform formula is shown in equation (7):

$$WT(a,\tau)=\frac{1}{\sqrt{a}}\int_{-\infty}^{\infty}f(t)*\psi\left(\frac{t-\tau}{a}\right)dt \tag{7}$$

where $WT(a,\tau)$ is the result of the wavelet transform, a represents the scale factor, and τ represents the equilibrium quantity.

II. C.Hybrid Neural Networks

According to the recognition rate of convolutional neural network model processing sensor data for human body recognition, when using convolutional neural network alone for human body action recognition, the recognition rate for simple actions is still high, but for some complex actions, the recognition rate of the convolutional neural network is relatively low, the possible reason is that convolutional neural network for the lack of spatial and temporal dependence on the data, which leads to the recognition rate of the complex actions is low. In order to solve this problem, this paper proposes a long short-term memory (LSTM) network to make up for this defect, through the fusion of the convolutional neural network and the LSTM network to carry out the recognition of human behavioral actions, and to improve its recognition rate.

II. C. 1) LSTM networks

LSTM network is a recursive neural network architecture designed to deal with gradient decay or gradient bursting problems, solving the problem of modeling temporal sequences and learning long-term dependencies. It is often used in fields such as machine learning and speech recognition, but with the development of technology, LSTM networks are widely and effectively used in the field of human recognition.

LSTM network is composed of multiple LSTM units, Figure 1 shows the LSTM network unit structure. The cell module is composed of three gates, which are input gate, output gate and forget gate, and the three gates represent the read, write and reset operations of the network cell respectively.

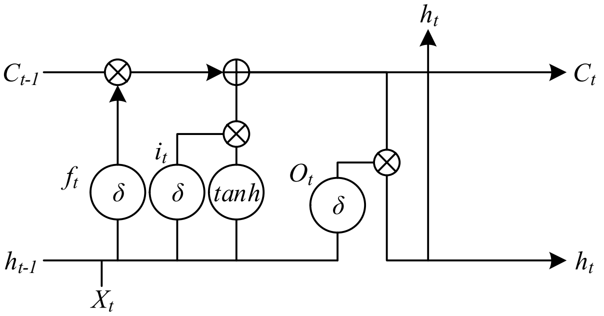


Figure 1: LSTM network unit structure

The working principle equations in this network unit are as follows:

$$\begin{gathered} f_t = \delta\left(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f\right) \\ i_t = \delta\left(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i\right) \end{gathered} \tag{8}$$

$$c_t = f_t c_{t-1} + i_t \delta \tanh\left(W_{xc}x_t + W_{hc}h_{t-1} + b_c\right) \tag{9}$$

$$o_t = \delta\left(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o\right) \tag{10}$$

$$h_t = o_t \delta \tanh\left(c_t\right) \tag{11}$$

where i_t is the input gate, f_t is the forgetting gate, o_t is the output gate, c_t is the activation vector, h_t and h_{t-1} are the outputs at the current and previous moments, respectively, δ is a nonlinear function, and W_{xf}, W_{hf}, W_{cf}, W_{xi}, W_{hi}, W_{ci}, W_{xc}, W_{hc}, W_{xo}, W_{ho}, and W_{co} are the weight matrices of vectors passing through the control gate, and b_f, b_i, b_c, and b_o are the bias vectors. Figure 2 shows the LSTM complex structure, which is added to the convolutional neural network to solve the problem of the lack of dependency of the convolutional neural network.

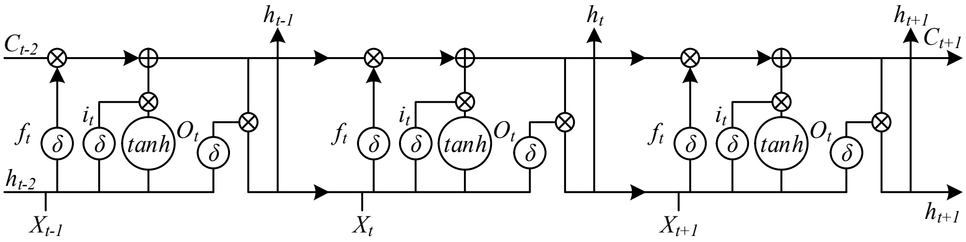


Figure 2: LSTM network structure

II. C. 2) LSTM Convolutional Neural Networks

As we know from the above, convolutional neural network has the defect of insufficient dependency, which leads to the low recognition rate of complex actions, in order to solve the problem, the previous subsection proposes LSTM network to solve the sub-problem, and the main task of this subsection is to fuse these two neural networks to establish a hybrid neural network model for the recognition of behavioral actions. Figure 3 shows the specific structure of the model.

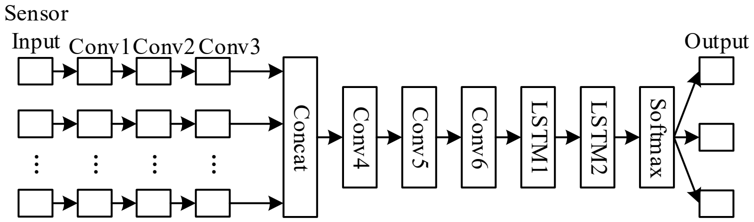


Figure 3: LSTM Hybrid Convolutional Neural Network

The number of layers of the convolutional network is 6, the first three convolutional layers contain multiple sub-convolution per layer, the number of sub-convolution is set according to the number of sensors, according to the number of sensors in the self-test dataset is 9, so there are 9 sub-convolution in each of the first three convolutional layers, the purpose of sub-convolution is to carry out the convolution operation of each sensor data respectively, the size of convolution kernel in the first three convolutional layers is set as $(1,6)$, the number of convolution kernels are 30, 62, 62, respectively, through the convolution operation of these three convolutional layers, through the activation function to activate the input Concat layer will be fused with the nine sensor data, to obtain a large feature map, this large feature map and then through the latter three layers of convolutional operations to learn the relationship between the sensing, after the latter three layers of convolutional layers convolution kernel size is set to $(1,6,3)$, the number of convolution kernels are set to 126, because the input of the last three convolutional layers is a large feature map, so the 3D convolution kernel is used for convolution, and activation pooling is carried out after each convolutional operation and then driven into the next layer, and after the end of the convolutional operation, the inputs are two 126-node LSTM networks to learn the data dependency, and then the tanh function is used to carry out the nonlinear processing, in order to avoid overfitting the data, dropout layer is also added behind the LSTM network, and finally the activation classification output is performed by softmax function.

III. Practice of human action recognition based on LSTM+convolutional neural network model

In this chapter, 2 human action datasets are selected to verify the human action recognition effect of this paper's model through comparative experiments and so on.

III. A. Human Movement Recognition Database

In order to verify the effectiveness of the recognition model established in this paper, we choose to conduct experiments on the public video dataset KTH, UCF101 to verify and test the recognition effect. In the later experiments, this paper converts the video into a frame of static images and selects a number of frames as samples from each set of images at intervals respectively, and then divides the samples into training samples and test samples. All the experiments follow the cross-validation principle and the results are expressed in terms of the correct recognition rates (ARRs).

III. A. 1) KTH database

The KTH database has 6 different human movements: walking, jogging, running, boxing, waving, and clapping. Each type of action is completed by 20 different people in 4 different scenes, including the motion target outdoors (S1), the motion target outdoors and there is a scale change (S2), the motion target is outdoors and there is a change in dress (S3), and the motion target is indoors but there is a light change (S4), so the database has a total of 6 * 20 * 4 = 480 videos, the image resolution is 180 * 140, and the video has 30 frames per second. In this paper, the training set and the test set are randomly assigned according to a certain proportion, which is done by first randomly selecting 12 individuals from each scenario as the training set, and the remaining 8 individuals are the test set.

III. A. 2) UCF101 database

The UCF101 database is composed of web videos filmed in real environments, mainly from various types of sports videos collected by BBC/ESPN radio and TV channels, and clips or movies edited from the Internet, especially YouTube, which contain 50 different categories of human movements, with a total of 7,500 videos, with an image resolution of 350*250.The database All actions can be categorized into 5 main categories: 1) human interaction, such as hair cutting, head massage, etc.; 2) musical instrument playing, such as playing the guitar, playing the violin, etc.; 3) human-object interaction, such as chopping vegetables, yo-yoing, etc.; 4) body actions without other object interaction, such as playing Tai Chi, etc.; 5) sports, such as playing volleyball, basketball, swimming, etc.. The difficulty of this database is that there are too many kinds of human actions, the background is more complex, the target is not single, and there are also some occlusions and camera movements, etc., so it is very challenging and unpredictable difficulties to conduct experiments on the UCF101 database.

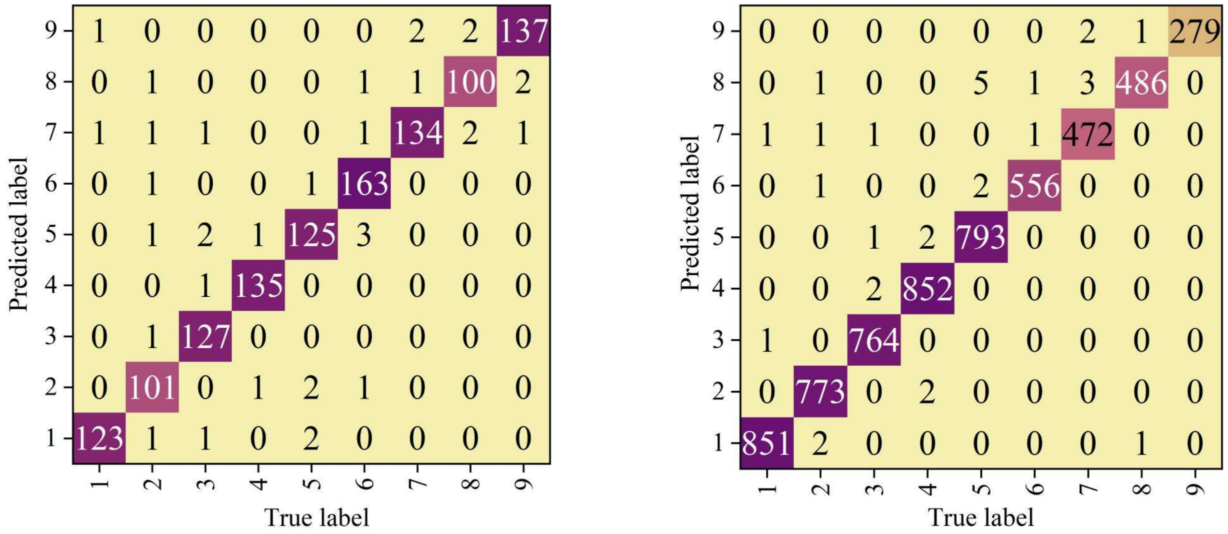
III. B. Comparison of human movement recognition performance

III. B. 1) Comparison of classification effects

In the evaluation of human action recognition classification effect, Random Forest Classifier (RFC) recognition classification method is chosen as a comparison to analyze this paper's model in human action recognition classification. The RFC, and this paper's model are trained and tested in the KTH database and UCF101 database, respectively, and the classification effect is demonstrated by the confusion matrix.

The confusion matrix is a tool often used in evaluating the performance of classification models, and is especially important in multi-category classification problems. The matrix presents the model's classification results for different categories of samples in a tabular form, with each row representing the actual category and each column representing the model's predicted category. The key terms True Positive Example (TP), True Negative Example (TN), False Positive Example (FP), and False Negative Example (FN) describe the positive and negative categories that are correctly categorized by the model and the positive and negative categories that are incorrectly categorized, respectively. Elements on the diagonal indicate cases where the model is correctly categorized, while off-diagonal elements indicate cases where the model is incorrectly categorized. The confusion matrix plot, as an intuitive visualization means, will further demonstrate the classification effect of the action classification model on different categories, providing visual support for in-depth understanding of the model performance.

Figure 4 shows the confusion matrix of this paper's model on the KTH database and the UCF101 database. Figure 5 shows the confusion matrix of RFC on KTH database and UCF101 database. Analyzing the non-diagonal misclassification in Fig. 4 and Fig. 5, it can be found that the actual number of misclassification of this paper's model is 16 in the classification of human action recognition on the KTH database, while the actual number of misclassification of the comparison method RFC is 113. In the human action recognition on the UCF101 database, the actual number of misclassification of this paper's model is 5, while the actual number of misclassification of the comparison method RFC is 646 From the comparison results of the actual misclassification number, it can be judged that the model of this paper has a better classification effect in human action recognition classification.



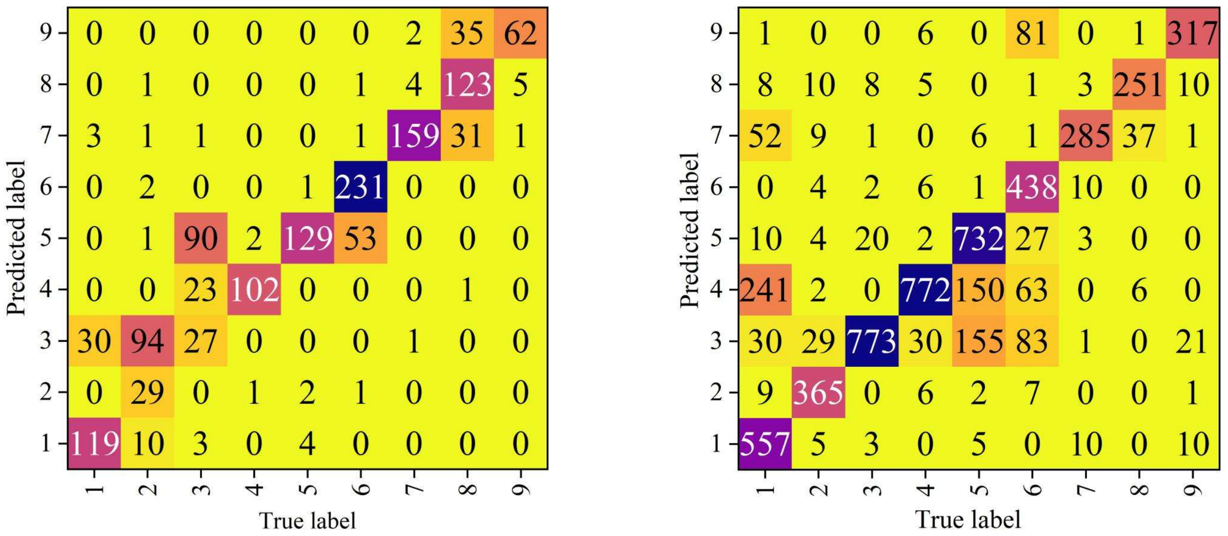
	1	2	3	4	5	6	7	8	9
9	1	0	0	0	0	0	2	2	137
8	0	1	0	0	0	1	1	100	2
7	1	1	1	0	0	1	134	2	1
6	0	1	0	0	1	163	0	0	0
5	0	1	2	1	125	3	0	0	0
4	0	0	1	135	0	0	0	0	0
3	0	1	127	0	0	0	0	0	0
2	0	101	0	1	2	1	0	0	0
1	123	1	1	0	2	0	0	0	0

(a) Confusion matrix of the model in this paper of KTH

	1	2	3	4	5	6	7	8	9
9	0	0	0	0	0	0	2	1	279
8	0	1	0	0	5	1	3	486	0
7	1	1	1	0	0	1	472	0	0
6	0	1	0	0	2	556	0	0	0
5	0	0	1	2	793	0	0	0	0
4	0	0	2	852	0	0	0	0	0
3	1	0	764	0	0	0	0	0	0
2	0	773	0	2	0	0	0	0	0
1	851	2	0	0	0	0	0	1	0

(b) Confusion matrix of the model in this paper of UCF101

Figure 4: Confusion matrix of the model in this paper



(a) Confusion matrix of RFC of KTH (b) Confusion matrix of RFC of UCF101

Figure 5: Confusion matrix of RFC

III. B. 2) Stability comparison

The changes of the model in this paper during the training process are further analyzed. The experimental results after model training provide several key indicators, which provide rich information for comprehensive analysis of target human action recognition detection performance.

Figure 6 shows the trend of the loss and mAP@0.45 (representing the average accuracy of the model recognition and detection calculated under the condition of 0.45 IoU threshold) during the training process. From this figure, it can be observed that during the training process of this paper's model, both the prediction frame and the category loss show a gradual decrease with the increase in the number of training rounds, and eventually stabilize. The value of predicted frame loss was reduced from 0.07733 to 0.00389 and category loss was reduced from 0.0288 to 0.00105. This indicates that the model gradually learns the features of the data during the training process and reaches a lower loss value in the final stage. In addition, the change in the loss value is relatively smooth, and there are no sharp fluctuations or oscillations. This indicates that the optimization during model training is relatively stable and continuous. According to the trend of mAP@0.45, with the increase of training times, the mAP@0.45 value increased steadily from 0.03891 to 0.76787, and there was no unstable fluctuation or obvious irregular change.

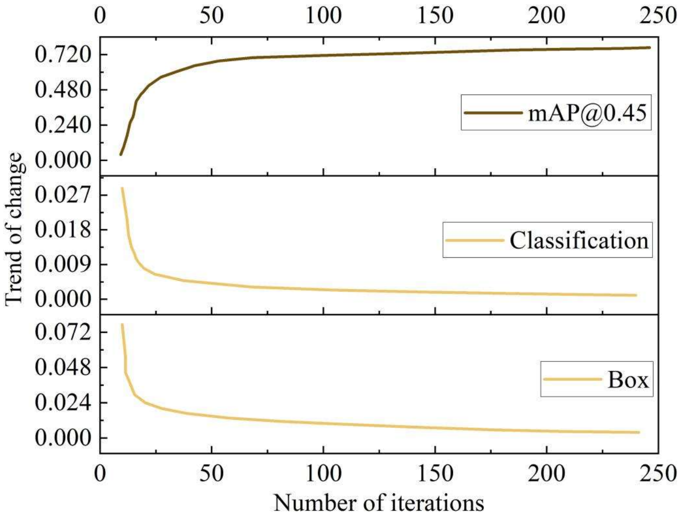


Figure 6: Loss values of the model in this paper and the variation of mAP@0.5

Figure 7 shows the PR curve of the model in this paper during the training process of the KTH database. The abscissa in the graph represents Recall, and the ordinate represents Precision. Since the task in this chapter is a multi-class task with 6 classes, the PR curve presents 7 curves (including a mAP@0.5 curve for all categories), each representing a different category. The trend of the curve shows the trade-off between precision and recall of the model. Typically, the PR curve should be smooth and sloping upwards to the right as much as possible, representing a high recall rate while maintaining high accuracy. The area at the bottom left of the PR curve represents the classification accuracy of the model on each category, and the closer its value is to 1, the better the performance of the model on that category. From the area size of the lower left of the PR curve of the six types of actions in Figure 7, the area of the six types of actions is greater than 0.85, which indicates that the model in this paper has high accuracy and high recall rate in the process of human action recognition and detection, and can stably and effectively complete the recognition and detection of human actions.

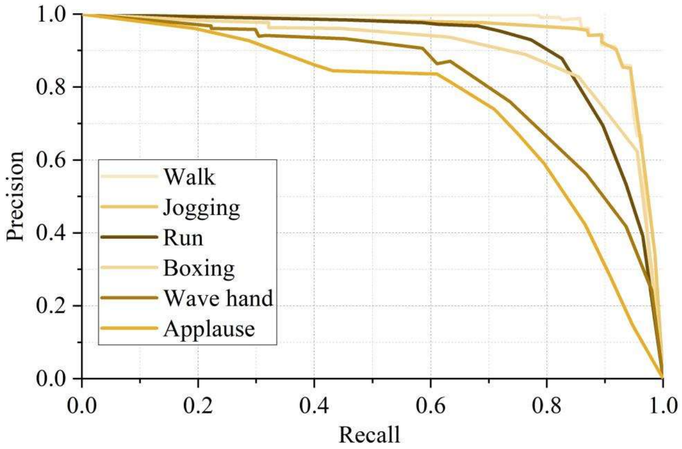


Figure 7: The PR curve in the model training process of this paper

III. B. 3) Accuracy comparison

In order to verify that the model in this paper has the advantage of action recognition classification accuracy in KTH database and UCF101 database, more same-type recognition methods are introduced as comparisons, and then the application value of this paper's model is examined under the premise of more data support. The same type of recognition methods include: traditional methods such as LieGroup, FeatureEnc, etc., CNN methods such as Clips+CNN+MTLN, etc., RNN methods such as ST-LSTM, HBRNN, DeepLSTM, etc., and GCN methods such as ST-GCN and TCN.

Table 1 shows the recognition accuracy of multiple methods on KTH database. Table 2 shows the recognition accuracy of multiple methods on UCF101 database. Comparing the recognition accuracies in Table 1 and Table 2, the recognition accuracies associated with the two metrics of CS and CV of this paper's model on the KTH database are 87.39% and 90.87%, respectively, which are higher than those of the comparison methods. In the UCF101 database, the recognition accuracy related to the two indicators of Top-1 and Top-5 of this paper's model is 41.76% and 60.63%, respectively, which is also higher than the comparison method. From the comparison results, it can be seen that this paper's LSTM+Convolutional Neural Network model achieves higher human action recognition accuracy on the 2 public datasets compared to other methods. This once again proves that the model in this paper has a greater potential for application in complex and diverse human action recognition.

Table 1: Recognition accuracy rate on the KTH database

Method	CS(%)	CV(%)
Lie Group	50.12	82.83
HBRNN	59.18	64.01
ST-lSTM	69.23	77.74
TCN	74.35	83.12
Clips+CNN+MTLN	79.61	84.85
ST-GCN	81.57	88.36
Article model	87.39	90.87

Table 2: Recognition accuracy rates on the UCF101 database

Method	Top-1(%)	Top-5
Feature Enc	14.91	25.82
Deep LSTM	16.44	35.31
TCN	20.36	40.07
ST-GCN	30.78	52.83
Article model	41.76	60.63

IV. Conclusion

In this paper, a human action recognition model based on the combination of long short-term memory network and convolutional neural network is constructed, and the recognition advantages of the model are verified through experiments. In the two databases, the actual number of misclassifications of the proposed model is 16 and 5, which is much lower than that of the comparison method of 113 and 646. With the increase of the number of trainings, the prediction box and category losses of the model in this paper steadily decreased to 0.00389 and 0.00105, and the mAP@0.45 value steadily increased to 0.76787. The loss in the learning process is small and stable, and the prediction accuracy is high. In the multi-method comparison of the model in this paper, the recognition accuracy of CS and CV indicators is 87.39% and 90.87%, respectively, and the recognition rates of Top-1 and Top-5 indicators are 41.76% and 60.63%, respectively. Among the same type of methods, the recognition accuracy is the highest. The model in this paper can be used to obtain high-precision human action recognition effect. In the future, we can further explore how to reduce the human action recognition time of the model, improve the recognition efficiency, and provide the possibility for the real-time application of the model.

References

[1] Wang, Z., Jiang, K., Hou, Y., Dou, W., Zhang, C., Huang, Z., & Guo, Y. (2019). A survey on human behavior recognition using channel state information. Ieee Access, 7, 155986-156024.

[2] Hu, K., Jin, J., Zheng, F., Weng, L., & Ding, Y. (2023). Overview of behavior recognition based on deep learning. Artificial intelligence review, 56(3), 1833-1865.

[3] Yuan, M., Wei, S., Zhao, J., & Sun, M. (2022). A systematic survey on human behavior recognition methods. SN Computer Science, 3(1), 6.

[4] Degardin, B., & Proenca, H. (2021). Human behavior analysis: A survey on action recognition. Applied Sciences, 11(18), 8324.

[5] Wang, Z., Jiang, K., Hou, Y., Huang, Z., Dou, W., Zhang, C., & Guo, Y. (2019). A survey on CSI-based human behavior recognition in through-the-wall scenario. IEEE Access, 7, 78772-78793.

[6] Jalal, A., Quaid, M. A. K., & Hasan, A. S. (2018, December). Wearable sensor-based human behavior understanding and recognition in daily life for smart environments. In 2018 International Conference on Frontiers of Information Technology (FIT) (pp. 105-110). IEEE.

[7] Dai, C., Liu, X., Lai, J., Li, P., & Chao, H. C. (2019). Human behavior deep recognition architecture for smart city applications in the 5G environment. IEEE Network, 33(5), 206-211.

[8] Qu, J., Qiao, N., Shi, H., Su, C., & Razi, A. (2020). Convolutional neural network for human behavior recognition based on smart bracelet. Journal of Intelligent & Fuzzy Systems, 38(5), 5615-5626.

[9] Wang, Z., Guo, B., Yu, Z., & Zhou, X. (2018). Wi-Fi CSI-based behavior recognition: From signals and actions to activities. IEEE Communications Magazine, 56(5), 109-115.

[10] Yousefi, S., Narui, H., Dayal, S., Ermon, S., & Valaee, S. (2017). A survey on behavior recognition using WiFi channel state information. IEEE Communications Magazine, 55(10), 98-104.

[11] Wang, L., Huynh, D. Q., & Koniusz, P. (2019). A comparative review of recent kinect-based action recognition algorithms. IEEE Transactions on Image Processing, 29, 15-28.

[12] Kong, Y., & Fu, Y. (2022). Human action recognition and prediction: A survey. International Journal of Computer Vision, 130(5), 1366-1401.

[13] Abro, I. A., & Jalal, A. (2024, December). Intelligent Multimodal Human Behavior Recognition using Inertial and Video Sensors. In 2024 International Conference on Frontiers of Information Technology (FIT) (pp. 1-6). IEEE.

[14] Gao, G., Li, Z., Huan, Z., Chen, Y., Liang, J., Zhou, B., & Dong, C. (2021). Human behavior recognition model based on feature and classifier selection. Sensors, 21(23), 7791.

[15] Xu, H., Li, L., Fang, M., & Zhang, F. (2018). Movement Human Actions Recognition Based on Machine Learning. International Journal of Online Engineering, 14(4).

[16] Lu, C. (2021). Multifeature fusion human motion behavior recognition algorithm using deep reinforcement learning. Mobile Information Systems, 2021(1), 2199930.

[17] Lu, J., Yan, W. Q., & Nguyen, M. (2018, November). Human behaviour recognition using deep learning. In 2018 15th IEEE international conference on advanced video and signal based surveillance (AVSS) (pp. 1-6). IEEE.

[18] Gulhane, M., & Sajana, T. (2021). Human behavior prediction and analysis using machine learning-A review. Turkish Journal of Computer and Mathematics Education, 12(5), 870-876.

[19] Ijjina, E. P., & Chalavadi, K. M. (2016). Human action recognition using genetic algorithms and convolutional neural networks. Pattern recognition, 59, 199-212.

[20] Zhu, J., Goyal, S. B., Verma, C., Raboaca, M. S., & Mihaltan, T. C. (2022). Machine learning human behavior detection mechanism based on python architecture. Mathematics, 10(17), 3159.

[21] Zahid, F. B., Ong, Z. C., Khoo, S. Y., & Salleh, M. F. M. (2021). Inertial sensor based human behavior recognition in modal testing using machine learning approach. Measurement Science and Technology, 32(11), 115905.

[22] Subasi, A., Khateeb, K., Brahimi, T., & Sarirete, A. (2020). Human activity recognition using machine learning methods in a smart healthcare environment. In Innovation in health informatics (pp. 123-144). Academic Press.

[23] Jaouedi, N., Boujnah, N., & Bouhlel, M. S. (2020). A new hybrid deep learning model for human action recognition. Journal of King Saud University-Computer and Information Sciences, 32(4), 447-453.

[24] An, F. P. (2018). Human action recognition algorithm based on adaptive initialization of deep learning model parameters and support vector machine. IEEE Access, 6, 59405-59421.

[25] Sun, L., Jia, K., Chen, K., Yeung, D. Y., Shi, B. E., & Savarese, S. (2017). Lattice long short-term memory for human action recognition. In Proceedings of the IEEE international conference on computer vision (pp. 2147-2156).

[26] Sarabu, A., & Santra, A. K. (2021). Human action recognition in videos using convolution long short-term memory network with spatio-temporal networks. Emerging Science Journal, 5(1), 25-33.