



Research on tennis serve drop prediction model based on attention mechanism

Sheng Liu¹, Yaxi Liu¹, Xingying Liu¹, Jiahao Hong¹ and Wenying Tian^{2,*}

¹ Department of Physical Education, Jingdezhen Ceramic University, Jingdezhen, Jiangxi, 333403, China

² Department of Physical Education and Health, Dasheng Primary School, Ganzhou, Jiangxi, 342500, China

SUMMARY: *When the current method predicts the landing point of tennis serve, it does not analyze the force of tennis ball in the process of movement, which leads to the problem of low accuracy of the prediction results. Based on this, this study adopts the YOLOv4 motion target detection algorithm incorporating the attention mechanism to realize tennis ball target detection. At the same time, a three-dimensional coordinate reconstruction method of the tennis ball is designed, and a kinetic model of the tennis ball is established based on the force analysis, and a continuous model of the tennis ball kinematics is obtained by solving the kinetic differential equations, which is used for predicting the trajectory and calculating the landing point. The experimental results show that the incorporation of the attention mechanism improves the detection accuracy of the target detection algorithm. The prediction of the landing point position by the landing point prediction model designed in this paper is not much different from the actual landing point position. It shows that the model can be put into the daily training of tennis players to help them explore the technical and tactical laws of tennis matches, assist people in making technical and tactical decisions, and improve the scientificity of decision-making.*

KEYWORDS: *Attention mechanism; YOLOv4; force analysis; serve drop prediction*

1 Introduction

The serve has occupied an important position in the tennis game since the beginning, and it is also an important technique controlled by the player himself and not restricted by the opponent [1]. With the rapid development of tennis technology, the tennis player's serve is more powerful, faster, and the angle of landing point is more tricky, and the serve brings more aggressiveness and oppression to the opponent [2]. Direct service points are more common, receiving becomes more and more difficult, the serving side dominates in the service game, and it is more and more difficult to break the serve. In modern tennis, both athletes and coaches pay more attention to the diversification and precision of serving techniques and tactics, and the speed, angle, rotation and landing point of the serve have an important impact on the serve, and whether it is the speed, angle, or rotation of the serve, it is closely related to the final landing point of the serve, and the landing point of the serve on the key points is particularly important [3-5].

In the current era, the rapid development of the Internet has led to an explosive growth in the number of tennis match videos, and network video software has gradually become a popular

*15779794118@168.com

<https://doi.org/10.65102/is2026649>

application for all people [6]. With the increase of players' participation opportunities, the increasingly abundant match videos make the data accumulation of techniques and tactics in the match increasing, how to explore the characteristics of excellent players' techniques and tactics from the abundant data, and provide a basis for the scientific decision-making of coaches and athletes has become one of the focuses and urgent problems in the theoretical research of tennis [7-9]. Machine learning techniques, represented by deep learning and attention mechanisms, provide excellent solutions for the task of athlete movement recognition [10, 11]. Among them, deep learning uses an end-to-end learning mechanism to automatically extract meaningful features from video data, and after training with large-scale data, it can effectively recognize character movements and scene changes in videos, automatically capture spatio-temporal features and process complex dynamic content [12, 13]. The performance of deep learning in video analytics far exceeds the traditional manual feature extraction methods and dramatically improves the efficiency and accuracy of action recognition. Using deep learning technology and attention mechanism to analyze the technical and tactical features and laws in athletes' matches can timely and accurately obtain and discover more valuable information and laws to make up for the shortcomings of the traditional statistical means, and further learn the historical data of tennis matches to predict the tennis serve landing point through classification models [14, 15]. Therefore, it is necessary for tennis players to adopt intelligent technology to assist them in making technical and tactical decisions.

As far as the sports field is concerned, numerical intelligence tools have been widely used in it. For example, Taha et al. [16] used classification algorithms for the first time in archery, analyzed the clustering of physiological and biochemical indicators for archery athletes of high and low skill levels, and successfully completed the experimental process of predicting archery potential from the combination of indicators by adopting the classification prediction method, and the results showed that the classification accuracy was as high as 97.5%. Chen, W et al [17] combined five data mining techniques such as extreme learning machine, multivariate adaptive regression curve, K nearest neighbor algorithm, extreme gradient boosting XGBoost and stochastic gradient boosting to construct a hybrid basketball game result prediction scheme for predicting the final score of an NBA game, and the empirical results show that the prediction model has high performance. Fruitful research results have also been achieved in tennis. For example, Niu et al [18] found that the method of analyzing tennis techniques and tactics based on mathematical statistics has problems such as strong subjectivity, amount of data up to, and slow analysis speed, for this reason, they proposed the method of data mining to analyze, and proved the feasibility of the method on tennis techniques and tactics analysis in real cases. Chen, H [19] established a mathematical model of tennis player's training movements and analyzed the player's three-dimensional data information base based on the number-wise algorithm, based on which an assessment model of tennis player's training movements was developed, which improved the tennis player's tactical level. Zhang, S et al [20] proposed an intelligent framework to analyze the physical fitness of tennis players, the framework embedded the XGBoost algorithm to achieve the micro-architecture event importance ranking at the same time reduces the big data dimensions of the tennis players, the main function of the framework is to monitor the physical condition of the athletes, technical, tactical skills and intelligence, the framework has been tested to improve the physical fitness of the athletes. Zhang, J et al [21] collected data from the men's singles final at Wimbledon in 2023 in order to predict the real-time scoring ability of tennis players and developed a player scoring prediction model by using a principal component analysis stepwise regression model, which achieved a scoring prediction accuracy of 95.27% in women's tennis matches. Wang et al [22] developed an intelligent system for decision support of tennis skills and tactics by using numerical intelligence tools. The system acquired the sports characteristics of tennis players and realized the quantitative analysis

of skills and tactics through tennis data mining, object detection and tracking techniques.

In terms of serving, Kaliffe et al [23] crawled some of the data from the Australian Open and Roland Garros tennis tournaments and used K-Means and Gaussian mixture models to cluster analyze the tennis serving skills in order to find the best serving strategies in tennis. Wei et al [24] proposed a tennis serve type prediction model based on a stylistic prior, which mines the potential association between a tennis player's serve motion and serve type and quantifies the similarity between players, and has good prediction performance on the Hawk-Eye dataset. In addition, Mecheri et al [25] collected three-dimensional kinematic data of tennis matches recorded by the Hawk-Eye system during 2003-2008 and analyzed tennis serve trajectories, serve characteristics, and scoring probabilities, and explored the effects of serve speed, serve location, court surface, gender differences, and rotational intensity on the probability of scoring in order to determine the optimal serving strategy.

As an emerging technology, machine learning has become a research hotspot in the field of tennis. Sampaio et al [26] reviewed the literature related to the application of machine learning in tennis, and concluded that machine learning has obvious advantages in the monitoring of athletes' psychological state, talent identification, match result prediction, spatial and tactical analysis, and injury prevention, which provides coaches with better tennis training programs. Huang et al [27] integrated the use of machine learning, data mining and mathematical statistics a variety of methods, aiming to mine the effective data related to the formulation of tennis tactics from the video of tennis matches, the research constructed a model has a certain value of application in the actual game, and improve the accuracy of the technical analysis of tennis matches. Zhu et al [28] designed a machine learning algorithm to predict the direction of tennis player's serve to improve the player's serve, the algorithm used had 49% and 44% accuracy in predicting the direction of tennis serve for male and female athletes, respectively, which may be due to the fact that athletes tend to use a mixture of serve strategies during the match and the background information has an effect on the prediction performance. Gao et al [29] used a random forest model to predict the outcome of a tennis match, the model is able to quickly learn and predict the outcome of the match by analyzing the processed tennis match information data with an accuracy of more than 80%, in which the model considers that the tennis serve skill is one of the key factors determining the outcome of the match. Vives et al [30] explored the key factors affecting serve scoring using a machine learning approach in order to improve tennis players' serve scoring rate, and found that a serve angle between 5.7° and 8.7° as well as a lateral bounce distance of the ball off the sideline during the serve of between 0 and 28 centimeters could substantially improve the serve scoring rate. Liu et al [31] used deep learning algorithms to build an intelligent classification model for tennis serve scoring, the model can automatically track and learn the characteristics of tennis player's serve, and then realize the accurate classification of serve position, landing position, and tennis skills, this model is of great help to improve the player's serve scoring rate. Ye et al [32] design tennis serve action recognition system, the system is based on local attention-based convolutional neural network as the driving technology, can accurately predict the tennis ball landing point, ball speed and identify the non-standard action of tennis serve, the system as an auxiliary technology can effectively improve the player's serve action to enhance the serving skills.

The study fully considers that the tennis ball size is too small, and firstly, the YOLOv4 network structure is improved to combine the attention mechanism for tennis ball detection. Then, in order to extract the position of tennis ball, a salient target detection algorithm that can retain the sub-pixel edge information is used to detect the tennis ball significantly, and the identification of the center of the tennis ball based on the least squares method. Then, the tennis ball trajectory is predicted by 3D reconstruction and force analysis of the tennis ball using a continuous motion model, and finally, the location of the landing point is calculated by solving

the kinetic differential equations.

2 Tennis target recognition and tracking based on the attention mechanism

2.1 Tennis ball detection based on improved YOLOv4

2.1.1 Feature map optimization

The 16×16 feature map of the original YOLOv4 is responsible for detecting large targets, while the 32×32 and 64×64 feature maps are responsible for predicting medium-sized and small-sized targets, respectively. Because tennis target recognition and tracking belongs to small targets, the network structure is improved accordingly in this paper. In the neck network part, a quadruple downsampled feature map is added for feature fusion with the three-scale feature maps of the original backbone network. For the detection head part, the 32-fold downsampled feature map possessed too large sensory fields to the extent that it was difficult to regressively predict targets of tennis ball size size, so it was removed. Similarly, in the neck network, a 4-fold downsampled feature map was introduced from the backbone network and added to the neck network to detect tennis balls; the 4-fold downsampled feature map has a larger scale and higher resolution to more clearly render smaller tennis ball sizes. The redirected network structure is represented as a 4-in-3-out neck network structure. The three scales of feature output layers are changed from 16×16 , 32×32 and 64×64 to 32×32 , 64×64 and 128×128 sizes to improve the tennis ball detection accuracy.

2.1.2 Attention mechanism module

The Attention Mechanism Module (AMB) consists of some special convolutional layers. It does not change the size of the feature map, but increases the representation of the target features to improve the detection capability. Therefore, the AMB can be easily added to existing tennis ball detection networks.

Taking the feature map $F \in \mathbb{R}^{H \times W \times C}$ as an input, the AMB sequentially arranges the channel attention feature map $M_C \in \mathbb{R}^{1 \times 1 \times C}$ and the spatial attention feature map $M_S \in \mathbb{R}^{H \times W \times C}$. The process of channel and spatial attention mechanisms can be summarized as follows:

$$\begin{aligned} F' &= F \otimes M_C(F) \\ F'' &= F' \otimes M_S(F') \end{aligned} \quad (1)$$

where \otimes denotes the intelligent product of elements and F'' denotes the final output.

The Channel Attention Module (CAM) in AMB is arranged before the Spatial Attention Module (SAM). The CAM is capable of suppressing background information and highlighting foreground features by adaptively re-weighting channel features. In this paper, the CAM is used to minimize false detections of tennis balls. Specifically, each channel feature map is transformed to size $1 \times 1 \times C$ by average pooling, and the channel attention feature map is obtained by two 1×1 convolutional layers. The entire M_C computation is represented as follows:

$$M_C = \sigma \left(C_{r=1/16}^{1 \times 1} \left(CM_{r=1/16}^{1 \times 1} \left(\text{AvgPool}(F) \right) \right) \right) \quad (2)$$

where $CM_r^{1 \times 1}$ denotes a convolution operation following a Mish activation function, where 1×1 denotes the convolution kernel size and r denotes the rate of decrement. $C_r^{1 \times 1}$ is a convolution operation with the same contextual information as $CM_r^{1 \times 1}$. The σ denotes the sigmoid function. The channel attention mechanism is shown in Fig. 1.

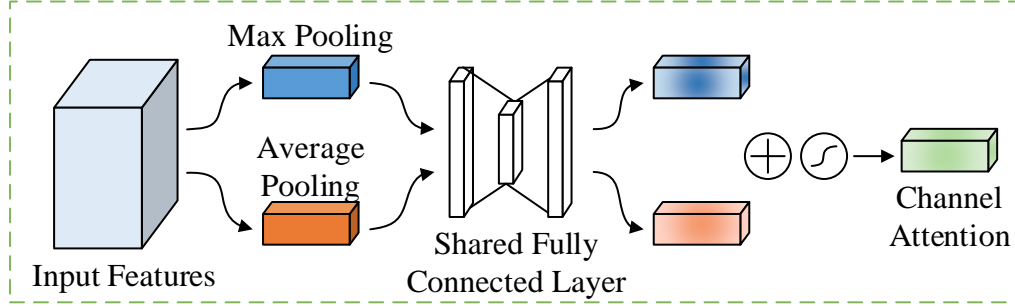


Figure 1: Channel attention mechanism

SAM is concerned with objects on a spatial scale. In general, the foreground takes up much fewer pixels than the background. Therefore, more attention should be given to the foreground region. In this paper, we use a mask of the same size and depth as the input feature map to generate the spatial attention map M_S . Specifically, the mask used to generate spatial attention is generated by the 1×1 convolutional layer. The computational process of M_S can be represented as follows:

$$M_S = \sigma \left(CB_{r=1}^{1 \times 1}(F') \right) \quad (3)$$

where F' denotes the provisional feature map of the CAM output. The $CB_r^{1 \times 1}$ denotes a convolution operation immediately followed by a batch normalization operation, which has the same contextual meaning as $CM_r^{1 \times 1}$. The spatial attention mechanism is shown in Fig. 2.

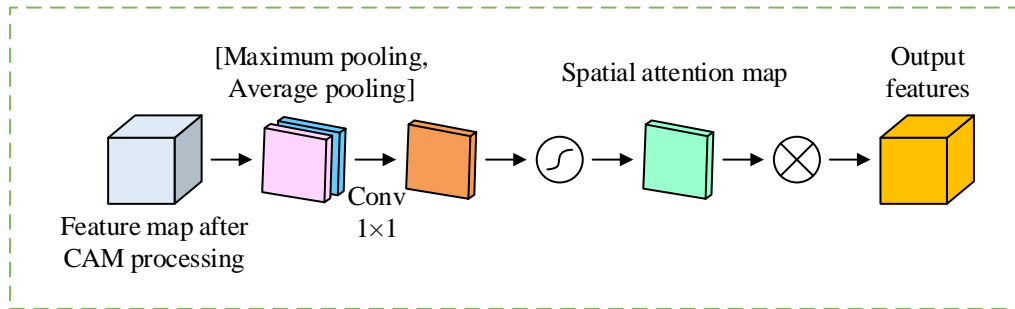


Figure 2: Spatial attention mechanism

2.2 Extraction of tennis ball contours based on BASNet

In order to recognize the center of the tennis ball circle, the image needs to be preprocessed in order to extract the location of the tennis ball. The preprocessing consists of two steps, namely image binarization and circular contour detection. Image binarization is the process of

converting each pixel in the image to black or white in order to distinguish the foreground from the background. Circular edge detection is to find the edge points in the binarized image that meet the conditions of a circle in order to determine the radius and center of the tennis ball. Binarization is performed by converting gray scale image pixels to 0 or 1 using an adaptive local thresholding method. However, the detected rectangular region usually contains background noise other than that of the tennis ball, which will affect the subsequent localization of the position of the center of the tennis ball. In this paper, a saliency detection algorithm that can focus on the edges is considered for saliency detection of tennis balls, eliminating the background noise while preserving the information of round subpixel boundaries. In this paper, a saliency target detection algorithm BASNet that preserves sub-pixel edge information is used for saliency detection of tennis balls. Using this method for binarization of the tennis ball region, compared with the adaptive thresholding method, it can effectively filter out the background noise while retaining the sub-pixel edge information of the tennis ball. Circle contour detection is performed on the resulting saliency map binarized image, and the least squares method is used to fit the circle contour and determine the coordinates of the tennis ball circle center.

The network of BASNet consists of two main modules: a prediction module (PM) and a residual refinement module (RRM). The role of the prediction module is to generate a preliminary saliency map, while the role of the residual refinement module is to further optimize and refine the saliency map.

2.3 Least squares based identification of tennis centers

The edge points of a circle can be quickly and accurately identified on a salient map using the Canny operator. In this paper, Least Squares Method (LSM) is used to fit the center of the ellipse for sub-pixel localization. The general expression for the ellipse equation is given below:

$$f(\alpha, X) = \alpha X = Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0 \quad (4)$$

where A, B, C, D, E and F are the five elliptic parameters, α stands for (A, B, C, D, E, F) , and $X_i = (x_i^2, x_i y_i, y_i^2, x_i, y_i, 1)$. According to the principle of least squares, the curve fitting problem can be solved by minimizing the sum of squares of algebraic distances:

$$f(A, B, C, D, E, F) = \sum_{i=1} (Ax_i^2 + Bx_i y_i + Cy_i^2 + Dx_i + Ey_i + F)^2 \quad (5)$$

The five elliptic parameters A, B, C, D, E and F in Eq. can be obtained by calculating the first order partial derivatives and setting them to zero. If the center coordinate of the ellipse is $P(x_0, y_0)$, the calculation formula can be expressed as:

$$\begin{cases} x_0 = (2BC - AD) / (A^2 - 4B) \\ y_0 = (2D - AD) / (A^2 - 4B) \end{cases} \quad (6)$$

3 Tennis serve drop prediction

3.1 3D reconstruction of tennis balls based on binocular camera models

3.1.1 Principles of Camera Calibration

Optical lens characteristics lead to radial aberrations in camera imaging, which are usually expressed in three parameters, denoted as k_1 , k_2 and k_3 . Whereas, the inability to achieve perfect parallelism between the sensor and the optical lens during the assembly process results in tangential aberrations in the camera imaging, which are usually expressed utilizing two parameters, denoted as p_1 and p_2 . Usually the calibration of a single camera is performed to obtain the internal parameters of the camera (camera focal length f , imaging origin c_x and c_y , aberration coefficients k_1 , k_2 , k_3 , p_1 , and p_2) to perform the corrective operations on the aberrated images. After obtaining the calibration parameters of a single camera, it is possible to correct the distorted images captured by the camera. In the experimental process, for the acquisition of the internal and external parameters of the camera, we adopt the tessellated grid-based camera calibration technique in the OpenCV open source library.

Using the acquired calibration parameters, we were able to correct the tennis balls imaged in the two cameras separately. Thus, the influence of image distortion on visual localization accuracy is excluded. After obtaining the corrected image of the aligned synchronous image, the center of the ball position of the tennis ball in the two images can be obtained by using the fusion color model and the center of mass method. The spatial position of the tennis ball can be determined by combining the imaging principle of the tennis ball in the camera. In order to improve the accuracy, after the monocular calibration of the cameras respectively, binocular calibration of the two synchronized cameras is needed to obtain the translation matrix T and relative rotation matrix R between the two cameras. Using the `reProjectImageTo3D` function in OpenCV combined with the synchronized calibration image can obtain the transformation matrix Q , which contains the geometric relationship parameters between the two cameras. Taking the left camera as the base coordinate system, the Q matrix is expressed as follows:

$$Q = \begin{pmatrix} 1 & 0 & 0 & -c_x \\ 0 & 1 & 0 & -c_y \\ 0 & 0 & 0 & f \\ 0 & 0 & \frac{-1}{T_x} & \frac{c_x - c_x^r}{T_x} \end{pmatrix} \quad (7)$$

After obtaining the transformation matrix Q , the spatial location information of the tennis ball can be deduced by combining the geometrical relations using the difference of its position in the synchronized images respectively. Assuming that the positions of the tennis ball in the left and right views are (x_l, y_l) and (x_r, y_r) respectively, the parallax of the tennis ball on the synchronized image can be defined as:

$$d = x_l - x_r \quad (8)$$

Then the spatial coordinates of the tennis ball (X, Y, Z) can be computed in the following

way:

$$X = x_l - c_x \quad (9)$$

$$Y = y_l - c_y \quad (10)$$

$$Z = \frac{-T_x f}{d - (c_x - c_x^r)} \quad (11)$$

3.1.2 Three-dimensional position reconstruction of tennis balls

The computation of the spatial position of the tennis ball relies on the acquisition of the calibration parameters of the cameras, therefore, the vision system has to design offline calibration experiments thus saving the parameters of the cameras in advance before performing real-time visual measurements of a real tennis ball. In this chapter, two smart cameras are calibrated by using the checkerboard grid calibration method in order to obtain the correction coefficients for a single camera and the transformation matrix between the two cameras.

Once the internal and external parameters and transformation matrix of the binocular camera are obtained through offline camera calibration experiments, the real-time vision system is able to localize the tennis ball in the scene in real time. First, the binocular camera acquires the scene image in real time and obtains the synchronized images containing the tennis ball, i.e., the left and right images. Then, the synchronized images are corrected and aligned by combining the acquired single target fixation parameters. Then, the centers of the tennis balls in the images are calculated using the fused color model and the center-of-mass method, which are denoted as (x_l, y_l) and (x_r, y_r) , respectively. Next, the tennis ball is reconstructed in three dimensions using the imaging parallax of the tennis ball in the synchronization image and the transformation matrix obtained from binocular calibration to obtain the spatial position of the tennis ball serve.

3.1.3 Coordinate system conversion

The coordinates of the tennis ball calculated by OpenCV-based 3D recalculation of the tennis ball are in the left camera coordinate system, i.e., the left camera is used as the coordinate origin. This position expression is not conducive to the study of tennis ball drop prediction, and in order to simplify the analysis of tennis ball drop, the position of the tennis ball relative to the ground of the effective playing area needs to be calculated. For this reason, in our study, the center of the ground of the effective playing area is used as the origin O of the world coordinate system, and the axis of the coordinate system Z is perpendicular to the ground upwards. Then, we utilize a standard piece to assist in calculating the mapping parameters for the conversion of the camera coordinate system to the world coordinate system. The tennis ball is placed at O , A (length direction of the ground in the effective playing area), B (width direction of the ground in the effective playing area), and C (height direction of the ground in the effective playing area), and the vision system is utilized to locate each of the four points, and to obtain a three-dimensional position representation of the four points in the left camera coordinate system. The three-dimensional position of these four points in the left camera coordinate system is denoted as $O(X_O, Y_O, Z_O)$, $A(X_A, Y_A, Z_A)$, $B(X_B, Y_B, Z_B)$ and $C(X_C, Y_C, Z_C)$.

Then for any position representation $P(X_P, Y_P, Z_P)$ in the camera coordinate system, the

corresponding position $P'(X_{P'}, Y_{P'}, Z_{P'})$ in the set world coordinate system can be represented as:

$$\begin{pmatrix} X_{P'} \\ Y_{P'} \\ Z_{P'} \end{pmatrix} = \begin{pmatrix} \frac{X_A - X_O}{L_X} & \frac{Y_A - Y_O}{L_X} & \frac{Z_A - Z_O}{L_X} \\ \frac{X_B - X_O}{L_Y} & \frac{Y_B - Y_O}{L_Y} & \frac{Z_B - Z_O}{L_Y} \\ \frac{X_C - X_O}{L_Z} & \frac{Y_C - Y_O}{L_Z} & \frac{Z_C - Z_O}{L_Z} \end{pmatrix} \cdot \begin{pmatrix} X_P - X_O \\ Y_P - Y_O \\ Z_P - Z_O \end{pmatrix} \quad (12)$$

3.2 Tennis ball drop prediction model

Through the previous three-dimensional position reconstruction of the tennis ball, the serve position and a series of continuous three-dimensional points of the tennis ball are obtained, and the next step to do is to get the trajectory of the tennis ball based on these three-dimensional points so as to predict the drop point. In this paper, the continuous motion model is used to predict the trajectory and solve the drop point.

3.2.1 Tennis force analysis

The forces on the tennis ball are shown in Figure 3. Where v is the speed of motion of the tennis ball, ω is the rotational speed of the tennis ball, F_g is the gravitational force on the tennis ball in the direction vertically downward, F_d is the air resistance, opposite to the direction of motion, F_b is the buoyancy force in the direction vertically upward, and F_m is the Magnus force due to selection, in a direction perpendicular to the direction of motion and the axis of rotation, satisfying the right-hand rule.

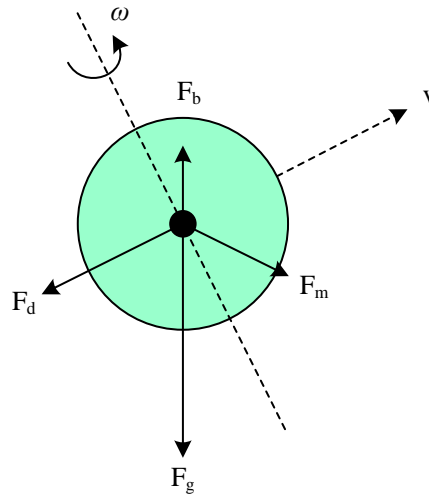


Figure 3: Force analysis of flying tennis

3.2.2 Continuous motion model

If the force to establish the tennis ball dynamics model is more complex, do the following simplification: tennis ball gravity relative to the exhaust air generated by the buoyancy, to several orders of magnitude larger, you can ignore the buoyancy. In addition, the software and

hardware in this paper can not do to observe the tennis ball rotating to calculate the rotational speed, so ignore the Magnus force generated by it, as far as possible to avoid the introduction of rotation in the experiment. Then the tennis ball only by gravity and air resistance, can be calculated by the following formula:

$$F_g = [0, 0, -mg]^T \quad (13)$$

$$F_d = -\frac{1}{2} C_D \rho_a A \|v\| v \quad (14)$$

where m is the mass of the tennis ball, g is the acceleration of gravity, C_D is the coefficient of air resistance, ρ_a is the air density, A is the maximum cross section of the tennis ball, and $v = [v_x, v_y, v_z]^T$ is the spatial velocity of the tennis ball. In addition to the state variables are physical constants, the reference values are shown, different experimental environment, the value will also change.

For the convenience of description, the constants in Eq. (14) are written as:

$$K_D = \frac{1}{2m} C_D \rho_a A \quad (15)$$

It follows from Newton's second law:

$$m \dot{v} = F_g + F_d \quad (16)$$

A slight transformation and written in matrix form has:

$$\begin{bmatrix} \dot{v}_x \\ \dot{v}_y \\ \dot{v}_z \end{bmatrix} = \begin{bmatrix} -K_D \|v\| v_x \\ -K_D \|v\| v_y \\ -K_D \|v\| v_z - g \end{bmatrix} \quad (17)$$

The above equation is the kinetic model of a tennis ball, which can also be written in the form of the following differential equation:

$$\dot{v} = \begin{bmatrix} -K_D \|v\| & 0 & 0 \\ 0 & -K_D \|v\| & 0 \\ 0 & 0 & -K_D \|v\| \end{bmatrix} v + \begin{bmatrix} 0 \\ 0 \\ -g \end{bmatrix} \quad (18)$$

In this differential equation, the coefficient matrix contains the variable v itself, so it is not possible to find an analytical solution. Observing the above equation, we will find that the flight process of the tennis ball is a Markov process, that is, as long as we know the state of motion of the tennis ball at any moment, we can determine its trajectory after all. Assuming that the tennis ball motion state has been obtained, the above formula can be used to calculate the landing point, although the analytical solution can not be obtained, but can be solved by Euler's method, Longecourt method and other numerical methods of iterative solution. Euler's

method assumes that within a certain step size Δt , \dot{v}_k is constant, then $v_{k+1} = v + \dot{v}_k \Delta t$, and keep iterating like this to find the desired position, the accuracy depends on the step size Δt , the accuracy is high when the step size is small, but the computation is large. Obviously, such an approximation will have a cumulative error, and the larger the distance, the larger the cumulative error. Since the dynamics modeled by Eq. (18) is relatively simple, we have a simpler and more accurate approximation than Euler's method: assume that within a small period of time ΔT , $\|v\|$ is a fixed value, the exact value is given by the initial velocity of the small section of the trajectory, i.e., let:

$$K = K_D \|v\| \quad (19)$$

Then it is obtained from Eq. (17):

$$\begin{bmatrix} \dot{v}_x \\ \dot{v}_y \\ \dot{v}_z \end{bmatrix} = \begin{bmatrix} -Kv_x \\ -Kv_y \\ -Kv_z - g \end{bmatrix} \quad (20)$$

3.2.3 Tennis drop prediction

It can be found that there is no more coupling in each direction of the kinetic model in Eq. (20), and the nonlinearity becomes linear. The differential equations in each direction in the above equation can be solved separately, in which the x , y direction is a first-order chi-square linear differential equation, and the z direction is a first-order non-chi-square linear differential equation, which can be obtained by solving this differential equation and substituting the initial condition v_0 :

$$\begin{cases} v_x(t) = v_{0x} \cdot e^{-Kt} \\ v_y(t) = v_{0y} \cdot e^{-Kt} \\ v_z(t) = v_{0z} \cdot e^{-Kt} + \frac{g}{K} (e^{-Kt} - 1) \end{cases} \quad (21)$$

The velocity is the first order derivative of the displacement with respect to time, so the above equation can be viewed as a differential equation for the displacement $p(t)$, which is a separable-variable type differential equation, solving and substituting the initial condition p_0 has:

$$\begin{cases} p_x(t) = p_{0x} - \frac{v_{0x}}{K} (e^{-Kt} - 1) \\ p_y(t) = p_{0y} - \frac{v_{0y}}{K} (e^{-Kt} - 1) \\ p_z(t) = p_{0z} - \frac{v_{0z}}{K} (e^{-Kt} - 1) - \frac{g}{K} t \end{cases} \quad (22)$$

Eq. (21) and Eq. (22) are the final continuous motion models used. It should be noted that the derivation of the above differential equation solutions are all under the assumption of a small segment of trajectory with a fixed value of K , so after each solution of this small segment of trajectory, it is necessary to update K and set up this equation at a new starting point. Where $[p(t), v(t)]^T$ is the state of the tennis ball after t time from the starting point in this small segment of the trajectory, and $[p_0, v_0]^T$ is the state of the tennis ball at the starting point.

As with the kinetic model, as long as we know the state of motion of the tennis ball at any moment, we can use the kinematic continuum model to determine all the trajectories after it, so we can use the above formula to iterate the calculation, and when the position state of the tennis ball's z -axis is sufficiently close to 0, we consider that this is the tennis ball's landing point.

4 Empirical analysis

4.1 Evaluation of the effectiveness of the tennis target test

4.1.1 Experimental setup

The operating system for the experiments in this chapter is Ubuntu 20.04, and the deep learning framework is Pytorch 1.13.0+cu117. In training the YOLOv4 target detector fused with AMB, the COCO pre-training weights are initialized in order to speed up the training using the data from the Tennis Ball-640 (Data 1) and Tennis Double-Entry ECCV'22 (Data 2) data sets for 100 rounds of training. The model was trained on 2 NVIDIA 2080Ti GPUs with a batch size of 10.

4.1.2 Evaluation indicators

Detection accuracy indicates the number of correctly classified samples as a proportion of the total number of samples, as opposed to error rate, which indicates the number of incorrectly classified samples as a proportion of the total number of samples. The accuracy rate is calculated by the formula:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (23)$$

The precision rate indicates the proportion of samples predicted to be positive cases that are actually positive cases. It is calculated by the formula:

$$Precision = \frac{TP}{TP + FP} \quad (24)$$

$F1$ is a metric used to measure the precision of a binary classification model. Precision and recall are contradictory metrics; in general, recall tends to be low when precision is high. In general, when precision is high, recall tends to be low, and when recall is high, precision tends to be low. The $F1$ score integrates the precision rate and recall rate, and is a weighted average of them. It is calculated by the formula:

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2TP}{2TP + FP + FN} \quad (25)$$

4.1.3 Analysis of ablation experiment results

We validated the detection performance on two test sets, Data 1 and Data 2, respectively. The training period is 100 cycles. The detection results are shown in Fig. 4, respectively. The results show that by the introduction of the channel attention module (CAM) and the spatial attention module (SAM), the improved YOLOv4 algorithm has significantly improved the detection performance on both test sets, $F1$ by 1.6 and 1.9 percentage points, and the detection accuracy by 2 and 1.9 percentage points respectively. Compared with the introduction of AMB only, the addition of BASNet and LSM improves the detection accuracy of the model on the two test sets by 0.8% and 1.9%, respectively. The tennis target recognition and tracking method based on the attention mechanism designed in this paper can ensure the detection of tennis serve point location.

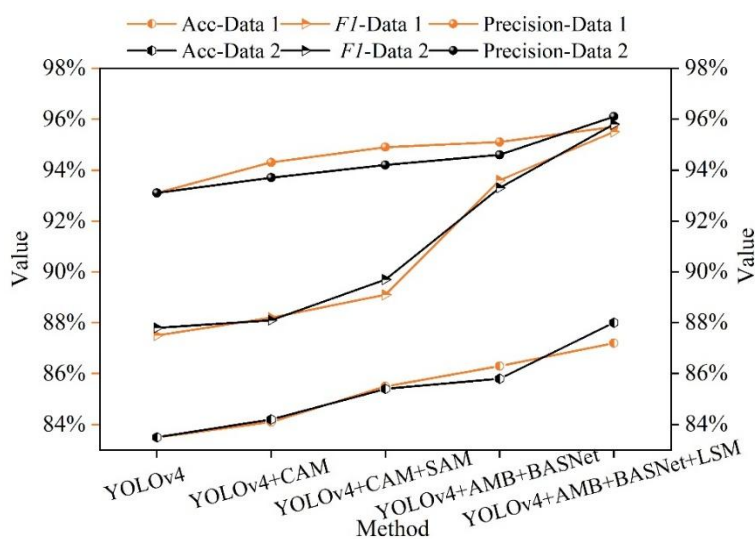


Figure 4: Ablation experiment results

4.2 Evaluation of the effectiveness of tennis ball drop prediction

In order to verify the overall effectiveness of the tennis serve drop prediction method, it is necessary to test the tennis serve drop prediction method. Taking the ATP singles tennis venue as an example, the length, width and height boundaries of the venue are 23.77 m, 8.23 m and 12.19 m. The prediction methods in this paper (three-dimensional coordinate reconstruction + force characterization + continuous motion model), the air resistance + gravity analytical model + least-squares fitting (Method 1) and LSTM sequence-to-sequence learning (Method 2) are used to predict the trajectory and landing point of the tennis serve, and the comparative results are shown in Figure 5.) methods are used to predict the trajectory and landing point of tennis serve, and the comparison results are obtained as shown in Fig. 5, and (a)~(c) represent the prediction effects of the prediction methods in this paper, Method 1 and Method 2, respectively. The locations of the serve points are all (-5,-2.8,2), and the actual drop point location is (3.7,3.4,0). It can be analyzed that when the proposed method is used to predict the trajectory and landing point of tennis serve, the predicted trajectory and landing point obtained are basically consistent with the actual trajectory and landing point, and the errors of X, Y and Z axes are very small and negligible. When the other two methods are used to predict the trajectory and landing point of tennis serve, the predicted trajectory and landing point obtained have errors in the X, Y and Z axes, and the locations of the landing points of Method 1 and Method 2 are (3.74,3.6,0) and (3.8,3.5,0), respectively. It shows that the above two methods cannot accurately predict the trajectory and landing point of tennis serve. Through the above

test results, it is verified that the proposed method has high prediction accuracy because the proposed method analyzes the force characteristics of the tennis ball before predicting the landing point of the tennis ball, and the trajectory is predicted on this basis, which improves the prediction accuracy.

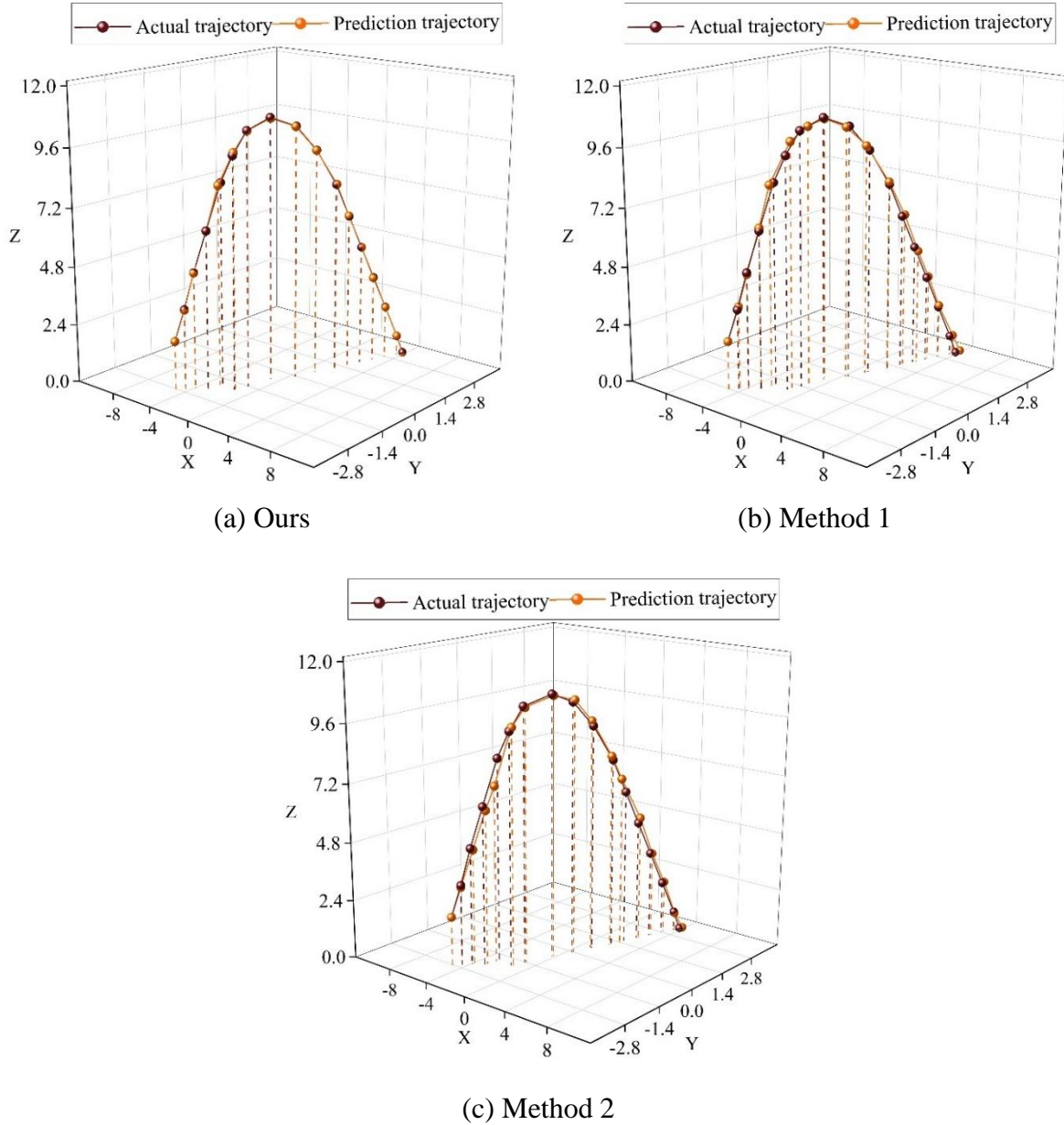


Figure 5: Comparison of the game of tennis

4.3 Case studies of model applications

This paper takes a match data of two tennis players (A and B) in the 15th National Games as an example, and uses the tennis target recognition and tracking method designed in this paper, as well as the established drop point prediction model for simulation and analysis. The characteristics of both sides' serve drop points and points scored and lost are shown in Table 1. In the actual match, players A and B served 80 and 63 balls respectively, and the actual scores were 62 and 83 points, respectively. The probability of both players A and B's tennis serves landing in the outer corner is the largest, which is 37.5% and 41.5% respectively. When the

landing area is in the outer corner, the scoring rate of athlete B is higher, which can reach 32.4%.

Table 1: The two sides serve and the gain and loss characteristics

Athlete	Serve area	Drop area coding	Drop probability	Scoring rate	Loss rate
A	Outer Angle	30	37.5	0	2.5
	Middle road	31	29.6	0	2.5
	Inner Angle	32	32.9	0	0
B	Outer Angle	30	41.5	32.4	0
	Middle road	31	26.3	0	0
	Inner Angle	32	32.2	15.6	0

Using the improved YOLOv4-based detector can successfully detect the position of 143 serves, and the prediction results of the drop point prediction model for the drop points of the serves of A and B players are shown in Figure 6. The prediction error range between the predicted drop point position and the actual drop point position is between 0.3% and 0.8%, indicating that the method of this paper predicts the trajectory of tennis balls on the basis of their force characteristics, which ensures the accuracy of the drop point prediction results, and facilitates the training of athletes.

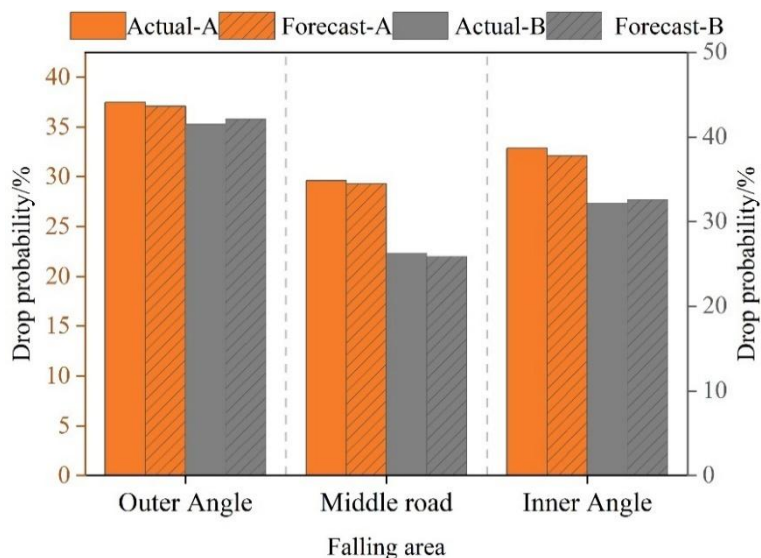


Figure 6: Analysis of the prediction results of the athletes' service

5 Conclusion

This paper explores the fast recognition of tennis ball and the prediction of trajectory landing point, and mainly draws the following conclusions:

(1) Based on the YOLOv4 method, this paper proposes a tennis ball target recognition and tracking method based on the fused attention mechanism, and introduces BASNet and LSM to further improve the detection accuracy. The detection results are validated on two datasets, Tennis Ball-640 and Tennis Double-Entry ECCV'22, and the results show that the inclusion of the attention mechanism improves the detection accuracy by 2% and 1.9%, respectively.

(2) In this paper, we use 3D coordinate reconstruction + force characterization + continuous motion model for tennis ball landing point prediction, and the method in this paper has good prediction performance compared with the air resistance + gravity resolution model + least

squares fitting method and LSTM sequence-to-sequence learning method.

In terms of the current research results of this study, some of the work is still to be improved by further in-depth research in the future, mainly reflected in the establishment of the dynamics model according to the tennis ball force, ignoring the effect of tennis ball rotation on the flight trajectory, for the trajectory prediction of the research can be carried out from this aspect.

References

- [1] Abrams, G. D., Sheets, A. L., Andriacchi, T. P., & Safran, M. R. (2011). Review of tennis serve motion analysis and the biomechanics of three serve types with implications for injury. *Sports Biomechanics*, 10(4), 378-390.
- [2] Palmer, K., Jones, D., Morgan, C., & Zeppieri Jr, G. (2018). Relationship between range of motion, strength, motor control, power, and the tennis serve in competitive-level tennis players: a pilot study. *Sports health*, 10(5), 462-467.
- [3] Roanes-Lozano, E., Casella, E. A., Sánchez, F., & Hernando, A. (2020). Diagnosis in tennis serving technique. *Algorithms*, 13(5), 106.
- [4] Sakurai, S., Reid, M., & Elliott, B. (2013). Ball spin in the tennis serve: spin rate and axis of rotation. *Sports biomechanics*, 12(1), 23-29.
- [5] Baiget, E., Corbi, F., & López, J. (2023). Influence of anthropometric, ball impact and landing location parameters on serve velocity in elite tennis competition. *Biology of Sport*, 40(1), 273-281.
- [6] Gong, X., & Wang, F. (2021). Classification of tennis video types based on machine learning technology. *Wireless Communications and Mobile Computing*, 2021(1), 2055703.
- [7] Terroba, A., Kusters, W., Varona, J., & Manresa-Yee, C. S. (2013). Finding optimal strategies in tennis from video sequences. *International Journal of Pattern Recognition and Artificial Intelligence*, 27(06), 1355010.
- [8] Chu, W. T., & Tsai, W. H. (2010). Modeling spatiotemporal relationships between moving objects for event tactics analysis in tennis videos. *Multimedia Tools and Applications*, 50(1), 149-171.
- [9] Renò, V., Mosca, N., Nitti, M., D’Orazio, T., Guaragnella, C., Campagnoli, D., ... & Stella, E. (2017). A technology platform for automatic high-level tennis game analysis. *Computer Vision and Image Understanding*, 159, 164-175.
- [10] Wu, F., Wang, Q., Bian, J., Ding, N., Lu, F., Cheng, J., ... & Xiong, H. (2022). A survey on video action recognition in sports: Datasets, methods and applications. *IEEE Transactions on Multimedia*, 25, 7943-7966.
- [11] Xiao, J., Tian, W., & Ding, L. (2022). Basketball action recognition method of deep neural network based on dynamic residual attention mechanism. *Information*, 14(1), 13.
- [12] Wu, D., Sharma, N., & Blumenstein, M. (2017, May). Recent advances in video-based

- human action recognition using deep learning: A review. In 2017 International joint conference on neural networks (IJCNN) (pp. 2865-2872). IEEE.
- [13] Chen, K., Zhang, D., Yao, L., Guo, B., Yu, Z., & Liu, Y. (2021). Deep learning for sensor-based human activity recognition: Overview, challenges, and opportunities. *ACM Computing Surveys (CSUR)*, 54(4), 1-40.
- [14] Whiteside, D., & Reid, M. (2017). Spatial characteristics of professional tennis serves with implications for serving aces: A machine learning approach. *Journal of sports sciences*, 35(7), 648-654.
- [15] Wu, D., & Xiao, A. (2022). Deep learning-based algorithm for recognizing tennis balls. *Applied Sciences*, 12(23), 12116.
- [16] Taha, Z., Musa, R. M., Majeed, A. P. A., Alim, M. M., & Abdullah, M. R. (2018). The identification of high potential archers based on fitness and motor ability variables: A Support Vector Machine approach. *Human movement science*, 57, 184-193.
- [17] Chen, W. J., Jhou, M. J., Lee, T. S., & Lu, C. J. (2021). Hybrid basketball game outcome prediction model by integrating data mining methods for the national basketball association. *Entropy*, 23(4), 477.
- [18] Niu, M. (2021, September). Data Mining Method of Tennis Technique and Tactics Association Analysis. In *International Conference on Cognitive based Information Processing and Applications (CIPA 2021) Volume 2* (pp. 284-290). Singapore: Springer Singapore.
- [19] Chen, H. (2022). A Data Mining-Based Model for Evaluating Tennis Players' Training Movements. *Discrete Dynamics in Nature and Society*, 2022(1), 8950732.
- [20] Zhang, S., & Mao, H. (2021). Optimization analysis of tennis players' physical fitness index based on data mining and mobile computing. *Wireless Communications and Mobile Computing*, 2021(1), 9838477.
- [21] Zhang, J., Zhu, Q., Xu, L., & Gong, H. (2024, July). Research on tennis match evaluation and prediction model based on data mining. In *Fourth International Conference on Applied Mathematics, Modelling, and Intelligent Computing (CAMMIC 2024)* (Vol. 13219, pp. 1169-1177). SPIE.
- [22] Wang, X. (2023, October). Tennis Skills and Tactics Intelligent Analysis Mechanism Based on Data Mining. In *2023 IEEE 5th Eurasia Conference on IOT, Communication and Engineering (ECICE)* (pp. 825-828). IEEE.
- [23] Kaliffe, K. S., & Santos, R. (2024, November). Finding the Best Tennis Serves with K-Means and GMM Clusters of Ball Tracking data to Interpret Serve Strategies. In *Symposium on Knowledge Discovery, Mining and Learning (KDMiLe)* (pp. 73-80). SBC.
- [24] Wei, X., Lucey, P., Morgan, S., Carr, P., Reid, M., & Sridharan, S. (2015, August). Predicting serves in tennis using style priors. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 2207-2215).

- [25] Mecheri, S., Rioult, F., Mantel, B., Kauffmann, F., & Benguigui, N. (2016). The serve impact in tennis: First large-scale study of big Hawk-Eye data. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 9(5), 310-325.
- [26] Sampaio, T., Oliveira, J. P., Marinho, D. A., Neiva, H. P., & Morais, J. E. (2024). Applications of machine learning to optimize tennis performance: a systematic review. *Applied Sciences*, 14(13), 5517.
- [27] Huang, H., & Deng, R. (2020). Analysis technology of tennis sports match based on data mining and image feature retrieval. *Complexity*, 2020(1), 8877161.
- [28] Zhu, Y., & Naikar, R. (2022, September). Predicting tennis serve directions with machine learning. In *International Workshop on Machine Learning and Data Mining for Sports Analytics* (pp. 89-100). Cham: Springer Nature Switzerland.
- [29] Gao, Z., & Kowalczyk, A. (2021). Random forest model identifies serve strength as a key predictor of tennis match outcome. *Journal of Sports Analytics*, 7(4), 255-262.
- [30] Vives, F., Lázaro, J., Guzmán, J. F., Martínez-Gallego, R., & Crespo, M. (2023). Optimizing sporting actions effectiveness: A machine learning approach to uncover key variables in the men's professional doubles tennis serve. *Applied Sciences*, 13(24), 13213.
- [31] Liu, J. W. (2024). Enhancing Tennis Serve Scoring Efficiency: An AI Deep Learning Approach. *International Journal of Interactive Multimedia and Artificial Intelligence*, 8(7), 48-54.
- [32] Ye, C., Zhu, R., Ma, J., Huang, H., Li, X., & Wen, J. (2024). Comprehensive tennis serve training system based on local attention-based CNN model. *IEEE Sensors Journal*, 24(7), 11917-11926.