ITC 3/55
Information Technology and Control
Vol. 55 / No. 3/ 2026
pp. 1295-1316
DOI 10.5755/j01.itc.55.3.44551

Shuttlecock Movement Detection and Evaluation Method Based on Improved YOLOv8 and Image Segmentation

Received 2026/03/16 | Accepted after revision 2026/06/18

HOW TO CITE: Qu, X., Zhu, F. (2026) Shuttlecock Movement Detection and Evaluation Method Based on Improved YOLOv8 and Image Segmentation. *Information Technology and Control*, 55(3), 1294-1315. https://doi.org/10.5755/j01.itc.55.3.44551

Shuttlecock Movement Detection and Evaluation Method Based on Improved YOLOv8 and Image Segmentation

Xiaoyu Qu*, Feng Zhu
College of Physical Education, Liaoning Petrochemical University; Fushun, 113001, China; e-mails: XiaoyuxyQu@outlook.com, tiyuren@163.com

Corresponding author: XiaoyuxyQu@outlook.com

In view of the problems existing in badminton sports analysis such as strong subjective judgment, difficulty in quantification, and separation of correlation analysis between action standardization and tactical effectiveness, this study proposes an improved YOLOv8 using the LAFF module. The network combines the shuttlecock intelligent detection and evaluation method of DBAG-SN, and builds an intelligent evaluation model for multi-source information fusion on this basis. First, the LAFF module and shape-aware loss are introduced in the detection network to enhance the ability to capture small targets. DBAG-SN is designed and combined with bone key point detection and action recognition technology to achieve refined analysis of key technical action links such as racquet introduction, swing, hitting, and follow-up swing. At the same time, multi-view geometric constraints are used to reconstruct the 3D trajectory, and the relationship between technical actions and tactical effects is established through the attention fusion network. Experiments show that in the self-built shuttlecock metric dataset, the improved model has a maximum mAP of 92.4% when the threshold is 0.5, an average increase of 7.1% compared with the baseline model before improvement, and the positioning accuracy improvement under high IoU thresholds is more significant. In the image segmentation task, the average IoU of DBAG-SN is 85.2%, which is better than mainstream algorithms such as Mask Region-Convolutional Neural Network. In actual application verification, the 3D trajectory reconstruction RMSEpos of this intelligent fusion evaluation model is reduced to 18.7 mm, and the Pearson correlation co-

efficient between the comprehensive score generated and the expert score is as high as 0.91. Compared with the single-dimensional evaluation method, the dual-stream fusion mechanism that combines action standardization and tactical effectiveness enables the evaluation's comprehensive evaluation index F1-Score to reach 85.5%, which is 12.6% and 17.7% higher than single action evaluation and single tactical evaluation. The main contribution of this study is to propose a closed-loop evaluation framework from multimodal perception to causal decision-making of "technical tactics". Through the dual stream fusion mechanism of action and landing point, it effectively solves the problem of separation between action and tactics in training analysis, and can provide objective data support for scientific badminton training.

KEYWORDS: YOLOv8, Dual-branch attention segmentation, Multi-source information fusion, Three-dimensional trajectory reconstruction, Shuttlecock

1. Introduction

Badminton is a competitive sport characterized by fast ball speed, complex movements and changeable tactics. Traditional training and game analysis relies heavily on coaches' empirical observations, which has limitations such as strong subjectivity, difficulty in quantification, and low efficiency. To break through this experience-led training bottleneck and provide athletes with more scientific, accurate, and efficient training and competition support, modern sports technology is accelerating its transformation to intelligence and digitalization [5]. This transformation process urgently requires the system to use advanced computer vision technology to capture athletes' body postures and joint motion trajectories in high-speed movements in real time through multi-angle cameras. Using bone key point detection and motion recognition technology, the whole process of technical movements such as racquet introduction, swing, hitting, and follow-up swing is meticulously dismantled. Calculating the accuracy, fluency, and coordination scores of movements by comparing them with standard movement models can help athletes pinpoint technical shortcomings [4]. At the same time, due to the small size and extremely fast flight characteristics of shuttlecock, the intelligent analysis system also needs to have the ability to conduct all-round and high-precision tracking of the flying ball. The system monitors the ball's 3D spatial position and landing point distribution in real time to eliminate the visual errors of human referees and ensure the fairness and accuracy of game decisions. In addition, combined with vision-based speed calculation technology, the system should also be able to evaluate batting power and explosive power in real time, providing a quantitative basis for analyzing the effectiveness of technology application [20, 25]. This study proposes an intelligent detection and evaluation method based on improved YOLOv8 and attention-enhanced image segmentation. The core innovation is to introduce the Lightweight Adaptive Feature Fusion (LAFF) module into the YOLOv8 network to improve the detection capabilities of small-sized shuttlecock. It designs a Dual-Branch Attention-Guided Segmentation Network (DBAG-SN) to achieve pixel-level accurate analysis of athletes and shuttlecock. Furthermore, an intelligent evaluation model based on multi-source information fusion is proposed to reconstruct the shuttlecock 3D trajectory through multi-view triangulation, and extract the athlete's skeletal key point sequence based on the improved posture estimation network. Subsequently, on this basis, the action normative evaluation sub-module and the landing point tactical evaluation sub-module are constructed, and the feature-level fusion decision-making mechanism is used to achieve collaborative optimization of multi-dimensional evaluation indicators. The innovation of this article mainly lies in: (1) It moves the lightweight feature fusion strategy from within the backbone network to the horizontal connections of the feature pyramid, and enhances small target response through adaptive alignment of shallow textures and deep semantics. (2) It designs a task decoupled dual branch network to meet the heterogeneous segmentation needs of shuttlecock and athletes, and

achieve pixel level accurate analysis through differentiated attention allocation. (3) A joint modeling framework for action norms and tactical effects has been constructed, replacing single dimensional independent evaluation with association analysis from multi-source temporal data. The main contribution of this study is: (1) This study proposes a unified perception architecture that integrates detection, segmentation, 3D trajectory reconstruction, and skeletal keypoint recognition. (2) This study constructs an intelligent evaluation model that integrates action standardization and tactical effectiveness, and establishes a causal relationship between "technical and tactical" aspects. (3) This study designs LAFF and shape aware loss, enhancing the perception ability of high-speed small targets and non-rigid human body deformations.

2. Related Work

The intelligent motion analysis system has important research significance in promoting the standardization and intelligent upgrading of the badminton match, and many scholars have conducted research on this. Haq and Nagawa [10] proposed a YOLO detection framework that combines multiple training heuristics to address the balance between speed and accuracy in visual tracking. It improved the average detection accuracy to 36.0 while maintaining real-time performance, optimizing the dynamic recognition efficiency [10]. Zhi et al. [29] aimed at the insufficient research on batting action recognition and proposed a temporal segmentation network that incorporated lightweight attention. The network achieved efficient recognition of four common movements such as forehand, with the average recall rate and accuracy exceeding 91% [29]. Lianju et al. [16] proposed an improved deep learning method that integrated static and dynamic features to address the difficulty in identifying take-off movements. This method used multi-feature weighted fusion to significantly improve the recognition accuracy and real-time performance [16]. Nugroho et al. developed an IoT based badminton agility and speed testing instrument, confirming the need for IoT integrated agility and speed testing instruments for badminton skill assessment [18]. Purnama et al. [22] artificially reduced the analysis error and proposed a temporal classification method that combined joint tracking and LSTM. This method achieved real-time and accurate classification of actions such as high and long shots, with a recognition accuracy of over 90%, optimizing the degree of automation of analysis [22].

The You Only Look Once Version 8 (YOLOv8) relies on the anchor-free paradigm and the C2f module to achieve Pareto optimality of accuracy and speed through multi-scale parallel search, and can lock high-speed tiny targets and non-rigid human body features at low cost and efficiently [2]. Perkasa et al. [21] verified the advantages of YOLOv8 in balancing accuracy and speed in football scenarios, with mAP@0.5 reaching 87.4%, proving its superiority as a motion detection base. Adam et al. [1] proposed an automatic archery scoring system based on YOLOv8, which realized real-time identification and distance estimation of tiny arrows, demonstrating its potential to handle high-speed small targets and precise positioning.

Image segmentation technology used pixel-level classification and Attention Mechanism (AM) to analyze deep features, overcame the geometric limitations of rectangular frames, and could accurately reconstruct the physical contours of targets under complex occlusions [30]. Shaker et al. [23] introduced efficient paired attention blocks in UNETR++, which reduced the number of parameters and improved accuracy, providing ideas for the design of lightweight attention segmentation networks. Tang et al. [26] proposed U-Net that fuses multi-scale attention, which significantly improved the segmentation accuracy of sports injury images and verified the effectiveness of AM in processing human body deformation.

In summary, although significant progress has been made in the visual detection and basic action classification of individual targets in shuttlecock, traditional detection algorithms still struggle to meet the strict requirements of high-level competitive analysis in complex backgrounds when faced with high-speed moving small shuttlecock. Weak feature enhancement and ultra high speed target recognition are common challenges in such scenarios [11, 14]. To this end, this study builds an intelligent evaluation system that integrates action normative evaluation, 3D trajectory tracking, and tactical effectiveness analysis.

3. Methodologies

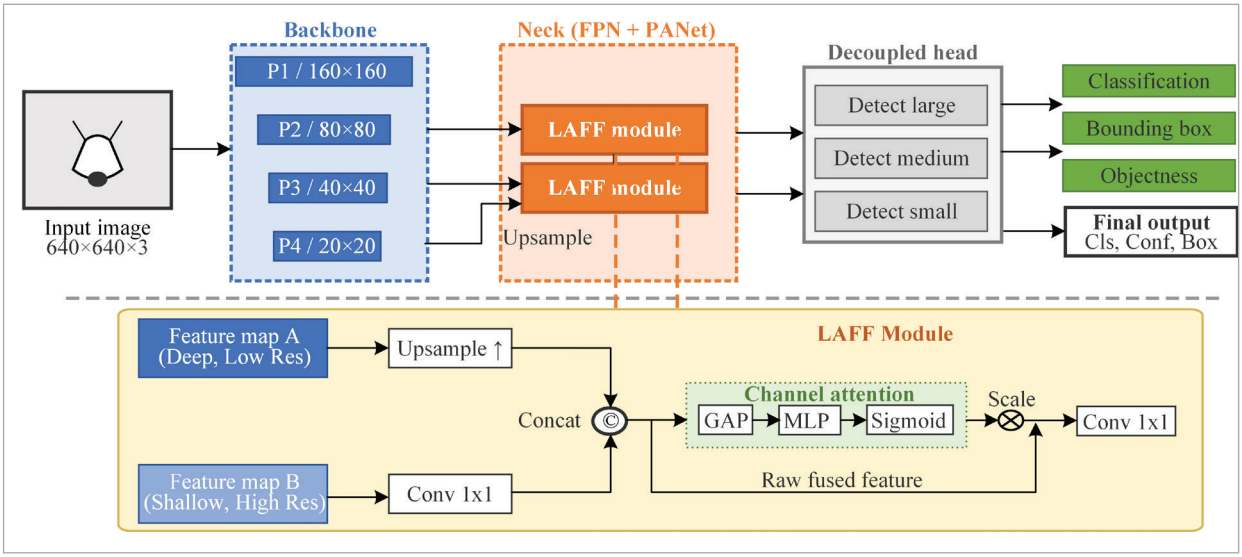
3.1. Improved YOLOv8 Detection Model for Body State and Motion Capture

In the AI intelligent evaluation system of swing sports, accurately capturing the athlete's physical condition and analyzing the entire process of technical movements is the basis for scientific training. The system not only needs to lock the high-speed flying shuttlecock in real time, but also needs to continuously track the athlete's body posture in various key technical links such as racket introduction, swing, hitting and follow-up swing [7]. However, when faced with motion blur and non-rigid deformation of the human body, general detection algorithms often have difficulty balancing the positioning accuracy of small targets (balls) and the feature integrity of key point areas of human bones [13]. To this end, this study constructs an improved YOLOv8 detection model as the visual perception core of the AI motion capture system. This model introduces LAFF and shape-aware loss to build an improved detection framework that can enhance shallow responses, explicitly model geometric consistency, and automatically adapt to motion features. It has crucial theoretical value and practical significance for realizing the paradigm transformation from general visual model to dedicated sports AI. The structure of the improved YOLOv8 detection model is shown in Figure 1.

In Figure 1, compared with the standard YOLOv8, the improvement of the improved YOLOv8 detection model is mainly reflected in the embedding of a LAFF module at the horizontal connection of the neck, replacing simple feature stacking. Specifically, the badminton match image at the input end is sent to the CSPDarknet backbone network, and four feature maps of different scales are extracted through layer-by-layer downsampling. Among them, shallow features retain the rich texture and geometric details of small objects such as shuttlecock, while deep features encode high-dimensional semantic information. The data flow then enters the core neck network. In order to overcome the problem of detail loss in the traditional FPN structure when processing high-speed small targets, this study embeds LAFF modules at the lateral connections of the feature pyramid. This module does not directly superimpose features, but aligns and splices the upsampling results of deep semantic features with shallow detail features through a dual-stream path. It utilizes channel AM dynamic reweighting, thereby explicitly enhancing the network's responsiveness to small target areas in the feature fusion stage [8, 9, 15]. In the dual stream fusion mechanism, the attention fusion network quantifies the statistical correlation strength

Figure 1

Structure of the improved YOLOv8 detection model.



between specific action deviations and tactical effectiveness attenuation by learning the cross-modal interaction weights between action features and tactical features. When the attention weight of a deviation node in a certain action is significantly high, the association is identified as a key factor affecting tactical scores, providing a quantifiable explanatory basis for the diagnostic chain. Although the framework of this study did not introduce a formal counterfactual reasoning model, the fusion network can reflect the weakened contribution of action features to tactical scores in the scene through the distribution shift of attention weights, thereby guiding the analysis focus to the landing point distribution and rhythm features of the tactical branch. This mechanism enables the model to not only output a comprehensive score highly correlated with expert ratings (r=0.91), but also trace the feature attribution path behind the score at the statistical association level, providing a traceable feature basis for the subsequent introduction of more rigorous intervention analysis tools such as structural causal models. Finally, the enhanced multi-scale features are fed into the decoupled detection head, which separates the classification task from the regression task, and outputs the target's category probability, bounding box coordinates and target existence confidence through parallel convolution branches. Finally, high-precision shuttlecock and player detection results are output through non-maximum suppression. Among them, the calculation of LAFF is shown in Equation (1).

$$F_{LAFF} = C_{fusion}(\left[U(F_{deep}); C_{trans}(F_{shallow})\right] \otimes \mathbf{M}_{attention} . \quad (1)$$

In Equation (1), [;] is the splicing operation along the channel dimension. $U(\cdot)$ is the upsampling operator, $\otimes$ is the element-wise Hadamard product, and C_{fusion} is the final feature fusion convolution layer. $\mathbf{M}_{attention}$ is the generated adaptive attention weight matrix. To allow the network to automatically focus on the small area containing the shuttlecock, the attention weight matrix $\mathbf{M}_{attention}$ is generated based on the spliced feature map F_{cat}. This study introduces global average pooling to compress spatial information, uses a multi-layer perceptron to capture the nonlinear dependence between channels, and finally activates it through the Sigmoid function $\sigma(\cdot)$, as shown in Equation (2).

$$\mathbf{M}_{attention} = \sigma(\mathbf{W}_2 \cdot \delta(\mathbf{W}_1 \cdot GAP(F_{cat}))) . \quad (2)$$

In Equation (2), $\delta(\cdot)$ represents the ReLU function. $\mathbf{W}_1$ and $\mathbf{W}_2$ are the learnable weight parameters of the dimensionality reduction layer and dimensionality enhancement layer in MLP. With this structure, the module achieves explicit enhancement of key channel features while remaining lightweight. Finally, to address the deformation problem caused by the high-speed movement of shuttlecock, this study introduces a shape-aware term into the regression loss to improve positioning accuracy by strengthening the consistency constraint on the aspect ratio between the predicted frame and the real frame. The total regression loss is specifically shown in Equation (3).

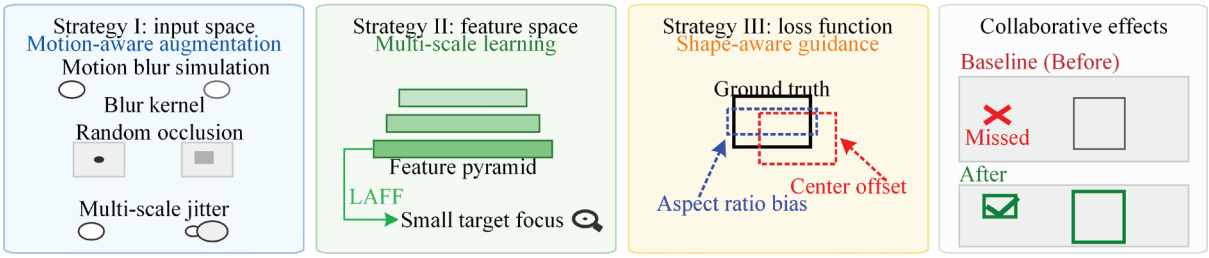
$$L_{reg} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha \cdot \frac{4}{\pi^2}(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h})^2. \quad (3)$$

In Equation (3), IoU is the intersection and union ratio. $\rho(\cdot)$ is the Euclidean distance between the predicted box center b and the true box center b^{gt}. c is the diagonal length of the minimum circumscribed rectangle. w and h are the width and height of the bounding box, and α is the balance coefficient. This loss function forces the network to focus not only on the degree of overlap during the regression process, but also on the precise matching of the target geometry.

The improved YOLOv8 architecture theoretically strengthens the feature representation capabilities of small targets. However, in real high-speed competitive scenes, motion blur, morphological distortion and complex background interference caused by the movement of the sphere still pose severe challenges to the generalization robustness of the detector. Stacking that relies solely on network depth is difficult to reach the performance bottleneck [17, 28]. To this end, this study uses motion sensing data enhancement to simulate real physical disturbances, uses multi-scale feature learning strategies to strengthen targeted guidance of small target representations, and uses shape-aware loss functions to refine the optimization goals of bounding box regression. This collaborative optimization strategy and its mechanism of action are shown in Figure 2.

Figure 2

Mechanism of motion-aware augmentation and multi-task loss synergistic optimization.



In Figure 2, first in the input space, motion blur simulation and mosaic enhancement strategies are introduced to address the unique physical degradation phenomenon of shuttlecock. It actively injects controlled image perturbations at the training end to force the model to learn robust feature expressions against fuzzy boundaries and local occlusions. Secondly, in the feature space, an improved feature pyramid structure is used to establish a "green channel" of small target information, ensuring that shallow high-resolution textures can be transmitted to the detection head losslessly, and solving the problem of detail annihilation caused by multi-layer convolution. Finally, at the level of loss function guidance, a composite objective function including shape-aware regression terms and zoom classification terms is constructed to constrain the descending direction of the gradient from the two dimensions of geometric accuracy and sample balance, thereby achieving precise positioning of rapidly deforming targets [24]. To mathematicalize the above optimization strategy, this study designs a total loss function for multi-task collaboration, whose formal definition is shown in Equation (4).

$$L_{tatal} = \lambda_{cls} L_{VFL} + \lambda_{box} L_{Shape} + \lambda_{dfl} L_{DFL} . \quad (4)$$

In Equation (4), λ_{cls}, λ_{box}, and λ_{dfl} are the hyperparameter weights of classification, bounding box regression, and distribution focus, which are used to balance the learning proportions of different tasks. Among them, the distribution focus loss further improves the accuracy of bounding box regression by modeling the probability distribution of the target box position, complementing the shape aware loss. To address the problem of sample imbalance (positive and negative samples and difficult and easy samples), the classification branch is optimized using zoom loss. This loss function uses the predicted quality score to dynamically adjust sample weights to make the model focus more on high-quality training samples, as shown in Equation (5).

$$L_{VFL}(p,q) = \begin{cases} -q\left[q\log(p) + (1-q)\log(1-p)\right] & q>0 \\ -\alpha p^{\gamma}\log(1-p) & q=0 \end{cases} . \quad (5)$$

In Equation (5), p is the predicted class confidence. q is the intersection score of the predicted box and the true box. γ is the modulation factor used to attenuate the dominant effect of large numbers of simple negative samples. At the same time, to address the problem of width and height distortion, the regression branch introduces a shape-aware loss term. On the basis of traditional geometric distance constraints, it adds a strong penalty on aspect ratio consistency to improve the model's geometric fitting ability for non-rigid deformation targets, as shown in Equation (6).

$$\begin{cases} L_{Shape} = 1 - IoU + \dfrac{\rho^2(b,b^{gt})}{c^2} + \beta v \\ v = \dfrac{4}{\pi^2}\left(\arctan\dfrac{w^{gt}}{h^{gt}} - \arctan\dfrac{w}{h}\right)^2 \end{cases} . \quad (6)$$

In Equation (6), v is the shape consistency penalty factor, and β is the dynamic balance coefficient. The shape aware term refers to the complete geometric constraint loss, while the shape consistency penalty factor specifically refers to the v factor used to calculate the square of the aspect ratio deviation in this term.

3.2. Dual-branch Attention-guided Image Segmentation Network Supporting Hawk-Eye

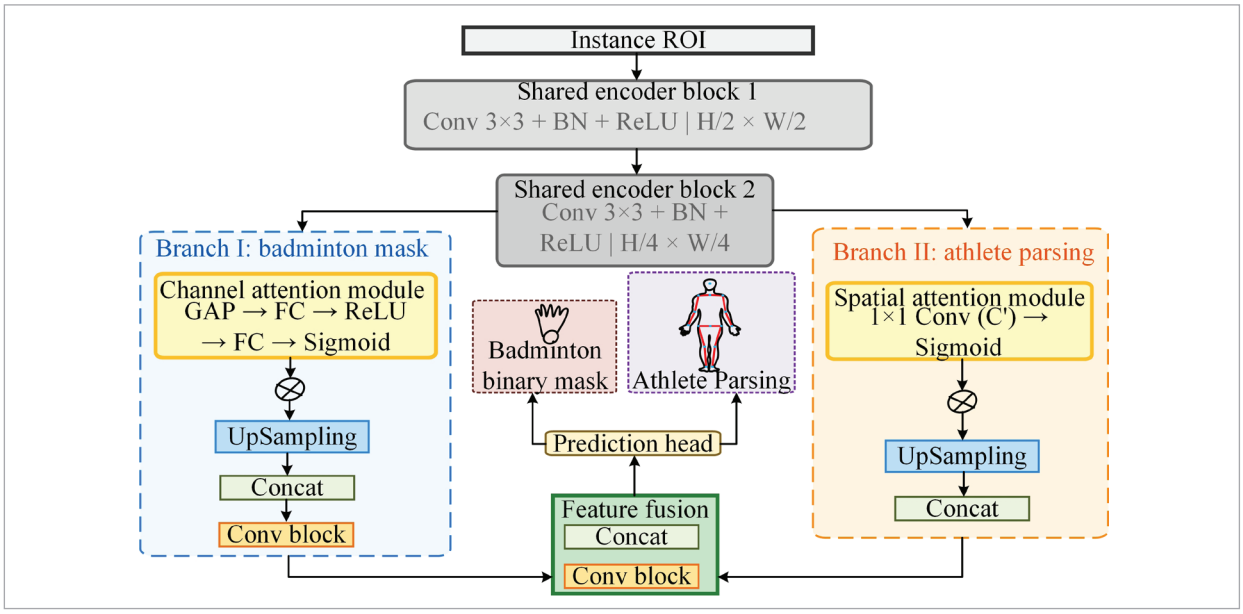
To overcome the limitations of general models in processing high-speed micro targets and complex human body deformations, this study constructs a DBAG-SN network based on the instance attention area generated by improved YOLOv8. The network adopts a parallel branch design with heterogeneous tasks to collaboratively optimize the weak texture features of shuttlecock and the non-rigid geometric structure of athletes. The core is to introduce a task-aware feature selection mechanism so that the network can adaptively focus on differentiated key information of different targets and achieve high-precision pixel-level analysis. The specific architecture is shown in Figure 3.

In Figure 3, DBAG-SN adopts the classic encoder-decoder topology to solve the heterogeneous segmentation problem where small targets and complex human body deformations coexist in badminton sports scenes. The common semantic features F_{enc} extracted by the shared feature encoder are shunted to two task-decoupled parallel branches at the bottleneck layer. Among them, the left branch specializes in shuttlecock mask generation. Given that the sphere texture is simple and the pixel ratio is extremely small, this branch embeds a channel attention module to explicitly model channel dependencies, thereby suppressing background noise channels and enhancing specific appearance feature responses. The design of this mapping is mainly based on the following considerations: small targets have sparse information in the channel dimension, channel attention compresses spatial information through global pooling and reweights the importance of each channel, which can effectively filter discriminative channels related to spherical texture. Meanwhile, for human targets with variable limb postures and occupying large spatial areas, spatial attention generates position sensitive weights on a two-dimensional plane, which is more helpful in capturing the geometric structure of bone edges and joint regions. The weighted feature V_{bad} is calculated as shown in Equation (7).

$$V_{bad} = F_{enc} \otimes \sigma(\Psi(GAP(F_{enc}))). \tag{7}$$

In Equation (7), Ψ is a multi-layer perceptron including a fully connected layer and ReLU, and GAP is a global average pooling. Correspondingly, the right branch is responsible for analyzing multiple parts of athletes. In view of the characteristics of changeable

Figure 3
DBAG-SN structure.



limb postures and occupying a large frame, a spatial attention module is introduced to enhance the geometric perception of bone edges and joint areas. This module extracts spatial features through parallel convolution transformation, and its output is shown in Equation (8).

$$V_{player} = F_{enc} \otimes \sigma(\phi_1(F_{enc}) \oplus (\phi_2(F_{enc})). \quad (8)$$

In Equation (8), ϕ and ϕ_2 both represent 1×1 convolution transformation. This mechanism enables the network to dynamically generate a two-dimensional spatial weight matrix while keeping the channel dimension unchanged. Both branches fuse the corresponding hierarchical features of the encoder stage through skip connections during the decoding process to restore the lost spatial details. Finally, the deep features with resolution restored through their respective decoding functions $D(\cdot)$ are spliced and blended in the fusion module to generate the final prediction map. The process is shown in Equation (9).

$$O_{final} = H_{head}(C_{fuse}(\left[D(V_{bad}); D(V_{player}) \right]. \quad (8)$$

In Equation (9), H_{head} is the final pixel-level classification prediction head. This feature selection through dual-branch AM and end-to-end optimization through multi-task loss can robustly output high-quality instance segmentation results, thus laying a solid foundation for subsequent precise motion analysis.

3.3. Hawk-Eye Trajectory 3D Reconstruction Method Combined with Ball Speed Detection

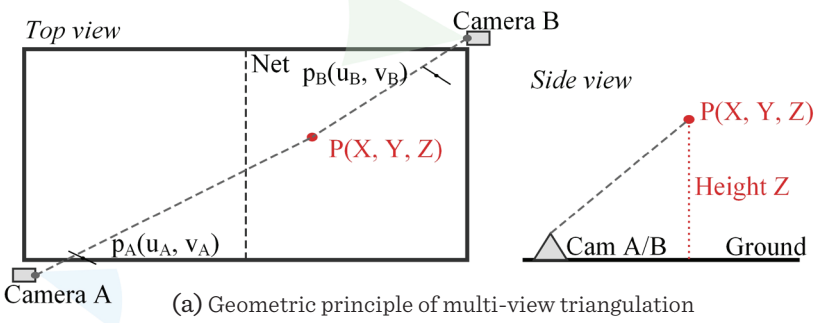
To achieve Hawk-Eye system-level free-throw percentage and achieve quantitative evaluation of batting quality, pure 2D visual observation can no longer meet the needs. It must have the ability to accurately reproduce the flight path of the ball in the 3D physical space, determine the inside and outside of the landing point (IN/OUT), and detect the flight speed of the ball in real time. To this end, this study proposes a trajectory reconstruction and parameter calculation method based on multi-view geometric constraints. It uses camera calibration parameters to map multi-view 2D observations to 3D space, and uses Kalman filtering to eliminate observation noise to generate a smooth 3D motion trajectory. On this basis, through time-series position difference and physical kinematics modeling, the instantaneous velocity and acceleration of the sphere are solved in real time. This computing framework can not only output millimeter-level impact point determination results, but also evaluate the athlete's explosive power through ball speed data, thereby deeply synergizing the Hawk-Eye trajectory tracking and ball speed detection functions. The calculation principle and algorithm flow are shown in Figure 4.

In Figure 4(a), the geometric nature of 3D space point reconstruction is revealed through the 2D perspective of top view and side view. The system relies on at least two sets of synchronized high-speed cameras with known spatial positions. The line of sight drawn by their optical centers passes through the 2D pixels (p_A, p_B) on the imaging plane and extends into space. According to the epipolar geometric constraints, the real 3D coordinate $P(X,Y,Z)$ of the shuttlecock is uniquely determined as the intersection point of multiple lines of sight in 3D space. This geometric constraint is the theoretical cornerstone of converting pixel-level observations into physical-level measurements, ensuring accurate recovery of depth information (Z-axis). Figure 4(b) presents the computing pipeline from video stream input to motion parameter output. The process covers pre-processing steps such as time-space coordinate correlation and lens distortion correction. The core link lies in multi-view triangulation solution. Due to the presence of measurement noise, lines of sight often do not intersect perfectly in space. Therefore, this study transforms the solution of 3D point coordinates into a nonlinear least squares optimization problem, aiming to find an optimal 3D point $\tilde{P}$ that minimizes the sum of reprojection errors on all camera planes. The optimization objective function is defined as shown in Equation (10) below.

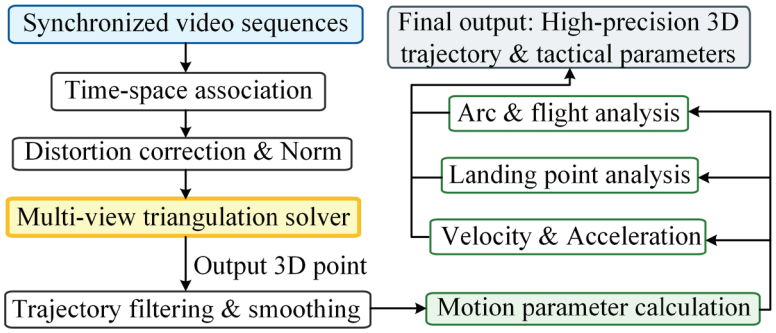
$$\tilde{P} = \arg\min_p \sum_{k=1}^{N} \left\| \pi(\mathbf{K}_k, \mathbf{R}_k, \mathbf{t}_k, P) - p_k \right\|^2. \quad (10)$$

In Equation (10), N is the number of cameras. p_k is the 2D pixel coordinate observed from the k-th viewing angle. $\mathbf{K}_k$ is the camera internal parameter matrix. $[\mathbf{R}_i|\mathbf{t}_k]$ is the external parameter matrix describing

Figure 4
Principle and parameter calculation flowchart of 3D trajectory reconstruction.



(a) Geometric principle of multi-view triangulation



(b) 3D trajectory reconstruction & parameter calculation flow

the camera posture. $\pi(\cdot)$ is the perspective projection function of the camera. By iteratively solving the above equation, the high-precision 3D coordinates of each frame can be obtained. Subsequently, the algorithm introduces Kalman filtering to perform temporal smoothing on the original point cloud sequence to eliminate outlier noise points and generate a continuous trajectory $T(t)$. Based on this, key tactical parameters such as instantaneous speed, flight arc characteristics, and landing point coordinates are further solved through parallel branches.

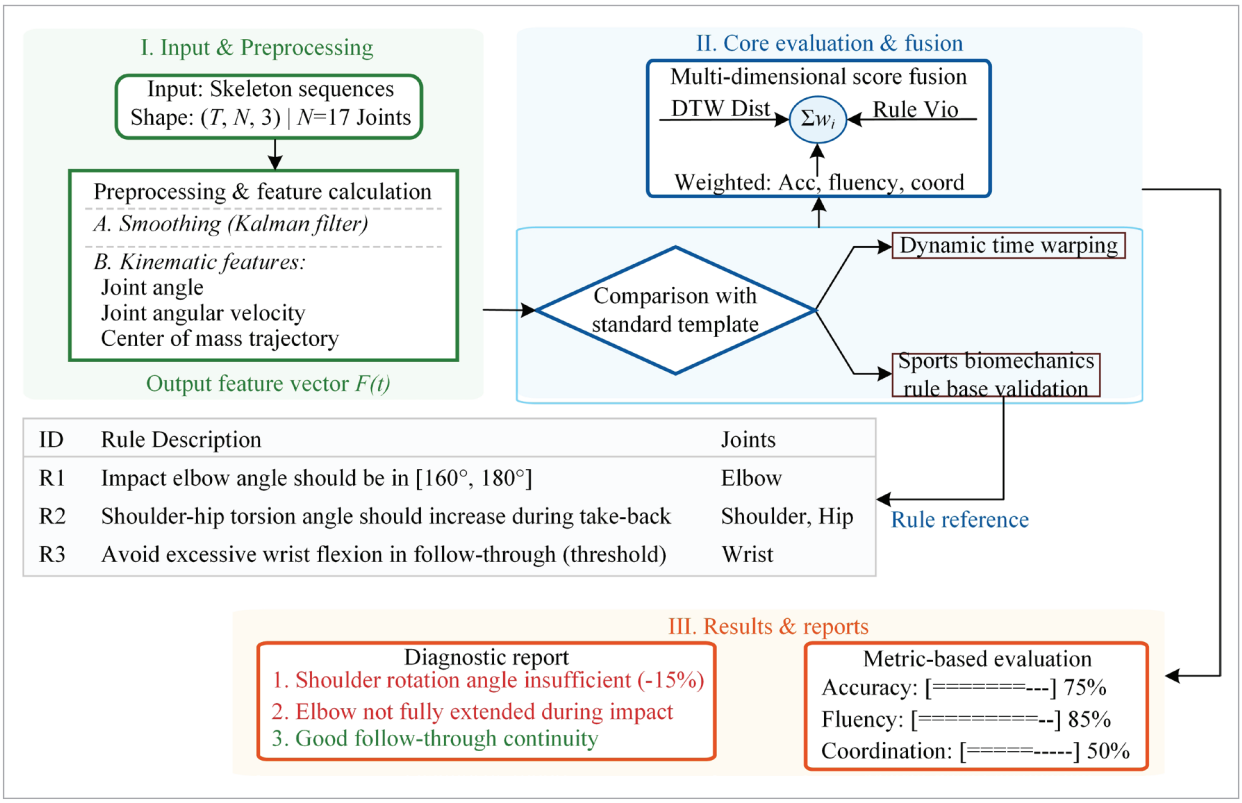
3.4. Multi-source Information Fusion Evaluation Model for Intelligent Coaching Decision-making

This study first integrates multi-source heterogeneous data such as time-series skeleton key points and 3D trajectory landing points on the basis of improvement, and proposes a hierarchical intelligent evaluation model. Subsequently, a dual-stream evaluation architecture is constructed that paralleled the standardization of movements and the tactical nature of landing points, and through a feature-level fusion decision-making mechanism, an integrated quantitative evaluation of shuttlecock performance from technical execution to tactical effectiveness is achieved.

3.4.1. Action Normative Evaluation Sub-model Based on Skeleton Point Sequence

As the biomechanical basis of the multi-source information fusion evaluation system, the movement normative evaluation sub-model aims to implement a refined quantitative analysis of the athlete's full-cycle technical links such as racket introduction, swing, hitting and follow-up swing. First, the high-precision time-series skeletal key points extracted in Section 2.2 are used as the data base, and then the spatial position, rotation angle and force timing characteristics of the human joint points are reconstructed from the kinematic dimension. Its core computing paradigm is to construct a mapping from discrete skeletal sequences to continuous action semantics: This study

Figure 5
Flowchart and decision rules of action normality assessment.



first solves high-order feature vectors with clear biomechanical significance based on original coordinates, covering joint angular velocity, limb end trajectory curvature and limb coordination. Then, it uses a dynamic time warping algorithm to perform time domain alignment and similarity measurement between the real-time action sequence and expert standard action templates, and combines the kinematics rule library to verify the consistency of the force sequence and joint activity range. The action normative evaluation process and core judgment rules based on the skeleton point sequence are shown in Figure 5.

In Figure 5, it follows a progressive mechanism from data input, feature quantification, double comparison to diagnostic output. First, the preprocessed time-series joint point coordinates are converted into high-order kinematic parameters (such as joint angles, angular velocity) to decouple the difference in the absolute position of the human body. Subsequently, the model uses a dynamic time warping algorithm and a sports biomechanics rule library for collaborative verification to calculate differences in timing shapes and rule violations at key force-generating nodes. Finally, the model integrates the above metrics to generate quantitative scores in the three dimensions of accuracy, fluency and coordination, and outputs a structured diagnostic report containing specific deviation locations. To achieve scientific conversion from original coordinates to quantifiable evaluation indicators, this study introduces key motion feature extraction and sequence similarity measurement algorithms. First, the accuracy of movements is highly dependent on the quantification of limb geometry. Taking the elbow joint as an example, to eliminate the influence of the shooting angle and human body scale, vector algebra is used to calculate the joint flexion and extension angle θ_{elbow} as the core feature, as shown in Equation (11).

$$\theta_{elbow} = \arccos\left(\frac{(E-S)\cdot(W-E)}{\|E-S\|\|W-E\|}\right). \tag{11}$$

In Equation (11), S, E, and W are the 3D space coordinate vectors of the shoulder joint, elbow joint, and wrist joint. Secondly, in view of the nonlinear differences in movement rates of different athletes, a dynamic time warping algorithm is used to calculate the minimum cumulative distance between the sequence Q to be evaluated and the standard template G, as an indicator of the difference in movement fluency and overall form. Its recursive solution formula is shown in Equation (12).

$$D(i,j)=d(q_i,c_i)+\min\begin{cases}D(i-1,j)\\D(i,j-1)\\D(i-1,j-1)\end{cases}. \tag{12}$$

In Equation (12), $D(i,j)$ is the minimum cumulative distance when frame i and frame j in the sequence are aligned. $d(q_i,c_i)$ is the Euclidean distance between the feature vectors of the two frames. By minimizing the final cumulative distance, the model achieves precise alignment and quantitative evaluation of heterogeneous duration action sequences.

3.4.2. Tactical Evaluation Sub-model Based on Trajectory and Landing Point

Following the normative evaluation of actions, the intelligent evaluation system needs to further objectively quantify the physical effectiveness of the shot. As a result-side analysis module, this sub-model uses the millimeter-level 3D trajectory and high-dimensional motion parameters obtained in Section 2.3 to build a mapping mechanism from physical characteristics to tactical semantics. The model adopts a hierarchical evaluation architecture that parallels rule adjudication and tactical quantification: First, based on the geometric constraints of the field boundary, it achieves millisecond-level automated determination of the landing point "IN/OUT", eliminating artificial parallax interference. Second, the speed attenuation rate, net passing height, and landing point spatial distribution entropy of the ball during its flight are analyzed in depth to quantitatively evaluate the suppressive force and tactical intention of a single shot. This transformation logic and overall analysis framework from spatio-temporal data to tactical value is shown in Figure 6.

Based on the multi-modal perception data output from the improved YOLOv8 detection model, the tactical evaluation submodule in Figure 6 further receives multidimensional physical parameters such as three-dimensional trajectory, landing point coordinates, and ball velocity curve, and then enters a parallel two-layer analysis module: The base layer performs strict "compliance determination" (IN/OUT determination) based on the geometric rules of the field. The core layer, under the guidance of the tactical knowledge base, deeply analyzes the tactical intention and suppression effect of the shot through the three dimensions of impact point distribution, rhythm changes and flight arc. Finally, the multi-source features are weighted and fused at the decision-making layer to generate a comprehensive evaluation report containing quantitative scores and structured semantic suggestions. To mine potential tactical space preferences (such as "repeatedly suppressing backhand backcourt") from discrete landing point data, this study introduces the kernel density estimation method to conduct non-parametric modeling of landing point distribution. Given a set of N historical landing point coordinates within an observation time window, the landing point probability density $\hat{f}(x)$ at any position x on the court plane is defined as shown in Equation (13).

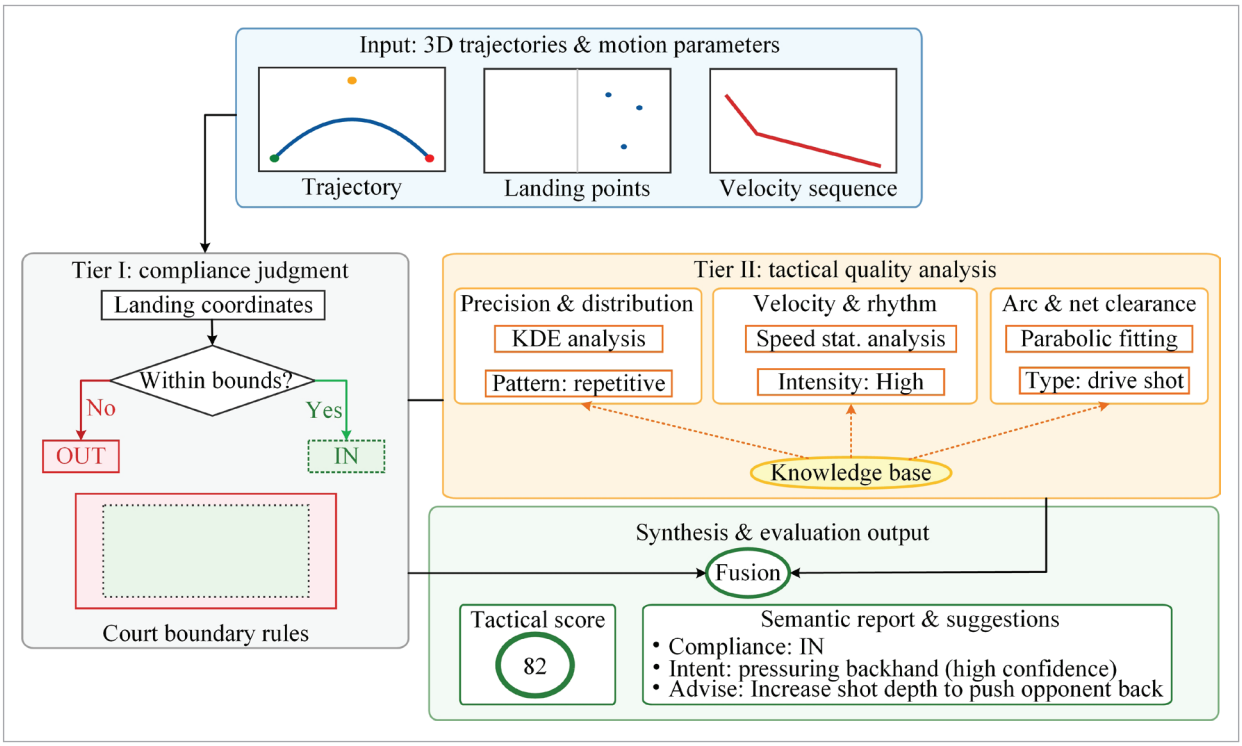
$$\hat{f}(x)=\frac{1}{Nh^2}\sum_{i=1}^{N}H\left(\frac{\|x-x_i\|}{h}\right). \tag{13}$$

In Equation (13), $H(\cdot)$ is the kernel function, which is used to smooth the influence range of discrete data points. By calculating the density response matrix of the entire field, the model can automatically identify high-frequency landing hot spots, thereby providing an objective mathematical basis for determining "tactical oppressiveness" and "landing accuracy."

3.4.3. Multi-dimensional Integrated Decision-making and Suggestion Generation

It is difficult to construct an overall evaluation system with causal explanation power from a single perspective based on the independent analysis of the human (action normativeness) and shuttlecock (tactical effectiveness) dimensions, and it is even more difficult to directly support the high-order decision-making needs of the intelligent coaching

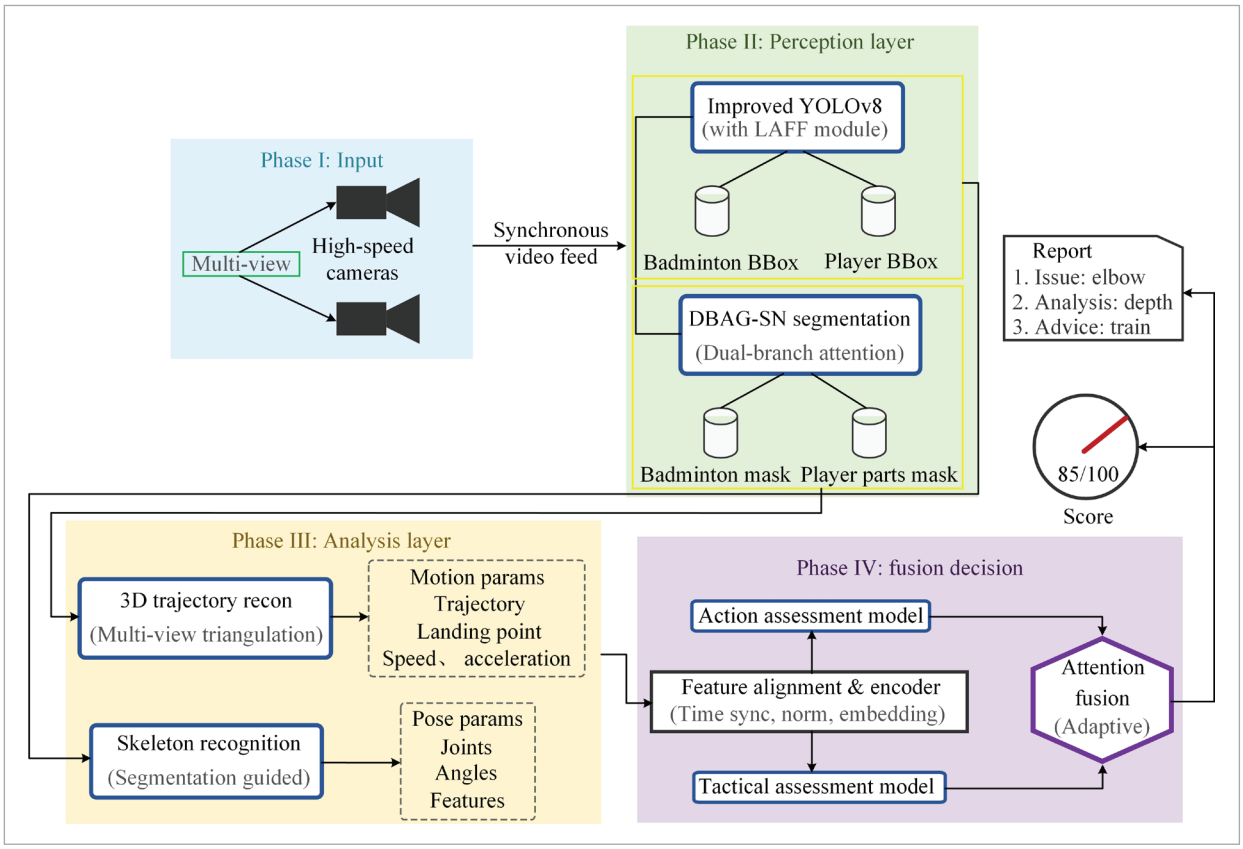
Figure 6
Logic and overall analysis framework of tactical assessment.



guidance system. To this end, this study proposes an intelligent fusion judgment model for badminton sports based on improved YOLOv8 and image segmentation, and names it YOLOv8-SegFusionJudge. This model uses an improved detection and segmentation network as the perception basis, and couples skeletal key point recognition and 3D trajectory reconstruction as the central axis of analysis, aiming to simulate the diagnostic logic of professional coaches. By integrating the decision-making layer in multiple dimensions, the model establishes a deep semantic correlation between technical action deviation and tactical effect attenuation. Ultimately, it achieves a leap from heterogeneous data characteristics to integrated intelligent judgment, providing athletes with structured guidance that includes causal analysis and executable strategies. Its architecture is specifically shown in Figure 7.

Figure 7 systematically shows the complete topology of the YOLOv8-SegFusionJudge model from multi-modal perception to intelligent decision-making. The integration decision-making layer on the right constitutes the core logical hub elaborated in this section. For the heterogeneous data output from the previous chapters, this layer introduces an attention fusion network to establish a deep semantic bridge between technical actions and tactical effects. Specifically, the fusion network first maps action normative vectors and tactical effectiveness vectors to a unified latent semantic space through feature embedding. Subsequently, it uses multi-head AM to dynamically solve the interaction weights between different modal features, thereby quantifying and mining the intrinsic causal relationship of biomechanical deviations-induced tactical quality attenuation. Based on this high-order fusion representation, the model finally outputs a comprehensive competitive performance score and a structured association analysis report in parallel. The report closely follows the logic of generation from deviation location to effect tracing to intervention strategies. The system can not only accurately locate technical shortcomings such as insufficient extension of the hitting elbow joint, but also explicitly deduce its neg-

Figure 7
YOLOv8-SegFusionJudge structure.



ative impact on actual combat consequences such as reduced return depth and speed, and match it to generate a prescription guidance program including specialized strength training and landing point control exercises. The generation of the diagnostic chain from "deviation positioning" to "efficiency degradation" and then to "intervention strategy" mainly includes three steps. Firstly, the deviation localization originates from the comparison results between the joint point kinematic deviation output by the action standardization evaluation submodule and the standard template. Secondly, effectiveness degradation is quantified through the cross-modal interaction weights of action and tactical features in the attention fusion network. When a specific action deviation node has significantly higher attention weights for predicting tactical scores, it is considered that there is a statistical correlation between the deviation and effectiveness degradation. Thirdly, intervention strategies are matched and recommended based on the predetermined biomechanical rule library and tactical knowledge graph. This process completely realizes the transition of the evaluation system from single-dimensional data description to interpretable and executable intelligent decision-making.

4. Results and Discussion

4.1. Experiment Setup and Dataset Construction

4.1.1. Experimental Environment and Parameter Configuration

To ensure the stability of improved model training and inference, all experiments were completed on a high-performance computing workstation. The hardware environment was equipped with AMD Ryzen Threadripper 7970X processor and NVIDIA Ge-

Force RTX 4090 (24GB) graphics card. The software environment was based on the Windows 11 operating system, using the PyTorch2.2 deep learning framework and Python 3.11 language to build the core algorithm library. In the training phase, to verify the convergence performance of the improved YOLOv8 (introducing the LAFF module) and DBAG-SN, the experiment uniformly used the AdamW optimizer for gradient descent, and introduced the cosine annealing strategy to dynamically adjust the learning rate to avoid the model falling into the local optimal solution. Table 1 is the specific key hyperparameters.

Table 1
Experimental parameter settings.

Parameter name	Parameter setting
Learning rate	1e-4
Weight decay	0.0005
Batch size	16
Epochs	300
Input resolution	640×640(Det) / 512 ×512(Seg)
Momentum	0.937

The relevant parameters were derived from image characteristics for the Shuttlecock-Metric dedicated data set. Table 1 shows the core hyperparameter configuration determined to achieve the best convergence effect and performance control during the model training and optimization process.

4.1.2. Dataset Construction

Existing public data lacked multi-view synchronization and fine instance segmentation annotation. To comprehensively evaluate the performance in the three dimensions of visual perception, 3D reconstruction, and intelligent judgment, this study constructed a composite verification set including public benchmarks and self-built multi-modal data. Firstly, this study introduced TrackNet V2 and ShuttleSet as auxiliary validation benchmarks for pre training weight initialization in the early stages of model training and auxiliary validation during the training process [12, 27]. On this basis, this study used multiple synchronous high-speed cameras in a professional badminton training hall to form an acquisition array to collect core data, and divided it into three dedicated subsets to correspond to different verification tasks. Subset A (Subset-DetSeg) served as the detection and segmentation benchmark and contained 8,000 selected key frames covering different lighting, viewing angles and motion blur levels. This subset used the LabelMe tool for manual fine labeling, and adopted a cross-quality control mechanism of two-person blind labeling and one-person review to ensure the true value accuracy of the shuttlecock bounding box and the pixel-level mask of human limbs. Subset B (Subset-3DTraj) was used to verify the accuracy of 3D trajectory reconstruction and contains 50 synchronized video segments of complete batting rounds. This subset used the standard geometric dimensions of the site (such as net height and edge distance) as physical constraints, and combined the 3D coordinate points calculated by multi-view manual triangulation as physical truth values. Subset C (Subset-Eval) was used as an intelligent judgment test set, including 200 typical tactical rounds. To ensure the objectivity and authority of the evaluation, this subset defined action standards according to the Badminton World Federation Technical Manual, and was independently marked by three professionals with senior coaching qualifications. Finally, the intersection of three people's annotations was selected as the gold standard to verify the accuracy of model recommendations.

4.2. Perceptual Performance Analysis of Body State Capture and Eagle Eye Vision Analysis

4.2.1. Improved YOLOv8 Detection Performance Evaluation for Motion Capture

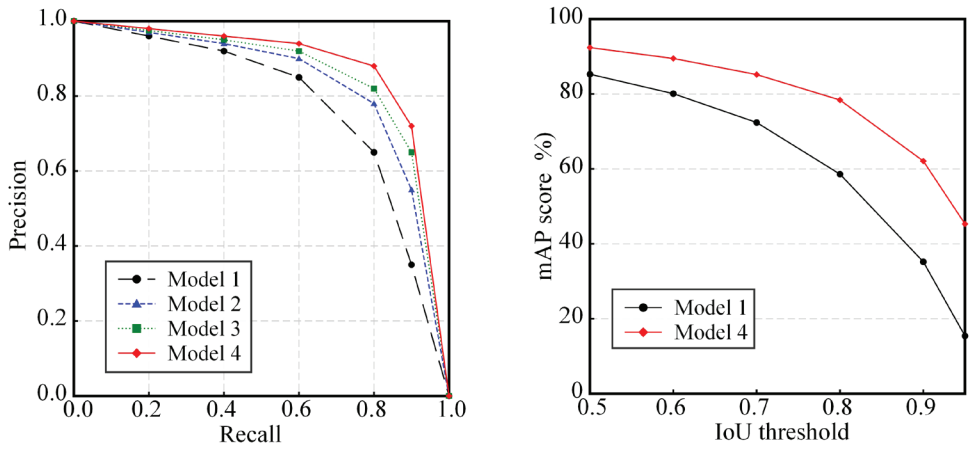
To verify the effectiveness of improving the detection performance of YOLOv8, this study took YOLOv8t as the baseline model and named it Model 1, added the LAFF module as Model 2, added only the loss function as Model 3, and improved the complete model of YOLOv8 as Model 4. And adopted YOLOv9-C, YOLOv10-S, and RT-DETR-R18 as baseline models. Table 2 shows the performance comparison of different models on the self-built test set.

In Table 2, in the object detection task, Model 4 achieved the best performance in mAP@0.5 (92.4%) and recall rate (89.6%), achieving the optimal overall detection trade-off. Although the Precision of Mod-

Table 2
Performance of object detection and image segmentation models.

Model ID	Precision (%)	Recall (%)	mAP@0.5 (%)	mAP@ 0.5:0.95 (%)	FPS (Hz)	Params (M)	GFLOPs	Model Size (MB)
Model 1	88.4	82.1	85.3	64.2	145	3.2	1.7	13.1
Model 2	89.1	86.5	88.7	66.8	132	3.5	1.9	14.2
Model 3	92.3	83.4	87.9	69.1	142	3.2	1.7	13.1
Model 4	91.8	89.6	92.4	71.5	128	3.6	2.0	14.7
YOLOv9-C	89.2	84.7	87.6	66.3	118	5.1	2.2	20.4
YOLOv10-S,	90.5	85.9	88.9	68.2	135	4.2	1.8	16.8
RT-DETR-R18	88.7	83.5	86.4	65.7	96	8.7	3.5	34.8

Figure 8
Analysis of improved YOLOv8 detection performance.



(a) P-R curve comparison

(b) mAP performance under different IoU thresholds

el 4 (91.8%) was slightly lower than that of Model 3 (92.3%), and the FPS (128 Hz) was slightly lower than that of Model 1 (145 Hz), it was compensated by a higher recall rate and positioning accuracy, so the overall performance remained the best. To further explore the robustness of the improved detection model under different thresholds, the detailed performance analysis curve of the improved YOLOv8 is shown in Figure 8.

In Figure 8(a), the area enclosed by the P-R curve of Model 4 was significantly larger than that of the baseline model, and it could still maintain a high precision rate especially in the high recall interval, indicating that the model's missed detection rate was extremely low. Figure 8(b) shows that as the IoU threshold increased from 0.5 to 0.95, the mAP of all models showed a downward trend, but the decline rate of Model 4 was significantly slower than that of the baseline model. Specifically, under the high threshold of IoU=0.8, Model 4 still maintained 78.4% mAP, while the baseline model has dropped to 58.6%, and the gap between the two widened to 19.8%. The

LAFF module effectively enhanced the characteristic response of small targets, while the shape-aware loss significantly improved the positioning stability under high IoU thresholds, making the model more adaptable to the stringent accuracy requirements of competitive sports.

4.2.2. Accuracy Verification of DBAG-SN Segmentation Model Supporting Eagle Eye Analysis

To verify the efficiency of DBAG-SN in processing small shuttlecock targets, this study introduced other image segmentation models for comparison. Mask R-CNN (Model A) added a target mask prediction branch in parallel to the architecture, which had the characteristics of taking into account target detection accuracy and pixel-level segmentation quality. It was often used as a strong benchmark model for instance segmentation tasks [3]. U-Net (Model B) was a semantic segmentation model based on a fully convolutional architecture. Through a symmetric encoder-decoder structure and a skip connection mechanism, it had the ability to achieve fine edge positioning under small sample conditions [6, 19]. DBAG-SN was Model C. The convergence characteristics and fine-grained category performance of different segmentation networks during the training process are shown in Figure 9.

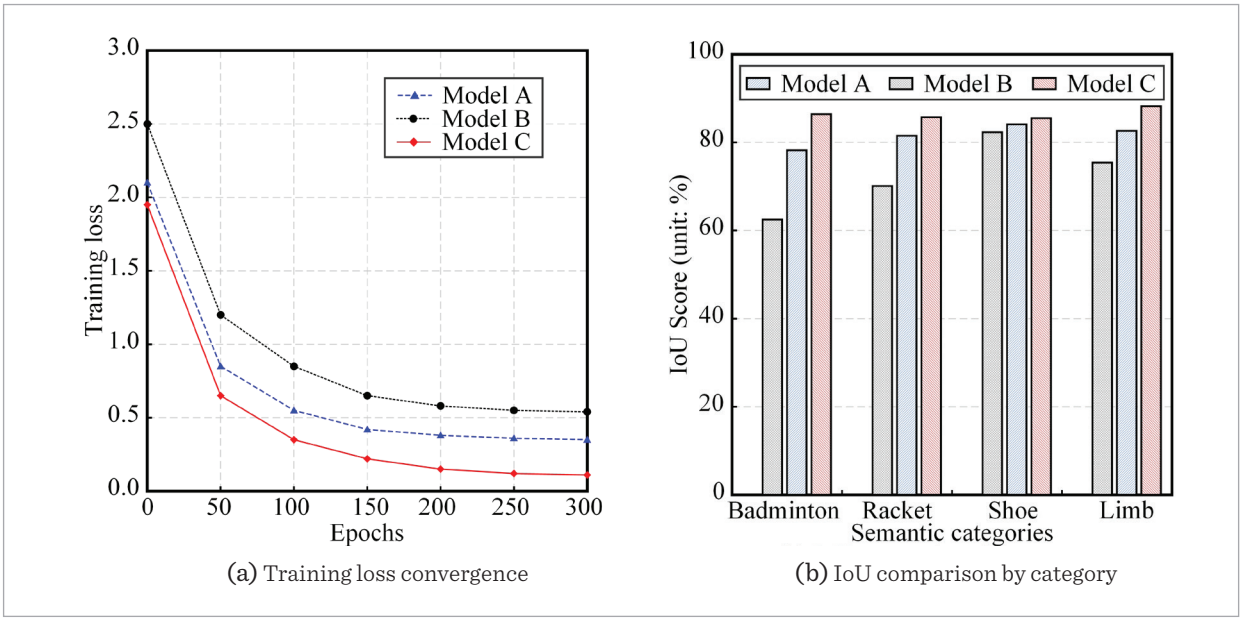
Figure 9(a) shows that Model C showed extremely fast convergence speed in the early stages of training. Its loss value dropped to around 0.22 around the 150th round and stabilizes, which was significantly lower than the 0.65 of U-Net and 0.42 of Mask R-CNN in the same period, indicating that dual-branch AM effectively accelerated feature fitting. The category IoU comparison in Figure 9(b) shows that DBAG-SN had significant advantages in the two extremely difficult categories of "shuttlecock" and "limbs", with IoU reaching 86.4% and 88.2%, which were 23.9% and 12.8% higher than U-Net. In Figures 9(a)-(b), DBAG-SN not only had efficient learning capabilities but also could effectively solve the problem of difficult segmentation of small targets and slender structures through task-aware attention allocation.

4.3. 3D Reconstruction and Motion Parameter Calculation Accuracy

To quantitatively evaluate the solving capabilities of the multi-view geometric reconstruction algorithm in real physical space, the error distribution characteristics of the research method and the traditional binocular vision method in 3D position and landing point prediction were first analyzed, as shown in Figure 10.

Figure 9

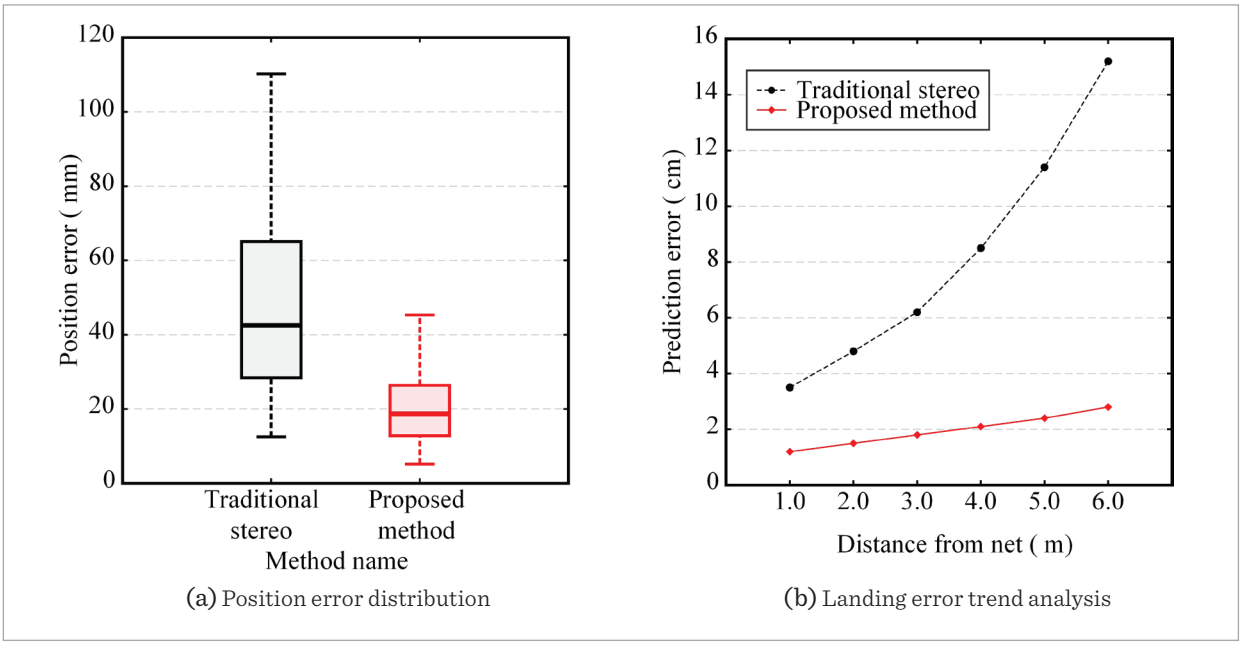
Performance analysis of DBAG-SN segmentation network.



(a) Training loss convergence

(b) IoU comparison by category

Figure 10
Error distribution and comparison of 3D trajectory reconstruction.



(a) Position error distribution

(b) Landing error trend analysis

Table 3
Accuracy verification of 3D trajectory reconstruction and intelligent assessment.

Method	3D Trajectory Reconstruction	Intelligent Assessment				
	RMSEpos (mm)	RMSEvel (m/s)	Landing Error (mm)	Pearson's r	Consistency (%)	IN/OUT Acc (%)
Traditional Stereo	42.5	1.85	68.4	0.76	78.2	88.5
Proposed Method	18.7	0.42	25.6	0.91	94.5	98.2

The boxplot in Figure 10(a) clearly showed the dispersion of the error distribution. The median error of the traditional binocular vision method was as high as 42.5 mm and the box was wide, indicating its poor stability. The median error of the research method was significantly reduced to 18.7 mm, and the interquartile range was extremely narrow, indicating that multi-viewing constraints effectively suppressed outliers. Figure 10(b) further revealed the variation pattern of landing point prediction error with flight distance: As the distance from the grid increased, the error of the traditional method showed a divergent trend, and the error reached 15.2 cm at the bottom line. Research method: Even at the far bottom line, the error was still stably controlled within 2.8 cm. The reconstruction algorithm based on multi-view geometric constraints had millimeter-level positioning accuracy and strong spatial robustness in the entire field. Table 3 shows the consistency verification results of 3D trajectory reconstruction accuracy and intelligent evaluation.

In Table 3, the RMSEpos of the research method in 3D trajectory reconstruction was only 18.7 mm, which was 56% lower than the traditional method, and the speed calculation error was also controlled at an extremely low level of 0.42 m/s. In the intelligent judgment dimension, the Pearson correlation coefficient between the proposed model and expert scores was as high as 0.91, which was higher than

the 0.76 in the control group. The accuracy of IN/OUT determination reached 98.2%. This showed that the system not only accurately calculated physical parameters, but also derived judgment logic that was highly aligned with the cognition of human experts.

4.4. Comprehensive Performance Evaluation of YOLOv8-SegFusionJudge Intelligent Fusion Judgment Model

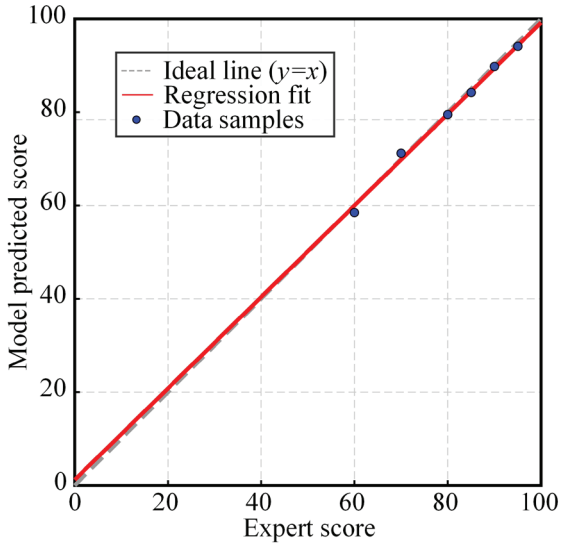
To verify the objectivity and authority of the intelligent evaluation system, this study first tested the consistency of the model scoring and the expert scoring through regression analysis and Bland-Altman method, as shown in Figure 11.

Figure 11(a) shows that the model prediction score was highly positively correlated with the expert score, the slope of the fitted straight line was close to 1, and the coefficient of determination R2 was as high as 0.925, indicating that the model could accurately reproduce the expert's evaluation criteria. The Bland-Altman plot in Figure 11(b) further verified the consistency between the two. The average score difference was only +0.42 points, and the vast majority of sample points fell within the 95% consistency limit (±1.96 SD). The YOLOv8-SegFusionJudge model had the ability to replace manual labor for objective and unbiased tactical judgment without significant systematic errors. Finally, to verify the necessity of the multi-dimensional fusion decision-making mechanism and its anti-interference ability, the ablation experiment and robustness analysis are shown in Figure 12.

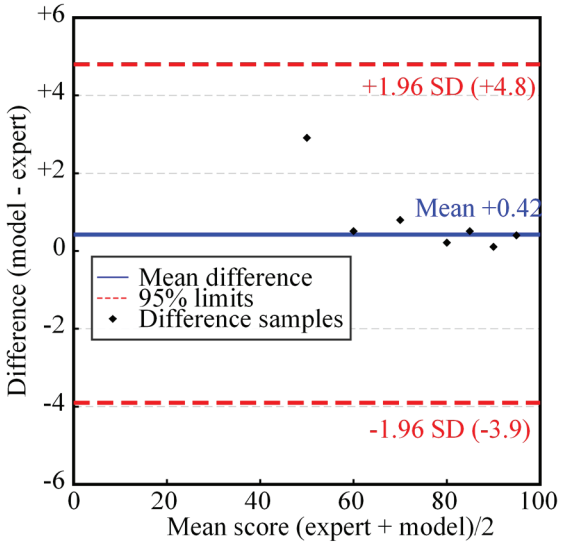
Figure 12(a) shows that there were limitations in single-dimensional evaluation. The F1-Score of the model that only relied on action or tactical features was 72.9% and 67.8%. The dual-stream fusion model significantly improved the F1-Score to 85.5%, proving the importance of "technical-tactical" causal correlation modeling. Figure 12(b) was the performance degradation curve of the model under different Gaussian noise levels. As the noise intensity increased from 0 to 0.5, the performance of the single-stream model dropped sharply to 42.6%, while the fusion model can still maintained a high availability of 74.8%. Feature fusion based on AM not only greatly improved the comprehensiveness of the evaluation but also gave the system strong robustness in complex and strong interference environments.

Figure 11

Correlation and consistency analysis of intelligent assessment scores.

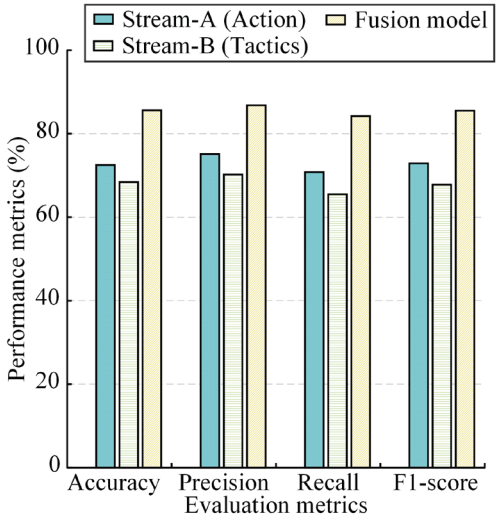


(a) Regression analysis

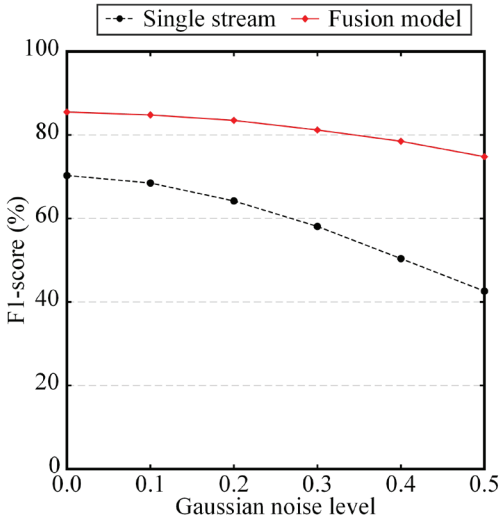


(b) Bland-altman consistency plot

Figure 12
Ablation study and robustness analysis of fusion decision model.



(a) Ablation study: single vs. dual stream



(b) Robustness analysis under noise

5. Discussion and Interpretation

This study proposed a YOLOv8-SegFusionJudge model. At the system architecture design level, this study organized the four modules of detection, segmentation, 3D reconstruction, and evaluation into a cascaded pipeline. This design choice had its theoretical rationality. Specifically, the above four types of tasks applied to different target spaces, with detection and segmentation losses defined in the two-dimensional image domain, three-dimensional reconstruction relying on multi view geometric constraints, and evaluation oriented towards temporal decision fusion. Although joint end-to-end training was theoretically attractive, it required simultaneous processing of gradient signals from 2D vision, 3D geometry, and high-level semantics in this scenario. The significant differences in magnitude and convergence rate of different task loss functions might lead to gradient conflicts and unstable optimization. Therefore, this article adopted a decoupled cascade paradigm to ensure that each module fully converges under its optimal independent hyperparameter configuration, and then achieved collaborative utilization of multi-source information through feature level fusion decision layer, providing a modular benchmark framework for exploring end-to-end joint optimization in the future. At the visual perception level, the improved model relied on the LAFF module and shape perception loss to increase mAP@0.5 to 92.4%, an increase of 7.1% compared with the baseline model. This was due to the LAFF module adaptively aligning shallow textures with deep semantics at the lateral connections of the feature pyramid, thereby prioritizing the enhancement of feature response for high-speed micro targets. At the decision-making application level, 3D reconstruction based on multi-view geometric constraints reduced the position error to 18.7 mm, and the Pearson correlation coefficient of the fusion of the judgment model and expert scoring was as high as 0.91. C. Wang et al. proposed a trajectory recognition algorithm based on M-YOLOv2 and Kalman filtering, which achieved a recognition accuracy of up to 98.6% in complex environments. However, this method mainly focused on target-level detection and trajectory prediction of the 2D plane. Different from this, this study expanded the analysis dimension from a simple "object (shuttlecock)" to a 3D space of "human-ball interaction" by introducing instance segmentation and 3D skeleton key point technology.

To sum up, the YOLOv8-SegFusionJudge model successfully achieved quantitative evaluation of badminton performance through high-precision visual perception and intelligent fusion decision-making.

6. Conclusion and Future Work

Given the problems existing in traditional badminton training and game analysis, such as strong subjectivity of judgment and separation of correlation analysis between action standardization and tactical effectiveness, this study designed a YOLOv8-SegFusionJudge model. It first optimized the visual perception capabilities by introducing the LAFF module and shape perception loss, and then relied on bone key point detection and action recognition technology to achieve refined analysis of the entire process of technical movements such as racket introduction, swing, hitting, and follow-up swing. On this basis, combined with DBAG-SN and multi-view geometric constraints, an integrated evaluation framework from multi-modal perception to "technical-tactical" causal decision-making was constructed. In the experiment, the improved detection model achieved a mAP@0.5 of 92.4% on the self-built test set, which was 7.1% higher than the baseline model. DBAG-SN was significantly better than mainstream algorithms such as Mask R-CNN with an MIoU of 85.2%, effectively solving the problem of accurate analysis of high-speed small targets. At the physical reconstruction and decision-making level, the system reduced the 3D trajectory reconstruction error to 18.7 mm, which was about 56% more accurate than the traditional binocular vision method. The Pearson correlation coefficient between intelligent evaluation results and expert scores was as high as 0.91, and the dual-stream fusion mechanism increased the F1-Score of the evaluation to 85.5%. In summary, the YOLOv8 SegFusionJudge model achieved high-precision quantification in the detection, segmentation, and 3D trajectory reconstruction modules, while the complete fusion evaluation of action standardization and tactical effectiveness, as well as tactical deep reasoning, was still in the conceptual verification stage and will continue to be improved in future work. However, current research is mainly based on data collected in controlled environments, and the robustness under extreme lighting changes and extremely crowded scenes still needs to be improved. Future work will be dedicated to building a larger-scale multi-scene competitive dataset and exploring the real-time deployment application of lightweight models on edge devices.

References

1. Adam, S., Fitri, N. A., Bibi, S., Sufandi, M. R. Real-Time Arrow Detection and Scoring on Archery Targets Using YOLOv8 with Euclidean Distance-Based Zone Estimation. Journal of Applied Informatics and Computing, 2025, 9(6), 3669-3680. https://doi.org/10.30871/jaic.v9i6.11086
2. Afifah, V., Erniwati, S. YOLOv8 for Object Detection: A Comprehensive Review of Advances, Techniques, and Applications. International Journal of Advanced Computing and Informatics, 2026, 2(1), 53-61. https://doi.org/10.71129/ijaci.v2i1.pp53-61
3. Bi, X., Hu, J., Xiao, B., Li, W., Gao, X. IEMask R-CNN: Information-Enhanced Mask R-CNN. IEEE Transactions on Big Data, 2023, 9(2), 688-700. https://doi.org/10.1109/TBDATA.2022.3187413
4. Chang, K.-S., Wang, W.-Y., Peng, W.-C. Where Will Players Move Next? Dynamic Graphs and Hierarchical Fusion for Movement Forecasting in Badminton. Proceedings of the AAAI Conference on Artificial Intelligence, 2023, 37(6), 6998-7005. https://doi.org/10.1609/aaai.v37i6.25855
5. Chang, Y.-T. A Method of Badminton Video Motion Recognition Based on Adaptive Enhanced AdaBoost Algorithm. International Journal of Biometrics, 2025, 17(1-2), 86-101. https://doi.org/10.1504/IJBM.2025.143715
6. Chen, H., He, Y., Zhang, L., Yao, S., Yang, W., Fang, Y., Liu, Y., Gao, B. A Landslide Extraction Method of Channel Attention Mechanism U-Net Network Based on Sentinel-2A Remote Sensing Images. International Journal of Digital Earth, 2023, 16(1), 552-577. https://doi.org/10.1080/17538947.2023.2177359
7. Du, C., Jin, R., Tang, H., Jiang, Q., He, Z. SKIP: Accurate Fall Detection Based on Skeleton Keypoint As-

sociation and Critical Feature Perception. IEEE Sensors Journal, 2024, 24(9), 14812-14824. https://doi.org/10.1109/JSEN.2024.3379167

8. Duan, K., Bai, S., Xie, L., Qi, H., Huang, Q., Tian, Q. CenterNet++ for Object Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(5), 3509-3521. https://doi.org/10.1109/TPAMI.2023.3342120
9. Farooq, J., Muaz, M., Khan Jadoon, K., Aafaq, N., Khan, M. K. An Improved YOLOv8 for Foreign Object Debris Detection with Optimized Architecture for Small Objects. Multimedia Tools and Applications, 2024, 83(21), 60921-60947. https://doi.org/10.1007/s11042-023-17838-w
10. Haq, M. A., Tagawa, N. Improving Badminton Player Detection Using YOLOv3 with Different Training Heuristic. JOIV: International Journal on Informatics Visualization, 2023, 7(2), 548-554. https://doi.org/10.30630/joiv.7.2.1166
11. He, J., He, J. Research and Implementation of Dynamic Path Tracking Method for Badminton Based on Multi-Feature Fusion. Discover Applied Sciences, 2025, 5(1), 135-150. https://doi.org/10.1007/s44163-025-00383-w
12. Hsu, Y.-H., Yu, C.-C., Cheng, H.-Y. Enhancing Badminton Game Analysis: An Approach to Shot Refinement via a Fusion of Shuttlecock Tracking and Hit Detection from Monocular Camera. Sensors, 2024, 24(13), 4372. https://doi.org/10.3390/s24134372
13. Inturi, A. R., Manikandan, V. M., Garrapally, V. A Novel Vision-Based Fall Detection Scheme Using Keypoints of Human Skeleton with Long Short-Term Memory Network. Arabian Journal for Science and Engineering, 2023, 48(2), 1143-1155. https://doi.org/10.1007/s13369-022-06684-x
14. Jiang, X. Identification of Badminton Players' Swinging Movements Based on Improved Dense Trajectory Algorithm. International Journal of Web Engineering and Technology, 2024, 19(3), 310-329. https://doi.org/10.1504/IJWET.2024.142213
15. Li, H., Chen, J., Xia, H. Multiperson Activity Recognition and Tracking Based on Skeletal Keypoint Detection. IEEE Transactions on Artificial Intelligence, 2024, 5(5), 2279-2292. https://doi.org/10.1109/TAI.2023.3318575
16. Lianju, L., Haiying, Z. Research on Badminton Take-Off Recognition Method Based on Improved Deep Learning. Journal of Ambient Intelligence and Humanized Computing, 2024, 15(9), 3317-3330. https://doi.org/10.1007/s12652-024-04809-8
17. Lin, Y. Lightweight CA-YOLOv7-Based Badminton Stroke Recognition: A Real-Time and Accurate Behavior Analysis Method. International Journal of Advanced Computer Science and Applications, 2025, 16(2), 324-331. https://doi.org/10.14569/IJACSA.2025.0160235
18. Nugroho, S., Komaini, A., Arga, M. W., Ihsan, F., Hidayat, R. A., Raharjo, S., Rusdiana, A. Motion-Based Technology Uses Sensors and the Internet of Things to Support Badminton Motor Skills Examination. Retos, 2025, 67, 173-181. https://doi.org/10.47197/retos.v67.109852
19. Ouyang, J., Liu, S., Peng, H., Garg, H., Thanh, D. N. LEA U-Net: A U-Net-Based Deep Learning Framework with Local Feature Enhancement and Attention for Retinal Vessel Segmentation. Complex & Intelligent Systems, 2023, 9(6), 6753-6766. https://doi.org/10.1007/s40747-023-01095-3
20. Özer, Ö., Soslu, R., Uyhan, O., Devrilmez, M., Devrilmez, E. Specific Change of Direction Test for Badminton Players: Badminton Shuttle Run Agility Test. Physical Education Studies, 2025, 29(6), 457-465. https://doi.org/10.15561/20755279.2025.0603
21. Perkasa, M. A., El Akbar, R. R., Al Husaini, M., Rizal, R. Visual Entity Object Detection System in Soccer Matches Based on Various YOLO Architecture. Jurnal Teknik Informatika, 2024, 5(3), 811-820. https://doi.org/10.52436/1.jutif.2024.5.3.2015
22. Purnama, B., Erfianto, B., Wirawan, I. R. Time Series Classification of Badminton Pose Using LSTM with Landmark Tracking. Journal of Electronics, Electromedical Engineering, and Medical Informatics, 2025, 7(1), 27-37. https://doi.org/10.35882/jeeemi.v7i1.488
23. Shaker, A., Maaz, M., Rasheed, H., Khan, S., Yang, M. H., Khan, F. S. UNETR++: Delving into Efficient and Accurate 3D Medical Image Segmentation. IEEE Transactions on Medical Imaging, 2024, 43(9), 3377-3390. https://doi.org/10.1109/TMI.2024.3398728
24. Sheng, Y., Liu, C., Yi, Q., Ouyang, W., Wang, R., Chen, P. Predicting Badminton Outcomes Through Machine Learning and Technical Action Frequencies. Scientific Reports, 2025, 15(1), 10575-10582. https://doi.org/10.1038/s41598-025-87610-7

25. Shih, C.-Y., Lin, Y.-T., Chen, W., Huang, J.-C. SVM-AdaBoost Based Badminton Offensive Movement Parsing Technique. Signal, Image and Video Processing, 2025, 19(4), 280-293. https://doi.org/10.1007/s11760-025-03865-7

26. Tang, X. Application Simulation Research Based on Visual Image Capture Technology in Sports Injury Rehabilitation. Information Technology and Control, 2024, 53(2), 429-441. https://doi.org/10.5755/j01.itc.53.2.35021

27. Wang, W.-Y., Huang, Y.-C., Ik, T. U., Peng, W.-C. ShuttleSet: A Human-Annotated Stroke-Level Singles Dataset for Badminton Tactical Analysis. Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2023, 5126-5136. https://doi.org/10.1145/3580305.3599906

28. Wu, Y., Chen, Z., Zhang, H., Yang, Y., Yi, W. Enhanced Pose Estimation for Badminton Players via Improved YOLOv8-Pose with Efficient Local Attention. Sensors, 2025, 25(14), 4446. https://doi.org/10.3390/s25144446

29. Zhi, J., Sun, Z., Zhang, R., Zhao, Z. Badminton Video Action Recognition Based on Time Network. Journal of Computer Methods in Sciences and Engineering, 2023, 23(5), 2739-2752. https://doi.org/10.3233/JCM-226889

30. Zhou, Z., Zhang, H., Effatparvar, M. Improved Sports Image Classification Using Deep Neural Network and Novel Tuna Swarm Optimization. Scientific Reports, 2024, 14(1), 14121-14140. https://doi.org/10.1038/s41598-024-64826-7

This article is an Open Access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 (CC BY 4.0) License (http://creativecommons.org/licenses/by/4.0/).