EAI Endorsed Transactions
on Scalable Information Systems

Research Article EAI.EU

Multi-target trajectory tracking in multi-frame video images of basketball games based on deep learning

Yong Gong[1] and Gautam Srivastava[2,*]

[1]Ministry of Sport, Wuhan Polytechnic University, Wuhan 430048, China
[2]Dept of Math and Computer Science, Brandon University, Brandon, Canada

Abstract

INTRODUCTION: There is occlusion interference in the multi-target visual tracking process of basketball video images, which leads to poor accuracy of multi-target trajectory tracking. This paper studies the multi-target trajectory tracking method in multi-frame video images of basketball sports based on deep learning.
OBJECTIVES: Aiming at the problem of target occlusion in the tracking process and the problem of trajectory tracking anomaly caused by target occlusion, a modified algorithm is proposed.
METHODS: The method is divided into two parts: detection and tracking. In the detection part, the YOLOv3 algorithm in deep learning technology is used to detect each target in the video, and the original YOLOv3 backbone network Darknet-53 is replaced by the lightweight backbone network MobileNetV2 to extract the target features.
RESULTS: Based on the target detection results, the Kalman filter is used to predict the next position and bounding box size of the target to obtain the target trajectory prediction results according to the current target position, then a hierarchical data association algorithm is designed, and multi-target tracking of the same category is completed based on the target appearance feature similarity and feature similarity.
CONCLUSION: The experimental results show that the method can accurately detect the targets in multi-frame video images in basketball sports and obtain high-precision target trajectory tracking results.

Keywords: deep learning, basketball sports video, multi-objective, trajectory tracking, YOLOv3 algorithm, data association.

Received on 24 August 2022, accepted on 12 October 2022, published on 18 October 2022

doi: 10.4108/eetsis.v9i6.2591

*Corresponding author. Email: srivastavag@brandonu.ca

1. Introduction

Among many sports events, basketball games have the largest audience and the highest attention. With the improvement of the quality of life and the rapid advancement of technology, the requirements for basketball videos are also getting higher and higher. For example, the current passive, flat viewing model of watching basketball will gradually fail to meet the needs of TV viewers. Broadcasters need to add various visual effects to meet the visual demands of the audience. In terms of game research and analysis, basketball coaches need to extract relevant data from basketball game videos to assist basketball players in tactical research. In terms of commercial applications, game broadcasters also need to fully tap the commercial value contained in basketball game broadcasts. All of these require the analysis of basketball video data and the processing of basketball video images according to different requirements to meet the segmentation and classification requirements of online video objects [1]. Therefore, object segmentation of moving online video objects and motion attributes has high practical value and understanding relevance [2]. Through the automatic extraction of visually salient image regions in video sequences, the accuracy and efficiency of moving object detection, extraction, localization and tracking can be effectively improved [3].

Some approaches have been proposed by some scholars for this area of target tracking. For multi-agent target tracking, Reference [4] formulated the tracking task as a distributed model predictive control (DMPC) problem, innovatively combined the adaptive differential evolution (ADE) algorithm with Nash optimization, and proposed a Nash combined ADE method. In the Reference [5], a multi-target tracking performance and trajectory prediction method for basketball players was proposed for the multi-target localization and tracking problem. The subjects tracked multiple targets moving on the computer screen which may disappear briefly and then re-appear during the movement, and the subjects were asked to continuously track the targets and report their final positions and the number of manipulated targets and the locations of the reappearance of the moving targets after they disappear.

To improve the tracking accuracy of spatially dense group targets, Reference [6] proposed a group target association and tracking algorithm based on the global nearest neighbour. Based on the principle of "global optimum", the closest group target and measurement are selected for priority association and updated to avoid association conflicts and reduce association errors, which can effectively solve the contradiction between association accuracy and tracking real-time. At the same time, the combination of track prediction and track forecast is proposed to solve the track intermittency and fusion problem in the tracking process. Reference [7] proposes a secure distributed detection system for industrial image and video data security based on IPFS and blockchain, and a decentralized peer-to-peer image and video sharing platform based on IPFS column (pHash) technology to detect copyright infringement of multimedia. When multimedia is uploaded to IPFS, the pHash of the same content is determined and checked against the existing pHash value in the blockchain network. Similarity to existing pHash values will result in multimedia being detected as tampered with. Blockchain technology offers the advantage of not participating in third parties, thus avoiding a single point of failure. Reference [8] proposes a sports safety information mining platform based on multimedia data sharing technology. The hardware part of the platform includes a teaching multimedia data sharing module, a shared server module, a shared client and a Web server. To realize low-latency and low-energy-consumption information transmission, ZigBee technology is introduced into the software design to realize the function of information communication and complete the evaluation of mining quality.

Although the research on target tracking has made great progress and breakthroughs to a certain extent in recent years, the effects of sequence target feature extraction, target joint detection and target trajectory tracking related methods are still not perfect due to the complexity of the external environment and the influence of noise factors and target deformation. The core problem of object tracking is feature expression. Appropriate features need to be selected according to the different application scenarios of features. However, the tracking effect is far from meeting the needs of practical applications.

In recent years, with the research of deep learning, researchers have found new ideas for target tracking. The field of computer vision has been rapidly developed since the emergence of deep learning technology, and deep learning technology has been used in image classification. In recent years, deep learning-based multi-target tracking algorithms have also made some breakthroughs. Multi-target tracking is a very challenging research direction in the field of computer vision and has a wide range of application scenarios, such as intelligent video monitoring and control, abnormal behavior analysis, mobile robot research and so on. Traditional multi-target tracking algorithms often have poor tracking results due to poor target detection. The detector based on deep learning can obtain a better target detection effect and improve the accuracy of target tracking. Therefore, the effective combination of object tracking and deep learning has become the focus of researchers in the tracking field. This paper studies the multi-target trajectory tracking method in multi-frame video images of basketball movement based on deep learning. The YOLOv3 algorithm in deep learning technology is used to detect each target in the video, and the original YOLOv3 backbone network Darknet-53 is replaced with a lightweight backbone network MobileNetV2 to extract target features. The Kalman filter is used to predict the next position and the size of the bounding box according to the current basketball target position, and the trajectory prediction result of the basketball target is obtained. The multi-target tracking of the same category is completed, and a correction algorithm is proposed for the problem of target occlusion in the tracking process and the abnormal trajectory tracking caused by target occlusion. Implemented a multi-target trajectory tracking method for basketball video based on deep learning.

2. Method

2.1. Overall framework of video multi-target trajectory tracking method

The overall framework of the video multi-target trajectory tracking method based on deep learning is shown in Figure 1, which is mainly divided into two parts: detection and tracking.

The detection mainly combines the YOLOv3 algorithm in deep learning to detect and recognize video multi-targets. Firstly, image preprocessing is performed on the video sequence. Then multi-scale point convolutional neural network [9] is used to achieve the target segmentation of the image and obtain the convolution feature map in the video image. The input video feature map is analyzed and filtered by the detection network. Finally, the frame of the optimal target is obtained by

confidence calculation and multi-scale prediction, the multi-target in the video is classified by the classifier, and the center point coordinates of the frame of the optimal target in the video are obtained.

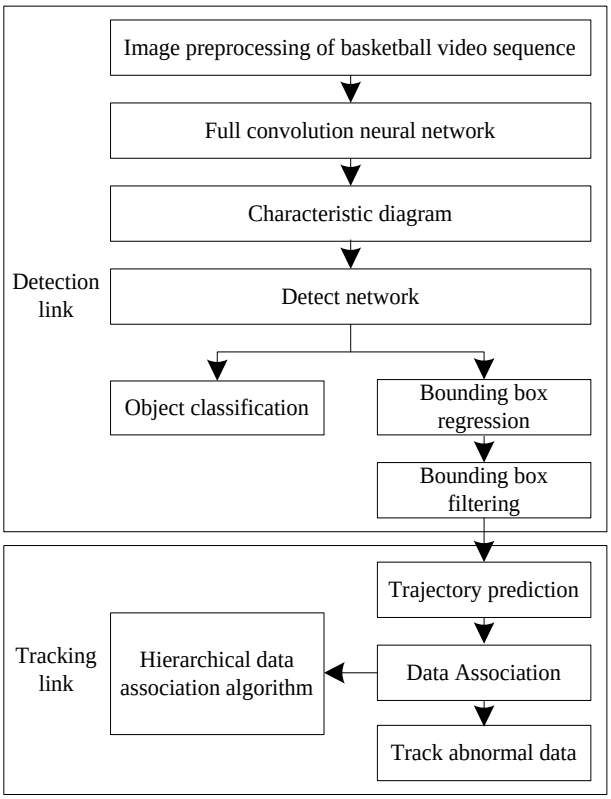


Figure 1. The overall framework of the video multi-target trajectory tracking method is based on deep learning

The detection mainly combines the YOLOv3 algorithm in deep learning to detect and recognize video multi-targets. Firstly, image preprocessing is performed on the video sequence. Then multi-scale point convolutional neural network [9] is used to achieve the target segmentation of the image and obtain the convolution feature map in the video image. The input video feature map is analyzed and filtered by the detection network. Finally, the frame of the optimal target is obtained by confidence calculation and multi-scale prediction, the multi-target in the video is classified by the classifier, and the center point coordinates of the frame of the optimal target in the video are obtained.

The tracking part is to combine the multi-target detection results of the detection link to conduct data association and tracking and input the optimal target center point coordinates of this kind of target into the Kalman Filter to predict the center point at the next time, that is, the multi-target trajectory prediction. The frame data of different target bounding boxes output by the detector is correlated to determine the number of different targets. The measurement values of the center points of different targets and the estimated values of the center points at this time are used to obtain the optimal estimates of the real states of different targets. If the data association fails due to occlusion, the hierarchical data association algorithm is used for data association, and the newly emerging target at that moment is associated with the target that disappears due to occlusion. If abnormal trajectory fluctuations are caused by partial occlusion, the trajectory anomaly correction algorithm is used to correct different target boxes and trajectories.

2.2. Multi-object detection based on deep learning

Multi-target detection methods based on deep learning can be divided into two-stage detection algorithms represented by the regional convolutional neural network in terms of detection modes. Among them, the idea adopted by the two-stage detection algorithm is as follows: Firstly, a proposal is used to provide location information, and then classifiers are used to provide category information. Real-time detection affected by detection mode cannot be guaranteed. However, the single-stage detection algorithm provides a new and more direct idea, that is, the whole image is used as network input, and the position and category of a compact rectangular bounding box containing objects are directly regress at the output layer, to transform the multi-object detection [10] problem into regression problem processing, which greatly improves the detection speed.

YOLOv3 algorithm design

YOLOv3 algorithm does not use classic Backbone networks such as VGG-16 and ResNet-50, but the YOLOv3 algorithm proposes its Backbone network——Darknet-53 feature extraction. There is no pooling layer or fully connected layer in the original network structure of YOLOv3. In the process of forwarding propagation, the size transformation of the tensor is realized only by changing the step size of the convolution kernel, that is, the step size is 2, which means that the side length of the video image is reduced by half and the area is reduced to 1/4 of the original one. Therefore, after 5 sampling times, the feature map is 1/32 of the original video image. The idea of FPN (Feature Pyramid Networks) is used for reference, the algorithm uses multi-scale to detect moving video objects of different sizes. Three feature maps of different scales are output, namely 13×13, 26×26 and 52×52. This makes the detection effect of YOLOv3 significantly improved compared with the previous version of the YOLO algorithm.

Compared with Faster, R-CNN and other two-stage detection algorithms, the YOLOv3 algorithm has obvious advantages in detection speed, but it has two shortcomings. First, the model obtained by training is large and not suitable for embedded equipment. Second, higher inference time is required in the model under CPU.

To solve the above problems, an optimization method is designed to optimize YOLOv3.

Optimization method design

Model optimization can be carried out from backbone network optimization, optimizer optimization and model pruning optimization. Considering the implementation steps and difficulties of various optimization methods, YOLOv3 is optimized from the perspective of backbone network adjustment in the multi-target detection process of YOLOv3.

From the perspective of the lightweight model, Mobile-NetV2 is used as the backbone network to replace the original network for feature extraction. Mobile NetV2 is an optimized version of Mobile Nets and an efficient model for mobile and embedded devices. Mobile Nets is a mainstream lightweight network based on streamlined architecture, which uses deep separable convolution to build deep neural networks. Mobile Nets decompose the standard convolution into deep convolution and point-by-point convolution to greatly reduce the number of parameters and computation.

The input feature map is F of size (Z_F, Z_F, M), the adopted standard convolution K is (Z_K, Z_K, M, N), and the output feature map is G of size (Z_C, Z_C, N). The standard convolution is calculated as follows:

$$G_{k,l,m} = \sum_{i,j,m} K_{i,j,m,n} \cdot F_{k+i-1,l+j-1,m} \quad (1)$$

If the number of input channels is M and the number of output channels is N, the corresponding calculation amount is:

$$\Omega = Z_K \cdot Z_K \cdot M \cdot N \cdot Z_F \cdot Z_F \quad (2)$$

The standard convolution $K(Z_K, Z_K, M, N)$ can be split into deep convolution and point-by-point convolution [11], Specifically, the processing part of deep convolution is the filter, the size of the filter is $(Z_K, Z_K, 1, M)$, and its output feature is (Z_G, Z_G, M). The processing part of Point-by-point convolution is to convert channels with a size of $(1,1,M,N)$, resulting in a final output of (Z_G, Z_G, N). The convolution formula of deep convolution is as follows:

$$\hat{G}_{k,l,n} = \sum_{i,j,m} \hat{K}_{k,l,m,n} \cdot \hat{F}_{k+i-1,l+j-1,m} \quad (3)$$

In the formula, $\hat{K}$ the deep convolution and the convolution kernel are $(Z_K, Z_K, 1, M)$, where, m_{th} convolution kernels are applied to the m_{th} channel F to produce the m_{th} channel output on $\hat{G}$.

The calculation amount of deep convolution and point-by-point convolution is:

$$\hat{\Omega} = Z_K \cdot Z_K \cdot M \cdot N \cdot Z_F \cdot Z_F + M \cdot N \cdot Z_F \cdot Z_F \quad (4)$$

The total calculation amount of the above calculation is reduced by $\frac{1}{NZ_K^2}$. It can be found that the number of parameters can be significantly reduced by using depth-separable convolution.

Mobile NetV2 introduces two improvements over Mobile Nets, linear bottlenecks and reverse residuals. Mobile Nets and Mobile NetV2 both use Depth Wise (DW) convolution combined with PointWise (PW) volume [12] to extract multi-object features from video images. The improvement of Mobile NetV2 is to add a PW convolution before DW convolution [13]. Due to the computational nature of DW convolution itself, it cannot change the number of channels, that is, the flow of the previous layer can only be equal to the output channel. If the number of upper channels is small, DW can only extract video multi-object features in low-dimensional space, which often leads to poor results. Therefore, the PW added before each DW is considered a tool to raise dimensions. Mobile NetV2 drops the activation function of the second PW, which is called as called as a linear bottleneck. Since the activation function can effectively increase the nonlinearity in the high dimensional space, it will destroy the multi-objective feature of the video in the low dimensional space, and the main function of the second PW is dimensionality reduction.

Mobile NetV2 uses a 1 x 1, 3 x 3,1 x 1 structure with a shortcut to add the output and the input. ResNet uses Standard convolution (SC) for multi-object feature extraction, while Mobile NetV2 always uses DW convolution for multi-object feature extraction. Intuitively, MobileNetV2 uses inverted residuals and DW convolution to effectively extract multi-object features.

2.3. Design of target tracking algorithm

After multiple detection targets for each frame of video image by YOLOv3, the same target in successive frames needs to be tracked and trajectories are generated in turn. The Kalman filter model [14] is used to predict the position of the target in the next frame.

The advantage of the Kalman filter is that the model can be applied to any dynamic system with uncertain information to make a reasonable prediction of the next direction of the system, and it can always point out the real situation even with noise interference.

When the Kalman filter is used for state prediction, the previous state is set as $x_{k-1}=(\phi_{k-1},v_{k-1})$, where ϕ and v represent position and velocity respectively. Then two parameters need to be introduced in the current state prediction under the Gaussian distribution, Mean $\hat{x}_{k-1}$ and covariance matrix ϕ_{k-1}. At the same time, considering the problem of external control quantity and external noise interference, the state prediction equation of the Kalman filter can be obtained according to the kinematic formula and relevant mathematical calculation, as shown below:

$$\begin{cases}\hat{x}_k = F_k\cdot\hat{x}_{k-1}+B_k\delta_k\\ \phi_k = F_k\cdot\phi_{k-1}\cdot F_k^{\ T}+\beta_k\end{cases} \tag{5}$$

In the formula, F_k is the kinematic coefficient matrix. B_k is the external control matrix. δ_k is the external control quantity. β_k is the covariance matrix of external noise.

Equation (5) shows that the current new optimal estimate is obtained by adding the known external control quantity to the previous optimal estimate, while the new uncertainty is obtained by adding the external environmental disturbance to the previous uncertainty.

To make the Kalman filter model work continuously, some parameters in the model need to be updated to ensure the real-time accuracy of video multi-target trajectory tracking.

The detection output of YOLOv3 is Q_k and the covariance matrix is R_k. By combining the previous predictions to get two Gaussian distributions. The prediction part is $(\mu_0,\Sigma_0)=(H_k\hat{x}_k,H_k\phi_k H_k^{\ T})$, where H_k is the covariance matrix of Gaussian distribution of the prediction part. The detection part is $(\mu_1,\Sigma_1)=(Q_k,R_k)$. Through a series of mathematical operations, the state update equation can be written:

$$\begin{cases}\hat{x}_k' = \hat{x}_k+K\cdot(Q_k-H_k\hat{x}_k)\\ \phi_k' = \phi_k-KH_k\phi_k\\ K=\phi_k H_k^{\ T}\cdot(H_k\phi_k H_k^{\ T}+R_k)^{-1}\end{cases} \tag{6}$$

In the formula, K is the Kalman gain. $\hat{x}_k'$ is the new optimal estimate, which can be iterated ϕ_k' into the next prediction and update equation.

2.4. Design of hierarchical data association algorithm

In previous multi-target tracking algorithms, the main problems in data association are as follows: Due to the target barrier, Kalman filter prediction uncertainty will increase, and when the appearance of being tracked target at high similarity, the forecast trajectory may not be that were the target of the need to track, to avoid to use other don't need to track target data association, associated with the method of hierarchical data.

When the tracked object continuously appears in the adjacent video frames with little appearance change greatly and no occlusion, the data association using only the appearance features of the tracked object can achieve better results. However, in general, the situation of the tracked object is not so simple, and it often faces the problems of appearance change and occlusion. To alleviate the occlusion and appearance change encountered in the tracking scene, this paper proposes a hierarchical data association method. This method uses the appearance features and motion features of the target to correlate. The purpose of the first-layer data association is to use the appearance feature to associate the target that is not occluded or has little appearance change in the video, that is, the appearance similarity of the target meets the confidence value, and the first-layer data association is conducted. If the appearance similarity is lower than the confidence value, it indicates that the appearance of the target has changed, or occlusion occurs in the tracking scene. Then the second layer of association is adopted to judge and associate multiple targets using motion features [15]. The appearance feature similarity and motion feature similarity mentioned above are described by cosine similarity.

Similarity calculation of target appearance features

The similarity of target appearance features is expressed by Equation (7):

$$E_{i,j}^{t}=\frac{\langle\partial_i,\partial_j\rangle}{\|\partial_i\|\|\partial_j\|} \tag{7}$$

∂_i represents the appearance feature of the existing target, and ∂_j represents the appearance feature of the target to be matched.

Calculation of feature similarity

The video frame rate is required to be high enough and the motion trajectory of the target is continuous and smooth. Considering the spatial information of the target, and the speed and direction of the target movement are used to represent the motion features of the tracking target.

However, the cosine similarity can only represent the direction consistency of the target. Therefore, the

modified similarity is considered to represent the similarity of the target direction and speed, and the feature similarity is expressed as:

$$M_i^j = \frac{\langle v_i - \bar{v}, v_j - \bar{v} \rangle}{\|v_i - \bar{v}\| \|v_j - \bar{v}\|} \tag{8}$$

Implementation of Hierarchical data Association

In the data association of the current frame of the video, the object detection box D^t has been obtained by the detector, and the target trajectory T^{t-1} of the previous frame is known. In the matching process, the detected candidate object is matched with the target trajectory, and the trajectory T of the current frame is obtained after the matching. The cost matrix is shown in Equation (9) :

$$c_{i,j}^t = \begin{cases} 1 - f_{i,j}^t, f_{i,j}^t \geq T_c \\ \infty \quad ,otherwise \end{cases} \tag{9}$$

$c_{i,j}^t$ represents the loss function of the match between target i and trajectory j, and $f_{i,j}^t$ is the similarity function between target i and trajectory j. As the similarity between target and trajectory increases, the loss function will decrease.

$aff_{i,j}^t$ can be calculated as follows:

$$aff_{i,j}^t = w_1 * E_{i,j}^t + w_2 * M_{i,j}^t \tag{10}$$

$E_{i,j}^t$ represents the appearance similarity between target i and trajectory j in a video frame t; $M_{i,j}^t$ represents the motion similarity between target i and trajectory j in the video frame t. w_1 represents the impact factor of target appearance. w_2 represents the impact factor of motion.

In data association, the problems of mutual occlusion and disappearance and re-appearance of objects usually occur. To reduce the occurrence of such problems, a hierarchical data association method is proposed to solve problems, and different functions are used for correlation between the two layers [16]. The previous layer only considers the matching value where the appearance similarity of the target is higher than the confidence T_a. If the appearance similarity of the target is high, it means that the target has not been occluded during the movement.

In the proposed algorithm, the influence factor w_1 which represents the appearance of the target is set to 1, and the influence factor w_2 of the motion is set to 0 in the similarity function of the first layer. However, when the appearance similarity of the target is low, it indicates that the target has occlusion or other problems that make the appearance change of the target in the process of movement. At this time, the association matching of the target cannot be carried out accurately only by relying on the appearance similarity of the target, so the motion feature is introduced for discrimination. After the experiment, the target appearance influence factor w_2 is set to 0.4, and the motion influence factor ω2 is set to 0.6, which can achieve the best test effect.

2.5. Video multi-target trajectory anomaly correction algorithm

When a partial occlusion occurs in the process of multi-target trajectory tracking in motion video [17-18], it is easy to fall into the problem that only the un-occluded part of multi-target can be obtained [19-20]. The trajectory fluctuates greatly when the target is occluded from the beginning to the end. Aiming at this situation, a new trajectory correction algorithm is proposed to avoid the trajectory deviation caused by the partial occlusion of the target. The algorithm uses the characteristic that the target frame will not change suddenly during the tracking process to correct the frame and trajectory.

The specific process of the video multi-target trajectory anomaly correction algorithm is shown as follows:

The numbers of average height and width of the targets $j \in \{0,1,2,\cdots,M\}$ in the tracking process are calculated and saved as $\bar{H}$ and $\bar{W}$. Compared with the border height b_j^h and width b_j^w of the target j at time T, the threshold is assumed to be $\mu \in (0,1)$.

The center points $U(T-1)_j$ and $U(T)_j$ of the target j at time $T-1$ and time T are also compared, The following situations will occur:

If the height $b_j^h \leq \mu\bar{H}$ and the width $b_j^w \geq \mu\bar{W}$, the center points $U(T-1)_j$ and $U(T)_j$ of the target j at time $T-1$ and time T is compared. If the coordinate of the center point of the target at time T is compared with that at time $T-1$, the change in the y direction Δy is much larger than the average change in the y direction $\overline{\Delta y}$. If Δy is positive, the lower body of the target is considered to be occluded. If Δy is negative, the upper body of the target is considered to be occluded. According to the frame height change Δh, update the coordinate

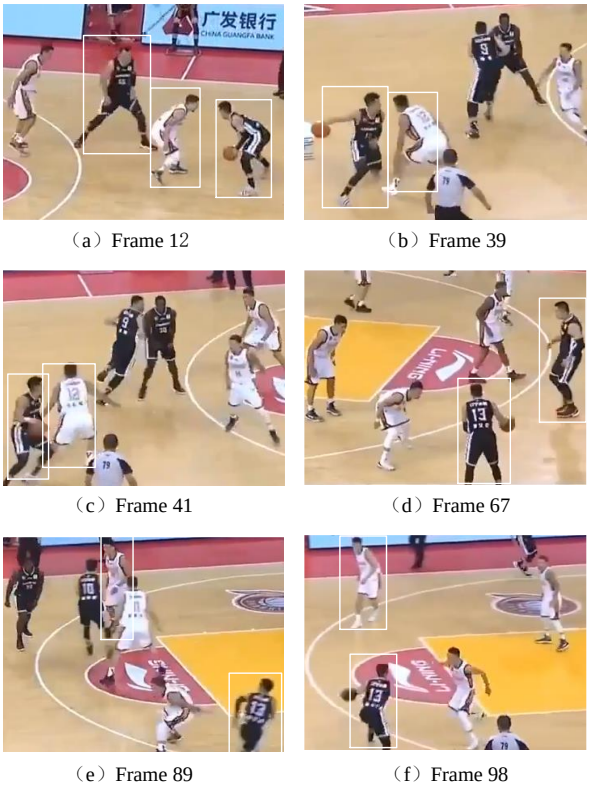
$U(T)_j^{(y)}$ of the center point of the target j in the y direction at time T;

If the height $b_j^h \geq \mu\bar{H}$ and the width $b_j^w \leq \mu\bar{W}$, the center points $U(T-1)_j$ and $U(T)_j$ of the border of the target j at time $T-1$ and time T is compared. If the coordinate of the center point of the target at time T is compared with that at time $T-1$, the change in the x direction $\overline{\Delta x}$ is much larger than the average change in the x direction $\overline{\Delta x}$ and $\overline{\Delta x}$ is positive, the left half of the body of the target is considered to be occluded. If $\overline{\Delta x}$ is negative, the right half of the body of the target is considered to be occluded. According to the frame width change Δw, the coordinate $U(T)_j^{(x)}$ of the center point of the target j in the x direction at time T is updated.

$$\begin{cases} U(T)_j^{(y)} = U(T-1)_j^{(y)} \pm \dfrac{\Delta h}{2} \\ U(T)_j^{(x)} = U(T-1)_j^{(x)} \pm \dfrac{\Delta h}{2} \end{cases} \quad (11)$$

3. Experimental Results

To verify the application effect of the multi-object trajectory tracking method of basketball video studied in this paper in actual basketball video, a CBA tournament is collected as the experimental data set, and a CBA game is selected as an application video. The proposed method is used to track multiple targets in the application video, and the target recognition and trajectory tracking results are as follows.



（a）Frame 12　（b）Frame 39

（c）Frame 41　（d）Frame 67

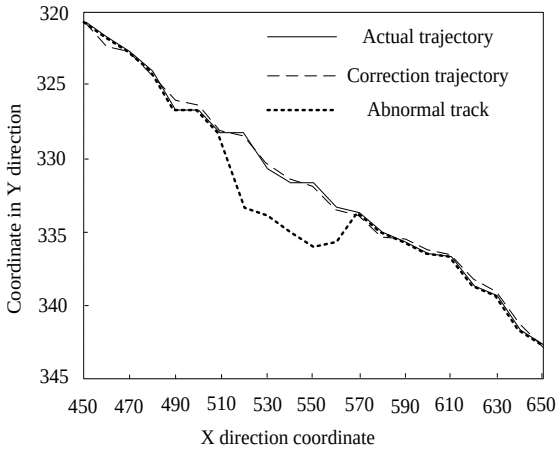
（e）Frame 89　（f）Frame 98

Figure 2. Tracking results of this method

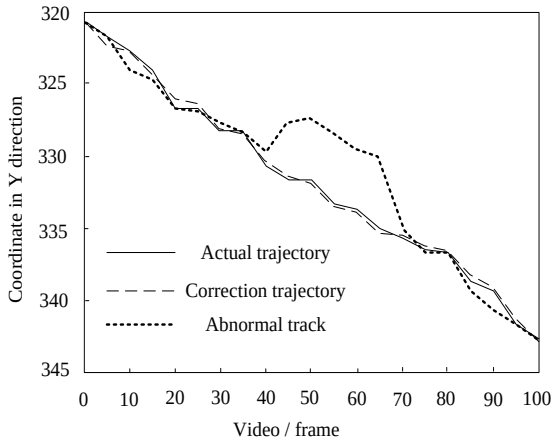
3.1. Tracking Results

In the experimental data set, a CBA basketball game video is arbitrarily selected as the application video, and the 12th, 39th, 41st, 67th, 89th and 98th frames are used as examples. The No. 12 white player, No. 13 black player, and No. 15 black player in the game performed multi-target recognition and tracked their trajectory. The target recognition results in different frames are shown in Figure 2.

As can be seen from Figure 2, the method proposed in this paper is used to carry out multi-target recognition for players in white No. 12, black No. 13 and black No. 15. player No.12 in white appears in frames 12, 39, 41, 89 and 98. Player No. 13 in black appears in frames 12, 39, 41, 67, 89 and 98. The No. 15 player in black appears in frames 12 and 67. The above results fully demonstrate that the proposed method can accurately track moving objects under both non-occluded and occluded conditions.



(a) Coordinate comparison diagram of abnormal track correction track and real track



(b) Comparison diagram of target change in the Y direction

Figure 3. Target trajectory correction

3.2. Analysis of trajectory tracking effect

In the process of analyzing the trajectory tracking effect of the proposed method, trajectory correction and trajectory tracking error are carried out, and the results are as follows.

Trajectory correction results

The proposed method is used to correct the abnormal trajectory in the process of target trajectory tracking, and the results are shown in Figure 3.

By analyzing Figure 3, it can be concluded that the trajectory of the target abnormal trajectory corrected by the proposed method is closer to the actual trajectory. The method in this paper uses the Kalman filter to predict the next position and bounding box size of the current basketball target position and obtains the trajectory prediction result of the basketball target. Similarity completes the tracking of multiple targets of the same category and corrects the problem of target occlusion in the tracking process and the problem of abnormal trajectory tracking caused by target occlusion, so the correction effect on abnormal target trajectories is better [21,22].

Trajectory tracking error analysis

In the application video, player No.13 in black is taken as an example to conduct experimental tests, and the target trajectory tracking error of the proposed method is compared under occluded conditions and non-occluded conditions. The results are shown in Figure 4.

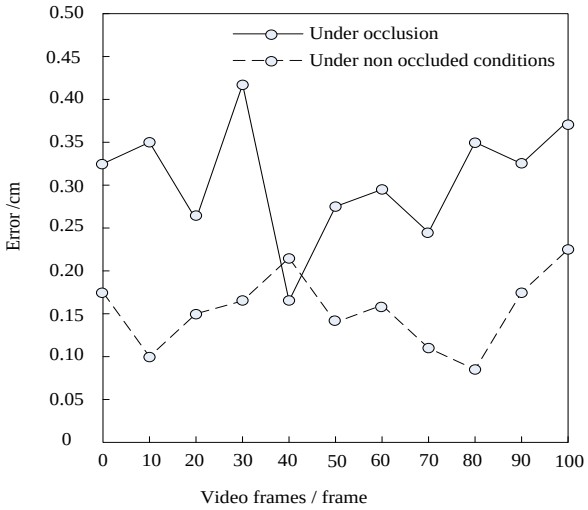


Figure 4. Comparison of tracking error of basketball target trajectory

From the analysis of Figure 4, it can be seen that under the condition that the target is occluded by the method in this paper, the trajectory tracking error range is in the range of 0.15-0.45cm, and the average difference is about 0.30cm. Under the condition that the target is not occluded, the error range of trajectory tracking is in the range of 0.05-0.25cm, and the mean value of the average difference is lower than 0.15cm. The above data fully demonstrate that the method in this paper can track moving targets more accurately, and the target trajectory correction can be significantly improved through abnormal trajectory correction.

3.3. Operation time

Taking the application video containing 100 frames of images used in the above experiment as an example, the size of the tracking target area is set as 32×32 pixels. Figure 5 shows the operation time of each frame image in the multi-target tracking process of the proposed method.

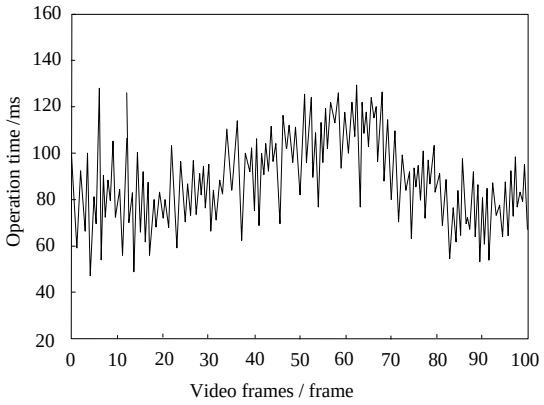


Figure 5. Operation time of each frame

According to the analysis of Figure 5, when the proposed method is used for target tracking, the average computation time of each frame is about 96ms, and the calculation results meet the requirements of conventional video image multi-target trajectory tracking, which indicates that the proposed method has better real-time tracking performance.

3.4. Comparison of center position error and coverage

Center position error and coverage describe the center deviation and the overlap ratio of the tracking box and the actual target box, respectively. To qualitatively evaluate the trajectory tracking effect of the proposed method, the above two indexes are taken as bid evaluation criteria, and the method of reference [5] and [6] are taken as comparison methods to compare the evaluation indexes of the proposed method and the two comparison methods. The results are shown in Table 1 and Table 2.

By analyzing Table 1 and Table 2, it can be seen that in the process of multi-target trajectory tracking, the calculation results of center position error and coverage obtained by the proposed method are significantly better than the two comparison algorithms, which indicates that the proposed method has a better tracking effect among the above methods.

Table 1. Comparison of center position error (pixels)

Moving target		Reference [5] method	Reference [6] method	The method in this paper
Under non-occluded conditions	1	1022.2	23.7	11.6
	2	84.8	140.0	12.1
	3	41.1	74.5	12.0
Under occlusion	1	114.2	90.9	17.6
	2	92.6	78.5	17.7
	3	83.4	116.3	16.9

Table 2. Comparison of coverage (%)

Moving target		Reference [5] method	Reference [6] method	The method in this paper
Under non-occluded conditions	1	27.4	9.4	83.1
	2	45.3	44.6	84.4
	3	36.4	37.1	68.0
Under occlusion	1	23.6	14.3	77.3
	2	37.5	36.0	66.2
	3	43.0	41.5	51.3

In summary, the method in this paper solves the interference caused by occlusion to the accuracy of trajectory and basketball moving target recognition to a certain extent, has high real-time performance, and has a certain application value in the field of moving video target tracking.

4. Conclusion

This paper studies a multi-target trajectory tracking method based on deep learning in basketball video, using high-performance detectors to detect multiple targets of the same type in basketball video, focusing on the relationship between frames before and after the same target and target tracking. Occlusion problems in the process. After experimental verification, the following conclusions are drawn:

(1) The proposed method solves the interference caused by occlusion to the accuracy of trajectory and basketball target recognition to a certain extent and improves the anti-interference and recognition accuracy of the target.

(2) The proposed method has good real-time performance and can meet the requirements of multi-target trajectory tracking in conventional video images.

(3) The proposed method has good calculation results of the target center position error and coverage rate in the process of basketball target tracking, indicating that the tracking effect of the method in this paper is good.

However, the current research is still carried out on the same type of target, and the next research will focus on breaking through the multi-target tracking problem of different types and refer to the feature information to improve the existing inter-frame relationship.

Acknowledgements

The paper was funded by the Research Project of Teaching Reform in Wuhan Polytechnic University with No.XM2015013.

References

[1] Ammar, S. , Bouwmans, T. , Zaghden, N. , & Neji, M. (2020). Deep detector classifier (DeepDC) for moving objects segmentation and classification in video surveillance. IET Image Processing, 14(8), 1490-1501.

[2] Liu, S., Wang, S., Liu, X., Lin, C. T., & Lv, Z. (2021) Fuzzy Detection aided Real-time and Robust Visual Tracking under Complex Environments. IEEE Transactions on Fuzzy Systems, 29(1), 90-102

[3] Li, B. (2021). Auto-Extraction Simulation of Visual Saliency Image Region in Video Sequence. Computer Simulation,38(07), 447-450.

[4] Yu, Y., Wang, H., Liu, S., Guo, L. & Li, Y. (2021). Distributed multi-agent target tracking: a nash-combined adaptive differential evolution method for uav systems. IEEE Transactions on Vehicular Technology, .2021, 70(8), 8122-8133.

[5] Zhang, Y. , Wei, Y F. , Tao, J T. , & Li, J. (2021). Multiple Object Tracking and Motion Trajectory Prediction of Basketball Players. Chinese Journal of Sports Medicine, 40(10), 800-809.

[6] Xiu, J j. , Han L L, . Dong, K. , Li, Q F. (2020). Study on Correlation and Tracking Algorithm of Space Dense Group Targets. Fire Control & Command Control, 45(8), 51-56.

[7] Rk, A. , Rt, A. , Nm, B. , Gsc, D. , Trg, E. , & Nnx, F. . (2021). A secured distributed detection system based on IPFS and blockchain for industrial image and video data security. Journal of Parallel and Distributed Computing,2(22), 128-143.

[8] Cao, X. Z. , & Gadekallu, T. R. . (2022). Construction of sports safety information mining platform based on multimedia data sharing technology, Mobile Networks and Applications, 5(2), 1-10.

[9] Ma, L. , Li, Y. , Li, J. , Tan, W. , Yu, Y. , & Chapman, M. A. . (2021). Multi-scale point-wise convolutional neural networks for 3d object segmentation from lidar point clouds in large-scale environments. IEEE Transactions on Intelligent Transportation Systems, 22(2), 821-836.

[10] Jin, Y. & Jin, L. Z. (2020). Multi-target (Pedestrian) Detection Algorithm Based on Mobile U-Net. Industrial Control Computer,33(03), 81-83.

[11] Alenezi, F. , & Ganesan, S. . (2021). Geometric-pixel guided single-pass convolution neural network with graph cut for image dehazing. IEEE Access, 9, 29380-29391.

[12] Lv, J. , Sun, Q. , Li, Q. , & Moreira-Matias, L. . (2020). Multi-scale and multi-scope convolutional neural networks for destination prediction of trajectories. IEEE Transactions on Intelligent Transportation Systems, 21(8), 3184-3195.

[13] Das, D. , Nayak, D. R. , Dash, R. , Majhi, B. , & Zhang, Y. D. (2020). H-WordNet: a holistic convolutional neural network approach for handwritten word recognition. IET Image Processing, 14(9), 1794-1805.

[14] Huang, Y. , Zhang, Y. , Zhao, Y. , Shi, P. , & Chambers, J. A. (2020). A novel outlier-robust Kalman filtering framework based on statistical similarity measure. IEEE Transactions on Automatic Control, 66(6), 2677-2692.

[15] Cao, L. , Zheng, D. , Zhao, Z. , Wang, T. , Wang, D. , Fu, C. , & Gu, J. (2021). Convex Variational Inference for Multi-Hypothesis Fractional Belief Propagation Based Data Association in Multiple Target Tracking Systems. IEEE Sensors Journal, 21(17), 19121-19133.

[16] Olech, U. P., Spytkowski, M., Kwanicka, H., & Michalewicz, Z. (2021). Hierarchical data generator based on tree-structured stick breaking process for benchmarking clustering methods. Information Sciences, 554, 99-119.

[17] Adhami, M. H. , & Ghazizadeh, R. (2021). Secure multiple target tracking based on clustering intersection points of measurement circles in wireless sensor networks. Wireless Networks, 27(2), 1233-1249.

[18] Liu, S., Wang, S., Liu, S., Gandomi, A. H., Daneshmand, M, Muhammad, K, & Albuquerque, V. H. C. (2021) Human Memory Update Strategy: A Multi-Layer Template Update Mechanism for Remote Visual Monitoring, IEEE Transactions on Multimedia, 23, 2188-2198

[19] Liu, S., Xu, X., Zhang, Y., et al. (2022) A Reliable Sample Selection Strategy for Weakly-supervised Visual Tracking, IEEE Transactions on Reliability, online first, 10.1109/TR.2022.3162346

[20] Dai, K. , Wang, Y. , Hu, J. S. , Nam, K. , & Yin, C. (2020). Intertarget occlusion handling in multiextended target tracking based on labeled multi-Bernoulli filter using laser range finder. IEEE/ASME Transactions on Mechatronics, 25(4), 1719-1728.

[21] Belhadi A, Djenouri Y, Srivastava G, Djenouri D, Cano A, Lin JC. A two-phase anomaly detection model for secure intelligent transportation ride-hailing trajectories. IEEE Transactions on Intelligent Transportation Systems. 2020 Sep 23;22(7):4496-506.

[22] Ahmed U, Lin JC, Srivastava G. Deep Fuzzy Contrast-Set Deviation Point Representation and Trajectory Detection. IEEE Transactions on Fuzzy Systems. 2022 Aug 10.