Badminton Tracking and Motion Evaluation Model Based on Faster RCNN and Improved VGG19

Jun Ou[1], Chao Fu[2]*, Yanyun Cao[3]
Physical Education Institute, Xinyu University, Xinyu 338000, Jiangxi, China[1, 2]
College of Physical Education and Health, Jiangxi Science and Technology Normal University, Nanchang 330000, China[3]

***Abstract*—Badminton, as a popular sport in the field of sports, has rich information on body motions and motion trajectories. Accurately identifying the swinging motions during badminton is of great significance for badminton education, promotion, and competition. Therefore, based on the framework of Faster R-CNN multi object tracking algorithm, a new badminton tracking and motion evaluation model is proposed by introducing a VGG19 network architecture and real-time multi person pose estimation algorithm for performance optimization. The experimental results showed that the new badminton tracking and motion evaluation model achieved an average processing speed of 31.02 frames per second for five bone points in the human head, shoulder, elbow, wrist, and neck. Its accuracy in detecting the highest percentage of correct key points for the head, shoulders, elbows, wrists, and neck reached 98.05%, 98.10%, 97.89%, 97.55%, and 98.26%, respectively. The minimum values of mean square error and mean absolute error were only 0.021 and 0.026. The highest resource consumption rate was only 6.85%, and the highest accuracy of motion evaluation was 97.71%. In addition, indoor and outdoor environments had almost no impact on the performance of the model. In summary, the study aims to improve the fast region convolutional neural network and apply it to badminton tracking and motion evaluation with higher effectiveness and recognition accuracy. This study aims to demonstrate a more effective approach for the development of badminton sports.**

Keywords—Faster RCNN; VGG19; badminton; target tracking; motion evaluation

I. Introduction

With the popularity and popularization of badminton in international sports events, its training methods have gradually become diversified [1]. In order to better track targets in badminton sports scenes, prevent injuries caused by non-standard technical motions, and promote more standardized training, it is necessary to conduct in-depth discussions on badminton tracking and motion evaluation methods. Currently, common object tracking algorithms include Visual Object Tracking (VOT), Multiple Object Tracking (MOT), and Multi-Camera Multi Object Tracking (MCMOT) [2]. The Fast Region Convolutional Neural Networks (Faster RCNN) multi object tracking algorithm based on deep learning is currently the mainstream object tracking method in the field of motion detection [3]. Numerous researchers both domestically and internationally have explored the Faster RCNN multi object tracking algorithm. For instance, J. Meza et al. developed a Faster RCNN method on the basis of transfer learning to significantly improve public transportation and achieve real-time localization in highly occluded scenes. The method could effectively shorten travel time, improve road smoothness, thereby controlling the fleet and reducing congestion [4]. T. Shimizu et al. analyzed the difficulty of target tracking in open surgeries such as plastic surgery. A method for analyzing and evaluating open surgical videos was created by combining Faster RCNN localization, Residual Network 18 (ResNet-18), and Long Short Term Memory (LSTM) modules. The experimental results showed that this method successfully detected two different open surgeries. It was superior to the commonly used two baseline methods [5]. H. Li et al. aimed to optimize the management efficiency of urban sports public services, facilitate residents to exercise, and increase their happiness index. A smart target tracking model was proposed using the Faster RCNN algorithm. The experimental results showed that Faster RCNN had good accuracy and low average time. This model could guide different populations to fully utilize public service facilities, improve quality of life, and achieve good behavior in national sports [6]. X. Yin et al. designed an image object detection method on the basis of Faster RCNN to address the incomplete image feature extraction and low classification accuracy in existing image object detection algorithms. The experimental results showed that the average accuracy was 91.04%, which had good image target detection ability [7].

Although the Faster RCNN performs well in object tracking and detection, badminton is a complex sports scene that is prone to occlusion and light interference during the motion process [8]. Therefore, to improve the accuracy of badminton target tracking and detection, and reduce the loss, it is necessary to deepen the network hierarchy of the Faster RCNN. Visual Geometry Group19 (VGG19) is an architecture in deep neural networks. It adds more convolutional layers and parameters than other architectures, which can not only better extract image features but also better process more complex image data. It has been used in various visual detection fields [9, 10]. X. Wan et al. found that traditional machine vision algorithms couldn’t successfully detect defects in various steel strips. Therefore, on the basis of fast image preprocessing algorithms and transfer learning theory, a complete improved VGG19 neural network strip defect detection process was proposed. The improved VGG19 had a recognition accuracy of 97.8%. Its performance in six types of defects outperformed the baseline VGG19 [11]. R. Mohan et al. proposed a VGG19 for diagnosing various lung diseases from chest CT images on the ground of customized medical image analysis and detection networks. The experimental results showed that in multi class classification tasks, the training accuracy and testing accuracy of VGG19 performed excellently [12]. A. Faghihi et al. analyzed the skin lesion classification using Convolutional Neural Network (CNN) technology. A pre-trained neural network application transfer learning framework was constructed using VGG19. Compared with other methods, the classification accuracy of the method reached 98.18% [13]. To develop the non-invasive diagnostic method for Obstructive Sleep Apnea Hypopnea Syndrome (OSAHS) patients, L. Ding et al. proposed a pre-trained VGG19 and LSTM fusion model to classify the snoring sounds of simple snorers and OSAHS patients. The experimental results showed that the VGG19+LSTM had the highest classification accuracy of 99.31% for simple snorers snoring and OSAHS patients snoring [14].

In summary, current target tracking and detection technologies both domestically and internationally still face many challenges in dealing with occlusions and similar object interference in complex dynamic environments. Although various studies have attempted to enhance multi-object detection in images by integrating deep learning network models with the VGG19 architecture, as well as using new algorithms such as CNN and Faster RCNN, there is still a significant gap between the current detection performance and the expectation in practical applications. These gaps are mainly reflected in insufficient robustness, making it difficult to stably track targets in environments with high occlusion or similar object interference; real-time performance has not yet met the requirements of some application scenarios, especially in sports motion analysis that requires rapid response; limited generalization ability, with poor adaptability to data under different environments and conditions; and high consumption of computing resources, which restricts the application on devices with limited resources. Therefore, an innovative badminton tracking and motion evaluation model based on Faster RCNN and improved VGG19 is proposed in the study. By combining the powerful object detection and tracking capabilities of Faster RCNN with the VGG19 feature extraction, it can solve the existing challenges in badminton tracking and motion assessment technologies and further improve the accuracy of badminton tracking and motion assessment, thus providing a more efficient and accurate motion assessment solution for the field of badminton sports. This study is divided into five sections, first being the introduction. The second section introduces how the Faster RCNN target tracking algorithm is improved and how the optimized design model is established. The third section is performance testing of the new model. The fourth section is the discussion of the results. The last section is a summary of the paper.

II. Methods and Materials

In response to the existing problems in badminton tracking and motion evaluation, such as the challenge of dealing with severe occlusion and similar appearance interference in complex sports environments, this study first introduces the basic framework of Faster RCNN algorithm from the perspective of badminton target tracking. The VGG19 architecture is introduced and significantly improved. In addition, from the perspective of evaluating the motions of badminton players, the Faster RCNN-VGG19 target tracking algorithm is used as the framework foundation, taking the real-time multi person pose estimation algorithm (OpenPose) for further optimization. Through these improvements, a new comprehensive badminton tracking and motion evaluation method is ultimately proposed, aiming to improve the accuracy and real-time performance of badminton tracking and motion evaluation.

A. *Construction of 3D Object Tracking Model Based on Faster RCNN and VGG19*

In order to enable athletes to master the basic motions of badminton in a standardized manner and achieve precise and real-time motion evaluation, it is necessary to quickly and accurately detect and track moving targets. The Faster RCNN is a target detection algorithm in the RCNN series, which has strong target recognition capabilities [15]. It mainly contains two parts, namely the Region Proposal Network (RPN) and the target classification network based on target feature classification [16]. The RPN network and the object classification network share weight parameters, and the two networks are trained collaboratively, which can promote the network to have good robustness and accuracy. The Faster RCNN is displayed in Fig. 1.

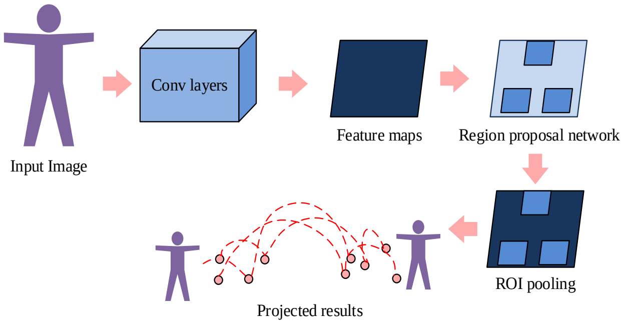


Fig. 1. Structure of the faster RCNN algorithm.

In Fig. 1, the Faster RCNN is mainly divided into three stages. Firstly, the motion image data is input into the network to obtain the corresponding feature image data. Secondly, the RPN is used to produce candidate boxes, mapping the candidate boxes generated by the RPN structure to the feature image data to obtain the relevant feature matrix. The obtained feature matrix is scaled to a size of 7×7 through the Region of Interest Pooling (ROI pooling) layer. Then, the 7×7 feature map is flattened and the final prediction result is obtained through fully connected layers. However, due to the limited appearance features of the small-sized shuttlecock, it is difficult to effectively distinguish the shuttlecock from similar small targets such as sneakers, light spots, and spectators' heads using the limited appearance features [17]. In addition, the traditional Faster RCNN uses the VGG16 framework, which cannot deepen the network hierarchy on the existing basis [18]. Therefore, in order to solve such problems, the study modifies the VGG16 framework in the Faster RCNN to the VGG19 framework. A new algorithm, namely the Faster RCNN-VGG19 object detection algorithm is proposed. The basic structure of the VGG19 framework is shown in Fig. 2.

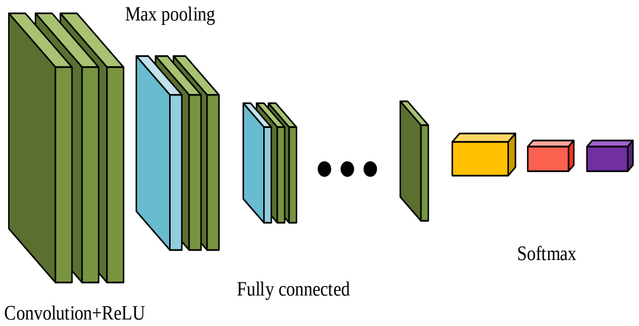


Fig. 2. Basic structure of the VGG19 framework.

In Fig. 2, the VGG19 framework mainly contains six parts: convolutional layer, batch normalization layer, ReLU activation function, maximum pooling layer, fully connected layer, and classifier [19]. The total depth of the network is 19 layers, including 16 CNN and 3 fully connected layers. The 16 CNN is further divided into 5 convolutional layers with varying numbers. The special structure of VGG19 can preserve all the features of the input image as much as possible, ensuring that the resolution of each layer's output and input is equal [20]. Batch normalization $y^{(q)}$ is shown in Eq. (1).

$$y^{(q)} = \gamma^{(q)} \frac{x^{(q)} - \mu^{(q)}}{\sqrt{(\sigma^{(q)})^2 + \varepsilon}} + \beta^{(q)} \tag{1}$$

In Eq. (1), $\gamma^{(q)}$ and $\beta^{(q)}$ represent learnable parameters. $x^{(q)}$ and $\mu^{(q)}$ represent the q-th dimensional input data and mean of the data, respectively. $\sigma^{(q)}$ represents the standard deviation. ε represents a number that prevents the denominator from being 0. Batch normalization can reduce the gradient vanishing and exploding, and accelerate the convergence speed of neural networks [21]. The ReLU is displayed in Eq. (2).

$$\mathrm{Re}\,LU(x) = \begin{cases} \max(x,0), x \geq 0 \\ 0, x < 0 \end{cases} \tag{2}$$

In Eq. (2), x signifies the input data. When x is greater than or equal to 0, $\max(x,0)$ is output. When x is less than 0, the output is 0. The ReLU (Rectified Linear Unit) activation function can perform a nonlinear transformation on the output of a neural network, thereby increasing the network's expressive and fitting capabilities. However, when the ReLU activation function encounters a situation where parameters need to be corrected during backpropagation, if the input is negative, the gradient becomes 0, which leads to an inability to adjust the parameters, resulting in the so-called "Dead ReLU" problem [22]. Therefore, the study introduces Leaky ReLU to address the DeadReLU of the ReLU in VGG19. The Leaky ReLU is shown in Eq. (3).

$$Leaky\,\mathrm{Re}\,LU(x) = \begin{cases} ax, x \geq 0 \\ x, x < 0 \end{cases} \tag{3}$$

In Eq. (3), a signifies the specified parameter, usually taking the smaller value. Leaky ReLU has a gradient even when the input is less than 0, and it possesses linear and non-saturating properties, which allows for fast convergence. It does not require exponential computations, making it computationally efficient and capable of addressing the issue of un-updatable weights in the standard ReLU activation function. The eigenvalue weight w_i is shown in Eq. (4).

$$w_i = \frac{e^{b_i}}{\sum_{j \in R} e^{b_j}} \tag{4}$$

In Eq. (4), b and R represent the feature map and local region, respectively. The eigenvalue weights ensure the transmission of important features, and during backpropagation, the features within the region will have a preset minimum gradient. Although the VGG19 framework can address the difficulties in tracking and recognizing small targets in the Faster-RCNN neural network algorithm, due to the effects of occlusion and lighting changes, there are still inevitable false positives and missed detections in the badminton detection results [23, 24]. Therefore, the study utilized the commonly employed triangulation algorithm from stereo vision matching to fuse the two-dimensional coordinates of the ball from multiple camera perspectives into three-dimensional coordinates, proposing a 3D target tracking model based on Faster RCNN and VGG19. The target tracking process framework of this model is shown in Fig. 3.

From Fig. 3, the target tracking process of the 3D target tracking model based on Faster RCNN and VGG19 is mainly divided into four stages: badminton 2D detection stage, badminton 2D tracking stage, badminton 3D coordinate fusion stage, and badminton 3D trajectory smoothing stage. The study first conducts badminton 2D detection in various camera perspectives based on the Faster RCNN-VGG19 object detection algorithm. Then, a 2D tracking algorithm based on Efficient Convolution Operators (ECO) is used to track the badminton balls in various camera perspectives. Secondly, the triangulation algorithm is used to effectively merge multiple 2D coordinates into one 3D coordinate. Finally, the Kalman filtering method is used to process and obtain smooth 3D badminton trajectories. The 3D coordinate point of badminton is shown in Eq. (5).

$$p_t = \frac{1}{N} \sum_{1 \leq i, j \leq n} (p_t^{ij} \left| e_t^{ij} < \tau \right.) \tag{5}$$

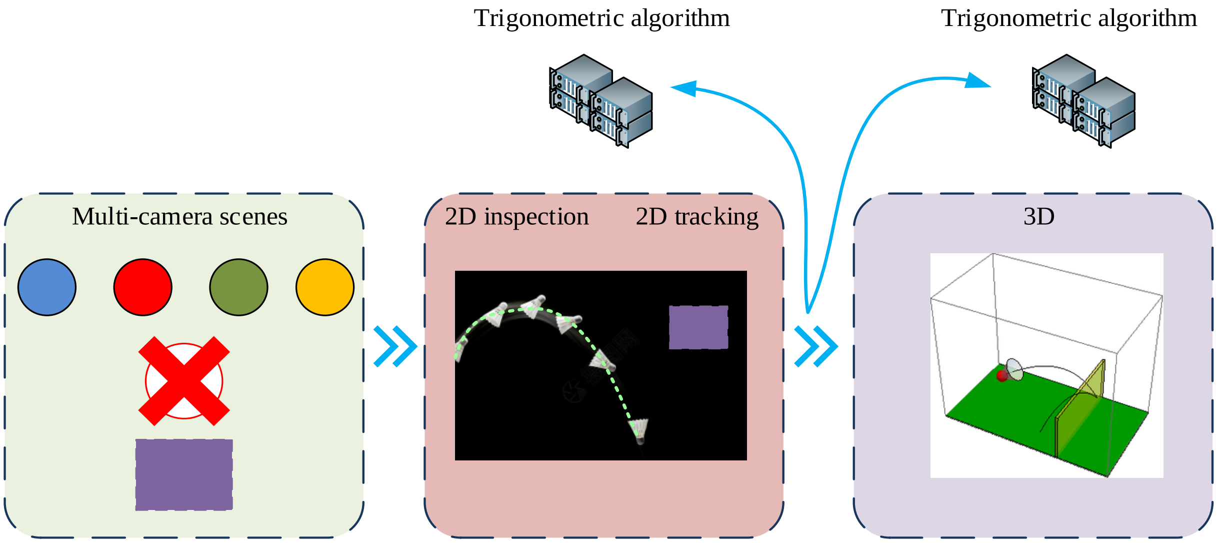


Fig. 3. Target tracking process framework for the proposed model.

In Eq. (5), p_t represents the three-dimensional coordinate fusion result of the badminton at time t. N signifies the number of camera matching pairs where the back projection error is less than the threshold τ. p_t^{ij} and e_t^{ij} represent the 3D coordinates and back projection errors calculated by the triangulation algorithm, respectively. The expression for maximizing probability $\hat{m}$ is shown in Eq. (6).

$$\hat{m} = \arg\max p(X_i^t = x_i^t, Y_i^t = Y_{j\in N(i)}^{t+1} \left| I^t \right.) \tag{6}$$

In Eq. (6), X_i^t and x_i^t refer to whether there is badminton players or not, taking 0 or 1. Y_i^t and $Y_{j\in N(i)}^{t+1}$ represent the appearance characteristics of badminton players. The common evaluation metric for multi-object tracking algorithms is the Multiple Object Tracking Accuracy (MOTA). The study primarily takes into account three types of tracking errors — missed detections, false positives, and identity switches—for subsequent performance assessment. MOTA is shown in Eq. (7).

$$MOTA = 1 - \frac{\sum_t (c_1 \cdot fn_t + c_2 \cdot fp_t + c_3 \cdot idsw_t)}{\sum_t g_t} \tag{7}$$

In Eq. (7), c_1, c_2, and c_3 represent constants. g_t represents the true value. fn_t, fp_t and $idsw_t$ represent the number of missed detections, false detections, and number of identity exchanges, respectively.

B. Construction of Motion Evaluation Model Based on Faster RCNN and VGG19

After constructing a 3D object tracking model on the ground of Faster RCNN and VGG19, this study aims to address the various drawbacks of the badminton training system and attempt to optimize the model from the perspective of badminton motion evaluation. The first step in evaluating the motions of badminton players is to effectively obtain their posture information. Traditional methods for obtaining pose information often have drawbacks such as weak real-time performance, complex operation, and poor pose estimation performance [25]. The OpenPose can estimate the posture of the human body by analyzing key points in images or videos, identifying various parts of the body, and inferring the posture information. It can maintain accuracy even in complex environments [26, 27]. Therefore, OpenPose is introduced into the Faster RCNN-VGG19 object detection algorithm to analyze and process the posture information of badminton players. The basic structure of the Faster RCNN-VGG19-OpenPose pose estimation algorithm is shown in Fig. 4.

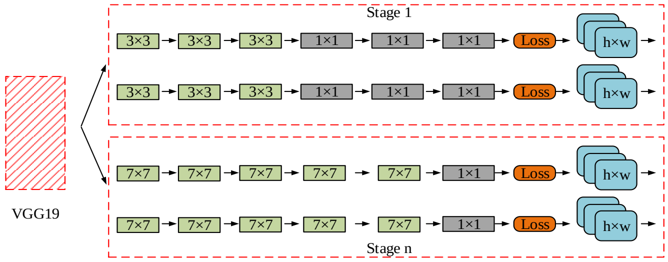


Fig. 4. The basic structure of OpenPose.

As shown in Fig. 4, the Faster RCNN-VGG19-OpenPose estimation algorithm is mainly divided into two parts, namely the limb confidence part and the site affinity vector field part. Firstly, the image is input into a parallel two branch Faster RCNN structure, and feature extraction is performed using the VGG19 architecture. Secondly, the feature maps of the image are obtained by clustering the bone points based on initialization operation and greedy algorithm. Finally, the limb confidence prediction data and site affinity vector field prediction data are output through the limb confidence section and site affinity vector field. The predicted limb confidence data S^t and the predicted site affinity vector field data L^t are displayed in Eq. (8).

$$\begin{cases} S^t = \rho^t(F) \\ L^t = \varphi^t(F) \end{cases} \tag{8}$$

In Eq. (8), F represents the feature mapping. ρ^t and φ^t represent the CNN used for inference. In each subsequent stage, the original features and the two branch predictions generated in the previous stage are jointly input into the next stage, which can fully utilize the original features of the image to accurately predict the image in each stage. In addition, in order to successfully identify the categories of badminton swing motions, the research also cascades a fast Support Vector Machine (SVM) classifier based on the Decision Tree (DT) at the backend of OpenPose. The DT-SVM classifier can decompose nonlinear optimization problems into multiple linear SVM problems for solution, making the method simple and easy to implement. The DT structure diagram for badminton swing is shown in Fig. 5.

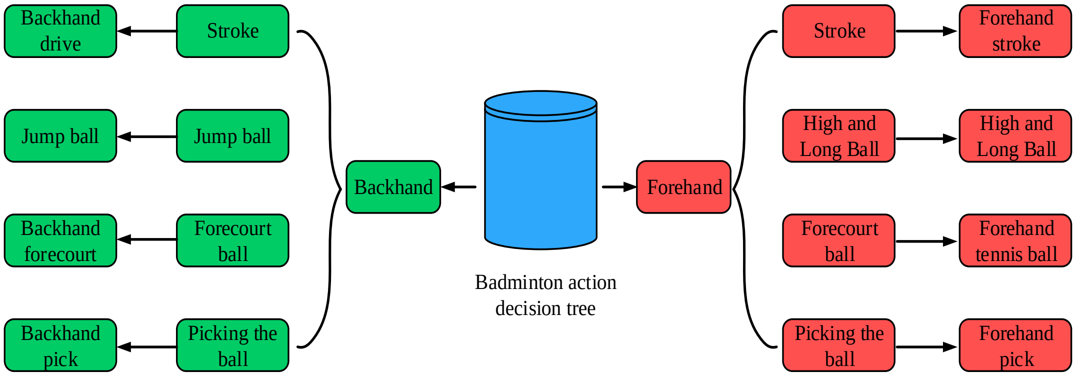


Fig. 5. Badminton swing decision tree.

From Fig. 5, the DT divides badminton swing motions into eight categories, namely high and long ball, Forehand stroke, forehand pick, Forehand tennis ball, jump ball, backhand drive, backhand pick, and backhand forecourt [28]. Firstly, the eight types of swing motions are combined to form a DT. During the classification process, each category needs to be selected before entering the next level, and unselected subtrees are deleted. This can effectively reduce the samples that need to be classified in the next step. The search step is repeated until the leaf node is reached. The final output result is obtained. However, the number of players in badminton training is generally large, and it is necessary to associate the results of the motion assessment with the target personnel; otherwise, the obtained sports assessment results will become meaningless [29]. Considering such situations, the study utilized the Particle Filter (PF) algorithm to deal with the non-Gaussianity of probability distributions in the tracking environment. The PF is shown in Eq. (9).

$$\begin{cases} x[t] = Ax[t-1] + v \\ y[t] = Cx[t] \end{cases} \tag{9}$$

In Eq. (9), $x[t]$ and $y[t]$ signify the coordinate values of the human skeletal neck. A represents the state transition matrix. v and C represent position random shift variables and diagonal matrices, respectively. The likelihood p^m of particles is shown in Eq. (10).

$$p^m = \frac{1}{\sqrt{2\pi\alpha^2}} \exp\left(-\frac{(d^m)^2}{2\alpha^2}\right) \tag{10}$$

In Eq. (10), d^m and α represent the degree difference and adjustable parameters, respectively. There are generally three annotation methods for human joint points: target instance annotation, target key point annotation, and image understanding annotation [30]. The key points of the human body during badminton motion are annotated based on target key point annotation. The annotation of key points in the human body is displayed in Fig. 6.

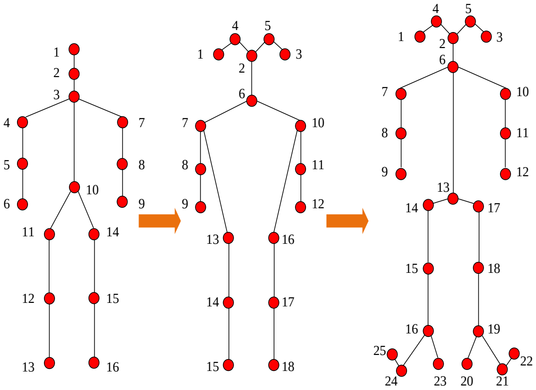


Fig. 6. Human body part key point labeling.

According to Fig. 6, there are three types of annotations for key points in the human body during badminton, namely 16 key points, 18 key points, and 25 key points. Among them, 18 key points increase facial key points compared with 16 key points, and 25 key points increase foot key points compared with 18 key points. The description of these three types of joint points can represent the skeletal information of the human body in detail, but these three methods will bring considerable computational complexity. Therefore, the study adjusts it by removing useless joints such as eyes and ears. The number of described joints is adjusted to 14. The distance $dis(x_i, x_j)$ between two joint points is shown in Eq. (11).

$$dis(x_i, x_j) = \sqrt{(x_i - x_j)^T M (x_i - x_j)} \tag{11}$$

In Eq. (11), x_i and x_j represent the positions of two joint points. M represents a symmetric positive semi-definite matrix. The rough pose representation metric is shown in Eq. (12).

$$E_{coarse}(W, B^*) = \left\| (x_n - x_n B^*)^T W_n \right\|_F^2 \tag{12}$$

In Eq. (12), B^* represents the coefficient matrix. x_n represents rough posture data, W then it represents the rank matrix of the entire column. Based on the above improvements, a motion evaluation model based on Faster RCNN and VGG19 is ultimately proposed. The motion evaluation process of this model is shown in Fig. 7.

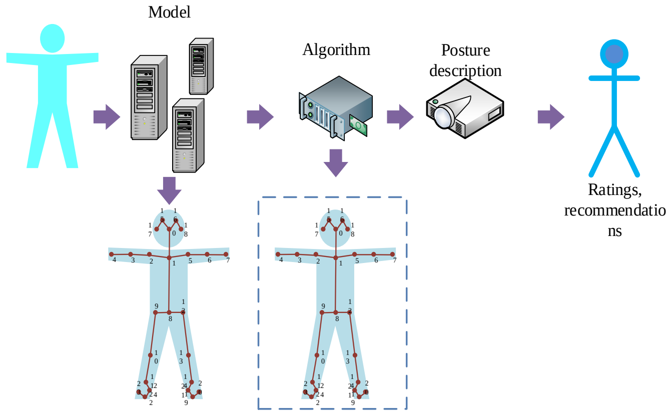


Fig. 7. The motion evaluation process of the proposed model.

From Fig. 7, the motion evaluation process mainly consists of four steps, namely the 3D target tracking model based on Faster RCNN and VGG19, posture estimation, granular posture description and scoring, and proposing motion suggestions. Firstly, a three-dimensional target tracking model based on Faster RCNN and VGG19 is used to quickly and accurately obtain the three-dimensional coordinates of badminton. Then, the position information of all human bones in the image is determined, and the prior standard parameters for each motion are obtained. Next, the posture image is described. Finally, the motion standard level is evaluated and motion suggestions are proposed.

III. Results

A. Faster RCNN-VGG19-OpenPose Algorithm Performance Testing

To verify the performance of the improved Faster RCNN-VGG19-OpenPose algorithm, a suitable experimental environment is established. The CPU is set to Intel Core i7 with a base frequency of 4.2Hz. The GPU is set to NVIDIA GeForce RTX 1660s, with 16GB of graphics memory and 16GB of memory. The Windows 10 is the operating system, Python as the algorithmic language. MPII and COCO datasets are used as the test data sources. The MPII dataset is a database of human body postures, containing approximately 25000 images and 40000 human body node information from different postures, and covering over 410 activities, all of which are sourced from YouTube videos. The COCO dataset is a dataset provided by the Microsoft team that can be used for image recognition, including various motion scenes. The study divided these datasets into training and testing sets in a ratio of 6:4, set the initial learning rate to 0.001, processed 100 frames per batch, and set the weight coefficient to 1. The study first conducts ablation tests on the Faster RCNN-VGG19-OpenPose algorithm using detection accuracy as an indicator. The test results are shown in Fig. 8.

Fig. 8(a) shows the ablation test results of each module of the Faster RCNN-VGG19-OpenPose algorithm on the training set. Fig. 8(b) shows the ablation test results of each module of the Faster RCNN-VGG19-OpenPose algorithm on the testing set. As shown in Fig. 8, with the continuous increase of iterations, the detection accuracy of each module showed an upward trend. The optimal detection accuracy of the basic Faster RCNN was 73.29%. After introducing the VGG19 architecture for optimization, the overall performance of the module improved by about 11%. The Faster RCNN-VGG19-OpenPose algorithm proposed in the study had the best performance, with the highest detection accuracy of 94.69% in the training set and a minimum of 298 iterations. The highest detection accuracy in the testing set was 93.08%, with a minimum of 243 iterations. From this, each module of the proposed algorithm has a positive effect on the overall performance, and the effect is significant. In addition, the study introduces popular target tracking algorithms for comparison, such as VOT algorithm, MOT algorithm, and MCMOT algorithm. A comparison test is conducted using tracking accuracy as an indicator, with a threshold of 200cm, as displayed in Fig. 9.

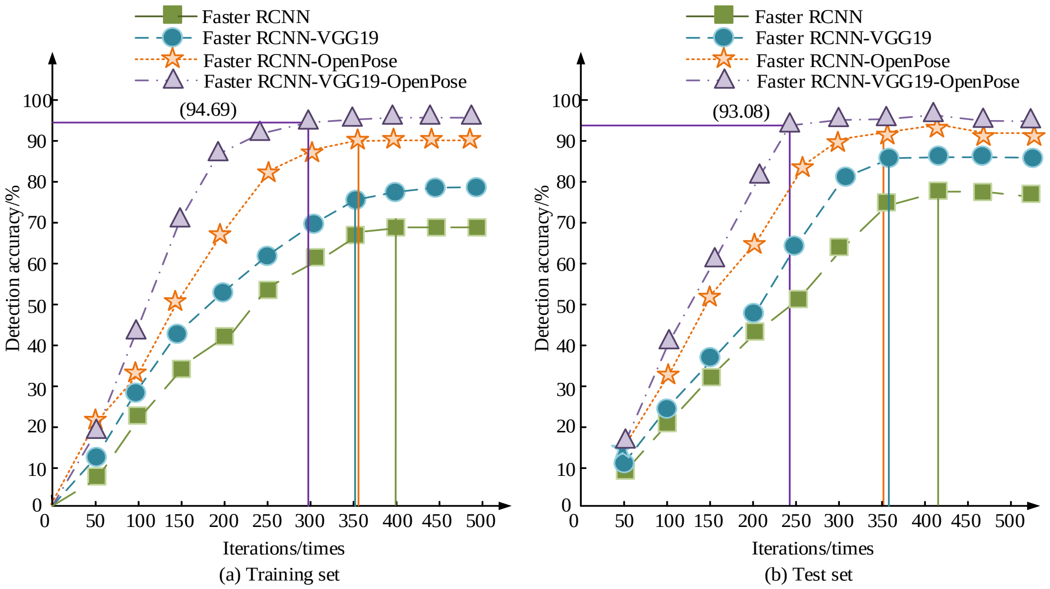


Fig. 8. Ablation test results of target tracking module with different datasets.

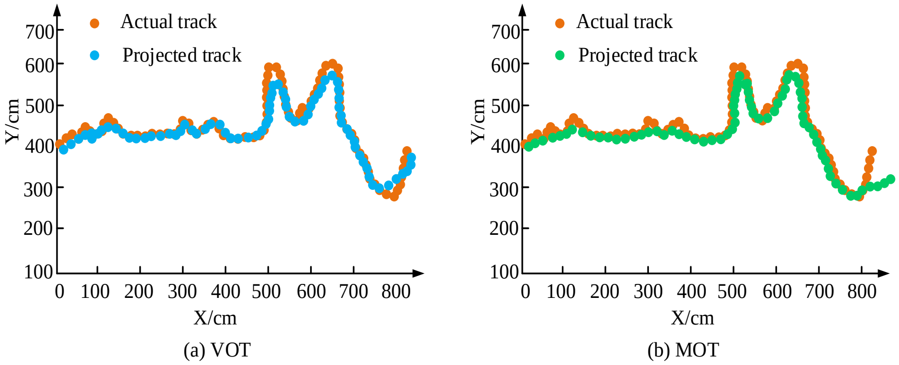
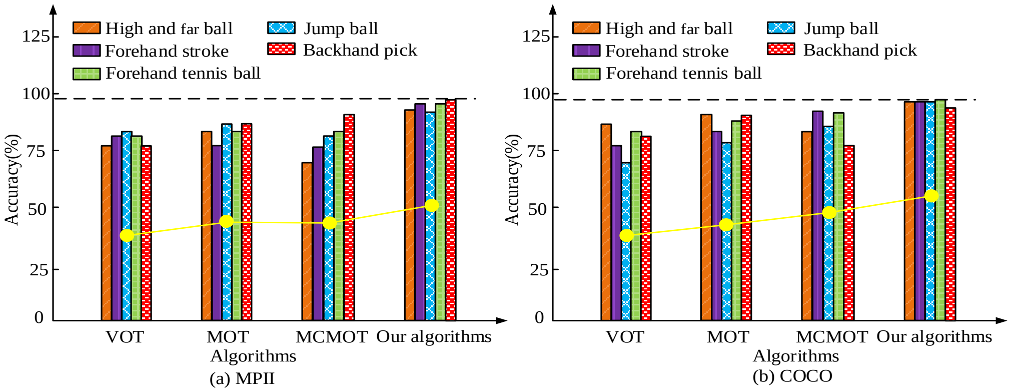


Fig. 9. Tracking accuracy of different target tracking algorithms.

Fig. 9(a) presents the comparison results of the VOT algorithm. Fig. 9(b) presents the MOT algorithm. Fig. 9(c) presents the MCMOT algorithm. Fig. 9(d) presents the Faster RCNN-VGG19-OpenPose. From Fig. 9, the Faster RCNN-VGG19-OpenPose algorithm had the highest overlap between the real trajectory and the tracked trajectory, followed by the MCMOT algorithm, and the worst overlap between the MOT algorithm and the VOT algorithm. This indicates that the proposed algorithm has strong robustness against common missed and false detections in detection. The VGG19 architecture deepens the hierarchy of Faster RCNN, alleviates tracking drift, and achieves stable tracking of badminton. The main differences between the models lie in their approaches to handling tracking problems and their adaptability to complex environments. The Faster RCNN-VGG19-OpenPose algorithm combines deep learning networks with human pose estimation technology, allowing it to more accurately capture the movement trajectory of a badminton shuttlecock, especially maintaining high tracking accuracy in situations where the shuttlecock is moving fast or there is occlusion. In contrast, the VOT algorithm is primarily designed for tracking a single target and lacks the capability to track multiple targets, resulting in poor performance when dealing with multiple shuttlecocks or complex scenes. The MOT algorithm faces challenges in dealing with occlusions and similarities between targets, leading to a lower overlap between tracking and actual trajectories. Although the MCMOT algorithm has been improved for multi-camera environments, its performance is still not as good as the Faster RCNN-VGG19-OpenPose algorithm proposed in this study when dealing with fast-moving and complex backgrounds. To ensure the classification accuracy of the Faster RCNN-VGG19-OpenPose algorithm, the study also tests the classification accuracy of the Faster RCNN-VGG19-OpenPose algorithm for different badminton swing methods. The classification accuracy curve is shown in Fig. 10.

Fig. 10(a) presents the classification performance in the MPII. Fig. 10(b) shows the classification performance in the COCO dataset. From Fig. 10, the Faster RCNN-VGG19-OpenPose algorithm had the best classification performance in the two datasets. The best classification accuracy for badminton high and far balls, forehand strokes, forehand tennis balls, jump balls, and backhand picks reached 98.26%, 98.33%, 98.35%, 98.21%, and 97.08%, respectively, all exceeding 95%. Compared with the VOT algorithm, the Faster RCNN-VGG19-OpenPose algorithm improved the classification accuracy of high and far balls, forehand strokes, forehand tennis ball, jump balls, and backhand picks by about 18% to 25%. The above data indicates that the characteristics of the DT-SVM classifier in solving multi linear problems enable the Faster RCNN-VGG19-OpenPose algorithm to accurately classify badminton swing motions at each stage. Compared with other target tracking algorithms of the same type, it has more stable and superior recognition ability. The study also conducts multi-indicator tests on the above four algorithms using precision, recall, F1 value, and average detection time as indicators, as displayed in Table I.



(a) MPII

(b) COCO

Fig. 10. Effectiveness of different algorithms in classifying badminton swing styles.

TABLE I. METRICS TEST RESULTS FOR VARIOUS ALGORITHMS

Data set	Algorithm	P/%	R/%	F1/%	Average detection time/s
MPII	VOT	88.86	87.38	88.57	5.18
	MOT	89.98	89.73	89.40	6.02
	MCMOT	91.87	88.64	90.26	5.38
	Faster RCNN-VGG19-OpenPose	95.41	94.28	95.85	2.74
COCO	VOT	89.28	88.16	88.37	4.61
	MOT	88.52	88.85	88.69	3.75
	MCMOT	91.43	90.79	91.11	3.08
	Faster RCNN-VGG19-OpenPose	95.56	93.68	95.32	2.51

According to Table I, among the four types of indicators detected, the VOT target tracking algorithm performed the worst, followed by the MOT target tracking algorithm, MCMOT target tracking algorithm, and the proposed Faster RCNN-VGG19-OpenPose algorithm. The highest P-value of the VOT target tracking algorithm was 89.28%, the highest R-value was 88.16%, the highest F1 was 88.37%, and the average detection time was 4.61s. The new Faster RCNN-VGG19-OpenPose target tracking algorithm proposed in the study had a maximum P-value of 95.56%, a maximum R-value of 93.68%, a maximum F1 value of 95.32%, and an average detection time of 2.51s. From this, the Faster RCNN-VGG19-OpenPose algorithm has relatively good performance, which is more suitable for the badminton tracking work at current stage.

B. Simulation Testing of Tracking and Motion Evaluation Model Based on Faster RCNN and VGG19

From the above test results, the Faster RCNN-VGG19-OpenPose algorithm performed excellently in badminton tracking and classification. However, this data is only feasible for the MPII and COCO datasets, and the persuasiveness and feasibility of the data results still need to be further strengthened. The performance of the tracking and motion evaluation model based on Faster RCNN and VGG19 has not been verified yet. Therefore, the study attempts to use a self-made dataset for testing, which includes eight swing styles of badminton high and far balls, forehand strokes, forehand picks, forehand tennis ball, jump balls, backhand drives, backhand picks, and backhand forecourts. The number of videos for each badminton swing motion is 200, totaling 1400 video sequences. Each video sequence has a frame rate of 25fps, a resolution of 160x120, and an average length of 30s. At this point, the fuzzy comprehensive evaluation method is used to assign weights to the above motions. In addition, the study tracks the usage of GPU and CPU to record the time and storage space required for the model to process data. Tools such as Nsight are used to analyze resource utilization, adjust model parameters and structure to optimize efficiency, and test the long-term stability of the model in actual deployment to determine the resource consumption of each model. Table II displays the weighting results.

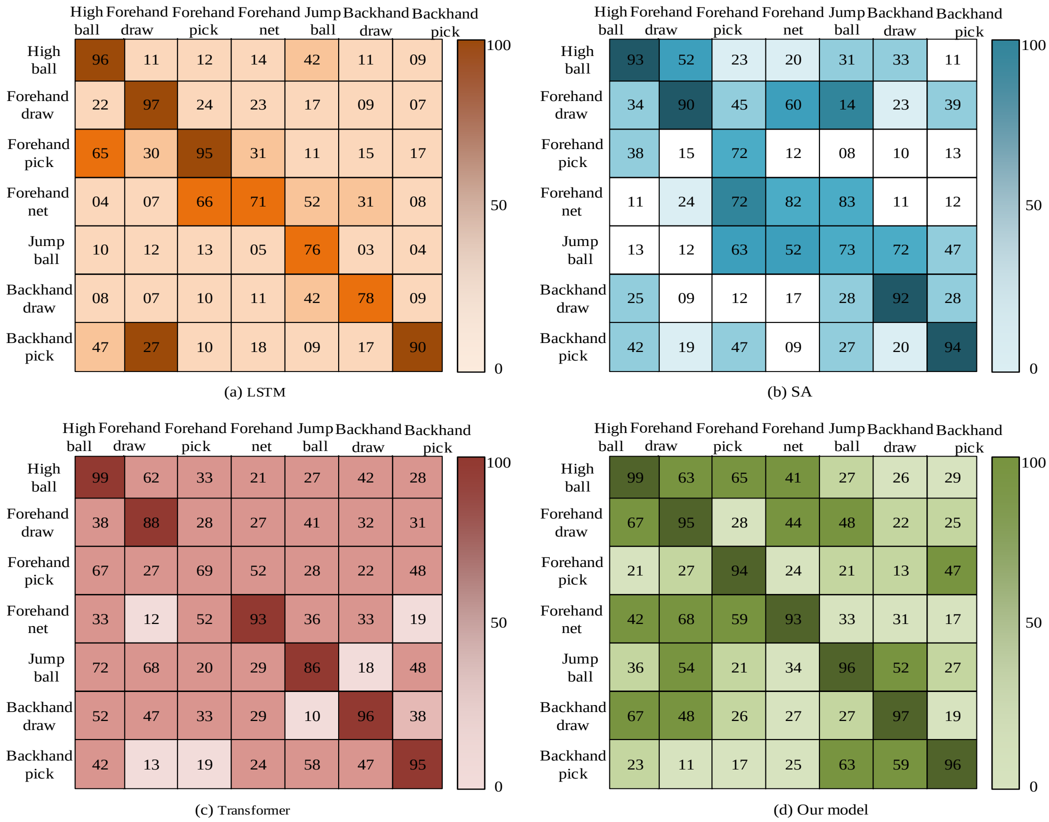
From Table II, the joint correlation under each motion, i.e. the weight score, was relatively reasonable and did not differ significantly from the actual physical sensation during motion. For example, in jump ball motions, there was a significant correlation between the weights of the legs, body, and knees. In forehand stroke, there is a significant correlation between the elbow, wrist, upper arm, and shoulder joints. The above motions are separately detected using confusion matrices. The same type popular motion recognition models are compared, including the LSTM model, Spatial Attention (SA), and Transformer network, as displayed in Fig. 11.

Fig. 11(a)-(d) show the confusion matrix results of LSTM, SA, Transformer, and the proposed model. From Fig. 11, the confusion matrix results for 7 different badminton swing motions under the LSTM model were poor, with only 4 groups scoring above 90 points. The same applies to the SA. There was a significant improvement in the confusion results of the Transformer, with 6 groups successfully paired for over 80 points. The confusion results of the model showed the best performance, with all 7 sets of badminton swing motions completed matching and scores above 90 points. Therefore, the designed method has certain effectiveness and performs better in similar models. In addition, to reflect the accuracy of the model in locating human skeletal points, the study also tests the model using the Percentage of Correct Key Points (PCK) and Frames Per Second (FPS) as indicators. The test results are shown in Fig. 12.

Fig. 12(a) shows the PCK performance of human bone point localization. Fig. 12(b) shows the FPS performance. As shown in Fig. 12, the proposed tracking and motion evaluation model based on Faster RCNN and VGG19 achieved an average FPS processing speed of 31.02 frames per second for five skeletal points in the human head, shoulder, elbow, wrist, and neck. Its highest PCK detection accuracy for the head, shoulder, elbow, wrist, and neck reached 98.05%, 98.10%, 97.89%, 97.55%, and 98.26%, respectively. The above data indicates that the new target tracking and motion evaluation model has unique advantages in recognizing human head, shoulder, elbow, wrist, and neck joint points. Finally, to explore the impact of the environment on the proposed model, the study also conducted comparative tests on the state-of-the-art motion assessment models under indoor and outdoor environments, using Mean Squared Error (MSE), Mean Absolute Error (MAE), resource consumption rate, and motion assessment accuracy as reference indicators. The models tested include detection methods based on the Deformable Parts Model (DFM), Tree-based Human Pose Estimation (TB-HPE), and Dual Source Deep Neural Network (DS-DNN) for human pose estimation. The test results are shown in Table III.

TABLE II. WEIGHTING VALUES FOR DIFFERENT MOTIONS AND JOINTS

Classification	High ball	Forehand stroke	Forehand pick	Forehand tennis ball	Jump ball	Backhand drive	Backhand pick	Backhand forecourt
Wrist	0.082	0.082	0.073	0.074	0.051	0.086	0.061	0.082
Elbow	0.083	0.074	0.085	0.072	0.053	0.087	0.078	0.079
Knee	0.035	0.042	0.059	0.052	0.067	0.044	0.047	0.048
Ankle	0.038	0.041	0.033	0.044	0.073	0.042	0.033	0.024
Hips and thighs	0.042	0.034	0.059	0.041	0.071	0.007	0.037	0.038
Crotch and body	0.043	0.043	0.048	0.036	0.082	0.024	0.024	0.035
Big arms and shoulders	0.074	0.067	0.054	0.051	0.011	0.055	0.038	0.031
Big arms and body	0.089	0.087	0.062	0.056	0.076	0.043	0.035	0.036



	High ball	Forehand draw	Forehand pick	Forehand net	Jump ball	Backhand draw	Backhand pick
High ball	96	11	12	14	42	11	09
Forehand draw	22	97	24	23	17	09	07
Forehand pick	65	30	95	31	11	15	17
Forehand net	04	07	66	71	52	31	08
Jump ball	10	12	13	05	76	03	04
Backhand draw	08	07	10	11	42	78	09
Backhand pick	47	27	10	18	09	17	90

(a) LSTM

	High ball	Forehand draw	Forehand pick	Forehand net	Jump ball	Backhand draw	Backhand pick
High ball	93	52	23	20	31	33	11
Forehand draw	34	90	45	60	14	23	39
Forehand pick	38	15	72	12	08	10	13
Forehand net	11	24	72	82	83	11	12
Jump ball	13	12	63	52	73	72	47
Backhand draw	25	09	12	17	28	92	28
Backhand pick	42	19	47	09	27	20	94

(b) SA

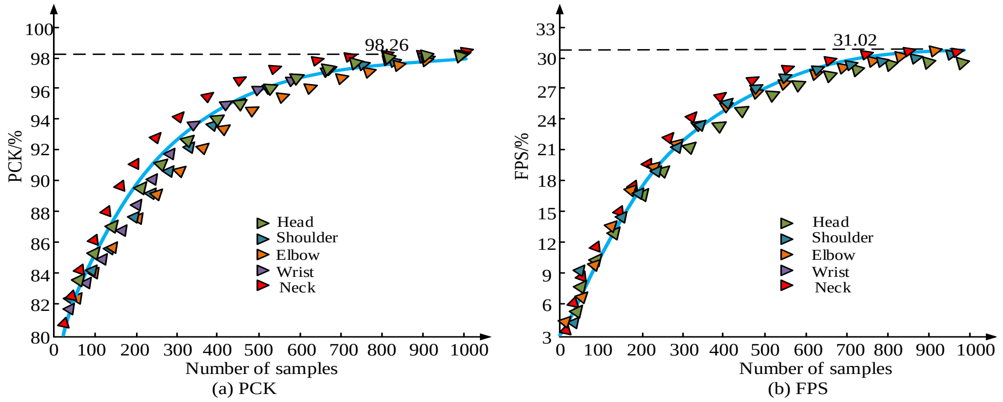
	High ball	Forehand draw	Forehand pick	Forehand net	Jump ball	Backhand draw	Backhand pick
High ball	99	62	33	21	27	42	28
Forehand draw	38	88	28	27	41	32	31
Forehand pick	67	27	69	52	28	22	48
Forehand net	33	12	52	93	36	33	19
Jump ball	72	68	20	29	86	18	48
Backhand draw	52	47	33	29	10	96	38
Backhand pick	42	13	19	24	58	47	95

(c) Transformer

	High ball	Forehand draw	Forehand pick	Forehand net	Jump ball	Backhand draw	Backhand pick
High ball	99	63	65	41	27	26	29
Forehand draw	67	95	28	44	48	22	25
Forehand pick	21	27	94	24	21	13	47
Forehand net	42	68	59	93	33	31	17
Jump ball	36	54	21	34	96	52	27
Backhand draw	67	48	26	27	27	97	19
Backhand pick	23	11	17	25	63	59	96

(d) Our model

Fig. 11. Confusion matrix results for different recognition models.



(a) PCK

(b) FPS

Fig. 12. The performance of modelling human skeletal point localization.

TABLE III. METRICS TEST RESULTS FOR DIFFERENT MODELS

Environments	Model	MSE	MAE	Resource consumption/%	Accuracy/s	References
Indoor	DFM	0.043	0.058	19.32	87.67	Lin L et al.
	TB-HPE	0.039	0.045	18.64	88.29	Gui W et al.
	DS-DNN	0.030	0.033	15.26	91.88	Xia Z et al.
	Our model	0.022	0.027	7.02	97.68	/
Outdoor	DFM	0.033	0.050	19.15	88.19	Lin L et al.
	TB-HPE	0.029	0.041	19.23	89.55	Gui W et al.
	DS-DNN	0.026	0.040	15.81	92.66	Xia Z et al.
	Our model	0.021	0.026	6.85	97.71	/

According to Table III, whether in indoor or outdoor environments, the DFM model had the worst performance in various indicators among the four models. The performance of TB-HPE, DS-DNN and the new target tracking and motion evaluation model increased from low to high. The proposed new target tracking and motion evaluation model had the lowest MSE value of 0.021, the lowest MAE value of 0.026, the lowest resource consumption rate of 6.85%, and the highest motion evaluation accuracy of 97.71%. Moreover, indoor and outdoor environments had almost no impact on the performance. In summary, the newly proposed model demonstrates the best overall performance and has more stable and superior recognition capabilities compared to the state-of-the-art motion assessment models currently available.

IV. Discussion

In the current field of sports motion target tracking and action assessment, deep learning technology, especially the application of Faster RCNN, has provided strong technical support for the automatic extraction of complex features from images. This enables algorithms to more accurately identify and track fast-moving sports targets, thereby significantly enhancing the accuracy and efficiency of target detection. However, due to the limited visual features of small-sized objects such as shuttlecocks, it is difficult for Faster RCNN to effectively distinguish them from similar small targets like sports shoes, light spots, and spectators' heads using limited visual features. In light of this, research has significantly improved Faster RCNN by introducing the VGG19 architecture, increasing the overall performance of the optimized Faster RCNN by about 11%. This indicates that the VGG19 architecture has a significant advantage in improving Faster RCNN's tracking and recognition of small targets. The result is consistent with the research findings of A Faghihi et al [13]. At the same time, in order to effectively obtain the posture information of the athletes, the study also used the improved Faster RCNN as the basic framework, and analyzed and processed the posture information of the badminton players through OpenPose and DT-SVM classifiers, finally proposing a tracking and action assessment model based on Faster RCNN and VGG19. Experimental results show that the proposed target tracking and action assessment model can achieve an average FPS processing speed of 31.02 frames per second for five key skeletal points of the human body: the head, shoulders, elbows, wrists, and neck. Moreover, its highest PCK detection accuracy rates for the head, shoulders, elbows, wrists, and neck reached 98.05%, 98.10%, 97.89%, 97.55%, and 98.26% respectively. This shows that OpenPose can process images in real-time and detect key points of multiple people, enhancing the model's joint point recognition ability. This is consistent with the research results of Chen C C et al. [31].

In summary, the research method has shown significant advantages in improving the accuracy, efficiency, and classification of badminton tracking and action assessment, which is consistent with the research conclusions of A Faghihi et al. and Chen C C et al., verifying the application potential and practical value of the method in target tracking and action assessment. Future work can further explore the integration of deep learning models, optimize algorithm efficiency, and expand the scope of applications to enhance the overall performance and applicability of sports motion analysis technology.

V. Conclusion

The research and development of sports tracking and motion evaluation have always been of great concern. The target recognition ability is crucial for athletes to achieve autonomous cooperation and optimized decision-making during the exercise process. In view of this, the Faster RCNN was used as the basic framework for target tracking. Then the VGG19 architecture and OpenPose algorithm were introduced for accuracy adjustment and pose estimation. Finally, a new badminton tracking and motion evaluation model based on Faster RCNN and improved VGG19 was proposed. Compared with other target tracking methods, the Faster RCNN-VGG19-OpenPose had the highest overlap between the real and tracking trajectories, which had good robustness, achieving stable tracking of badminton. The Faster RCNN-VGG19-OpenPose object tracking algorithm had a maximum P-value of 95.56%, R-value of 93.68%, and F1-value of 95.32%, with an average detection time of only 2.51s. It has relatively good performance, which is more suitable for the badminton tracking work at current stage. The simulation results showed that the proposed tracking and motion evaluation model based on Faster RCNN and VGG19 achieved an average FPS processing speed of 31.02 frames per second for five bone points in the human head, shoulder, elbow, wrist, and neck. Its highest PCK detection accuracy for the head, shoulder, elbow, wrist, and neck reached 98.05%, 98.10%, 97.89%, 97.55%, and 98.26%, respectively. The impact of the environment on the model was relatively small, and its overall performance was the best. Compared with similar motion recognition models, it has more stable and excellent recognition ability. In summary, the model has certain advantages and feasibility in the recognition and evaluation of badminton swing motions. However, this study only identifies a single badminton swing motion. Future research can add multiple combinations of badminton motions to improve the technical integrity. Additionally, although the study optimized the model through posture estimation technology, improving the estimation accuracy of the shoulder, elbow, and wrist joints, the estimation accuracy of other joint points has slightly decreased compared to the original model. To apply the research method to other more complex sports, future research can achieve more accurate human posture estimation by adjusting network structures, improving training strategies, or applying data augmentation techniques. At the same time, multimodal data, such as depth information or inertial sensor data, can be considered to enhance the model's ability to capture complex movements, thereby further improving the model's performance and providing more robust technical support for the rest of the sports training and assessment.

References

[1] Oh N. and Rodrigue H.. "Toward the development of large-scale inflatable robotic arms using hot air welding," Soft Rob, vol. 10, no. 1, pp. 88-96, January, 2023, DOI:10.1089/soro.2021.0134.

[2] RPohan M. A. and Utama J.. "Efficient Sampling-based for Mobile Robot Path Planning in a Dynamic Environment Based on the

Rapidly-exploring Random Tree and a Rule-template Sets," Int. J. Eng, vol. 36, no. 4, pp. 797-806, April, 2023, DOI:10.5829/IJE.2023.36.04A.16.

[3] Tian D., Han Y., Wang S., 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, June, 2022, DOI:10.1016/j.neucom.2022.06.018.

[4] Meza J. , Delpiano J. , S. Velastín, R. Fernández and S. Awad. "Multiple Object Tracking for Robust Quantitative Analysis of Passenger Motion While Boarding and Alighting a Metropolitan Train," International Conference of Pattern Recognition Systems, vol. 43, no. 2, pp. 231-238, March, 2021, DOI:10.1049/icp.2021.1468.

[5] Shimizu T., Hachiuma R. and H. Kajita. "Hand Motion-Aware Surgical Tool Localization and Classification from an Egocentric Camera," Journal of Imaging,vol. 7, no. 2, pp. 15-19, March, 2021, DOI:10.3390/jimaging7020015.

[6] Li H. and Li D. "Recognition and Optimization Analysis of Urban Public Sports Facilities Based on Intelligent Image Processing," Hindawi, vol. 8, no. 9, pp. 42-48, March, 2021, DOI:10.1155/2021/8948248.

[7] Yin X, and Chen L.. "Image Object Detection Method Based on Improved Faster R-CNN," Journal of Circuits, Systems and Computers, vol. 33, no. 7, pp. 54-59, March, 2024, DOI:10.1142/S0218126624501305.

[8] Kandhro I. A., Manickam S. and Fatima K. "Performance evaluation of E-VGG19 model: Enhancing real-time skin cancer detection and classification," Heliyon, vol. 10, no. 10, pp. 35-41, March, 2024, DOI:10.1016/j.heliyon.2024.e31488.

[9] Awan M. J., Masood O. A. and Mohammed M. A.. "Image-Based Malware Classification Using VGG19 Network and Spatial Convolutional Attention," Electronics, vol. 10, no. 19, pp. 2444-2454, 2021, DOI:10.3390/electronics10192444.

[10] Verma S., Singh G. and Warishpat. "Detection of Traffic Sign using Inception V3 in Comparison with VGG-19 to Measure Accuracy," International Conference on Advance Computing and Innovative Technologies in Engineering, vol. 32, no. 4, pp. 730-733, March, 2023, DOI:10.1109/ICACITE57410.2023.10183243.

[11] Wan X., Zhang X. and L. Liu. "An Improved VGG19 Transfer Learning Strip Steel Surface Defect Recognition Deep Neural Network Based on Few Samples and Imbalanced Datasets," Applied Sciences, vol. 11, no. 6, pp. 2606-2616, March, 2021, DOI:10.3390/app11062606.

[12] Mohan R., Rama A. and K. Ganapathy. "Comparison of Convolutional Neural Network for Classifying Lung Diseases from Chest CT Images," International Journal of Pattern Recognition and Artificial Intelligence, vol. 8, no. 14, pp. 25-30, April, 2022, DOI:10.1142/S0218001422400031.

[13] Faghihi A., Fathollahi M. and R. Rajabi. "Diagnosis of skin cancer using VGG16 and VGG19 based transfer learning models," Multimedia Tools and Applications, vol. 83, no. 19, pp. 57495-57510, April, 2024, DOI:10.1007/s11042-023-17735-2.

[14] Ding L., Peng J. and Song L. "Automatically detecting apnea-hypopnea snoring signal based on VGG19+LSTM. Biomed," Signal Process. Control. vol. 80, no. 10, pp. 351-370, April, 2023, DOI:10.1016/j.bspc.2022.104351.

[15] Fan J., Yang X., Lu R., W. Li, and Y. Huang,"Long-term visual tracking algorithm for UAVs based on kernel correlation filtering and SURF features," Vis. Comput., vol. 39, no. 1, pp. 319-333, Jan. 2023. DOI: 10.1007/s00371-021-02331-y.

[16] Gali V., Babu B. C., R. B. Mutluri, M. Gupta, and S. K. Gupta,"Experimental investigation of Harris Hawk optimization-based maximum power point tracking algorithm for photovoltaic system under partial shading conditions," Opt. Control Appl. Methods, vol. 44, no. 2, pp. 577-600, Aug. 2023. DOI: 10.1002/oca.2773.

[17] He Q., Li X., and Li W. ,"Common Sports Injuries of Track and Field Athletes Using Cloud Computing and Internet of Things," Int. J. Comput. Intell. Syst., vol. 16, no. 1, p. 70, May 2023. DOI: 10.1007/s44196-023-00257-y.

[18] Murugan R. A., and Sathyabama B.,"Object Detection for Night Surveillance Using Ssan Dataset Based Modified Yolo Algorithm in Wireless Communication," Wireless Personal Commun., vol. 128, no. 3, pp. 1813-1826, Sep. 2023. DOI: 10.1007/s11277-022-10020-9.

[19] Chen Y., Zheng W., Zhao Y., Song T. H., and Shin H.,"Dw-yolo: an efficient object detector for drones and self-driving vehicles," Arabian J. Sci. Eng., vol. 48, no. 2, pp. 1427-1436, Feb. 2023a. DOI: 10.1007/s13369-022-06874-7.

[20] Chen Y., Xu H., X. Zhang, P. Gao, Z. Xu, and X. Huang,"An object detection method for bayberry trees based on an improved YOLO algorithm," Int. J. Digit. Earth, vol. 16, no. 1, pp. 781-805, 2023b Mar. DOI: 10.1080/17538947.2023.2173318.

[21] Yang Y., Chen L., Zhang J., Long L., and Wang Z., "UGC-YOLO: Underwater Environment Object Detection Based on YOLO with a Global Context Block," J. Ocean Univ. China, vol. 22, no. 3, pp. 665-674, May 2023. DOI: 10.1007/s11802-023-5296-z.

[22] Zhang J., He Y., Feng W., J. Wang, and N. N. Xiong,"Learning background-aware and spatial-temporal regularized correlation filters for visual tracking," Appl. Intell., vol. 53, no. 7, pp. 7697-7712, Jul. 2023. DOI: 10.1007/s10489-022-03868-8.

[23] Aygül K., Cikan M., Demirdelen T., and M. Tumay,"Butterfly optimization algorithm based maximum power point tracking of photovoltaic systems under partial shading condition," Energy Sources, Part A: Recovery, Util. Environ. Effects, vol. 45, no. 3, pp. 8337-8355, Oct. 2023. DOI: 10.1080/15567036.2019.1677818.

[24] Chessa A., Urso P. D', L. De Giovanni, V. Vitale, and A. Gebbia,"Complex networks for community detection of basketball players," Ann. Oper. Res., vol. 325, no. 1, pp. 363-389, Oct. 2023. DOI: 10.1007/s10479-022-04647-x.

[25] Rodríguez-Fernández A., 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," Proc. Inst. Mech. Eng., Part P: J. Sports Eng. Technol., vol. 237, no. 3, pp. 188-196, Jul. 2023. DOI: 10.1177/17543371211031160.

[26] Rahimian P. and Toka L.,"Optical tracking in team sports: A survey on player and ball tracking methods in soccer and other team sports," J. Quant. Anal. Sports, vol. 18, no. 1, pp. 35-57, Mar. 2022. DOI: 10.1515/jqas-2020-0088.

[27] Liu H., Duan X., H. Chen, H. Lou, and L. Deng,"DBF-YOLO: UAV Small Targets Detection Based on Shallow Feature Fusion," IEEJ Trans. Electr. Electron. Eng., vol. 18, no. 4, pp. 605-612, Jan. 2023. DOI: 10.1002/tee.23758.

[28] Gai R., Chen N., and H. Yuan,"A detection algorithm for cherry fruits based on the improved YOLO-v4 model," Neural Comput. Appl., vol. 35, no. 19, pp. 13895-13906, Jul. 2023. DOI: 10.1007/s00521-021-06029-z.

[29] Yang D., "Research on multi-target tracking technology based on machine vision," Appl. Nanosci., vol. 13, no. 4, pp. 2945-2955, Apr. 2023. DOI: 10.1007/s13204-021-02293-6.

[30] Hasanvand M., Nooshyar M., E. Moharamkhani, and A. Selyari. "Machine Learning Methodology for Identifying Vehicles Using Image Processing," AIA, vol. 1, no. 3, pp. 170-178, Apr, 2023, DOI: https://doi.org/10.47852/bonviewAIA3202833

[31] Chen C. C., Chang C., Lin C. S., et al., "Video Based Basketball Shooting Prediction and Pose Suggestion System," Multimedia Tools and Applications, vol. 82, no. 18, pp. 27551-27570, May 2023. Doi: 10.1007/s11042-023-14490-2.