International Journal of Industrial Engineering, 30(2), 451-462, 2023

SPACE-TIME GRAPH-BASED CONVOLUTIONAL NEURAL NETWORKS OF STUDY ON MOVEMENT RECOGNITION OF FOOTBALL PLAYERS

Siming Tian[1], Bingcheng Yin[2], Lei Wang[3], and Li Sun[4,*]

[1]College of Humanities
Chongqing Metropolitan College of Science and Technology
Chongqing, China

[2]Department of Sport and Healthcare
Namseoul University
Cheonan, Republic of Korea

[3]Department of Physical Education
Tangshan Normal University
Tangshan, China

[4]Police Sports, Teaching Department
Hebei Vocational College for Correctional Police
Shijiazhuang, China
[*]Corresponding author's e-mail: shwh23@163.com

Behaviour recognition technology is an interdisciplinary technology, integrating many research achievements in computer vision, deep learning, pattern recognition and other fields. The key information of bone data on human behavior can not only accurately describe the motion posture of the human body in three-dimensional space, but also its rigid connection structure is robust to various external interference factors. However, the behavioral recognition algorithm is influenced by different factors such as background, light and environment, which is easy to lead to unstable recognition accuracy and limited application scenarios. To address this problem, in this paper, we propose a noise filtering algorithm based on data correlation and skeleton energy model filtering, construct a set of football player data sets, using the ST-GCN algorithm to train the skeleton characteristics of football players, and construct a behavior recognition system applied to football players. Finally, by comparing the accuracy of Deep LSTM, 2s-AGCN and the algorithm in this paper, the accuracy of TOP1 and TOP5 is 39.97% and 66.34%, respectively, which are significantly higher than the other two algorithms. It can realize the statistics of athletes and analyze the technical and tactical movements of players on the football field.

Keywords: Spatiotemporal Map; Convolutional Neural Network; Football Movement; Action Recognition.

(Received on September 26, 2022; Accepted on March 2, 2023)

1. INTRODUCTION

The behavioral recognition technique is to extract and analyze the three-step information of video images. The core technical idea is to detect the movement target first, and extract the action characteristics of the target second, and finally to classify the extracted action features to judge the human action category (Liu *et al.*, 2021a). In various application scenarios, RGB video is usually used as the information input of the behavior recognition system, which is increasingly convenient for the collection of RGB data with the continuous development of mobile device technology. The high-dimensional description of skeleton data as human pose information is usually composed of lines of key points and points of the human body. Ordinary RGB video data is susceptible to many unstable factors in the external environment, such as redundant information factors such as light, color and clothing type (Duan *et al.*, 2022). External redundancy factors have little influence on skeleton data because skeleton data is the numerical data that record the point position and connection weights between human joints. Moreover, the size of RGB data, increasing with resolution and image quality, will occupy a lot of storage and computing resources and affect the performance of the system (Yang *et al.*, 2021a). Therefore, the current study of the posture information based on the skeletal information in the time series is of great significance for the study of human behavior

DOI: 10.23055/ijietap.2023.30.2.8581 **ISSN 1943-670X**

recognition topics.

Human behavior recognition is an important research direction in the field of video content understanding. At present, the mainstream idea of behavior recognition is to use a convolutional neural network and deep learning technology to design a network model for classifying human behavior characteristics in video (Zhu, 2021). Behavioral understanding can be simply considered as a classification problem for time-varying data. The test sequence is matched to a pre-calibration reference sequence representing a typical behavior. Through the collation of most of the current behavior identification research literature: The study of research into the understanding of human beings generally follows several basic processes, such as feature extraction and motor representation, behavior recognition, high-level behavior and scene understanding. Among them, feature extraction and motion representation are the motion state or feature descriptions of targets extracted from video or image information, which often use the results of target detection and tracking as pre-information. Behavioral recognition is a sequence of information matching the motor representations and reference used to determine which behavioral model the current action is in; High-level behavior and scene understanding combine the scene information on behavior occurrence and related domain knowledge to identify complex behaviors and realize the understanding of events and scenes (Cao *et al.*, 2022; Huang *et al.*, 2022; Xu *et al.*, 2022; Zhang, 2021).

A computer understands behavior is to classify it into a certain class action. The action is often neither a verb nor a separate noun. It belongs to a combination of verbs and nouns, such as: playing basketball, playing badminton, fencing and other sports. So for nouns, the research focus is on extracting their special feature information. Feature extraction of verbs is performed in the time domain. The deep learning network designed by the researcher needs to consider both the functionality and compatibility of the two to complete the task of feature extraction (Lv *et al.*, 2021; Yang *et al.*, 2022; Zhang, 2019). The main difficulties of this problem are as follows: 1) The influence of background and other factors in the environment are the biggest difficulties in all fields of computer vision. There is mainly a diversity of perspectives; that is, the same action from different perspectives will get different two-dimensional images (Meng *et al.*, 2022; Huang, 2018; Li *et al.*, 2021; Zhang *et al.*, 2022). The mutual occlusion between people and the background also makes it difficult for computers to classify the extracted features of actions. At present, the problem of multiple vision and occlusion is solved at present (Wang *et al.*, 2019). 2) Data acquisition and annotation. These data are video data, and the location and time of each action in the video are uncertain while considering the different manifestations of the same action and the difference between different actions, that is, the diversity and comprehensiveness of the data (Li, 2018a; She *et al.*, 2022; Xu *et al.*, 2023; Zhou *et al.*, 2021a).

In this paper, the pose estimation of the characters in the scene is used to extract the bone information in the human body, eliminate the external interference factors, and then use the human bone information to conduct deep learning model training to complete the classification of human behavior. Since the robustness of the OpenPose algorithm is susceptible to external interference, a set of behavior recognition algorithms applied to football players is constructed. And verify the accuracy of the algorithm through experiments, and determine that the algorithm can be applied to the technical and tactical movements on the football field.

2. RESEARCH STATUS

In the study of human movement recognition, in the research results of sports athletes. With the implementation of related technologies in the display application scenarios, human behavior recognition technology and posture estimation technology gradually began to be applied in the intelligent sports analysis system (Meng *et al.*, 2021; Zhou *et al.*, 2018). Hoop Tracker is a smart basketball analysis system equipped with smart wearables. The Hoop Tracker watch has an acceleration sensor that can detect every shot action. The shot detector is made into patches, fixed to the basket box, and the basketball passes through the patch after each goal. Each time a pitcher shoots, the watch and the shot detector communicate in real-time to detect the distance of the pitcher from the basket and whether the ball hit (Li *et al.*, 2017). The numbers the system can present include CIC, 3-pointer, free throw scores, shooting percentage and goal points. Krossover is a motion analysis company based in New York City. Coaches can upload players' sports videos to the website, and the ball sports event analysis platform will make a complete and detailed analysis of every movement of the athletes. ShotTracker's smart system, called ShotTracker Team, has its core principle to install smart sensors on basketball and players' shoes. In addition, the basketball court is surrounded sensors around and at the top of the court, and the players are like sensing by sensors in a blind space. These sensors give feedback to the player and basketball in real time. Through intelligent devices, conduct real-time analysis according to the on-court moving data of the players and the ball on the court, and show the comparative information about the players in the form of data, including analyzing the number of shots, turnovers, assists, steals, dunk and other actions. These data can not only show the status of the players on the court but also provide tactical arrangements with data support (Tian *et al.*, 2021a; Zhao *et al.*, 2022).

In the action recognition algorithm based on graph convolutional neural network, in recent years, due to the development of graph neural network and the maturity of the pose estimation algorithm, the human bone topology map extracted by the pose estimation algorithm has rich spatial domain and time domain characteristics, and the data has strong robustness.

Therefore, many researchers have devoted themselves to the research work of action recognition algorithms based on graph convolutional neural network. We propose a spatiotemporal graph convolutional network model, which first uses the graph convolutional neural network in the field of action recognition, defines the bone sequence as a spatiotemporal graph, and proposes a spatiotemporal graph convolution model that can be used to extract the features of the spatiotemporal graph (Yan *et al.*, 2018). Proposed a temporal routing algorithm can adaptive learning the connection between bone points, using the encoder-decoder structure to obtain the implied joint correlation and high-order neighborhood information. In addition, by extending the skeleton diagram, the high-order dependence and potential dependence, respectively as structure link and action link, can compensate for the input in the convolution of the graph are fixed topology structure, the model has made great progress compared with the previous methods (Li *et al.*, 2019).

In conclusion, according to the above introduction of the research status in the field of motion recognition and motion assessment, there are still many problems: the existing public motion recognition data set and the existing motion recognition network based on skeleton data are difficult to learn the connection between joint points. This paper starts with the action recognition of football players and designs and realizes a motion recognition system based on human skeleton data.

3. HUMAN BEHAVIOR RECOGNITION ALGORITHM BASED ON SKELETON EXTRACTION

3.1 Common Processes of Behavior Recognition Algorithms

When the time series of features is long, the problem of gradient disappearance occurs during the traditional model training (Song, 2019). The long and short-term memory network structure improves the neuronal structure of the traditional recurrent neural network and realizes the memory of the forward input information on the time domain (Tian *et al.*, 2021b; Wang *et al.*, 2003). A certain behavioral action of a human being is often closely related to the set of specific joint points of the skeleton and the interaction of the nodes in this set. This property, where several joint points simultaneously influence and determine discrimination, is called a co-occurrence feature. LSTM structures are often used to describe the co-occurrence of joint points in human behavior recognition (Dai, 2020). An LSTM layer consists of several LSTM neurons that are divided into several groups. Each neuron in the same group shared larger connection weights with some joint points and smaller connection weights with the others (Li, 2019; Wu *et al.*, 2020).

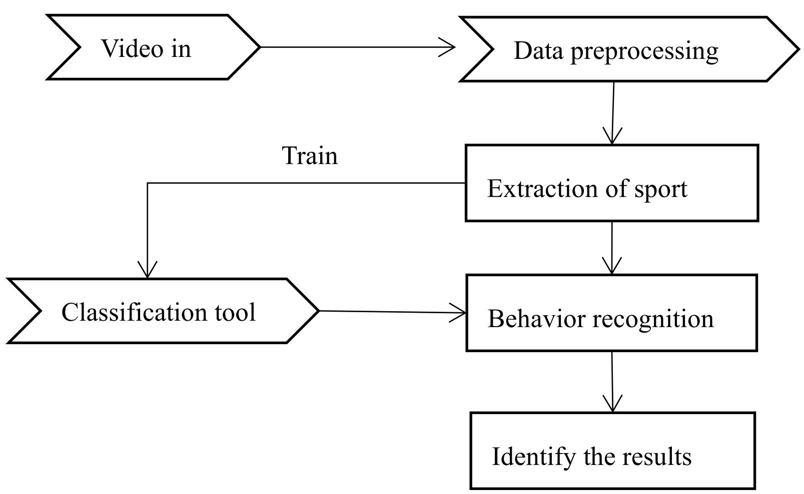


Figure 1. Common Processes of Behavior Recognition Algorithms

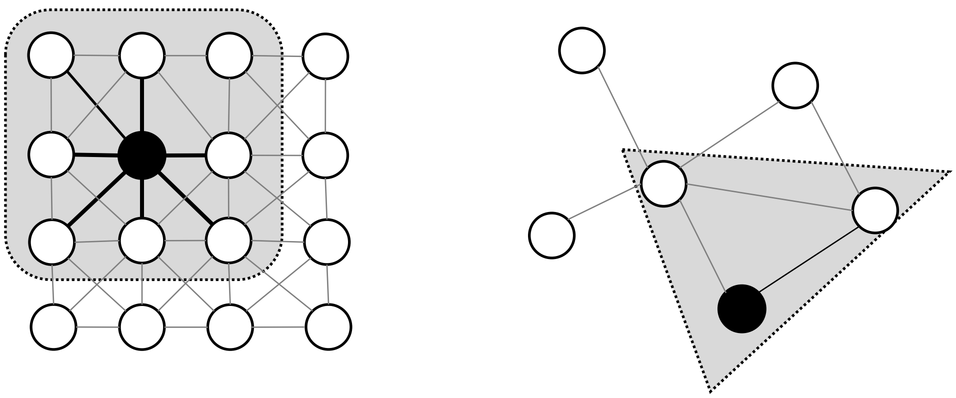
Figure 1 is the general flow of the current behavior recognition algorithm scheme. In order to make and stabilize the identification results of the algorithm efficient, this step of feature extraction of the moving target is crucial (Zhang *et al.*, 2019). Based on the pose estimation of targets in images or videos, the study finds that the information on the key points of the human skeleton has the advantages of simple structure and high robustness. The key points of the human skeleton are composed of rigid segments of a series of key points. In terms of data structure, each key point containing (x, y, score) 3-D information can be regarded as a graph structure (Li, 2018b). The graphical convolutional neural network is effective in performing the feature learning and classification work in the time domain for such graph structure data, including airspace information (Zhou *et al.*, 2021b). Therefore, this paper uses a space-time one-graph convolutional neural network for the learning and training of behavior recognition.

3.2 Time Domain Characteristics

3.2.1 Graph Convolutional Network

Convolutional neural networks are often used for image feature extraction (Han *et al.*, 2021). Since the image is a two-dimensional data structure, the main idea of CNN is to design a suitable convolution core to form a sliding window model on the image and then perform the convolution operation by repeated translation sliding window to extract the features in the image data (Liu, 2021). Where the process of sliding window convolution is based on the translation invariance of the structure space of the image, that is, a small window is exactly the same internal structure no matter which position it moves to the image, so the CNN can realize parameter sharing (Wang, 2018).

The objects distinguishing graph convolutional neural network from convolution operation are graphs and images, respectively. As shown in Figures 2a and 2b, Figure 2a can be seen as a very dense picture; Figure 2b is a normal picture. The shaded part represents a convolution kernel, with a conventional one on the left and a graph one on the right.



(a) (b)

Figure 2. Image and Graph Convolution

Through the above graph structure, we can find that the traditional convolution kernel cannot be directly used to extract the characteristics of the nodes on the graph because the neighbor nodes in the graph are not fixed. At present, the mainstream research solves this matter from two ways: one is to propose a way to transform the graph of non-European space into European space. The second is to find a controllable convolution kernel of variable long neighbor nodes to extract features on the graph. These two principles are also, the design principles of subsequent graph convolutional neural networks. The essence of graph convolution is to find learnable convolutional kernels applicable to graphs, as shown in Figure 3.

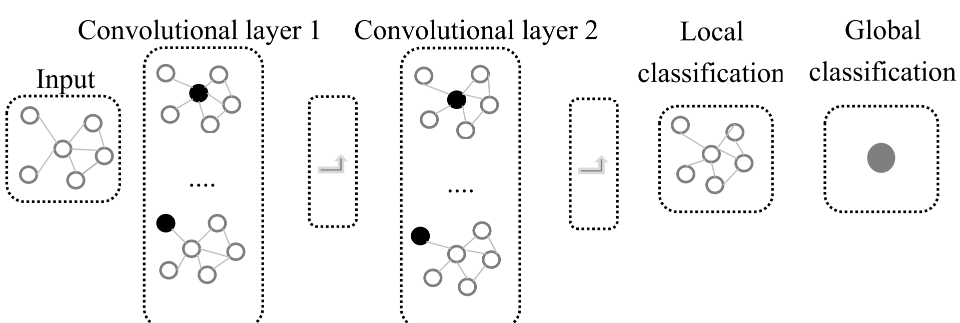


Figure 3. The Framework of the Graph Convolutional Neural Network

The input is the whole graph, in the convolution layer 1, the neighbor of each node performs a convolution operation, and the node is updated with the result of the convolution; Then, through the activation function, and then over a convolution layer 2 and a layer of activation function; Repeat this process until the number of layers reaches the desired depth. The graph convolutional neural networks also have a local output function for converting the state of the nodes into task-related labels. For the behavior recognition algorithm in this paper, the procedure belongs to a global classification task. The graph convolutional neural network mainly extracts the key points of the human skeleton. The space convolution method refers to the decomposition of a node into two processes: Messaging and status update operations, completed by the $M_t(\cdot)$and $U_t(\cdot)$. After taking the feature x_v of the junction v as the $h_v{}^0$ initial state of its hidden state, The update of airspace convolution shows the following equation:

$$h_v{}^{l+1} = U_{l+1}\left(h_v, \sum_{u \in ne[v]} M_{l+1}\left(h_v{}^l, h_u{}^l, x_{vu}\right)\right), \tag{1}$$

where l represents layer l of the graph convolution, the physical meaning of the above formula is that: How does each node update its own status after receiving a message M_{l+1} from each neighbor.

3.2.2 Attention Model

RNN and LSTM technologies in deep learning have been the mainstream way to solve sequence learning and sequence transformation since 2014. But the attention model saves more training and running resources. Figure 4 shows the network framework diagram. The time domain attention subnetwork learns a time domain attention model to assign appropriate importance to different frames and to fuse information from different frames based on this. The spacial attention subnetwork integrates different frame information for the human key points of the input network. The network employs a time-domain attention model to assign different weights to different keyframes. The spacial attention and the time-domain attention-specific values in the sequence are not in reference, which means that the true value needs to automatically acquire the attention to optimize the final classification performance. The continuous development of actions during the learning process causes the increase of the time domain attention weight and the increase in the importance of adjacent frames. Time-domain attention models will give greater attention weights to keyframes with higher importance. Similar to human perception, the network gives large spatial attention weights to different key points for different sets of behavioral actions.

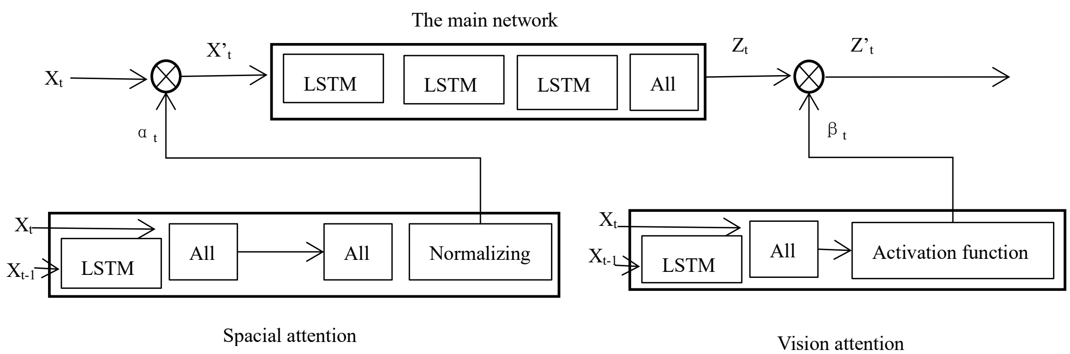


Figure 4. Space-Time Attention Network Structure

3.2.3 Behavioral Recognition Algorithm Based on Space-time Map Convolutional Neural Network

Dynamic human skeleton model with important information on action recognition, traditional methods often use manual features or ergodic rules to model skeletons, this limits the expressive ability, and it is difficult to de-generalize. On the basis of the investigation of domestic and foreign research, the dynamic skeleton model ST-GCN was used. It can automatically learn the modes of space and time from the data, which gives the model strong expressive and generalization capabilities. The algorithm designs a general representation of the skeleton sequence for action recognition, called a spatiotemporal graph convolutional network. The model is built in the sequence composed of the skeleton diagram, where each point corresponds to each node in the structure of the human skeleton.

As mentioned above, the network of behavior recognition algorithm mainly uses the airspace convolutional properties of the graph convolutional neural networks to directly apply the convolutional kernels to the graph nodes and their neighbor nodes, implicitly combining positional information and temporal dynamic information through the graph convolutional network. Given the skeleton sequence information of an action video, the graph structure representing the skeleton sequence information is first constructed. The input of ST-GCN is the joint coordinate vector on the graph node, and then a series of spatiotemporal graph convolutional operations to extract the features of the top level, and finally, the SofMax classifier is used to obtain the corresponding action classification. The whole process achieves end-to-end training.

The specific implementation steps of ST-GCN are divided into the following parts. First, remember the spatiotemporal map of a bone sequence with N nodes and T frames as $G = (V, E)$, the set of nodes is $V = v_{ti}|t = 1,,,T, i + 1, \dots, N$, the set of nodes is $F(v_{ti})$, it consists of the coordinate vector and the estimated confidence of the node. The graph structure consists of two parts: one is connecting the nodes of each frame into edges according to the body structure of the football player, which forms spatial edges, and two is connecting the same nodes in two consecutive frames into edges, which form temporal edges. Taking the common 2D convolution of an image as an example, the convolution output for a certain position x can be written as follows:

$$f_u t(x) = \sum_{h=1}^{K} \sum_{w=1}^{K} f_{in}(p(x, h, w) \cdot w(h, w) \ . \tag{2}$$

Characteristic plot f_{in} with the number of input channels as c, convolutional kernel size $K \times K$, sampling function $p(x, h, w)$, and weighting function. The number of channels is c. In the figure, the set of neighborhood pixels is defined as:

$$B(v_{ti}) = v_{ti} \big| d(v_{tj}, v_{ti}) \le D \ , \tag{3}$$

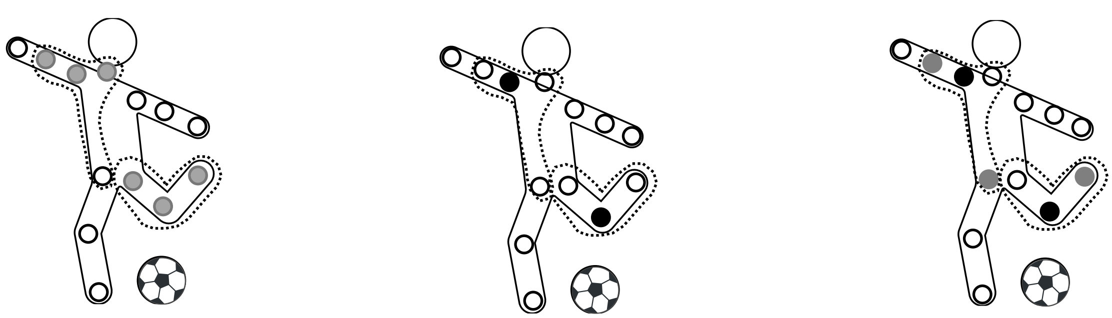
where $d(v_{tj}, v_{ti})$ refers to the shortest distance from v_{tj} to v_{ti}, the sampling function can be written as $p(v_{tj}, v_{ti})$. In the 2D convolution, the neighbor pixels are regularly arranged around the central pixels so that the convolution operations can be checked with regular convolutions according to the spatial order. By 2D convolution, the neighbor pixels obtained by the sampling function in the graph are divided into different subsets. Each subset has a number label, and a neighbor node is mapped to the corresponding subset label. The weight equation is:

$$\omega\big((v_{tj}, v_{ti})\big) = \omega'\big(l_{ti}(v_{tj})\big) \ . \tag{4}$$

After the above derivation, the expression of the spatial graph convolution is:

$$f_{out}(v_{ti}) = \sum_{v_{tj} \in B(v_{ti})} \frac{1}{z_{ti}(v_{tj})} f_{in}(p(v_{ti}, v_{tj})) \cdot \omega(v_{ti}, v_{tj}) \ . \tag{5}$$

After passing through the airspace convolution, the key points are divided to design the convolution kernel. There are three ways to divide subsets: (1) Unique structure division: divide the 1 neighborhood of nodes into a subset, as shown in Figure 5(a). (2) Distance-based division: the 1 neighborhood of nodes is divided into two subsets. The subset of nodes themselves and the subset of neighboring nodes are shown in Figure 5(b). (3) The spatial configuration division: node 1 neighborhood is divided into three subsets, the first subset connects the spatial location than the root node away from the skeleton neighbor node, the second subset connects the closer to the center of the neighbor node, the third subset for the root node itself, respectively represents the centrifugal movement, centripetal movement and static movement characteristics, as shown in Figure 5(c).



(a) (b) (c)

Figure 5. Subset Partitioning of the ST-GCN

Finally, add the attention mechanism. The importance of different torso varies among football players. For example, leg and hand movements may be more important than the neck, and the legs can even tell football players as running, walking and jumping, but neck movements often contain less effective information. Thus, the ST-GCN weighted the different torsos.

Each ST-GCN unit also has its own weight parameters for training. The overall process of the algorithm constructed in this paper is as follows: the input data is first batch normalized, then passed through 9 ST-GCN units, followed by a pooling layer to obtain the 256-dimensional feature vector of each sequence, and finally classified with SoftMax function to obtain the final label.

4. EXPERIMENT AND APPLICATION

This chapter focuses on the motion identification of football players based on convolutional neural networks. In this paper, the space-time graph convolutional neural network is used to identify the players in the football scenes, which includes the training process and performance evaluation of this network model in the standard database. In this paper, we study the original network for classifying the features under some complex motion and scene by modifying the network parameter size and integrating the skeleton information and the confidence scores of the voting judgment. This paper tests the program robustness of the procedure to the cross-perspective and partial misdetection cases when the model runs correctly.

4.1 Training Model Evaluation

The proposed algorithm is mainly based on behavior recognition techniques for skeleton extraction and pose estimation. In this paper, experiments on video data for the ST-GCN network. First, the model training experiment is conducted, and the training experimental configuration parameters are shown in Table 1.

Table 1. ST-GCN Training Experimental Parameters

Training Parameters	Parameter Values
Training	32
Test dataset	32
Learning rate	0.1
iterations	50
GPU	2*T100

The training test set uses the Kinetics-skeleton skeleton behavior dataset. This dataset uses the OpenPose algorithm to extract 2D skeleton sequences in the Kinect-400 dataset and then uses the confidence as the third-dimensional data of the z-axis so as to obtain the 3D-like skeleton sequences extracted later. The training data consist rom about 240,000 video clips covering 400 soccer players' movements, with at least 600 video clips in each action category. Each clip lasts about 10 seconds and is marked with a separate category. The cross-validation set contains 20,000 short video behavior information. All clips were subjected to multiple rounds of manual annotation, each coming from a unique YouTube video. These moves cover a wide range of footballer movements. In this paper, the model is trained using a deep learning server deployed environment, with a total of 60 iterations trained on the training set. Data from the cross-validation sets were evaluated at 10 intervals, and the probability of showing true results in the TOP 1 and TOP5 classification is shown in Table 2. To test whether the training results in different perspectives were generalized, the 37,920 videos from 2 perspective cameras and the 18,960 videos from the third Kinect v2 camera were trained as a validation test set.

Table 2. Training Results for the Kinect-600 Dataset

Iterations	Average Loss	Top1 Precision	Top5 Precision
10	4.502641	11.18%	29.35%
20	4.426153	16.84%	33.89%
30	3.765894	25.97%	45.24%
40	3.452671	28.08%	49.98%
50	3.103489	28.66%	50.66%
60	3.354687	28.93%	52.03%

4.2 Action Recognition Results of Football Players

In this paper, video acquisition data from real scenes and official public standard behavior identification data are used to conduct performance tests and result analysis. The results of action recognition are displayed in the video file output, it consists of four parts: Original Video, Pose Estimation, Attention+Predictioin, and Attention+RGB. The algorithm first uses the OpenPose algorithm to conduct pose estimation and skeleton extraction on the input video data and then inputs the ST-GCN network to do behavior classification. After adding the attention model mechanism, it outputs the current action category in real-time and finally votes on the action category of the whole video.

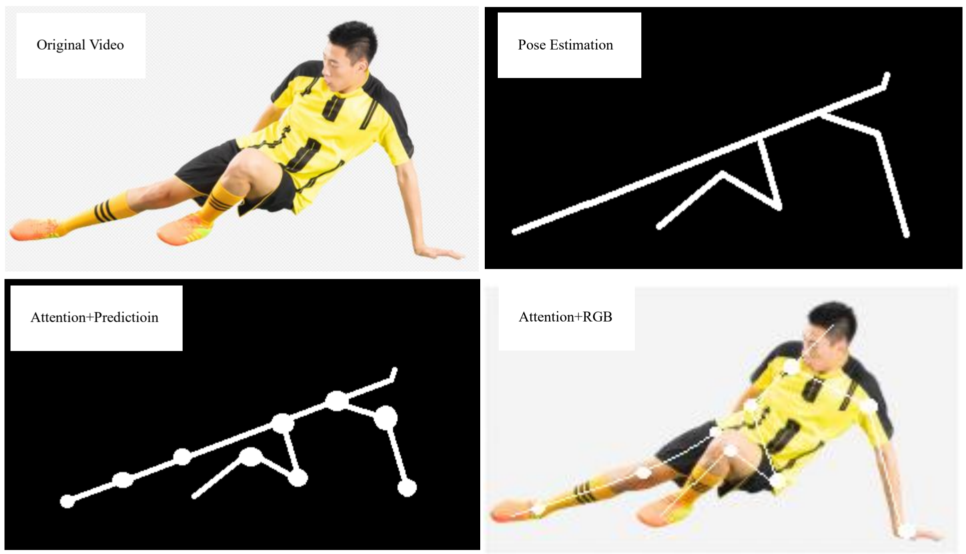


Figure 6. Recognition Results of Football Players

As shown in Figure 6, the test data are the football player tackle movements in the data taken in this paper. The upper left picture shows the original video data. The top right panel shows the skeleton extraction results. The elbow and waist joint points in the human skeleton in the bottom left figure are more weighted by the attention model because they are the key joint points in this movement. The classification result of the video-added attention model is pushup, and it is identified correctly.

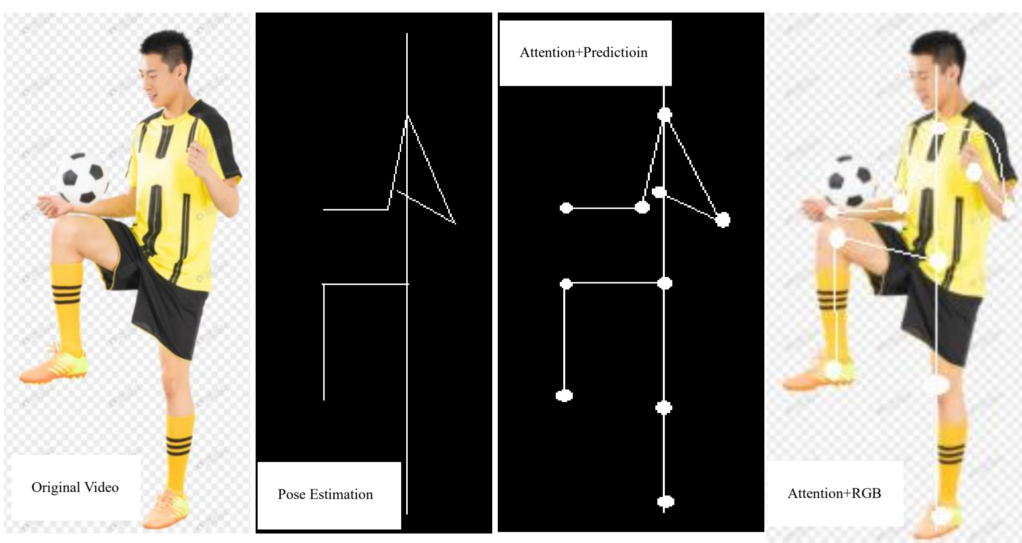


Figure 7. Football Players Recognize the Ball

As shown in Figure 7, the test data are for the football players in the data taken in this paper. The left is the original video data. The middle two plots are the skeleton extraction results and the leg joint points in the human skeleton, respectively, because the key joint points are added weighted by the attention model in this action. The classification result of video with attention model is: ride a bike, correct identification.

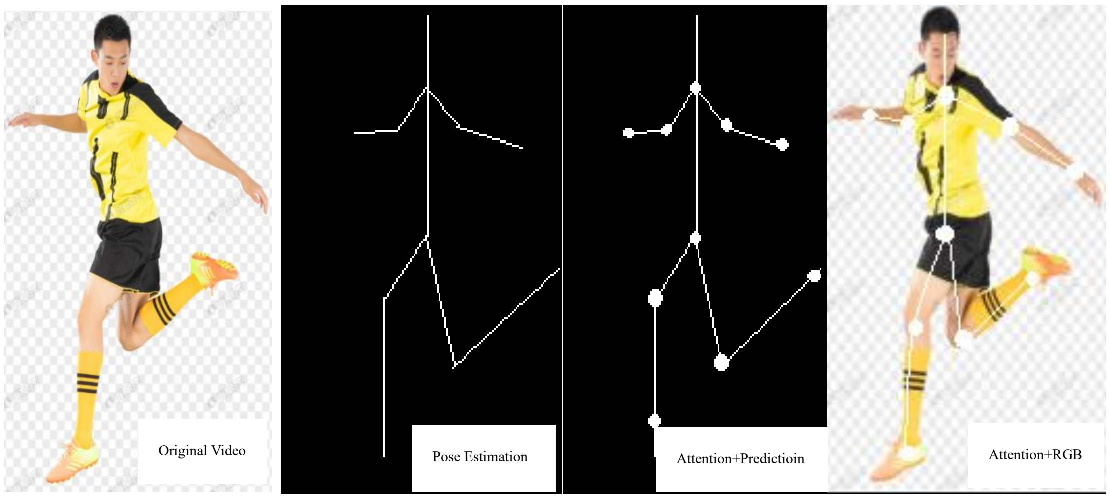


Figure 8. Running Identification Results

As shown in Figure 8, the test data are the running movements in this captured data. The hand and leg joint points in the human skeleton on the left are further weighted by the attention model because they are the key joint points in this movement. The classification result of the video-added attention model is jogging, and it is identified correctly. The experimental performance analysis of Kinect-skeleton behavior action recognition resolution compares with the current typical. The experimental results of the skeleton-based behavior recognition algorithm are shown in Table 3.

Table 3. Performance Assessment of the Behavior Recognition Algorithm

Algorithm Type	TOP1 Precision	TOP5 Precision
DeepLSTM	16.39%	35.30%
2s-AGCN	35.59%	58.72%
The algorithm in this paper	39.97%	66.34%

The 2s-AGCN algorithm is an adaptive behavior recognition algorithm based on skeleton information based on ST-GCN. The algorithm was identified with 35.59% accuracy in TOP1 and 58.72% in TOP5. The LSTM algorithm was identified with an accuracy of 16.39% in TOP1 and 35.30% in TOP5. The comparative performance analysis shows that the performance of the ST-GCN algorithm is still relatively high compared with the theoretical algorithm proposed by the researchers. Compared to the classical LSTM-based algorithm, etc. The accuracy of the algorithm constructed in this paper is 39.97% and 66.34%, respectively, which is higher than the other two algorithms, which shows that the algorithm constructed in this paper is more effective.

5. CONCLUSION

For human behavior recognition algorithm in the background, light, environment, and other diversity factors lead to unstable identification accuracy and application scenarios, this paper adopts the estimation of the scene, extract the skeleton information of the human body, eliminate external interference factors, then use human skeleton information deep learning model training, completed the classification of human behavior. In this paper, the bottom-up detection idea is used for human skeleton detection and posture estimation. Since the robustness of the OpenPose algorithm is susceptible to external interference, this paper proposes a noise-noising method based on data correlation and skeleton energy model filtering. This paper constructs a set of football data set based on this algorithm, uses the ST-GCN algorithm to train the characteristics of the human skeleton, builds a set of behavior recognition systems applied to football scenes, and realizes the statistics and analysis of the technical and tactical movements of football players in football scenes. This paper uses a space-time graph convolutional neural network to identify moving bodies in different scenarios. The algorithm is robust on multi-perspective human behavior recognition tasks by testing on shot video and standard datasets. Through training experiments and model performance evaluation on different data sets, the results show that the proposed algorithm is more accurate than the other two algorithm theories.

The research limitation of this paper is that the algorithm currently only identifies football actions and proves the effectiveness of the study, but the identification of other actions still needs more experiments to verify. In addition, the prospect of the follow-up work includes: The modification of this problem has been explained in the conclusion section of the article. Specifically, the performance of the OpenPose bottom-up pose estimation algorithm for skeleton detection still needs to be improved, and the network is needed to achieve full real-time high-resolution video speed. Current ST-GCN-based behavior-recognition algorithm, the training target is still single or two-person behavior-recognition. For football, such as a team sport, many techniques and tactics need to be completed by multiple players. This paper may have insufficient computational accuracy, so increasing the number of training samples can be considered in subsequent studies to train a behavioral recognition model with more targets. The behavioral recognition system of football sports can extract other information rather than bone information to improve performance, such as a decent person to the other side or us, the player's position, and so on. Follow-up studies could consider adding the capture of additional information to improve behavioral identification accuracy in this setting.

REFERENCES

Cao, B., Zhao, J., Liu, X., Arabas, J., Tanveer, M., Singh, A.K., and Lv, Z. (2022). Multiobjective Evolution of the Explainable Fuzzy Rough Neural Network with Gene Expression Programming. *IEEE Transactions on Fuzzy Systems,* 1. DOI: https://dor.org/10.1109/TFUZZ.2022.3141761

Dai, D. (2020). A Method Based on the Action Behavior Recognition of the Neural Network. CN110781847A.

Duan, H., Zhao, Y., Chen, K., Lin, D., and Dai, B. (2022). Revisiting skeleton-based action recognition. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, LA, USA.

Han, K.L., Zhang, Y.X., and Chen, X.Z. (2021). Action Recognition and Evaluation of A Gymnast from A Deep Learning View. *Modern Scientific Instruments,* (2): 234-238.

Huang, C., Jiang, F., Huang, Q., Wang, X., Han, Z0,. and Huang, W. (2022). Dual-Graph Attention Convolution Network for 3-D Point Cloud Classification. *IEEE Transactions on Neural Networks and Learning Systems*. DOI: 10.1109/TNNLS.2022.3162301

Huang, J. (2018). Application Study of — Based on Artificial Neural Network and Basketball Player Training Strategy. *Agricultural mechanization Research,* 40(7): 5.

Li B., Li, X., Zhang, Z.F., and Wu, F. (2019). Spatio-temporal Graph Routing for Skeleton-based Action Recognition. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(1): 8561-8568.

Li, D., Ge, S.S., and Lee, T.H. (2021). Simultaneous-Arrival-to-Origin Convergence: Sliding-Mode Control through the Norm-Normalized Sign Function. *IEEE Transactions on Automatic Control,* 67(4): 1966-1972.

Li, H. (2018b). Construction and Simulation Study of Athlete Injury Early Warning Model Based on Attribute Reduction Algorithm. *Automation & Instrumentation,* (9): 24-27.

Li, J., Xu, K., Chaudhuri, S., Yumer, E., Zhang, H., and Guibas, L. (2017). GRASS: Generative Recursive Autoencoders for Shape Structures. *ACM Transactions on Graphics,* 36(4): 1-14.

Li, S. (2018a). Design of Athletes' Special Performance Prediction System Based on Hybrid Genetic Neural Network. *Modern Electronic Technology,* 41(8): 4.

Li, S. (2019). Research on Human Action Recognition Based on Convolutional Neural Network. Master's Thesis. University of Electronic Science and Technology.

Li, Z., Peng, X., Hu, G., Zhang, D., Xu, Z., Peng, Y., and Xie, S. (2022). Towards Real-Time Self-Powered Sensing with

Ample Redundant Charges by A Piezostack-Based Frequency-Converted Generator from Human Motions. *Energy Conversion and Management,* 258: 115466.

Liu, J., Shi, M., Chen, Q., Fu, H., and Tai, C.-L. (2021a). Normalized human pose features for human action video alignment. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, QC, Canada.

Liu, L. (2021). Automatic Motion Detection System for Aerobics Athletes Based on Pose Estimation. *Automation and Instrumentation,* (3): 119-122.

Lv, Z., Li, Y., Feng, H., and Lv, H. (2021). Deep Learning for Security in Digital Twins of Cooperative Intelligent Transportation Systems. *IEEE Transactions on Intelligent Transportation Systems,* p. 1-10. DOI: https://doi.org/10.1109/TITS.2021.3113779

Meng, F., Cheng, W., and Wang, J. (2021). Semi-supervised Software Defect Prediction Model Based on Tri-training. *KSII Transactions on Internet and Information Systems,* 15(11): 4028-4042.

Meng, F., Xiao, X., and Wang, J. (2022). Rating the Crisis of Online Public Opinion Using a Multi-Level Index System. *The International Arab Journal of Information Technology*, 19(4): 597-608.

She, Q., Hu, R., Xu, J., Liu, M., Xu, K., and Huang, H. (2022). Learning High-DOF Reaching-and-Grasping via Dynamic Representation of Gripper-Object Interaction. *ACM Transactions on Graphics*, 41(4): 97.

Song, Y. (2019). An Athlete-Assisted Training Data Acquisition Method, Device and Electronic Equipment. CN109635925A.

Tian, H., Qin, Y., Niu, Z., Wang, L., and Ge, S. (2021a). Summer Maize Mapping by Compositing Time Series Sentinel-1A Imagery Based on Crop Growth Cycles. *Journal of the Indian Society of Remote Sensing,* 49(11): 2863-2874.

Tian, H., Wang, Y., Chen, T., Zhang, L., and Qin, Y. (2021b). Early-Season Mapping of Winter Crops Using Sentinel-2 Optical Imagery. *Remote sensing (Basel, Switzerland),* 13(19): 3822.

Wang, L.J., Wang, Y.T., Gong, M.X., Ma, A.D., Zou, N.X., and Niu, W.X. (2019). Construction and application of the competitive ability structure evaluation model of trampoline athletes based on artificial neural network. *Abstract compilation of the 11th National Sports Science Conference,* Nanjing, China.

Wang, L.L., Wang, S.X., Sun, X.Y., and Hu, F.Y. (2003). Beam Space-Time Block Coding for Wireless Communications. *Journal of Jilin University (Engineering and Technology Edition),* (4): 63-67.

Wang, X. (2018). Assessment of Skill Level in Badminton Players. *Computers and Digital Engineering,* 46(7): 1311.

Wu, Z., Cao, J., Wang, Y., Wang, Y., Zhang, L., and Wu, J. (2020). hPSD: A Hybrid PU-Learning-Based Spammer Detection Model for Product Reviews. *IEEE Transactions on Cybernetics,* 50(4): 1595-1606.

Xu, J.W., Pan, S.C., Sun, P.Z.H., Park, S.H., and Guo, K. (2023). Human-Factors-in-Driving-Loop: Driver Identification and Verification via a Deep Learning Approach using Psychological Behavioral Data. *IEEE Transactions on Intelligent Transportation Systems*, 2022, 24(3): 3383-3394.

Xu, S., He, Q., Tao, S., Chen, H., Chai, Y., and Zheng, W. (2022). Pig Face Recognition Based on Trapezoid Normalized Pixel Difference Feature and Trimmed Mean Attention Mechanism. *IEEE Transactions on Instrumentation and Measurement*, 72: 1-13.

Yan, S., Xiong Y., and Lin D. (2018). Spatial Temporal Graph Convolutional Networks for Skeleton-based Action Recognition. *Thirty-second AAAI conference on Artificial Intelligence*, 22: 625.

Yang, S., Tan, J., and Chen, B. (2022). Robust Spike-Based Continual Meta-Learning Improved by Restricted Minimum Error Entropy Criterion. *Entropy,* 24(4):455.

Yang, Y., Ren, Z., Li, H.X., Zhou, C.L., Wang, X.C., and Hua, G. (2021a). Learning dynamics via graph neural networks for human pose estimation and tracking. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, TN, USA.

Zhang, G. (2021). *A Method of Badminton Players Combining Multimodal Feature Analysis and Neural Network*. CN112396018A.

Zhang, J., Tang, Y., Wang, H., and Xu, K. (2022). ASRO-DIO: Active Subspace Random Optimization Based Depth Inertial Odometry. *IEEE Transactions on Robotics*. DOI: 10.1109/TRO.2022.3208503

Zhang, Y. (2019). Athletes' Cardiac Function Analysis Model Using the BP Neural Network. *Journal of Ningde Normal University: Natural Science Edition,* 31(4): 5.

Zhang, Y.S., Zhang, L.M., and Jiang, S.Y. (2019). Name Nationality Identification Based on Character-Level Truncated Recurrent Neural Network. *Pattern Recognition and Artificial Intelligence,* 32(4): 369-375.

Zhao, H., Zhu, C., Xu, X., Huang, H., and Xu, K. (2022). Learning Practically Feasible Policies for Online 3D Bin Packing. *Science China Information sciences,* 65: 112105.

Zhou, W., Liu, J., Lei, J., Yu, L., and Hwang, J. (2021a). GMNet: Graded-Feature Multilabel-Learning Network for RGB-Thermal Urban Scene Semantic Segmentation. *IEEE Transactions on Image Processing*, 30: 7790-7802.

Zhou, W., Yu, L., Zhou, Y., Qiu, W., Wu, M., and Luo, T. (2018). Local and Global Feature Learning for Blind Quality Evaluation of Screen Content and Natural Scene Images. *IEEE Transactions on Image Processing*, 27(5): 2086-2095. DOI: 10.1109/TIP.2018.2794207

Zhou, Y.L., Li, H.J., Yao, T.Q., and Sun, K.Y. (2021b). Prediction of Lateral Shear Direction Dynamics in Football Players Based on Neural Networks. *Journal of Medical Biomechanics,* 36(S1): 148.

Zhu, X. (2021). Neural Network Prediction Model of Traditional Technical Action Speed in Cross-Country Skiers. *Journal of Beijing Sport University,* 44(12): 11.