Research Article

Extraction and Recognition Method of Basketball Players' Dynamic Human Actions Based on Deep Learning

Qiulin Wang,¹ Baole Tao,¹ Fulei Han,¹ and Wenting Wei^{ID}²

¹Physical Education Institute, Yangzhou University, Yangzhou 225000, Jiangsu, China

²Physical Education Institute, Kunming University, Kunming 650000, Yunnan, China

Correspondence should be addressed to Wenting Wei; kmxywwt0930@yzu.edu.cn

Received 20 May 2021; Revised 10 June 2021; Accepted 19 June 2021; Published 28 June 2021

Academic Editor: Sang-Bing Tsai

Copyright © 2021 Qiulin Wang 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.

The extraction and recognition of human actions has always been a research hotspot in the field of state recognition. It has a wide range of application prospects in many fields. In sports, it can reduce the occurrence of accidental injuries and improve the training level of basketball players. How to extract effective features from the dynamic body movements of basketball players is of great significance. In order to improve the fairness of the basketball game, realize the accurate recognition of the athletes' movements, and simultaneously improve the level of the athletes and regulate the movements of the athletes during training, this article uses deep learning to extract and recognize the movements of the basketball players. This paper implements human action recognition algorithm based on deep learning. This method automatically extracts image features through convolution kernels, which greatly improves the efficiency compared with traditional manual feature extraction methods. This method uses the deep convolutional neural network VGG model on the TensorFlow platform to extract and recognize human actions. On the Matlab platform, the KTH and Weizmann datasets are preprocessed to obtain the input image set. Then, the preprocessed dataset is used to train the model to obtain the optimal network model and corresponding data by testing the two datasets. Finally, the two datasets are analyzed in detail, and the specific cause of each action confusion is given. Simultaneously, the recognition accuracy and average recognition accuracy rates of each action category are calculated. The experimental results show that the human action recognition algorithm based on deep learning obtains a higher recognition accuracy rate.

1. Introduction

In the field of sports and athletics, the standard of the basketball player's action is the key to determine the athlete's performance. The traditional scoring method relies on the human eye to score, which may cause a large error or injustice. Using human action extraction and recognition technology to capture and analyze the actions of athletes can provide more accurate action information data as a scoring reference, which is very effective in improving the fairness and accuracy of scoring, can also train athletes, and provide accurate action data for the coach's reference to improve the athlete's action level.

Many scholars at home and abroad have conducted related researches on three aspects: feature collection, deep convolutional neural networks, and human action

recognition. Holden uses the acceleration sensor embedded in the smartphone to collect data and finds that the placement of the sensor has a great influence on the accuracy of the experiment and uses algorithms to extract and learn data features. However, in the classification algorithm, the classifier cannot distinguish certain similar motion states, which is also a problem that many algorithms currently have [1]. Scott uses deep convolutional neural networks for feature learning, mainly by exploring the intrinsic features of each activity and one-dimensional time series signals, while providing a method for automatically extracting robust data features from the original data. Experiments show that although the complexity level of each layer of features decreases, each layer of the convolutional neural network can still distinguish complex features [2]. Lecun uses a new general human body state recognition algorithm, which uses

Kalman filtering to filter the data, real-time judgment of human body movement, stillness, and state transition. Experimental results show that the algorithm has achieved good performance on mobile devices with limited computing and storage capabilities [3].

From the perspective of deep learning, this article extracts and recognizes the dynamic human movements of basketball players, and uses Bi-LSTM neural network and TensorFlow to identify and simulate human basketball training, calculates the accuracy of simulation experiment results, and finally collects basketball players' movement data. The innovation of this article is to apply the convolutional neural network in deep learning to the action analysis of basketball players, which can improve the efficiency of athletes' training and improve their physical fitness and competitive ability.

2. Proposed Method

2.1. Deep Convolutional Neural Network VGG Structure. Deep convolutional neural network is a special type of neural network. Its super learning ability is mainly realized by using multiple nonlinear feature extraction stages, which can automatically learn hierarchical representations from data. This kind of convolutional neural network is mainly composed of input layer, convolution layer, activation function, pooling layer, fully connected layer, and output layer.

VGG-16 is a 16-layer convolutional neural network model, which contains 13 convolutional layers (conv), 525 pooling layers (pool) and three fully connected layers (FC), by dividing the convolutional layer group. In operation, the 16 layers are divided into five convolutional layer groups and three fully connected layer network structures [4, 5]. The network structure is as follows.

2.1.1. Convolutional Layer Group A. In the structure of convolutional layer group A, it contains two convolutional layers and a maximum pooling layer. Each convolutional layer uses 64 convolution kernels with a size of 3×3 . The step size of the convolution kernel is 1, and the padding is 1. The step size (s) and fill size (p) satisfy the following relationship [6, 7]:

$$y = \frac{x + 2 \times p - k}{s} + 1. \quad (1)$$

2.1.2. Convolutional Layer Group B. Convolutional layer group B contains two convolutional layers and a maximum pooling layer. Each convolutional layer uses 128 convolution kernels of size 3×3 . The step size of the convolution kernel is 1, and the padding is 1. Therefore, the size of the feature map is $(112 + 2 \times 1 - 3) \div 1 + 1 = 112$; the pooling layer uses maximum pooling [8, 9]. The size is 2×2 , and the step size is 2; then, the size of the feature map after maximum sampling is $(112 - 2) \div 2 + 1 = 56$.

2.1.3. Fully Connected Layer Group. The fully connected layer group has three fully connected layers, of which the first two fully connected layers contain 4096 neurons, and the last fully connected layer's neuron output must be consistent with the number of categories to be divided [10, 11]. The last fully connected layer uses Softmax classifier to classify the data. Softmax is a normalized function that can convert a set of ratings of $(-\infty, +\infty)$ into a set of probabilities, and let their sum be 1, and this function is order-preserving. The original large rating is converted The probability is large, and the small score corresponds to the small probability.

The formula of the Softmax classifier is as follows:

$$P(i) = \frac{\exp \theta_i^T x}{\sum_{k=1}^K \exp(\theta_k^T x)}, \quad (2)$$

where $\theta_i^T x$ is multiple inputs. It can be seen from the previously mentioned formula that the Softmax classifier can obtain multiple values, and the sum of these numbers is 1, and the output results are in the interval $[1, 0]$. Such a form can be seen. It becomes a question of probability [12, 13].

2.2. Deep Convolutional Neural Network Training

2.2.1. Data Processing. The datasets used in this article are the KTH and Weizmann datasets. [14]. KTH includes 2391 video samples of six types of actions performed by 25 people in four different scenarios. It makes it possible to use the same input data to systematically evaluate the performance of different algorithms. KTH and Weizmann datasets are the most cited databases in the field of behavior recognition, and they have greatly promoted the research of behavior recognition. Both KTH and Weizmann are public datasets, where KTH includes six types of human motions, and Weizmann includes ten types of human motions [15, 16]. First, the Matlab platform is used to intercept the video data of two datasets every five frames as the marker set of the dataset. Then, valid frames that can correctly express the characteristics of the action in the allocation set are selected, and they are processed into sizes through geometric changes. It is a 224×224 image dataset [17, 18].

2.2.2. Bi-GRU Neural Network. The process of obtaining a linear sum between the existing state and newly calculated state is similar to the LSTM unit [19, 20]. However, GRU has no mechanism to control the degree of exposure of its state, and the entire state is displayed every time [21, 22]:

$$\begin{aligned} z_t &= \sigma(w_z \cdot [h_{t-1}, x_t]), \\ r_t &= \sigma(w_r \cdot [h_{t-1}, x_t]), \\ h_t &= \tanh(w \cdot [r_t * h_{t-1}, x_t]), \\ h_t &= (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t, \end{aligned} \quad (3)$$

where z_t represents the value of the update gate; h is the value of the hidden layer; σ represents the activation function; $\tanh$

represents the activation function; w represents the weight; r_t represents a set of reset gate values [23, 24].

The previously mentioned is an introduction to the neural network used in this article. Each neural network has its own advantages and characteristics. Because smartphones are easy to carry, it is very convenient for coaches to test the athletes' movements, and TensorFlow can be used on the phone. TensorFlow officially supports iOS and Android, and the display of this model is very clear. Transplanting the trained model to a smartphone to test the recognition of the neural network model is the best way to test the model.

3. Experiments

3.1. State Extraction and Recognition of Human Basketball. This article mainly uses the acceleration sensors and gyroscopes in Android smartphones to collect the status data of common human basketball sports and builds a neural network architecture through the TensorFlow deep learning platform, uses a variant of the recurrent neural network to build a network model, and selects and recognizes through experimental results. The neural network model with better performance is transplanted into the smartphone. The gyroscope is used to measure angular velocity, and the acceleration sensor is used to measure linear acceleration. The former is the principle of inertia, and the latter is the principle of force balance. The measured value of the acceleration sensor is correct for long periods, but for short periods, due to the existence of signal noise, there is an error. The gyroscope is more accurate in a short period of time, and there is an error with drift in a longer period of time.

3.2. Human Basketball Status Recognition Steps. First, collect acceleration and gyroscope sensor data under different basketball motion states of the human body. Here, PhyPhox sensor data acquisition software is used for data acquisition. Second, preprocess the collected data, including data denoising, filtering, and marking the human basketball game state of motion; third, divide the length and data structure of each piece of data to achieve data segmentation; fourth, use the convolutional neural network in deep learning to extract the features of each collected human action; fifth, continuously adjust the hyperparameters and neural network structure to obtain a neural network model; the sixth step is to combine the neural network model with TensorFlow technology, and transplant Android technology to mobile phones for viewing and analysis.

4. Discussion

4.1. Bi-LSTM Neural Network Recognition Analysis. Use Table 1 to calculate the accuracy, recall, and F values. Precision, also known as accuracy and correct rate, represents the proportion of related documents retrieved among all documents retrieved. Recall represents the ratio of retrieved related documents among all related documents. F-measure is the weighted harmonic average of accuracy (P) and recall (R), where 0.867 is the final accuracy of classification. From the experimental results, it can be seen that

TABLE 1: Bi-LSTM neural network confusion polyline result.

		Predicted class			
		WK	WU	SI	ST
True class	WK	518	26	0	4
	WU	5	521	0	3
	SI	0	2	286	115
	ST	3	2	43	421

the Bi-LSTM neural network recognizes WK and WU better, but the recognition probability of SI and ST is lower, mainly because the two actions of SI and ST are relatively similar and belong to the motion of the static category. It is difficult to distinguish between similar features with high feature similarity, which reduces the overall recognition accuracy. Recognizing similar motion states has always been the difficulty of state recognition. You can consider increasing the amount of data or adding different types of sensors to increase unique characteristics of similar motion states. It can be seen from Figure 1 that the position corresponding to WK-WK on the line chart is the number of a class that is predicted to be correctly classified. The larger the number at the corresponding position, the better the classification result. Except for the numbers in the corresponding positions, the others are misclassified.

4.2. Analysis of Training Accuracy Generated by TensorFlow for Human Basketball. Table 2 is a table of training accuracy rates generated by TensorFlow. In Figure 2, the gray bars represent the recognition accuracy of the training dataset during training. The red curve represents the recognition accuracy of the test dataset. It can be found that, from the beginning of the training to the end of the data, the overall slope of the two curves is getting smaller. Simultaneously, the accuracy of the training data is high, indicating that the model recognizes the data that was learnt better. It also has a good ability to recognize unknown data, indicating that the model has strong generalization ability and can recognize unknown data well. In the first 200 iterations, the two curves rise faster, indicating that the loss function has a good ability to deal with the model, can quickly converge, and extract the key features of the state of human basketball. After 200 times, relatively slow learning started, and more detailed extraction and learning of feature details began.

In general, the training accuracy table generated by TensorFlow can quickly extract the key features of the athlete's actions, and the more the number of times, the more detailed the feature extraction.

4.3. Feature Extraction Analysis. As shown in Figure 3, the difference between the RUN and WAL signals is very obvious, and the periodicity of each signal is relatively regular. The difference is mainly the length of each cycle and the size of the highest and lowest values. Through the comparison between the signals, the differences between the signals and their respective movement characteristics are found. In the feature extraction, through the analysis of the acceleration

FIGURE 1: Bi-LSTM neural network confusion matrix results.

TABLE 2: Bi-LSTM neural network training accuracy.

		Experimental parameters						
	0.00	200.0	400.0	600.0	800.0	1.00k	1.20k	1.40k
Train	56	847	889	919	953	971	979	982
Test	49	783	832	852	875	879	872	873

FIGURE 2: Bi-LSTM neural network training accuracy rate graph.

and angular velocity signals, the action state can be better extracted.

4.4. Using APP to Collect Basketball Player Movement Data. The number of times each motion state can be correctly identified after 100 tests. For example, after collecting 100

pieces of WAL data on a smartphone and performing average filtering of 5-bit, 7-bit, and 9-bit filtering, the number of times correctly recognized by the neural network model is 95, 90, and 88 respectively. Through experimental comparison, it is found that the number of one-time filtering bits of mean filtering also has a certain influence on the experimental results. The experiment is shown in Figure 4.

FIGURE 3: Two kinds of human bodies playing basketball signals. (a) WAL acceleration signal. (b) WAL angular velocity signal.

FIGURE 4: Comparison of mean and median filter recognition.

5. Conclusions

The use of deep learning-based dynamic human action extraction and recognition technology for basketball players can realize the standardization and determination of athletes' movements, which is very helpful for future Chinese athletes to achieve better results on the playing field.

In the process of neural network model structure design and parameter tuning, it is necessary to use deep learning techniques to optimize, including regularization and discarding hidden layer units, and gradually adjust the neural network structure through continuous experiments to obtain a model with good performance.

On using the public human body motion dataset as the object, the data is preprocessed to construct a data format that meets TensorFlow requirements. The network structures of convolutional neural networks, classic neural

networks, unidirectional recurrent neural networks, and bidirectional recurrent neural networks are compared.

In the process of constructing the neural network structure, the long-term and short-term memory neural network of the recurrent neural network and its variants are analyzed and compared, and the neural network model with better performance is obtained by continuously adjusting the neural network model and parameters.

The design of the neural network model structure in this article relies on deep learning, but the angle of the experiment in this article is not comprehensive enough, and this model has not been put into the training of actual basketball players. With in-depth learning of the neural network, we can continue to optimize it.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] D. Holden, J. Saito, and T. Komura, "A deep learning framework for character motion synthesis and editing," *ACM Transactions on Graphics*, vol. 35, no. 4, pp. 1–11, 2016.
- [2] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [3] S. R. Granter, A. H. Beck, D. J. Papke, and P. AlphaGo, "AlphaGo, deep learning, and the future of the human microscopist," *Archives of Pathology & Laboratory Medicine*, vol. 141, no. 5, pp. 619–621, 2017.
- [4] S. R. Granter, A. H. Beck, and D. J. Papke Jr, "Straw men, deep learning, and the future of the human microscopist: response to 'artificial intelligence and the pathologist: future frenemies?'" *Archives of Pathology & Laboratory Medicine*, vol. 141, no. 5, p. 624, 2017.

- [5] J. Miyake, Y. Kaneshita, S. Asatani, S. Tagawa, H. Niioka, and T. Hirano, "Graphical classification of DNA sequences of HLA alleles by deep learning," *Human Cell*, vol. 31, no. 2, pp. 102–105, 2018.
- [6] H. Shim, K.-E. Ko, and K.-B. Sim, "Object-human interaction pattern generating system using deep learning," *Journal of Institute of Control, Robotics and Systems*, vol. 23, no. 5, pp. 317–322, 2017.
- [7] S. Narejo, E. Pasero, and F. Kulsoom, "EEG based eye state classification using deep belief network and stacked AutoEncoder," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 6, no. 6, p. 3131, 2016.
- [8] J. Wei, Q. Fang, X. Zheng, W. Lu, Y. He, and J. Dang, "Mapping ultrasound-based articulatory images and vowel sounds with a deep neural network framework," *Multimedia Tools and Applications*, vol. 75, no. 9, pp. 5223–5245, 2015.
- [9] J. Bruineberg and T. Froese, "Surfing uncertainty: prediction, action and the embodied mind," *International Journal of Performance Arts and Digital Media*, vol. 12, no. 2, pp. 215–217, 2016.
- [10] K. H. Wood, "Curative pedagogy in the undergraduate theatre historiography classroom," *Theatre Topics*, vol. 27, no. 3, pp. 187–196, 2017.
- [11] C. Shen, X. Huang, and Q. Zhao, "Attention in low resolution: learning proto-object representations with a deep network," *Journal of Vision*, vol. 15, no. 12, p. 898, 2015.
- [12] F. Wang, W. Gong, J. Liu et al., "Channel selective activity recognition with WiFi: a deep learning approach exploring wideband information," *IEEE Transactions on Network Science & Engineering*, vol. 99, p. 1, 2018.
- [13] R. M. Nishikawa and K. T. Bae, "Importance of better human-computer interaction in the era of deep learning: mammography computer-aided diagnosis as a use case," *Journal of the American College of Radiology*, vol. 15, no. 1 Pt A, pp. 49–52, 2017.
- [14] P. Wang, H. Liu, L. Wang, and R. X. Gao, "Deep learning-based human motion recognition for predictive context-aware human-robot collaboration," *CIRP Annals*, vol. 67, no. 1, pp. 17–20, 2018.
- [15] J. P. Davis and W. A. Price, "Deep learning for teaching university physics to computers," *American Journal of Physics*, vol. 85, no. 4, pp. 311–312, 2017.
- [16] M. Koohzadi and N. M. Charkari, "Survey on deep learning methods in human action recognition," *Iet Computer Vision*, vol. 11, no. 8, pp. 623–632, 2017.
- [17] K. Kiruba, D. S. Elizabeth, and C. S. R. Raj, "Deep learning for human action recognition survey," *International Journal of Computer Sciences and Engineering*, vol. 6, no. 10, pp. 323–328, 2018.
- [18] S. Zhang, S. Gao, and G. Wang, "Robust NLOS error mitigation method for TOA-based localization via second-order cone relaxation," *IEEE Communications Letters*, vol. 19, no. 12, p. 1, 2015.
- [19] K. Fernandes, D. Chicco, J. S. Cardoso, and J. Fernandes, "Supervised deep learning embeddings for the prediction of cervical cancer diagnosis," *PeerJ Computer Science*, vol. 4, no. 8, p. e154, 2018.
- [20] W. Lee, "A deep learning analysis of the KOSPI's directions," *Journal of the Korean Data and Information Science Society*, vol. 28, no. 2, pp. 287–295, 2017.
- [21] G. Boccignone, D. Conte, and V. Cuculo, "Deep construction of an affective latent space via multimodal enactment," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 99, p. 1, 2018.
- [22] H. K. Zhang, Y. Li, and Y. N. Jiang, "Deep learning for hyperspectral imagery classification: the state of the art and prospects," *Zidonghua Xuebao/acta Automatica Sinica*, vol. 44, no. 6, pp. 961–977, 2018.
- [23] D. Wang, "Deep learning reinvents the hearing aid," *IEEE Spectrum*, vol. 54, no. 3, pp. 32–37, 2017.
- [24] S. Yao, Y. Zhao, A. Zhang et al., "Deep learning for the internet of things," *Computer*, vol. 51, no. 5, pp. 32–41, 2018.