https://doi.org/10.31449/inf.v48i23.6696 Informatica **48** (2024) 17–32

Optimized Method for Basketball Game Judging by Integrating Faster-RCNN with L-K Algorithm

Zhaohui Xie[1], Guangyu Wu[2*]
[1]School of Physical Education, Guangzhou College of Commerce Guangzhou 511363, China
[2]School of Information Technology and Engineering, Guangzhou College of Commerce Guangzhou 511363, China
E-mail: xiezhaohui799@163.com, wuguangyu2002@163.com
*Corresponding author

Keywords: computer vision, image processing, faster-RCNN, l-k, basketball

Received: July 17, 2024

The accuracy and stability of real-time target detection in computer vision during basketball games has always been a challenge. Against this background, this study first examines the shortcomings of computer vision systems for target tracking and detection. On this basis, it also introduces the faster region convolutional neural network algorithm to optimize target detection. At the same time, to further improve the target tracking capability of the model, the study incorporates the improved pyramid optical flow algorithm and refines it by applying Kalman filtering. Finally, a breakthrough target detection and tracking model is proposed by integrating the optical flow algorithm and the faster region convolutional neural network. The experimental results indicated that among 3000 actual dataset samples, the newly proposed model in the study demonstrated relatively outstanding object tracking and positioning effects, with the error between the true trajectory and the tracking trajectory being less than 3%. Furthermore, the model proposed in the study demonstrated the highest scores in the testing of three indicators: object location accuracy, average of precision and recall, and object tracking accuracy. The model achieved an object location accuracy of 93.57% and an object tracking accuracy of 91.57%, with the highest average of precision and recall reaching 95.02%. In conclusion, the novel target detection and tracking model proposed in the study demonstrates the capacity to markedly enhance the detection and recognition performance of the existing target detection and tracking model. Furthermore, it offers substantial support for the advancement of optimization methods for evaluating basketball games.

Povzetek: Predlagana metoda združuje algoritma Faster-RCNN in Lucas-Kanade (L-K) za optimizacijo sodniške presoje v košarki. Model omogoča izboljšano zaznavanje in sledenje v realnem času ter podporo pravičnosti in analizi tekem.

1 Introduction

The interaction between computer vision (CV) technology and sports is becoming more and more intimate due to the rapid development of this field. Among them, basketball game is popular among the public for its frequent offensive and defensive contents. Therefore, the establishment of an efficient optimization method of basketball game judging has profound practical significance for the realization of scientific training methods and game fairness [1]. In basketball, game judging optimization methods generally need to satisfy both real-time recognition and real-time tracking of basketballs for decision making and judging. However, this involves complex CV and image processing (IP) problems, including target detection (TD), real-time requirements, and environment adaptability. Currently, TD algorithms such as you only look once (YOLO) algorithm and faster region convolutional neural networks (Faster-RCNN) algorithm have been widely used in the field of sports [2]. Based on the enhanced YOLO v5, Liu R et al. presented a target detecting system for basketball robots. The target identification approach that the study team presented reduced the number of floating-point operations by 48.13% at 1 billion times per second, according to the testing results [3]. Fu X B et al. analyzed the feasibility of convolutional neural networks (CNNs) in real basketball sports scenarios. The study combined CNN with YOLO neural network to propose a camera-based basketball score detection method. The experimental results indicated that the method had good TD ability and has been used in several basketball courts in Beijing with good application results [4]. According to L Liu et al., technology limited the ability of traditional motion target recognition, making it impossible to obtain the desired outcomes while analyzing complex motion. Therefore, the research team combined Faster-RCNN and sports big data to propose a motion image target recognition detection system. The experimental results demonstrated that the TD system for difficult basketball movements proposed by the research team had a certain degree of effectiveness and robustness, and the detection accuracy was high [5].

However, in basketball, players often undergo drastic changes in their motion postures during the scrambling process, resulting in collisions as well as occlusions, so pure TD algorithms cannot achieve effective tracking of image targets [6]. A technique called Lucas-Kanada (L-K) can track target motion information based on image brightness while avoiding the impact of the surrounding environment on target tracking (TT). It has been explored to varying degrees by many researchers [7]. Based on L-K, Xin C et al. created a markerless measuring technique. Based on the testing results, it was possible to measure the target motion effectively and correctly without spraying high-contrast markers or speckles on the surface of the reinforced concrete structure. It could quickly diagnose the reinforced concrete structures affected by earthquakes [8]. To address the shortcomings of radiation exposure of healthy tissues around the patient's tumor due to breathing motion and delays in the linear gas pedal system, Pohl M et al. successfully tracked the initial 3D image of the patient's tumor and the predicted 3D tumor image in real time using L-K with recurrent neural networks (RNNs). Experimental results indicated that the maximum prediction error of the method was only 1.51 mm and the average cross-correlation between the original and predicted images was as high as 0.955 [9]. In their analysis of neurological illnesses resulting from reduced visual perception in adults, Jaiseeli C et al. combined the L-K approach with the effective subsampling optokinetic nystagmus (OKN) optical flow (OF) method. Experimental outcomes indicated that the OKN gain in most cases was equivalent to 1/4 of the subsampled image and was computationally efficient [10]. The current microparticle tracking velocimetry method has the disadvantage of being computationally expensive or requiring a large amount of instrumentation when recording video clips of microfluidic flow. Devasagayam J et al. used the L-K OF feature tracking technique to create a microparticle tracking velocimetry application in order to solve this issue. The program was shown to be very dependable and computationally efficient for pressure driving and end of file (EOF) in microfluidic devices based on experimental data [11]. In light of the aforementioned pertinent studies, Table 1 provides a comprehensive overview of the research theme, principal index methods, and limitations of the relevant studies.

Table 1: Summary of relevant information of relevant studies.

Author	Research theme	Main index	Insufficient	Method improvement
Liu R et al. [3]	Improved YOLO v5 Basketball Robot Target Detection Algorithm	Increased inference speed, reducing the number of floating-point operations per second by 48.13%	Performance is limited in small object detection and complex backgrounds	Introduced Faster-RCNN and multi-scale feature extraction to enhance the accuracy of small target detection
Fu X B et al. [4]	CNN-YOLO basketball scoring detection method	Good target detection capability, already applied in multiple basketball courts	There is a higher rate of false detections for targets moving at high speeds	Combined L-K optical flow algorithm with KF (Kalman Filter) to enhance the tracking effect of high-speed moving targets
L Liu [5]	Faster R-CNN neural network and motion big data detection system	Demonstrates excellent performance in detecting high-difficulty movements, with a high detection accuracy	Exhibits poor detection stability when dealing with occlusions and changes in lighting	Improved robustness by introducing ECO (efficient convolutional operator) for efficient convolutional computations and triangular algorithms
Xin C et al. [8]	Markerless vibration measurement method based on the L-K optical flow algorithm	Capable of efficient measurement under markerless conditions	The edge tracking for high-speed moving targets is not smooth enough	Improved to a pyramid L-K optical flow algorithm, and combined with a Kalman filter to enhance tracking effects
Pohl M et al. [9]	L-K-RNN	Achieved real-time tracking of tumors with a prediction error of less than 1.51mm	Sensitive to environmental lighting changes, leading to unstable tracking effects	Combined the L-K optical flow algorithm with Faster-RCNN, using Kalman filtering to enhance the stability of optical flow tracking
Jaiseeli C et al. [10]	Subsampled OKN optical flow algorithm	Improved computational efficiency, suitable for the detection of neurological diseases	Limited recognition capability for complex movements	Enhanced the recognition capability for complex movements by integrating the pyramid L-K optical flow algorithm

Devasagayam J et al. [11]	Microparticle tracking velocimetry program (L-K optical flow algorithm).	Demonstrates excellent performance in microfluidic systems, with high reliability and computational efficiency	When dealing with high-speed moving targets, the computational complexity is high	Improved L-K algorithm to a pyramid optical flow algorithm, reducing computational complexity and increasing processing speed.

To summarize, existing TT and TD methods still face many challenges. Especially in coping with the loss problem induced by the high-speed motion of the target, and the situation where the target suffers from occlusion during the motion. Although various studies have tried to enhance TD and TT in images by integrating YOLO neural network and Faster-RCNN algorithm (FRCNNA), as well as adopting new algorithms such as L-K, there is still a significant gap between the current TT and TD results and the expectations of people in real-world application scenarios such as basketball games. Therefore, the study innovatively adopts the FRCNNA with higher robustness for training and detection. Moreover, the pyramid L-K is improved by utilizing Kalman filtering (KF). The improved pyramid L-K is integrated with the FRCNNA, and finally an optimization method for judging basketball games based on CV and IP is proposed. The study aims to improve the accuracy, fairness, and real-time performance of TT and action evaluation in basketball games. This research is divided into three parts. The first part describes how the Faster-RCNN TT algorithm with L-K is improved and how the optimal design model is built, respectively. The second part is the performance test of the new model. The last part summarizes the article.

2 Methods and materials

Aiming at the existing problems of TT and action judging in basketball games, the study first introduces the basic framework of FRCNNA. Moreover, by introducing multi-scale features, the basketball in each camera viewpoint is detected in two dimensions. In addition, the study also adopts the pyramid L-K as the basis of the framework from the perspective of basketball game action judging, and introduces the KF for further optimization. With these improvements, the study finally proposes a novel optimization method for basketball game judging based on CV and IP.

2.1 Establishment of target detection algorithm based on Faster-RCNN

The existing sports TD process is roughly sequenced as target feature extraction, target recognition, and target localization. Each link is interconnected and inseparable [12-13]. However, picture occlusion has always been a major challenge in the field of sports TD. Especially in basketball games, tracking targets often encounter temporary disappearance [14]. Therefore, how to effectively and reasonably recognize and detect the moving targets in the temporary disappearing state is the key point to overcome the occlusion problem during TT [15]. FRCNNA inherits the advantages of the first-generation regions with convolutional neural network (RCNN) features and the second generation based on Faster-RCNN, which has a powerful target recognition and TD capability [16-17]. Figure 1 depicts the FRCNNA structure.

In Figure 1, the structure of the FRCNNA is composed of four main parts, namely the regional proposal network (RPN) layer, the convolutional layer, the classification and regression layer, and the region of interest pooling (ROI pooling) layer [18]. First, the convolutional layer performs feature extraction on the input motion image data to obtain the corresponding feature image data. Then, the RPN structure is utilized to generate candidate frames. Secondly, the ROI pooling layer solves the problem of inconsistent feature image sizes through the pyramid structure and utilizes the downsampling operation to change the different sized region proposals into the same sized output. Finally, the classification and regression layer transfer the adjusted feature maps to the fully-connected layer and obtains

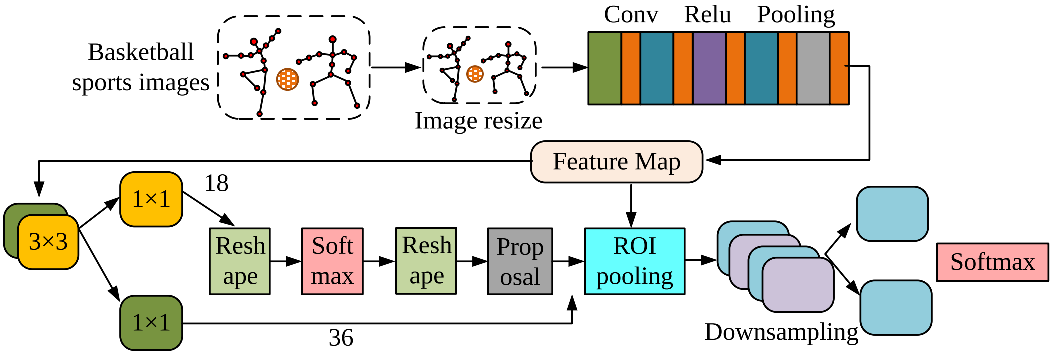


Figure 1: Structure of the FRCNNA.

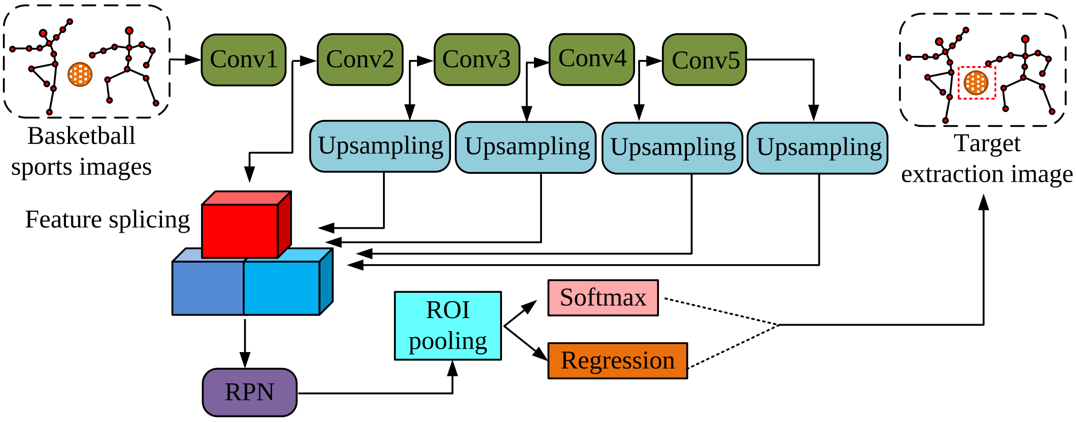


Figure 2: Flowchart of 2D inspection framework.

the final prediction results through a series of fully-connected layer outputs. Equation (1) displays the formula for the convolution operation.

$$g(i) = f\left(\sum_{x=1}^{n}\sum_{y=1}^{n} a_{x,y} \times W_{x,y}^{i} + b^{i}\right) \tag{1}$$

In Equation (1), $g(i)$ is the convolution of the i th node of the output feature mapping matrix. $a_{x,y}$ and b represent the pixel values and bias terms of the input image coordinates as (x, y), respectively. W and f then represent the weights and Sigmoid activation function, respectively. The Sigmoid activation function is calculated as shown in Equation (2).

$$f(x) = \frac{1}{1+e^{-x}} \tag{2}$$

In Equation (2), e represents the numerical value and x represents the input of the fully connected layer. However, tracking balls in sporting events, such as basketball, soccer and table tennis, is a challenging task. Small-sized basketballs cannot be effectively distinguished from other small-sized objects such as sneakers, light reflection points, or spectators' heads with only a limited number of visible features due to their fewer surface features. In addition, the difficulty of accurate tracking will be further increased when the basketball is controlled by a player or obscured in motion [19-20]. Therefore, in order to solve such problems, the study introduces multi-scale features to improve the Faster-RCNN TD algorithm and performs 2D detection of basketballs. The flowchart of the 2D detection framework is shown in Figure 2. In Figure 2, the visual geometry group-16 (VGG 16) network in the Faster-RCNN TD algorithm is firstly divided into five layers. After layer-by-layer feature extraction, the output size of the last image layer is reduced to 1/16 of the original image size. Secondly, multiple features of different scales are then spliced and input into the RPN layer to obtain candidate features. Ultimately, the fully linked classification and regression layer outputs the findings. The feature value weight w_i is calculated as shown in Equation (3).

$$w_i = \frac{e^{d_i}}{\sum_{j\in R} e^{d_j}} \tag{3}$$

In Equation (3), d and R represent feature maps and local regions, respectively. The eigenvalue weights ensure the transfer of important features. In the reverse transfer, the feature values in the region are predefined with a minimum gradient [21-22]. However, due to the effect of occlusion, there will still always be the condition of misdetection and missed detection. Efficient convolution operators (ECO) is an advanced convolution operation method, which can not only improve the efficiency of convolution operation, but also effectively deal with partial occlusion as well as short-time occlusion [23-24]. Therefore, the study introduces the ECO method to improve the Faster-RCNN TD algorithm. Furthermore, a TD algorithm based on Faster-RCNN, i.e., Faster-RCNN-ECO-TA, is proposed by fusing the 2D coordinates of basketball into 3D coordinates through triangle algorithm (TA). Figure 3 displays the TA schematic diagram.

In Figure 3, the points in two different camera planes can be identified as one 3D spatial point. Among them, points $p_1^{'}$ and $p_2^{'}$ represent the counter-projection points of p_{12} in different cameras, respectively. Their distances from the original points p_2 and p_1 are referred to as the back-projection errors. The study takes the average of these two back-projection distances as the error of the 3D coordinate points obtained by the TA. The expression for calculating the 3D coordinates of the basketball is shown in Equation (4).

$$p_t = \frac{1}{N}\sum_{1\le i,j\le n} (p_t^{ij} \left| e_t^{ij} < \tau \right.) \tag{4}$$

In Equation (4), p_t represents the 3D coordinate fusion result of the basketball at the t moment. N represents the number of all camera-matched pairs with inverse projection error less than the threshold τ. p_t^{ij} and e_t^{ij} represent the 3D coordinates and the inverse projection error calculated by TA, respectively. The probability maximization $\hat{m}$ expression is shown in Equation (5).

$$\hat{m} = \arg\max p(X_i^t = x_i^t, Y_i^t = Y_{j\in N(i)}^{t+1} \,|\, I^t) \quad (5)$$

In Equation (5), X_i^t and x_i^t represent the presence or absence of a basketball player and are taken as 0 or 1. Y_i^t and $Y_{j\in N(i)}^{t+1}$ represent the appearance characteristics of the basketball player. The flow framework diagram of Faster-RCNN-ECO-TA TD algorithm is shown in Figure 4.

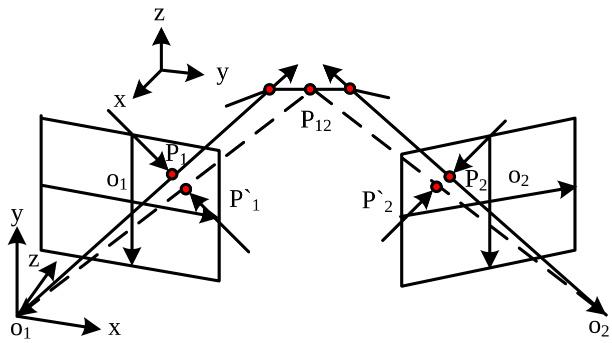


Figure 3: Schematic diagram of the TA.

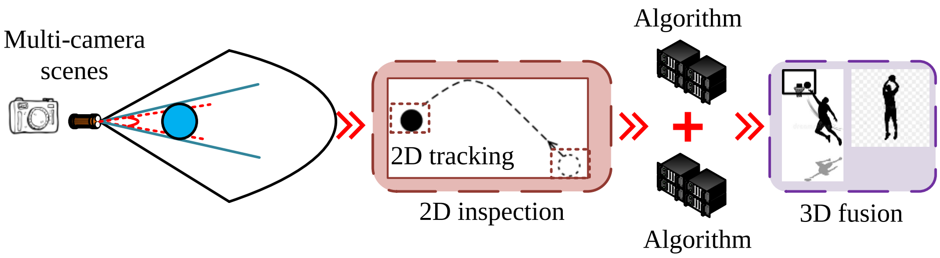


Figure 4: Process framework diagram of Faster-RCNN-ECO-TA target detection algorithm.

In Figure 4, the flow of the Faster-RCNN-ECO-TA TD algorithm proposed by the study is mainly divided into four phases, that is, the basketball 2D detection phase, the basketball 2D tracking phase, the basketball 3D coordinate fusion phase and the basketball 3D trajectory smoothing phase. The study firstly detects the basketball in 2D based on the Faster-RCNN TD algorithm. Then, ECO efficient convolutional operation is used to track the basketball in different camera views in 2D. Next, TA is employed to effectively fuse multiple 2D coordinates into one 3D coordinate. Finally, the optimal linear state estimation method is utilized to process to obtain a smooth 3D trajectory of the basketball.

2.2 Modeling of target detection and tracking based on L-K and Faster-RCNN

Although the Faster-RCNN-ECO-TA TD algorithm can provide accurate information about the initial position of the target for player TT operations, it is not yet able to effectively solve the excessive interference due to changes in ambient lighting. L-K can calculate an object's motion information between adjacent frames, mitigating the impact of changes in ambient lighting [25-26]. The expression for calculating the conditional constraint ∇Egw of L-K is shown in Equation (6).

$$\nabla Egw = \frac{\partial E}{\partial x}\frac{dx}{dt} + \frac{\partial E}{\partial y}\frac{dy}{dt} \quad (6)$$

In Equation (6), $\frac{dx}{dt}$ and $\frac{dy}{dt}$ denote the moving components of the OF in the x horizontal and y vertical directions, respectively. E represents the corresponding point brightness. However, when L-K tracks a fast-moving target, the OF at the edge of the object tends to be unsmooth, which leads to unsatisfactory TT results [27-28]. To address this situation, the pyramid OF algorithm is introduced into the L-K sparse OF algorithm to reduce the computational complexity. The structure of the L-K pyramid optical flow algorithm (L-K-POFA) is shown in Figure 5.

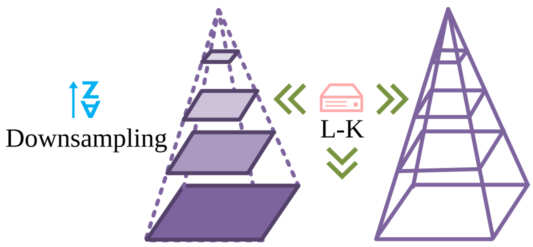


Figure 5: Structure of the L-K-POFA.

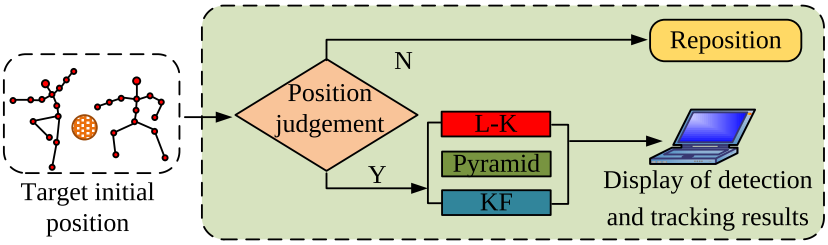


Figure 6: KF-L-K-POFA flow structure.

In Figure 5, the L-K-POFA employs continuous downsampling of images to form a collection of images from all the image data and processes the images using a multi-resolution approach [29]. It promotes the highest image resolution at the bottom and the lowest resolution at the top. In this way, the fast-moving targets can be reduced with the resolution to finally meet the calculation range, which makes the detection and TT effect of fast and displaced targets to be improved substantially. The expression for calculating the OF $I(x)$ of a pixel point is shown in Equation (7).

$$I(x) = I^L\left(x + u / 2^L\right) \tag{7}$$

In Equation (7), L and u represent the neighborhood and mobile components, respectively. The horizontal and vertical bias formulas are shown in Equation (8).

$$\begin{cases} I_x = \left[-I(x-1, y) + I(x+1, y)\right] / 2 \\ I_y = \left[-I(x, y-1) + I(x, y+1)\right] / 2 \end{cases} \tag{8}$$

In Equation (8), I_y and I_x represent the OF deflections in the vertical and horizontal directions, respectively. The spatial gradient matrix G is calculated as shown in Equation (9).

$$G = \sum_{x=-W_x}^{W_x} \sum_{y=-W_y}^{W_y} \begin{bmatrix} I_x^2 & \cdots & yI_xI_y \\ \vdots & \ddots & \vdots \\ yI_xI_y & \cdots & y^2I_x^2 \end{bmatrix} \tag{9}$$

In Equation (9), W_x and W_y represent the weight functions of each pixel in the window in the horizontal and vertical neighborhoods, respectively. The error vector $b*$ for two consecutive image frames is calculated as shown in Equation (10).

$$b* = \sum_{x=-W_x}^{W_x} \sum_{y=-W_y}^{W_y} \begin{bmatrix} I_x\delta I \\ \vdots \\ yI_y\delta I \end{bmatrix} \tag{10}$$

In Equation (10), δ represents the image gray value, and the rest of the algebraic meanings are the same as before. In order to meet the real-time requirements when tracking moving targets, the study also introduces the KF to improve the L-K-POFA. The flow structure of the improved KF-L-K-POFA is shown in Figure 6.

First, the basketball game image in Figure 6 determines the target player's starting position. All the images are formed into an image collection by the mechanism that the

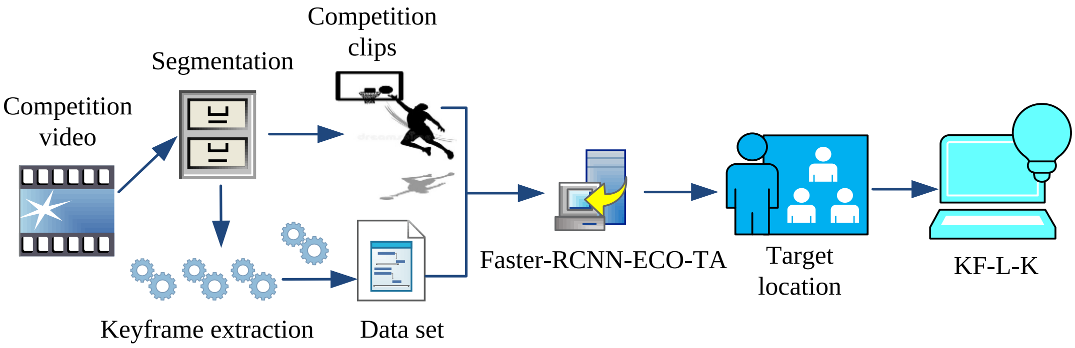


Figure 7: L-K and Faster-RCNN based model structure flow.

image pyramid can continuously downsample the original image. Secondly, the target image tracking is performed using L-K-POFA, and the KF filter enhances the OF tracking effect. Finally, the TT results are output. The state equation x_k and observation equation Z_k computational expressions of KF are shown in Equation (11).

$$\begin{cases} x_k = A_k x_{k-1} + \mu_k \\ Z_k = H_k x_{k-1} + v_k \end{cases} \tag{11}$$

In Equation (11), A_k and H_k represent the state transfer matrix and observation matrix, respectively. μ_k and v_k represent the Gaussian noise covariance. The study synthesizes the improvements of the above modules and finally proposes a novel target detection and tracking model (TDTM) based on L-K and Faster-RCNN, aiming to provide new insights in the area of action evaluation in the field of basketball games. The flow structure of this TDTM is shown in Figure 7.

In Figure 7, the proposed novel TDTM based on L-K and Faster-RCNN is mainly composed of four phases, i.e., video image preprocessing phase, labeling data and building dataset phase, TD phase and TT phase. Among them, the video image preprocessing stage is divided into video shot segmentation and video key frame extraction. First of all, the study adopts a clustering-based video lens segmentation algorithm to judge the boundary of the lens through the differences between consecutive frame images for segmentation. Furthermore, the study adopts the content-based key frame extraction method to reflect a number of key frames that reflect the main content of the current shot. Then, the acquired images are size normalized to build the dataset. Next, TD using Faster-RCNN-ECO-TA is performed to determine the initial position of the target player. Finally, real-time TT is experimented by the KF-L-K-POFA. A common evaluation metric for TT algorithms is multiple object tracking accuracy (MOTA) [30]. Equation (12) displays the formula for calculating MOTA.

$$MOTA = 1 - \frac{\sum_t (c_1 \cdot fn_t + c_2 \cdot fp_t + c_3 \cdot idsw_t)}{\sum_t g_t} \tag{12}$$

In Equation (12), c_1, c_2 and c_3 represent constants. g_t represents the true value. fn_t, fp_t and $idsw_t$ represent the missed detection, the false detection and the identity exchange, respectively.

3 Results

To validate the performance of the proposed Faster-RCNN-ECO-TA TD algorithm and the TDTM based on L-K and Faster-RCNN, the study firstly builds a suitable experimental environment and preprocesses the test data. A portion of the data is utilized to train the model. Secondly, the performance and simulation experiments are conducted to test the Faster-RCNN-ECO-TA TD algorithm and the TDTM based on L-K and Faster-RCNN, respectively.

3.1 Performance test of Faster-RCNN-ECO-TA target detection algorithm

The study utilizes the Ubuntu 16.04 LTS operating system, equipped with an Intel Core i7 CPU, NVIDIA GeForce GPU, 64GB of RAM, and programmed with TensorFlow-GPU. The iteration is set to 300, the IoU (Intersection over Union) threshold to 0.5, the learning rate to 0.001, and the batch size to 16. The Chinese basketball match live video image dataset and the NTU RGB+D60 dataset are selected as the sources of test data. The Chinese basketball match live video image dataset contains a total of 3000 images. The NTU RGB+D60 dataset, introduced by Nanyang Technological University in Singapore, is a multimodal behavior recognition collection that includes 60 types of action categories. It is particularly suitable for the analysis of human skeletal motion, covering daily activities and sports actions including basketball. The research divides these datasets into training and testing sets in an 8:2 ratio. Furthermore, the test set is evenly divided into three categories: Test

Set 1 is for images under normal lighting and clarity conditions. Test Set 2 is for increased image blurriness with constant lighting. Test Set 3 is for changed lighting intensity with constant image blurriness, including strong and weak light conditions. To verify the impact of each module in the Faster R-CNN-ECO-TA object detection algorithm on the overall performance, the study first conducted an ablation test using the object detection accuracy as the metric. The results are shown in Figure 8.

Figure 8(a) shows the detection accuracy curve of each module in the Faster-RCNN-ECO-TA TD algorithm in the TrS. Figure 8(b) shows the detection accuracy curve of each module in the Faster-RCNN-ECO-TA TD algorithm in the TeS. The FRCNNA alone performs poorly, with the highest detection accuracy close to 92%. The inclusion of the TA 3D coordinate fusion module and the ECO efficient convolution module greatly enhances the algorithm's overall performance. Compared to Faster-RCNN or ECO-TA, the performance is improved by about 6%. The research proposed Faster-RCNN-ECO-TA has the best overall performance. Its best performance in the TrS is 98.28% and in the TeS is 98.51%. In summary, ECO has demonstrated excellent performance in handling partial and temporary occlusions encountered in basketball motion, which has significantly improved the model's TD capabilities. In addition, TA's integration and upgrade of the basketball's two-dimensional coordinate data to three-dimensional coordinates further improved the model's accuracy and reliability. Each component of the modules proposed by the research had a positive impact on the final model, effectively improving the accuracy of the recognition model. This comprehensive optimization approach has proven to be effective in improving model performance, especially in terms of target recognition accuracy. In addition, in order to verify the performance difference between the research-proposed Faster-RCNN-ECO-TA TD algorithm and the more popular algorithms of the same type, the study introduces adaptive boosting (Adaboost), YOLO TD algorithm and RCNN TD algorithm with the average detection time as a reference index. The comparison results of each algorithm in the TrS and TeS are shown in Figure 9.

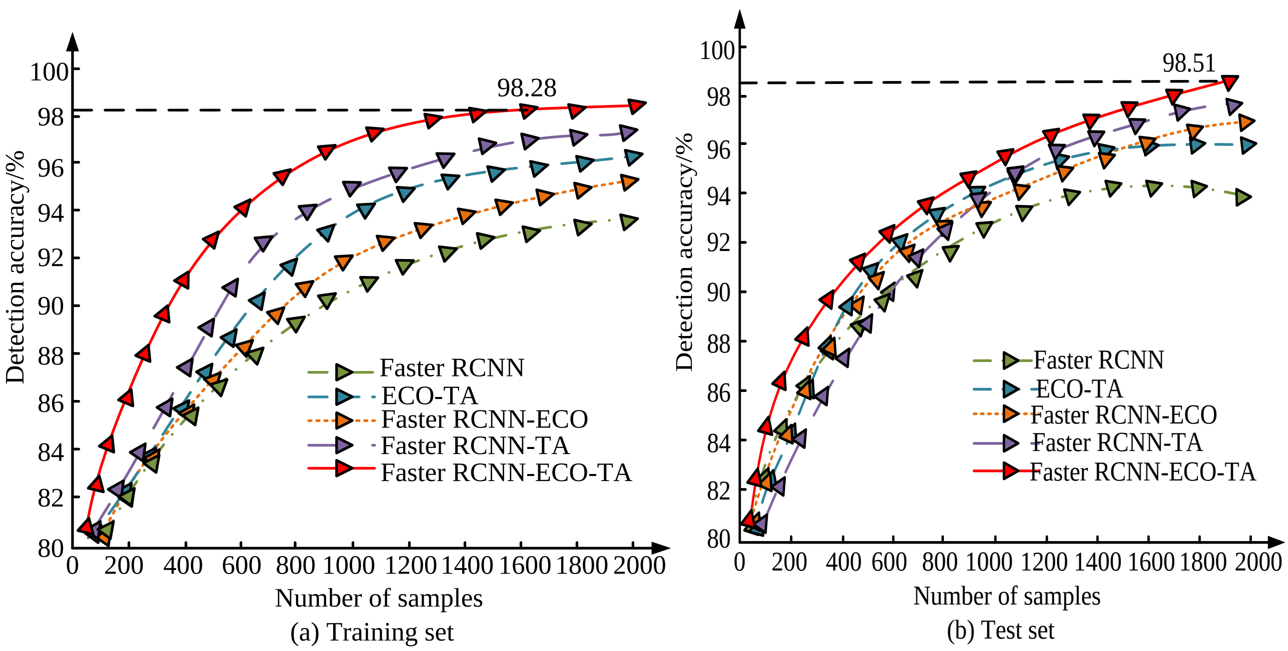
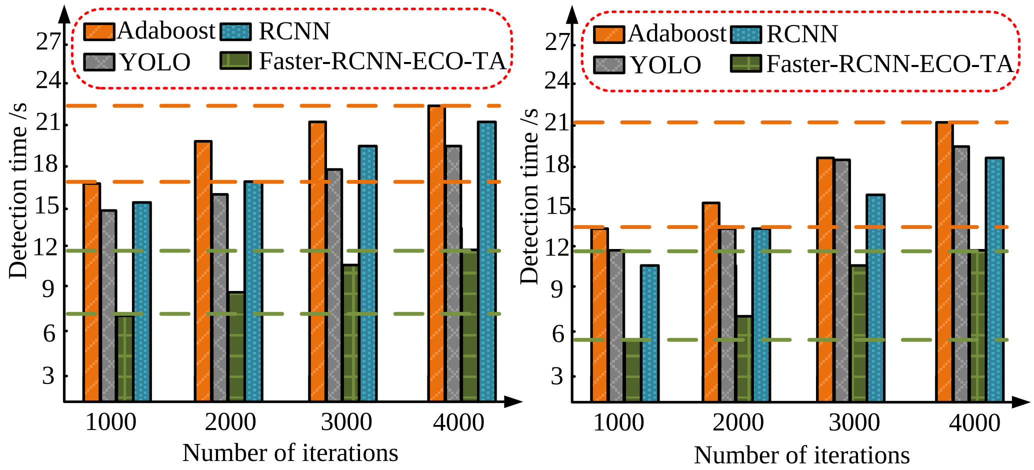


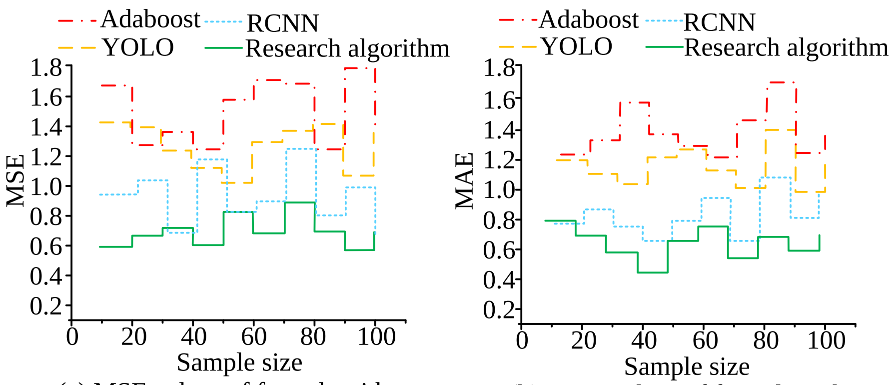
Figure 8: The performance impact of different modules on the model.



(a) Training set (b) Test set

Figure 9: Detection time comparison results of different target detection algorithms.

Figure 9(a) shows the detection time comparison results of different TD algorithms in the TrS. Figure 9(b) shows the detection time comparison outcomes of different TD algorithms in the TeS. The proposed Faster-RCNN-ECO-TA TD algorithm has the least detection time. Its average running time is 5.6s, 7.1s, 10.2s, and 11.8s for 1000, 2000, 3000 and 4000 iterations, respectively. Since the TeS has the fewest number of parameters, its corresponding running time is the shortest. However, the Adaboost TD algorithm takes the longest time to run, which is due to the fact that the Adaboost algorithm is susceptible to different degrees of interference from the complex environment during the detection process, which leads to a longer inference time. It can be concluded that the proposed Faster-RCNN-ECO-TA TD algorithm combines good operational efficiency with



(a) MSE values of four algorithms with different data (b) MAE values of four algorithms with different data

Figure 10: Error performance of different algorithms.

Table 2: Comparison test results of multiple indicators.

Data set	Algorithm	P/%	R/%	F1/%	Accuracy/%
Test set 1	Adaboost	72.58	70.05	67.98	63.87
	YOLO	74.67	74.27	75.28	76.75
	RCNN	79.09	77.27	76.87	76.41
	Faster-RCNN-ECO-TA	95.68	93.67	95.16	92.98
Test set 2	Adaboost	68.15	67.85	67.66	60.37
	YOLO	70.26	69.78	70.11	68.24
	RCNN	73.25	73.06	73.58	72.65
	Faster-RCNN-ECO-TA	92.65	91.09	92.88	92.03
Test set 3	Adaboost	65.58	64.37	65.87	58.68
	YOLO	66.89	65.26	66.32	60.07
	RCNN	69.78	70.01	69.54	66.37
	Faster-RCNN-ECO-TA	90.66	90.27	90.23	91.27

detection accuracy. To explore the error performance of different algorithms, the study also tested the four algorithms in terms of mean squared error (MSE) and mean absolute error (MAE), respectively. The test results are shown in Figure 10.

Figure 10(a) shows the variation curves of MSE values for four TD algorithms with different samples. The MSE values of the four algorithms change with the increase of the samples. Among them, the proposed Faster-RCNN-ECO-TA TD algorithm has the smallest range of variation in MSE value, floating between 0.5 and 0.9. The other three commonly used TD algorithms all have larger error variations. Figure 10(b) shows the variation curves of MAE values of the four TD algorithms with different samples. The MAE values of the four algorithms varied as the samples increased. Among them, the MAE value fluctuation of the Faster-RCNN-ECO-TA TD algorithm proposed in the study is also the smallest, fluctuating between 0.4 and 0.8. This indicates that the improved Faster-RCNN can have a more accurate measurement with less error impact. To quantify the performance comparison results of each TD algorithm more accurately, the study continues to test the above TD algorithms using precision, recall, F1-score, and accuracy as reference indexes. Table 2 displays the test results.In Table 2, both increasing blurring and stronger and weaker light reduce the precision of the TD algorithm. However, in the three datasets with different light intensities and blurring levels, the performance of the Faster-RCNN-ECO-TA TD algorithm proposed in the study performs optimally. Its P-value, R-value, and F1-score under normal illumination are 95.58%, 93.67% and 95.16%, respectively. Even when the blurring level increases and the light becomes stronger and weaker, its P-value, R-value, and F1-score decrease by less than 5%. Compared with the Adaboost TD algorithm, the P-value, R-value and F1-score of the proposed Faster-RCNN-ECO-TA TD algorithm under normal lighting conditions are improved by 23.10%, 23.62%, and 27.18%, respectively. The above data indicate that the fast movement of players during basketball games can cause the performance of the TD algorithm to be affected to different degrees. Nonetheless, the Faster-RCNN-ECO-TA TD algorithm proposed in the study is almost unaffected and adapted to target localization recognition during basketball games.

3.2 Simulation test of target detection and tracking model based on L-K and Faster-RCNN

From the above test results, the Faster-RCNN-ECO-TA TD algorithm performs well in target localization and detection. However, the data is limited to image data collected from video, and the persuasiveness and feasibility of the results for dynamic video data still need to be further strengthened. Moreover, the performance of the proposed TDTM based on L-K and Faster-RCNN has not yet been verified. Therefore, the study attempts to test this using a homemade video dataset. The action samples in the homemade dataset include eight common basketball action patterns in basketball games: crotch dribbling, layup with the ball, hitting foul, pulling foul, resting the ball into the basket, dribbling behind the back (switching hands), and rebounding outside the basket with an airball and a long pass. The number of videos for each basketball maneuver is 180, totaling 2000 video sequences. Each video sequence has a frame rate of 25fps, a resolution of 160x120, and an average length of 25 seconds. In accordance with the 6:4 ratio, it is further separated into TrS and TeS. Moreover, the off-threshold is set to 200cm, and the study first tests the model with tracking precision as a metric. Figure 11 presents the test findings.

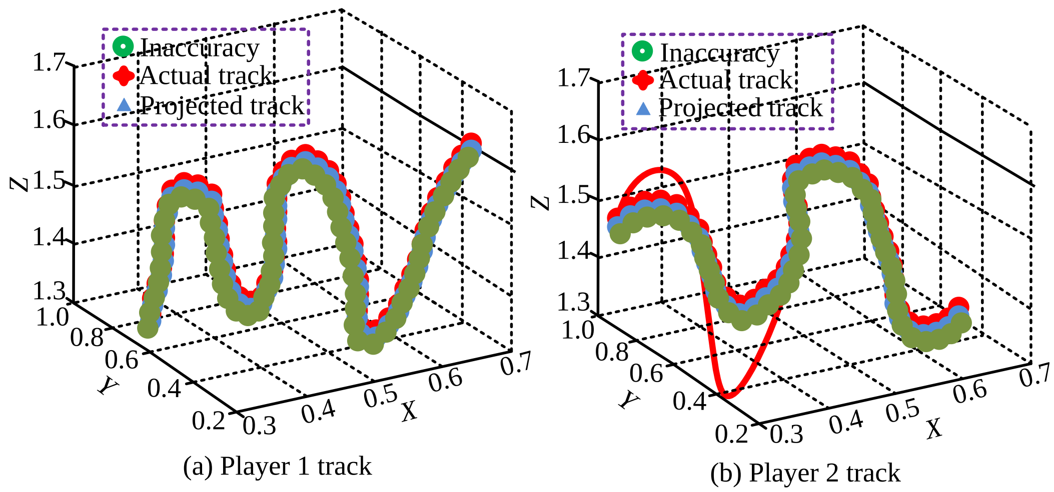


Figure 11: 3D plot of the trajectory of the player's centroid position over time.

over time in a basketball game. Figure 11(b) shows the 3D plot of the trajectory of the center point position of player 2 over time in a basketball game. The real and tracked trajectories of the proposed L-K and Faster-RCNN based TDTM have a high degree of overlap. The confrontation between the two players also does not affect the tracking and localization effects of the model, and the error between the true trajectory and the tracking trajectory is less than 3%. It shows that the proposed model of the study is robust to the phenomena of omission and misdetection that often occur in TD. The L-K-POFA mitigates tracking drift and achieves stable TT during basketball games. In addition, the study also introduces the same types of more popular TT models for comparison, such as visual object tracking (VOT), multiple object tracking (MOT), and multi-camera multi-object tracking (MCMOT). The classification effect of different basketball actions is tested with classification accuracy and classification efficiency as the reference index. The classification effect is shown in Figure 12.

Figure 12(a) and 12(b) shows the classification performance of different models on basketball movements in the TrS and TeS, respectively. Both in the TrS and TeS, the TDTM based on L-K and Faster-RCNN proposed by the research performs the best, followed by MCMOT and MOT models, and VOT model performs the worst. The

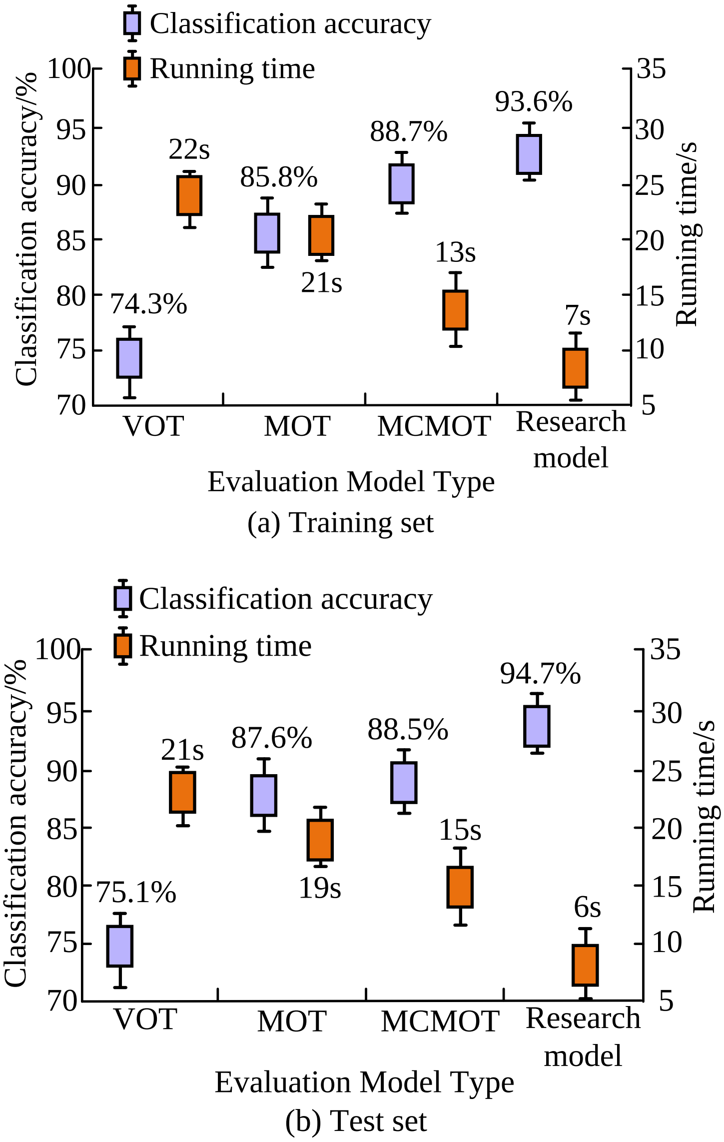


Figure 12: Effectiveness of different models for classification of basketball movements.

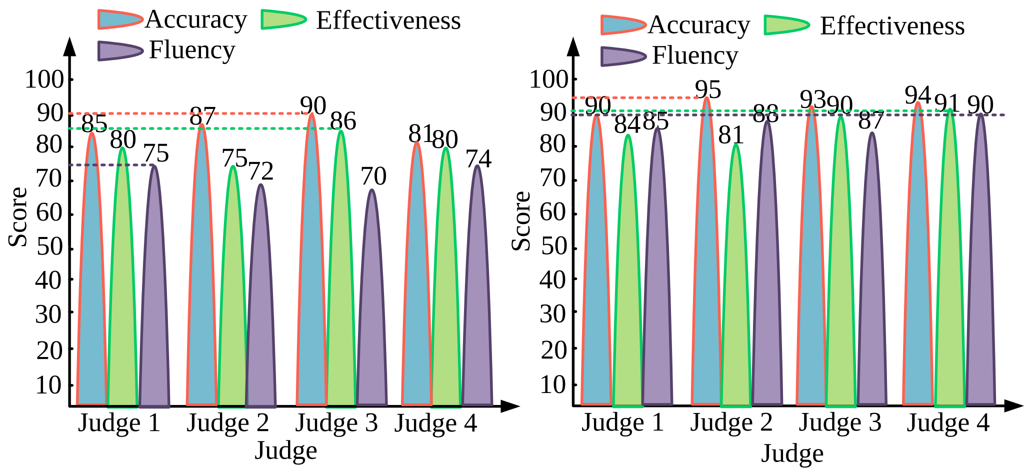
Table 3: Comparison test results of multiple indicators.

Data set	Method	MOTP/%	IDF1/%	MOTA/%
Training set	VOT	70.83	74.43	73.35
	MOT	75.64	75.18	76.98
	MCMOT	78.85	76.39	76.28
	Research model	91.27	93.58	89.76
Test set	VOT	73.73	76.66	76.87
	MOT	78.86	78.67	78.58
	MCMOT	80.09	80.11	79.99
	Research model	93.57	95.02	91.57

classification accuracy of the TDTM, MCMOT model, MOT model, and VOT model based on L-K and Faster-RCNN in the TrS are 74.3%, 85.8%, 88.7%, and 93.6%, respectively. The classification times are 22s, 21s, 13s, and 7s, respectively. The above data indicates that the proposed method has certain effectiveness and performs better in models of the same type. The study also conducts tests using IDF1, which harmonizes the mean of accuracy and recall, as well as multiple object tracking precision (MOTP) and MOTA, which measure the accuracy of target location. Table 3 displays the test results.

In Table 3, among the three indicators of MOTP, IDF1, and MOTA, the proposed object detection and tracking model based on L-K and Faster-RCNN has the highest overall performance score. The TrS and TeS also shows the best recognition performance of the proposed model. The highest MOTP value of the object detection and tracking model based on L-K and Faster-RCNN is 93.57%. The highest IDF1 value is 95.02%, and the highest MOTA value is 91.57%. In summary, the proposed object detection and tracking model based on L-K and Faster-RCNN has certain feasibility and effectiveness in target recognition in basketball games, which can efficiently perform object detection and improve the efficiency of object detection. The research ultimately uses recognition accuracy, effectiveness, and fluency as reference indicators. Moreover, 24 evaluation results from 4 judges, including basketball experts, basketball coaches, and basketball players, are collected through expert scoring method. The average scores of each indicator are shown in Figure 13.

Figures 13(a) and 13(b) show the scoring results of the MCMOT and the model proposed in the research. The recognition accuracy of the MCMOT model is relatively high, but its recognition fluency score is low. This rating indicates that the model still faces challenges in terms of computational coherence and video frame processing when recognizing video actions. The three indicators of the model proposed in the research have relatively average scores. Among them, the highest score for recognition accuracy is 95 points, the highest score for effectiveness is 91 points, and the highest score for fluency is 90 points. In summary, the public has a higher preference for the model proposed in the research, with better applicability and model effectiveness.



(a) MCMOT (b) Our model

Figure 13: Average rating results for each indicator.

4 Discussion

The fairness and entertainment value of basketball games largely depend on the accuracy of the referees. Therefore, to achieve automatic tracking and detection of key events during basketball games, the research proposed a faster R-CNN-ECO-TA object detection algorithm using multi-scale features, efficient convolutional operators, and triangular algorithms. The Faster-R-CNN algorithm's superior detection speed and accuracy, in conjunction with the capacity of efficient convolutional operators to augment detection efficiency, have resulted in the proposed algorithm sustaining P-values, R-values, and F1 scores above 90% even in the presence of increased blur and varying light intensities. The research proposed by Martin P E et al. also confirmed the powerful capability of the Faster-R-CNN algorithm in detecting small target ball sports, with an accuracy rate as high as 92.6% [31]. However, a single object detection algorithm still has limitations when dealing with rapid motion and occlusion issues in basketball games [32]. To address this issue, the research introduced a pyramid L-K OF algorithm integrated with a KF filter on the basis of the Faster R-CNN-ECO-TA object detection algorithm. This resulted in the proposal of a TD and tracking model based on L-K and Faster-R-CNN. By combining the player position information detected by Faster-R-CNN and the OF tracking of the L-K algorithm, the model could more accurately capture the motion state of players on the field, maintaining a high tracking accuracy even under conditions of rapid motion or partial occlusion. Thus, the research model achieved a 93.6% classification accuracy, with an error of less than 3% between the actual trajectory and the tracking trajectory. The research proposed by Fang N et al. also showed that the L-K algorithm achieved an action classification accuracy of 88.9%, further verifying the advantage of the L-K OF algorithm in tracking effects. It can be concluded that the method proposed by the research not only provides a novel technical approach for the assessment of basketball games but also serves as a reference for other fields that necessitate the tracking of targets in a rapid and precise manner. With the continuous development and improvement of technology, automated sports game judging methods will play a greater role in the future.

5 Conclusion

With the popularization and popularity of basketball games, the demand for automated target action recognition is becoming increasingly urgent. However, due to the complexity and diversity of actions during basketball games, traditional action recognition algorithms face significant challenges in video evaluation of basketball games. In view of this, a design framework for object detection algorithm based on FRCNNA was studied. On this basis, a pyramid L-K fused with KF filter was introduced for precision adjustment and object tracking. Finally, a novel object detection and tracking model based on L-K and Faster-RCNN was proposed. The outcomes indicated that compared with similar object detection algorithms, the Faster-RCNN-ECO-TA object detection algorithm proposed in the research had the least detection time, with an average running time of only 5.6s at 1000 iterations. The MAE and MSE values of this algorithm also achieved the smallest range of variation, fluctuating only between 0.4-0.9. Even with an increase in blur level and a decrease in light intensity, the P-value, R-value, and F1 score of the Faster-RCNN-ECO-TA object detection algorithm did not exceed 5%. The simulation test outcomes revealed that the TDTM proposed in the research based on L-K and Faster-RCNN had a high degree of coincidence between the real trajectory and the tracking trajectory. The error between the real trajectory and the tracked trajectory was less than 3%, and the accuracy of identifying and classifying different basketball movements reached 93.6%. In addition, after unanimous evaluation by professional basketball judges, it was found that the recognition accuracy, effectiveness, and fluency scores of the model proposed in the research were generally high, with the highest scores reaching 95 points, 91 points, and 90 points, respectively. The results of this study are anticipated to propel the advancement of TD technology in basketball games to a new frontier, offering novel avenues for enhancing the precision of TT and action assessment in basketball games. However, this study employs a relatively limited set of scenarios. In the future, an expansion to more diverse and complex scenarios will be undertaken with the aim of further enhancing the universality and comprehensiveness of object detection and tracking.

References

[1] G. Eberle, J. Bourlot, and C. Martínez, "Automatic extraction of heat maps and goal instances of a basketball game using video processing," Computer Science - CACIC 2021, vol. 1584, pp. 95-105, 2022. https://doi.org/10.1007/978-3-031-05903-2_7

[2] K. Host, M. Pobar, and M. Ivasic-Kos, "Analysis of movement and activities of handball players using deep neural networks," Journal of Imaging, vol. 9, no. 4, pp. 1-18, 2023. https://doi.org/10.3390/jimaging9040080

[3] R. Liu, H. Wang, S. Zhang, and C. Chen, "Object detection algorithm based onimproved YOLOv5 for basketball robot," Chinese Intelligent Systems Conference, vol. 951, pp. 49-63, 2022. https://doi.org/10.1007/978-981-19-6226-4_7

[4] X. B. Fu, S. L. Yue, and D. Y. Pan, "Camera-based basketball scoring detection using convolutional neural network," International Journal of Automation and Computing, vol. 18, no. 2, pp. 266-276, 2021. https://doi.org/10.1007/s11633-020-1259-7

[5] L. Liu, "Objects detection toward complicated high remote basketball sports by leveraging deep CNN architecture," Future Generation Computer Systems, vol. 119, no. 6, pp. 31-36, 2021. https://doi.org/10.1016/j.future.2021.01.020

[6] M. Hu, and Q. Hu, "Design of basketball game image acquisition and processing system based on machine vision and image processor," Microprocessors and Microsystems, vol. 82, no. 4, pp. 103-114, 2021. https://doi.org/10.1016/j.micpro.2021.103904

[7] D. Tian, Y. Han, S. Wang, X. Chen, and T. Guan, "Absolute size IoU loss for the bounding box regression of the object detection," Neurocomputing, vol. 500, no. 8, pp. 1029-1040, 2022. https://doi.org/10.1016/j.neucom.2022.06.018

[8] C. Xin, C. Wang, Z. Xu, M. Qin, and M. He, "Marker-free vision-based method for vibration measurements of RC structure under seismic vibration," Earthquake Engineering & Structural Dynamics, vol. 51, no. 8, pp. 1773-1793, 2022. https://doi.org/10.1002/eqe.3637

[9] M. Pohl, M. Uesaka, K. Demachi, and R. B. Chhatkuli, "Prediction of the motion of chest internal points using a recurrent neural network trained with real-time recurrent learning for latency compensation in lung cancer radiotherapy," Computerized Medical Imaging and Graphics, vol. 91, no. 2021, pp. 101941, 2021. https://doi.org/10.1016/j.compmedimag.2021.101941

[10] C. Jaiseeli, and N. R. Raajan, "SLKOF: Subsampled lucas-kanade optical flow for opto kinetic nystagmus detection," Journal of Intelligent & Fuzzy Systems, vol. 41, no. 5, pp. 5265-5274, 2021. https://doi.org/10.3233/JIFS-189849

[11] J. Devasagayam, R. Bosma, and C. M. Collier, "A velocity program using the Kanade-Lucas-Tomasi feature-tracking algorithm with demonstration for pressure and electroosmosis conditions," Electrophoresis, vol. 43, no. 7-8, pp. 865-878, 2022. https://doi.org/10.1002/elps.202100177

[12] B. Yang, Z. Chang, and Y. Chen, "Falling-point recognition and scoring algorithm in table tennis using dual-channel target motion detection," Journal of Healthcare Engineering, vol. 7, no. 2, pp. 9892847, 2021. https://doi.org/10.1155/2023/9892847

[13] H. Li, and D. Li, "Recognition and optimization analysis of urban public sports facilities based on intelligent image processing," Computational Intelligence and Neuroscience, vol. 2021, pp. 8948248.1-8948248.13, 2021. https://doi.org/10.1155/2021/8948248

[14] X. Yin, and L. Chen, "Image object detection method based on improved faster R-CNN," Journal of Circuits, Systems and Computers, vol. 33, no. 7, pp. 54-59, 2024. https://doi.org/10.1142/S0218126624501305

[15] I. A. Kandhro, S. Manickam, K. Fatima, M. Uddin, U. Malik, A. Naz, and A. Dandoush, "Performance evaluation of E-VGG19 model: Enhancing real-time skin cancer detection and classification," Heliyon, vol. 10, no. 10, pp. e31488.1-e31488.20, 2024. https://doi.org/10.1016/j.heliyon.2024.e31488

[16] M. J. Awan, O. A. Masood, M. A. Mohammed, A. Yasin, A. M. Zain, R. Damaševičius, and K. H. Abdulkareem, "Image-based malware classification using VGG19 network and spatial convolutional attention," Electronics, vol. 10, no. 19, pp. 2444-2454, 2021. https://doi.org/10.3390/electronics10192444

[17] S. Verma, G. Singh, H. M. Salman, A. Pavani, and S. Srikanth, "Detection of traffic sign using inception V3 in comparison with VGG-19 to measure accuracy," ICACITE, vol. 32, pp. 730-733, 2023. https://doi.org/10.1109/ICACITE57410.2023.10183243

[18] N. Darapaneni, P. Kumar, and N. Malhotra, "Detecting key soccer match events to create highlights using computer vision," Computer Vision and Pattern Recognition, vol. 1, no. 1, pp. 3-10, 2022. https://doi.org/10.48550/arXiv.2204.02573

[19] X. Fang, F. Hu, M. Yang, T. Zhu, R. Bi, Z. Zhang, and Z. Gao, "Small object detection in remote sensing images based on super-resolution," Pattern Recognition Letters, vol. 153, no. 1, pp. 107-112, 2022. https://doi.org/10.1016/j.patrec.2021.11.027

[20] S. A. Ghasemzadeh, G. Van Zandycke, and M. Istasse, "DeepSportLab: A unified framework for ball detection, player instance segmentation, and pose estimation in team sports scenes," Computer Vision and Pattern Recognition, vol. 16, no. 1, pp. 1-13, 2021. https://doi.org/10.48550/arXiv.2112.00627

[21] R. A. Murugan, and B. Sathyabama, "Object detection for night surveillance using ssan dataset based modified yolo algorithm in wireless communication," Wireless Personal Communications, vol. 128, no. 3, pp. 1813-1826, 2023. https://doi.org/10.1007/s11277-022-10020-9

[22] A. Malini, P. Priyadharshini, and S. Sabeena, "An automatic assessment of road condition from aerial imagery using modified VGG architecture in Faster-RCNN framework," Journal of Intelligent & Fuzzy Systems, vol. 40, no. 6, pp. 11411-11422, 2021. https://doi.org/10.3233/JIFS-202596

[23] D. Yang, "Research on multi-target tracking technology based on machine vision," Applied Nanoscience, vol. 13, pp. 2945-2955, 2023. https://doi.org/10.1007/s13204-021-02293-6

[24] J. Zhang, Y. He, W. Feng, J. Wang, and N. N. Xiong, "Learning background-aware and spatial-temporal regularized correlation filters for visual tracking," Applied Intelligence, vol. 53, no. 7, pp. 7697-7712, 2023. https://doi.org/10.1007/s10489-022-03868-8

[25] K. Aygül, M. Cikan, T. Demirdelen, and M. Tumay,

"Butterfly optimization algorithm based maximum power point tracking of photovoltaic systems under partial shading condition," Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, vol. 45, no. 3, pp. 8337-8355, 2023. https://doi.org/10.1080/15567036.2019.1677818

[26] Y. Liu, and Z. Zhou, "Optical flow-based stereo visual odometry with dynamic object detection," IEEE Transactions on Computational Social Systems, vol. 10, no. 6, pp. 3556-3568, 2023. https://doi.org/10.1109/TCSS.2022.3205015

[27] A. Rodríguez-Fernández, R. Ramirez-Campillo, J. Raya-González, D. Castillo, and F. Y. Nakamura, "Is physical fitness related with in-game physical performance? A case study through local positioning system in professional basketball players," Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology, vol. 237, no. 3, pp. 188-196, 2023. https://doi.org/10.1177/1754337121103116

[28] P. Rahimian, and L. Toka, "Optical tracking in team sports: A survey on player and ball tracking methods in soccer and other team sports," Journal of Quantitative Analysis in Sports, vol. 18, no. 1, pp. 35-57, 2022. https://doi.org/10.1515/jqas-2020-0088

[29] H. Liu, X. Duan, H. Chen, H. Lou, and L. Deng, "DBF-YOLO: UAV small targets detection based on shallow feature fusion," IEEJ Transactions on Electrical and Electronic Engineering, vol. 18, no. 4, pp. 605-612, 2023. https://doi.org/10.1002/tee.23758

[30] M. Hasanvand, M. Nooshyar, E. Moharamkhani, and A. Selyari, "Machine learning methodology for identifying vehicles using image processing," Artificial Intelligence and Applications, vol. 1, no. 3, pp. 170-178, 2023. https://doi.org/10.47852/bonviewAIA3202833

[31] P. E. Martin, J. Benois-Pineau, R. Péteri, and J. Morlier, "Three-stream 3D/1D CNN for fine-grained action classification and segmentation in table tennis," Proceedings of the 4th International Workshop on Multimedia Content Analysis in Sports, vol. 9, no. 14, pp. 35-41, 2021. https://doi.org/10.1145/3475722.3482793

[32] N. Fang, and Z. Zhan, "High-resolution optical flow and frame-recurrent network for video super-resolution and deblurring," Neurocomputing, vol. 489, no. 6, pp. 128-138, 2022. https://doi.org/10.1016/j.neucom.2022.02.067