Basketball Fixed-point Shooting Hit Prediction Based on Human Pose Estimation Algorithm

Xi Li[1], Jiao Hua[2]

[1]Department of Physical Education, Wuxi Taihu University, Wuxi 214000, China

[2]Physical Education Group, Wuxi Yangming Center Primary School, Wuxi 214000, China

E-mail: 000058@wxu.edu.cn, tiandiboy2000@163.com

As computer vision and artificial intelligence develop, the research on basketball fixed-point shooting hit prediction based on human pose estimation algorithm becomes a topic of great concern. To construct a basketball fixed-point shooting hit prediction model, a new object detection algorithm was designed, and the You Only Look Once version 5 algorithm was optimized on the basis of GIoU loss function and Convolutional Block Attention Module. Then, a new human pose estimation algorithm was designed based on the OpenPose algorithm. Results showed that the average accuracy of the improved YOLOv5 algorithm reached 95.34% when the number of iterations was 50. In the comparison among improved OpenPose and other algorithms, improved OpenPose performed better, with a recall rate of 96.23%, an accuracy of 87.16%, a precision of 89.75%, and an F1 value of 88.19%. In the comparison with other models, the area of receiver operating characteristic curve was the largest, reaching 0.974, and the F1 value, accuracy and recall rate of the research model were the highest, reaching 95.54%, 96.39% and 98.25%, respectively. Results show that it effectively predicts the shoot hit of basketball fixed-point shooting, which provides a useful reference for tactical analysis and player performance evaluation in basketball games.

Povzetek: Raziskava predstavlja napovedni model za zadetke pri košarkarskem metu s fiksne točke na osnovi algoritma za ocenjevanje človeške drže, kar je koristno za taktično analizo in ocenjevanje igralcev v košarki.

1 Introduction

Basketball is a competitive sport that combines speed, strength, skill and tactics and requires a high degree of skill and strategic awareness. In basketball, fixed-point shooting is a common means of scoring, and predicting the shooting rate is of great significance for the team's tactical layout and the improvement of players' skills [1]. Human pose estimation is the recognition and analysis of human pose information in images or videos through computer vision technology [2]. Through human pose estimation, key information such as the player's body posture, action characteristics, and movement trajectory during the shooting process can be obtained [3, 4]. This information can be further used to analyze the relationship between a player's skill level and shooting rate [5]. The purpose of this study is to use the human pose estimation algorithm to predict the hit rate of basketball fixed-point shooting. In view of this, this study analyzes the posture and action characteristics of players during the shooting, and establishes a prediction model using machine learning. The significance of the study is that by predicting the shooting hit rate, coaches and players can adjust their shooting strategies to improve the winning rate of the game. In addition, it also provides new ideas and methods to apply computer vision and artificial intelligence in sports competition. The research content includes four parts. The first part is a detailed introduction of the human pose estimation algorithm and the fixed-point shooting hit prediction model, and the second part is the You Only Look Once version 5 (YOLOv5) algorithm using object detection, and its optimization. Then, based on the OpenPose algorithm, a new human pose estimation algorithm is designed. Finally, a basketball fixed-point shooting hit prediction system is designed. The third part verifies the validity and reliability of the research model through experimental design and data analysis. The fourth part summarizes and prospects the research content.

2 Related works

Human pose estimation is important in computer vision in recent decades, and it plays a crucial role in understanding people in images and videos. At present, there have been many studies on human pose estimation. Dubey and Dixit reviewed the key research and recent

advances in human pose estimation, including 2D and 3D pose estimation techniques and their traditional and deep learning methods. The results showed that these different pose assessment methods effectively improved the accuracy and efficiency of pose prediction [6]. Liu et al. proposed a lightweight pose estimation network using the polarimetric self-attention mechanism. Firstly, ghost convolution reduced feature extraction network parameters. Secondly, polarimetric self-attention module solved the pixel-level regression task, reduce the insufficient feature extraction caused by parameter reduction, and improve human key point regression accuracy. Finally, a new coordinate decoding method reduced the error during heat map decoding and improve the accuracy of key point regression. Results showed that it reduced model parameters while ensuring a small loss of accuracy [7]. Qin et al. proposed a lightweight human pose estimation network and named it CVC-Net, aiming to reduce the complexity and improve human pose detection speed. CVC-Net was based on the stacked hourglass network architecture, which used Res2Net_depth residual blocks to reduce the parameters, and combined channel attention mechanism and PixelShuffle upsampling methods to optimize performance. Results showed that CVC-Net significantly reduced the model parameters while maintaining high accuracy, which was especially suitable for devices with limited computing power [8]. Xu et al. proposed a novel network called multi-scale position augmentation network to improve small and medium-scale key point detection and semantic confusion discrimination in human pose estimation. The network adopted a multi-scale adaptive fusion unit and a position enhancement module to emphasize the real joint position characteristics and improve the detection accuracy. Experimental results showed that network performance was significantly improved in pose estimation tasks, and results were more accurate and reliable [9].

Jiang proposed a deep learning-based basketball game monitoring system to collect data non-invasively to identify player behavior during the game. The system used a video frame to analyze the rebounding situation and predict the player's position to get the rebound. In addition, the position of the player was determined by traditional regression techniques and their movement towards the point of landing. Simulation analysis of the feasibility, performance, and system efficiency demonstrated the reliability of the framework [10]. Özkan studied the electrophysiological basis of predicting free throw hits in wheelchair basketball players. Their predictions of hitting or missing free throws were observed by conducting EEG tests on semi-professional wheelchair basketball players and non-professionals. The results of the study showed that expert players exhibited significant negative amplitude in the 100 milliseconds before the release of the free throw, and this electrophysiological response was regarded as a valid indicator for predicting the effect of the action [11]. Siemon and Jörn used Twitter data mining technology to predict the performance of NBA basketball players. The study conducted automated personality mining on the tweets of 185 professional players and collected their top five personality traits and player statistics. Correlation and multiple linear regression analyses found that personality traits such as extraversion, agreeableness, and conscientiousness were associated with basketball performance and was used to predict future performance [12]. Naik and Hashmi discussed the ability to predict the trajectory of dynamic objects in a dynamic sports environment, especially in basketball, and proposed a dual-mode exponential normal distribution processing method with relational network to accurately predict the trajectory of basketball. Results showed a good shooting status prediction of athletes [13]. The results of the literature survey are shown in Table 1.

Table 1: Results of the literature survey

Literature number	Accuracy (%)	Recall (%)	Advantage	Insufficient
[6]	/	/	Summarize the current progress in pose estimation research	/
[7]	84	94	Reduce in the algorithm parameters	The algorithm loss is not effectively controlled
[8]	78	94	Improve the applicability of the model	The model accuracy is reduced
[9]	91	95	The algorithm performance has improved significantly	The model increases
[10]	89	92	Can accurately predict the location of the	The model calculation is more

			athletes	time-consuming
[11]	/	/	New indicators of action prediction are found	No method is proposed according to the new indicators
[12]	/	/	Found potential associations of personality traits with basketball performance	/
[13]	84	96	Players' shooting status can be predicted according to the basketball movement trajectory	The model calculation is complicated and less efficient

In summary, significant progress was made in human pose estimation, especially with the support of deep learning. From Dubey and Dixit's [6] review to specific algorithm innovations such as Liu et al.'s [7] lightweight network, Qin et al.'s [8] CVC-Net, and Xu et al.'s [9] multi-scale position enhancement network, these studies have demonstrated efforts to improve estimation accuracy and computational efficiency. In addition, Jiang and Özkan [10, 11] showed the practical application value of the technique by applying the human pose estimation technique to specific scenarios, such as basketball motion monitoring and wheelchair basketball player behavior prediction. These studies have improved the accuracy and efficiency of pose estimation and expanded its application in sports, medicine, and other fields. In view of this, this paper constructs a basketball fixed-point shooting hit prediction model based on the human pose estimation algorithm.

3 Construction of basketball fixed-point shooting hit prediction model based on human pose estimation algorithm

To construct a basketball fixed-point shooting hit prediction model, this study first introduced the YOLOv5 algorithm based on object detection and introduced the GIoU loss function to optimize YOLOv5 algorithm.

Combined with Convolutional Block Attention Module (CBAM) attention mechanism, environment interference on object detection was reduced. Then, a new human pose estimation algorithm was designed based on the OpenPose algorithm. Finally, a basketball fixed-point shooting hit prediction system was designed.

3.1 Basketball object detection method using deep learning

In sports, human object detection can identify athletes on the field, and even can detect and identify referees and spectators. In the competition, the performance and technique of athletes are judged, and a technology that can track athletes and detect and identify athletes' movements is needed. YOLOv5 is an object detection algorithm with fast speed, high accuracy and strong real-time performance. YOLOv5 is more accurate in detecting small objects and can adapt to a variety of different scenarios and mission requirements [14]. Compared with other object detection algorithms, YOLOv5 is more concise, efficient, stable, and easy to expand and optimize [15-17]. In view of this, the YOLOv5 network is used to detect fixed-point shooting in basketball games, hoping to complete the detection and positioning of basketball players and further realize the posture recognition of athletes, as shown in Figure 1.
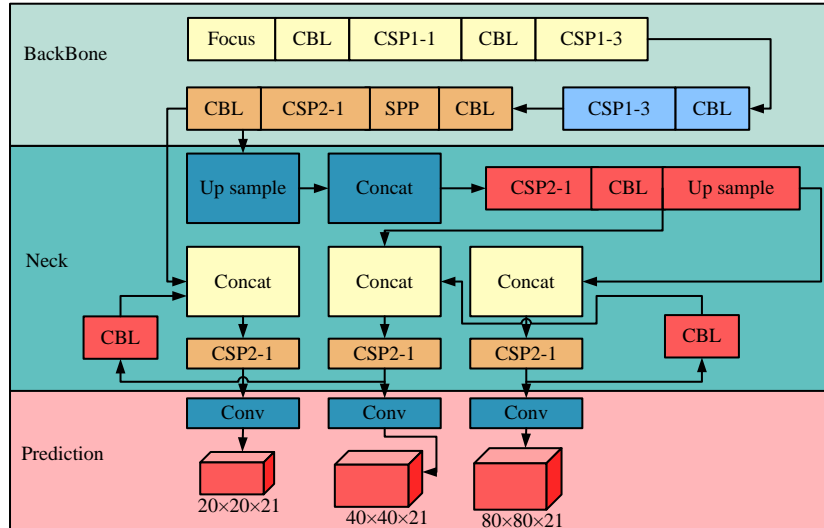
Figure 1: YOLOv5 network architecture

In Figure 1, YOLOv5 is an object detection algorithm including backbone, neck, and prediction structure. In the backbone structure, it uses a feature pyramid structure for feature extraction. This structure can effectively fuse shallow features and deep features and detect targets of different sizes [18, 19]. In addition, YOLOv5 also uses lightweight convolutional neural networks as backbones, such as MobileNetV3, to further reduce network complexity and improve the running speed. In the neck structure, YOLOv5 adopts a top-down path fusion structure to shorten low-level feature flow path to the prediction layer. This structure can effectively reduce computing and improve the efficiency of network operation. In the prediction structure, YOLOv5 integrates the features, and fuses different low-level features with three paths: prediction1, prediction2 and prediction3, respectively, and outputs the defect target bounding box information and category information. This design can improve the network detection accuracy and robustness against targets of different sizes and attitudes [20]. In this study, the GIoU loss function optimizes the traditional YOLOv5, and IoU loss function expression used by the traditional YOLOv5 algorithm is shown in equation (1).

$$L_{IoU} = 1 - IoU(A, B) \qquad (1)$$

In equation (1), if the target box and the prediction box do not intersect, IoU's value is 0, but if there is no intersection between the two boxes, the relationship between the two boxes cannot be measured, and the training and learning cannot be carried out, so that the regression effect cannot be evaluated [21]. Therefore, the GIoU function can solve the gradient problem due to the disjoint of two boxes, which is expressed as shown in equation (2).

$$L_{GIoU} = 1 - IoU(A, B) + \frac{|C - A \cup B|}{|C|} \qquad (2)$$

In equation (2), GIoU pays attention to other blank areas in addition to the overlap between the target box and the prediction frame, which can better reflect the overlap [22]. However, to improve the convergence speed, a penalty term is added, as shown in equation (3).

$$L_{DIoU} = 1 - IoU(A, B) + \frac{\rho^2(A_c, B_c)}{l^2} \qquad (3)$$

In equation (3), A_c and B_c represent the center point of the prediction and the target box, and l represents the diagonal distance in the minimum area of the target box and the prediction box, the convergence speed can be optimized if the two do not coincide [23]. The prediction box aspect ratio is taken into account, as shown in equation (4).

$$L_{CIoU} = 1 - IoU(A, B) + \frac{\rho^2(A_c, B_c)}{l^2} + \alpha v \qquad (4)$$

In equation (4), α represents the weight function. The expression of v is shown in equation (5).

$$v = \frac{4}{\pi^2}(\arctan \frac{w_A}{h_A} - \arctan \frac{w_B}{h_B})^2 \qquad (5)$$

In equation (5), v represents the similarity between the detection frame and the target frame. The purpose of this penalty term is to quickly complete the approximation of the length and width of the prediction box to the target box [24]. Additionally, to improve detection accuracy, CBAM is introduced, which is expected to reduce the environment interference on target detection. CBAM structure is shown in Figure 2.
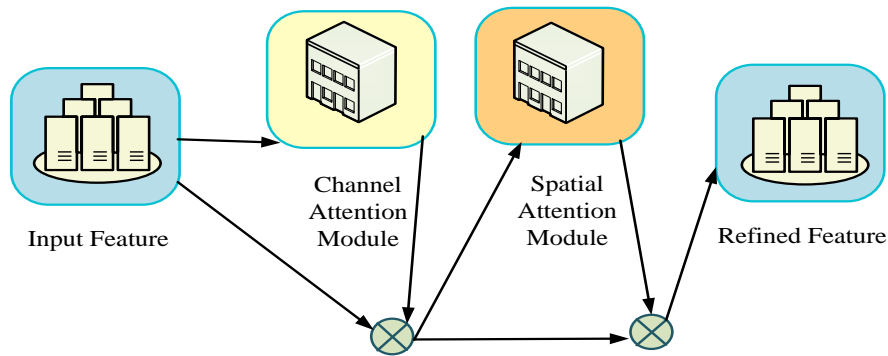
Figure 2: CBAM attention mechanism structure

In Figure 2, CBAM is an attention model using convolutional neural network, which realizes the weighted fusion and selection of input features by applying the self-attention mechanism and the channel attention mechanism on the feature map, to improve object detection accuracy and efficiency. The attention mechanism of CBAM consists of multiple sub-modules, such as spatial and channel attention module [25]. After feature map is input, weight attention map is generated through the two modules, and then weight values in the graph are multiplied by the values for the prominent feature map. Channel attention module is shown in Figure 3.
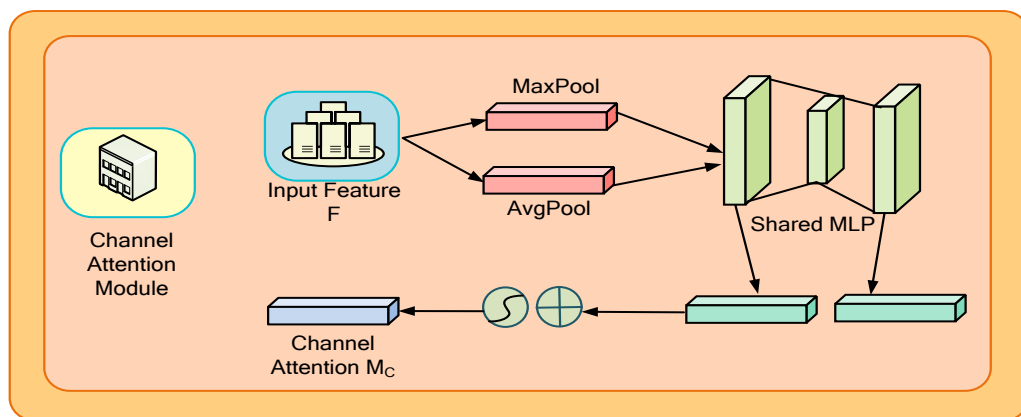

Figure 3: Channel attention module

Figure 3 illustrates a technique enhancing the performance of convolutional neural networks. In CBAM, channel attention mechanism first performs the maximum pooling and average pooling operations in the spatial dimension respectively to obtain the maximum and average values. Then, two vectors are connected and added together. Finally, the results are mapped to [0,1] through the sigmoid function to obtain channel attention vector. This vector can be used to weight the input feature map on important features. Spatial attention module is shown in Figure 4.
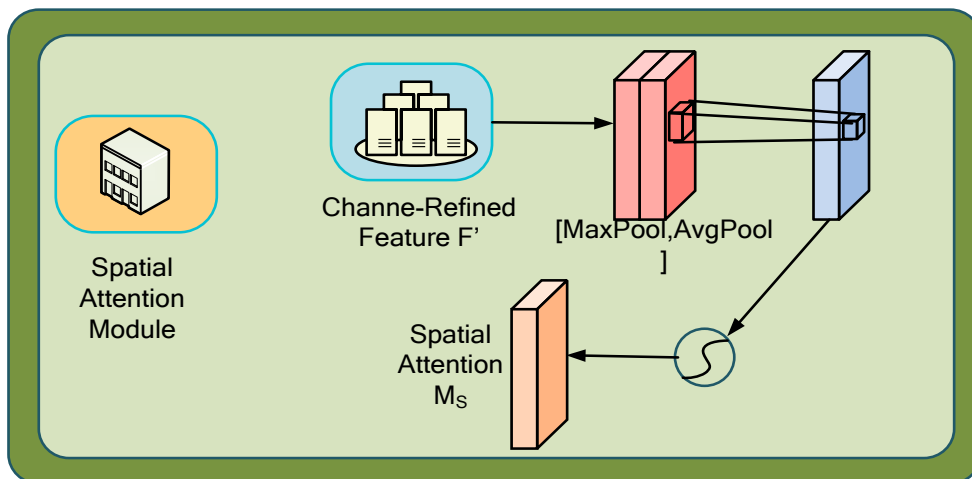
Figure 4: Spatial attention module

In Figure 4, two two-dimensional spatial feature maps are generated by pooling input feature map maximum value and the average value in channel dimension through spatial attention mechanism. Two spatial maps are spliced together according to the channels to form a new feature map with 2 channels. Then, a convolutional layer is used to reduce feature map dimensionality to generate a one-dimensional spatial attention map [26]. Finally, sigmoid function is used to process the spatial attention map to obtain final weight. Attention mechanism introduction on the basketball court can make the recognition algorithm cover more details of the player's body, and can improve the attention of the recognition algorithm to key information, so as to achieve a higher accuracy of the player's fixed-point shooting.

3.2 Construction and system design of basketball fixed-point shooting prediction model

Based on the object detection of basketball players, the pose estimation of the human body will be further studied, and the basketball fixed-point shooting hit prediction model will be constructed. OpenPose algorithm is a method based on deep learning, developed by Carnegie Mellon University in the United States. It can detect the joint points of all people in an image or video and connect these joint points to form a skeleton map of the human body. This algorithm has excellent robustness, is suitable for single and multi-person people and various scenarios like behavior recognition [27]. In view of this, based on the OpenPose algorithm, a trajectory optimization recognition method is proposed, which detects and recognizes the posture of basketball players, and finally uses the support vector machine algorithm for classification. Firstly, 18 human body joint points are selected as the output to predict the pose of the basketball player, and the labeling result of OpenPose is shown in Figure 5.
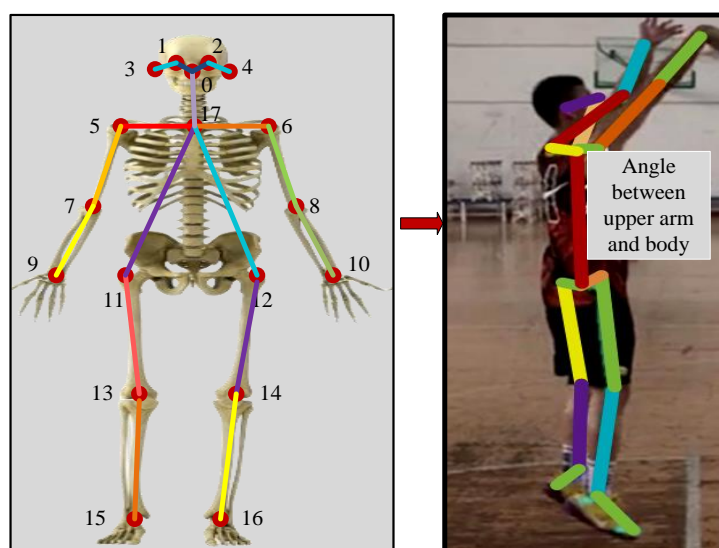

Figure 5: OpenPose joint output points

Figure 5 shows all the object data in the skeletal key annotation, where the skeletal keys are represented as an array of 2×K in length, where K is the total number of skeletal keys defined by that category. These 18 skeletal keys are connected by 19 connecting lines that form the torso, and these skeletal keys correspond to specific locations in the joints of the human body, including the nose, ears, elbows, eyes, wrists, shoulders, knees, hips, and ankles. Figure 6 shows the OpenPose network structure.
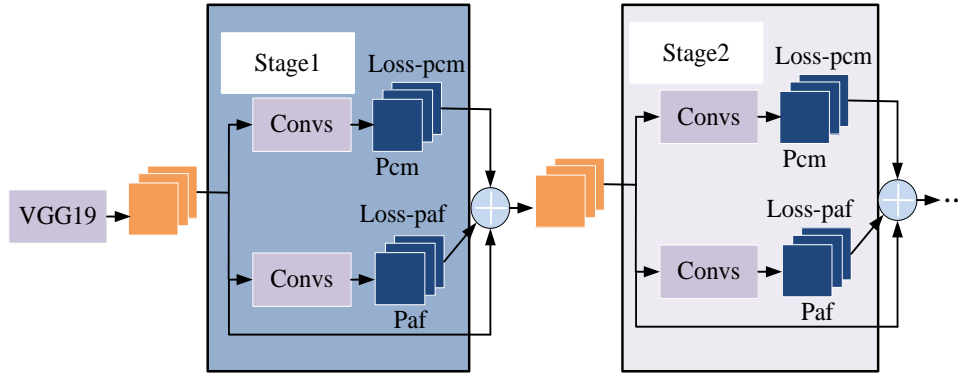

Figure 6: OpenPose network architecture diagram

In Figure 6, the VGG19 backbone network extracts image features, which are then passed through a series of stage modules. Each module has the same structure and function. These modules include two branches: one generates PCM, and the other generates PAF. Loss is calculated for PCM and PAF at each stage. While the first stage can theoretically output complete information, multiple stages are utilized in practice. This is because there is semantic information shared between key points, and later stages can optimize detection results based on the information extracted from previous stages [28]. Each network branch can be iterated, with the next branch using information from the previous stage as input to make predictions, as shown in equation (6).

$$\begin{cases} S^t = \rho^t(F, S^{t-1}, L^{t-1}), \forall t \geq 2 \\ L^t = \phi^t(F, S^{t-1}, L^{t-1}), \forall t \geq 2 \end{cases} \quad (6)$$

In equation (6), ρ and ϕ represent the network branches, which outputs a joint connection matching degree vector field $L = (L_1, L_2, L_3...L_m)$ and a key point prediction confidence graph $S = (S_1, S_2, S_3...S_j)$. The loss function of the two branch networks in the t stage is shown in equation (7).

$$\begin{cases} f_S^t = \sum_j^J \sum_P W(P). \| S_j^t(p) - S_j^*(p) \|_2^2 \\ f_L^t = \sum_{m=1}^M \sum_P W(P). \| L_m^t(p) - L_m^*(p) \|_2^2 \end{cases} \quad (7)$$

In equation (7), S_j^* represents the reliability diagram of the athlete's body position in the real data, L_m^* represents the affinity vector field, and W represents the base mask to avoid false penalties in special cases. To avoid gradient's disappearance, it is supplemented at each stage, as shown in equation (8).

$$f = \sum_{t=1}^T (f_S^t + f_L^t) \quad (8)$$

Equation (8) represents the final objective function. The human pose estimation method is prone to recognition errors, and the background and occlusion are regarded as joint nodes, so the misidentified joint points need to be repaired, and the expression is shown in equation (9).

$$d(k_1, k_2) = \sum_i \frac{y_i}{x_i} \quad (9)$$

In equation (9), k_1 and k_2 represent the gesture of adjacent frames, B_1^i and B_2^i represent the extracted bounding box of the body part. The feature points extracted from B_1^i are x_i and the feature points extracted from B_2^i are y_i. The similarity between the previous frame pose of the body and the current pose is expressed by equation (10).

$$Sc_{g,h}^i = \alpha * \sum \frac{n_i}{m_i} + (1-\alpha)* \| H_g - H_h \|_2 \quad (10)$$

In equation (10), m_i represents the number of feature points of the i joint point in the g frame, and n_i represents the number of feature points of the i joint point in the h frame. When the similarity is higher than the threshold set by the threshold, the joint points in the previous sequence are used as candidates, and if the similarity is lower than the set threshold, the joint point data of this frame is cleared. This study is carried out on the basis of the human posture recognition method to identify joint points position, and such methods need to record the body shape information of basketball players to avoid the inconsistency between the body shape information and the extracted feature points, resulting in the instability of the prediction model. In view of this, this study analyzes the angle change of athletes in the

process of fixed-point shooting to improve the prediction accuracy. In the process of shooting, the arm plays an important role, so the right wrist, elbow, and shoulder are used as the main feature information, and the coordinates of their joint points are expressed as $Rt(x_0, y_0), Rb(x_1, y_1), Rm(x_2, y_2)$ and the vector representation of the right forearm and right arm is shown in equation (11).

$$l_1 = (x_0 - x_1, y_0 - y_1), l_2 = (x_2 - x_1, y_2 - y_1) \quad (11)$$

In equation (11), l_1 represents the right forearm vector and l_2 represents the right arm vector. The angle between two vectors is denoted by equation (12).

$$\beta = \cos^{-1}(\frac{l1.l2}{\| l1 \| \| l2 \|}) \quad (12)$$

In equation (12), β represents the angle of the right arm. When the joint point of the hand is at the highest point, the angle between the small arm and the torso of the large arm reach the maximum, and the angle feature data is input into the classifier for prediction. The problem with the fixed-point shooting results is a classification problem, classified by hits or misses, and the study will employ a support vector machine algorithm. Support vector machines excel when working with complex datasets, especially for high-dimensional and large-scale data. Categorical learning can find a hyperplane, and the hyperplane that the sample divides in space is shown in equation (13).

$$\omega^T x + b = 0 \quad (13)$$

In equation (13), the $x = (x_1, x_2, x_3...x_n)$ represents angular feature data of the athlete's limb is denoted with a dimension of 7, $\omega = (\omega_1, \omega_2, \omega_3...\omega_n)$ represents the normal vector in the hyperplane, and b represents the bias term. The distance between the sample and the hyperplane in each shooting pose sample is shown by equation (14).

$$l = \frac{| \omega^T x + b |}{\| \omega \|} \quad (14)$$

In equation (14), l represents the distance between any sample X and the hyperplane. If the classification is correct, then equation (15) exists.

$$\begin{cases} \omega^T x_i + b \geq 1, y_i = +1 \\ \omega^T x_i + b \leq -1, y_i = -1 \end{cases} \quad (15)$$

In equation (15), when a fixed-point shooting is hit, the specimen will be above the superplane and vice versa. The basic expression of the support vector machine is shown in equation (16).

$$\begin{cases} \min_{\omega, b} \frac{1}{2} \| \omega \|^2 \\ s.t. y_i(\omega^T x_i + b) \geq +1, i = 1, 2, ...n \end{cases} \quad (16)$$

In equation (16), SVM satisfies the points in the sample set in equation (15), and the distance from these points to the hyperplane is the spacing. In this study, a prediction system will be designed based on the fixed-point shooting prediction model, and the system will be designed and implemented according to the system demand analysis. The main goal is to apply the object detection and pose estimation algorithms in the process of athletes' fixed-point shooting [29]. The system can be applied not only to the team, but also to individual basketball training to help athletes improve their set shooting skills. In addition, non-functional requirements such as operability, reliability, scalability, ease of maintenance, ease of use, and security of hardware devices need to be considered. The system adopts the B/S architecture to build the system, and its framework composition is shown in Figure 7.
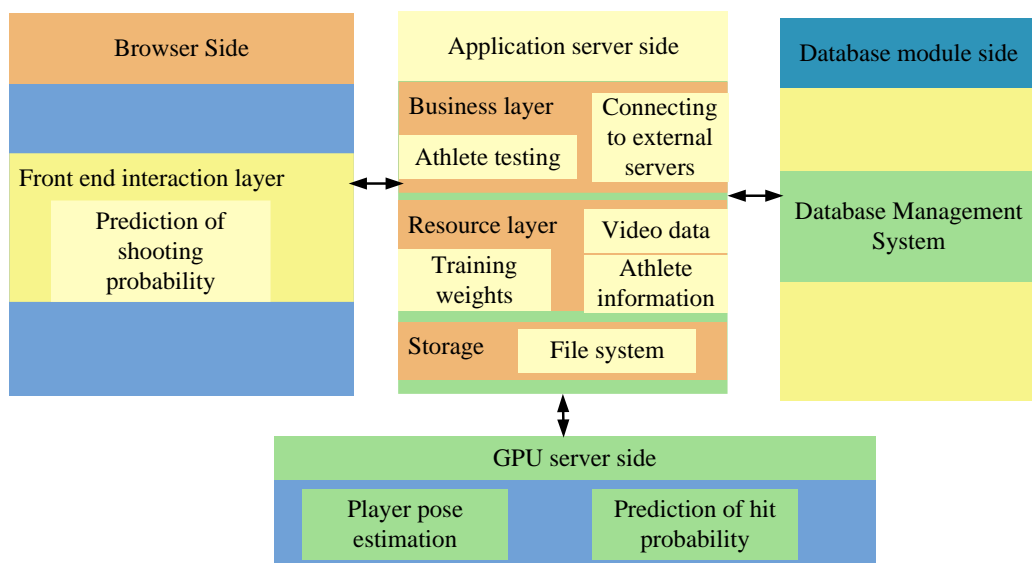

Figure 7: Framework diagram of fixed-point shooting probability prediction system

The framework in Figure 7 is mainly composed of an application service module, a browser module, a database module and a GPU server module, wherein the browser module contains a front-end interaction layer, which is mainly a fixed-point shooting probability prediction system. The fixed-point shooting hit prediction system is mainly composed of information input, result prediction, and recording prediction results, which can be browsed by users. The application server includes a business, a resource, and a storage layer [30]. The resource layer contains predictive training weights, shooting video data, and athlete information. The database module is used to store the prediction results.

4 Model performance evaluation and discussion of basketball fixed-point shooting hit prediction model

To verify the applicability and superiority of the prediction model, the performance was first tested, then the performance of the human pose estimation recognition algorithm was tested, and then prediction accuracy was verified. Finally, the functional module of fixed-point shooting prediction system was tested.

4.1 Target detection and target tracking performance test

In chapter 3, a basketball fixed-point shooting hit prediction algorithm based on human pose estimation algorithm was constructed. To verify the feasibility of the algorithm, a simulation experiment environment was built to analyze the algorithm. The simulation experimental environment of the research construction was based on the laboratory basic equipment, and the detailed information is shown in Table 2.

Table 2: Experimental environment setting

Hardware configuration		Software configuration	
CPU	AMD Ryzen 9 5950X	Operating system	CentOS 7
GPU	NVIDIA RTX3090	Programming environment	Python 3.8
RAM	Corsair Vengeance LPX 32GB (2 x 16GB) DDR4 3200	Simulation software	MATLAB R2021a
Storage device	Samsung 970 EVO Plus 1TB NVMe M.2 Internal SSD	Data set	NBA Player Movement Data

In the experiment, the NBA Player Movement Data was used as the experimental training and detection data set. The data was provided by NBA officials, including five types of basic information: timestamp, player position, position of the ball, player identity information and game information. The content of this data set would be constantly updated with the competition, in which there were about 50000

data related to the player posture, 30000 data as the network training data set and 20000 data as the network test set. YOLOv5 is a common deep learning object detection algorithm, and the human pose estimation algorithm is designed as the core. In the simulation experiment, the YOLOv5 deep neural network built was a deep learning network with three layers, and the specific parameters are shown in Table 3.

Table 3: Parameter setting of the YOLOv5 algorithm

Name	Value	Name	Value
Input image size	640	Pre data augmentation	True
Batch size	16	Anchor box	Automatically match datasets
Learning rate	0.01	Loss function	GIoU
Weight decay	0.005	Confidence threshold	0.25
Optimizer	Adam	Non maximum suppression threshold	0.45
Learning rate scheduler	Cosine LR schedule	Iterations	300

In the constructed YOLOv5 network, the lightweight convolutional neural network was used as the backbone, and it extracted the character action characteristics of the input image through the pyramid feature network structure. When extracting the action features, the high-level features of the image were combined with the low-level features, and the high-level semantic information and task action details of the image were extracted, which could effectively improve the extraction ability of the characters. In the above experimental environment, the model was analyzed with the parameters set in Table 3, and the results of the model are shown in Figure 8.

(a) Before improvement

(b) After improvement

(c) Before improvement

(d) After improvement

Figure 8: Detection results of the model before and after improvement

Figure 8 (a) and Figure 8 (b) represent the detection and identification results before and after the improvement of the YOLOv5 method, respectively, and the green boxes represent the detected athletes. The confidence level before the improvement was low, and the recognition effect was not significant. In motion, the basketball player in motion in the second frame was not detected by the previous method, while the improved method clearly captured the player's body. Figure 8(c) and Figure 8(d) illustrate the dynamic character tracking effect before and after the improvement of the YOLOv5 method, respectively. In Figure 8(c), one basketball player was missed due to the visual overlap of the two basketball players, while in Figure 8(d), the partially occluded basketball player was still detected. Results showed that the improvement effect was obvious. To further verify improved research method's superiority, it was compared with YOLOv5 and YOLOv5+DloU methods, and the loss value and average accuracy were compared in Figure 9.

(a) Loss curve of AC/DC ratio

(b) Target detection loss curve
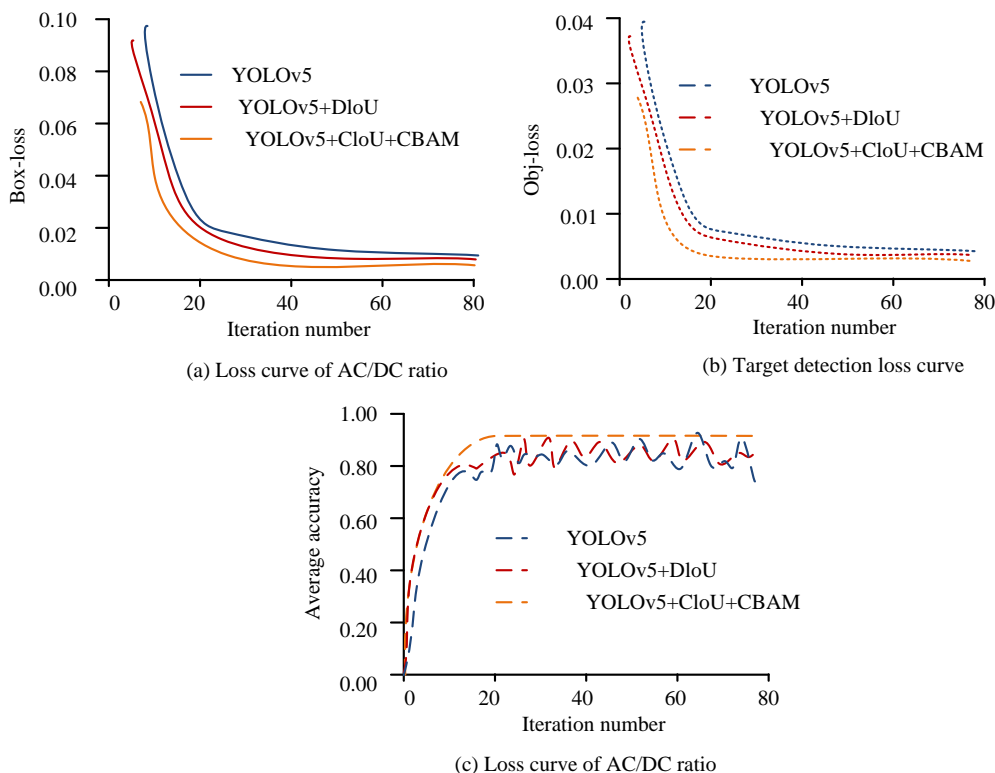
(c) Loss curve of AC/DC ratio

Figure 9: Loss curve and average accuracy curve

Figure 9 (a) and Figure 9 (b) represent the intersection and union ratio loss curve and the object detection loss curve, respectively. The abscissa counts iterations, and the ordinate represents the matching and detection loss of the object and the prediction frame, respectively. From the graph, the improved convergence speed was faster and the loss value was lower. Figure 9(c) represents the average accuracy curve which increased as iterations increased. When iteration of the research method was 50, the average accuracy tended to be stable, which was 95.34%. Compared with the two types of methods before the improvement, the research method fluctuation amplitude was significantly smaller, and the average accuracy was improved, which verified the rationality of the model improvement. Continuing to further compare these three methods, they were applied to the four-basketball game video Q, W, E, and R in the dataset, the changes in the recall rate are shown in Figure 10.
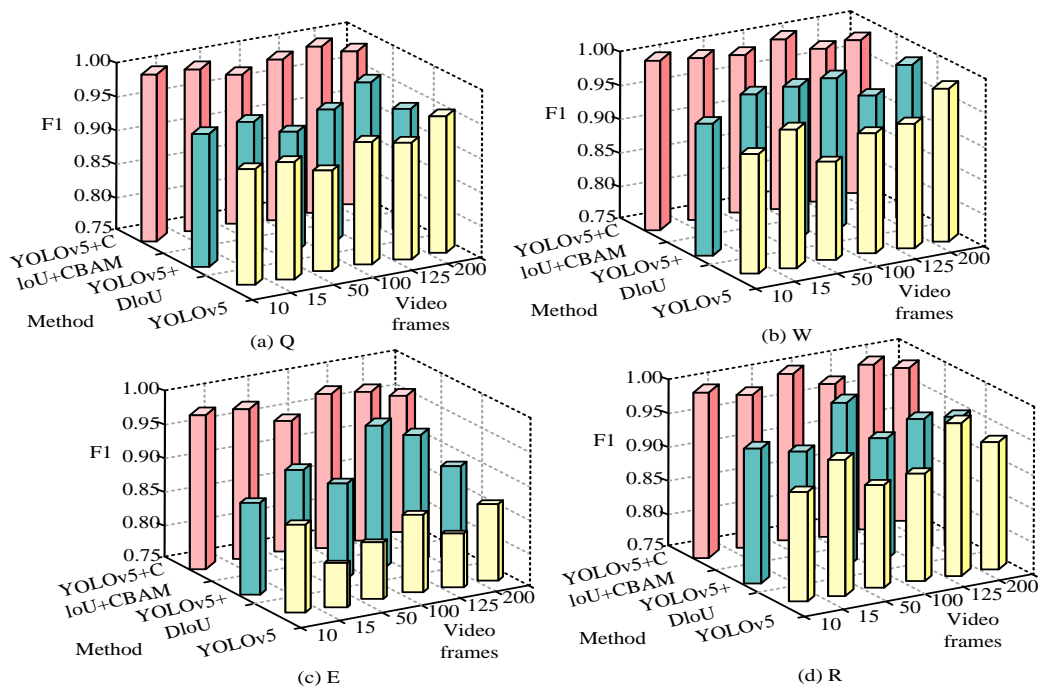
Figure10: Comparison chart of recall rate of three methods

In Figure 10, the abscissa represents the method category and the frames of the video, and six frames in the four videos were selected for analysis. The frame numbers were 10, 15, 50, 100, 125, and 200, respectively. The recall rates of Figure 10(c) were the lowest, which may be due to the fact that there are more spectators in the basketball game of E-video, and there was a certain occlusion of basketball players, but the research method still showed a high recall rate in Figure 10(c). The recall rate of the improved method was highest before the improvement, and the average recall rates in the four videos of Q, W, E, and R were 97.73%, 96.72%, 95.34%, and 97.98%, respectively.

4.2 Performance test of human posture algorithm

To test the human posture algorithm, the key frames in the fixed-point shooting process of basketball players were used as the input features of limb angles, and the precision, recall, F1 value and accuracy were selected as the evaluation indexes. The improved OpenPose method was compared with the evaluation indexes of OpenPose algorithm, PoseNet algorithm and Hourglass network algorithm in Table 4.

Table 4: Comparison of evaluation indicators using four methods

Method	Prediction	Recall	Precision	Accuracy	F1 value
OpenPose	In	85.45%	72.31%	81.23%	79.51%
	Out	76.27%	88.14%	81.34%	84.31%
PoseNet	In	84.48%	75.36%	83.45%	79.56%
	Out	77.23%	87.24%	87.63%	80.94%
Hourglass	In	88.45%	81.24%	86.35%	83.18%
	Out	82.57%	88.73%	83.68%	85.05%
Improved OpenPose	In	96.23%	87.16%	89.75%	88.19%
	Out	85.13%	95.72%	90.86%	89.53%

In Table 4, the improved effect was better, and all indicators were better than the improved OpenPose algorithm, with a recall rate of 96.23% for hit prediction and 85.13% for miss prediction, 87.16% for hit prediction and 95.72% accuracy for miss prediction, and 89.75% accuracy for hit prediction and 90.86% accuracy for miss prediction after improvement. The improved F1 value for hit prediction was 88.19% and 89.53% for miss prediction. PoseNet predicted the same as the OpenPose method, and Hourglass predicted better than them. The improved research method was superior, which illustrated the superiority of the research method. The experiment observed that different limb angle characteristics would affect the shooting percentage, so the study analyzed the effect of different characteristics on the prediction rate, as shown in Figure 11.

(a) The impact of the characteristics of various
parts of the limbs on accuracy
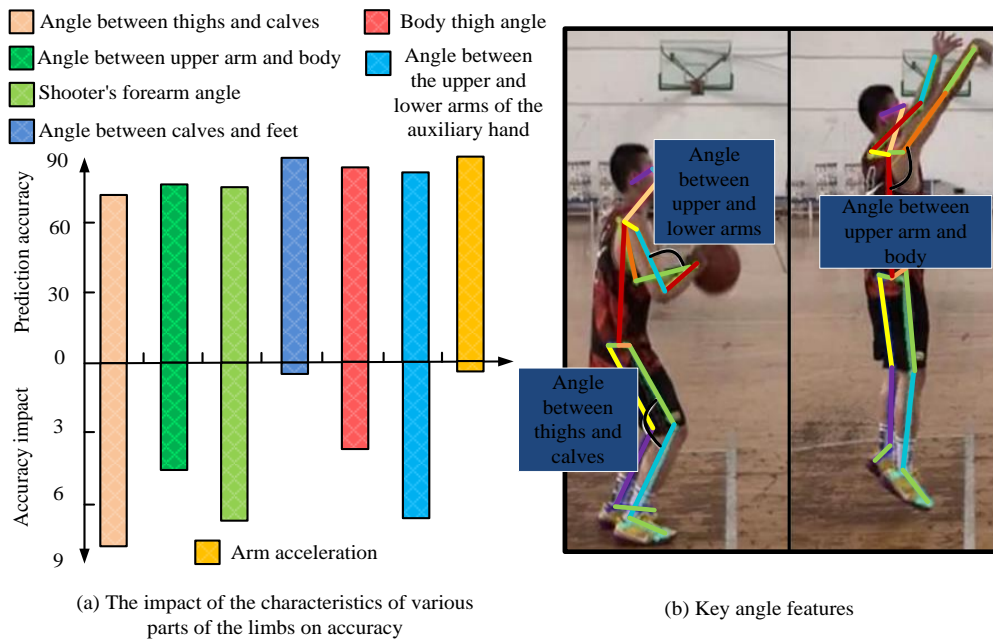
(b) Key angle features

Figure 11: The impact of different features on prediction rate

Figure 11(a) shows the characteristics effect of each part of the limb on the accuracy, and it can be seen that the calf foot angle and arm acceleration had the least impact on the accuracy, and their effect on the accuracy was within 1%. The limb characteristics that greatly influenced the accuracy of fixed-point shooting prediction were the angle between the shooting arm and the lower arm and the angle between the thigh and calf, reaching 7% to 8%. The second was the angle between the upper arm and the lower arm of the auxiliary hand, and the influence rate reached 6%. In the training process, if athletes want to improve the accuracy of fixed-point shooting, athletes need to pay attention to these three characteristics and control the size of the angle. Figure 11(b) shows the three types of limb features that have a greater impact on accuracy during shooting, and the angles of these three types of limb features can be clearly

seen. In summary, results not only showed a high accuracy of the research method, and performance of the research design human posture algorithm is superior, but also showed that the research method can point out the key characteristics of fixed-point shooting and provide suggestions for the improvement of basketball players' shooting training.

4.3 Performance test of fixed-point shooting prediction model

To verify research model superiority, Receiver Operating Characteristic (ROC) curve compared the predictive models: Extreme Gradient Boosting (XGBoost) model, Bagging model, and K-Nearest Neighbors (KNN) model. The comparison results are shown in Figure 12.
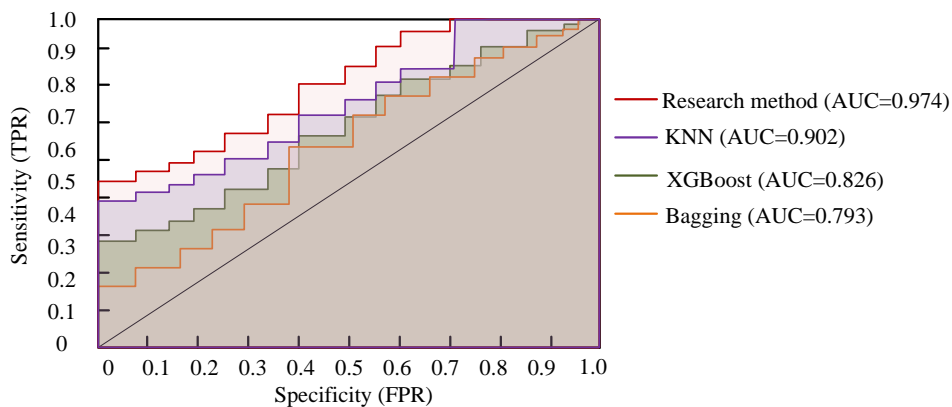

Figure12: Comparison of ROC curves for four methods

In Figure 12, ROC area of the indicator was between 0.1 and 1, which can intuitively evaluate the model accuracy, and the larger the Area Under Curve (AUC) value, the higher the model accuracy. The AUC value of the research model was the largest, reaching 0.974, which was very close to 1. The second was KNN, which had an AUC value of 0.902 and a higher accuracy. The AUC values of the remaining models ranged from 0.70 to 0.85, with average accuracy. The dataset was divided into five parts, and the F1 values of the four models were compared with the accuracy and recall rates, as shown in Figure 13.

(a) F1 value change
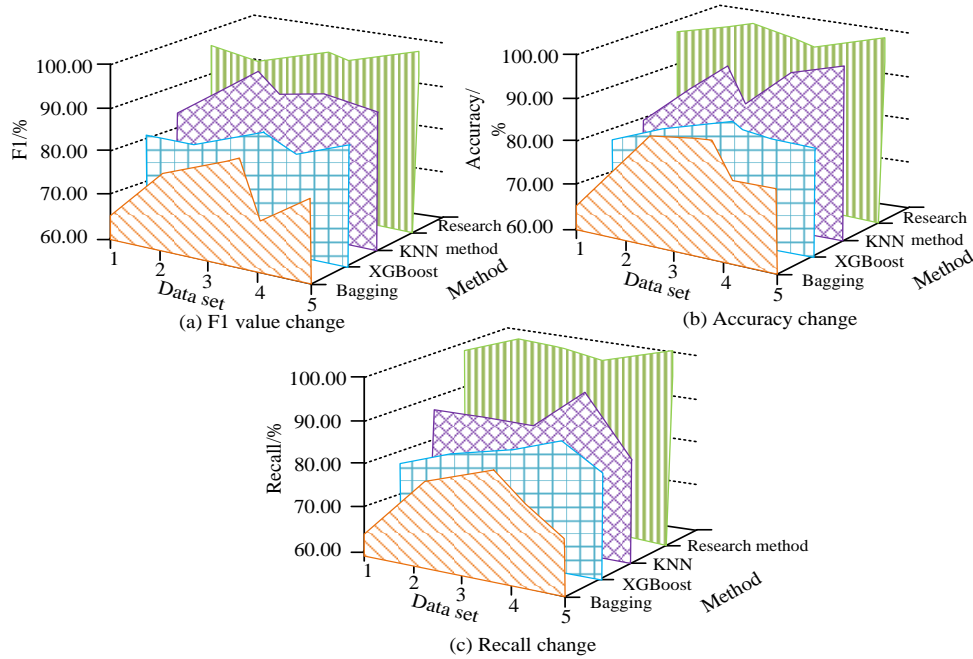
(b) Accuracy change

(c) Recall change

Figure 13: Comparison of F1 value, accuracy and recall rate of four models

In Figure 13, the F1 value, accuracy and recall of the research model were the highest. Specifically, the average values of the research model on the dataset were 95.54%, 96.39%, and 98.25%, which indicated that its performance was excellent. The performance from good to poor was: research model, KNN model, XGBoost model, and Bagging model. This once again proved the superiority of the research model in the prediction task. To verify the applicability of the fixed-point shooting prediction system of the study design, several modules of the system were tested and the browser results were displayed, as shown in Figure 14.

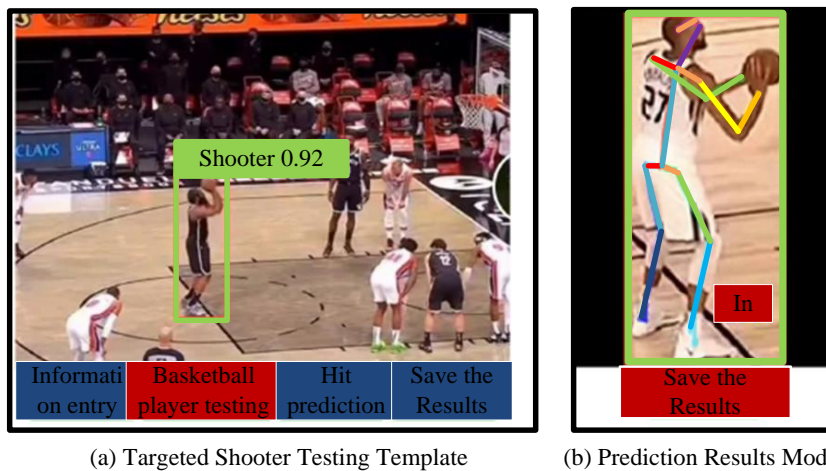(a) Targeted Shooter Testing Template   (b) Prediction Results Module

Figure 14: Performance testing of fixed-point shooting hit prediction system

Figure 14 (a) is the display of the shooting athlete detection module. There were options for information entry, athlete detection, hit prediction, and saving results in this interface, and the athlete prediction was marked in

red, which means that this option is selected, and the athlete in the picture was marked by a green box, and showed that the shooting hit rate was 0.92. Figure 14(b) shows the prediction results module. Because the prediction hit rate was 0.92, which is close to 1, and the system recognized this as a hit. The experimental results showed that the fixed-point shooting prediction system of the research design could operate normally, and the user could predict and save the results of the athlete's fixed-point shooting. The module is clear and easy to understand, which is extremely user-friendly.

4.4 Practical application and robustness analysis of the model

To verify the robustness of the study constructed model, two test experiments were designed for full system memory load and abnormal input, and the results are shown in Figure 15.
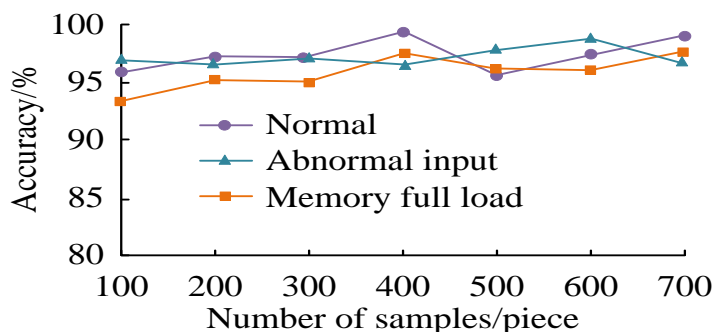

Figure 15: The robustness analysis of the algorithm

From Figure 15, the accuracy of this model was still in a steady state even in the environment of memory high pressure or abnormal data input. In the normal state, the highest accuracy of the model prediction could reach 98%. When there was abnormal data input, the accuracy of the model decreased, but it was still higher than 90%, and the highest accuracy could also reach 96%. The full load environment of the system had no effect on the model, and the overall prediction accuracy was about 95%. Gate Control Loop Unit (GRU) and Support Vector Machine (SVM) are the commonly used prediction algorithm, to compare the analysis of the algorithm in practical application, research in different games, using different shooting prediction algorithm, with the athletes shooting accuracy prediction analysis results as shown in Table 5.

Table 5: The practical application analysis of pose recognition field hit prediction algorithm

Game time	Proposed method		GRU		SVM	
	Time (s)	Accuracy (%)	Time (s)	Accuracy (%)	Time (s)	Accuracy (%)
1	1.5	95	3.1	88	3.3	84
2	1.2	94	2.5	87	3.1	85
3	1.3	93	2.6	85	3.2	86
4	1.4	95	2.8	89	3.5	89
5	1.3	96	2.8	85	3.4	88

In Table 5, whether in that game, the prediction accuracy of the proposed algorithm was always above 90% and reached 96%, while the prediction accuracy of GRU and SVM was only 89%. From the time-consuming comparison that the algorithm proposed by the study could give the prediction results within 2s, which greatly improved the prediction of the direction of the competition results, while both GRU and SVM algorithms needed a certain time to give the prediction results.

4.5 Discussion

Basketball is a ball game widely distributed in the world. This sport takes the number of shootings as the outcome judgment. In the continuous development of basketball, scholars have found that the athletes' shooting percentage can be predicted. Ozkan found that the athletes will have a certain degree of excitement changes before shooting, and different degrees of excitement changes correspond to different shooting posture. By analyzing the muscle excitement changes of the athletes during the shooting, it can predict the shooting posture that the athletes are about to adopt. Naik and Hashmi et al. found that the landing point of the object could be predicted by the initial

direction when analyzing the motion trajectory of the object. In basketball, by analyzing the shooting posture of the athletes, people can understand the movement state of the basketball, so as to speculate its landing point and determine whether the ball can be hit accurately. Therefore, the study proposed to use deep learning network to build a human posture recognition algorithm to analyze the movement state of basketball given by athletes and analyze the shooting percentage of athletes. The human posture recognition algorithm designed by the research could reach more than 90% with the accuracy and the recall rate of more than 95% when judging the posture of athletes, showing excellent recognition ability. The accuracy of the shooting prediction model based on the human posture recognition algorithm could reach more than 95%. The accuracy analysis of the shooting rate of the players can help the team coach to arrange tactics in advance. For example, according to the shooting simulation training of the players such as warm-up before the game, first is to analyze the shooting state of the players. Then, the depressed players are temporarily replaced to rest.

5 Conclusion

To predict the probability of fixed-point shooting, a prediction model of athletes' fixed-point shooting was constructed based on the YOLOv5 algorithm and OpenPose algorithm, and a prediction system was designed. Results showed that the YOLOv5 algorithm had a significant improvement effect, and the average accuracy of the improved YOLOv5 algorithm reached 95.34% when the number of iterations was 50. Applied to the research dataset, the average recall rates in the four videos of Q, W, E, and R were 97.73%, 96.72%, 95.34%, and 97.98%, respectively, and the detection and tracking effect on athletes was good. In the comparison of improved OpenPose with OpenPose, PoseNet, and the network, improved OpenPose performed better, with a recall rate of 96.23%, an accuracy of 87.16%, a precision of 89.75%, and an F1 value of 88.19%. In the influence analysis of different characteristics on the prediction rate, it is found that the influence rate of the three types of limb features, namely the angle between the shooting arm and the lower arm, the angle between the thigh and the calf, and the angle between the auxiliary arm and the lower arm, was larger, exceeding 6%. The results showed that basketball players need to pay attention to these three characteristics, control the size of the angle, and improve the shooting rate of fixed-point shooting. In comparison with the KNN model, XGBoost model and Bagging model, ROC area was used as the evaluation index, and research model's AUC value was the largest, which was 0.974. The F1 value, accuracy and recall rate of the research model were the highest, reaching 95.54%, 96.39% and 98.25%, respectively. The study designed system properly and clearly reflected the predicted results. The results verified research model's superiority, and

indicated that the research can provide a useful reference for tactical analysis and player performance evaluation in basketball games. However, there are still shortcomings in this study. The dataset selected for this study comes from the video of the competition, which contains a lot of redundant information and occupies a lot of training time.

6 Ethical compliance

The dataset used in the experiment is a publicly available dataset that does not involve athletes' personal privacy and data leakage and has no potential impact on athletes.

References

[1] A. Rodríguez-Fernández, R. Ramirez-Campillo, J. Raya-González, D. Castillo, and F. Y. Nakamura, "Is physical fitness related with in-game physical performance? A case study through local positioning system in professional basketball players. Proceedings of the Institution of Mechanical Engineers Part P Journal of Sports Engineering and Technology," vol. 237, no. 3, pp. 188-196, 2023. https://doi.org/10.1177/17543371211031160

[2] P. Soltani, and A. H. Morice, "A multi-scale analysis of basketball throw in virtual reality for tracking perceptual-motor expertise," Scandinavian Journal of Medicine and Science in Sports, vol. 33, no. 2, pp. 178-188, 2023. https://doi.org/10.1111/sms.14250

[3] Y. Ren, Z. Wang, Y. Wang, S. Tan, Y. Chen, and J. Yang, "GoPose: 3D human pose estimation using WiFi," Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 6, no. 2, pp. 1-25, 2022. https://doi.org/10.1145/3534605

[4] W. Liu, Q. Bao, Y. Sun, and T. Mei, "Recent advances of monocular 2D and 3D human pose estimation: A deep learning perspective," ACM Computing Surveys, vol. 55, no. 4, pp. 1-41, 2022. https://doi.org/10.1145/3524497

[5] L. Lonini, Y. Moon, K. Embry, R. J. Cotton, K. McKenzie, S. Jenz, and A. Jayaraman, "Video-based pose estimation for gait analysis in stroke survivors during clinical assessments: A proof-of-concept study," Digital Biomarkers, vol. 6, no. 1, pp. 9-18, 2022. https://doi.org/10.1159/000520732

[6] S. Dubey, and M. Dixit, "A comprehensive survey on human pose estimation approaches," Multimedia Systems, vol. 29, no. 1, pp. 167-195, 2023. https://doi.org/10.1007/s00530-022-00980-0

[7] S. Liu, N. He, C. Wang, H. Yu, and W. Han, "Lightweight human pose estimation algorithm based on polarized self-attention," Multimedia Systems, vol. 29, no. 1, pp. 197-210, 2023. https://doi.org/10.1007/s00530-022-00981-z

[8] X. Qin, H. Guo, C. He, and X. Zhang, "Lightweight human pose estimation: CVC-net," Multimedia Tools and Applications, vol. 81, no. 13, pp. 17615-17637, 2022. https://doi.org/10.1007/s11042-022-12245-z

[9] J. Xu, W. Liu, W. Xing, and X. Wei, "MSPENet: multi-scale adaptive fusion and position enhancement network for human pose estimation," The Visual Computer, vol. 39, no. 5, pp. 2005-2019, 2023. https://doi.org/10.1007/s00371-022-02460-y

[10] H. Jiang, "Application of deep learning method in automatic collection and processing of video surveillance data for basketball sports prediction," Arabian Journal for Science and Engineering, vol. 48, no. 3, pp. 4111-4112, 2023. https://doi.org/10.1007/s13369-021-05884-1

[11] D. G. Özkan, "Predicting the fate of basketball throws: An EEG study on expert action prediction in wheelchair basketball players," Experimental Brain Research, vol. 237, no. 12, pp. 3363-3373, 2019. https://doi.org/10.1007/s00221-019-05677-x

[12] D. Siemon, and W. Jörn, "Performance prediction of basketball players using automated personality mining with twitter data," Sport, Business and Management: An International Journal, vol. 13, no. 2, pp. 228-247, 2023. https://doi.org/10.1108/sbm-10-2021-0119

[13] B. T. Naik, and M. F. Hashmi, "LSTM-BEND: Predicting the trajectories of basketball," IEEE Sensors Letters, vol. 7, no. 4, pp. 1-4, 2023. https://doi.org/110.1109/LSENS.2023.3253863

[14] X. Wang, J. Tong, and R. Wang, "Attention refined network for human pose estimation," Neural Processing Letters, vol. 53, no. 4, pp. 2853-2872, 2021. https://doi.org/10.1007/s11063-021-10523-9

[15] Y. Liu, and X. Hou, "Fixed-resolution representation network for human pose estimation," Multimedia Systems, vol. 28, no. 5, pp. 1597-1609, 2022. https://doi.org/10.1007/s00530-022-00919-5

[16] X. Wang, R. Feng, H. Chen, R. Zimmermann, Z. Liu, and H. Liu, "Personalized motion kernel learning for human pose estimation," International Journal of Intelligent Systems, vol. 37, no. 9, pp. 5859-5879, 2022. https://doi.org/10.1002/int.22817

[17] J. Shi, and S. Kai, "A discrete-time and finite-state markov chain based in-play prediction model for NBA basketball matches," Communications in Statistics- Simulation and Computation, vol. 50, no. 11, pp. 3768-3776, 2021. https://doi.org/10.1080/03610918.2019.1633351

[18] X. Cong, S. Li, F. Chen, C. Liu, and Y. Meng, "A review of YOLO object detection algorithms based on deep learning," Frontiers in Computing and Intelligent Systems, vol. 4, no. 2, pp. 17-20, 2023.

https://doi.org/10.54097/fcis.v4i2.9730

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

[20] S. Pastel, J. Marlok, N Bandow, and K. Witte, "Application of eye-tracking systems integrated into immersive virtual reality and possible transfer to the sports sector-A systematic review," Multimedia Tools and Applications, vol. 82, no. 3, pp. 4181-4208, 2022. https://doi.org/10.1007/s11042-022-13474-y

[21] Q. He, X. Li, and W. Li, "Common sports injuries of track and field athletes using cloud computing and internet of things," International Journal of Computational Intelligence Systems, vol. 16, no. 1, pp. 70, 2023. https://doi.org/10.1007/s44196-023-00257-y

[22] S. Velugoti, and M. P. Vani, "An Approach for Privacy Preservation Assisted Secure Cloud Computation," Infomatica, vol. 47, no. 10, pp. 41-52, 2023. https://doi.org/10.31449/inf.v47i10.4586

[23] J. Fan, X. Yang, R. Lu, W. Li, and Y. Huang, "Long-term visual tracking algorithm for UAVs based on kernel correlation filtering and SURF features," The Visual Computer, vol. 39, no. 1, pp. 319-333, 2023. https://doi.org/10.1007/s00371-021-02331-y

[24] V. Gali, B. C. Babu, R. B. Mutluri, M. Gupta, and S. K. Gupta, "Experimental investigation of harris hawk optimization-based maximum power point tracking algorithm for photovoltaic system under partial shading conditions," Optimal Control Applications and Methods, vol. 44, no. 2, pp. 577-600, 2023. https://doi.org/10.1002/oca.2773

[25] M. Dunnhofer, A. Furnari, G. M. Farinella, and C. Micheloni, "Visual object tracking in first person vision," International Journal of Computer Vision, vol. 131, no. 1, pp. 259-283, 2023. https://doi.org/10.1007/s11263-022-01694-6

[26] D. Yang, "Research on multi-target tracking technology based on machine vision," Applied Nanoscience, vol. 13, no. 4, pp. 2945-2955, 2023. https://doi.org/10.1007/s13204-021-02293-6

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

[28] K. Aygül, M. Cikan, T. Demirdelen, and M. Tumay, "Butterfly optimization algorithm based maximum

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

[29] A. Chessa, P. D'Urso, L. De Giovanni, V. Vitale, and A. Gebbia, "Complex networks for community detection of basketball players," Annals of Operations Research, vol. 325, no. 1, pp. 363-389, 2023. https://doi.org/10.1007/s10479-022-04647-x

[30] H. Mokayed, T. Z. Quan, L. Alkhaled, V. Sivakumar, "Real-time human detection and counting system using deep learning computer vision techniques," Artificial Intelligence and Applications, vol. 1, no. 4, pp. 221-229, 2023. https://doi.org/10.47852/bonviewAIA2202391