scientific reports

Check for updates

OPEN

Analysis of baseball behavior recognition model based on Dual-GCN improved by motion weights

Ji Li

This research aims to address the poor performance in baseball behavior recognition, insufficient connection between characters, and low accuracy in baseball behavior recognition. A motion weight improvement model based on dual-graph convolutional network is proposed. The new model takes a dual-graph convolutional network for behavior recognition and key region segmentation of baseball video images, and enhances the correlation and contribution between characters through motion weights. The research results indicated that the new model performed best when the image frame rate was 1/2w, the width of the key area was 1/2H, and the number of key areas was 2. The highest accuracy was 94.84%, which was 12.06% higher than that of the hierarchical temporal depth model. After adding motion weights, the accuracy improved by 3.45%. The accuracy of baseball recognition using different models has been effectively improved. The new model can more effectively recognize baseball behavior, which has important guiding significance for behavior recognition in baseball sports.

Keywords Motion weight, Dual-GCN, Baseball sport, Behavior recognition, Character

It is of great significance to accurately identify and analyze motion behaviors in professional baseball games[1]. This can not only improve the efficiency of athlete training and optimize tactical layout, but also quantitatively evaluate athlete performance, prevent sports injuries, and provide rehabilitation training[2]. With the development of computer vision and artificial intelligence technology, more network models are demonstrating powerful data processing capabilities in graph structured data, making it possible to automatically recognize and analyze motion behaviors from video image data[3]. However, existing motion behavior recognition models still have problems such as insufficient recognition accuracy and poor recognition performance in dealing with complex motion and multi-character scenes. Graph Convolutional Networks (GCNs) have made significant progress in processing graph structured data, particularly in fields such as social network analysis, knowledge graphs, and bio-informatics[4]. GCN provides powerful representation capabilities for graph data by effectively capturing complex relationships between nodes. In sports analysis, GCN can extract athletes' motion patterns and behaviors from video data.

In previous studies, many studies have taken GCN for motion behavior recognition analysis. Li Z et al. designed a recognition model based on NAM-STGCN, which combined normalized attention mechanism and ReLU optimization to overcome existing models neglecting spatial information and insufficient aggregation of skeleton key point features in worker fall recognition. The model had a recognition accuracy of 96.72% on self-built datasets. It also performed well on standard datasets, which could accurately identify abnormal behaviors[5]. There is still a significant research gap in group feature recognition and skeleton point analysis by combining attention mechanisms and effectively improving model recognition performance. Hu K et al. conducted in-depth analysis on behavior recognition algorithm based on deep learning, focusing on image and skeletal data to improve the accuracy and robustness of human behavior recognition. The research results indicated that the algorithm was significantly superior to traditional methods, demonstrating strong recognition capabilities and broad application potential[6]. Different improvement methods can improve the accuracy of traditional recognition, but there are still significant problems in the research on improving the effectiveness of specific models. Lovanshi M et al. proposed a customized ST-GCN model to effectively utilize 3D skeleton depth information to improve the accuracy of human activity recognition, focusing on fine spatial and temporal features. The model performed better than state-of-the-art methods on large datasets such as NTU-RGB-D, Kinetics Skeleton, and Florence 3D, with significant improvements in Top-1 and Top-5 accuracy[7]. 3D technology can effectively improve the accuracy of behavior recognition, but existing research has not yet analyzed the relationship between multi-target features and skeleton points. Hasan M B et al. designed a multi-scale graph convolution method called "Gait GCN++" to overcome over-fitting of spatial and temporal maps and incomplete

School of Sports Training, Jilin Sport University, Changchun 130022, China. email: 0274@jlsu.edu.cn

long-term relationships between joints in gait recognition. The research results indicated that the Gait GCN + + performed better than common techniques in diverse and challenging scenes on the CASIA-B and GREW benchmark datasets, displaying significant advantages in improving gait recognition accuracy and robustness[8]. The current research shows good results in recognizing character action spaces and joint capture, but there are still significant issues in recognizing different character behaviors. Myung W et al. proposed a deformable GCN to overcome the insufficient intra class diversity and redundant connections in traditional GCN for skeleton sequence processing in skeleton action recognition. This network adaptively learned variable sampling positions on spatial and temporal graphs, effectively capturing the joints with the highest information content and perceiving different receptive fields. Meanwhile, considering the continuity of human behavior, temporal features in a continuous latent space were defined. The new network achieved better performance on multiple datasets, verifying the effectiveness[9]. Character skeleton point recognition and analysis can also improve the performance of models in behavior recognition, but existing research still needs to analyze multi-character recognition in motion scenes. Ma L et al. proposed a method based on CS-Transformer, CLIP, and transfer learning techniques for ball motion recognition, which could effectively process the visual information of different ball sports scenes to accurately recognize the batting movements. The results showed that this method could more effectively identify the hitting action of ball sports and had good recognition accuracy[10]. The transfer learning method was used to identify the hitting actions of ball sports, but further exploration was necessary for other action recognition. Therefore, this study took a Dual-Graph Convolutional Network (Dual-GCN) model to identify different ball sports actions and scenes. N. J. Bommagani et al. proposed a human action recognition method based on artificial butterfly optimization algorithm, which used a dual branch neural network for data extraction, greatly optimizing the convolution kernel size and enhancing the model's performance and recognition accuracy in action recognition. The research results indicated that this method could significantly improve the accuracy of human action recognition[11]. Although this method can recognize human movements, its effectiveness in recognizing motion movements still needs further exploration. Therefore, Dual-GCN is used to analyze human motion. To achieve high-precision 3D pose estimation and motion optimization for track and field athletes, Hauri S et al. proposed an Internet of Things enhanced pose optimization network (IE PONet), which integrated modules such as C3D, OpenPose, and Bayesian optimization. The research results showed that on the NTURGB + D and FineGYM datasets, the average precision scores of the network were 90.5 and 91.0, respectively, and the mean average precision scores were 74.3 and 74.0, respectively. The ablation study confirmed the important role of each module, providing a powerful tool for sports performance analysis, etc[12]. However, this study cannot achieve motion segmentation and recognition optimization for different motion scenes. Ren F et al. proposed a new deep learning method, NETS, for Group Activity Recognition (GAR) in team sports to automatically identify team activities from basketball game tracking data to assist players and coaches in formulating strategies. This method combined Transformer-based architecture with LSTM embedding, and took a team pool layer to identify team activities. To alleviate the high labeling cost, self-supervised pre-training combined with strong and weak label fine-tuning was adopted. The research results indicated that NETS could learn group activities with high accuracy, and self-and weakly supervised training had a positive impact on the accuracy of GAR[13]. However, further exploration should be conducted on how this method can recognize and analyze the motions of different individuals. To improve the effectiveness of working memory training and enhance the transferability of daily life, Chen A et al. proposed a VR sandbox MEM Box for working memory training and evaluation, which simulated daily scenes and adaptively adjusted task difficulty. Compared with the PC-based training control group, the group trained with MEM Box showed greater improvement in working memory ability. Physiological data such as decreased HRV and SDNN, as well as increased sympathetic nervous system activity, further confirmed the effectiveness of MEM Box and provided physiological data support for constructing adaptive VR systems[14]. However, further investigation is necessary to determine whether the model is suitable for motion scene image recognition. In summary, traditional models mainly recognize and analyze information such as single character skeleton points in motion behavior recognition. There is still a significant research gap in motion behavior recognition for multi-character scenes. In baseball, the positional information of two characters cannot fully explain the association between the current characters, but the Dual-GCN model is able to analyze the association between different characters by analyzing the association between different individuals and adding the Pose-Guided Pruning (PGP) module to analyze the association between different characters. Secondly, baseball action has unique action characteristics and rules, which requires more detailed spatial and temporal characteristics and interactions between characters for action capture of baseball. The Dual-GCN model can utilize its powerful graphic data processing capabilities to extract athletes' motion patterns and behaviors from video data, making it suitable for baseball scene recognition. The GCN model can utilize its powerful graphic data processing capabilities to extract motion patterns and behaviors from video data. Therefore, the Dual-GCN model can be applied to baseball scene recognition. Dual-GCN can segment the key region to enhance the model's segmentation effect on the characters' behaviors and actions. Meanwhile, key characters are defined as the most focused characters among other characters in baseball scene recognition. Therefore, Dual-GCN can be applied to character and behavior segmentation in baseball scenes. The current progress in research on human activity recognition indicates that deep learning has brought new opportunities and challenges to HAR. By designing more advanced model structures and utilizing techniques such as unsupervised and semi-supervised learning, the performance of HAR is further improved. Future research can explore this framework to handle multi-modal sensor data, achieve real-time processing, and optimize for specific applications, thereby further expanding its practicality in various HAR scenes.

Therefore, this study innovatively integrates Motion Weight Measure (MWM) and PGP into the Dual-GCN framework, demonstrating moderate innovation in the field of baseball behavior recognition. MWM assigns weights by analyzing the motion amplitude of characters between adjacent frames, enhancing the model's attention to motion salient regions, while PGP module determines key regions through bone point information,

improving the model's recognition ability for key regions. The concept of semi-circular key areas is different from traditional interest area methods, which is more in line with the characteristics of baseball. Finally, the model segments the key regions in the baseball video image using Dual-GCN to enhance the segmentation effect on character behavior and action. The Dual-GCN model can analyze athlete movements through video streams, provide real-time feedback to coaches, and optimize training efficiency. Moreover, this model can also identify tactical strategies to assist in tactical design during competitions, monitor athletes' abnormal movements in real time, and warn potential injury risks in advance. Dual-GCN processes dynamic propagation graphs and dynamic knowledge graphs using two parallel graph convolution network branches to capture the motion relations and static features of characters, respectively. Each branch extracts node features through dual-layer graph convolution and fuses them in a time-series fusion unit to learn time-series data, which is ultimately fed into a classifier for behavior recognition. To recognize critical regions, motion weights are introduced, which are calculated based on the character's motion amplitude between adjacent frames. A larger motion amplitude indicates a higher weight. The PGP module extracts key regions of characters in videos, analyzes skeletal point information, identifies key regions, and fuses their features with global features to further improve recognition performance. The study aims to propose a sports character recognition model that can be applied to multi-character complex scenes, providing new research ideas for motion character action recognition.

Methods

Analysis of baseball behavior recognition

Baseball is a ball sport that involves multiple players. The behavior recognition of baseball players is a group behavior recognition process aimed at multiple people. In current research, group recognition methods can be analyzed by analyzing the correlation between people in the scene. In baseball, the position information of two characters cannot fully explain the correlation among the current characters. The Dual-GCN model consists of two parallel GCN branches that process dynamic propagation graphs and dynamic knowledge graphs, respectively. The dynamic propagation graph is used to capture the motion relationships between characters, while the dynamic knowledge graph captures static features and background information of characters. Therefore, a Dual-GCN is proposed to analyze the correlation among different individuals, and a PGP module is added to the Dual-GCN to analyze the correlation between different characters[15]. Equation (1) shows the network recognition of motion behavior regions.

$$\{FC_x\}_{x=1}^{L} = PGP((M_1, M_2, M_3, \cdots, M_n), I_x) \tag{1}$$

.

In Eq. (1), FC represents the identified key region. I_x represents the frame number of the video image as x. M_n signifies the number of n individuals in the current I_x video image. PGP represents the action pruning module. L signifies the number of key areas. $PGP(\cdot)$ signifies the result of processing character relationships under different character coverage situations, as well as the number of characters obtained. By combining the relationships and features of different baseball players in Dual-GCN, key information in the video image area is obtained. The final video image representation data is obtained through scene accumulation, as shown in Eq. (2)[16].

$$F^x = \sum_{j=1}^{L} CN(C(I_x), N(FC_j)) \tag{2}$$

In Eq. (2), F^x represents the representation data after processing the x-th frame video image. CN represents Dual-GCN video image processing operation. $C(I_x)$ represents the network recognition process of character representation data in video images when the frame quantity is x. $N(FC_x)$ represents the operation of connecting different characters in the network when the frame quantity of the video image is x. The PGP model is used to locate different key characters in baseball video images, and the final character relationships are obtained through Dual-GCN processing[17]. The Dual-GCN model is displayed in Fig. 1. (The image data in the picture is sourced from http://www.dataworld.co.za/public dataset).

From Fig. 1, in the Dual-GCN, when obtaining baseball sports video images, the PGP model processes the video images, obtains key character regions in the current scene based on the skeletal features of the characters in the image, and extracts key data features through representation data to input into the Dual-GCN. Cluster learning is performed on information data such as relationships, positions, and actions of different characters in key areas through a Dual-GCN. Different character association data information is obtained. The final video image data information is input into the data processing layer to process and analyze the video image characters, obtaining the final character relationships and data classification results. Dual-GCN consists of two parallel GCN branches, and the features extracted from these two branches are fused in the temporal fusion unit. The final fused features are input into the classifier, which outputs the final behavior category. Figure 2 shows the cutting effect of key areas in the PGP module.

From Fig. 2, there is a motion trend due to the large limb movements. The distribution of skeletal points is relatively scattered. The skeletal points of characters without any motion trend tend to be in a straight line. In the process of processing skeletal point video images, due to the significant difference in skeletal points between people in motion and non-motion states, this model divides the skeletal points of athletes in motion and non-motion states[18]. The basis for judgment is that if the angle between the skeletal points of the current character is greater than 30°, the current character is in motion. If the angle is less than 30°, the character is non-motion[19]. The change in joint angle is a key indicator for distinguishing between rest and exercise in human

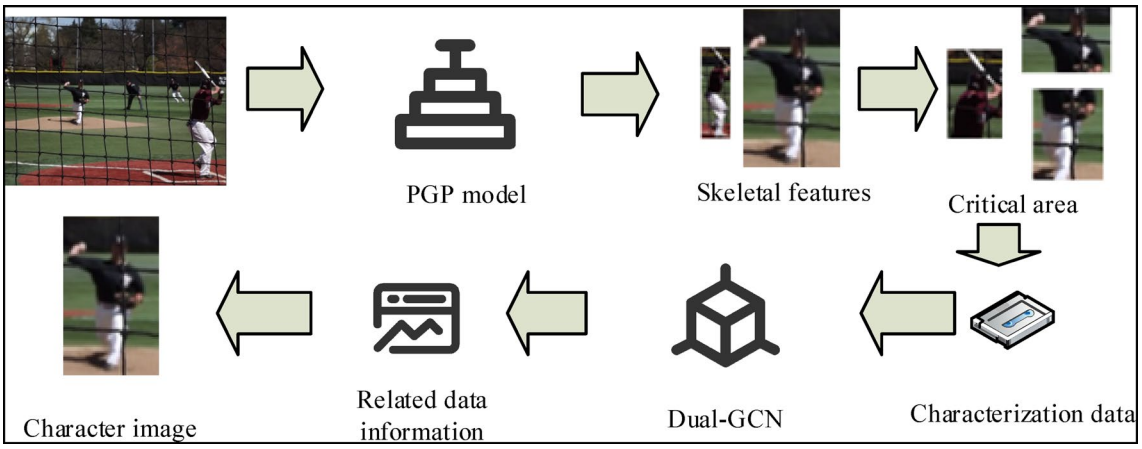


Fig. 1. The structure of the Dual-GCN model. (Source from: https://www.1001freedownloads.com/free-photo/baseball-sports-plate-home-batter-up-park-field).

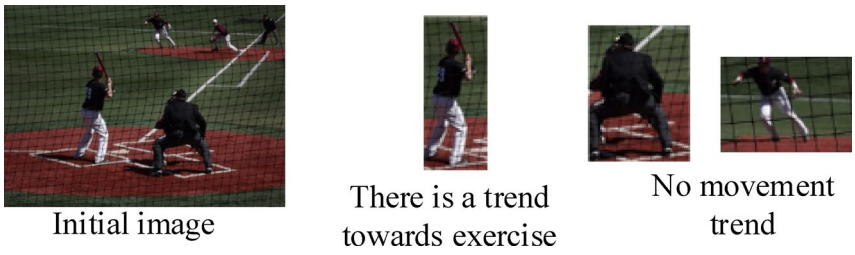


Fig. 2. The cutting effect of key areas in the PGP module. (Source from: https://www.1001freedownloads.com/free-photo/baseball-sports-field-plate-home-park-batter-up).

motion. Actions such as arm swinging and leg stepping can significantly alter the angle between skeletal points. Experiments have shown that 30° is the critical value that can effectively differentiate between minor postural adjustments and clearly defined motions. The joint angle usually remains stable at rest, but changes more during motion. 30° is a threshold that filters out small daily actions and focuses on significant action behaviors[20]. As shown in Eq. (3), the change in the area above the neck of a person in motion is represented .

$$a_w^i = \frac{(a_r^i - a_t^i)}{(b_r^i - b_t^i)} * (b - b_t^i) + a_t^i \tag{3}$$

.

In Eq. (3), a_w^i represents the bone points above the neck of the i-th person in the video image. (a_r^i, b_r^i) represents the horizontal and vertical coordinates of the nasal skeletal points in the video image. (a_t^i, b_t^i) signifies the horizontal and vertical coordinates of the neck skeletal points in the video image. $*$ represents multiplication symbols. For non-motion individuals, the change above the neck is shown in Eq. (4)[21].

$$a_w^i = -\frac{(b_r^i - b_t^i)}{(a_r^i - a_t^i)} * (b - b_t^i) + a_t^i \tag{4}$$

In Eq. (4), the changes in the skeletal points of the character are basically the same in both motion and non-motion states. Therefore, when analyzing the range below the neck of the character, the changes in the skeletal points of the knee and ankle joints are analyzed, as shown in Eq. (5)[22].

$$a_q^i = -\frac{(b_u^i - b_g^i)}{(a_u^i - a_g^i)} * (b - b_g^i) + a_u^i \tag{5}$$

.

In Eq. (5), a_q^i represents the skeletal points below the neck of the i-th person in the video image. (a_u^i, b_u^i) signifies the horizontal and vertical coordinates of the knee joint skeleton points of the character in the video image. (a_g^i, b_g^i) signifies the horizontal and vertical coordinates of the ankle joint skeleton points in the video image. After determining the skeletal points, key areas are drawn based on the skeletal points of the character, where the character in the baseball scene that is most covered by the attention of other characters is defined

as the key character. The semi-circular area of the key character is the key area. The size and number of key areas determine the scene recognition accuracy for the entire network model. The changes in appearance characteristics in character relationships can affect the changes in relationships between characters. Therefore, based on the model, this study analyzes the geographical and appearance relationships of characters, as shown in Eq. (6)[23].

$$h(V,G) = \frac{Ge^{V}}{\sum Ge^{V}} \tag{6}$$

In Eq. (6), G represents the change in the positional information of the character in the scene. V represents the appearance change of the character in the scene. $h(V,G)$ denotes the weight of final location information change V and appearance change G. The appearance change of the character on the internet is shown in Eq. (7)[24].

$$V(h_{s_1}, h_{s_2}) = \frac{(h_{s_1})^{T} \cdot h_{s_2}}{\sqrt{d}} \tag{7}$$

In Eq. (7), h_{s_1} represents the appearance vector of the representation data of the first character in the key region s. h_{s_2} represents the appearance vector of the representation data of the second character in the key region s. d represents the sum of squares of two appearance vectors. Simultaneously, the distance characteristic of characters based on their distance and attention relationships is determined. The attention coverage relationship is shown in Eq. (8)[25].

$$G(J_{s_1}, J_{s_2}) = \begin{cases} 1 \\ 0 \end{cases} \tag{8}$$

In Eq. (8), when the coverage range of J_{s_1} includes J_{s_2}, the distance is judged as 1. When the coverage range of J_{s_1} does not include J_{s_2}, the distance is judged as 0. J_{s_1} represents the first character in the key area s. The distance relationship is shown in Eq. (9)[26].

$$Z((a_{s_1}, b_{s_1}), (a_{s_2}, b_{s_2})) = \mathrm{Re}\, LU(W\delta((a_{s_1}, b_{s_1}), (a_{s_2}, b_{s_2})) + \theta) \tag{9}$$

In Eq. (9), $Z((a_{s_1}, b_{s_1}), (a_{s_2}, b_{s_2}))$ represents the distance between the first and second characters in the key area s. W signifies the weight matrix. θ signifies the bias vector. δ represents the distance between two different characters when tiling. Based on the distance and appearance of the character, the key area of the current character is finally analyzed. The skeletal point image processing process focuses on extracting and analyzing key motion features of characters in baseball videos to more accurately identify and analyze their behaviors. Changes in skeletal points are computed to distinguish the movement states to analyze key areas and quantify the spatial relationships and interactions between the characters. The Dual-GCN model is also provided with rich contextual information to enhance the model's recognition ability in complex multiplayer scenes, thus optimizing the model's performance to more accurately recognize various behavioral patterns in baseball games. The PGP module enhances the model's ability to recognize character behaviors by extracting key area features and fusing them with global features. In a baseball game, the PGP module can extract the key region features of the batter's swinging action and the runner's running action, thus improving the recognition performance.

Dual-GCN model improved based on motion weights

Due to significant differences in recognition between individuals in key areas and all individuals in the scene, the contribution of individuals in a single key area varies among different individuals[27]. Therefore, a MWM module is added to the Dual-GCN in the previous section, as shown in Fig. 3.

From Fig. 3, after obtaining the video image, the model mainly processes the video image through the PGP module. After processing, the motion weight module is used to quantify the contribution of the processed data. Different input data information is quantitatively considered based on different contribution degrees. Then, the quantized data is convolved using a convolutional neural network to obtain the video image representation data. Finally, the final video image data information is obtained through the Dual-GCN. The PGP model mainly consists of a skeletal point analysis module, a key region extraction module, a feature fusion module, and a motion weight module. The MWM module compares the position change of a character between neighboring frames. If the change exceeds a certain threshold, the character is considered to be in motion and the motion weights are increased accordingly. These weights are normalized by the Softmax activation function to enhance the model's focus on motion-significant regions. In this way, the MWM module is able to effectively distinguish between motion and stationary states. When analyzing the contribution of key regions in the model, the MWM module mainly measures and quantifies the impact of key regions. The quantification process is shown in Eq. (10)[28].

$$E_I^K = O(MWM(S_K, A_I)) \tag{10}$$

.

In Eq. (10), E_I^K represents the motion weight of the character in the K-th key area. MWM represents performing motion weight analysis on video image data. S_K represents the K-th key area. A_I represents the I-th character in the current designated key area. O represents the activation function Softmax. The main difference among MWM, SENet, and CBAM is that MWM focuses on action changes in time series data, while SENet

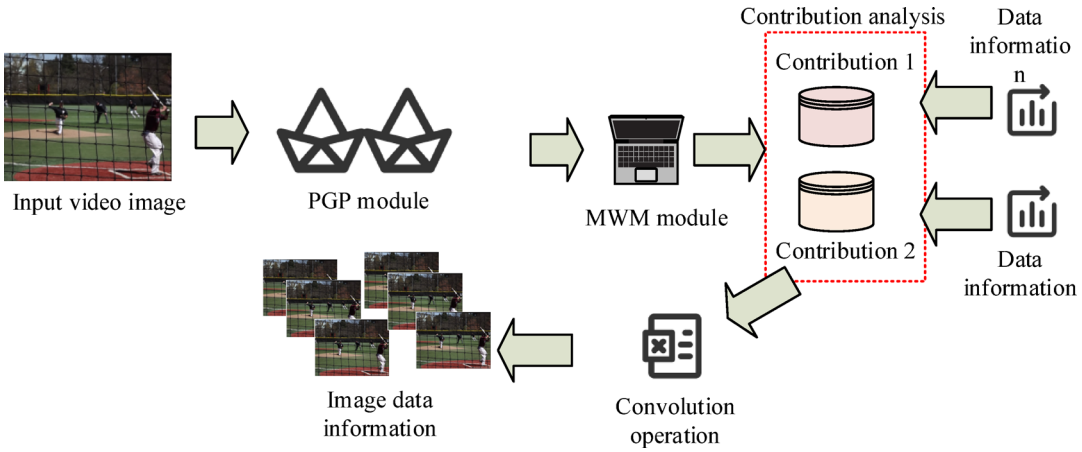


Fig. 3. The model with added motion weight module. (Source from:https://www.1001freedownloads.com/free-photo/baseball-sports-plate-home-batter-up-park-field).

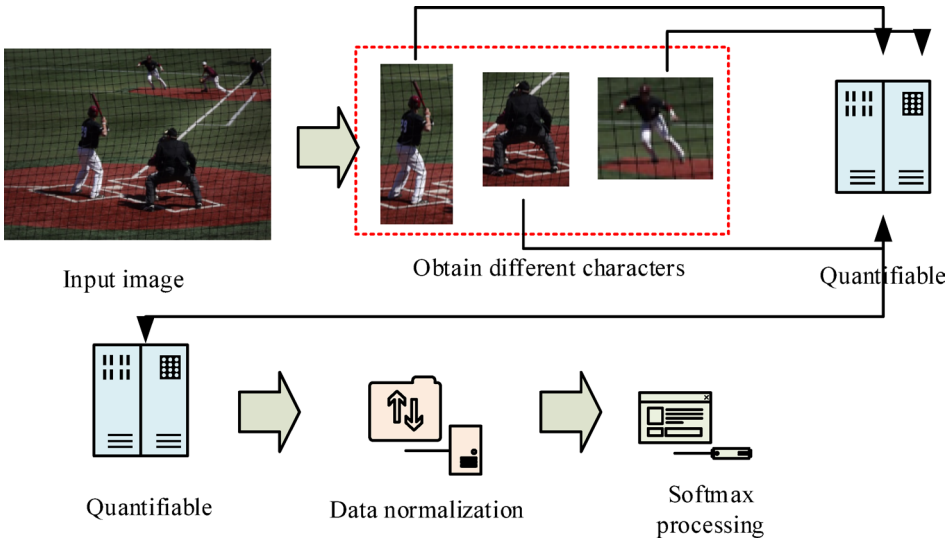


Fig. 4. Structure of motion weights. (Source from:https://www.1001freedownloads.com/free-photo/baseball-sports-field-plate-home-park-batter-up).

and CBAM focus on enhancing spatial features and relationships between channels. The final motion weight structure based on character characteristics is shown in Fig. 4. (The image data in the picture is sourced from www.dataworld.co.za/public dataset).

From Fig. 4, after obtaining the video image data, the MWM module divides the tasks in the video image, obtains the information of the video image in different regions, and quantifies them according to the different motion situations and contribution sizes of the character video image. Furthermore, the quantized data is normalized using the Softmax activation function. Finally, the final motion weight size of the module is obtained through activation function processing, and the data is initialized with numerical values by analyzing the characterization data of key areas[29]. When using MWM, due to the different scene movements between characters, the positions of characters in adjacent frames of video images will change. Therefore, according to the differences in position changes between adjacent images, the current motion amplitude between video images can be reflected.

In summary, every frame of video image data is undergoing changes. Therefore, the difference in frame rate changes between two images determines the magnitude of the character's motion changes. If the motion amplitude difference between the current character images is large, it indicates that the number of frames between adjacent images is relatively large. Therefore, the image pixels are converted into 1 or 0. At this point, the pixel values of the obtained image should be less than the set value. The current image motion situation has changed little[30]. If the pixel size is greater than the set threshold, it indicates that the motion amplitude at the current position is relatively large, as displayed in Eq. (11).

$$D_n(x,y) = |f_n(x,y) - f_{n-1}(x,y)| \tag{11}$$

.

In Eq. (11), $D_n(x,y)$ signifies the converted numerical value of the image. $f_n(x,y)$ represents the image in frame *n*. (x,y) signifies the pixel coordinates of the image. The threshold size of the image pixels set after processing is shown in Eq. (12)[31].

$$T_n(x,y) = \begin{cases} 255 \\ 0 \end{cases} \tag{12}$$

In Eq. (12), $T_n(x,y)$ represents the image after pixel conversion, where the current image size is 255 when the pixel value is greater than the set threshold. If the pixel value is less than the set threshold, the current image size is 0. The model loss function is shown in Eq. (13)[32,33].

$$Loss = L_a(U^a, \bar{U}^a) \tag{13}$$

.

In Eq. (13), $Loss$ represents the loss function. L_a represents the size of the cross-entropy loss function of the model. U^a represents the prediction results of group action models in the video image. $\bar{U}^a$ represents the actual result of group actions in the video image. This study takes a pose estimation method based on GCN to extract skeleton points, but inaccurately describing key areas can reduce recognition ability and affect the accuracy of behavior recognition. Meanwhile, inaccurate feature fusion can also weaken the model's overall understanding of character behavior, thereby affecting recognition results. To reduce these impacts, more advanced attitude estimation techniques such as OpenPose, as well as methods such as data augmentation, preprocessing, multimodal data fusion, and model optimization, have been introduced to improve the accuracy of skeleton point extraction. Motion weighting enhances the model's ability to identify key regions by focusing on important areas of motion. In a baseball game, the swinging motion of the batter and the running motion of the runner are critical regions, and motion weights can enhance the feature representation of these regions. There are significant differences between professional athletes and amateur athletes in physical conditions, skill levels, tactical coordination, and training conditions. Professional athletes demonstrate higher competitiveness and professional abilities. These differences are manifested in the game visuals, as professional players are able to more effectively utilize their skills to break through defenses and create scoring opportunities, while amateur players may perform less well in these areas. The standard size diagram of a baseball field shows that the infield is a right angled fan-shaped area, while the outfield is the area enclosed by the extension of the first and third baselines. However, baseball may be played at different scales or under different rules, and these venues may not follow the standard size of professional baseball fields, resulting in the need to adjust the definition and application of key areas accordingly. Therefore, although the model may be developed based on standardized field dimensions, the impact of different competition environments on the definition of key areas should be considered, and the model parameters should be adjusted accordingly to ensure accuracy and reliability. The lightweight design of the model introduces the EfficientNet architecture, which is also optimized for embedded devices and mobile devices GPU/CPU. Finally, low-bit quantization is used to convert floating-point weights and activation values into integer or lower bit values, reducing memory footprint and computational overhead. It also uses a specialized AI acceleration chip TPU to improve the model inference speed on edge devices. The Dual-GCN model performs well in baseball sports behavior recognition and has the potential to be extended to sports with more complex group interactions. However, to achieve different motion extensions, the model needs to be further optimized to cope with more complex motion scenes and multi-person interactions. The generalization ability and real-time performance of the model need to be improved through data augmentation, multi-modal data fusion, and model optimization.

Dual-GCN processes dynamic propagation graphs and dynamic knowledge graphs through two parallel branches to capture the motion relations and static features among characters, respectively. The dynamic propagation graph is used to analyze the motion associations between characters, while the dynamic knowledge graph extracts static features and background information of characters. The Dual-GCN model can more comprehensively capture the motion behaviors in the scene by combining the information from these two graphs.

Results

Analysis of the recognition effect of baseball behavior recognition model

The model training process starts with weight and bias initialization settings, and then propagates the input data forward through each layer of the model. The cross-entropy loss function is used to label the differences. Then, the weights and biases of the model are updated by back-propagation algorithm. Finally, these parameters are updated using Adam optimizer. To investigate the effect of the model on baseball behavior recognition, the MLB-YouTube[34] Baseball Dataset is used for the study. The dataset contains a large number of video frames, with video duration exceeding 42 h. The estimated total frame rate is in the millions, with videos containing 25–30 frames per second. The size of the entire dataset ranges from tens to hundreds of GB. The MLB-YouTube Baseball Dataset consists of two small datasets. The Segmented Video Dataset contains 4,290 video clips, each labeled with multiple action categories related to baseball. The Continuous Video Dataset contains 2,128 1–2 min long video segments, each densely labeled with an average of 7.2 activity instances, totaling over 15,000 action instances. The research takes the PyTorch framework and designs the model through a system server. The system takes Xeon E5 CPU as the chip, Nvidia TITAN as the graphics card, and 64GB of system memory.

The encoder is Adam, with a model learning rate set to 0.001 and batch sizes set to 16 and 8. The hidden layer dimension of the Dual-GCN model is set to 64, and the number of convolutional layers is 2. In the study, the dataset is divided into a training set and a test set in a ratio of 2:8. The actions involved in the study include daily activities such as eating and reading, as well as various hitting actions in complex sports such as baseball, including fastballs, changeovers, sliders, curveballs, and spirals. All methods used in the study were conducted under the same conditions to ensure the specificity of the research. The NTU RGB + D + I dataset contains 60 action categories, while the PKU-MMD dataset contains 51 action categories. Meanwhile, this study uses data augmentation to over-sample a few categories, under-sample the majority of categories, or use category weights to mitigate the impact of category imbalance on model performance. The number of iterations for the dataset is set to 150 and 120. To compare the effectiveness of different models in baseball action recognition, the Inflated 3D ConvNet (I3D)[35], InceptionV3[36], and Temporal Convolutional Networks (TCNs)[37] are compared with the model proposed in the study. The attention sizes of different models are analyzed for their effects on video image recognition. The results are shown in Fig. 5. The attention size is set to w, 1/2w, 1/4w, 1/6w, and 1/8w, where w represents the frame width in the video image. Image frame rate refers to the sampling rate or resolution representation of an image, rather than the traditional number of images processed per second.

From Fig. 5 (a), the accuracy of the research model was relatively higher compared with other models. The highest accuracy was 97.68% when the image frame rate was 1/2w. The highest recognition accuracy of the I3D was 94.85%. Compared with the I3D, the research model was improved by 2.83%. The highest accuracy in the research model was achieved when the image frame rate was 1/2w. This indicates that the recognition effect of video images was best at this frame rate. A large image frame rate difference indicates poorer image recognition performance. From Fig. 5 (b), there were differences in the recognition accuracy of different models in the Continuous Video Dataset. The TCNs model had the worst recognition accuracy, with a minimum accuracy of 82.75%. The maximum recognition accuracy of the research model was achieved at 1/2w image frame rate, with an accuracy of 97.43%. This indicates that different models have differences in image recognition across different datasets. This may be due to differences in image frame rates and baseball actions in the dataset. To analyze the impact of key regions on the model, the study sets the key region sizes to H, 1/2H, 1/4H, 1/6H, and 1/8H, where H represents the width of the key region in the video image, as shown in Fig. 6.

From Fig. 6 (a), the model used in the Segmented Video Dataset had the best recognition accuracy when the width of the key area in the image was 1/2H, with a maximum recognition accuracy of 95.47%. Compared with

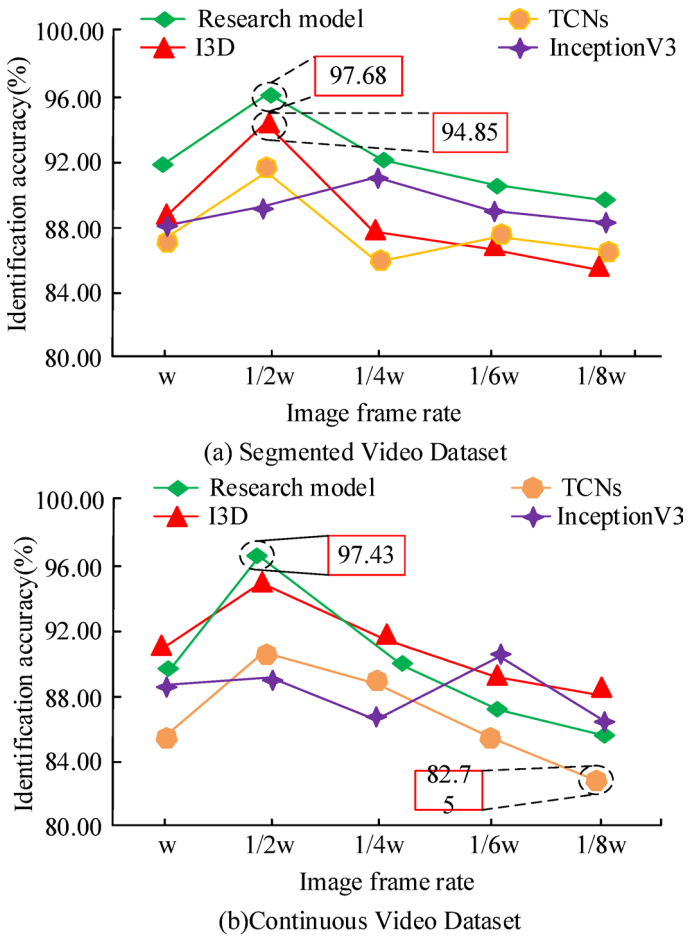


Fig. 5. Comparison of frame rate accuracy for image recognition using different algorithms.

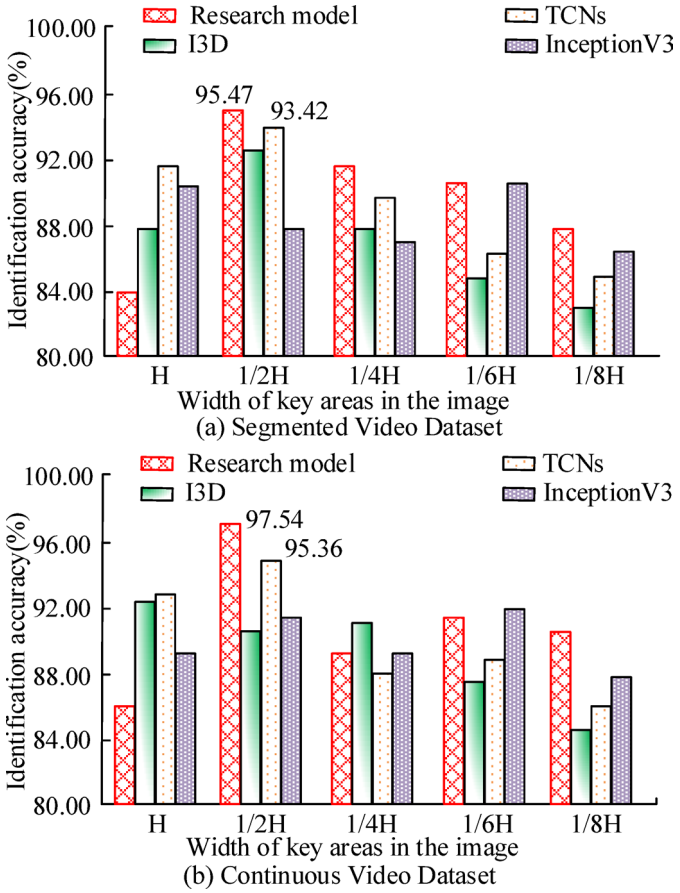


Fig. 6. The influence of key area width on model recognition accuracy.

the TCNs, the recognition accuracy was improved by about 2.05%. This indicates that the recognition accuracy is higher when comparing different models, which may be due to the motion weights in the model. In Fig. 6 (b), the recognition accuracy of the model used in different datasets also reached its best performance when the key region width was 1/2H. Compared with the TCNs model, the model recognition accuracy was improved by 2.18%. In the comparative analysis of different datasets, the recognition accuracy of the research model was better, and the size and width of key regions also determine the recognition effect of the model. Therefore, it is best to set the width of the key area to 1/2H. To analyze the impact of the number of key regions on the model, different models are compared, as shown in Fig. 7. The current basic recognition accuracy is set to 85.00%. If it does not exceed the set line, the model has not met the basic recognition requirements.

From Fig. 7 (a), in the comparison of the number of key regions on the Segmented Video Dataset, the best model performance was achieved when the number of key regions was 2, and the recognition accuracy was 92.78%. Compared with the InceptionV3 model's 85.34%, the accuracy of the research model was increased by 7.44%. From Fig. 7 (b), in the comparison test of the Continuous Video Dataset, the model achieved the highest recognition accuracy of 91.54% when the number of key regions was 2. However, the recognition accuracy of the InceptionV3 model remained less than 85.00% as the number of key regions increased. Therefore, the InceptionV3 model cannot meet the basic recognition accuracy requirements in this dataset. This may be due to poor performance in this dataset. From the above results, different models require different parameter data for baseball recognition. Therefore, to compare the actual recognition performance under the optimal parameters of different models, the recognition accuracy, model recall, frame rate, and Intersection over Union (IoU) of the four models are analyzed, as displayed in Table 1.

In Table 1, a high frame rate indicates a high recognition frequency of video images and better model performance. Moreover, a larger IoU indicates a smaller deviation between the real situation and the actual situation. From Table 1, the highest recognition accuracy of the model used in the study was 92.51%, which was 13.66% higher than TCNs. The recall rate, IoU, frame rate, and other parameter indicators of the model used in the study were higher than the other three models. The research model has better recognition performance in baseball action recognition. This may be due to the model adopting more methods to improve its performance.

Performance analysis of baseball behavior recognition model

To compare the performance of the baseball sports recognition network with different methods, the model performance of different methods for MWM processing is compared. The comparison models include Hierarchical Temporal Deep Model (HTDM)[38], Cross-Environment Recurrent Neural Network-2 (CERN-2)[39],

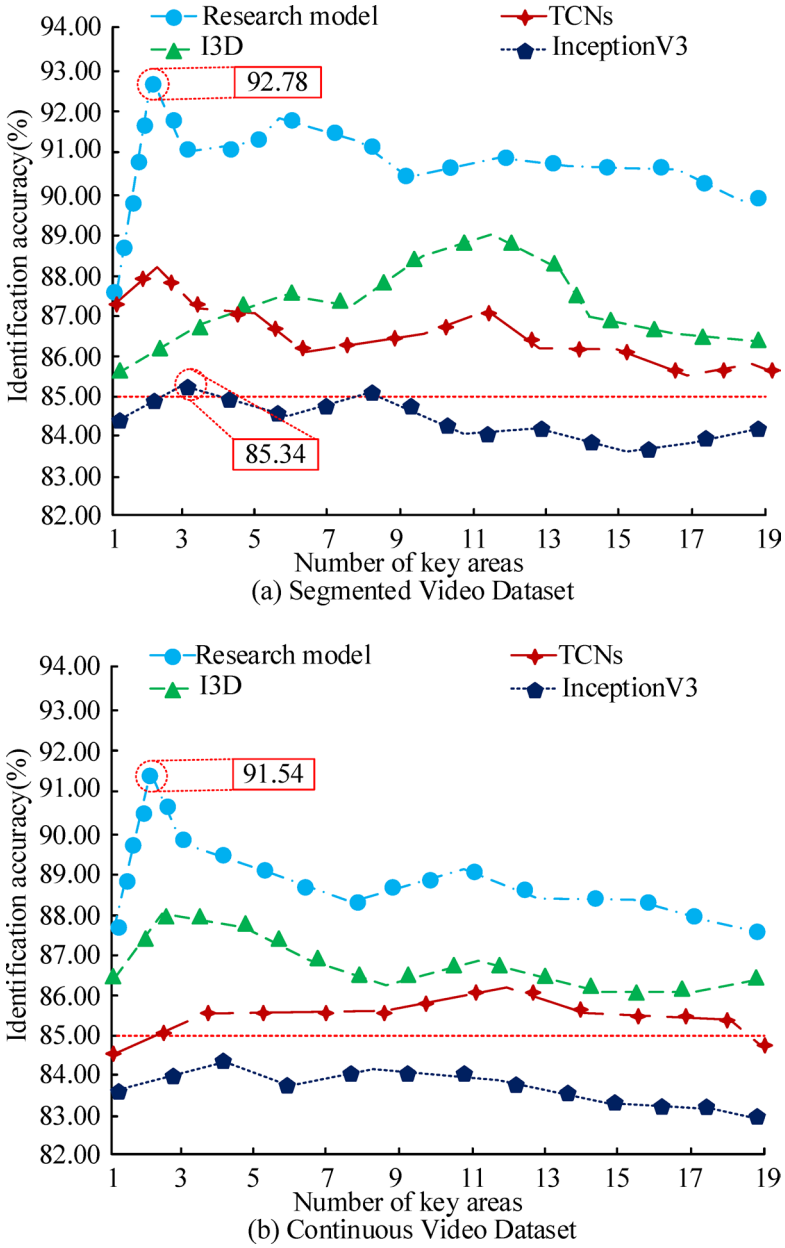


Fig. 7. Comparison results of key area quantities.

Model	Accuracy	Frame rate	Recall	Intersection over union
I3D	85.51%	35	83%	0.75
InceptionV3	84.84%	44	80%	0.72
TCNs	78.85%	45	76%	0.79
Research model	92.51%	56	85%	0.88

Table 1. Comparison of recognition indicators for different models.

Action Recognition Guided (ARG)[40], Pose Regression Learning (PRL)[41], and Dynamic Interaction Network (DIN)[42], as demonstrated in Table 2.

From Table 2, when different methods were used to process baseball sports images, the actual processing effect of the model was better. The highest recognition accuracy for video images reached 94.84%, which was 12.06% higher than that of the HTDM. The research model had better model recognition accuracy. From the number of individuals recognized by the model in the dataset, the number of individuals recognized in the

Method	Accuracy	Number of recognized individuals in dataset	Number of video frames in the dataset
HTDM	81.90%	500	100,000
CERN-2	83.30%	550	120,000
ARG	92.30%	642	150,000
PRL	91.40%	620	140,000
DIN	93.30%	650	160,000
Research model	94.84%	768	170,000

Table 2. Comparison of model effects of different methods.

study reached 768 in the dataset. Finally, in the comparison of the number of video frames in the dataset, the number of image frames using the model in the study reached 170,000. For the same dataset, the model can decompose into more detailed images and have better recognition and processing effects on images. To compare the performance of the model after adding different module, the performance after adding different modules is compared, as shown in Fig. 8. The added modules are Region Distance Graph (RDG), Region Overlapping Graph (ROG), and MWM, respectively.

From Fig. 8 (a), the accuracy of the model increased first and then tended to a relatively stable state with the increase of iterations when different modules were added. The model with MWM module had the highest accuracy of 92.98% after stabilization. Compared with the Dual-GCN model without any added modules, its accuracy was improved by about 3.45%. The accuracy of the model was effectively improved through different modules for model improvement, indicating that using modules can effectively enhance the recognition performance. From Fig. 8 (b), when using the Continuous Video Dataset for model testing, the accuracy change was consistent with that in Fig. 8 (a). This indicates that the changes in the dataset did not have a significant impact on the model performance. However, the recognition accuracy in different models has significantly decreased. This indicates that the actual performance of the model is worse when tested on the Continuous Video Dataset, which may be due to the larger video image data in the dataset. The parameter performance after adding different modules is compared, as displayed in Table 3. Recognition quantity refers to the number of correctly recognized instances, frameworks, or actions.

From Table 3, in the Segmented Video Dataset, the recognition quantity of Dual-GCN + MWM model was 94,860, with a shorter running time of only 1.35s, and a smaller number of model parameters of only 8,623. The Dual-GCN + MWM model had better performance and faster data processing. The Dual-GCN + MWM model has higher data volume and runtime parameters in continuous video datasets than in segmented video datasets, indicating that larger continuous video datasets lead to poorer model performance. The actual running effect of the research model is displayed in Fig. 9. (The image data in the picture is sourced from www.dataworld.co.za/public dataset).

From the comparison between Fig. 9 (a) and Fig. 9 (b), it can be seen that when identifying the combination of baseball base running and hitting actions, the model recognizes the characters separately and obtains different nodes based on their motion points. The recognition accuracies for different actions in Fig. 9(a) and Fig. 9(b) were 93.62% and 94.22%, respectively. The current set of actions was recognized according to the node situation. From the comparison between Fig. 9 (c), in the recognition of baseball batting actions, it was similar to running base recognition. The model divided the group of characters into key areas for action recognition and disassembly. The model can more effectively recognize and analyze character actions when recognizing baseball actions through disassembly. The recognition accuracy for different actions in Fig. 9(c) was 94.21%.

To test the actual system operation, the study tests the system and obtains the test results, as shown in Table 4.

Identification effectiveness is the ratio of the number of correctly identified samples to the total number of samples. Motion weight improvement indicates the improvement in the recognition accuracy by introducing the MWM. The width of key region indicates the impact of the key area width setting on the model recognition accuracy during the recognition process. The number of key regions is the effect of the number of key regions on the recognition accuracy. Frame rate indicates the number of video frames that can be processed per second. Recall rate represents the ratio of the actual number of positive samples to the total number of positive samples in a model. The IoU indicates the overlap between the region recognized by the model and the real labeled region. The number of video frames indicates the total number of video frames contained in the test dataset, reflecting the dataset size. Running time indicates the total time required by the model to process the entire test dataset. In Table 4, the running time of the system was about 1.35s, and the number of running parameters of the system was about 8,623. The research method has high recall and recognition rates for different videos, with good operational effects. IoU (Key area) measures the overlap between the predicted area and the real area, which is used to measure the accuracy of the model in identifying the key area. Temporal localization accuracy (%) refers to the accuracy of the model in identifying the point in time at which an action occurs, i.e., the matching between the predicted action occurrence time and the actual occurrence time. Multiple Object Tracking Accuracy (MOTA) measures the overall accuracy of the model in tracking multiple targets. Different model metrics, including IoU, temporal localization accuracy, and MOTA are tested, as shown in Table 5.

From Table 5, the model used in the study had a higher IoU value of 0.78 in different indicator tests. The model also had the best performance on temporal localization accuracy and multiple object tracking accuracy, reaching 92.3% and 0.85, respectively. This shows that the model used in the study has a better recognition effect in motion action recognition. The pseudo-code used for the research model is shown in Fig. 10.

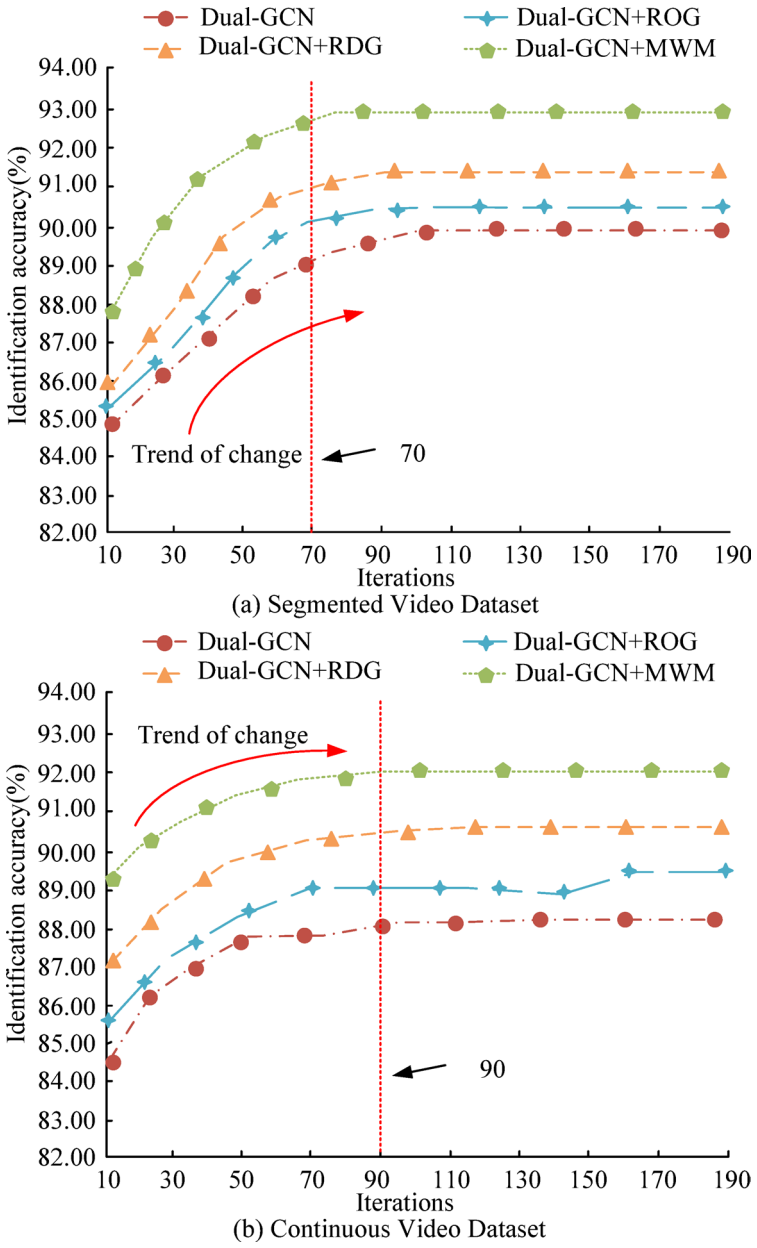


Fig. 8. Comparison of model accuracy changes.

Data set	Method	Recognition quantity	Running time(s)	Number of video frames in dataset
Segmented Video Dataset	Dual-GCN	15,214	3.12	13,845
	Dual-GCN + ROG	26,458	2.03	11,848
	Dual-GCN + RDG	28,641	1.98	9,458
	Dual-GCN + MWM	94,860	1.35	8,623
Continuous Video Dataset	Dual-GCN	26,344	3.68	15,742
	Dual-GCN + ROG	56,843	2.31	12,548
	Dual-GCN + RDG	68,451	2.43	13,654
	Dual-GCN + MWM	103,524	1.85	9,485

Table 3. Performance comparison of different models.



Fig. 9. Analysis of action recognition. (Picture (**a**) Source from:https://www.1001freedownloads.com/free-photo/baseball-sports-plate-home-batter-up-park-field; Picture (**b**) Source from:https://www.1001freedownloads.com/free-photo/baseball-sports-field-plate-home-park-batter-up; Picture (**c**) Source from:https://www.1001freedownloads.com/free-photo/little-league-baseball-batter-up).

Evaluation metric	Data parameter
System stability	Stability exceeding 0.990 at a image frame rate of 1/2w
Recognition effectiveness	Highest accuracy: 94.84%
Motion weight improvement	Accuracy improvement: 3.45%
Key area width	Best width: 1/2H, accuracy: 95.47%
Key area quantity	Best quantity: 2, accuracy: 92.78%
Frame rate	Frame rate: 56
Recall rate	Recall rate: 85%
Intersection over union	Intersection over union: 0.88
Recognition quantity	Recognition quantity: 768
Number of video frames	Number of video frames: 170,000
Running time	Running time: 1.35s (Dual-GCN + MWM model in Segmented Video Dataset)
Number of model parameters	Number of parameters: 8,623 (Dual-GCN + MWM model in Segmented Video Dataset)

Table 4. System test results.

Independent sample t-test

Verify whether there is a significant difference in accuracy between the Dual-GCN + MWM model and other models (I3D, InceptionV3, and TCNs). Accuracy: Dual-GCN + MWM: 92.51%. I3D: 85.51%. InceptionV3: 84.84%. TCNs: 78.85%.

Hypothesis testing: Null hypothesis (H_0): There is no significant difference in accuracy between models. Alternative hypothesis (H_1): Dual-GCN + MWM has significantly higher accuracy than other models.

Method: Use independent sample t-test, assuming that each group of data is independent and follows a normal distribution. Calculate the t-value and p-value, with a significance level of $\alpha = 0.05$.

Result: Dual-GCN + MWM vs. I3D: $t = 5.32$ and $p = 0.001$ ($p < 0.05$, rejecting H_0).

Dual-GCN + MWM vs. InceptionV3: $t = 6.15$ and $p = 0.000$ ($p < 0.05$, rejecting H_0).

Dual-GCN + MWM vs. TCNs: $t = 8.94$ and $p = 0.000$ ($p < 0.05$, rejecting H_0).

Model name	Intersection over union	Temporal localization accuracy (%)	Multiple object tracking accuracy
Dual-GCN + MWM	0.78	92.3	0.85
NAM-STGCN [5]	0.65	88.1	0.72
Customized ST-GCN [7]	0.71	89.5	0.68
Gait GCN++ [8]	0.63	85.2	0.79
Deformable GCN [9]	0.69	90.1	0.75
CS-Transformer + CLIP [10]	0.74	91.8	0.81
Hierarchical Temporal Model	0.58	82.4	0.64

Table 5. Comparison results of different models with different test indicators.

Conclusion: The accuracy of Dual-GCN + MWM is significantly better than that of comparison models.

Conclusion

The research aims to address the low accuracy in behavior recognition and limited character recognition in current baseball sports. A Dual-GCN baseball behavior recognition model based on motion weight improvement was proposed. The new model used Dual-GCN to recognize characters in baseball scenes and divided key areas based on their skeletal points. The MWM module was added to enhance the connections between characters. Finally, a baseball video image recognition system was developed to recognize and store baseball behavior videos. The research results showed that among different models, the model had the highest accuracy of 94.85% at an image frame rate of 1/2w, which was 2.83% higher than the accuracy of the I3D. The highest accuracy was 95.47% when the key region was 1/2H, which was 2.05% higher than that of the TCNs. The accuracy of the model varied in different datasets. When the key quantity was 2, the highest accuracy was 92.78%, which was 7.44% higher than the InceptionV3 model. A high image frame rate indicates good model performance. The highest accuracy of the model used was 94.84%, which was 12.06% higher than the HTDM. After adding the MWM module, the model accuracy increased by 3.45%. The added MWM module model significantly improved the recognition quantity and running time of system parameters. From this, the research model has higher accuracy and better model performance for baseball behavior recognition. Although some achievements have been made, there are still some shortcomings. The study only focuses on baseball. Therefore, future research will analyze more diverse sports and video images. The study only adds one module to improve model performance. Future research will use different modules to explore the effectiveness of the model. Since the model is designed for baseball, its training data and feature extraction cannot be applied to all sports. For some non-sports such as violence recognition, the model may need to be retrained, making the model unpredictable. Therefore, in subsequent research, the model will also recognize multiple different sports behaviors and optimize the model to improve the model's recognition of other sports. Finally, this study only takes some baseball datasets for testing, and further validation and analysis are necessary for the testing effectiveness of different ball sports. Therefore, in future research, more datasets with different ball sports will be tested to explore the recognition performance. The MLB-YouTube dataset may have focused primarily on male athletes, resulting in a demographic bias. Therefore, this model needs to be further validated and possibly adjusted in subsequent research to ensure its applicability to female athletes or youth leagues with different biomechanical characteristics. To enhance the generalizability of the model, future research can further explore cross-motion validation and cross-dataset validation, such as extending the model to other sports projects for testing, while analyzing multi-module collaborative optimization, model light-weighting, and real-time performance, thereby further improving the performance and application scope of the model.

```
# Import necessary libraries
import torch
import torch.nn as nn
import torch.optim as optim
from torch.utils.data import DataLoader, Dataset
import torchvision.transforms as transforms
from sklearn.metrics import accuracy_score, recall_score
import numpy as np

# Define the Dual-GCN model with Motion Weight Measure (MWM) and Pose-Guided Pruning (PGP)
class DualGCNModel(nn.Module):
    def __init__(self, num_classes, hidden_dim, num_conv_layers):
        super(DualGCNModel, self).__init__()
        self.dual_gcn = DualGCN(hidden_dim, num_conv_layers)
        self.mwm = MotionWeightMeasure()
        self.pgp = PoseGuidedPruning()
        self.classifier = nn.Linear(hidden_dim, num_classes)

    def forward(self, x):
        # Process input through Dual-GCN
        x = self.dual_gcn(x)

        # Apply Motion Weight Measure
        x = self.mwm(x)

        # Apply Pose-Guided Pruning
        x = self.pgp(x)

        # Classify the processed features
        x = self.classifier(x)
        return x

# Define the Dual-GCN component
class DualGCN(nn.Module):
    def __init__(self, hidden_dim, num_conv_layers):
        super(DualGCN, self).__init__()
        self.gcn_branch1 = nn.ModuleList([GCNLayer(hidden_dim) for _ in range(num_conv_layers)])
        self.gcn_branch2 = nn.ModuleList([GCNLayer(hidden_dim) for _ in range(num_conv_layers)])
        self.temporal_fusion = TemporalFusionUnit(hidden_dim)

    def forward(self, x):
        # Process through two parallel GCN branches
        x1 = x
        x2 = x
        for layer in self.gcn_branch1:
            x1 = layer(x1)
        for layer in self.gcn_branch2:
            x2 = layer(x2)

        # Fuse features from both branches
        x = self.temporal_fusion(x1, x2)
        return x
```

Fig. 10. Pseudo-code example.

Data availability

The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

Received: 19 February 2025; Accepted: 4 July 2025
Published online: 15 July 2025

References

1. Zhang, Y. & Hou, X. Application of video image processing in sports action recognition based on particle swarm optimization algorithm. *Prev. Med.* **173**(6), 107592. https://doi.org/10.1016/j.ypmed.2023.107592 (2023).
2. Afsar, M. M. et al. Body-worn sensors for recognizing physical sports activities in exergaming via deep learning model. *IEEE Access* **11**(1), 12460–12473. https://doi.org/10.1109/ACCESS.2023.3239692 (2023).

3. Furley, P. The nature and culture of nonverbal behavior in sports: Theory, methodology, and a review of the literature. *Int. Rev. Sport Exerc. Psychol.* **16**(1), 448–473. https://doi.org/10.1080/1750984X.2021.1894594 (2023).
4. Zhao, T., Zhang, J., Wang, Z. & Alturki, R. An improved deep learning mechanism for EEG recognition in sports health informatics. *Neural Comput. Appl.* **15**(6), 1–3. https://doi.org/10.1007/s00521-021-06118-z (2023).
5. Li, Z., Zhang, A., Han, F., Zhu, J. & Wang, Y. Worker abnormal behavior recognition based on spatio-temporal graph convolution and attention model. *Electron.* **12**(13), 2915–2916. https://doi.org/10.3390/electronics12132915 (2023).
6. Hu, K., Zheng, F., Weng, L. & Ding, Y. Overview of behavior recognition based on deep learning. *Artif. Intell. Rev.* **56**(3), 1833–1865. https://doi.org/10.1007/s10462-021-10116-x (2023).
7. Lovanshi, M. & Tiwari, V. Human skeleton pose and spatio-temporal feature-based activity recognition using ST-GCN. *Multimed. Tools Appl.* **83**(5), 12705–12730. https://doi.org/10.1007/s11042-023-17529-6 (2024).
8. Hasan, M. B., Ahmed, T., Ahmed, S. & Kabir, M. H. GaitGCN++: Improving GCN-based gait recognition with part-wise attention and DropGraph. *J. King Saud. Univ. Comput. Inform. Sci.* **35**(7), 101641–101642. https://doi.org/10.1016/j.jksuci.2023.101641 (2023).
9. Myung, W., Su, N., Xue, J. H. & Wang, G. DeGCN: Deformable graph convolutional networks for skeleton-based action recognition. *IEEE Trans. Image Process.* **33**(1), 2477–2490. https://doi.org/10.1109/TIP.2024 (2024).
10. Ma, L. & Tong, Y. TL-CStrans Net: A vision robot for table tennis player action recognition driven via CS-Transformer. *Front. Neurorobotics* **18**, 1443177–1443178. https://doi.org/10.3389/fnbot.2024.1443177 (2024).
11. Bommagani, N. J. et al. Artificial butterfly optimizer based two-layer convolutional neural network with polarized attention mechanism for human activity recognition. *Math. Modell. Eng. Probl.* **11**(3), 1–5. https://doi.org/10.18280/mmep.110306 (2024).
12. Hauri, S., & Vucetic, S. Group activity recognition in basketball tracking data–neural embeddings in team sports (Nets). In *ECAI 2023*, 1012–1019. https://doi.org/10.3233/FAIA230373 (IOS Press, 2023).
13. Ren, F., Ren, C. & Lyu, T. IoT-based 3D pose estimation and motion optimization for athletes: Application of C3D and OpenPose. *Alex. Eng. J.* **115**, 210–221. https://doi.org/10.1016/j.aej.2024.10.079 (2025).
14. Chen, A., Li, M. & Gao, Y. Mem-Box: VR sandbox for adaptive working memory evaluation and training using physiological signals. *Vis. Comput.* **40**(11), 7559–7573. https://doi.org/10.1007/s00371-024-03539-4 (2024).
15. Li, X. & Cao, X. Human motion recognition information processing system based on LSTM recurrent neural network algorithm. *J. Ambient Intell. Human. Comput.* **14**(7), 8509–8521. https://doi.org/10.1007/s12652-021-02617-0 (2023).
16. Raza, M. A., Chen, L., Nanbo, L. & Fisher, R. B. EatSense: Human centric, action recognition and localization dataset for understanding eating behaviors and quality of motion assessment. *Image Vis. Comput.* **137**, 104762. https://doi.org/10.1016/j.imavis.2023.104762 (2023).
17. Kobdeh, S. B., Yami, M. R. & Sareshkeh, S. K. Basketball action recognition based on the combination of YOLO and a deep fuzzy LSTM network. *J. Supercomput.* **80**(3), 3528–3535. https://doi.org/10.3233/JSC2024.80.3.3528 (2024).
18. Ghosh, I., Ramasamy Ramamurthy, S., Chakma, A. & Roy, N. Sports analytics review: Artificial intelligence applications, emerging technologies, and algorithmic perspective. *Wiley Interdiscipl. Rev. Data Min. Knowl. Discov.* https://doi.org/10.1002/widm.1496 (2023).
19. Liu, J. et al. A survey on location and motion tracking technologies, methodologies and applications in precision sports. *Expert Syst. Appl.* **229**, 120492. https://doi.org/10.1016/j.eswa.2023.120492 (2023).
20. Fitze, D. C., Mast, F. W. & Ertl, M. Human vestibular perceptual thresholds—A systematic review of passive motion perception. *Gait Posture* **107**, 83–95. https://doi.org/10.1016/j.gaitpost.2023.09.011 (2024).
21. Kumar, R. & Kumar, S. A survey on intelligent human action recognition techniques. *Multimed. Tools Appl.* **83**, 52653–52709. https://doi.org/10.1007/s11042-023-17529-6 (2024).
22. Khan, M. A. et al. Human action recognition using fusion of multiview and deep features: An application to video surveillance. *Multimed. Tools Appl.* **83**, 14885–14911. https://doi.org/10.1007/s11042-020-08806-9 (2024).
23. Ustad, A. et al. Validation of an activity type recognition model classifying daily physical behavior in older adults: The HAR70 + model. *Sens.* **23**(5), 2368. https://doi.org/10.3390/s23052368 (2023).
24. Duan, A., Hu, B., Liu, W. & Song, J. Motion capture for sporting events based on graph convolutional neural networks and single target pose estimation algorithms. *Appl. Sci.* **13**(13), 7611. https://doi.org/10.3390/app13137611 (2023).
25. Yang, H., Li, Z. & Guo, X. A novel source code clone detection method based on Dual-GCN and IVHFS. *Electron.* **12**(6), 1315. https://doi.org/10.3390/electronics12061315 (2023).
26. Chen, J., Shi, J. & Xu, R. Dual subspace manifold learning based on GCN for intensity-invariant facial expression recognition. *Pattern. Recogn.* **148**, 110157. https://doi.org/10.1016/j.patcog.2023.110157 (2024).
27. Alchihabi, A. & Guo, Y. Learning robust graph neural networks with limited supervision. *Int. Conf. Artif. Intell. Statist.* (2023).
28. Xu, A., Wang, X., Liu, C., Liu, Z. & Kang, L. Dual-channel spatio-temporal wavelet transform graph neural network for traffic forecasting. In *2023 Int. Joint Conf. Neural Netw.* https://doi.org/10.1109/IJCNN54540.2023.10191900 (2023).
29. Zhang, Y., He, X., Chan, Y. H., Teng, Q. & Rajapakse, J. C. Multi-modal graph neural network for early diagnosis of Alzheimer's disease from sMRI and PET scans. *Comput. Biol. Med.* **164**, 107328. https://doi.org/10.1016/j.compbiomed.2023.107328 (2023).
30. Lukas, F. et al. Graphitic carbon nitride as a photocatalyst for decarboxylative C (sp2)–C (sp3) couplings via nickel catalysis. *Angew. Chem. Int. Ed.* https://doi.org/10.1002/anie.202405902 (2024).
31. Liu, X. & He, Y. Power transformer fault diagnosis using dynamic multiscale graph modeling and M2SGCN network based on statistical fusion. *Meas. Sci. Technol.* https://doi.org/10.1088/1361-6501/ad3308 (2024).
32. Grubbs, J. B. & Kraus, S. W. Sports wagering in the context of addictive disorders: Results from a census-matched US sample. *Cogent Ment. Health* https://doi.org/10.1080/28324765.2023.2231497 (2023).
33. Seçkin, A. Ç., Ateş, B. & Seçkin, M. Review on wearable technology in sports: Concepts, challenges and opportunities. *Appl. Sci.* **13**(18), 10399. https://doi.org/10.3390/app131810399 (2023).
34. Martín-Rodríguez, A. et al. Sporting mind: The interplay of physical activity and psychological health. *Sport.* **12**(1), 37. https://doi.org/10.3390/sports12010037 (2024).
35. Huesca-Flores, S., Benitez-Garcia, G., Nakano-Miyatake, M. & Takahashi, H. Skeleton-based baseball pitch classification on broadcast videos. In *Int. Workshop Adv. Imaging Technol. (IWAIT)*, Vol. 13510, 158–163, https://doi.org/10.1117/12.3057901 (SPIE, 2025).
36. Cheng, C. & Xu, H. A 3D motion image recognition model based on 3D CNN-GRU model and attention mechanism. *Image Vis. Comput.* **146**, 104991–104992. https://doi.org/10.1016/j.imavis.2024.104991 (2024).
37. Rohitha, K., Supriya, K., Kuchoor, S. K. & Bodapati, J. D. Optimizing deep convolutional neural networks with progressive unfreezing for enhanced sports activity recognition. In *2024 Int. Conf. Integr. Intell. Commun. Syst. (ICIICS)*, 1–6 https://doi.org/10.1109/ICIICS63763.2024.10859751 (IEEE, 2024).
38. Zaidi, Z. & Gombolay, M. Multi-camera asynchronous ball localization and trajectory prediction with factor graphs and human poses. In *2024 IEEE Int. Conf. Robot. Autom. (ICRA)*, 13695–13702 https://doi.org/10.1109/ICRA57147.2024.10610631 (IEEE, 2024).
39. Zeng, W., He, J., Ren, Z., Ao, C. & Ma, T. Cross-regional crop identification using the hypothesis testing distribution method. In *EGU General Assembly Conf. Abstracts*, 14085–14086 https://doi.org/10.5194/egusphere-egu24-14085 (2024).
40. Jiao, L. et al. 6GAI: active IPv6 address generation via adversarial training with leaked information. In *27th Int. Conf. Comput. Supported Cooperative Work Design (CSCWD)*, 1079–1085 https://doi.org/10.1109/CSCWD61410.2024.10580769 (IEEE, 2024).

41. Pei, Y. et al. ARGNet: using deep neural networks for robust identification and classification of antibiotic resistance genes from sequences. *Microbiome* **12**, 84–85. https://doi.org/10.1186/s40168-024-01805-0 (2024).
42. Guo, Z., Kan, H., Zhang, J. & Li, Y. Neuromorphic visual computing with ZnMgO QDs-based UV-responsive optoelectronic synaptic devices for image encryption and recognition. *Small* **21**, 2412531–2412532. https://doi.org/10.1002/smll.202412531 (2025).

Author contributions

J.L. processed the numerical attribute linear programming of communication big data, and the mutual information feature quantity of communication big data numerical attribute was extracted by the cloud extended distributed feature fitting method. J.L. combined with fuzzy C-means clustering and linear regression analysis, the statistical analysis of big data numerical attribute feature information was carried out, and the associated attribute sample set of communication big data numerical attribute cloud grid distribution was constructed. J.L. did the experiments, recorded data, and created manuscripts. All authors read and approved the final manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025