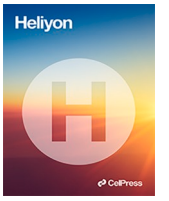
Heliyon 9 (2023) e18124

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Convolutional neural network-based recognition method for volleyball movements

Hua Wang [a], Xiaojiao Jin [b], Tianyang Zhang [c], Jianbin Wang [b,*]

[a] Physical Education Department, Xinxiang Institute of Engineering, Xinxiang, 453000, China
[b] Physical Education Department, Xingtai Medical College, Xingtai, 054000, China
[c] Sports Work Department, Hebei Vocational University of Technology and Engineering, Xingtai, 054000, China

ARTICLE INFO

Keywords:
Convolution neural network
Physical education curriculum
Motion recognition
Accuracy
Complexity

ABSTRACT

With the development of network technology and computer intelligent monitoring technology, a large number of video data came into being. In view of the analysis of specific targets in video, the traditional artificial analysis method can not meet the existing needs. In volleyball teaching in college physical education, because each student has different movements, intelligent processing of video data has become a key issue. Through collecting and sorting out the relevant research results on behavior recognition, it is found that in the research of deep learning, the algorithm structure applied in this paper is representative, and its strong learning ability, especially compared with the traditional algorithm, can more accurately identify human movements, natural language processing and so on, but the research on volleyball action recognition is still less. Therefore, this paper constructs a data set, improves the convolution neural network model, subsequently, new models are constructed through the neural network structure to improve the accuracy of the nonlinear expression and optimize the content of the input data. In order to more accurately analyze the effectiveness of this algorithm, the new data are obtained by grouping the volleyball games in college sports courses. Compared with the original paper, the accuracy of the improved 3D network is improved by 3.3%–88.5%, and the complexity is reduced by 33.6%.

1. Introduction

Human motion recognition has always been a hot and difficult point in computer vision research, and it also has important application value in real life [1]. Human motion recognition is widely used in video surveillance, virtual reality, human-computer intelligent interaction and other fields. With the attention of education departments to sports, most of the current sports research starts from three aspects: moving object detection, action feature extraction and action feature understanding, and basically solves the problem of human motion recognition in simple scenes [2]. Because the research process will encounter many different factors, such as the difference of action performance, the difference of environment, the difference of time and so on [3]. In order to solve these differences, some excellent feature representation methods have been proposed, among which the most popular and advanced one is deep learning representation method [4]. Especially in motion recognition, different motion types show great differences in appearance and motion model. Manual setting needs to rely on experience and luck to obtain better characteristics, so it is difficult to ensure the acquisition of its essential characteristics in college physical education curriculum volleyball [5].

* Corresponding author.
E-mail address: wjbhlf@163.com (J. Wang).

https://doi.org/10.1016/j.heliyon.2023.e18124
Received 27 March 2023; Received in revised form 7 July 2023; Accepted 7 July 2023
Available online 12 July 2023

Therefore, we need a method that can learn features automatically, and solve the blindness and one-sidedness of the time-consuming manual extraction method of volleyball action features in college physical education courses. Deep learning is developed on the basis of machine learning [6]. It adopts hierarchical processing mechanism and can automatically learn advanced features from input data layer by layer. Therefore, deep learning successfully replaces manual feature extraction methods through unsupervised or semi-supervised feature learning and hierarchical feature extraction algorithms [7]. Convolution neural network (CNN) is a depth model, which directly and alternately performs convolution and sub-sampling operations on the original input image, and gradually obtains layered complex features [8]. In the volleyball action recognition of college physical education curriculum, the convolution neural network can get better performance in the visual target recognition task by properly adjusting it.

Study the space-time characteristics of sports video. In order to make the recognition method is more practical value, this paper proposed a multi-resolution 3 d convolution neural network model. In the premise of high resolution of the original input stream, increase a contains the action of low resolution input stream, form a new shuangliu 3 d convolution neural network framework. So it can be extracted by using 3 d convolution kernels continuous space and time information of a video frame, and accelerated the speed of the network. Experiments show that this method without any priori information and the traditional algorithm of similar results have been achieved. Combined with the above contents, it is found that it is of great significance to establish a general, stable and excellent motion recognition system by using the advantages of deep learning model in feature extraction. At the same time, in the process of human motion recognition in video, there is also a very important problem is real-time. In this paper, a deep learning model is introduced to obtain the essential features of human motion in video and improve the real-time performance of human motion recognition. In addition, the theoretical research and practical exploration of the deep learning model can find and improve the shortcomings of the deep learning model in the application of computer vision, which is helpful to develop and improve the deep learning model and provide reference for volleyball action recognition in college physical education courses.

This research aims to improve the application of deep learning algorithms in volleyball action recognition. It has significant theoretical implications for promoting the use of deep learning in sports education and expanding the application areas of deep learning algorithms. By constructing a new dataset and enhancing the convolutional neural network model, this study improves the accuracy of nonlinear expression and optimizes the content of input data. This is of great theoretical significance for improving the accuracy and efficiency of action recognition and providing new ideas and methods for related research fields. By intelligently processing video data, it is possible to accurately recognize students' actions and provide personalized guidance and feedback. The proposed improved algorithm in this study can enhance the accuracy of volleyball action recognition and provide teachers with more refined teaching assistant tools, thereby improving teaching effectiveness. By optimizing the algorithm structure and the content of input data, this research enhances the precision of behavior recognition while reducing complexity. This offers feasibility for resource and cost savings in practical applications and makes behavior recognition technology more practical and feasible.

2. Related work

2.1. Research status of traditional motion recognition

2.1.1. Research based on template matching

Template matching is a method in image recognition, which is a method to distinguish behavior types according to the sum of Euclidean distances between various features. The advantage of template matching method is that the concept is simple, and the recognition effect is better for images with large difference in attributes, but the disadvantage is that the recognition effect is poor for images with small or only subtle difference in attributes [9]. The method based on template matching is a method to identify behavior through contour features. The earliest method to calculate the difference is Motion Energy Image (MEI), which can identify actions according to the difference of contour features and the energy map of motion. After improvement, a motion history image (MHI) generated by time series frames is proposed, which can identify motion information by calculating the brightness of moving pixels [10]. By extracting the contour features of a single angle of view, scholars calculate the sum of contour differences between consecutive frames, and use MEI and MHI to construct motion effects, so as to calculate the type discrimination behavior of geometric moment invariants matching [11]. Euclidean distance is the most commonly used distance, but it has obvious defects when calculating the distance between feature vectors of time series [12]. Therefore, a new time series distance calculation method is used, and a matching algorithm for time-varying data series is proposed, that is, dynamic time warping method. This method has obvious advantages, simple concept and robust algorithm, and can classify image sequences. However, the calculation involved is relatively complex, and the dynamic correlation between adjacent sequences of motion sequences in time and space is not considered.

2.1.2. Research based on state space

Considering the continuous change of motion sequence in time and space, the state space method which can express the characteristics of motion time series is adopted. In this method, one of the representative methods is Markov network. In behavior recognition, the joint probability of action sequence is calculated to recognize actions. In the following research, many improvements have been made to the Hidden Markov Model [13]. There are many improved models based on Markov Model, but the Hidden Markov Model can deal with two independent processes at most.

Dynamic Bayesian network based on state space method can be regarded as an extension of time-hidden Markov model. Dynamic Bayesian network holds that the connection between actions has a certain sequence. According to this sequence, the timing characteristics of actions in video can be judged [14]. The advantage of dynamic Bayesian network is that it can change the topological structure without affecting the algorithm training, and can also change the correlation between visual variables by adding or deleting

variables. Through the visual variable relationship, the network principle can be clearly defined. At the same time, because the network is a directed graph describing stochastic process, this structure makes Bayesian network have more explicit probability meaning [15]. But relatively speaking, Bayesian network has many parameters, so the calculation is complicated when training the network.

2.2. Research status of motion recognition based on deep learning

2.2.1. Human action recognition method based on depth image information

Depth image generally use depth measured the distance from the camera said the size of the pixels of the pixel value, to light, texture, etc [16,17]. Have strong robustness. In based on depth image studies, some scholars use depth camera to the human body movement for multiple directions of orthogonal projection, the difference between the two frames under more than as a means of motion description to do the three directions of the depth of the orthogonal projection image four dimensional space-time sampling, statistics under different scales of time and space information, statistical son the number of pixels in the air, again after the sparse coding, the method of using the SVM classification 1 [18,19]. In the process of action recognition can be through the feature extraction of from bottom to top, to extract the characteristics of local community has carried on the related inquiry have taken different approaches, such as word bag model [20–22]. Sparse Coding model (Sparse Coding. SC), naive bayesian nearest Neighbor (Native BayesNearest Neighbor, NBNN) and other methods. In depth information, the 3 d point cloud is also regarded as an effective method. Rahmani and others related directly with their 3 d point cloud processing, using 3 d point cloud principal component histogram as a descriptor by the action of recognition, the method can change to the noise and movement performance differences and perspective is not sensitive [23,24].

2.2.2. Motion recognition method based on convolution neural network

The first step of this recognition method is to segment the motion in the video and divide the whole motion video into segments which are easy to be recognized. Then, spatio-temporal coding and decoding are used for recognition. There are four coding methods in action recognition method based on convolution neural network. The first one is to process video frames separately, and then integrate the information extracted from each frame, and finally take the integrated information as the coded information; The second is to use three-dimensional convolution neural network to extract feature information in time domain and space domain; The third is to encode the video directly into an image, and the processed image includes the spatio-temporal information in the original video [25]. The processed image is sent to the convolution neural network to train and learn the features in the image; The fourth is to express temporal and spatial information by multi-channel network.

2.2.3. Motion recognition method based on cyclic neural network and related algorithms

Because the elements in the existing neural network are independent of each other and have no memory, the existing neural network can not show the expected effect for some problems that need to be connected with context. Therefore, in order to better deal with some sequence models, researchers proposed a new network structure, namely cyclic neural network [26]. In the later improvement, the unidirectional neural network is changed into bidirectional neural network and deep bidirectional neural network, but these improvements have the disadvantage of cyclic neural network, that is, gradient dissipation. Therefore, researchers put forward long-term memory network, and after improvement, put forward long-term cyclic neural network.

2.3. Comparison and difficulties between traditional and deep learning recognition methods

2.3.1. Comparison of traditional methods and deep learning methods in motion recognition

Traditional method is an algorithm to construct features according to specific tasks. The traditional method is effective in dealing with some simple classification problems or small data volume problems. Deep learning can be regarded as a feature learner and a data-driven algorithm in essence [27]. Here, we will compare traditional methods with deep learning methods from the following aspects:

Performance: Deep Web has achieved far more accuracy than traditional methods, including voice, natural language, vision and game playing. Taking the image classification accuracy of different methods on ILSVRC, a large-scale visual recognition challenge of ImageNet data set, as an example, the Top-5 error rate of ILSVRC in recent years is selected to illustrate the performance of deep learning.

Model extensibility: Traditional methods build models based on domain-specific and application-specific algorithms and feature engineering. The knowledge base of traditional methods is very different for different fields and applications, and usually requires extensive professional research in each separate area. If there is a higher requirement for accuracy, the traditional method needs more complex algorithms to complete. Compared with traditional methods, deep network can be expanded better by using more data, and the accuracy of network can be improved by using more data. When adapting to different fields and applications, deep learning technology is adaptable and easy to transform, and transfer learning can meet the requirements more easily.

2.3.2. Difficulties in motion recognition based on deep learning

It is difficult to train the model with large parameters. In the learning of network model, with the increase of model complexity, the number of parameters is multiplied. In order to better train the parameters in the model and avoid the lack of network learning, the amount of data required for network training is also greatly increased. In a certain layer of the model, the parameter quantity is calculated as the product of the filter size, the size of the feature map and the depth of the feature map, while the parameter quantity is

not increased in the pooled layer [28]. The fully connected layer is the layer that produces the largest number of parameters in a network. The parameter calculation of the fully connected layer is the product of the number of input channels and the number of output channels. Because there are forward propagation and back propagation in training, the parameter quantity should be multiplied by 2.

Limitations of characterization features. The ability of network characterization is expressed by learning parameters in the network. The more parameters, the more accurate the network characterization, the stronger the ability to express. For a network whose structure has been determined, the parameters are limited, the accuracy of corresponding characterization features is limited, and the expression ability is also limited. If the depth of the network is increased and the number of parameters used in the network is increased, the expression ability of the network will be improved, but at the same time, the data set with huge amount of data is required to take the training network parameters [29,30]. For the current situation, high-quality data sets are limited, so the network with huge parameters will be limited in expression ability due to incomplete training.

3. Volleyball action recognition algorithm in college physical education curriculum based on 3D convolution neural network

3.1. Overview of convolution neural network

The basic structure of convolution neural network is composed of three network layers, which are convolution layer for convolution operation, pooling layer for feature screening and full connection layer for feature fusion. Among them, the convolution layer of each layer is composed of many neurons, and a weight value will be obtained between neurons during training or learning. The neurons between convolution layers map the data through nonlinear function, and pass the simplified data to the neurons in the next layer according to the weight. Finally, the results are obtained in the output layer through the full connection layer, which is the classification category. The function of convolution layer is to convolve the input image to obtain the underlying features in the image. The input of the convolution layer can be the original video frame or image of the input, or it can be the result of the previous convolution layer. In the process of convolution operation, the convolution object is input and convolution kernel, in which one convolution kernel should be a convolution result, that is, a feature, and each convolution kernel has corresponding weight coefficient and carried offset. The convolution node is shown in Fig. 1.

In the past many researches, 2DCNN was used to extract the features of each video frame, and then various related algorithms were used to combine the features of video frames. This method can extract the motion features of each static frame, but for a motion action, the motion is a continuous process, and an action will last for several frames and dozens of frames, while the data processing of static video frames does not consider the motion information of motion action in time dimension. Spatio-temporal features extracted by 3D convolution include not only static spatial semantic information, but also motion time sequence information. The principle of using 2D convolution kernel and 3D convolution kernel for image sequence is shown in Fig. 2.

3.2. Improved C3D network structure design and network hyperparameter selection

In the research of volleyball action recognition in college physical education curriculum, in order to enhance the learning ability of network for features and enrich the feature space, this topic uses two 1*3 and 3*1 long convolution kernels for convolution operation, as shown in Fig. 3. Because this network will be used as the backbone network of feature extraction in the follow-up research, the C3D full connection layer will remain the original connection without processing. In this topic, for the convenience of writing, this improved network is called improved C3D. The improved C3D network structure is shown in Fig. 3.

Every pixel in a video frame or image is a mixture of three primary colors. During network training, the data read at one time is large, which seriously affects the computing speed of the computer during network training. Each pixel in the gray image has a range of change, but it can still keep the gradient and other information in the original image. When the network reads gray-scale data, it saves memory and makes it easier to process gray-scale features. In the recognition of volleyball movements in college physical education courses, the change of light and different colors of clothes will affect the recognition, while the gray characteristics of human movements can ignore the influence of light and color, so the gray characteristics of movements are a factor that needs to be considered [31].

In the process of network training, the setting of parameters in the model is an important factor affecting the recognition effect of the model. In this chapter, in the process of determining the parameters in this paper, the collected volleyball movements in the

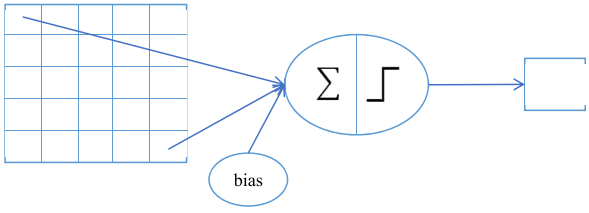


Fig. 1. Convolution node connection.

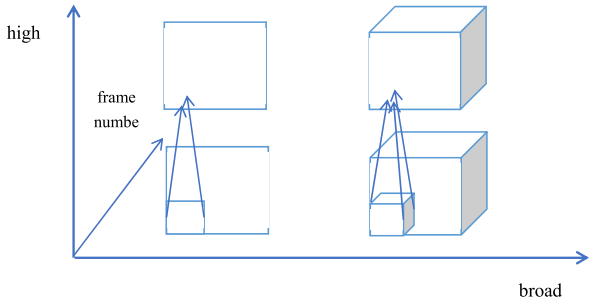


Fig. 2. Schematic diagram of two-dimensional convolution and three-dimensional convolution.

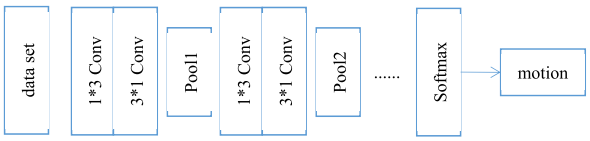


Fig. 3. Improved C3D structure diagram.

university physical education course were processed through several iterative processing, and the subsequent training process was completed. After sorting out the data, it was found that the videos of the volleyball movements in the university physical education course had a similar size, so the grayscale image was taken as the training parameter.

3.3. Improvement of data input based on network

3.3.1. Moving object detection and region segmentation based on GMM

The purpose of target location is to detect the movement area, but it doesn't pay attention to the specific volleyball movement or whether the detection results contain holes and whether they are complete. Because we only need to know the range of motion to detect the moving region, we choose Gaussian mixture model to detect and locate the target. In this subject, through the area detection of college sports volleyball action recognition, firstly, the human body motion area in the original video frame is detected, and then the human body area in the video frame is segmented according to the detection results, which is used as the input data for extracting gray features. The steps of target detection and region segmentation are shown in Table 1.

3.3.2. Feature fusion based on data enhancement

According to the target detection and gray feature analysis, considering comprehensively, the input data of the network is adjusted as follows: First, the fusion of gray image and RGB image: adding gray feature in the input, adding the video frame after gray processing as the input data of gray feature to the network, training, and extracting gray feature as recognition reference. For gray feature extraction, the 3D CNN network structure here uses a single-channel convolution network. For color image feature extraction, the same network structure is used. In feature fusion, two kinds of feature vectors are averaged as the final extracted feature vectors. At this time, the feature vector dimension is 4096, which is the same as the feature vector dimension extracted by single feature. Finally, the decision maker Softmax is used to classify and output the behavior tags identified by the network. Secondly, the fusion of Crop color map and original color map: detect and extract the key areas of human motion from video frames, add them to training as Crop features, add the fusion of Crop features and video frame features after feature extraction, and finally identify actions through decision-making [32–34].

Table 1
Target detection and region segmentation.

Steps	Project	Specific content
1	Target detection	Using GMM to detect and mark the main moving area of the target
2	Binary target motion region	The center point of the moving region is calculated by binary processing of the image
3	Get the moving area	The position of the moving region in the original image is obtained according to the center point
4	Segmented image	Use 112*112 size bounding box to intercept the moving area

4. Results analysis of experimental accuracy and complexity

4.1. Accuracy recognition result

For target detection, the experimental results are explained and analyzed by taking the result graph of volleyball in college physical education course in the training set of THUMOS14 and UCF101 data set as an example. When determining the size of the motion region, considering that it contains the motion region in most video frames, this paper adopts 125*125 bounding frames respectively.

4.1.1. Analysis of recognition result of RGB video frame

As a comparative experiment, RGB video frames are first used for recognition. The experimental results are shown in Fig. 4. The accuracy rate on the test data set is 0.96, and the accuracy rate on the verification data set is 0.85.

4.1.2. Analysis of recognition results based on fusion features

From the experimental results, we can see that different feature fusion has a certain impact on later motion recognition, and the accuracy of gray image features and RGB color image is quite different. Based on the fusion of gray features and RGB features, the final experimental results prove that the accuracy is between the accuracy of gray and RGB color images, as shown in Fig. 5. It is proved that the gray image features will have great interference to the recognition effect of the original image in the scene with complex background.

The experimental results after using the fusion of crop features and RGB features show that using GMM to extract crop features after target detection and fusing RGB video frames, the effect is better, and the accuracy rate is 0.88, as shown in Fig. 6. Compared with the recognition accuracy rate of 0.85 according to the network structure in the original paper, the recognition effect of combined features is better than that of single feature.

Through the experimental analysis of the above single feature and combined feature, the results of the above experiments are summarized as shown in Table 2. According to the experimental data, in the format of data input, it can be seen that the accuracy of the model for action recognition is the lowest when using gray image to train the network. Through the identification results, the identification accuracy of the two images is 0.8, which is significantly improved than the accuracy without combination. The subsequent comparison with the original image shows that the latter is more accurate. Raw images are taken in in the training process through data optimization to determine the accuracy of the analysis results. The recognition effect of the obtained model is due to the first three data combination methods. The loss value of the model decreased from 6.7 to 3.8, and the accuracy rate increased to 88%, which proved the effectiveness of the improvement.

4.2. Model complexity detection results

The analysis of complexity can be conducted through two aspects, namely, time and space, where the former refers to the time of volleyball sports recognition in college physical education courses. When the complexity is relatively high, it indicates that more time is consumed during training and less efficient. Spatial complexity is the number of training participants. So that the number of data sets should be reasonably controlled during the training process, so as to improve the accuracy of the training results as a whole.

4.2.1. Time complexity

Time complexity is the number of operations of the model, which can be measured by FLOPs, that is, the number of floating-point operations. The time complexity of a single convolution layer is calculated as shown in Equation (1):

$$Time \sim 0\left(Map_{wide} \times Map_{height} \cdot Kernel_{wide} \times Kernel_{height} \cdot Channel_{in} \times Channel_{out}\right) \tag{1}$$

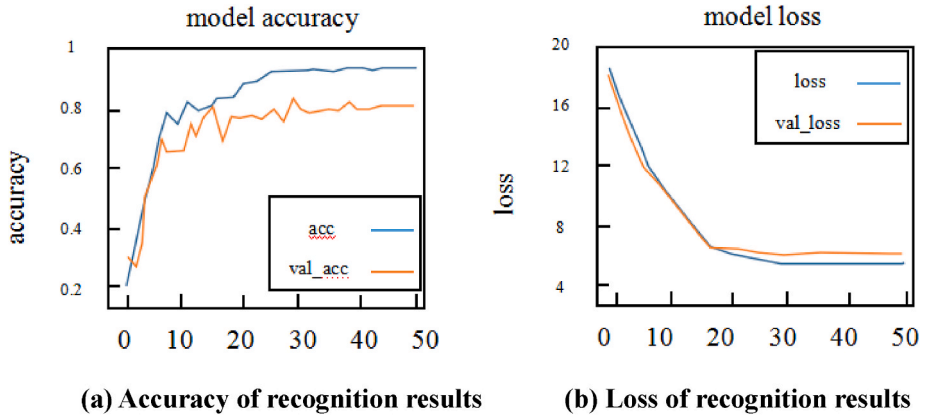


Fig. 4. Recognition results based on RGB.

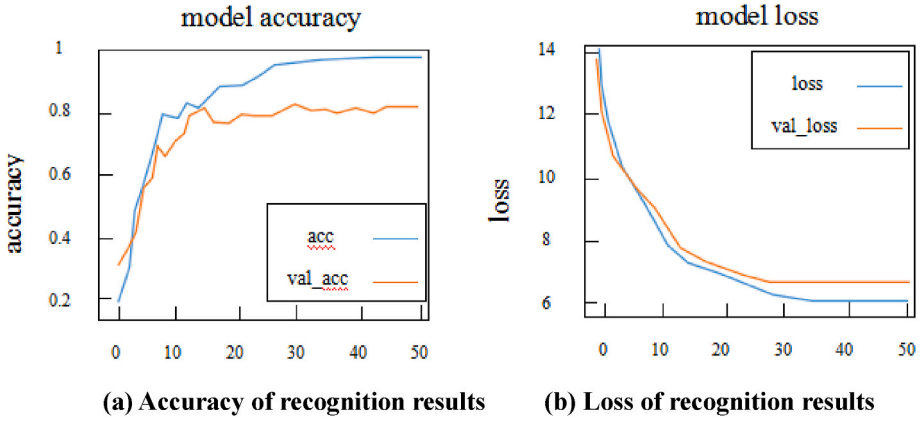


Fig. 5. Recognition results based on fusion of gray features and RGB features.

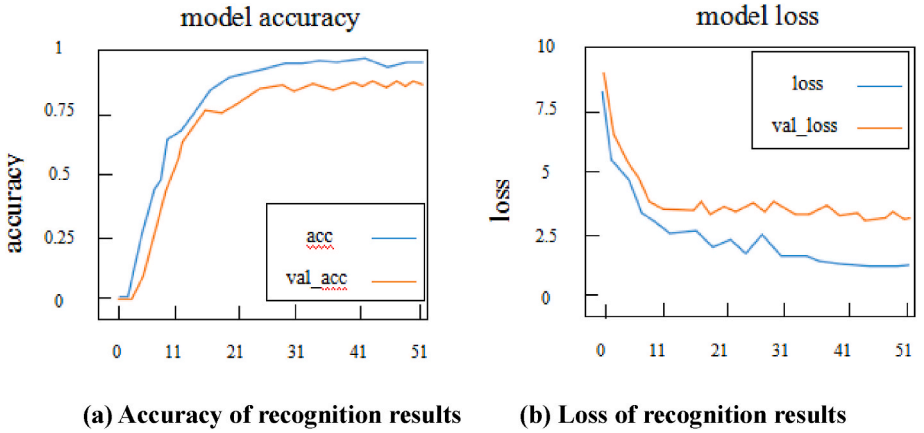


Fig. 6. Recognition results based on RGB feature and CROP feature fusion.

Table 2
Summary of experimental results.

Feature selection	Test set accuracy	Verify set accuracy	Test set loss	Verification set loss
GRAY	0.97	0.72	0.1	1.7
RGB	0.97	0.85	6.1	6.7
RGB + GRAY	0.96	0.8	2.1	5.6
RGB + CropRGB	0.97	0.88	1.9	3.8

Among them, Map_{wide} and Map_{height} output the width and height of the feature map of each convolution kernel; $Kernel_{wide}$, $Kernel_{height}$ the width and height of each convolution kernel; Therefore, the time complexity of each convolution layer is completely determined by the area of output feature map, the area of convolution kernel and the number of input and output channels. Among them, the output feature map itself is determined by four parameters: input matrix size X, convolution kernel size K, Padding and Stride, which are shown in Equation (2):

$$Map = \frac{(X - K + 2 * padding)}{Stride} + 1 \tag{2}$$

The overall time complexity calculation of convolution neural network is shown in Equation (3):

$$Time \sim 0\left(\sum_{l=1}^{Dep} Map_{l,wide} \times Map_{l,height} \cdot Kernel_{l,wide} \times Kernel_{l,height} \cdot Channel_{l-1} \times Channel_{l}\right) \tag{3}$$

For the l and convolution layer, the input channel number $Channel_{l-1}$ is the output channel number of the $l-1$ convolution layer.

4.2.2. Spatial complexity

Spatial complexity consists of two parts: the total number of parameters and the output characteristic map of each layer. The calculation formula is shown in Equation (4).

$$Space \sim O(parama + Map) \tag{4}$$

The parameter quantity is the total weight parameter of all layers with parameters in the model, as shown in Equation (5).

$$parama = \sum_{l=1}^{Dep} Map_{l,wide} \times Kernel_{l,height} \cdot Channel_{l-1} \times Channel_l \tag{5}$$

Feature map refers to the output feature map size calculated by each layer during the real-time operation of the model, as shown in Equation (6).

$$Map = \sum_{l=1}^{Dep} Map_{l,wide} \times Map_{l,height} \cdot Channel_l \tag{6}$$

The parameter is limited by multiple factors. In terms of the specific content, it is not only related to the number of convolution cores, but also can change the parameter, but it is not closely related to the size of the input data in the training.

In addition, some layers (such as the activation layer ReLU) can actually be completed by in-situ operation, so there is no need to output the feature map statistically. According to the improved C3D network structure, the time complexity of the two networks is calculated.

Through calculation and experiments, the data shown in Table 3 are obtained. The comparison results show that the model complexity of the improved network is obviously reduced, and the total training time is reduced by 33.6%.

5. Conclusion

It is one of the hottest research topics to identify the content analysis of volleyball movements in college physical education courses by video. On the basis of reading a large number of papers at home and abroad, this paper sorts out the research methods of human behavior recognition technology. Inspired by the above research, based on C3D model, in order to more accurately identify volleyball movements in college sports courses, the accuracy of training results is optimized through small convolution core, especially at the input data level, and found that a better data format can be determined through experiments. According to the analysis of network structure and experimental results, the recognition performance of the network has been improved, and the complexity of the network structure has been reduced to a certain extent, that is, the algorithm proposed in this paper can more effectively identify volleyball movements in college physical education courses. With the development of technology, scholars put forward solutions and improvement methods from different angles. Although the results have been improved, there are still many problems to be solved and optimized in the research of video-based behavior recognition. The follow-up research will be based on weak supervision timing action detection method, which can reduce the labor and time cost of timeline labeling, but also increase the difficulty of timing detection. It has achieved good results in detection and recognition, and video timing detection based on weak supervision has certain research value.

This research also has some limitations that need to be acknowledged. Firstly, despite the improvements in deep learning algorithms and dataset construction in this study, it is still challenging to avoid the limitations of the dataset. The size and diversity of the dataset can potentially impact the performance of the algorithm, necessitating larger and more diverse datasets to validate and improve the algorithm's effectiveness. Secondly, the model used in this research is specific to volleyball actions and may not directly apply to other sports or motion scenarios. Adjustments and training of the model may be required for different sports projects. Additionally, this study focuses solely on action recognition in volleyball and does not consider more advanced action analysis and decision-making processes. For more complex behavior understanding and decision problems, additional algorithms and methods may need to be introduced. Lastly, the experimental results of this research are primarily based on data from simulated environments or controlled experimental conditions, while noise, lighting variations, and other factors in real-world scenarios may affect the algorithm's robustness. Therefore, future research can involve validating and optimizing the algorithm in real-world environments. In conclusion, despite making some progress, this research still has limitations and areas for improvement, requiring further research and experimentation to address these issues.

Funding statement

No funding was received.

Ethics statement

The studies titled "Volleyball Movement Recognition Method in College Physical Education Curriculum Based on Convolution Neural Network" involving human participants were reviewed and approved by Xingtai Medical College. The patients/participants provided their written informed consent to participate in this study. Participants were assured of confidentiality and anonymity of the information relating to the survey.

Table 3
Complex comparison of network models.

Network model	Model parameter quantity	Time complexity	Training duration min/epoch	Total training time
C3D	11.6 × 106	11.1 m	92	16H
After improvement	7.7 × 106	7.4 m	61.3	10H

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

Acknowledgments

Not applicable.

References

[1] M. Boukens, A. Boukabou, M. Chadli, et al., Robust adaptive neural network-based trajectory tracking control approach for nonholonomic electrically driven mobile robots, Robot. Autonom. Syst. (2017).
[2] C. Long, E. Smith, A. Basharat, et al., A C 3D-Based Constructive Neural Network for Frame Dropping Detection in a Single Video Shot, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, 2017.
[3] H. Fan, X. Niu, L. Qiang, et al., F-C3D: FPGA-Based 3-dimensional Universal Neural Network, in: International Conference on Field Programmable Logic and Applications, 2017.
[4] Z. Lu, X. Xia, H. Wu, et al., Violence detection with two-stream neural network based on C3D, Int. J. Cognit. Inf. Nat. Intell. (2021) 15.
[5] X.D. Liao, X.X. Jia, E.S. Department, Action Recognition Technology Based on Improved C3D Neural Network, in: Computer and Modernization, 2019.
[6] M. Xu, J. Chen, H. Wang, et al., C3DVQA: Full-Reference Video Quality Assessment with 3D Conventional Neural Network, in: ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019.
[7] B.A. Er, M.S. Odabas, N. Senyer, et al., Evaluation of deep sea discharge systems efficiency in the eastern black sea using artificial neural network: a case study for Trabzon, Turkey, Braz. Arch. Biol. Technol. (2022) 65.
[8] M. Boukens, A. Boukabou, M. Chadli, Robust adaptive neural network-based trajectory tracking control approach for nonholonomic electrically driven mobile robots, Robot. Autonom. Syst. (2017).
[9] Y. Ren, C. Chen, S. Li, et al., Context-assisted 3D (C3D) object detection from RGB-D images, J. Vis. Commun. Image Represent. 55 (AUG) (2018) 131–141.
[10] B. Liang, Y. Tian, X. Chen, et al., Prediction of radiation pneumonitis with dose distribution: a Constructive Neural Network (CNN) based model, Front. Oncol. 9 (2020) 1500.
[11] Y. Li, Q. Miao, K. Tian, et al., Large-scale gesture recognition with a fusion of RGB-D data based on optical flow and the C3D model, Pattern Recogn. Lett. 119 (MAR.) (2019) 187–194.
[12] J. Yang, F. Wang, J. Yang, A review of action recognition based on conventional neural network, J. Phys. Conf. 1827 (1) (2021), 012138, 10pp.
[13] M.I. Fanany, A. Arinaldi, End-to-End multi-resolution 3D capsule network for people action detection, Int. J. Pattern Recogn. Artif. Intell. 36 (8) (2022), 2255015.
[14] L. Yu, K. Saida, S. Hirano, et al., Application of neural network-based hardness prediction method to HAZ of A533B steel produced by laser temper bead welding, Weld. World 61 (2017) 483–498.
[15] I. Ben-Bassat, B. Chor, Y. Orenstein, A deep neural network approach for learning intrinsic protein-RNA binding preferences, Bioinformatics 34 (17) (2018) i638–i646.
[16] R. Liu, Z. Liu, S. Liu, Recognition of basketball player's shooting action based on the convolutional neural network, Sci. Program. (2021), 3045418.
[17] G. Wang, D. Li, S. Jia, Mix-hops Graph Convolutional Networks for Skeleton-Based Action Recognition, in: International Joint Conference on Neural Networks, IEEE, 2021.
[18] K. Xu, F. Ye, Q. Zhong, et al., Topology-aware Convolutional Neural Network for Efficient Skeleton-Based Action Recognition, 2021 arXiv:2112.04178 [cs.CV].
[19] J. Yang, F. Wang, J. Yang, A review of action recognition based on Convolutional Neural Network, J. Phys. Conf. 1827 (1) (2021), 012138, 10pp.
[20] Q. Wang, M. Wang, Aerobics action recognition algorithm based on three-dimensional convolutional neural network and multilabel classification, Sci. Program. (2021), 3058141.
[21] X. Li, W. Zhai, Y. Cao, A tri-attention enhanced graph convolutional network for skeleton-based action recognition, IET Comput. Vis. (2021) 15.
[22] Y.M. Seo, Y.S. Choi, Graph Convolutional Networks for Skeleton-Based Action Recognition with LSTM Using Tool-Information, in: SAC '21: the 36th ACM/SIGAPP Symposium on Applied Computing, ACM, 2021.
[23] H. Qiao, S. Liu, Q. Xu, et al., Two-stream Convolutional Neural Network for Video Action Recognition, 2021.
[24] N. Heidari, A. Iosifidis, On the Spatial Attention in Spatio-Temporal Graph Convolutional Networks for Skeleton-Based Human Action Recognition, in: International Joint Conference on Neural Networks, IEEE, 2021.
[25] Q. Xuan, F. Li, Y. Liu, et al., MV-C3D: a spatial correlated multi-view 3D conventional neural networks, IEEE Access 7 (2019) 92528–92538.
[26] H. Liu, S. Parajuli, J. Hostetler, et al., Dynamically Throttleable Neural Networks (TNN), in: Computer Science, 2020.
[27] Y. Miao, J. Han, Y. Gao, et al., ST-CNN: spatial-temporal conventional neural network for crowd counting in videos, Pattern Recogn. Lett. 125 (JUL.) (2019) 113–118.
[28] L. Wu, S. Zhang, M. Jian, et al., Two stage shot boundary detection via feature fusion and spatial-temporal constitutional neural networks, IEEE Access 7 (2019) 77268–77276.
[29] A.A. April, J.K. Basu, D. Bhattacharyya, et al., Use of Artistic Neural Network in Pattern Recognition, Springer Berlin Heidelberg, 2020.
[30] R. Xin, Z. Jiang, Y. Shao, Complex network classification with conventional neural network, J. Tsinghua Univ.: Natural Science Edition (English Edition) (4) (2020) 11.
[31] G. Jin, M. Wang, J. Zhang, et al., STGNN-TTE: travel time estimation via spatial-temporary graph neural network, Future Generat. Comput. Syst. 126 (2022) 70–81.
[32] H.R. Liu, Z. Zuo, P. Li, et al., Anti-noise performance of the pulse coupled neural network applied in division of Neutron and gamma-ray, Nucl. Technol.: English Edition 33 (6) (2022) 13.
[33] J.C. Jang, E.H. Sohn, K.H. Park, et al., Estimation of daily potential evapotranspiration in real-time from GK2A/AMI data using artistic neural network for the Korean Peninsula, Hydrology 8 (3) (2021) 129.
[34] H. Shan, R. Vimieiro, L.R. Borges, et al., Impact of loss functions on the performance of a deep neural network designed to restore low-dose digital mammography, Electrical Engineering and Systems Science (2021).