Improved Generative Adversarial Network and Particle Swarm Optimization Support Vector Machine for Tennis Serving Behavior Analysis

Haibo Cao
School of Physical Education, Xinyang Normal University, Xinyang, 464000, China
E-mail: caohb@xynu.edu.cn

Keywords: video images, generate adversarial networks, particle swarm optimization algorithm, support vector machine, behavior analysis

Received: October 9, 2024

This study proposes a behavior analysis model based on an improved generative adversarial network and particle swarm optimization support vector machine algorithm for deblurring and feature extraction in tennis video serving behavior. The model first improves the generative adduction network by introducing a multi-layer convolution structure and a variety of activation functions, including three-layer convolution. The activation function is selected to use ReLU and Leaky ReLU alternately to enhance the generator in capturing image details. During model training, the generator optimizes the output image by minimizing the Wasserstein distance, and the discriminator evaluates the difference between the generated image and the real image. Then, to further extract features, the particle swarm optimization algorithm was used to dynamically optimize the feature extraction of each frame in the feature space, and dynamically adjust the inertia weights. The initial value was 0.9 and the final value was 0.4. After feature extraction, the data were input into SVM for classification. The penalty parameter of SVM was set to 1 and the accuracy was set to 0.001. The results of the comparative experiments demonstrated that the proposed method exhibited superior performance in deblurring images, with an average subjective score of 81.16 points, a notable advantage over the comparison algorithm. In the objective evaluation, the average Peak Signal-to-Noise Ratio (PSNR) and structural similarity value of the image after defuzzing by the research method reached 35.12dB and 0.93, respectively. The PSNR and structural similarity of the image increased by 13.56% to 18.29% and 8.33% to 19.90%, respectively. In the feature extraction and classification experiments, the accuracy of the proposed algorithm reached 91.24%, which was significantly higher than the traditional algorithm. The convergence speed was faster than the particle swarm optimization algorithm, the ant colony optimization algorithm, and the simulated annealing algorithm, reducing the number of iterations by 35.33%, 40.52%, and 51.55%, respectively. The data validate that the designed method has good application prospects in improving video image quality and feature extraction.

Povzetek: Raziskava se ukvarja z analizo teniškega servisa, ki z izboljšanimi algoritmi GAN in PSO-SVM poboljša ugotovitve iz slik.

1 Introduction

In light of the accelerated advancement of digital media and video technology, the analysis and processing of video content have become increasingly important in many fields. Especially in sports science and sports analysis, how to extract valuable information from video images has become a research focus. Tennis, as a dynamic and complex sport, often suffers from blurring in its video images due to various factors such as the athlete's rapid movement, different shooting angles, and changes in lighting. These factors can lead to a decrease in image quality, affecting game analysis and technical evaluation [1-2]. Therefore, improving the clarity and information extraction ability of Tennis Video Images (TVI) has important practical significance for athlete performance evaluation and tactical analysis [3-4]. Currently, many industry scholars have researched the image processing technology and action behavior analysis. Ding Q et al. proposed a camera-based long-term trajectory tracking technique to improve the effectiveness of multi-target tracking technology in sports game feature recognition. This study first improved the Tracking-Learning-Detection (TLD) algorithm and then integrated machine learning methods into the improved algorithm. This method has significantly improved performance and can be effectively applied to feature extraction in sports events [5]. Mulimani D et al. developed a video preprocessing technique under a new framework. This method first calibrated players and classified occlusions to ensure accurate identification of athletes in complex game environments. Subsequently, by utilizing the system to

track and label athletes on the court, the framework significantly improved the tracking accuracy of basketball players and provided more reliable technical support for fairness in the game [6]. Zhang J et al. developed a new spatial attention and temporal dilation GCN that uses a self-attention mechanism to select human joints that are beneficial for action recognition, thereby reducing the impact of data redundancy and noise. Extensive experiments on NTU-RGB+D and Kinetics Skeleton have shown that this method completes State-of-the-art (SOTA) performance in skeleton-based action recognition [7]. Zhu X et al. introduced a skeleton attention module in the action recognition data system, which shoots the skeleton sequence onto a single-RGB frame to assist focusing on the limb motion area. Experiments on the NTU RGB+D and SYSU benchmarks have shown that compared to SOTA methods, this model achieved competitive performance while reducing network complexity [8]. The results of related work are summarized in Table 1.

While the aforementioned research has yielded promising outcomes in the domains of video image processing and action

Table 1: Research status analysis

Method	Accuracy (%)	PSNR (dB)	SSIM	Major deficiency
Ding Q et al. [5]	82	28.5	0.76	Multi-target tracking accuracy is insufficient in complex background.
Mulimani D et al. [6]	80.5	29	0.79	The real-time and accuracy of rapidly changing scenes are insufficient.
Zhang J et al. [7]	86	30.8	0.82	Data redundancy and noise are likely to affect the processing of diverse actions.
Zhu X et al. [8]	84	31.1	0.81	The recognition ability in complex motion scenes needs to be improved.

behavior analysis, it has not yet fully addressed the intricacies of the fuzzy states that arise in motion scenes. When faced with information redundancy and feature overlap, it is easy to encounter the problem of misidentification of action behavior. This study aims to propose a method based on the combination of improved a Generative Adversarial Network (GAN) and Particle Swarm Optimization - Support Vector Machine (PSO-SVM) for fuzzy removal and feature extraction in TVI processing. The innovation of this method lies in combining improved GAN and PSO algorithms to optimize the overall process of deblurring and feature extraction, making it suitable for complex motion scenes. By dynamically adjusting the inertia weight, the convergence speed and accuracy of the PSO algorithm have been improved, effectively enhancing the performance of the model at the feature extraction level.

2 Methods and materials

2.1 Deblurring processing based on video images

The goal of this research is to propose a method based on the combination of improved GAN and PSO-SVM to improve the deblurring effect and feature extraction ability of TVIs, thereby improving the accuracy of tennis serve motion recognition. The research specifically focuses on solving the problem of image blur and information redundancy caused by dynamic scenes. Assuming that the improved GAN structure will effectively enhance feature extraction capability through multi-layer convolution and multiple activation functions. Therefore, the goal of improving Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) indicators, and achieving more accurate recognition of tennis serving actions can be achieved. The analysis of tennis serve movements is not only important for improving athletes' technical level but also has significant implications in sports training, match judgment, and tactical development. Therefore, this study is expected to provide effective support for sports science and intelligent sports applications.

Tennis is a popular competitive sport, divided into singles and doubles forms. Participants throw the ball and hit it with a racket to make it land on the opponent's court. In tennis matches, serving is the way the game begins and an important part that determines the pace and strategy of the game. Serving is a highly technical action in tennis that involves multiple key elements, including preparation, pitching, hitting, and swinging. Serving is not only a technical action but also a strategic behavior. Athletes can choose different types of serve and make different choices based on their opponents' weaknesses and court conditions. Choosing which side of the court to serve on can affect the opponent's receiving angle and preparation. When

serving, the athlete's psychological state can also affect their performance. At the same time, the serving behavior of athletes is influenced by various factors, including physical fitness, proficiency in serving skills, the athlete's judgment of the game progress, understanding of the opponent, and scientific training methods and feedback mechanisms. In terms of physical fitness, technical proficiency, and tactical awareness, athletes need to continuously improve through their own efforts. In terms of scientific training methods and feedback mechanisms, this study believes that using video image processing technology can provide a detailed analysis of athletes' serving movements. By capturing the athlete's serve process at a high frame rate, analyzing the movements in great detail, and using computer vision algorithms to process the video, it is possible to extract keyframes, measured serve angles, speeds, and technical elements. By using machine learning to analyze serving data, patterns and performance characteristics of athletes during training can be identified.

The occurrence of blurry video images in TVI can be attributed to a number of factors, including the camera's shutter speed being inadequate to capture fast-moving objects during rapid motion, such as the swing of a racket or the trajectory of a ball. The shaking or vibration of the shooting equipment can cause the entire image to blur. Shooting tennis matches in low light environments can also increase the risk of motion blur [9]. The fuzzy model of TVI can be represented by equation (1).

$$I_B = K * I_S + N \tag{1}$$

In equation (1), I_B is the blurred image, which is the object that needs to be restored or studied during the analysis process. I_S is the original image, which represents the true state of the athlete at the moment of serving. By restoring the original image, the serving behavior can be analyzed more accurately. K is the convolution kernel. In serve analysis, convolutional kernels can be used to simulate the trajectory of an athlete's swing and ball flight, aiming to understand the source of ambiguity. N is additive noise, which may be interference caused by device noise or changes in ambient light. Analyzing this noise can improve the accuracy and reliability of serving moments. * is a convolution operation. In the analysis of serving behavior, this operation helps identify and simulate the cause of blurring, thereby restoring the original image and analyzing the behavior. In the context of blurry images, the Camera Response Function (CRF) is an important tool. It helps to understand and process image data by describing how the camera converts light into image pixel values, as given by equation (2) [10].

$$g(I_{S(i)}) = I_{S'(i)}^{\gamma} \tag{2}$$

In equation (2), g is the CRF approximation function, which describes the conversion of light rays into pixel values. The significance of CRF lies in its capacity to rectify discrepancies in image brightness resulting from disparate camera models or shooting conditions, thereby enhancing image uniformity and comparability. γ is a constant value, usually set to 2.2 by default. $I_{S(i)}$ is a potential clear image, representing the original unaffected image. It is an ideal image that can be generated through deblurring and restoration, thereby helping to establish accurate feature representations in image retrieval. $I_{S'(i)}$ is a clear image observed. The calculation of blurred image I_B is shown in equation (3).

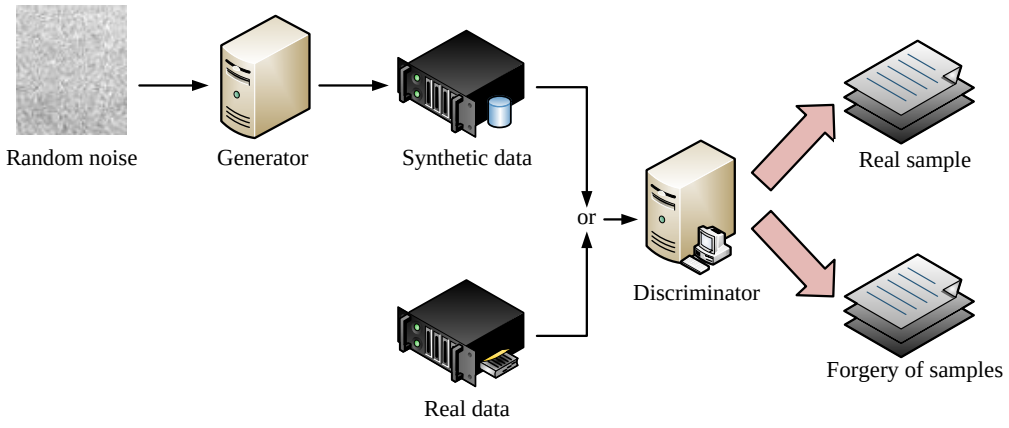


Figure 1: GAN structure.

$$I_B = g(\frac{1}{M}\sum_{t=1}^{M} I_{S(t)}) \quad (3)$$

In equation (3), t is the time in the video image. M is the number of clear frames used to generate blurry images. In image retrieval, the quantity of M affects the quality of blurry images. Collecting multiple clear images can help generate more stable and rich image features. This study calculates the actual blurred image using equation (4).

$$I_B = g(\frac{1}{T}\int_{t=0}^{T} I_{S(t)} dt) \quad (4)$$

In equation (4), T is the exposure time period, which indicates the reception time of the light line in the captured image. In image retrieval, the corresponding exposure time can affect the brightness and details of the image. This study adopts a non-blind deblurring method for blurry images. This method can obtain information about the blur kernel used in the restoration process of blurred images. The blur kernels may be the result of a number of factors, including camera motion, object motion, inaccurate focusing, and other causes [11-12]. This study assumes that the noisy and original images are Y and X, and the blur kernel stands for Z. The non-blind deblurring process is given by equation (5).

$$\left\{\hat{X}, \hat{Z}\right\} = \arg\min \left\| Z \otimes X - Y \right\|_2^2 + \varphi(X) + \theta(Z) \quad (5)$$

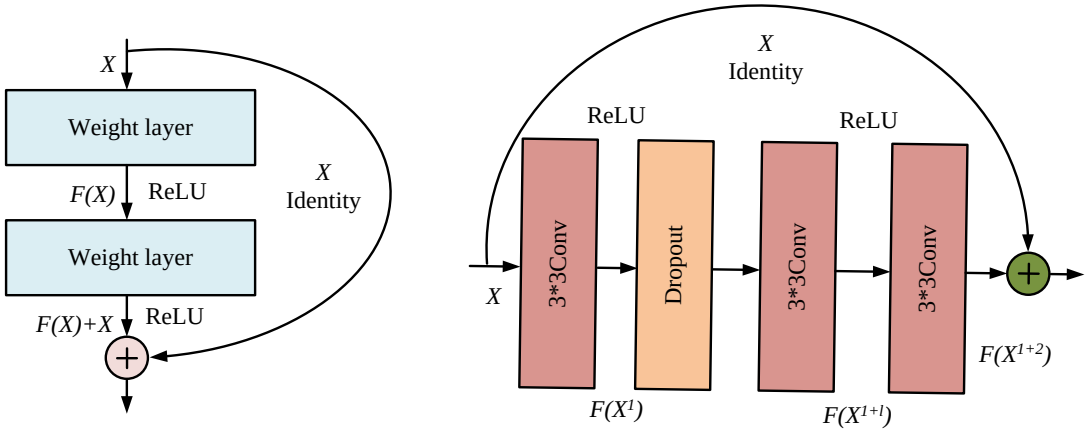
In equation (5), $\varphi(X)$ is the regularization term for the expected clear image. $\theta(Z)$ represents possible fuzzy kernels. Due to the excellent performance of GAN, this study applies it to denoising sports images. Figure 1 shows the GAN structure.

In Figure 1, the task of the GAN generator is to receive blurry images as input and attempt to generate outputs corresponding to real clear images. The generator gradually adjusts its parameters by learning the mapping relationship between blurry and clear images to achieve the goal of generating high-quality clear images. At the same time, the discriminator receives the image and outputs a probability value representing the probability that the input image is "real". The discriminator strives to improve its judgment ability to accurately identify the differences between the generator's output and the real image [13-15]. The objective function for the confrontation between the generator and discriminator is shown in equation (6)

.

$$\min_G \max_D V(D,G) = E_{x\sim P_{data(x)}}[\log D(x)] + E_{z\sim P_{z(z)}}[\log(1 - D(G(z)))] \quad (6)$$

In equation (6), x is a real sample from $P_{data(x)}$ used to train the discriminator and help it learn how to recognize the characteristics of clear images. $E_{x\sim P_{data(x)}}$ is the expectation for



(a) Traditional residual block structure diagram

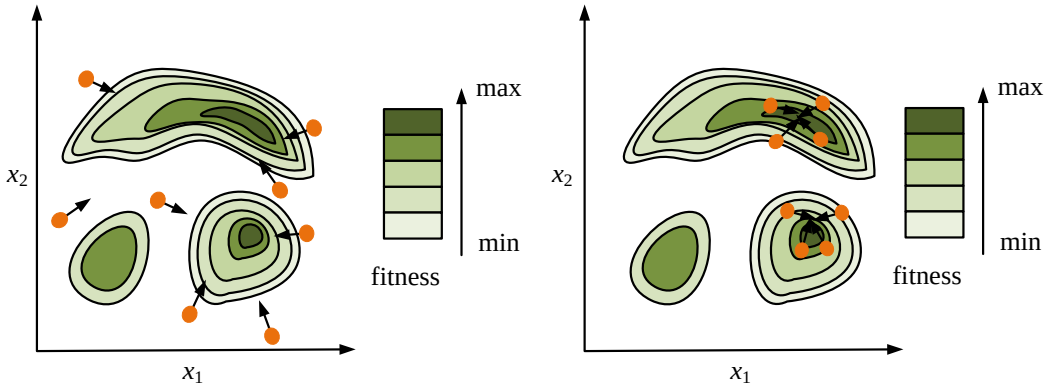
(b) Structure diagram of improved residual block

Figure 2: Traditional residual block structure.

clear input images. In the field of image retrieval, the concept of "expectation" plays a pivotal role in the development of effective models. By effectively calculating expectations, a more balanced approach to generation and discrimination can be achieved, ultimately enhancing the clarity and quality of both

generated and blurred images. $D(\cdot)$ is the output of discriminator D. The discriminator continuously optimizes its classification ability by comparing real samples with generated samples. In image retrieval, the performance of the discriminator affects the quality of the generated images by the generator, which further affects the accuracy of the retrieval results. $G(\cdot)$ is the output of generator G. When the image retrieval system faces a fuzzy query, the generator can transform the fuzzy image into a clear image. Wasserstein GAN (WGAN) is a variant of GAN. In image deblurring tasks, gradient vanishing is a common problem due to image degradation and other reasons, which affects the convergence of the model. This study uses Wasserstein distance to quantify the difference between the generator and the real data distribution, which can better handle the problem of gradient vanishing and help the generator learn the data distribution better [16]. To improve the deblurring effect of images, this study improves the structure of traditional residual blocks. The Residual Bock Structure (RBS) is displayed in Figure 2.

In Figure 2 (a), RBS extracts feature from the input blurred image through multiple convolutional layers. For TVI, important features include the trajectory of the ball and the movements of the athletes. Meanwhile, with the help of residual connections, the network can converge to the optimal solution faster. However, traditional residual blocks typically contain fewer convolutional layers or simpler structures, which may limit the model's capacity to learn complex features and details. In the face of more complex models, the lack of effective regularization may lead to overfitting. Therefore, this study improves the traditional RBS, as shown in Figure 2 (b). The improvement of residual blocks in this study mainly includes increasing the depth of convolutional layers, introducing multiple activation functions, applying Dropout, implementing skip connection modules, and removing batch normalization. The improved residual block consists of three convolutional layers, each using a 3x3 convolution kernel. This design enhances the expressive power of



(a) time 1 (a) time 2

Figure 3: Schematic diagram of PSO particle motion.

the model, enabling it to capture more complex feature representations. Introducing two ReLU activation functions between two convolutional layers can accelerate convergence and help the model learn nonlinear features. The final skip connection module is retained to alleviate gradient vanishing and explosion problems, ensure the flow of important information, and maintain the training stability of the model. Removing the batch normalization layer makes the model more flexible during small batch training and reduces computational burden. The loss function is shown in equation (7).

$$W(P_{data}, P_g) = \inf_{\gamma \sim \prod(P_{data}, P_g)} E_{(x,y)\sim\gamma}\left[\|x-y\|\right] \quad (7)$$

In equation (7), x and y are real samples and generated samples. In image retrieval, real samples are the basis for training network models. Generating samples is the key to successful high-quality image retrieval. $\prod(P_{data}, P_g)$ is the set of joint distributions of P_{data} and P_g. By comparing the joint distribution of real samples and generated samples, the generator can more effectively capture the features of real data when generating images, ensuring the similarity between generated samples and real samples in the feature space. $(x, y) \sim \gamma$ is one of the samples, which supports adversarial training between the generator and

discriminator, improving the model's generalization ability and optimization performance. $\inf$ is the expected distance. The loss function of the model generator is equation (8).

$$L_x = \frac{1}{W_{i,j}H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} \left(\begin{matrix} \Phi_{i,j}(I^S)_{x,y} \\ -\Phi_{i,j}(G_{\theta G}(I^B))_{x,y} \end{matrix} \right)^2 \quad (8)$$

In equation (8), $\Phi_{i,j}$ is a feature map that can capture important semantic information and high-level features in the image. $W_{i,j}$ and $H_{i,j}$ are the width and height of the feature map.

2.2 Feature extraction and classification of video images based on PSO-SVM algorithm

The above study extracts and recognizes key features of video images by improving the residual network structure. Due to the complexity of dynamic scenes in tennis, video images often contain a large amount of information, resulting in the problem of information redundancy between features, which has a negative impact on feature extraction. To improve the effectiveness of feature extraction, this study introduces PSO-SVM [17-18] into the model. In the PSO-SVM algorithm, PSO is responsible for feature extraction in tennis, while SVM is responsible for feature recognition and classification. In PSO, individuals update their position and velocity to find the optima to the problem. The concept of employing PSO for the purpose of extracting features from motion images entails the treatment of each frame within the image sequence as a particle. Through the optimization of the particle's motion trajectory, it is possible to selectively capture salient features of object motion [19]. The PSO motion is shown in Figure 3.

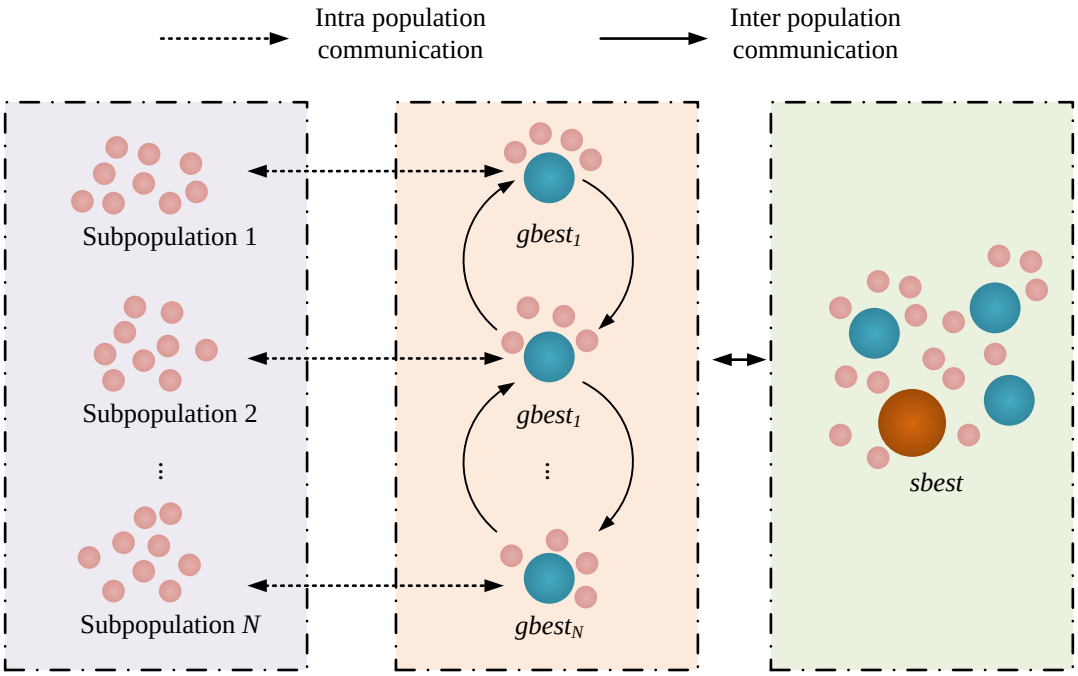


Figure 4: MPSO structure diagram.

Figure 3 shows the motion of PSO particles. In the initial stage of the PSO, a set of particles is generated at random, each representing a possible solution value. The motion of each particle mainly updates its current position and velocity. After each update, the fitness of each particle's current position is calculated to evaluate the quality of that position. The particle speed update is shown in equation (9) [20].

$$V_i^{t+1} = \omega \times V_i^t + c_1 \times rand() \times (pbest_i - X_i^t) + c_2 \times rand() \times (gbest - X_i^t) \quad (9)$$

In equation (9), ω is the inertia weight, which is usually a non-negative value. c_1 and c_2 are acceleration factors. The position of the i-th particle is represented as vector X_i. $rand()$ represents a random number between [0, 1]. $pbest_i$ is the optimal

position of particle i. The best position of the population is represented by $gbest$. When the ω value is large, the global Search Ability (SA) is strong and the local SA is weak. When the ω value is low, the local SA is strong and the global SA is weak. The expression for particle position is shown in equation (10).

$$X_i^{t+1} = X_i^t + V_i^{t+1} \tag{10}$$

The above process indicates that the PSO owns the characteristics of simplicity, ease of implementation, and strong universality. However, due to the fact that particles in PSO belong to the same population, it is easy to form local optimal positions, resulting in the inability to obtain global optimal solutions and the problem of excessive dependence on

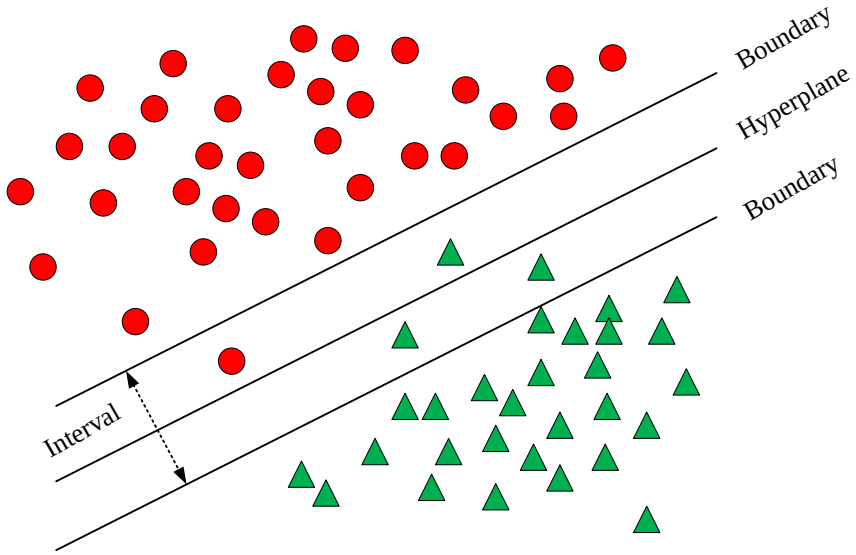


Figure 5: SVM optimal classification hyperplane.

parameters. In response to the above issues, this study adopts a combination of multiple groups and adaptive adjustment of acceleration coefficients to optimize PSO and proposes the MPSO optimization algorithm [21-22]. The MPSO algorithm can effectively improve the division of labor and cooperation among populations. In MPSO, a population contains multiple sub-populations, which in turn contain multiple particles. The MPSO algorithm structure is shown in Figure 4.

In Figure 4, each sub-population is a complete communication and interaction system. All particles within the population can communicate. During algorithm iteration, it is required to discover the optimum $gbest$ for each subgroup, and then find the optimal solution $sbest$ for the entire particle swarm, which is $sbest = \max(gbest_1, gbest_2, ..., gbest_N)$. N is the number of sub-populations. Therefore, in MPSO, the velocity update formula for particles is shown in equation (11).

$$V_i^{t+1} = \omega \times V_{id}^t + c_1 \times rand() \times (pbest_{id}^t - X_{id}^t) + c_2 \times rand() [\lambda (gbest_{kd}^t - X_{id}^t) + \theta (1-\lambda)(sbest_d^t - X_{id}^t)] \tag{11}$$

In equation (11), ω is the inertia weight. $pbest_{id}^t$ and $sbest_{id}^t$ are the historical and global best position of particle i and particle swarm. $pbest_{kd}^t$ is the optimal position of sub-population k so far. c_1, c_2 are the acceleration factors. $rand()$ is a random number between [0, 1]. λ is the classification accuracy of the sample. θ is the number of sub-populations. Due to the impact of inertia weight on the balance of local and global SAs, the value of inertia weight is crucial in PSO algorithm. However, the control effect of fixed inertia weights on global and local search capabilities is limited. Accordingly, this study employs a linear differential descent method to dynamically adjust the inertia weight, thereby enhancing the algorithm's overall SA in the initial stage of iteration and enhancing its local SA in the subsequent phase of iteration. The dynamically adjusted inertia weight is shown in equation (12).

$$\omega(t) = \omega_{\max} - \frac{(\omega_{\max} - \omega_{\min}) * t^2}{t_{\max}^2} \quad (12)$$

In equation (12), t is the current iteration count. $t_{\max}$ is the maximum iterations. $\omega_{\max}$ is the initial inertia weight, which is set to 0.9 in the study. $\omega_{\min}$ is the inertia weight at the maximum iteration count, set to 0.4. The improved particle velocity update formula is shown in equation (13).

$$V_i^{t+1} = \omega(t) \times V_{id}^t + c_1 \times rand() \times (pbest_{id}^t - X_{id}^t) + c_2 \times rand()[\lambda(gbest_{kd}^t - X_{id}^t) + \theta(1-\lambda)(sbest_d^t - X_{id}^t)] \quad (13)$$

In equation (13), $\omega(t)$ is the dynamically adjusted inertia weight. The above is a feature model construction built on the improved PSO. The MPSO first randomly divides the particle swarm into K sub-populations and initializes the sub-population particles randomly. Then, the individual extremum of the particles, the optimal value of each sub-population, and the global optima of the entire PSO are selected. The next step is to determine the optimal value found in the search. If the conditions are met, running is stopped. Otherwise, the speed and location of the particles are continued to update, and the best value is selected to continue running the algorithm until the conditions are met. Finally, the optimal solution of the optimization problem is outputted. After extracting features from video images as described above, this study uses SVM for feature recognition and classification [23-25]. SVM is an algorithm that maps low dimensional data to high-dimensional data to minimize functional. The model is defined as shown in Figure 5.

In Figure 5, SVM is a binary classification model. This model is the nonlinear classifier with the largest interval in the feature space. The learning strategy of SVM is to maximize the interval, which can be formalized as a convex quadratic programming problem. This is equivalent to the minimization problem of a regularized hinge loss function [26-27]. The learning algorithm of SVM is the optimization algorithm for solving convex quadratic programming. This study assumes a sample size of (x_i, y_i). x_i is the input vector. y_i is the corresponding output target. The SVM model initially employs a high-dimensional mapping feature space, which facilitates the identification of a superior hyperplane for the separation of diverse categories of data. Subsequently, it utilizes linear functions within the feature space for function approximation. According to statistical theory, the SVM minimum optimization objective function yields a fitted regression function as shown in equation (14) [28].

$$\min(W, b): \frac{1}{2}W^2 + C\sum_{i=1}^{n} |y_i - [W, \varphi(X) - b]| \quad (14)$$

In equation (14), W is vector data. b is the function threshold. y is the function value after dot product processing. $\varphi(X)$ is an approximation function. C is the penalty coefficient for training model complexity and controlling model loss. The classification of the model can be completed through equation (14). This model can effectively solve problems regardless of the sample size or whether the linear fitting conditions are met, and due to its global strategy, it will not be unable to obtain the optimal solution due to local optima. In the PSO-SVM model, particles dynamically adjust their position and speed through interactions to efficiently capture features in video frames. Each particle represents a potential solution whose position corresponds to the selection of key attributes in feature extraction. The fitness value of the particle is evaluated based on the classification accuracy feedback. Additionally, the particle is able to learn which features are most relevant in tennis serve recognition by comparing with the historical optimal position and the global optimal position. The specific particle parameters, including inertial weights and acceleration factors, influence the exploration and development capabilities of particles, which assist in balancing global search and local search in complex feature spaces. Ultimately, this improves the accuracy and recognition rate of feature extraction.

3 Results

3.1 Analysis of deblurring effect based on video image processing and PSO-SVM

To verify the performance and effectiveness of the research algorithm, this study conducts comparative experiments for analysis. The comparative methods include GAN, Attention Mechanism-GAN (AGAN), and Multi-Scale Convolution (MSC) algorithm. In the improved GAN structure, the number of convolutional layers has been improved to enhance the extraction capability of intricate features. Each convolutional layer employs a 3×3 convolution kernel to facilitate the capture of image details with greater precision. In addition, ReLU and Leaky ReLU activation functions are used interchangeably, aiming to avoid the "dead neuron" phenomenon and improve the model's ability to learn from non-linear features. By removing batch normalization, the study enhances the flexibility of the generator and reduces the introduction of noise. It is imperative to ensure that each algorithm is configured with identical initial settings to minimize variation. Furthermore, the number of iterations must be trained up to 3,000 times. The learning rate of GAN is set to 0.0002. The learning rate of PSO is set to 0.5, and the number of PSO particles is 100. The initial value of inertia weight is 0.9 and dynamically adjusted to 0.4. The acceleration factor is set to 2.0. Performance evaluation indicators include relative subjective and objective indicators. The relative subjective indicator is the subject's evaluation of image quality. Objective indicators are selected as PSNR and SSIM. The higher the PSNR value, the closer the SSIM value is to 1, indicating

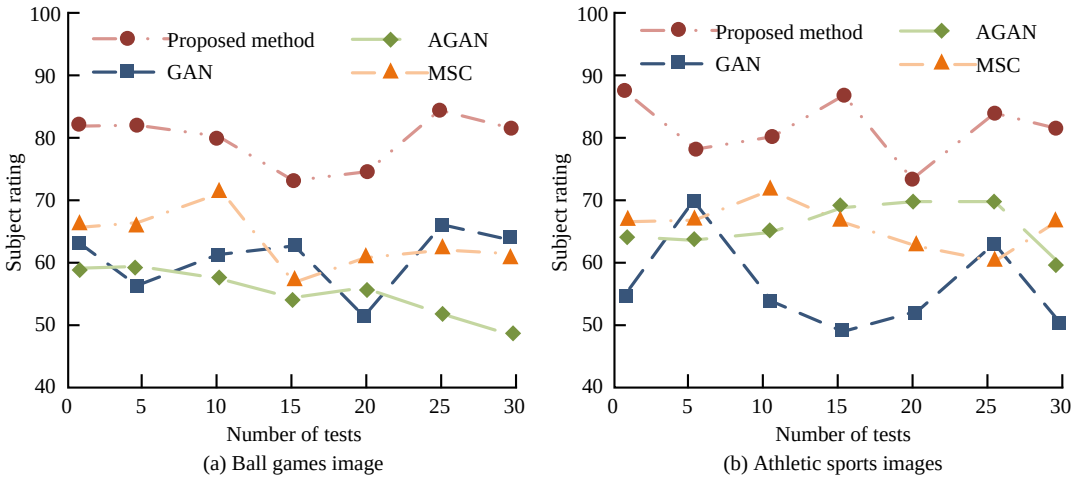


Figure 6: Subjective evaluation score results of the model.

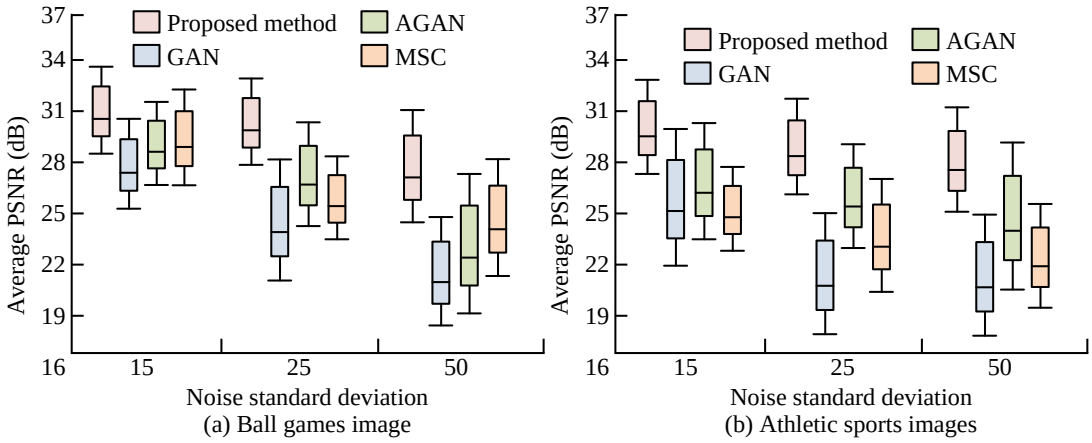


Figure 7: Comparison of average PSNR of different deflurring algorithms.

better image quality, that is, better deblurring effect. The experiment uses sports images from the GOPRO dataset for testing. This dataset divides images into ball sports and track and field sports. In subjective evaluation, the model score results are shown in Figure 6.

Figures 6 (a) and (b) show the subjective evaluation scores of subjects on the deblurring effect of images in ball sports and track and field sports. In Figure 6 (a), the deblurred image under the research method scores the highest, with an average score of 81.16 out of 30 experiments. In the evaluation of 50 subjects, the research method shows better deblurring effect on ball sports images. In Figure 6 (b), the subjective evaluation obtained by the research method is better, and the mean score of the deblurred image is 86.94. The average scores of AGAN, GAN, and MSC do not exceed 75. To compare the PSNR and SSIM performance, this paper adds noise with standard deviations of 15, 25, and 50 to the test original images, respectively, to generate test images for testing the algorithm's deblurring ability. Following the implementation of a research adjustment, the image quality has been markedly enhanced, as evidenced by elevated PSNR and SSIM values. This signifies that the generated image is more closely aligned with the authentic image in terms of structural similarity and clarity. It is particularly well-suited for the processing of dynamic motion scenes. Figure 7 is a boxplot of the mean PSNR of various algorithms.

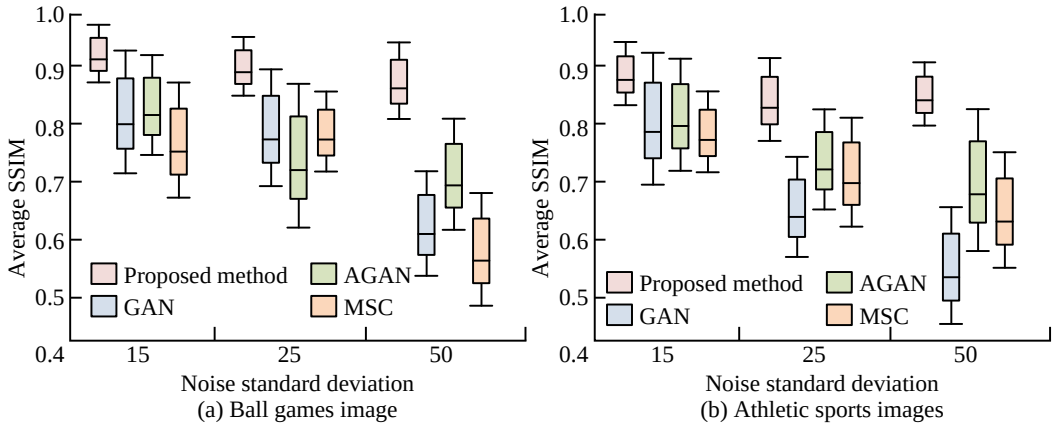


Figure 8: Average SSIM of different deblurring algorithms.

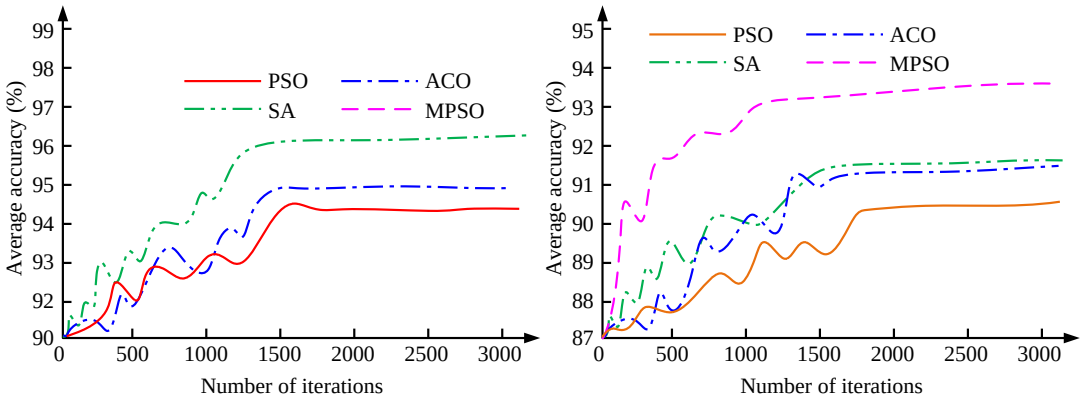
Figures 7 (a) and (b) show the average PSNR performance of algorithms in image deblurring tests for ball sports and track and field sports. The PSNR values of the research methods are consistently the highest. Compared with AGAN, MSC, and GAN, the research method increases the average PSNR by 13.56%, 15.02%, and 18.29% in Figure 7 (a), and by 12.82%, 14.02%, and 22.72% in Figure 7 (b), respectively. This indicates that the difference between the original and the deblurred images is smaller under the research method, indicating that the deblurring effect is better. Figure 8 shows the mean SSIM boxplots of four algorithms.

In the deblurring tests of ball sports and track and field sports images in Figures 8 (a) and (b), the SSIM values of the research method are the highest, and the overall performance is more stable. Compared with AGAN, MSC, and GAN algorithms, in Figure 8 (a), the average SSIM of the research method increases by 8.33%, 12.24%, and 19.90%, while in Figure 8 (b), it increases by 12.34%, 19.38%, and 22.20%. Overall, the research method has the best deblurring effect, followed by AGAN, MSC, and finally GAN. The experimental results validate the effectiveness of this study.

3.2 Feature extraction based on video image processing and PSO-SVM model

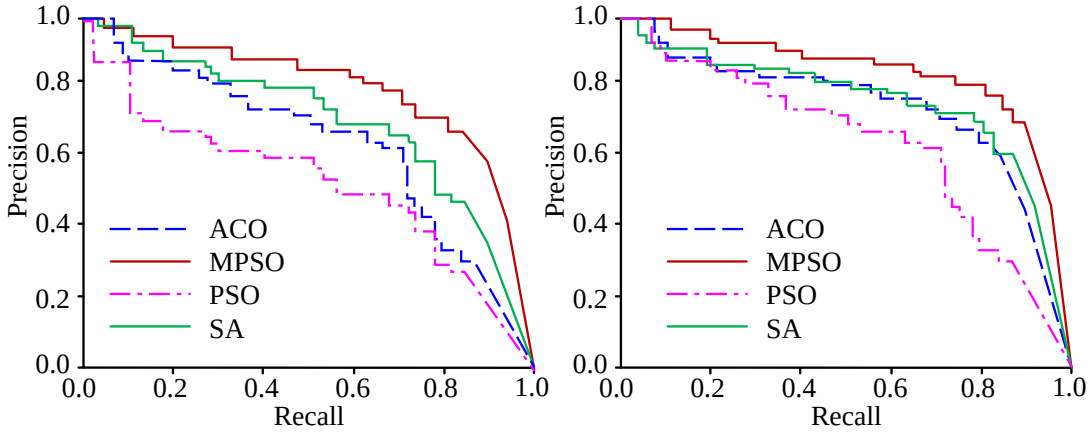
To validate the proposed sports image feature extraction algorithm, this study selects the difficult to solve single peak function Rosenbrock and the easy to trap algorithm in local optima multi-peak function Griebank as the standard test functions. To verify the convergence performance of MPSO, this study compares PSO, Ant Colony Optimization (ACO), and Simulated Annealing (SA). In the experiment, all algorithms are set with the same common parameters, namely population size and dimensionality. The iterations are 3000. Each algorithm is independently run 100 times on each test function, and statistical analysis is conducted on the results of the 100 runs. Figure 9 shows the convergence curve of the obtained algorithm.

Figure 9 shows the convergence curves of the algorithm on Griebank and Rosenbrock. MPSO achieves the highest average accuracy in the shortest number of iterations. In Figure 9 (a), at 850 iterations, MPSO approaches convergence with an average accuracy of 97.31%. Compared to the other three algorithms, MPSO has reduced the number of iterations during convergence by 35.33%,



(a) Test results on Griewank (b) Test results on Rosenbrock

Figure 9: Comparison of algorithm convergence curves.



(a) Ball game images (b) Athletic sports images

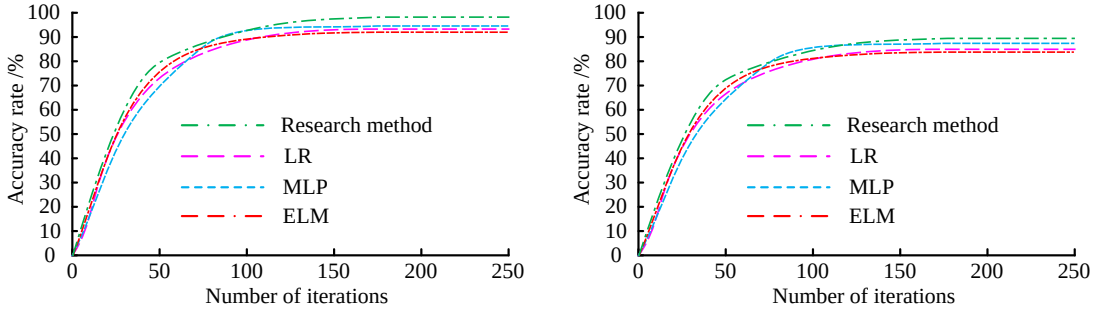
Figure 10: Comparison of algorithm PR curve.

40.52%, and 51.55%. In Figure 9 (b), after 1100 iterations, MPSO approaches convergence with an average accuracy of 92.94%. In contrast, the number of iterations during MPSO convergence decreases by 25.37%, 26.54%, and 41.89%. This verifies that the iteration speed and accuracy of MPSO are superior to SA, ACO, and PSO. To validate the performance of various feature extraction algorithms, this study tests the Precision-Recall (PR) curves of different algorithms, as exhibited in Figure 10.

Figure 10 shows the PR curves of various algorithms in the images of ball sports and track and field sports. In Figure 10 (a), the performance of each algorithm from best to worst is MPSO, SA, ACO, and PSO, respectively. In Figure 10 (b), MPSO has the best feature extraction performance, while ACO and SA have similar performance. ACO has the relatively worst feature extraction performance. Finally, in sports image feature extraction, feature extraction time is also a commonly used evaluation metric, which can be used to determine the efficiency of feature extraction methods.

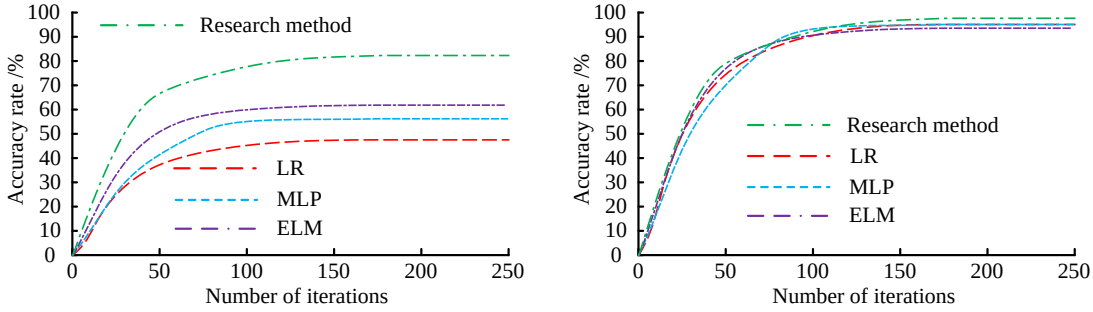
3.3 Behavior analysis based on video image processing and PSO-SVM Model

To verify the effectiveness of the research



(a) Prepare posture behavior analysis accuracy (b) Analysis accuracy of service swing action

Figure 11: Analysis accuracy of preparation posture and serving and swinging behavior.



(a) Batting action behavior analysis accuracy (b) With wave as behavior analysis accuracy

Figure 12: Accuracy analysis of hitting action and swing action behavior.

model in analyzing action behavior in video images, this study collects match videos of tennis players with a total length of 5.6 hours. This study categorizes the serving actions of tennis players in videos into five behaviors, including preparation posture, serving swing, hitting action, swing action, and recovery posture. This study sets the SVM related parameters, with a penalty coefficient of 1 and SVM error accuracy of 0.001. The experiment recognizes five types of actions and uses Logistic Regression (LR), Extreme Learning Machine (ELM), and Multi-Layer Perceptron (MLP) as comparison methods. The analysis results of the preparation posture and serve swing behavior are shown in Figure 11.

Figure 11 (a) shows the accuracy results of the "preparation posture" behavior analysis. The behavior analysis accuracy of the research algorithm has reached 97.2%, which is 5.84%, 7.62%, 8.04%, and 8.16% higher than LR, ELM, and MLP methods. Figure 11 (b) shows the accuracy results of the analysis of the "serve swing" behavior. The accuracy of the research algorithm reaches 90.4%, which is about 1% higher than MLP. This indicates that the comparative algorithm may have difficulty capturing complex motion features due to the limitations of its linear model, and may still be insufficient in capturing diverse and delicate features. The research method adopts advanced feature extraction techniques, which can better identify and classify a small number of posture changes. Figure 12 shows the accuracy analysis results of hitting and swinging movements in tennis serving behavior.

In Figure 12 (a), the behavior analysis accuracy of the research algorithm reaches 82.8%, which is 33.6% higher than LR, 20.8% higher than ELM, and 24.5% higher than MLP. In Figure 12 (b), the behavior analysis accuracy of all four methods reaches over 98%. Therefore,

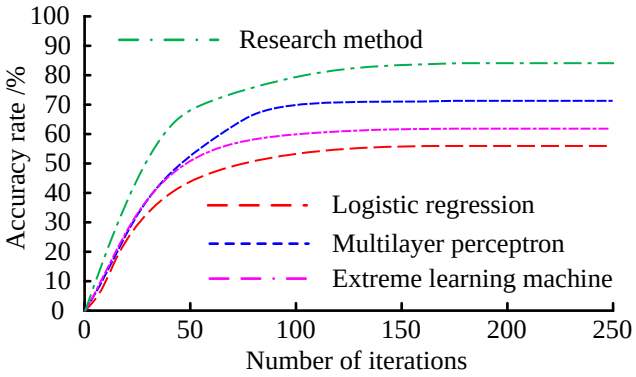


Figure 13: Accuracy of posture recovery behavior analysis.

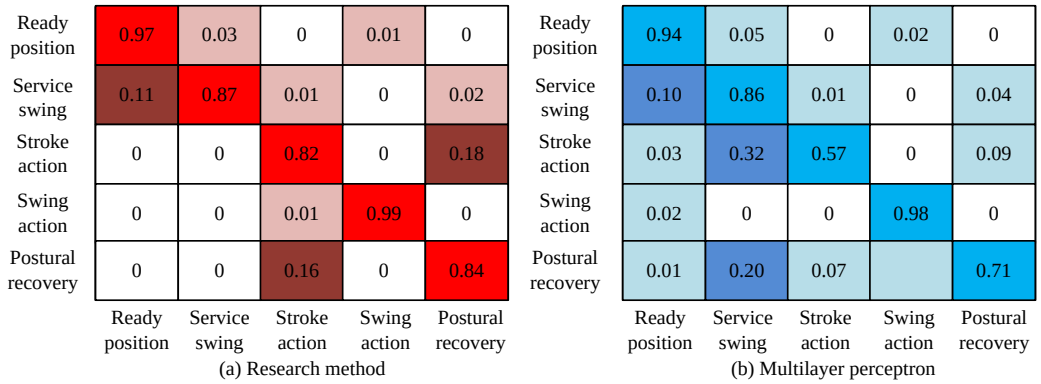


Figure 14: Confusion matrix results of different algorithms.

the research algorithm significantly improves the recognition ability of "hitting action", possibly due to its sensitivity to action details and dynamic changes. Figure 13 shows the accuracy results of posture recovery analysis in tennis serving behavior.

In Figure 13, the analysis accuracy of the research algorithm for "posture recovery" behavior reaches 84.7%, LR is 58.2%, ELM is 62.3%, and MLP is 70.3%. The research algorithm has obvious advantages in behavior analysis, indicating that it has better feature extraction and action classification performance. This study proposes methods to analyze the recognition performance of MLP, as shown in Figure 14.

Figure 14 shows the confusion matrix results of the research method and MLP. In Figure 14 (a), the overall recognition accuracy of the research method reaches 91.24%, and in the dynamic classification effect, the main manifestation is mutual interference. There is an 18% probability that the "hitting action" will be misidentified as "posture recovery". There is a 16% chance that the "posture recovery" action will be misidentified as a "hitting action". This shows that there is some overlap and confusion in the recognition between these two actions. This phenomenon may be attributed to the fact that the player's posture during the act of serving is analogous to that observed during the subsequent recovery phase. This results in an insufficient degree of feature extraction, which in turn hinders the ability to distinguish between the two movements. To reduce misclassification, it would be beneficial to consider optimizing the classifier threshold in SVM to adjust the decision boundary, thereby improving the ability to distinguish between these actions. In Figure 14 (b), the MLP method has lower recognition performance than the research method in dynamic actions. Although its dynamic actions have significant interference, the recognition accuracy of the research method is over 80%. This indicates that the research model owns good anti-interference ability and high recognition accuracy in action recognition. This study compares the computational efficiency of different models. The details are listed in Table 2, using model runtime, GPU usage, and memory usage as evaluation metrics.

Table 2: Comparison results of model calculation efficiency.

Algorithm	Run time (s)	GPU Usage (%)	Memory usage (MB)
Research method	45.2	75.3	512
AGAN	55.6	80.1	600
GAN	62.8	82.4	650
MSC	50.4	78.5	580
LR	40.1	72.0	500
ELM	42.3	71.5	490
MLP	43.0	73.2	495

In Table 2, the research method shows superior performance in terms of running time, significantly reducing computation time compared to other algorithms. Its GPU usage rate is 75.3%, which is relatively lower compared to algorithms such as AGAN and WGAN, indicating that the research model is more efficient in utilizing computing resources. In terms of memory usage, the research method has also shown good optimization ability, maintaining a usage of 512MB, which is reduced compared to other methods. To further reduce processing requirements, this study can implement effective memory management strategies to optimize the allocation of computing resources, thereby reducing memory consumption and improving computing speed. For example, small batch processing and dynamic memory allocation. These strategies not only improve the operational efficiency of the model but also ensure that good performance can be maintained when processing large-scale data. The research method is further compared with the Recurrent Neural Network (RNN) and Convolutional Neural Network Ensemble Long Short-term Memory Network (CNN-LSTM) models. The results are shown in Table 3.

The results in Table 3 show that the research method outperforms RNN and CNN-LSTM in several performance indexes. Specifically, the accuracy of the research method is as high as 91.24%, which is significantly higher than the 85.1% of RNN and 87.3% of CNN-LSTM. In terms of image quality, the PSNR and SSIM of the research method are 35.12 dB and 0.93, respectively, indicating a robust defuzziness effect and the capacity to preserve structural details. In comparison, the corresponding indexes of RNN and CNN-LSTM are 32.5 dB and 0.87, and 33.8 dB and 0.89, respectively. It shows that the latter is relatively weak in image quality. In addition, the research method has a relatively low runtime and memory usage (45.2s and 512MB), showing better computational efficiency. Overall, these results validate the advances in accuracy and efficiency of the research method, indicating its application potential in complex dynamic scenarios.

4 Discussion

When deblurring TVI, the average subjective score of the research method in the GOPRO dataset was 81.16 points, and the PSNR value increased by 13.56%, 15.02%, and 18.29% compared to AGAN, MSC, and traditional GAN. The potential benefit of studying the model lies in optimizing the adversarial learning mechanism between the generator and discriminator. This could result in the generator paying greater attention to image details and effectively reducing blurring phenomena. By introducing multiple convolutional layers and activation functions, the model's ability to learn complex features was enhanced, thereby improving the clarity and realism of generated images. This study demonstrated significant advantages in feature extraction using PSO-SVM. The accuracy of the MPSO algorithm reached 97.31%, and its convergence speed improved by 35.33%, 40.52%, and 51.55

Table 3: Comparative results of advanced nature

Method	Accuracy (%)	PSNR (dB)	SSIM	Run time (s)	GPU Usage (%)	Memory Usage (MB)
Research method	91.24	35.12	0.93	45.2	75.3	512
RNN	85.1	32.5	0.87	55.0	78.2	520
CNN-LSTM	87.3	33.8	0.89	50.5	76.5	510

% Compared to traditional PSO, ACO, and SA. The advantage of the research method lied in the introduction of multiple population PSO strategies, which enable different sub-populations to share information and optimal solutions. By using linear differential descent, the inertia weight of particles was dynamically adjusted, which could enhance the global SA in the early stage and local SA in the later stage, thus improving the stability and efficiency of the feature extraction process. In the action recognition experiment, the recognition accuracy of the research method reached 91.24%, and in feature recognition such as hitting action and posture recovery, it was higher than classical algorithms such as LR, ELM, and MLP. The recognition accuracy for the "preparation posture" behavior reached 97.2%, while the recognition accuracy for the "hitting action" was as high as 82.8%. By improving the residual network structure, the model could effectively extract deep dynamic features from action sequences, enhancing its ability to capture complex motion trajectories. This enabled the action classification model to more accurately identify various behaviors of different athletes. In similar studies, Fréjus et al. proposed a behavior recognition model based on neural networks, which showed good recognition accuracy in posture recognition [29]. Luo Z et al. proposed a behavior recognition model grounded on multi-layer LSTM, which demonstrated good performance in specific applications [30]. However, the above research cannot effectively capture small differences between actions when dealing with complex dynamic sequences, which can easily lead to misidentification in similar actions. Compared with it, the research method can comprehensively handle complex dynamic scenes and has high flexibility and robustness, providing a new approach for motion behavior analysis.

In complex dynamic scenes, the recognition challenges of moving images are often caused by motion blur, background interference, and illumination changes. The proposed model shows good robustness in various complex situations, mainly due to the design of adaptive feature extraction and adversarial learning mechanism. The introduction of PSO algorithm makes the feature extraction process more flexible and adaptive. When dealing with dynamic scenes, the particle can dynamically adjust its position to capture the key movement trajectory of the player and the ball, which improves the relevance and usability of the feature. By optimizing the adversarial training process between generator and discriminator, the enhanced GAN model is capable of not only generating high-quality images but also of producing more robust mapping relationships in the feature extraction process. This mechanism serves to mitigate the impact of noise and background changes on the results, thereby enhancing the overall robustness of the model.

5 Conclusion

With the rapid growth of video technology, the processing and behavior analysis of TVI have become increasingly important. To address the challenges of video image blurring and feature extraction, this paper constructed a comprehensive method built on improved GAN and PSO-SVM algorithms. For fuzzy processing, the improved GAN significantly improved the ability to recover details through the introduction of multiple convolutional structures and various activation functions. In terms of feature extraction, a combination of PSO-SVM algorithm was adopted, which integrated the advantages of PSO and SVM to further enhance the efficiency and accuracy of feature extraction. The experimental test results demonstrated that the research method could more accurately capture important features in motion images and effectively reduce blurring phenomena. Moreover, the model could maintain high accuracy even in fast dynamic scenarios. This indicates that the research model has strong anti-interference ability, which is helpful for accurate behavior analysis in complex environments. Although this study has achieved significant success in improving the quality of video image processing, there are still some shortcomings. Under complex background interference, the recognition performance of the model may be limited, and improving the algorithm's performance in complex background scenes is still a problem to be solved. Therefore, future research directions can focus on combining deep learning and reinforcement learning methods to enhance processing capabilities for complex dynamic scenes.

References

[1] Deqiang Cheng, Jiansheng Qian, Xingge Guo, Qiqi Kou, Feixiang Xu, Jun Gu, Yachao Gao, and Jinsheng Zhao. Review on key technologies of AI recognition for videos in coal mine. Coal science and technology, 51(2):349-365, 2023. https://doi.org/10.13199/j.cnki.cst.2022-0359

[2] Chandravva Hebbi, and H. R. Mamatha. Comprehensive dataset building and recognition of isolated handwritten kannada characters using machine learning models. Artificial intelligence and applications, 1(3):179-190, 2023. https://doi.org/10.47852/bonviewAIA3202624

[3] Teymoor Ali, Deepayan Bhowmik, and Robert Nicol. Domain-specific optimisations for image processing on FPGAs. Journal of signal processing systems, 95(10):1167-1179, 2023. https://doi.org/10.1007/s11265-023-01888-2

[4] Yunzhong Hou, Zhongdao Wang, Shengjin Wang, and Liang Zheng. Adaptive affinity for associations in multi-target multi-camera tracking. IEEE transactions on image processing,

31(10):612-622, 2021. https://doi.org/10.48550/arXiv.2112.07664

[5] Lintong Zhang, David Wisth, Marco Camurri, and Maurice Fallon. Balancing the budget: Feature selection and tracking for multi-camera visual-inertial odometry. IEEE robotics and automation letters, 7(2):1182-1189, 2021. https://doi.org/10.48550/arXiv.2109.05975

[6] Qinglong Ding, and Zhenfeng Ding. Machine learning model for feature recognition of sports competition based on improved TLD algorithm. Journal of intelligent & fuzzy systems, 40(2):2697-2708, 2021. https://doi.org/10.3233/JIFS-189312

[7] Jiaxu Zhang, Gaoxiang Ye, Zhigang Tu, Yongtao Qin, Qianqing Qin, Jinlu Zhang, and Jun Liu. A spatial attentive and temporal dilated (SATD) GCN for skeleton-based action recognition. CAAI transactions on intelligence technology, 7(1):46-55, 2022. https://doi.org/10.1049/cit2.12012

[8] Xiaoguang Zhu, Ye Zhu, Haoyu Wang, Honglin Wen, Yan Yan, and Peilin Liu. Skeleton sequence and RGB frame based multi-modality feature fusion network for action recognition. ACM transactions on multimedia computing, communications, and applications (TOMM), 18(3):1-24, 2022. https://doi.org/10.48550/arXiv.2202.11374

[9] Arti Ranjan, and M. Ravinder. VAEWGAN-NCO in image deblurring framework using variational autoencoders and Wasserstein generative adversarial network. Signal, image and video processing, 18(5):4447-4456, 2024. https://doi.org/10.1007/s11760-024-03085-5

[10] Chuang Li, and Zhizhong Mao. Generative adversarial network-based real-time temperature prediction model for heating stage of electric arc furnace. Transactions of the institute of measurement and control, 44(8):1669-1684, 2022. https://doi.org/10.1177/01423312211052213

[11] Yupeng Song, Xu Hong, Jiecheng Xiong, Jiaxu Shen, and Zekun Xu. Probabilistic modeling of long-term joint wind and wave load conditions via generative adversarial network. Stochastic environmental research and risk assessment, 37(2):2829-2847, 2023. https://doi.org/10.1007/s00477-023-02421-4

[12] Zhiwu Shang, Jie Zhang, Wanxiang Li, Shiqi Qian, Jingyu Liu, and Maosheng Gao. A novel small samples fault diagnosis method based on the self-attention wasserstein generative adversarial network. Neural processing letters, 55(5):6377-6407, 2023. https://doi.org/10.1007/s11063-022-11143-7

[13] Priyanshu Mahey, Nima Toussi, Grace Purnomu, and Anthony Thomas Herdman. Generative adversarial network (GAN) for simulating electroencephalography. Brain topography, 36(5):661-670, 2023. https://doi.org/10.1007/s10548-023-00986-5

[14] Angelo Lorusso, Barbara Messina, and Domenico Santaniello. The use of generative adversarial network as graphical support for historical urban renovation. ICGG 2022 - proceedings of the 20th international conference on geometry and graphics, 146(1):738-748, 2022. https://doi.org/10.1007/978-3-031-13588-0_64

[15] Deepa Kumari, S. K. Vyshnavi, Rupsa Dhar, B. S. A. S. Rajita, Subhrakanta Panda, and Jabez Christopher. Smart GAN: A smart generative adversarial network for limited imbalanced dataset. The journal of supercomputing, 80(14):20640-20681, 2024. https://doi.org/10.1007/s11227-024-06198-3

[16] Yaxiang Fan, Gongjian Wen, Fei Xiao, Shaohua Qiu, and Deren Li. Detecting anomalies in videos using perception generative adversarial network. Circuits, systems, and signal processing, 41(2):994-1018, 2022. https://doi.org/10.1007/s00034-021-01820-8

[17] Canran Zhang, Jianping Dou, Shuai Wang, and Pingyuan Wang. Hybrid particle swarm optimization algorithms for cost-oriented robotic assembly line balancing problems. Robotic intelligence and automation, 43(4):420-430, 2023. https://doi.org/10.1108/RIA-07-2022-0178.

[18] Naveed ur Rehman, and Muhammad Uzair. Concentrator shape optimization using particle swarm optimization for solar concentrating photovoltaic applications. Renewable energy, 184(5):1043-1054, 2022. https://doi.org/10.1016/j.renene.2021.12.015

[19] Yue Li, Jianfang Qi, Xiaoquan Chu, and Weisong Mu. Customer segmentation using K-means clustering and the hybrid particle swarm optimization algorithm. The computer journal, 66(4):941-962, 2022. https://doi.org/10.1093/comjnl/bxab206

[20] Vahid Goodarzimehr, Fereydoon Omidinasab, and Nasser Taghizadieh. Optimum design of space structures using hybrid particle swarm optimization and genetic algorithm. World journal of engineering, 20(3):591-608, 2023. https://doi.org/10.1108/WJE-05-2021-0279

[21] Fanyi Duanmu, Dian Ning Chia, and Eva Sorensen. A combined particle swarm optimization and outer approximation optimization strategy for the optimal design of distillation systems. Computer aided chemical engineering, 49(3):1315-1320, 2022. https://doi.org/10.1016/B978-0-323-85159-6.50219-0

[22] Yongjie Zhu, Jiajun Chen, Ling Mao, and Jinbin Zhao. A noise-immune model identification method for lithium-ion battery using two-swarm cooperative particle swarm optimization algorithm based on adaptive dynamic sliding window. International journal of energy research, 46(3):3512-3528, 2022. https://doi.org/10.1002/er.7401

[23] Min Gi, Shugo Suzuki, Masayuki Kanki, Masanao Yokohira, Tetsuya Tsukamoto, Masaki Fujioka, Arpamas Vachiraarunwong, Guiyu Qiu, Runjie Guo, and Hideki Wanibuchi. A novel support vector machine-based 1-day, single-dose prediction model of genotoxic hepatocarcinogenicity in rats. Archives of toxicology, 98(8):2711-2730, 2024. https://doi.org/10.1007/s00204-024-03755-w

[24] Vamsi Alla, Upendra Kumar Sahoo, and Rabi Narayan Behera. Seismic liquefaction analysis of MCDM weighted SPT data using support vector machine classification. Iranian journal of science and technology, transactions of civil engineering, 48(4):2293-2303, 2024. https://doi.org/10.1007/s40996-023-01293-6

[25] Laith Abualigah, Saba Hussein Ahmed, Mohammad H. Almomani, Raed Abu Zitar, Anas Ratib Alsoud, Belal Abuhaija, Essam Said Hanandeh, Heming Jia, Diaa Salama Abd Elminaam, and Mohamed Abd Elaziz. Modified aquila optimizer feature selection approach and support vector machine classifier for intrusion detection system. Multimedia tools and applications, 83(21):59887-59913, 2024. https://doi.org/10.1007/s11042-023-17886-2

[26] Ning Chu, Weimin Kang, Xinhua Yao, and Jianzhong Fu. Online roundness prediction of grinding workpiece based on vibration signals and support vector machine. The international journal of advanced manufacturing technology, 126(5/6):2733-2743, 2023. https://doi.org/10.1007/s00170-023-11206-6

[27] Hossein Moosaei, Ahmad Mousavi, Milan Hladík, and Zheming Gao. Sparse L1-norm quadratic surface support vector machine with Universum data. Soft computing, 27(9):5567-5586, 2023. https://doi.org/10.1007/s00500-023-07860-3

[28] Jagadeesh Basavaiah, and Audre Arlene Anthony. A pragmatic approach for infant cry analysis using support vector machine and random forest classifiers. Wireless personal communications, 137(4):2269-2280, 2024. https://doi.org/10.1007/s11277-024-11491-8

[29] Fréjus A. A. Laleye, and Mikaël A. Mousse. Attention-based recurrent neural network for automatic behavior laying hen recognition. Multimedia tools and applications, 83(22):62443-62458, 2024. https://doi.org/10.1007/s11042-024-18241-9

[30] Zhenmin Luo, Lidong Zhang, and Zeyang Song. Multistep prediction of CO in the extraction zone based on a fully connected long short-term memory network. Journal of tsinghua university (science and technology). 64(6):940-952, 2024. https://doi.org/10.16511/j.cnki.qhdxxb.2024.22.011