Occlusive Basketball Player Detection Using Posture-Guided Feature Alignment and Deep Learning

Jiqiao He, Jiahong Luo*, Chao Fu
Physical Education Institute, Xinyu University Xinyu 338000, China
E-mail: hhjjqq112233@126.com, 15180029966@163.com, fuchao0785@126.com
*Corresponding author

Keywords: player detection, occlusion, re-identification, motion trajectory, posture recognition, appearance features

Received: July 17, 2024

In response to the problems of fast transition between offense and defense and severe occlusion in basketball, this study focuses on complex occlusion video scenes in basketball. Firstly, a pre-constructed fully convolutional encoding decoding network is used to extract player features, and a basketball player detection algorithm based on adaptive keypoint heatmap is proposed. Then, by combining global and local features for athlete deep appearance feature extraction, useful features are separated from occlusion using posture information, and a posture recognition-assisted feature alignment athlete re-identification algorithm is proposed. Finally, a basketball player trajectory prediction algorithm under complex occlusion conditions is proposed by integrating athlete spatial information and depth appearance features. The experimental results show that the athlete detection algorithm proposed in the study can improve accuracy to a certain extent on three different backbone networks, with accuracy improvements of 3.6%, 2.5%, and 2.9% compared to traditional algorithms. The highest processing speed of the algorithm studied on DLA-34 is 23.4 frames per second, which is significantly improved compared to other algorithms. The athlete re-identification algorithm achieves 0.7851 on the rank-1 index, with an average accuracy of 0.445. The cumulative matching curve results indicate that the re-identification algorithm has the highest recognition accuracy in severely occluded environments. In the performance test results of the occlusion basketball player detection algorithm based on spatial information and deep appearance feature fusion, it is found that MOTR still lags behind the research algorithm by 0.027 on HOTA. The correlation order of "1+2+3" in the research design achieves the best results in HOTA, DetA, AsSA, MOTA, and IDF1 indicators, with values of 0.510, 0.746, 0.348, 0.872, and 0.557, respectively. The research results have important application value for predicting the motion trajectory of basketball players in complex occlusion scenes.

Povzetek: Raziskava uvaja metodo za zaznavanje zakritih košarkarjev z uporabo poravnave značilnosti na osnovi drže in globokega učenja, kar poveča prepoznavanje gibanja.

1 Introduction

With the continuous improvement of China's comprehensive national strength, China is moving from a sports powerhouse to a sports powerhouse. Adopting technological means for tactical analysis, diagnosis, and scientific training is currently the main method of physical fitness training in China [1-2]. The traditional method mainly relies on the experience and visual observation of instructors to guide athletes, enabling them to understand the key points through repeated practice. This is an urgent problem that needs to be solved at present [3-4]. In recent years, video analysis methods based on image processing have gradually become well-known, which can help people better understand and master the competitive state [5]. Player detection is the foundation of image re identification and tracking, which requires detecting all athletes in the image and determining their types and locations, unlike general pedestrian detection. In a sports environment, multiple people work together to move. There is obvious occlusion and high-speed movement between each other. How to accurately identify target athletes in severely occluded environments has become a research hotspot in current competition video analysis [6-7]. Long et al. addressed the difficulty of tracking player postures on the field and utilized an improved neural network to extract trajectory features of football players in match videos. This algorithm achieved good results in the field of football, with an accuracy rate of over 90%, which was superior to traditional methods [8]. Hao et al. proposed a detection and recognition algorithm for basketball player movements based on visual images and Harris corner extraction. This method improved the player detection and recognition efficiency, had good performance, and had certain practical effects [9]. In the context of sports, athletes often experience the phenomenon of many targets disappearing from the screen and then returning to the screen during high-speed sports. Therefore, it should re-identify the same athlete within the same frame. However, in complex backgrounds, it is difficult to

accurately obtain the overall features of moving targets, which poses great challenges to feature matching. The difficulty is that the clothing colors of the participating athletes are similar. There is significant occlusion between them. Therefore, it should achieve recognition of different athletes through fine-grained information. Zhang et al. developed a wearable inertial sensor for athlete motion capture system based on BP neural network to address the athlete motion state detection. A sensor-based attitude angle setting and action recognition system was constructed. This algorithm was effective for athlete state detection [10]. The core problem of the sports trajectory prediction method for athletes is to extract all athletes from a period of time through a series of sports videos. Then, these athletes' movement characteristics are linked to their appearance features, thereby obtaining their movement trajectories. Traditional methods suffer from low accuracy and long-time consumption in tracking facial images of aerobics athletes. Yang et al. proposed a facial feature tracking algorithm based on Kalman filtering and mean shift. The tracking accuracy of this method was 97%. The shortest tracking time was about 1.5 s [11]. Zhang et al. proposed a computer vision-based technique to track the movement trajectory of athletes to improve their athletic performance. This algorithm reduced trajectory prediction errors and improved operational efficiency [12]. The research achievements of previous scholars are summarized, as shown in Table 1.

In summary, this study proposes a video analysis method based on deep learning. The current research on video analysis mainly focuses on the individual or individual behavior of multiple individuals. However, in competitions such as basketball and football, athletes cooperate with each other and are not completely independent, which increases the video analysis difficulty. Therefore, this study conducts in-depth research on predicting the

Table 1: Summary of the achievements of previous scholars on models related to target detection.

Scholar	Methodology	Key results	Limitations
Long et al. [8]	The improved neural network was used to extract the trajectory characteristics of the players in the game video of football players	The algorithm has achieved good results in football, with an accuracy of more than 90%, better than the traditional convolutional neural networks	The algorithm has a narrow application field, and although the accuracy is improved, the running time has increased
Hao et al. [9]	Based on the visual image and the Harris corner point extraction algorithm	It is found that the algorithm performs well and has a certain practical effect	The accuracy of the algorithm is not further improved
Zhang et al. [10]	Based on BP neural network algorithm, the sensor of athlete motion capture system based on wearable inertia is developed to build the attitude angle setting and motion recognition based on the sensor system	The proposed algorithm is effective for the state detection of the athletes	Improved the dynamic identification effect, but also improved the running time
Yang et al. [11]	Face feature tracking algorithm based on Kalman filtering and mean drift	It solves the problem of low accuracy and long time consuming of face image tracking of aerobics athletes	The algorithm has a poor timeliness
Zhang et al. [12]	Computer vision-based techniques to track the movements of athletes	The proposed algorithm can reduce the trajectory prediction error and improve the operational efficiency	Dynamic recognition ability and algorithm accuracy need to be improved

movement trajectory of basketball players in severely occluded environments by integrating spatial information and deep appearance features.

It is hoped to provide convenient and fast training methods for athletes. There are three main points of research innovation. Firstly, a basketball player detection algorithm in complex occlusion environments is proposed to improve basketball players' detection performance in the presence of occlusion. Secondly, a basketball player re-identification algorithm based on posture

recognition-assisted feature alignment is designed. This can achieve precise separation of basketball player specific appearance information under occlusion. Meanwhile, this can enhance the feature re recognition of basketball players in densely occluded sports scenarios. Finally, an occlusive basketball player detection algorithm based on posture recognition assisted feature alignment is constructed. It is hoped to solve the real-time prediction of basketball player's movement trajectory in severely occluded environments.

2 Methods and materials

2.1 Occlusive basketball player detection algorithm based on adaptive key point heat maps

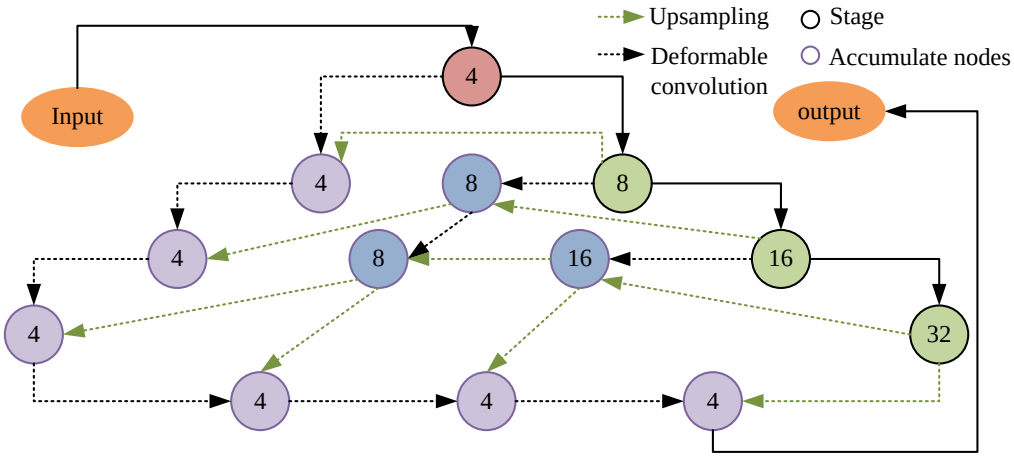


Figure 1: Overall structure of improved DLA-34 feature extraction network.

In the basketball game background with strong occlusion features, the temperature distribution characteristics of feature points are utilized to extract the target position within the target area. On this basis, ResNet, DLA, and Hourglass backbone networks are introduced. Firstly, deformable convolution is combined with up sampling layering to enhance the perception field calculation accuracy through deformable convolution. Then, the target area's accurate restoration is achieved by iteratively convolving the up sampled data. The combination of these two improves spatial resolution [13-15]. Figure 1 shows the overall structure of the improved DLA-34 feature extraction network.

The input image consists of 2 consecutive Hourglass overlays and 4 down samples. Before up sampling, this experiment introduces deformation convolution to make the perception field more accurate. Then, inverse convolution is utilized to up sample the signal, thereby improving the output image's resolution. Although three different feature extraction networks are used, the results still consist of three parts: heat map, height, and regression network. The heat map sub-network's output represents the Gaussian distribution of the object's center points. The width and height sub-network analyze the height and width of the input end. The regression sub-network's output is aligned with the bias estimation [16-18]. It is assumed that a heat map of a key point is obtained during the image prediction. For a certain category, its center point coordinates can be obtained through its coordinates in the upper left and lower right. Then, the center point coordinates can be sampled downwards. On this basis, a center point is drawn on the feature map using Gaussian kernel function, represented by formula (1).

$$Y = \exp\left(-\frac{(x-x_0)^2}{2\varepsilon_x^2} - \frac{(y-y_0)^2}{2\varepsilon_y^2}\right) \quad (1)$$

In formula (1), $\varepsilon_x = \frac{\mu w}{6}$. $\varepsilon_y = \frac{\mu h}{6}$. ε refers to the parameters of the Gaussian kernel function. w and h correspond to the input image's width and height. μ refers to the scaling factor that adjusts the Gaussian kernel function's rendering range. The point at $Y=1$ is the target center point. The designed image reconstruction algorithm based on Gaussian model has the characteristic of changing with parameters such as image width and height. It can draw images based on the size of the object, thereby reducing the blurriness in the image. The overall loss function of the improved DLA-34 feature extraction network consists of a weighted sum of three parts, represented by formula (2).

$$L = L_k + w_{size} L_{size} + w_{off} L_{off} \quad (2)$$

In formula (2), L_{size} refers to the target width and

height loss function. w_{size} refers to the target's height loss weight. L_k refers to the center point position loss function. w_{off} refers to the center point offset loss's weight. L_{off} refers to the center point offset loss.

2.2 Basketball player re-identification algorithm based on posture recognition assisted feature alignment

In player detection, there are shortcomings in target tracking methods that rely solely on continuous motion space. There are issues such as identity conversion and trajectory reset that are prone to occur in target recognition. Therefore, the posture-based target tracking method achieves effective separation between target features and occlusions. Matching and association are performed on athletes who reappear after occlusion to improve the basketball player detecting accuracy in occluded scenes. This algorithm includes two parts: local feature branch and pose guided global feature branch. For the posture guided global feature branch, the study adopts an improved ResNet50 with moderate network depth as the backbone network to extract global feature information in Figure 2.

The study optimizes the traditional ResNet50. The average pooling layer and fully connected layer in the backbone network are removed. A convolution operation is performed at level 4. The $W \times H$ image is imported into the improved ResNet50, resulting in a feature map $\frac{W}{16} \times \frac{H}{16}$ with a larger spatial dimension, labeled as $F \in R^{w \times h \times c}$. w refers to the output feature map's width. h refers to the output feature map's height. c refers to the channels number. In terms of pose acquisition, this study utilizes the COCO-based human pose estimator Alphapose to obtain 18 joint points and generate corresponding coordinate $LM_j = (cx_j, cy_j)$ and confidence score for each joint. On this basis, a new method for detecting human key points is proposed. By setting a pose critical point threshold γ, this model is filtered to obtain human key points with lower credibility. After processing with Gaussian kernel function, a two-dimensional Gaussian heat map M_j can be obtained. Due to the presence of multiple pose critical points in the target box, this study aims to identify target players by identifying the key points for each pose. From this, the overall process of the branch algorithm is obtained in Figure 3.

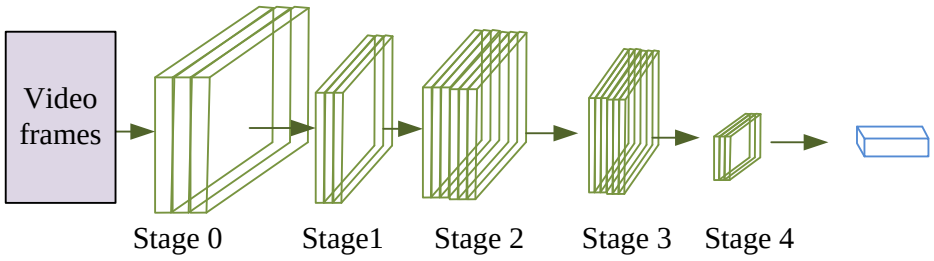


Figure 2: Improved ResNet50 structure.

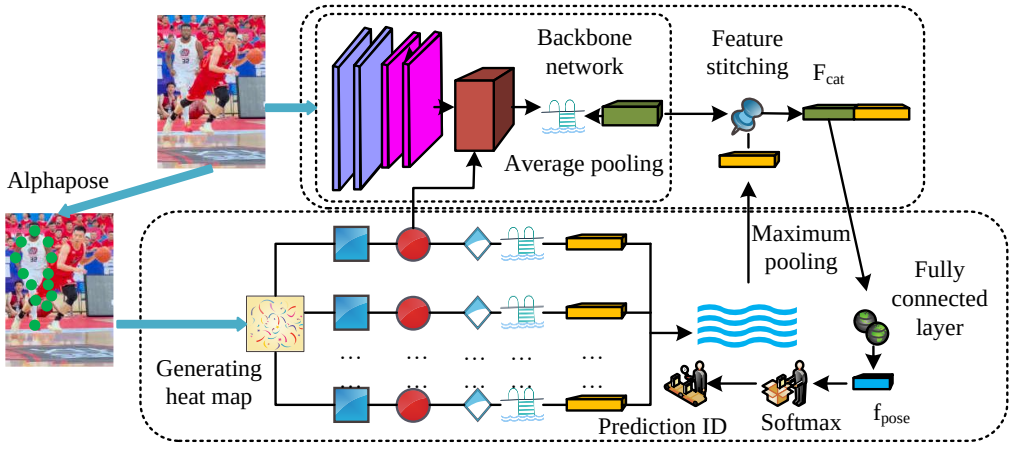


Figure 3: Posture guided global feature branching process.

Firstly, the global shared feature map F is averaged and pooled to obtain the global feature vector f_g. Then, F is multiplied with M_j to obtain the pose guided feature map $M_j^{'}$ of the target athlete. Secondly, after average pooling of each $M_j^{'}$, a feature vector with a dimension of 2048 is obtained. All vectors are passed through the max pooling layer and then concatenated with f_g to obtain the feature vector f_{cat}. Then, f_{cat} is input into the fully connected layer, reducing its dimension to 256, resulting in the global feature vector f_{pose}. Finally, each input image's predicted positions are obtained through softmax. The loss function utilized this time is the cross-entropy function $LOSS_c$, represented by formula (3).

$$LOSS_c\left(\bar{y}, y\right) = -\sum_{n=1}^{N_b}\sum_{r}^{C} y^n \log_2 \bar{y}_r^n \tag{3}$$

In formula (3), y refers to the true position. y^n refers to the n th target's true position. $\bar{y}$ refers to the position prediction probability value corresponding to the global feature branch guided by posture. N_b refers to the data amount in a batch. C refers to the total categories. $\bar{y}_r^n$ refers to the probability that the r th target's predicted position is category r. Figure 4 shows the local feature branching process.

In this branch, F is horizontally divided into p parts. Then, the parts processed by the average pooling layer are transformed into partial feature vectors f_i. After being processed by the fully connected layer and softmax layer, f_i can obtain the predicted position of its local features [19-20]. The local feature branch loss function is represented by formula (4).

$$L_{part} = \sum_{i=1}^{p} LOSS_c\left(\bar{y}_i, y\right) \tag{4}$$

In formula (4), p refers to the quantity of parts divided. The total loss function is the proportional weighted sum of these two branches' loss functions, expressed by formula (5).

$$L_{tol} = \eta L_{part} + (1-\eta) L_{pose} \tag{5}$$

In formula (5), η determines the loss function proportion of the global and local feature branches guided by pose in the overall loss function. The depth appearance features of different frames need to be correlated after

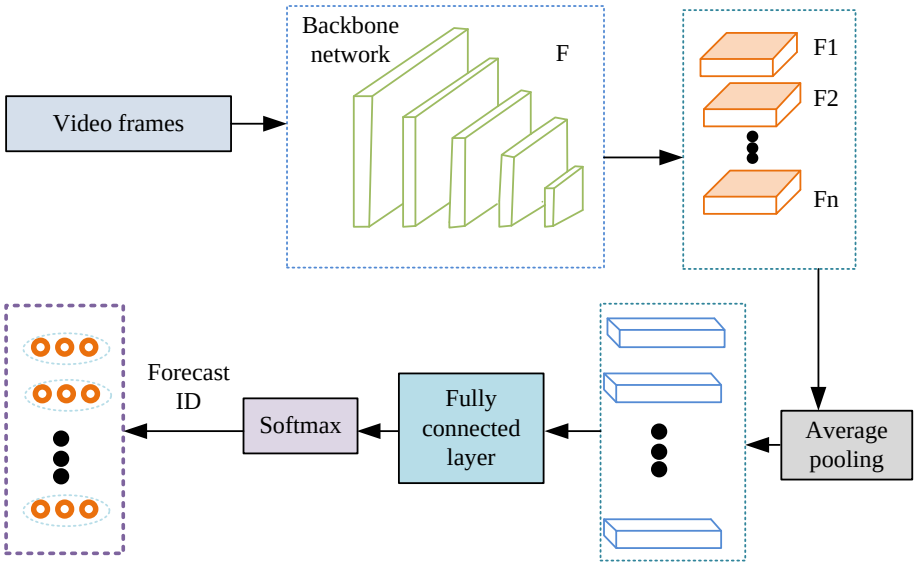


Figure 4: Local feature branch process.

being extracted. Therefore, this study adopts cosine distance to calculate depth appearance features' similarity among different athletes. This can measure the differences between features and match the same basketball players, represented by formula (6).

$$\cos\beta = \frac{x_1 x_2 + y_1 y_2}{\sqrt{x_1^2 + y_1^2} \bullet \sqrt{x_2^2 + y_2^2}} \tag{6}$$

In formula (6), the cosine distance measurement method is denoted as AB. The angle between vectors (x_1, y_1) and (x_2, y_2) is $\cos\beta$. The overall feature distance h guided by posture is represented by formula (7).

$$h = AB\left(f^q, f^g\right) \tag{7}$$

In formula (7), f^q refers to the query image's vector in the overall feature branch guided by pose. f^g refers to candidate images' vector in the overall feature branch guided by pose. However, this algorithm is related to local feature branches. The calculation results of similar distances only show the unobstructed parts. For each part of i, the distance is denoted as h_i and expressed using formula (8) [21-22].

$$h_i = \left(l_i^q \cdot l_i^g\right) AB\left(f_i^q, f_i^g\right) \tag{8}$$

In formula (8), l_i^q refers to the corresponding part's value of the query image. l_i^g refers to the corresponding part's value of the candidate image. f_i^q refers to the query image's vector in the local feature branch. f_i^g refers to the candidate image's vector in the local feature branch. The final distance $hist$ required for the research algorithm is calculated and expressed using formula (9).

$$hist = \frac{\sum_{i=1}^{p} h_i + h}{\sum_{i=1}^{p}\left(l_i^q \cdot l_i^g\right) + 1} \tag{9}$$

In formula (9), p refers to the quantity of unobstructed parts. $hist$ is positively correlated with query and candidate images' similarity.

2.3 Occlusive basketball player detection algorithm based on spatial information and deep appearance feature fusion

To further predict the movement trajectory of basketball players in severely occluded environments, this study delves into the temporal connections of basketball players. Based on the spatial information and depth appearance features of athletes obtained from the previous text, Figure 5 shows the overall process of the designed algorithm.

Firstly, by training one frame of each image, the region and range of the participating athletes in each image are obtained. Then, the image data of each athlete are imported into the re-identification network. The overall and partial images of each athlete are obtained separately. On this basis, three motion predictions are made for the trajectory fragments. It is associated with athletes in the trajectory fragments three times, generating a subsequent trajectory fragment. After each completion of the correlation, the trajectory is updated. The next image is processed and repeated until the process is completed [23-25]. The study combines Kalman filtering to establish a real-time motion trajectory prediction model for basketball players in Figure 6.

In formula (10), x refers to the average of the athlete's previous moment. x' refers to the average of the athlete at that moment. D refers to the state transition matrix. The prediction of athlete covariance is represented by formula (11).

$$M' = DMD^T + Q \tag{11}$$

In formula (11), M' refers to the athlete's covariance at the predicted time. M refers to the athlete's covariance at the previous moment. Q refers to the entire system's noise matrix. When updating, based on the real-time detection of athletes, the corresponding sports status is updated to obtain more accurate sports status. Firstly, the average error of the motion trajectory segment and the currently detected athletes is calculated using formula (12).

$$y = z - Hx' \tag{12}$$

In formula (12), H refers to a measurement matrix. z and x' correspond to the current detected athlete and the athlete's motion path segment's average vectors. The vector S is obtained by combining and expressed using formula (13).

$$S = HM'H^T + R \tag{13}$$

In formula (13), R represents the noise matrix of the detection space. Then the Kalman gain K can be calculated using formula (14).

$$K = M'H^T S^{-1} \tag{14}$$

Finally, the updated mean vector x and covariance matrix M can be obtained, represented by formula (15).

$$\begin{cases} x = x' + Ky \\ P = \left(1 - KH\right)P' \end{cases} \tag{15}$$

The study also proposes a method of triple data association to achieve global information correlation between motion trajectories and object detection boxes. This can improve the prediction accuracy and robustness of athletes' movement paths under occlusion. The Hungarian algorithm ensures that each trajectory fragment can only correspond to at most one trajectory with the current detection target at the lowest cost. A cost matrix is constructed between the target and trajectory fragments to achieve the fusion of trajectory fragments and target trajectories. Figure 7 shows the data association algorithm.

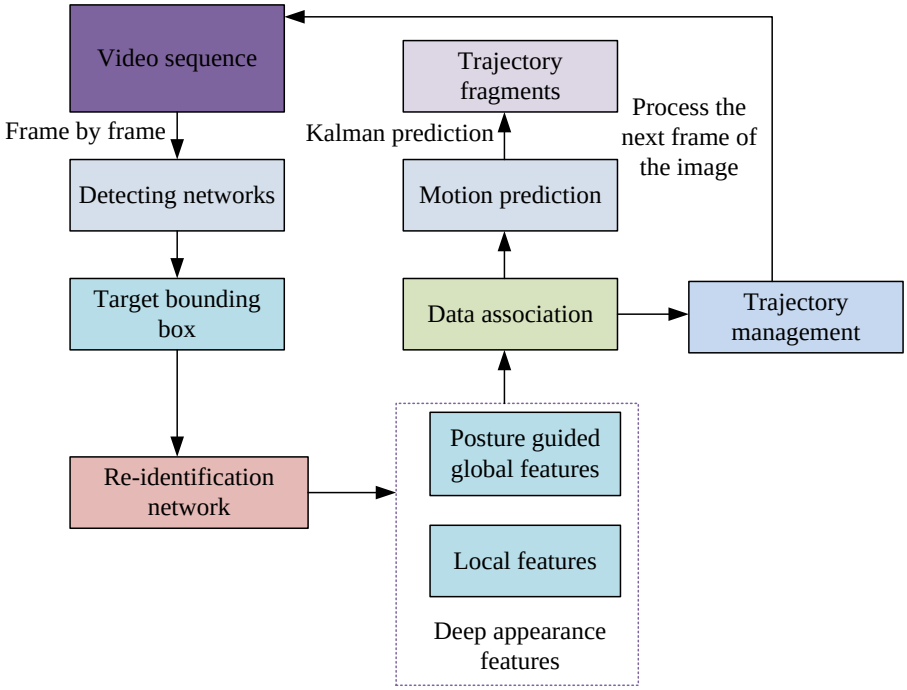


Figure 5: Player time-series association algorithm.

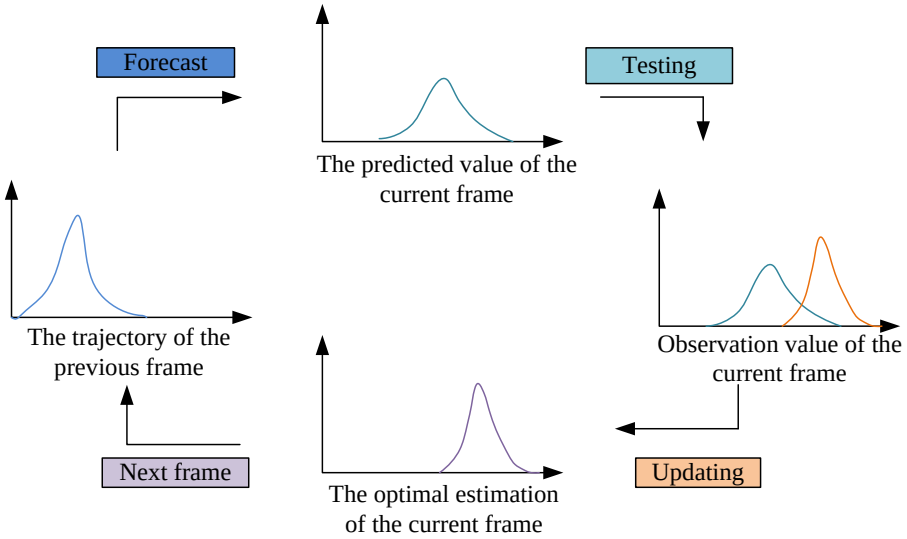


Figure 6: Motion trajectory prediction model.

For each frame of the video, a new detection method can be utilized to obtain an area that includes all participating athletes. In this framework, two credibility thresholds are set, high and low. If the credibility score is higher than the high credibility threshold, it will be treated as a high score box. On the contrary, it is a low score box. Players in low scoring boxes have strong occlusion. To reduce the impact of this error, only depth table appearance features are extracted from images in high score boxes. The algorithm first utilizes a Kalman filter to estimate the motion trajectories between frames. Then, after three rounds of correlation, the Kalman filter algorithm is utilized to correct the motion trajectories, ultimately achieving tracking of

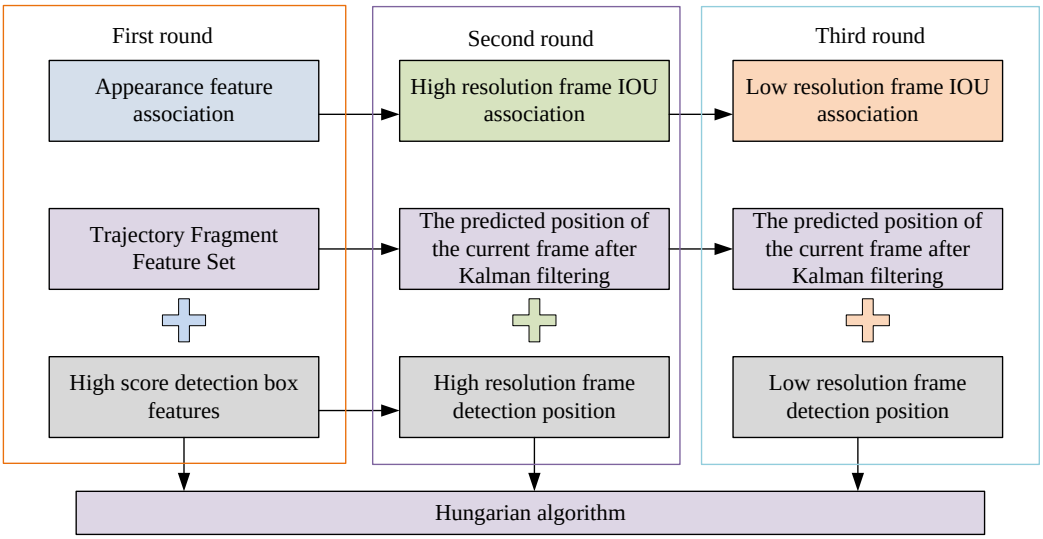


Figure 7: Data association algorithm process.

Table 2: Experimental operating environment and parameter settings.

Experimental parameters	Settings
Graphics card environment	Tesla V100 GPU
Operating system	Ubuntu16.04
Experimental framework	PyTorch
Python version	3.6.2，CUDA 10.0，CUDNN 7.6
Batch size	16
Total training cycle	300epoch、60epoch
Optimizer	Adam
Key point threshold	0.75
Loss weight	wsize=0.1，woff=1
Gaussian kernel scaling factor	1.15

moving targets.

3 Results

3.1 Experimental parameter settings

The study was conducted using a TeslaV100GPU graphics processor as a platform and a storage space of 32 GB. Based on the PyTorch architecture, Ubuntu 16.04 was utilized as the testing platform. In the performance testing experiments of player detection algorithm, player re-identification algorithm, and athlete trajectory prediction algorithm, the input image sizes of the model were 1920×1080, 384×128, and 1920×1080, respectively. Three datasets suitable for researching algorithms were constructed for the practical application scenarios of basketball. Experimental verification was conducted on the above research content. Table 2 shows the operating environment and parameters for this experiment.

3.2 Performance testing of occlusive basketball player detection algorithm

An evaluation was conducted on the impact of player detection algorithms, using test data collected from a basketball game video developed by the research team. LabelImg image labeling software was utilized to label the data, resulting in 5141 images. After data augmentation, 7958 images were obtained, including 5049 images in the training set, 1552 images in the validation set, and 1357 images in the testing set, respectively. The study utilized the weights from the established large-scale public transportation flow database COCO as the initial weights and trained the overall model with updated weights. Figure 8 shows the basketball player detection results based on multiple feature extraction networks such as ResNet, DLA, Hourglass, etc.

In Figure 8 (a), compared with the conventional critical point heatmap drawing method, the three backbone networks utilized improved the athlete monitoring accuracy. In the Hourglass-104 backbone network, the Average Precision (AP) reached 58.2%, ResNet-18 backbone network reached 46.8%, and DLA-34 backbone network reached 53.7%, which was 3.6%, 2.5%, and 2.9% higher than conventional methods. In Figure 8 (b), the proposed method reached better AP than other methods on backbone networks based on DLA-34 or Hourglass-104. This method's maximum processing speed on the DLA-34 platform reached 23.4 frames per second, which was a significant improvement

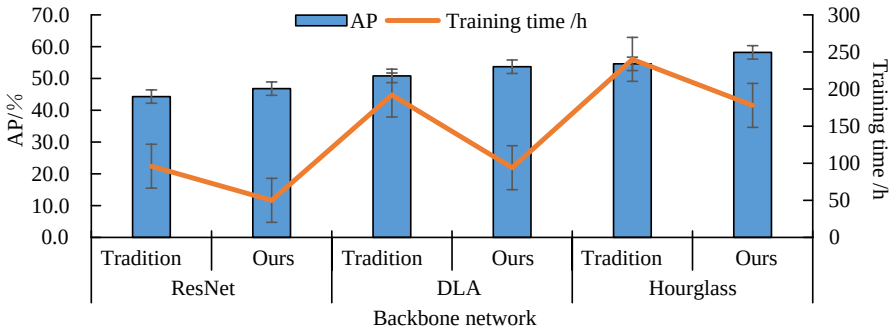
compared to YOLOv3 and RetinaNet. The network structure of Hourglass-104 was relatively large, and its processing speed was slightly lower than DLA-34. However, its computational efficiency was still higher than object monitoring methods such as CornerNet and ExtremeNet that utilized Hourglass-104 as the backbone network. In the ResNet-101-FPN backbone network experiment, the proposed method achieved AP of 65.42%, which was 5.7% and 1.5% higher than FasterR-CNN and FCOS. On the ResNet-101-FPN backbone network, the proposed method had a processing speed of 12.78 frames per second, which was significantly improved compared to methods such as FasterR-CNN and FCOS. Figure 9 shows the iteration situation and loss function during the training of the improved algorithm.

In Figure 9, compared with the traditional algorithm, the improved algorithm converged more quickly on the bounding box regression loss function, classification loss function, and target probability loss function. The initial decline speed of the network training was faster. The calculated loss value after improvement was smaller compared to the original model. The improved model reached the iteration inflection point faster and tended to stabilize after about 225 iterations. After data analysis and comparison with other algorithms, the improved adaptive keypoint heatmap occlusive basketball player detection algorithm had better detection performance. It accelerated video frame processing time, improved algorithmic accuracy, and improved algorithmic performance.

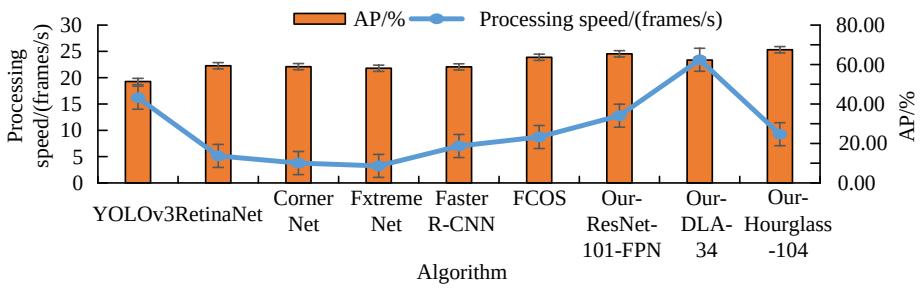
3.3 Basketball player re-identification algorithm based on posture recognition assisted feature alignment

To evaluate the player re-identification algorithm's effectiveness, a basketball player re-identification dataset was constructed for subsequent training and experiments. This dataset selected five games, each with ten starting players and one to two substitute players. There were a total of 56 athlete IDs and 9736 labeled images. To address overfitting, the study utilized methods such as random erasure, random blocks, and Cutout to expand existing labeled image data and obtained 1476 player re-identification samples. 26 athlete images were utilized as training samples, 10 athlete images were utilized as test samples, and the remaining 20 athlete images were utilized as detection samples. The detection set was divided into candidate images and search images, with 1-3 images taken from each athlete, and the remaining images were candidate images.

Before training, first, the alpha Persian matrix on COCO samples was utilized to calculate the pose and heatmap distribution of each sample. Subsequently, the threshold values of attitude key points, the selection of



(a) Different feature extraction network detection results



(b) Different algorithm detection results

Figure 8: Performance test results of player detection algorithm.

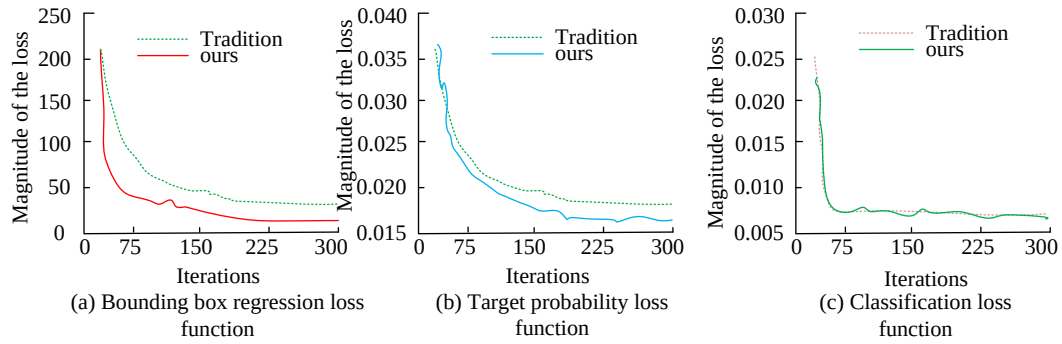


Figure 9: Comparison results of loss functions before and after improvement of player detection algorithm.

local feature parts, and the values of loss weights were all validated on the validation set in Figure 10.

In Figure 10 (a), a threshold value that was too large or too small led to poor prediction performance. When the critical point threshold was high, most of the attitude key points were abandoned, which reduced the overall algorithmic performance. When the attitude critical point's threshold was 0.2, this method performed the best. In Figure 10 (b), the fine-grained nature of local characteristics was determined by the quantity of partitions. In the case of a local number of 1, the learned characteristics were the object's overall characteristics. The experimental results indicated that it was very important to study the local characteristics of segmentation during the segmentation process. Due to this model's best performance when the partitions were 5, the study set the quantity of partitions to 5. In Figure 10 (c), the loss weight was utilized to measure the effect of two feature extraction branches. When the loss weight was 0, only the pose was utilized to guide the global feature branch. When the loss weight was 1, only local characteristic branches were utilized. When using two different branch characteristics, the prediction performance was better than a single branch. Due to the best performance of the model with a loss weight of 0.2, the study chose a loss weight of 0.2. To see the superiority of the research model more intuitively, the research visually compared the precision rate, recall rate, Mean Average Precision (mAP) and other data indicators of the traditional model and the research model in the process of training the data set. The results are shown in Figure Figure 11.

From Figure Figure 11, both the precision and recall of the study model stabilized after 51 iterations, while the mAP stabilized after 25 iterations. Both the precision and recall of the traditional algorithm stabilized after 101 iterations, and the mAP stabilized after 51 iterations. Compared with the traditional algorithm, the research model was more efficient and accurate. Figure 12 shows the comparison between the research algorithm and advanced algorithms in the pedestrian re-identification.

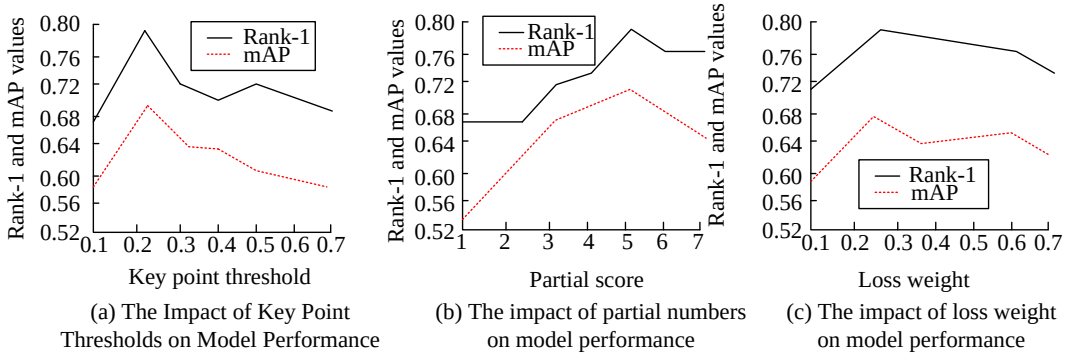


Figure 10: Influence of posture key point threshold, partitions number, and loss weights.

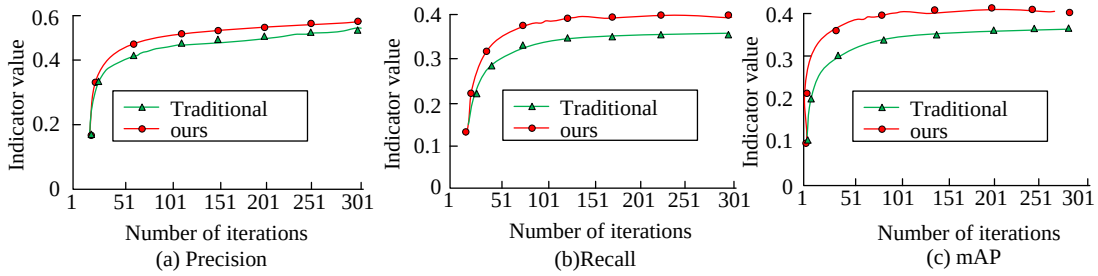


Figure 11: Comparison results of different parameters.

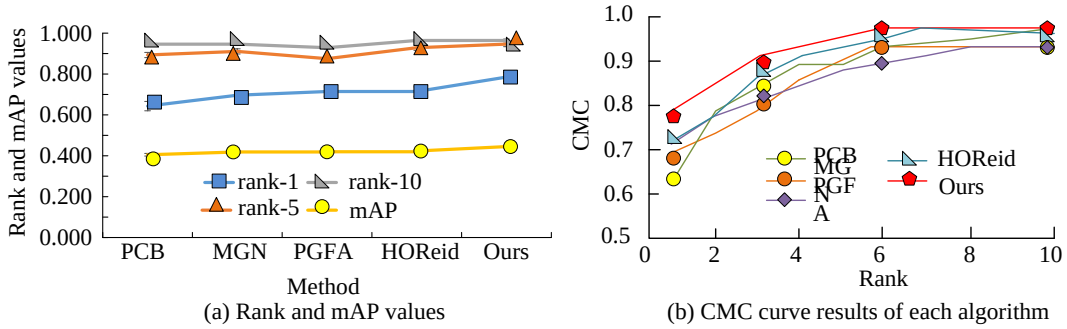


Figure 12: Comparison results of pedestrian re-identification.

In Figure 12, the research method had a rank-k accuracy (rank-1) of 0.7851. However, PCB utilized local features of athletes and divided them into 6 regions, with a rank-1 of 0.643. Compared with other methods, its mAP decreased by 0.445. Compared with PCB that only used partial features, MGN's mAP reached 0.418. PGFA trained other athletes based on their appearance characteristics, resulting in an mAP of only 0. 419. The study focused on basketball games with obstacle targets as the human body and incorporated a criterion for judging multiple critical points in appearance feature matching, which was 0.708 higher than PGFA and 0.255 higher than mAP. Although HOReid exhibited highly efficient performance on public occlusion data, up to 423, its extraction of higher-order semantic features in basketball player re-identification was only slightly improved compared to PGFA. By analyzing the Cumulative Match Characteristic Curve (CMC) of different algorithms, the proposed method outperformed current mainstream re-identification methods in multiple dimensions such as rank-1 to rank-10. It achieved similar appearance feature re-identification performance to SOTA in basketball games.

3.4 Performance testing of occlusive basketball player detection algorithm based on posture recognition-assisted feature alignment

The study evaluated the occlusive basketball player detection algorithm based on posture recognition assisted feature alignment, using 12 basketball games of Chongqing University of Posts and Telecommunications as the research object. Five parts were selected as materials, which were marked according to the format of the MOT database. The remaining seven were processed visually. 4728 frames of 5 labeled images were recorded, including 1 segment for parameter validation, 4 segments for testing. A quantitative evaluation index was utilized. To test the algorithmic effectiveness, training and testing were conducted on the public multi-objective tracking database DanceTrack. These results were compared with existing experimental results. The sample of DanceTrack includes 100 videos, including 40 training segments, 25 validation segments, and 35 testing segments. The study evaluated the basketball player tracking algorithm using five evaluation indicators: HOTA, DetA, AssA, MOTA, and IDF15.

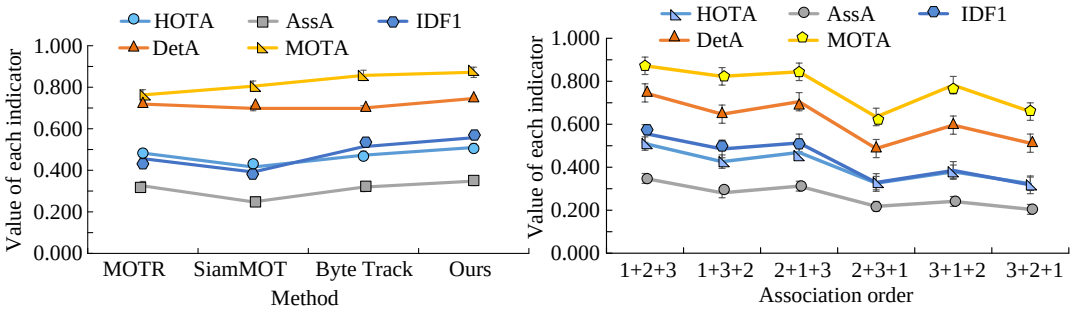
An occlusive basketball player detection algorithm based on the fusion of spatial information and deep appearance features was tested. The study selected the most mature method in current multi-objective tracking technology as the control experiment. According to the original experimental conditions, retrain and test in the basketball player tracking database. In addition, the experiment verified whether the correlation order between the three data points would have an impact on the tracking effect. Therefore, the correlation between each data was compared in Figure 13.

In Figure 13(a), MOTR had an AssA performance of 0.763, which effectively reduced the ID conversion, but it was still 0.027 worse than existing methods. Due to

SiamMOT being a real-time based, more precise local correlation method, its AssA performance was only 0.248. ByteTrack was implemented through two steps: one was to perform IoU correlation on high bounding boxes. The other was to perform IoU correlation on low classification frameworks for trajectories that do not match. In IDF1, the performance of restoring missing objects and reducing ID conversion rate was close to the existing research results of 0.514. However, this method did not consider apparent features and relied only on action information, which was insufficient. In Figure 13 (b), "1", "2", and "3" correspond to appearance feature associations, high bounding box IoU associations, and low bounding box IoU associations. The designed correlation sequence of "1+2+3" achieved the optimal results in various indicators such as HOTA, DetA, AssA, MOTA, and IDF1, with values of 0.510, 0.746, 0.348, 0.872, and 0.557, respectively. For the association order of "1+3+2", due to the irregularity of the motion, a low score framework IoU correlation method was adopted to introduce a large amount of background information, resulting in a decrease in tracking accuracy. The HOTA was only 0.426, a decrease of 0.084 compared to existing research results. For the correlation order of "2+1+3" and "3+1+2", compared with the research algorithm, the AssA index was reduced by 3.6%. Low frame IoU correlation was first performed. Then, appearance feature correlation was performed on unmatched information. Compared with the proposed correlation order, the AssA index was reduced by 10.7%. Overall, the proposed algorithm was effective and superior to traditional algorithms for predicting the movement trajectory of basketball players under complex occlusion conditions. This proved the set "1+2+3" association order's necessity and effectiveness. The experiment further demonstrated the occlusive basketball player detection algorithm's generalization ability based on posture recognition assisted feature alignment. The proposed method's effectiveness was verified through experiments using the current DanceTrack-based multi-target tracking experimental platform in Table 3.

The proposed method performed significantly better than other SOTA methods on the DanceTrack dataset, with detection accuracy of HOTA, DetA, AssA, MOTA, and IDF1 reaching 0.48, 0.736, 0.314, 0.885, and 0.476, respectively. This was significantly higher than the MOTR with the best performance on the HOTA indicator, which was 0.003 higher. The proposed method's AssA was 0.314, which performed the best compared to other methods and effectively reduced the ID conversion. In terms of algorithm complexity index, the performance of time complexity, spatial complexity and Big-O index was 3.1,35 and 19.05, respectively, which were better than the comparison algorithm, and the calculation complexity of the algorithm was significantly reduced. Further enhancing the long-term data correlation of target motion on the DanceTrack dataset could improve this model's tracking performance. Finally, to test the impact of each part of the research algorithm on the final performance, ablation experiments were set up, as denoted in Table 4.

In Table 4, after removing the appearance information module, all indicator values were the lowest, with HOTA, DetA, AssA, MOTA,



(a) Performance comparison of various algorithms

(b) The impact of different data association orders on algorithm performance

Figure 13: Verification of the occlusive basketball player detection algorithm for posture recognition assisted feature alignment.

Table 3: Validation results of generalization ability.

Method	HOTA	DetA	AssA	MOTA	IDF1	Time complexity	Space complexity	Big-O
Center Track	0.418	0.781	0.226	0.868	0.357	15.8	54	34.9
FairMOT	0.397	0.667	0.238	0.822	0.408	7.3	62	34.65
QDTrack	0.457	0.721	0.292	0.830	0.448	7.4	60	33.7
Trans Track	0.455	0.759	0.275	0.884	0.452	5.1	74	39.55
TraDes	0.433	0.745	0.254	0.862	0.412	7.3	56	31.65

MOTR	0.481	0.718	0.227	0.792	0.461	7.9	53	30.45
ByteTrack	0.477	0.710	0.221	0.896	0.539	6.3	71	38.65
Ours	0.484	0.736	0.314	0.885	0.476	3.1	35	19.05

Table 4: Results of ablation experiment.

Experimental setup	HOTA	DetA	AssA	MOTA	IDF1	Weight Size (MB)	Total parameter
Remove appearance information module	0.448	0.702	0.387	0.845	0.437	13.7	7020913
Remove spatial information module	0.491	0.774	0.413	0.876	0.483	9.17	4635537
Remove the low bounding box association module	0.465	0.713	0.405	0.851	0.469	9.2	4747258
Ours	0.521	0.767	0.459	0.903	0.568	6.33	3154806

and IDFA of 0.448, 0.702.0.387, 0.845, and 0.437, respectively. This indicated that when the video frame rate was low, analyzing only through spatial information was unreliable, confirming the significant impact of appearance information on the research algorithm. After removing the spatial information module, only the results between the current frame and the previous frame detection boxes were calculated. In terms of AssA and IDF1, they were 4.8% and 8.6% lower than those with spatial information modules, greatly reducing the tracking effect. After removing the low score box association module, all indicators were significantly reduced, with HOTA, DetA, AssA, MOTA, and IDFA corresponding to 0.465, 0.713, 0.405, 0.851, and 0.469, respectively. After removing the appearance information module, spatial information module and low frame correlation module, the weight size of each parameter were 13.7, 9.17, and 9.2 respectively, and the total amount of each parameter was 7020913, 4635537, and 4747258 respectively, while the weight size and total parameters of the research model were 6.33 and 3154806 respectively, which performed better. In short, the three modules were closely related to the final performance. The absence of any one might obviously affect the final result.

4 Discussion

The study focused on basketball sports scenes with occlusion, and added a judgment condition for multi-person keypoint occlusion in the appearance feature matching strategy. Compared with the PGFA algorithm, it improved the rank-1 index by 7.08% and the mAP index by 2.55%. Chaudhuri et al. conducted experiments on multiple pedestrian re-identification datasets and found that the selected evaluation metrics had high accuracy, making them widely recognized as the most powerful open-source library for target re-identification. However, it performed the worst on the basketball player dataset, with a mAP of only 35.89%. This is due to the lack of algorithm processing specifically for occlusion, which indirectly confirms that the basketball player dataset is more difficult to verify compared to the pedestrian dataset [31]. He et al. predicted the features of occluded areas based on spatial contextual information of non occluded areas, and the recognition accuracy was second only to the research algorithm. However, when athletes occluded for a long time, the prediction results based on context had great uncertainty, which reduced the accuracy [32].

The SORT model studied by Videira et al. lagged behind other algorithms in tracking performance due to the use of only IoU association strategy and the lack of feature extraction for appearance. Compared with the HOTA algorithm, it decreased by 14.1%, and athlete ID switching was the most frequent. Meanwhile, the model did not specifically process the occluded features and performed well on pedestrian tracking datasets, but performed poorly on basketball player tracking datasets with severe occlusion and similar appearance [33]. Pan et al. explored the use of twin multi-target tracking networks for motion modeling, optimizing the Kalman filter motion model based on simple geometric features in SORT into a more complex twin tracking network. This model learned a matching function to find a matching detection box within a local range. Compared with the SORT algorithm that only used a simple Kalman filter

motion model, it improved AssA indicators by 3.4%. However, due to its online algorithm, it focused more on accurate local correlations rather than global optimal correlations, so its performance was still poor in complex basketball scenes [34]. Wang et al. made breakthroughs and innovations in data association, designing two rounds of data association. Firstly, high bounding box IoU association was performed, and low bounding box IoU association was performed on unmatched trajectories, which can recognize targets under short-term occlusion conditions and reduce ID switching. The IDF1 index reached 51.4%, second only to the research algorithm. However, its model robustness was poor because it did not introduce appearance features for data association and only relied on motion information [35].

5 Conclusion

In the current analysis of basketball game videos, there are problems such as severe occlusion and difficulty in tracking the movements of athletes. The study combines the spatial and deep appearance information of athletes to explore efficient, fast, and intelligent deep learning video analysis methods. A basketball player trajectory prediction algorithm under complex occlusion conditions is proposed. The player detection algorithm improved AP on three different backbone networks, with an increase of 3.6%, 2.5%, and 2.9%, respectively, compared to traditional methods. The research algorithm's maximum processing speed on DLA-34 was 23.4 frames per second, which was significantly improved compared to other algorithms. The player re-identification algorithm achieved 0.7851 on rank-1, with an mAP of 0.445. The CMC results indicated that in severely occluded environments, the re-identification algorithm's recognizing accuracy was the highest. In the occlusive basketball player detection algorithm's performance test based on spatial information and deep appearance feature fusion, MOTR still lagged behind the research algorithm by 0.027 on HOTA. The designed correlation sequence of "1+2+3" achieved the best performance in various indicators such as HOTA, DetA, AssA, MOTA, and IDF1, with values of 0.510, 0.746, 0.348, 0.872, and 0.557, respectively. This proved the set sequence's necessity and effectiveness of three rounds of data association. Therefore, this research algorithm has high prediction accuracy for predicting the movement trajectory of athletes under complex occlusion conditions. However, there are still certain limitations to the research. In future research, the performance and application effects of the algorithm in mobilization discrimination with severe occlusion can be optimized to achieve more fine-grained feature extraction.

References

[1] Matthew T.O. Worsey, Hugo G. Espinosa, Jonathan B. Shepherd, and David V. Thiel. One size doesn't fit all: Supervised machine learning classification in athlete-monitoring. IEEE sensors letters, 5(3):1-4, 2021. https://doi.org/10.1109/LSENS.2021.3060376

[2] Tang Lei, Zhu Cai, and Luo Hua. Training prediction and athlete heart rate measurement based on multi-channel PPG signal and SVM algorithm. Journal of intelligent & fuzzy systems, 40(4):7497-7508, 2021. https://doi.org/10.3233/JIFS-189571

[3] Perumalsamy Gunasekaran, A. Azhagu Jaisudhan Pazhani, and T. Ajith Bosco Raj. A novel method for multiple object detection on road using improved YOLOv2 mode. Informatica, 46(4):7497-7508, 2022. https://doi.org/10.31449/inf.v46i4.3884

[4] Ming Li, Hui Dong, Fei Zhang, and Xiaoxiao Liu. A method for top view pedestrian flow detection based on small target tracking. Informatica, 48(11):3717-3732, 2024. https://doi.org/10.31449/inf.v48i11.6033

[5] Xin Wang, and Junxiang Yang. Marathon athletes number recognition model with compound deep neural network. Signal, image and video processing, 14(7):1379-1386, 2020. https://doi.org/10.1007/s11760-020-01677-5

[6] Yuzhong Liu, and Yuliang Ji. Target recognition of sport athletes based on deep learning and convolutional neural network. Journal of intelligent & fuzzy systems, 40(2):2253-2263, 2021. https://doi.org/ 10.3233/JIFS-189223

[7] Gang Chen, Guipeng Zhang, Zhenguo Yang, and Wenyin Liu. Multi-scale patch-GAN with edge detection for image inpainting. Applied intelligence, 53(4):3917-3932, 2023. https://doi.org/10.1007/s10489-022-03577-2

[8] Teng Long. Research on application of athlete gesture tracking algorithms based on deep learning. Journal of ambient intelligence and humanized computing, 11(9):3649-3657, 2020. https://doi.org/10.1007/s12652-019-01575-w

[9] Zongshuai Hao, Xin Wang, and Shoucun Zheng. Recognition of basketball players' action detection based on visual image and Harris corner extraction algorithm. Journal of intelligent & fuzzy systems, 40(4):7589-7599, 2021. https://doi.org/10.3233/JIFS-189579

[10] Ling Zhang, and Faze Liang. Monitoring and analysis of athletes' local body movement status based on BP neural network. Journal of intelligent & fuzzy systems, 40(2):2325-2335, 2021. https://doi.org/10.3233/JIFS-189229

[11] Shu Yang (2022). Face feature tracking algorithm of aerobics athletes based on Kalman filter and mean shift. International journal of biometrics, 14(3-4):394-407, 2022. https://doi.org/10.1504/IJBM.2022.124679

[12] Lina Zhang, and Haidong Dai (2023). Motion trajectory tracking of athletes with improved depth information-based KCF tracking method. Multimedia tools and applications,

82(17):26481-26493, 2023. https://doi.org/10.1007/s11042-023-14929-6

[13] Mingyue Wu, Ran Wang, Yang Hu, Mengjiao Fan, Yufan Wang, Yanchun Li, and Shengyuan Wu. Invisible experience to real-time assessment in elite tennis athlete training: Sport-specific movement classification based on wearable MEMS sensor data. Proceedings of the institution of mechanical engineers, part P: Journal of sports engineering and technology, 237(4):271-282, 2023. https://doi.org/10.1177/17543371211050312

[14] He Xu, Mingtao Guo, Nadia Nedjah, Jindan Zhang, and Peng Li. Vehicle and pedestrian detection algorithm based on lightweight YOLOv3-promote and semi-precision acceleration. IEEE transactions on intelligent transportation systems, 23(10):19760-19771, 2022. https://doi.org/10.1109/TITS.2021.3137253

[15] Jing Zhuang, Jianli Sun, and Guoliang Yuan. "Arrhythmia diagnosis of young martial arts athletes based on deep learning for smart medical care," Neural computing and applications, 35(20):14641-14652, 2023. https://doi.org/10.1007/s00521-021-06159-4

[16] Vishesh Kumar, and Marcelo Worsley. Scratch for sports: athletic drills as a platform for experiencing, understanding, and developing AI-driven apps. Proceedings of the AAAI conference on artificial intelligence, 37(13):16011-16016, 2023. https://doi.org/10.1609/aaai.v37i13.26901

[17] Brandon Giles, Stephanie Kovalchik, and Machar Reid (2020). A machine learning approach for automatic detection and classification of changes of direction from player tracking data in professional tennis. Journal of sports sciences, 38(1):106-113, 2020. https://doi.org/10.1080/02640414.2019.1684132

[18] Gengyuan Wang, Xiaolong Zeng, Guanquan Lai, Guoqing Zhong, Ke Ma, and Yu Zhang. Efficient subject-independent detection of anterior cruciate ligament deficiency based on marine predator algorithm and support vector machine. IEEE journal of biomedical and health informatics, 26(10):4936-4947, 2022. https://doi.org/10.1109/JBHI.2022.3152846

[19] Tristan Equey, Antoni Pastor, Rafael de la Torre Fornell, Andreas Thomas, Sylvain Giraud, Mario Thevis, Tiia Kuuranne, Norbert Baume, Osquel Barroso, and Reid Aikin. Application of the athlete biological passport approach to the detection of growth hormone doping. The journal of clinical endocrinology & metabolism, 107(3):649-659, 2022. https://doi.org/10.1210/clinem/dgab799

[20] Matthew T.O. Worsey, Hugo G. Espinosa, Jonathan B. Shepherd, and David V. Thiel. One size doesn't fit all: Supervised machine learning classification in athlete-monitoring. IEEE sensors letters, 5(3):1-4, 2021. https://doi.org/10.1109/LSENS.2021.3060376

[21] Chunmin Dai, and Yang Lu. Improved biological image tracking algorithm of athlete's cervical spine health. Revista brasileira de medicina do esporte, 27(3):274-277, 2021. https://doi.org/10.1590/1517-8692202127032021_0129

[22] Hao Ren. Sports video athlete detection based on deep learning. Neural computing and applications, 35(6):4201-4210, 2023. https://doi.org/10.1007/s00521-022-07077-9

[23] Zhipeng Li, Xiaolan Li, Ming Shi, Wenli Song, Guowei Zhao, Ruizhu Yang, and Shangbine Li. Tracking algorithm of snowboard target in intelligent system. Journal of intelligent & fuzzy systems, 40(2):3117-3125, 2021. https://doi.org/10.3233/JIFS-189350

[24] Tianyi Wang, and Cuiping Shi. Basketball motion video target tracking algorithm based on improved gray neural network. Neural computing and applications, 35(6):4267-4282, 2023. https://doi.org/10.1007/s00521-022-07026-6

[25] Alec Buttfield, and Kevin Ball. The practical application of a method of analysing the variability of within-step accelerations collected via athlete tracking devices. Journal of sports sciences, 38(3):343-350, 2020. https://doi.org/10.1080/02640414.2019.1699987

[26] Susanne Ellens, Daniel Hodges, Sean McCullagh, James J. Malone, and Matthew C. Varley. Interchangeability of player movement variables from different athlete tracking systems in professional soccer. Science and medicine in football, 6(1):1-6, 2021. https://doi.org/10.1080/24733938.2021.1879393

[27] Brandon Giles, Stephanie Kovalchik, and Machar Reid. A machine learning approach for automatic detection and classification of changes of direction from player tracking data in professional tennis. Journal of sports sciences, 38(1):106-113, 2020. https://doi.org/10.1080/02640414.2019.1684132

[28] Hao Ren. Sports video athlete detection based on deep learning. Neural computing and applications, 35(6):4201-4210, 2023. https://doi.org/10.1007/s00521-022-07077-9

[29] Mingyue Wu, Ran Wang, Yang Hu, Mengjiao Fan, Yufan Wang, Yanchun Li, and Shengyuan Wu. Invisible experience to real-time assessment in elite tennis athlete training: Sport-specific movement classification based on wearable MEMS sensor data. Proceedings of the institution of mechanical engineers, part P: Journal of sports engineering and technology, 237(4):271-282, 2023. https://doi.org/10.1177/17543371211050312

[30] Jing Zhuang, Jianli Sun, and Guoliang Yuan. Arrhythmia diagnosis of young martial arts athletes based on deep learning for smart medical care. Neural computing and applications, 35(20):14641-14652, 2023.

https://doi.org/10.1007/s00521-021-06159-4
[31] Arindam Chaudhuri. Hierarchical modified fast R-CNN for object detection. Informatica, 45(7):2201-2210, 2021. https://doi.org/10.31449/inf.v45i7.3732
[32] Lingxiao He, Xingyu Liao, Wu Liu, Xinchen Liu, Peng Cheng, and Tao Mei. Fastreid: A pytorch toolbox for general instance re-identification. Association for computing machinery, 45(7):9664-9667, 2023 https://doi.org/10.1145/3581783.3613460.
[33] João Videira, Pedro D. Gaspar, Vasco N. G. J. Soares, and João M. L. P. Caldeira. A mobile application for detecting and monitoring the development stages of wild flowers and plants. Informatica, 48(6): 3217-3225, 2024. https://doi.org/10.31449/inf.v48i6.5645
[34] Cheng Pan, Haiyan Zhao, and Meijiao Sun. Real-time target detection system in scenic landscape based on improved YOLOv4 algorithm. Informatica, 48(8):5117-5125, 2024. https://doi.org/10.31449/inf.v48i8.5700
[35] Gaoang Wang, Yizhou Wang, Haotian Zhang, Renshu Gu, and Jenq-Neng Hwang. Exploit the connectivity: multi-object tracking with trackletnet. Association for computing machinery, 48(8):482-490, 2019. https://doi.org/10.1145/3343031.3350853