Hawk-Eye Deblurring and Pose Recognition in Tennis Matches Based on Improved GAN and HRNet Algorithms

Weixin Zhao

Department of Physical Education and Sports Science, Fuzhou University, Fuzhou, 350108, China

Abstract—In tennis matches, the Hawk-eye system causes blurry trajectory judgment and low accuracy in player posture recognition due to rapid movement and complex backgrounds. Therefore, the research improves the backbone network and iterative attention feature fusion mechanism of deblur generative adversarial network version. At the same time, Ghost, Sandglass module, and coordinate attention mechanism are used to optimize the high-resolution network, and a new model for deblurring and pose recognition of Hawk-eye images in tennis matches is proposed by integrating the improved generative adversarial network and high-resolution network. The new model achieved an information entropy value of 11.2, a peak signal-to-noise ratio of 29.74 decibels, a structural similarity of 0.89, a minimum parameter size of 4.53, and a running time of 0.25 seconds on the tennis tracking dataset and the Max Planck Society human posture dataset, which was superior to current advanced models. The highest accuracy of deblurring and pose recognition for the model under different lighting intensities was 92.44%, and the highest improvement rate of video frame quality was 18%. From this, the model has significant advantages in deblurring effect, posture recognition accuracy, parameter quantity, and running time, and has high practical application potential. It can provide an advanced theoretical reference for tennis match refereeing and technical training.

Keywords—DeblurGANv2; HRNet; tennis; hawk-eye system; deblurring; pose recognition

I. INTRODUCTION

In tennis matches, the Hawk-eye system, as a high-precision technology application, has been widely used in referee decision-making, motion analysis, and other fields worldwide. The Hawk-eye system captures image data within the field through multiple high-speed cameras and uses image processing algorithms for real-time calculation and analysis, providing judgments on whether the ball is within or outside the boundary. In addition, the movements of players in tennis matches are complex and have high spatiotemporal variations. Accurately capturing the players' motion posture is the key to improving the accuracy of Hawk-eye system judgment [1-2]. Pose recognition technology can accurately capture the movement information of athletes by analyzing their body position, movement trajectory, etc., providing a more precise player behavior model for the Hawkeye system and improving the reliability of judgment results. F. Meng et al. introduced a hybrid neural network to optimize target feature extraction and constructed a novel Hawk-eye detection model to improve the visual detection level of tennis in the sports industry. The

model achieved a tennis motion tracking accuracy of 0.694 under grayscale feature conditions, which was the highest among all testing methods [3]. Y. Zhao et al. proposed a lightweight tennis Hawk-eye detection scheme combining "You Only Look Once version 5" (YOLOv5) to address the inefficiency of traditional tennis detection algorithms. Compared with traditional methods, experimental results showed that this algorithm reduced model parameters by 42% and computational complexity by 44%, while improving detection accuracy by 2% [4]. Y. Yang et al. built a new tennis trajectory prediction method by combining artificial neural network detection algorithm and stereo vision. The experiment showed that this method had high reliability and robustness, effectively improving the prediction ability of tennis trajectory [5]. D Gao et al. built a deep learning driven small object automatic detection method to address the difficulty of small object detection in tennis videos. The experiment showed that this method performed well in the integrity, recognition accuracy, and detection speed [6]. Y. Ke et al. proposed an object detection algorithm on the basis of deep learning aimed at handling advanced visual tasks such as tennis. This algorithm combines prior knowledge of tennis impact areas. The experimental results showed that it could provide high detection accuracy and faster detection speed, effectively improving the accuracy and stability of tennis impact detection [7].

Generative Adversarial Networks (GANs) are powerful deep learning models that have shown great potential in tasks such as image generation, denoising, and deblurring [8]. Bian J et al. argued that detecting dense movements from fast-moving objects in sports videos remained challenging. To this end, a novel table tennis detection model by combining GAN and P2ANet was proposed. The model could achieve an average accuracy of 88.47% for the localization and recognition of 8 types of table tennis movements, while improving the detection robustness [9]. Ghezelsefloo H R et al. proposed an auxiliary calibration model for Hawk-eye detection after improving the GAN algorithm to effectively reduce the error in Hawk-eye detection in sports events. The model had a success rate of 92.17% in assisting correct judgments in 130 sports, and had significant practical value [10]. Peng X et al. constructed a video pose detection model for ball players by combining sensor image acquisition with GAN and Modbus. The performance of the model on Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM) was about 4.5 and 0.143 higher than other algorithms, respectively [11]. In

addition, High-Resolution Network (HRNet) is an excellent deep learning model that excels in processing detailed information in high-resolution images and is widely used in fields such as image segmentation, object detection, and pose recognition. Nguyen H C et al. proposed an automatic combined human pose estimation model by combining HRNet and YOLOv5 to improve the accuracy of human motion pose estimation. The processing time on a 3.3-megapixel dataset was 55FPS, and the highest accuracy of human keypoint detection was 98.24% [12]. Li Y combined HRNet to construct a monocular video motion capture method, which optimized it for human motion reconstruction problems such as floating, ground penetration, and sliding. This method achieved a good balance between accuracy and frame rate, and had significant detection advantages [13]. Fitzpatrick A et al. In order to strengthen the accuracy of the hawk-eye monitoring system under different serve and return strategies, the researchers proposed a hawk-eye-assisted detection model with multimodal data training and convolutional graph neural network processing. The experimental results show that the method achieves higher detection accuracy and greater stability for a variety of different serving and hitting motions [14].Ning T et al. In order to address the limitations of computer vision-assisted table tennis ball detection, the researchers proposed a real-time computational method for determining the landing point of a table tennis ball. The experimental results showed that the method achieved a detection speed of 45.3 fps, and the key frame extraction method correctly recognized the landing point frames with an accuracy rate of more than 93.3% [15].

In summary, traditional Hawk-Eye systems mostly rely on classical deblurring algorithms, but these methods usually cannot effectively deal with fast motion and multi-angle shooting conditions, resulting in image distortion and inaccurate pose estimation. In addition, while existing pose recognition methods are able to achieve better results in static or slow scenes, they still exhibit large errors in dynamic tennis match scenarios, especially when players are moving fast. Although several approaches have been dedicated to solving this problem, existing solutions usually face certain limitations. For example, traditional algorithms are computationally inefficient when dealing with large-scale data and insufficiently adaptable when facing complex environments. In order to overcome these limitations, the study innovatively proposes a novel hawk-eye deblurring and pose recognition model for tennis matches, which incorporates the improved deblur generative adversarial network version 2 (DeblurGANv2) and HRNet algorithms, respectively, and introduces a lightweight Mobilenetv2 backbone network, Ghost module and Sandglass module are introduced to improve the computational efficiency, and Iterative Attentionla Feature Fusion (IAFF) and Coordinate Attention (CA) mechanisms are adopted to enhance the feature extraction capability. Enhance the feature extraction ability, and at the same time significantly improve the processing speed and robustness of the algorithm, especially in the complex environment of the adaptability of the excellent performance. Among them, the improved deblurring technique of GAN can better handle blurred images under different motion states while ensuring image quality. Combined with the high-resolution feature of HRNet, the accuracy of pose recognition is further improved, especially in the capture of complex motion and action details. The research aims to significantly improve image clarity and pose recognition accuracy in dynamic scenes by combining these innovative designs, providing an effective solution for efficient and real-time tennis match Hawk-eye systems. This research is divided into four parts, the first part is the analysis and summary of others' research, the second part describes how the Hawk-Eye image deblurring algorithm for tennis matches and the tennis match stance recognition model were designed, respectively, while the third part tests the performance of the model, and the last part is the summary of the article.

II. METHODS AND MATERIALS

In response to the challenges of image blur and athlete pose recognition in tennis matches, this study first introduces IAFF based on DeblurGANv2 and uses Feature Pyramid Network (FPN) to achieve bidirectional fusion of multi-scale features. Secondly, based on HRNet, Ghost, Sandglass, CA mechanism, and Transformer-based object tracking module are sequentially introduced to propose a new Hawk-eye analysis model that integrates deblurring and pose recognition.

A. Deblurring Algorithm for Hawk-eye Images in Tennis Matches Based on Improved GAN

Image blur is one of the main issues affecting the accuracy of Hawk-eye system judgment, especially during the rapid movement of players and the high-speed flight of the ball [16-17]. The traditional image degradation model mainly generates degraded images from the original image after degradation function and noise processing, while the restoration model restores clear images close to the original image by applying restoration functions to the degraded image. Image degradation is usually caused by factors such as motion blur and poor lighting [18-20]. Through this approach, a classic algorithm for image motion blur, DeblurGANv2, is introduced into the study. This algorithm efficiently removes motion blur through the improved GAN. Compared with traditional deblurring algorithms, DeblurGANv2 has adaptability to complex backgrounds and multi-scale feature extraction ability, which can more comprehensively restore image details and is suitable for dynamic motion scenes with large changes [21-23]. In order to adapt to tennis motion detection and improve universality, the structure of DeblurGANv2 is improved, and an improved DeblurGANv2 tennis match Hawk-eye image deblurring algorithm is proposed. The framework of this algorithm is shown in Fig. 1.
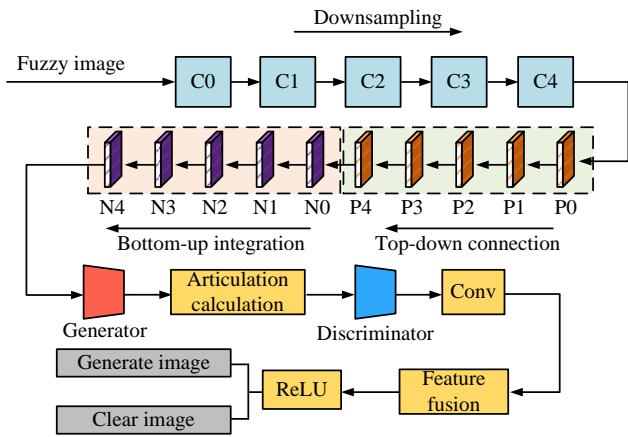
Fig. 1. Improved framework of the Hawk-eye image deblurring algorithm for tennis match in DeblurGANv2.

In Fig. 1, the algorithm framework mainly has generator, discriminator, and attention mechanism modules, and includes specific operations such as convolution, upsampling, downsampling, feature fusion, stacked convolution, batch normalization, and ReLU activation. Firstly, the improved backbone network is used to downsample and extract five feature maps of different scales step by step, from C0 to C4. Then, these feature maps are generated using the top-down connection of the FPN, namely P0 to P4. Next, P0 to P4 gradually perform bottom-up feature fusion to obtain feature maps, namely N0 to N4. Afterwards, N0 to N4 are fused with the original image to generate the final deblurred image. The generated image is then input into the discriminator along with the clear image to calculate the clarity probability of the generated image, in order to optimize the generator. Finally, the generator and discriminator are alternately trained and output after convolution, feature fusion layer, and ReLU activation. Compared with the improved DeblurGANv2, a bottom-up feature fusion branch is added, allowing low-level features to fully interact with high-level features. Specifically, the feature fusion path is designed through the bidirectional connection of FPN, which first performs top-down connection and then performs bottom-up fusion. The calculation formula is shown in Eq. (1).

$$P_i = Conv_{1\times1}(C_i) + Upsample(P_{i+1}) \tag{1}$$

In equation (1), P_i represents the intermediate feature map fused from top to bottom. C_i represents feature maps of different scales extracted from the backbone network. $Conv_{1\times1}$ signifies a 1×1 convolution operation. *Upsample* represents upsampling operation. Next, P_i is subjected to bottom-up feature fusion to enhance the information transmission of cross layer features, as shown in Eq. (2).

$$N_i = Conv_{3\times3}(P_i) + Downsample(N_{i-1}) \tag{2}$$

In Eq. (2), N_i represents the fused feature map. *Downsample* represents downsampling operation. To further enhance the ability to focus on key regions, the improved

algorithm adopts the IAFF mechanism. IAFF calculates attention weights through multiple iterations to focus on important regions in the image [24-26]. The calculation process is shown in Eq. (3).

$$\begin{cases} F_{IAFF}^t = F^t + \Gamma \cdot Attention(Q^t, K^t, V^t) \\ Attention(Q^t, K^t, V^t) = Soft\max(\dfrac{Q^t K^{t^T}}{\sqrt{d}})V^t \end{cases} \tag{3}$$

In Eq. (3), F_{IAFF}^t represents the fused feature map after the t-th iteration. F^t signifies the input feature of the t-th iteration. Γ represents the fusion coefficient. Q^t, K^t and V^t signify the query, key, and value matrices for the t-th iteration, respectively. $\sqrt{d}$ represents the scaling factor, used to avoid excessive attention weights. In addition, to accelerate image deblurring processing, the improved algorithm uses Mobilenetv2 as the backbone network, replacing the traditional heavy convolutional network. Mobilenetv2 uses depthwise separable convolution, which contains two parts: depthwise convolution and pointwise convolution. The calculation formula is shown in Eq. (4) [27-29].

$$DSConv(x) = DepthwiseConv(x) + Po\operatorname{int}wiseConv(x) \tag{4}$$

In Eq. (4), $DSConv(x)$ represents a depthwise separable convolution operation. $DepthwiseConv(x)$ represents convolution only in the spatial dimension. $Po\operatorname{int}wiseConv(x)$ represents using 1×1 convolution to fuse features in the channel dimension. This convolution method significantly reduces the computational and parameter complexity, as shown in Eq. (5).

$$FLOPs_{DSConv} = \frac{1}{k^2} \times FLOPs_{S\tan dardConv} \tag{5}$$

In Eq. (5), k signifies the size of the convolution kernel. In summary, the model generator and discriminator network structure of the improved DeblurGANv2 are shown in Fig. 2.

Fig. 2 (a) displays the improved generator structure of DeblurGANv2, and Fig. 2(b) displays the improved discriminator structure of DeblurGANv2. In Fig. 2 (a), the generator includes multiple layers of feature extraction modules. Firstly, the main network performs downsampling to extract feature maps of different scales layer by layer. Then, FPN is used for multi-scale feature fusion, adopting a bidirectional connection design of top-down and bottom-up. In the feature map processing at each scale, convolutional layers, Batch Normalization (BN) layers, and ReLU activation functions are used to enhance feature extraction performance, and IAFF mechanism is adopted to highlight key regions. Finally, the fused feature map is upsampled and residual connected to reconstruct a blurred image. As shown in Fig. 2 (b), the discriminator structure includes a series of convolutional layers, Leaky ReLU activation functions, and BN layers, which are used to extract high-level features of the input image. The discriminator adopts a layer by layer downsampling design, gradually compressing the image size

through multiple convolutional layers. Finally, the fully connected layer is applied to calculate the probability score between the generated image and the real clear image, to determine whether it is a real image.

To further improve the restoration effect and image detail preservation ability of DeblurGANv2 in deblurring tasks. A mixed loss function is designed, including adversarial loss,

perceptual loss, and image reconstruction loss. Firstly, the adversarial loss is used to optimize the game between the generator and discriminator, making the output image of the generator more realistic. Secondly, the perceptual loss is applied to measure the differences in high-level semantic features between generated images and real clear images, as shown in Fig. 3.

(a) Generator structure

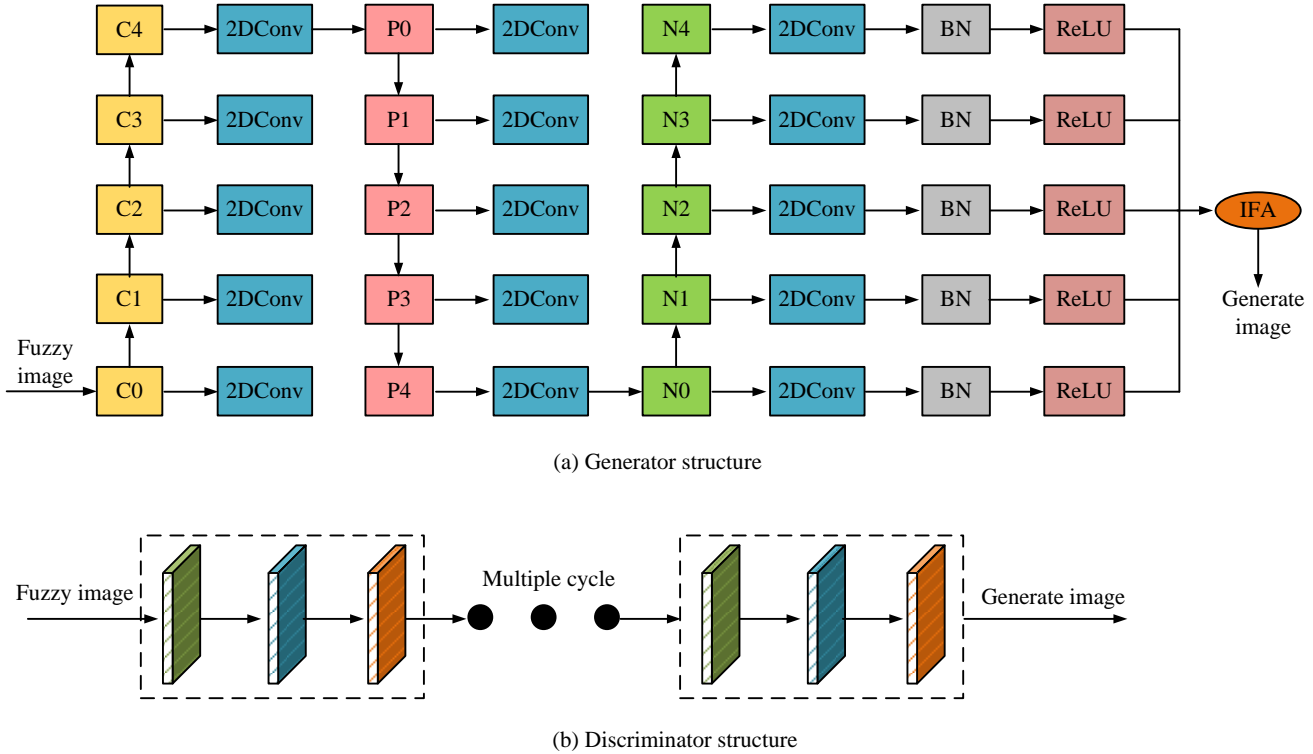(b) Discriminator structure

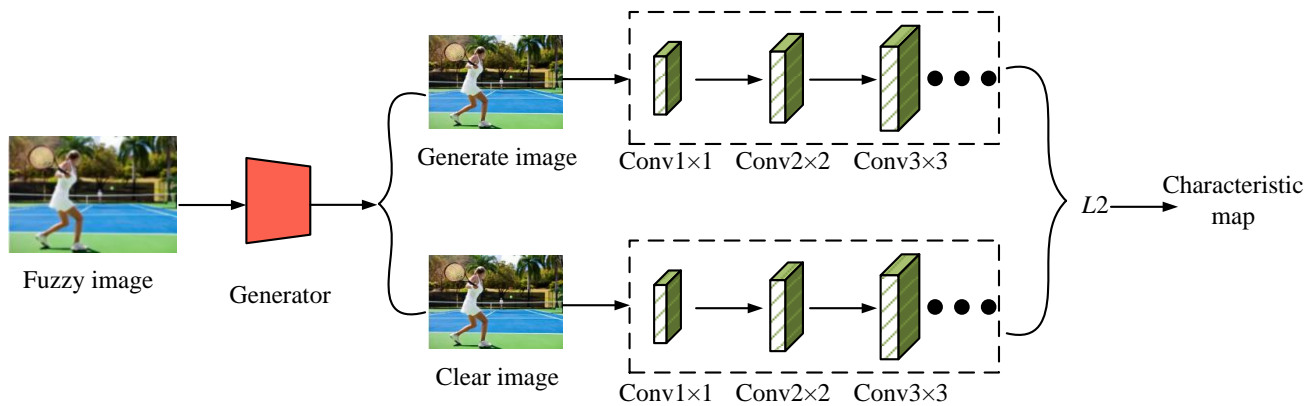Fig. 2. Improved generator and discriminator structure for DeblurGANv2.

Fig. 3. Schematic diagram of perceptual loss.

As shown in Fig. 3, the calculation of perceptual loss is achieved by introducing a pre-trained deep Convolutional Neural Network (CNN) to extract high-level features, and comparing the feature differences between the generated image and the real clear image at different convolutional layers. Specifically, the input blurred image is first deblurred by a generator to generate a restored image. Then, the generated

image and the real clear image are input into a deep CNN, and feature maps are extracted through several layers of convolution. In these feature maps, the perceptual loss calculates the difference in $L2$-norm between the generated image and the real image on each layer of the feature map, as shown in Eq. (6).

$$L_{perc} = \sum_l \lambda_l \left\| \phi_l(G(x)) - \phi_l(y) \right\|_2^2 \qquad (6)$$

In Eq. (6), L_{perc} represents perceptual loss, which is applied to measure the difference in high-level features between the generated image and the real image. λ_l signifies the weight coefficient. ϕ_l represents the feature mapping of the l-th layer of the deep CNN. $G(x)$ represents the deblurred image output by the generator. y represents the real and clear image. $\|\cdot\|_2$ represents the $L2$-norm. In order to capture more levels of semantic information, perceptual loss usually selects feature maps from multiple convolutional layers for calculation, and the comprehensive formula is shown in Eq. (7).

$$L_{perc} = \sum_{l=1}^{L} \left(\frac{1}{H_l W_l C_l} \sum_{h=1}^{H_l} \sum_{w=1}^{W_l} \sum_{c=1}^{C_l} (\phi_l(G(x))_{h,w,c} - \phi_l(y)_{h,w,c})^2 \right) \qquad (7)$$

In Eq. (7), L represents the number of selected convolutional layers. H_l, W_l and C_l respectively represent the height, width, and number of channels of the l-layer feature map. $f_l(G(x))_{h,w,c}$ represents the feature values at position (h,w) and channel c in the l-layer feature map. $\phi_l(y)_{h,w,c}$ signifies the feature value of the corresponding position of the real image in the l-th layer feature map.

B. Construction of Tennis Match Pose Recognition Model Integrating Improved HRNet Algorithm

After improving the DeblurGANv2 structure, the blurring effect of Hawk-eye monitoring images is effectively avoided. This research further focuses on the task of athlete posture recognition in tennis matches. Pose recognition is an important part of technical analysis in tennis matches, which is crucial for the standardization analysis of player movements, optimization of game strategies, and monitoring of potential rule violations

[30-31]. However, the rapid changes in player movements, complex postures, and dynamic background interference in tennis matches make traditional posture recognition methods difficult to cope with [32-33]. To address these issues, the HRNet algorithm is combined with research. Compared with other advanced methods, it consistently maintains high-resolution feature maps throughout the entire feature extraction process and fully utilizes multi-scale information through layer by layer fusion of multi-resolution features. In addition, targeted optimization and improvement are carried out on the basis of the standard HRNet architecture to make it more suitable for the scene requirements of tennis matches. The improved HRNet is displayed in Fig. 4.

In Fig. 4, the improved HRNet has four stages, each stage achieving the extraction and fusion of multi-scale features through parallel resolution branches. The first stage uses standard convolution operations for preliminary feature extraction, generating high-resolution feature maps. In the second stage, while retaining high-resolution branches, low resolution branches are applied to capture deeper feature information through downsampling. In the third stage, more resolution branches are added to achieve multi-scale feature alignment and complementarity from high resolution to low resolution. In the fourth stage, a cross resolution feature fusion strategy is used to effectively combine feature information from different resolutions, generating a multi-scale feature map with global context awareness capability. Specifically, there are four major improvements. First, Ghost and Sandglass have been introduced to replace the Bottleneck and Basicblock modules in HRNet, reducing the running parameter. Second, the introduced CA enhances the feature extraction capability of the model. Third, the ability to enhance data has been improved through unbiased data augmentation methods. Fourth, the effectiveness of object detection is improved through a separate object tracking module. The target tracking module is shown in Fig. 5.
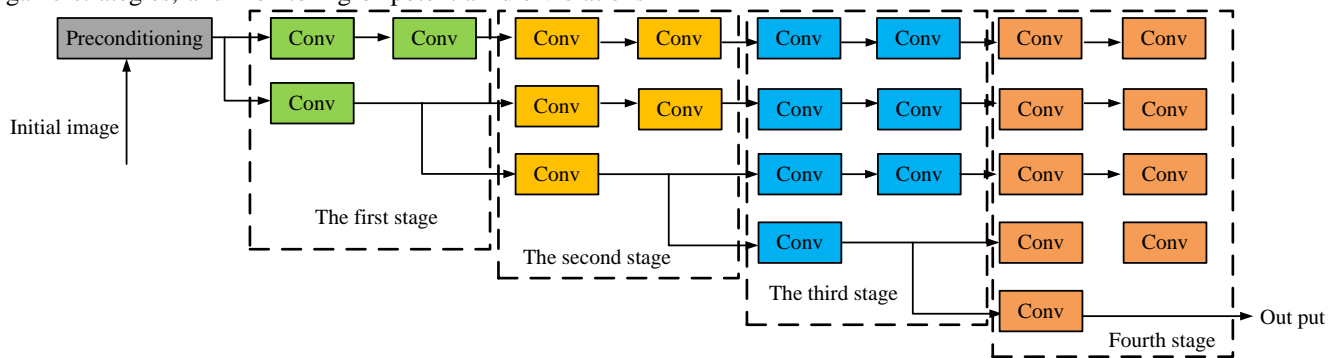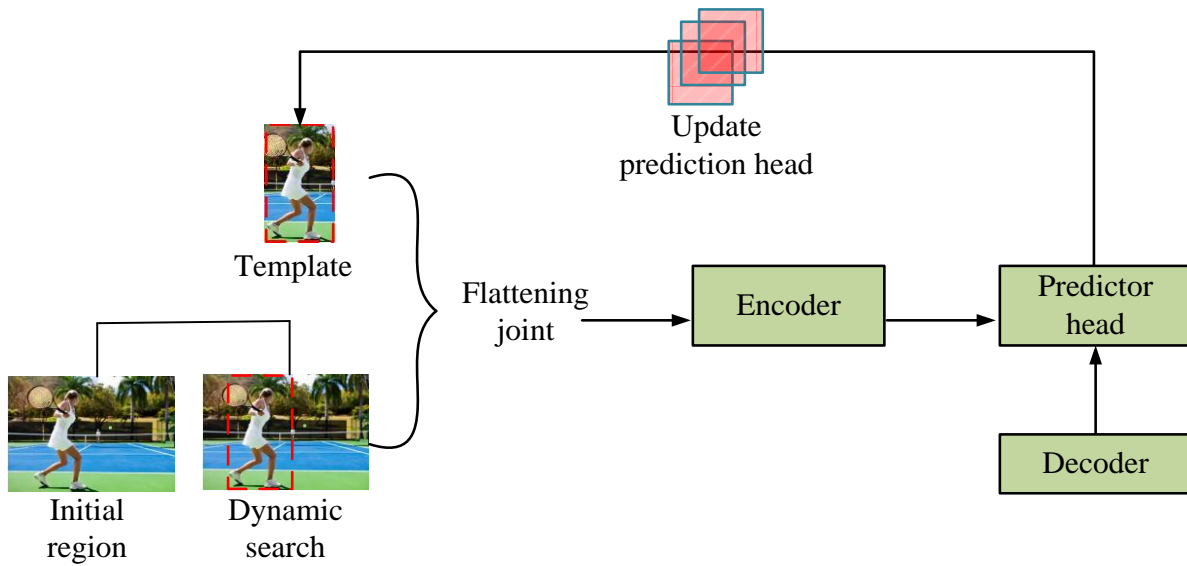

Fig. 4. Improved HRNet structure.

Fig. 5. Target tracking module diagram.

As shown in Fig. 5, the target tracking module mainly includes three core parts: target detection, target association, and trajectory update, which are also key steps based on the Transformer target tracking algorithm. Firstly, the input video frames are processed by an object detection network to generate an initial detection box for the target, and key information such as the target category and confidence level is annotated. Subsequently, the target association module combines the appearance features of the target, such as color, texture, and motion features, such as speed and trajectory, to match the target in the current frame with the tracked target in the previous frame, ensuring the continuity and consistency of the trajectory. Among them, the target association module achieves matching by calculating the similarity matrix between targets, where the similarity comprehensively considers the appearance and motion features of the targets, as shown in Eq. (8).

$$S_{ij} = \alpha \cdot IoU(B_i, B_j) + \beta \cdot \cos(f_i, f_j) \tag{8}$$

In Eq. (8), S_{ij} represents the similarity score between target i and target j. α and β both represent weight parameters. $IoU(B_i, B_j)$ represents the intersection over union ratio of the bounding boxes of target i and target j. f_i and f_j represent the appearance feature vectors of target i and target j, respectively. $\cos(f_i, f_j)$ represents the cosine similarity between appearance feature vectors. After the association is completed, the trajectory update module uses Kalman filtering to dynamically estimate the position and velocity of the target, in order to smooth the tracking results, as displayed in Eq. (9).

$$\begin{cases} x_t = Fx_{t-1} + E(r_t - Rx_{t-1}) \\ E = P_{t-1}H^T(RP_{t-1}R^T + \zeta)^{-1} \end{cases} \tag{9}$$

In Eq. (9), x_t and x_{t-1} represent the target state variables at the current time and the previous time, respectively. F represents the state transition matrix. r_t represents the observation vector at the current time. R represents the observation model matrix. E represents the Kalman gain. P_{t-1} represents the covariance matrix of the previous time state. ζ represents the covariance matrix of observed noise. In addition, improving the CA mechanism in the HRNet encodes the global directional information of the input feature map, and then generates a weight distribution through embedding coordinate information. Finally, the feature map is adjusted using weighting, as displayed in Eq. (10).

$$\begin{cases} z_c^h = \dfrac{1}{H}\sum_{i=1}^{H} X(i,j,c), \quad z_c^w = \dfrac{1}{w}\sum_{j=1}^{w} X(i,j,c) \\ f_c = \sigma(Conv_{1\times1}^h(z_c^h) + Conv_{1\times1}^w(z_c^w)) \\ Y(i,j,c) = f_c \cdot X(i,j,c) \end{cases} \tag{10}$$

In Eq. (10), z_c^h and z_c^w represent the global information encoding of feature map X in the height and width directions, respectively. $X(i,j,c)$ represents the feature values of the input feature map at position (i,j) and channel c. f_c represents the generated channel weight. σ represents the activation function. $Y(i,j,c)$ represents the output feature map. $Conv_{1\times1}^h$ and $Conv_{1\times1}^w$ represent 1×1 convolution operations in the height and width directions, respectively. Regarding the original Bottleneck and Basicblock modules in HRNet, Ghost and CA modules are respectively integrated for optimization. The schematic diagram of the optimized Bottleneck and Basicblock modules is shown in Fig. 6.

Fig. 6 (a) displays the Bottleneck module structure before and after optimization. Fig. 6 (a) shows the Basicblock module

structure before and after optimization. As shown in Fig. 6 (a), the two 1×1 convolutions and 3×3 convolutions in the original module have been replaced by the Ghost module, which efficiently reduces feature redundancy by generating primary and auxiliary features. In addition, CA modules are inserted between Ghost modules. By modeling the interaction between space and channels, the model's ability to express features of the target area has been enhanced. In Fig. 6 (b), the structure originally composed of two stacked 3×3 convolutions has been

replaced with a lightweight convolution combination implemented by the Sandglass module. The Sandglass module significantly reduces the number of parameters and computational complexity while retaining feature information. At this point, the calculation formula for Ghost in Bottleneck is shown in Eq. (11).

$$Y = Concat(X * W_m, \sigma(X * W_m) * W_a) \qquad (11)$$

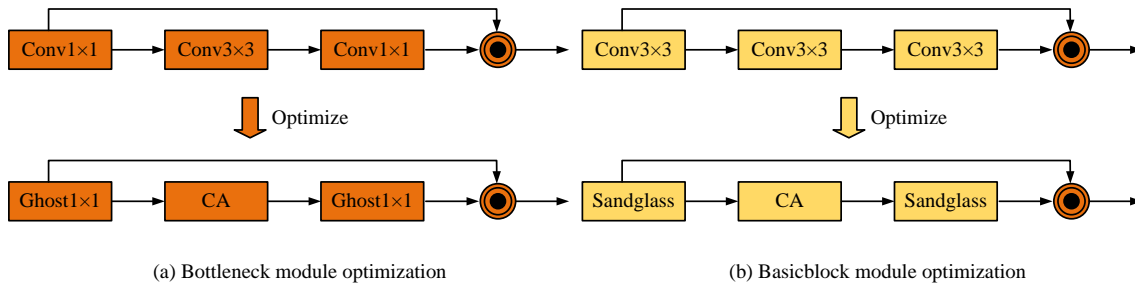(a) Bottleneck module optimization   (b) Basicblock module optimization

Fig. 6. Schematic diagram of Bottleneck and Basicblock modules before and after optimization.

In Eq. (11), W_m and W_a represent the convolution kernel parameters of the main feature and auxiliary feature, respectively. The Sandglass in Basicblock is shown in Eq. (12).

$$Y = X + \sigma(DepthwiseConv(Point\,wiseConv(X))) \quad (12)$$

In Eq. (12), $Point\,wiseConv(\_)$ represents a 1×1 point convolution. $DepthwiseConv(\_)$ represents deep convolution. Based on the improvement of HRNet structure and the comprehensive improvement of DeblurGANv2, a new tennis match Hawk-eye deblurring and pose recognition model is proposed. The process is shown in Fig. 7.
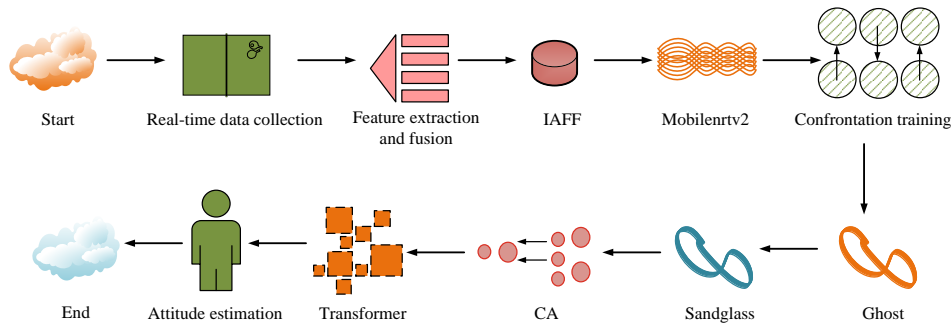

Fig. 7. New model flow of Hawk-eye deblurring and pose recognition in tennis match.

As shown in Fig. 7, firstly, the improved DeblurGANv2 is used for multi-scale feature extraction and fusion of blurred images. The IAFF mechanism is introduced to focus on key regions, and the Mobilenetv2 backbone network is taken to reduce computational overhead. The generator is optimized through adversarial training to generate high-quality and clear images. Subsequently, based on the improved HRNet for pose recognition, Ghost and Sandglass modules are introduced to replace the original Bottleneck and Basicblock modules to reduce the number of parameters, while combining CA mechanism to enhance the feature expression of key regions. Finally, through object detection and Transformer-based object tracking modules, target association and trajectory updates are achieved, outputting clear images, pose keypoints, and motion trajectories.

III. RESULTS

The study first establishes an experimental environment and conducts hyperparameter tuning, with deblurring effect and pose recognition accuracy as the core indicators for testing. The experiment covers two classic datasets and conducts ablation testing, comparative testing, and multi-scenario experiments on lighting, number of people, etc., to verify the robustness and adaptability of the model. Compared with multiple advanced models, the proposed model has achieved good results, especially showing significant advantages in parameter quantity and inference time. In complex lighting and multi-target scenes, the proposed model also demonstrates excellent performance and practical application potential.

A. Performance Testing of Hawk-eye Deblurring and Pose Recognition Model in New Tennis Matches

The study selects two classic public datasets as data sources, namely the Tennis Tracking Dataset (TTD) and the Max Planck Institute for Informatics Human Pose Dataset (MPII). Among them, TTD is a dataset focused on tennis match analysis, which includes key point annotations of players such as head, shoulder, elbow, knee and other joint points, tennis trajectories, as well as action annotations on the court such as serving, returning, running, etc. The MPII dataset is a high-quality dataset focused on human pose estimation, containing 25000 images covering over 40000 human instances. The images in this dataset are from real-life scenarios and provide rich annotations for 16 joint points, including head, shoulders, elbows, knees, etc. The detailed experimental environment parameters are displayed in Table I.

TABLE I. EXPERIMENTAL PARAMETER TABLE

Experimental equipment	Value
CPU	AMD Ryzen 9 5950X
GPU	NVIDIA RTX 4090
Memory	64GB DDR5
Graphics Memory	24GB GDDR6X
Development Environment	Ubuntu 20.04, Python 3.9
Programming Tools	PyTorch 1.10
Initialize learning rate	0.0005
Learning rate batch size	64
Momentum parameters	0.95
Training period	200 epochs

The study first conducts value selection tests on the feature fusion coefficients in the deblurring stage and the convolutional kernel layers in the pose recognition stage, to achieve the optimal state and facilitate subsequent testing. Taking information entropy as an indicator, Fig. 8 displays the test results.
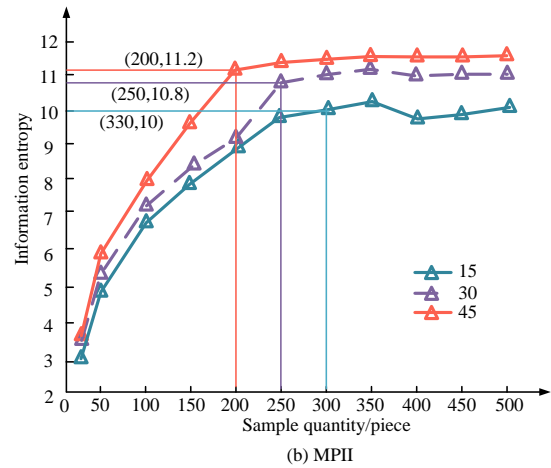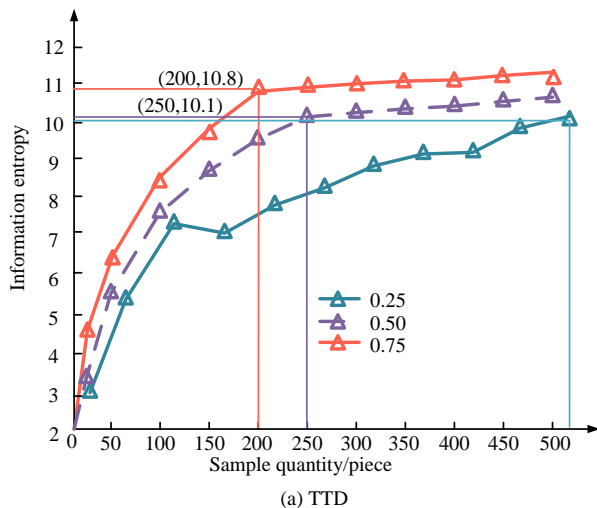

(a) TTD

(b) MPII

Fig. 8. Hyperparameter selection test result.

Fig. 8 (a) shows the selection test of different feature fusion coefficients in the TDD dataset, and Fig. 8 (b) shows the selection test of different convolutional kernel layers in the MPII dataset. According to Fig. 8 (a), when the fusion coefficient was 0.75, the information entropy grew the fastest and tended to stabilize at a sample size of 200, reaching a maximum value of 10.8. When the fusion coefficients were 0.25 and 0.50, the information entropy tended to stabilize at 250 samples, with values of 10.1 and 9.5, respectively. Overall, a fusion coefficient of 0.75 can significantly improve the deblurring effect. In Fig. 8 (b), with the increase of sample size, the information entropy gradually increased. When the number of convolutional kernel layers was 45, the information entropy reached its maximum value of 11.2 at a sample size of 200 and tended to stabilize. When the number of convolutional kernel layers was 30, the information entropy tended to stabilize at 250 samples, reaching 10.8. When the number of convolutional kernel layers was 15, it increased to 330 samples to reach a stable state, with an information entropy value of 10.0. In summary, when the fusion coefficient was 0.75 and the number of convolutional kernels was 45, the deblurring effect of the model was optimal. The study conducts ablation testing on the final model using the Mean Average Precision (mAP) of keypoint detection as the indicator, as presented in Fig. 9.
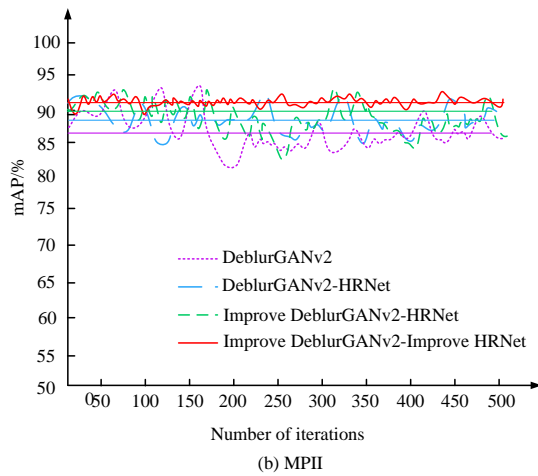

(a) TTD

(b) MPII
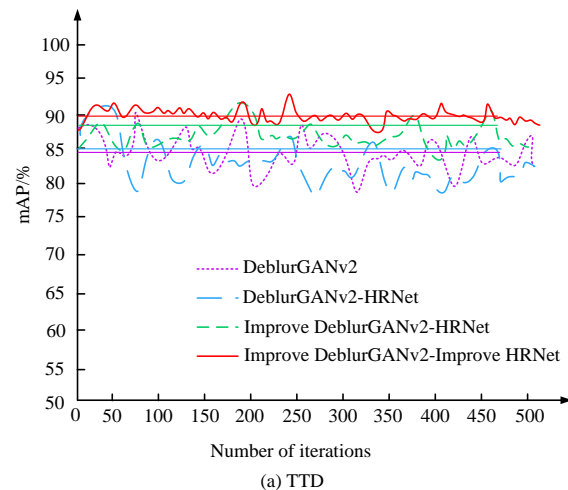
Fig. 9.   Ablation test results.

Fig. 9 (a) displays the ablation test results on the TTD dataset, and Fig. 9 (b) displays the ablation test results on the MPII dataset. As shown in Fig. 9 (a), the mAP value of the basic model DeblurGANv2 fluctuated significantly during the iteration process, stabilizing at around 80.18%. After adding HRNet, the model performance improved and mAP remained stable at around 85.37%. After further improving DeblurGANv2 and introducing improved HRNet, the mAP value increased to around 88.74%, demonstrating better stability. The improved DeblurGANv2 and improved HRNet models were ultimately integrated, with mAP values reaching the highest level. It stabilized at around 92.48%, with minimal fluctuations throughout the entire iteration process, demonstrating the best deblurring and pose recognition performance. According to Fig. 9 (b), the mAP value of the basic model DeblurGANv2 fluctuated greatly and remained stable at around 78.77%. After joining HRNet, mAP increased to 83.21%. The improved DeblurGANv2 and HRNet models showed improvements in both stability and accuracy. The final integrated improved model performed the best, with mAP values stable above 90.49% and minimal fluctuations, demonstrating the strongest robustness and consistency. Other advanced deblurring and pose detection models are introduced

for comparison. For example, Scale-Recurrent Network (SRN), Multi-Stage Progressive Restoration Network (MPRNet), Deep Blind Generative Adversarial Network (DBGAN), High-Resolution Transformer (HRFormer), Pose Estimation Network (PoseNet), and Dynamic Encoder for Keypoint Regression (DEKR) are used for comparison. The test results are shown in Table II, using PSNR, SSIM, parameter count, and runtime as indicators.

TABLE II.   INDEX TEST RESULTS OF DIFFERENT MODELS

Model	PSNR (dB)	SSIM	Parameter quantity (M)	Running time (s)
SRN	30.05	0.91	6.82	4.35
MPRNet	29.56	0.88	20.63	1.17
DBGAN	28.87	0.87	15.58	0.95
HRFormer	29.85	0.96	25.41	0.66
PoseNet	27.92	0.85	12.74	0.75
DEKR	28.51	0.86	18.88	0.58
Our model	29.74	0.89	4.53	0.25

According to Table II, the model exhibited good comprehensive performance. In terms of PSNR index, the proposed model achieved 29.74dB, which was close to SRN and HRFormer and better than most comparative models. SSIM was 0.89, slightly lower than HRFormer's 0.96, but still stable. The most significant advantage lies in the parameter count and running time. The parameter count of the proposed model was only 4.53M, significantly lower than MPRNet's 20.63M and HRFormer's 25.41M. The inference time was 0.25, which was 78%-94% faster than SRN and MPRNet. This indicates that the model has high efficiency while balancing effectiveness, making it very suitable for real-time image processing tasks.

B. Simulation Testing of Hawk-eye Deblurring and Pose Recognition Model for New Tennis Matches

To verify the practical application effect of the new model, two sets of photos are randomly selected from two types of datasets for testing the deblurring and pose estimation effects of different models, as presented in Fig. 10.

(a) SRN   (b) MPRNet   (c) DBGAN   (d) Our model

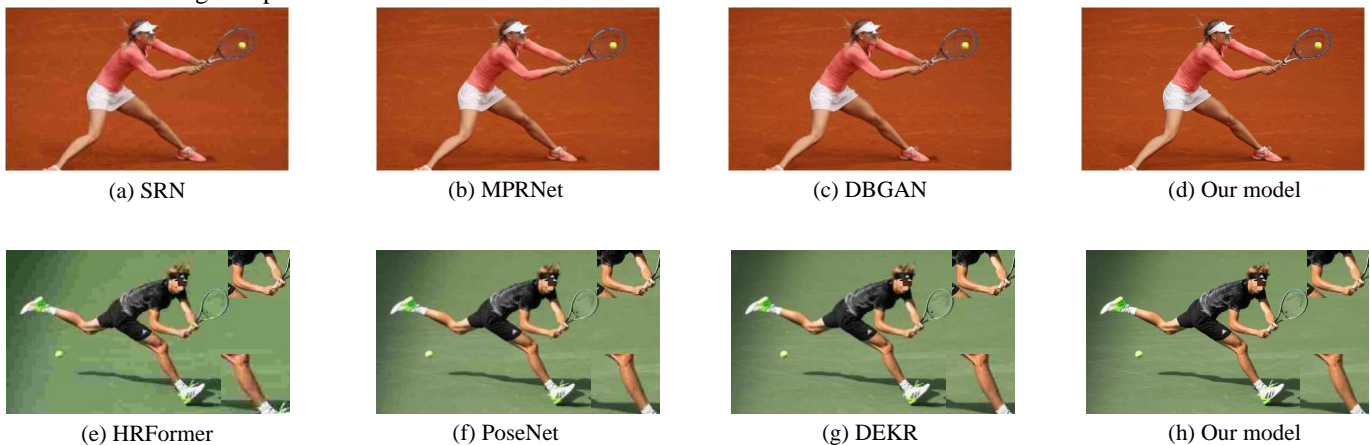(e) HRFormer   (f) PoseNet   (g) DEKR   (h) Our model

Fig. 10. Comparison of deblurring and pose recognition effect of different models.

Fig. 10 (a)-(d) show the actual comparison results of deblurring applications between SRN, MPRNet, DBGAN, and the proposed model. Fig. 10 (e)-(f) show the comparison results of pose recognition applications between HRFormer, PoseNet, DEKR, and the proposed model. From Fig. 10 (a), SRN and MPRNet had similar deblurring effects, but SRN's restoration details were slightly insufficient, while MPRNet had slight artifacts in the texture part. The DBGAN model performed poorly in handling high dynamic blur, with obvious edge blurring. In contrast, the model performed the best, with clear image details, better overall restoration performance than other models, and more natural texture parts. From Fig. 10 (b), the HRFormer and DEKR models could effectively detect the key points of tennis players. However, the HRFormer model was susceptible to interference in complex backgrounds, and the PoseNet model may miss detection, especially in inaccurate recognition during rapid limb movement. The model can accurately identify all key points and has strong robustness to complex poses and motion blur, demonstrating higher recognition accuracy. The performance of the model under different lighting conditions is tested using video frame quality improvement rate and fuzzy image recognition rate as indicators. The results are shown in Fig. 11.
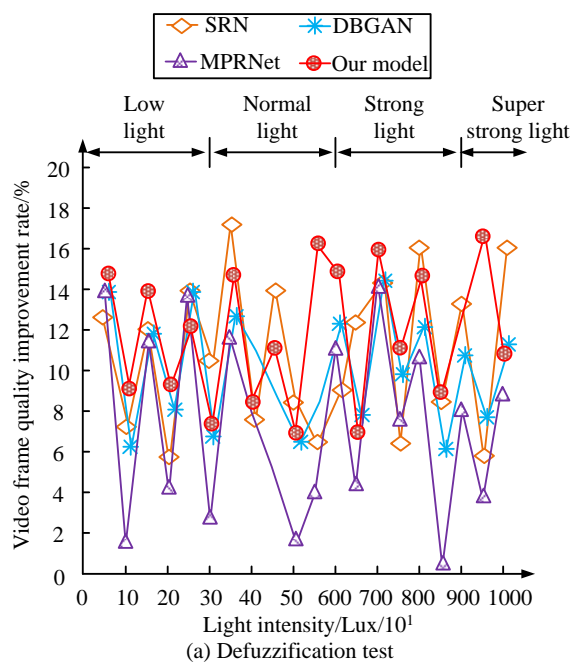
Fig. 11 (a) shows the video frame quality improvement rate test results of four models, and Fig. 11 (b) shows the fuzzy image recognition rate test results of the four models. According to Fig. 11 (a), SRN and MPRNet exhibited relatively stable performance under weak and normal light conditions, with improvement rates ranging from 10% to 12%. However, in strong and ultra strong light environments, the effectiveness of DBGAN and SRN significantly decreased. In contrast, the model showed good improvement rates under various lighting conditions, especially in strong and ultra strong light environments, with a video frame quality improvement rate of 16%-18%, indicating that it could still effectively deblur under severe lighting changes. According to Fig. 11 (b), the recognition rates of HRFormer and PoseNet were relatively high under weak light and normal light, stable between 80.73%-85.46%, respectively. However, the recognition rate of DEKR fluctuated greatly in strong and ultra strong light environments. In contrast, the model maintained a high recognition rate under all lighting conditions, especially in ultra strong light environments, with a recognition rate of up to 92.44%, which was significantly better than other models. Overall, the model demonstrates stronger robustness and stability in deblurring and pose recognition tasks, and has superior adaptability to changes in lighting conditions. The research tests the accuracy of model deblurring and pose recognition in multi-player scenarios, and the results are shown in Fig. 12.

(a) Defuzzification test

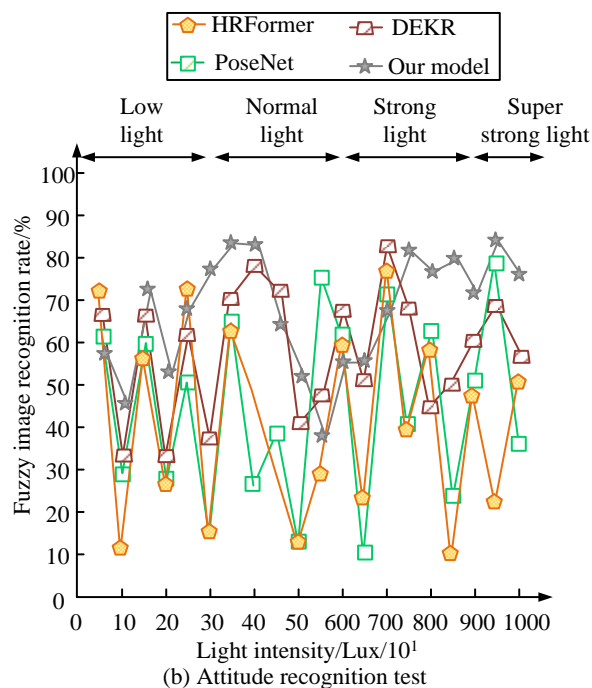
(b) Attitude recognition test

Fig. 11. Test results of video frame quality improvement rate and fuzzy image recognition rate in different modes.
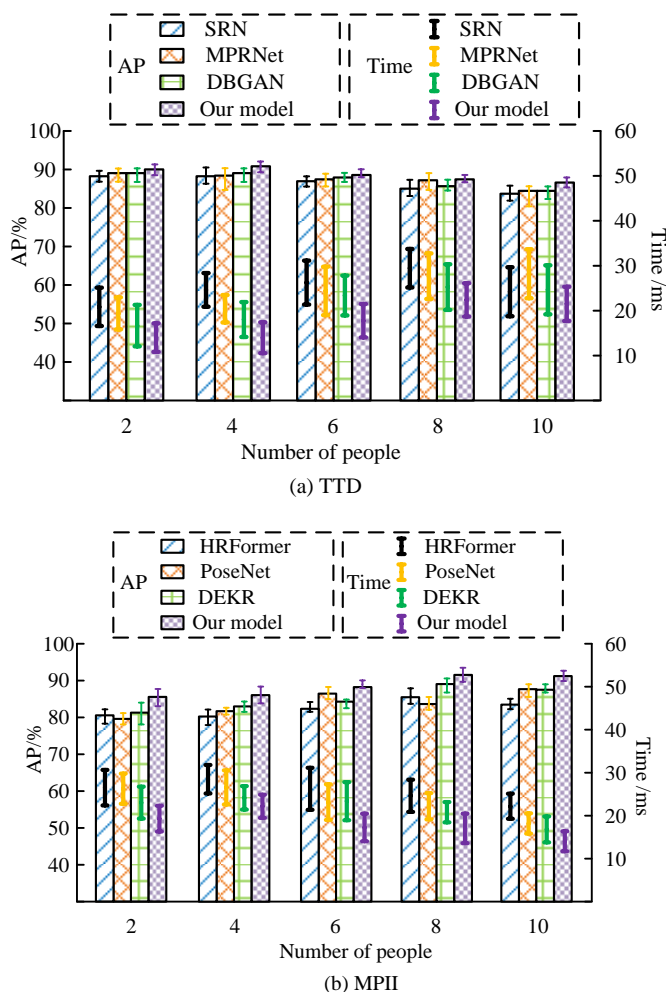
Fig. 12. Results of model deblurring and pose recognition accuracy under different numbers of people.

Fig. 12 (a) displays the deblurring accuracy and time for different models on the TTD dataset, and Fig. 12 (b) displays the pose recognition accuracy and time for different models on the MPII dataset. From Fig. 12 (a), SRN and MPRNet had higher AP values in the 2-person scenario, reaching 90.08% and 88.84% respectively. However, as the number of people increased, the AP values gradually decreased, especially in the 10 person scenario, dropping below 80.96%. The performance of DBGAN in multi-player scenarios was relatively unstable, with large fluctuations in accuracy and longer inference time. In contrast, the model proposed maintained a high AP value of 85.17%-92.38% for all participants, and had the shortest inference time, stabilizing at around 20ms, demonstrating good real-time performance and accuracy. As shown in Fig. 12 (b), HRFormer and PoseNet exhibited high AP values of over 90.02% in both 2-person and 4-person scenarios, but their accuracy significantly decreased in the 10-person scenario. The recognition accuracy and time performance of DEKR in multi-target scenes were unstable and exhibit significant fluctuations. In contrast, the posture recognition accuracy of the model remained stable under different numbers of people, with AP values consistently above 88.74% and inference time controlled at around 30 ms, significantly better than other models. Four types of pose recognition models are tested using

tracking error, target overlap rate, and decision accuracy as indicators, as displayed in Table III.

TABLE III. POSE RECOGNITION TEST RESULTS OF DIFFERENT MODELS

Data set	Model	Tracking error/%	Overlap rate/%	Decision accuracy/%
TTD	HRFormer	6.83	78.57	88.92
	PoseNet	8.93	75.23	85.37
	DEKR	7.62	77.19	87.13
	Our model	5.27	82.35	92.53
MPII	HRFormer	7.12	79.83	89.21
	PoseNet	9.27	74.67	84.71
	DEKR	8.03	76.32	86.83
	Our model	5.31	81.93	93.07

According to Table III, the tracking error of the proposed model on the TTD dataset was 5.27%, significantly lower than HRFormer's 6.83% and PoseNet's 8.93%. The overlap rate was 82.35%, which was about 5%-7% higher than other models. The decision accuracy was 92.53%, significantly better than DEKR's 87.13%. On the MPII dataset, the tracking error of the proposed model was 5.31%, which performed the best. The overlap rate was 81.93%, slightly higher than DEKR's 76.32%. The decision-making accuracy was the highest, at 93.07%, which was significantly improved compared with HRFormer. The above data shows that the model exhibits superior performance and robustness in all indicators.

IV. CONCLUSION

In tennis matches, image blur and pose recognition errors are the main issues affecting the accuracy of the Hawkeye system. To address this challenge, the research improved DeblurGANv2 and HRNet, proposing a novel tennis game image deblurring and pose recognition model. When the fusion coefficient was 0.75 and the number of convolutional kernels was 45, the deblurring effect of the model was optimal, achieving an information entropy value of 11.2. At the same time, after sequentially improving DeblurGANv2 and HRNet, the mAP value of the combined model reached 92.48%, indicating that the improvement and fusion of each module in the study were effective. Compared with other deblurring and pose recognition models, this new model had a PSNR of up to 29.74dB, SSIM of up to 0.89, minimum parameter size of 4.53, and shortest running time of 0.25s, which was 78%-94% faster than SRN and MPRNet. Under different lighting intensities, the proposed model had strong robustness to complex poses and motion blur, showing a recognition accuracy of up to 92.44% and a video frame quality improvement rate of 16% - 18%. In a multi-person scenario, the model had the highest recognition AP value of 92.38%, and the shortest stable inference time was around 20ms. The lowest pose recognition tracking error was 5.27%. Although the overlap rate was higher than other models, the decision accuracy was 92.53%, far exceeding other methods. In summary, the model has significant advantages in both processing effectiveness and efficiency. However, the performance of the model still

fluctuates to some extent under extreme lighting conditions, such as ultra-low light or severe lighting environments. Future research will further optimize the robustness of the model and explore methods that combine multi-modal data to enhance its adaptability and generalizability in practical applications.

REFERENCES

[1] J. Zhang, "Evaluation of the effect of artificial intelligence training equipment in physical training of table tennis players from a biomechanical perspective," Mol. Cell. Biomech., vol. 21, no. 1, pp. 319-319, November 2024.

[2] Y. M. Deng, and S. Y. Wang, "Biological eagle-eye inspired target detection for unmanned aerial vehicles equipped with a manipulator," Mach. Intell. Res., vol. 20, no. 5, pp. 741-752, March 2023.

[3] F. Meng, "Tennis video target tracking based on mobile network communication and machine learning algorithm," Int. Trans. Elec. Energ. Syst., vol. 2022, no. 1, pp. 7447121-7447125, September 2022.

[4] Y. Zhao, L. Lu, W. Yang, Q. Li, and X. Zhang, "Lightweight tennis ball detection algorithm based on Robomaster EP," Appl. Sci., vol. 13, no. 6, pp. 3461-3462, March 2023.

[5] Y. Yang, D. Kim, and D. Choi, "Ball tracking and trajectory prediction system for tennis robots," J. Comput. Design Eng., vol. 10, no. 3, pp. 1176-1184, June 2023.

[6] D. Gao, Y. Zhang, and H. Qiu, "Automatic detection method of small target in tennis game video based on deep learning," J. Intell. Fuzzy Syst., vol. 45, no. 6, pp. 9199-9209, December 2023.

[7] Y. Ke, Z. Liu, and S. Liu, "Prediction algorithm and simulation of tennis impact area based on semantic analysis of prior knowledge," Soft Comput., vol. 26, no. 20, pp. 10863-10870, April 2022.

[8] B. T. Naik, M. F. Hashmi, and N. D. Bokde, "A comprehensive review of computer vision in sports: Open issues, future trends and research directions," Appl. Sci., vol. 12, no. 9, pp. 4429-4434, April 2022.

[9] J. Bian, X. Li, T. Wang, O. Wang, J. Huang, and C. Liu, "P2ANet: A large-scale benchmark for dense action detection from table tennis match broadcasting videos," ACM Trans. Multimedia Comput., Commun. Appl., vol. 20, no. 4, pp. 1-23, January 2024.

[10] H. R. Ghezelsefloo, and S. H. Alavi, "The Impact of event-based sports technologies on the training and career development of referees in Iran volleyball super league," Technol. Educ. J. (TEJ), vol. 16, no. 2, pp. 351-362, May 2022.

[11] X. Peng, L. Tang, "Biomechanics analysis of real-time tennis batting images using Internet of Things and deep learning," J. Supercomput., vol. 78, no. 4, pp. 5883-5902, October 2022.

[12] H. C. Nguyen, T. H. Nguyen, and J. Nowak, "Combined YOLOv5 and HRNet for high accuracy 2D keypoint and human pose estimation," J. Artif. Intell. Soft Comput. Res., vol. 12, no. 4, pp. 281-298. October 2022.

[13] Y. Li, "Visualization of movements in sports training based on multimedia information processing technology," J. Ambient Intell. Humanized Comput., vol. 15, no. 4, pp. 2505-2515, March 2024.

[14] A. Fitzpatrick, J. A. Stone, S. Choppin, and J. Kelley, "Analysing Hawk-Eye ball-tracking data to explore successful serving and returning strategies at Wimbledon," Int. J. Perform. Anal. Sport, vol. 24, no. 3, pp. 251-268, December 2024.

[15] T. Ning, C. Wang, M. Fu, and X. Duan, "A study on table tennis landing point detection algorithm based on spatial domain information," Sci. Rep., vol. 13, no. 1, pp. 20656-20659, November 2023.

[16] E. J. Jeong, J. Kim, and S. Ha, "Tensorrt-based framework and optimization methodology for deep learning inference on jetson boards," ACM Trans. Embed. Comput. Syst., vol. 21, no. 5, pp. 1-26, October 2022.

[17] C. Li, H. Li, L. Liao, Z. F. Liu, and Y. Dong, "Real-time seed sorting system via 2D information entropy-based CNN pruning and TensorRt acceleration," IET Image Process., vol. 17, no. 6, pp. 1694-1708, January 2023.

[18] P. Liu, Q. Wang, H. Zhang, J. Mi, and Y. Liu, "A lightweight object detection algorithm for remote sensing images based on attention mechanism and YOLOv5s," Remote Sens., vol. 15, no. 9, pp. 2429-2430, April 2023.

[19] S. Yan, C. Yang, and L. Guo, "Accuracy improvement in motion tracking of tennis balls using nano-sensors technology," Adv. Nano Res., vol. 14, no. 5, pp. 409-419, May 2023.

[20] L. Li, and A. Yang, "Correction algorithm of tennis dynamic image serving path based on symmetric algorithm," Symmetry, vol. 14, no. 9, pp. 1833-1834, Aug 2022.

[21] C. Shen, and Z. Sun, "Research on target localization recognition of automatic mobile ball-picking robot," J. Opt., vol. 51, no. 4, pp. 866-873, January 2022.

[22] G. C. Domínguez, E. F. Álvarez, A. T. Córdoba, and D. G. Reina, "A comparative study of machine learning and deep learning algorithms for padel tennis shot classification," Soft Comput., vol. 27, no. 17, pp. 12367-12385, February 2023.

[23] C. Z, Q. Jiacheng, and B. Wang, "YOLOX on embedded device with CCTV & TensorRT for intelligent multicategories garbage identification and classification," IEEE Sens. J., vol. 22, no. 16, pp. 16522-16532, August 2022.

[24] X. Luo, Y. Wu, and F. Wang, "Target detection method of UAV aerial imagery based on improved YOLOv5," Remote Sens., vol. 14, no. 19, pp. 5063-5067, September 2022.

[25] Q. Wang, and N. Yao, "Light imaging detection based on cluster analysis for the prevention of sports injury in tennis players," Opt. Quantum Electron., vol. 56, no. 2, pp. 191-192, December 2024.

[26] S. Vancurik, and D. W. Callahan, "Detection and identification of choking under pressure in college tennis based upon physiological parameters, performance patterns, and game statistics," IEEE Trans. Affect. Comput., vol. 14, no. 3, pp. 1942-1953, July 2022.

[27] M. Niu, "Research on tennis-assisted teaching assessment technology based on improved dense trajectory algorithm," Int. J. Netw. Virtual Organ., vol. 28, no. 2, pp. 154-170, September 2023.

[28] W. Ren, "A novel approach for automatic detection and identification of inappropriate postures and movements of table tennis players," Soft Comput., vol. 28, no. 3, pp. 2245-2269, January 2024.

[29] W. Wu, "Multimodal emotion detection of tennis players based on deep reinforcement learning," Int. J. Biometrics, vol. 16, no. 5, pp. 497-513, September 2024.

[30] J. Yao, X. Fan, B. Li, and W. Qin, "Adverse weather target detection algorithm based on adaptive color levels and improved YOLOv5," Sens., vol. 22, no. 21, pp. 8577-8579, October 2022.

[31] M. Skublewska-Paszkowska, P. Powroznik, and E. Lukasik, "Tennis patterns recognition based on a novel tennis dataset–3DTennisDS," Adv. Sci. Technol. Res. J., vol. 18, no. 6, pp. 159-176, May 2024.

[32] X. Song, "Physical education teaching mode assisted by artificial intelligence assistant under the guidance of high-order complex network," Sci. Rep., vol. 14, no. 1, pp. 4104-4109, February 2024.

[33] A. Abba Haruna, L. J. Muhammad, and M. Abubakar, "Novel thermal-aware green scheduling in grid environment," Artif. Intell. Appl., vol. 1, no. 4, pp. 244-251, November 2022.