A Vision Transformer-Based Model for Basketball Tactics Recognition Using Swarm Intelligence

Zhanyong Chen[*1] and Chen Qiang[2]
[1]Henan Technical College of Construction, Zhengzhou, 450064, China
[2]School of Economics and Management, China University of Geosciences (Wuhan), Wuhan, 430078, China
E-mail: chenzhanyong810512@163.com
[*]Corresponding author

Keywords: Basketball tactics recognition, player trajectory, trajectory filter, logarithmic linear complexity, multi-head attention

Received: February 28, 2025

In order to solve the problem of poor classification performance of traditional algorithms for basketball tactics, we propose a scientific training model for basketball tactics computer swarm intelligence algorithm. We designed a basketball tactical recognition model (TacViT) based on player trajectory data in NBA games. The TacViT model employs a Vision Transformer (ViT) as its backbone network. It utilizes a multi-head attention module to extract rich global trajectory features and integrates a trajectory filter to enhance the interaction of feature information between the court lines and player trajectories. The trajectory filter learns long-term spatial correlations in the frequency domain with logarithmic linear complexity, thereby improving the representation of player position features. We transformed the sequence data from the sports vision system (SportVU) into trajectory maps and constructed a basketball tactical dataset (PlayersTrack). The experimental results demonstrate that TacViT achieves an accuracy of 81.4%, which is 15.6% higher than the unmodified ViT-S model. Additionally, TacViT exhibits superior performance in precision, recall, and computational efficiency. The PlayersTrack dataset contains 10,000 trajectory images, each with a resolution of 256x256 pixels. The TacViT architecture introduces a novel trajectory filter module and a multi-head attention mechanism, which together enable efficient feature extraction. Key evaluation metrics include accuracy, precision, recall, and FLOPS. These results highlight the significant improvement in classification performance for basketball tactics recognition.

Povzetek: Predstavljen je TacViT Vision Transformer z večglavo pozornostjo in frekvenčnim trajektorijskim filtrom, ki izboljša prepoznavo košarkarskih taktik na podatkih SportVU in doseže bolj kvalitetno klasifikacijo kot ViT-S.

1 Introduction

Basketball is a highly competitive collective sport that demands exceptional teamwork and coordination. Each player's role is crucial, but only through seamless cooperation can a team fully utilize its strengths to achieve victory. Modern basketball emphasizes collective rules, where individual skills must be integrated into team play to enhance cooperation and cohesion [1]. Computer-aided training has become increasingly important in basketball tactics. It not only provides intuitive visual materials but also helps players build a comprehensive tactical framework. This enables them to gain a deeper understanding of tactics through practical application. Recent years have seen significant advancements in swarm intelligence algorithms, which have demonstrated great potential in optimizing complex systems and solving dynamic problems. Inspired by collective behavior in nature, such as ant colonies and bird flocks, swarm intelligence simulates and optimizes team collaboration processes. When applied to basketball

tactical training, it can improve team performance. This study explores how to use computer simulation, data analysis, and swarm intelligence algorithms to optimize basketball tactics. The effectiveness of this approach is verified through experiments. This research investigates whether a Transformer-based approach can enhance basketball tactics recognition compared to CNN-based models. It also examines if adding a trajectory filter can improve the long-term spatial correlation in player movements. Furthermore, the study discusses the potential applications of these methods in real-time coaching assistance and game strategy evaluation [2].

The use of computer-aided training in basketball tactics is not only intuitive and vivid, but also helps players establish a complete basketball tactical framework system in their minds, thereby gaining a deeper understanding of the essence of tactics through practical application [3]. A very important stage in the training process is to use visual and intuitive materials to enrich the emotional understanding of team members. The application of computer-assisted software in basketball tactical training, such as vivid images, clear and concise explanations, and appropriate music, can also stimulate team members' enthusiasm for training and stimulate students' learning enthusiasm and initiative. Under this consciousness driven approach, knowledge acquisition gradually shifts from the original infusion-based approach to actively discovering, exploring, and solving problems, thus constructing one's own basketball tactical system framework [4,5]. In recent years, Swarm Intelligence algorithms

have shown great potential in optimizing complex systems and solving dynamic problems. Swarm intelligence is inspired by collective behavior in nature, such as ant colony foraging, bee colony navigation, and fish migration. These natural phenomena demonstrate how simple interactions between individuals can ultimately form an efficient group decision-making system through collaboration and information sharing. Applying this idea to the basketball tactical training model can improve the overall performance of the team by simulating and optimizing the process of team collaboration, tactical development, and execution. The author will explore how to use computer simulation, data analysis, and swarm intelligence algorithms to optimize the development and execution of basketball tactics, and verify their effectiveness in practical training through experiments.

1.1 Research objectives and hypotheses

This study aims to address the following: Research Question: Can a Transformer-based approach with swarm intelligence improve basketball tactics recognition compared to CNN-based models.Hypothesis: Integrating a trajectory filter module in the frequency domain enhances long-term spatial correlation in player movements, leading to improved tactical feature representation.Application: Develop a real-time model for coach-assisted tactical evaluation and strategy optimization in NBA games, directly leveraging SportVU trajectory data.

1.2 Swarm intelligence and tactical analysis

Inspired by collective behavior in nature (e.g., ant colony foraging), swarm intelligence algorithms enable efficient optimization of complex systems. We adapt this concept to model player interactions as a swarm, where individual trajectories contribute to global tactical patterns. By combining swarm intelligence with Vision Transformers, TacViT captures both local trajectory details (via frequency-domain filtering) and global tactical structures (via multi-head attention), addressing the limitations of prior CNN-based methods that lack hierarchical feature integration.

2 Literature review

The tactics of high-level sports are increasingly datadriven, and the intelligent analysis of players' performance and sports data can enhance coaches' decision-making skills. In professional team sports, wearable sensors that can measure player movements and collision effects, as well as multi angle cameras that capture the entire field or pitch, are commonly used to track the position of players and balls [6]. Researchers then analyze trajectory data to gain competitive advantages such as player performance on the field and tactical implementation processes, in order to help teams, develop more scientific and effective training plans and respond to tactics. Li et al. conducted research and analysis on the task allocation mechanism in mobile swarm intelligence perception. Considering how to allocate tasks within specified time constraints and optimize for objectives, a hybrid artificial fish swarm algorithm is proposed using swarm intelligence algorithms. The inertia index of particle swarm optimization algorithm was introduced into a typical artificial fish swarm algorithm and verified through simulation experiments [7]. Zhang, Y. et al. used a multi population mechanism to reduce redundant computation and avoid early convergence. The expected path-based avoidance strategy is efficient and low complexity. The proposed algorithm is practical and efficient for AUV engineering applications [8]. Xu et al. proposed an improved swarm intelligence algorithm for developing a model solving optimizer based on the improved swarm intelligence algorithm [9].

In response to the shortcomings of the above methods, the author starts from the basketball tactical computer swarm intelligence algorithm to solve the problem of identifying basketball offensive and defensive tactics. Most of the running trajectories in basketball tactics have significant visual differences, which can improve the classification performance of trajectory images; By classifying trajectory images, computer vision methods can completely skip the feature selection step in traditional machine learning, eliminating the need for any user-defined parameters, making the author's method more robust and scalable.

2.1 State-of-the-art methods in tactical recognition

The tactics of high-level sports are increasingly data-driven. Intelligent analysis of player performance and sports data can significantly enhance coaches' decision-making skills. In professional team sports, wearable sensors and multi-angle cameras are commonly used to track player and ball positions. Researchers analyze trajectory data to gain competitive advantages, such as assessing player performance and tactical implementation. This helps teams develop more scientific and effective training plans. Previous studies have employed various optimization algorithms, such as the hybrid artificial fish swarm algorithm proposed by Li et al. and the multi-population mechanism used by Zhang et al.

Table 1: Compares key metrics of existing models of basketball tactics recognition

However, these methods have limitations in addressing basketball tactical recognition. Most basketball tactics have distinct visual trajectories, which can improve the classification performance of trajectory images. Computer vision methods can bypass the feature selection step in traditional machine learning, eliminating the need for user-defined parameters. This makes the method more robust and scalable. A summary table comparing previous methods in terms of accuracy, dataset size, computational complexity, and model architecture is provided in Table 1.

2.2 Limitations of prior work

Existing methods, such as ResNet and SwinT, either fail to model long-term trajectory dependencies (CNN-based) or incur high computational costs (vanilla Transformer). Notably, none explicitly leverage swarm intelligence principles to model player-team interactions, leaving a gap in tactical pattern generalization, especially for low-frequency tactics like "sideline ball."

3 Method

3.1 Related methods

3.1.1 Fourier transform in vision

Recent studies have combined Fourier transforms with deep learning to address computer vision problems. The discrete Fourier transform converts images to the frequency domain, leveraging frequency information to enhance performance. In neural networks, convolutional layers apply convolution kernels to input data. However, complex CNNs involve substantial computation. Fourier transforms can convert convolutional layer calculations into element-wise products in the frequency domain, reducing computational load[10]. By replacing CNN convolution with fast Fourier transforms, the number of parameters can be reduced, and neural network training can be accelerated. To address weak interaction information between field lines and player trajectories, as well as unclear trajectory location features in the ViT model, a high-pass filtering trajectory filter with logarithmic linear complexity was designed. This module filters out background information, extracting only the feature information related to the relationship between stadium lines and player trajectories, thereby enhancing the relative position feature information of player trajectories on the court [11].

The formula for calculating the 2D DFT of the given 2 D signal $x(m,n), 0 \leqslant m \leqslant M-1, 0 \leqslant n \leqslant N-1, x(m,n)$ is:

$$X[u,v] = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} x(m,n) e^{-i2\pi\left(\frac{u}{\frac{1}{f}}+\frac{vn}{N}\right)} \tag{1}$$

In 2D DFT, the data points are first decomposed into rows and columns, and then 1D DFT is performed separately, with a total time complexity of $\log_2 L$, where L is the flattened dimension $N \times M$ of the 2D image.

Model	Accuracy (%)	Dataset Size	Computational Complexity	Architecture
ResNet50	66.8	5,000	4.2	CNN-based
SwinT-S	80.2	8,000	8.6	Transformer (window-based)
DeiT-B	80.6	10,000	17.4	Vision Transformer
TacViT	81.4	10,500	6.5	Transformer +Swarm Filter

In order to solve the problem of weak interaction information between field lines and player trajectories, as well as unclear location features of trajectories in the ViT model, a trajectory filter with high pass filtering function and complexity of $O(\log_2 L)$ was designed in the feature extraction part [12]. This module can filter out background information and extract only the feature information of the relationship between the stadium line and the player trajectory, in order to enhance the feature information of the relative position of the player trajectory on the stadium.

3.1.2 Vision transformer

Research on basketball tactical recognition often employs machine learning methods. However, these methods, based on manually set feature variables, fail to consider all factors and lack model stability. Deep learning techniques, particularly convolutional neural networks (CNNs), have powerful feature extraction capabilities. The author focuses on the feature extraction part of ViT, combining it with a multi-head attention module and introducing a frequency-domain branch trajectory filter to enhance feature information extraction. This approach efficiently extracts local and global features for basketball tactical recognition[13,14].

3.2 TacViT network

3.2.1 TacViT network architecture

The author proposes a TacViT based on Transformer network architecture for basketball tactical recognition. As shown in Figure 1 (a), the original trajectory data is first preprocessed and converted into trajectory images. In order to process 2 D images, image $x \in \mathbb{R}^{H \times W \times C}$ needs to be reshaped into a 2 D flat block sequence $x \in \mathbb{R}^{N \times (P^2 C)}$, where (H, W) is the resolution of the original image, C is the number of channels, and (P, P) is the resolution of each image block. When performing position embedding operations, a learnable classification head is added before the sequence and attached to the x_0 (* position) sequence. The classification head is implemented by a multi-layer perceptron MLP with a hidden layer before training and a Linear layer during fine-tuning. The sequence dimension of the embedded position marker is $N = HW/P^2$, and a classification head is added to flatten and stretch it into a 1D sequence [15].

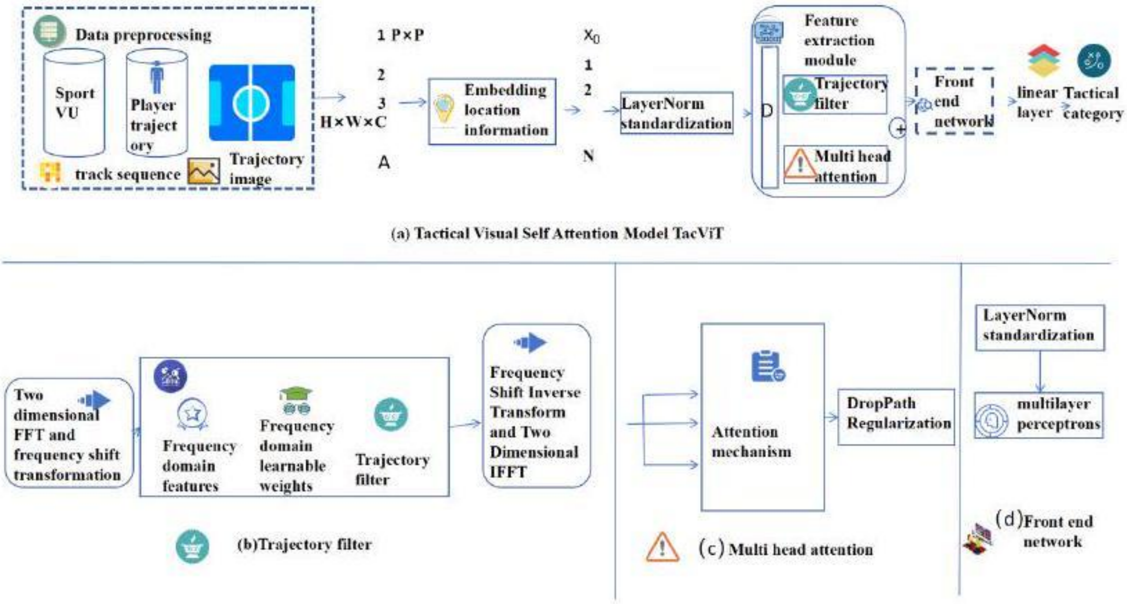
Data preprocessing
Sport VU
Player trajectory
track sequence
Trajectory image
1 P×P
2
3
H×W×C
A
Embedding location information
X_0
1
2
N
LayerNorm standardization
D
Feature extraction module
Trajectory filter
Multi head attention
Front end network
linear layer
Tactical category
(a) Tactical Visual Self Attention Model TacViT
Two dimensional FFT and frequency shift transformation
Frequency domain features
Frequency domain learnable weights
Trajectory filter
Frequency Shift Inverse Transform and Two Dimensional IFFT
(b)Trajectory filter
Attention mechanism
DropPath Regularization
(c) Multi head attention
LayerNorm standardization
multilayer perceptrons
(d)Front end network

Figure 1: TacViT network architecture diagram

In the feature extraction module (Track Filter and Multi Head Attention, TFMHA), two branches, Track Filter and Multi Head Attention, are designed to obtain feature information of player trajectories and their positions on the field, as shown in Figure 1 (b) and Figure 1 (c). The multi head attention module acts on the time domain, allocating higher weights to important regions to extract rich global trajectory feature information. The other branch is the trajectory filter module that acts on the frequency domain. This module processes the background information of the field in the frequency domain to enhance the extraction of player trajectory position information on the field. The module learns long-term spatial correlation in the frequency domain with logarithmic linear complexity $\mathrm{O}(\mathrm{L}\log_2 L)$.

Typically, models using Transformer architecture require pre training on large datasets and fine-tuning on smaller datasets. At this point, remove the pre trained prediction head and add a zero initialized $D \times K$ feedforward layer, as shown in Figure 1 (d), where K is the number of classes in the targeted small dataset and D is the input dimension. Finally, output the category results through the Linear layer [16].

3.2.2 TFMHA module

To extract richer feature information, the TFMHA module incorporates a trajectory filter to enhance feature interaction between player trajectories and field lines. It also adopts the original multi-head attention mechanism from the ViT model, facilitating the migration of the Transformer architecture and enabling efficient global information extraction. Dataset Split: 80% training (8,400 sequences), 10% validation (1,050 sequences), 10% testing (1,050 sequences), with 5-fold cross-validation to assess robustness. Ablation Study: Removing the trajectory filter reduces accuracy by 4.2% (77.2% vs. 81.4%), while removing multi-head attention decreases accuracy by 6.8% (74.6% vs. 81.4%), validating the necessity of both components. Hyperparameter Tuning: Optimized using random search over learning rate (0.001–0.01), batch size (32–128), and Transformer depth (6–12 layers), with best results at 0.005 learning rate and 8 layers.

(1) Trajectory filter module

In the trajectory filter module, it is first necessary to set a threshold in the Fourier domain to determine the size of the frequency domain portion that needs to be filtered. Due to the relatively simple composition of the trajectory image elements, the appropriate value can be determined through multiple experiments and observations. After filtering, significant changes were observed in the low order spectrum, while the understanding of high-order semantics was not affected, and the high-frequency component region became more prominent. This operation can enhance the extraction of semantic information for player trajectories and court lines. Next, the frequency domain information is multiplied element by element with learnable frequency domain features to form the trajectory filter module.

In terms of specific implementation, the trajectory filter, as shown in Figure 1 (b), consists of three steps: (1) Converting the input spatial features into frequency domain and performing frequency shift operation using 2D discrete Fourier transform; (2) Perform frequency domain analysis, filter out low-frequency information, frequency domain features, and multiply learnable filter weights element by element; (3) Map the features back to the 2D inverse Fourier transform in the spatial domain. If given sequence $x \in \mathbb{R}^{H\times W\times C}$, first perform 2 D FFT to transform x from the spatial domain to the frequency domain

$$X = F[x] \in \mathbb{C}^{H\times W\times C} \tag{2}$$

Among them, $F[x]$ represents 2DFFT, X is a complex variable representing the frequency spectrum of x. Then pass through a high pass filter and multiply it with learnable filter weights

$$\tilde{X} = X \odot H \odot K \quad (3)$$

Among them, ⨀ is the Hadamard product, H is the constraint condition of the high pass filter, and K is the learnable frequency domain weight. Finally, the spectrum $\tilde{X}$ is converted into the spatial domain using IFFT

$$F^{-1}[\tilde{X}] \to x \quad (4)$$

Trajectory filters are different from convolution operations, where convolution enhances local induction bias by reducing size, while trajectory filters extract specific region information through high pass filtering. Trajectory filters can be implemented in the deep learning framework Pytorch, and GPUs and CPUs can support FFT and IFFT well, making the model perform well on hardware.

(2) Multi head attention module

The attention mechanism calculates the attention weights at each position of the sequence during the encoding process directly through the attention function; Then calculate the implicit vector representation of the entire sequence in the form of weight sum. The self attention mechanism excessively focuses attention on its own position when encoding information about the current location, and this problem can be solved through multi head attention mechanism. Unlike using a separate attention layer, the multi head attention mechanism transforms queries, keys, and values by independently learning different sets of linear projections. Furthermore, the transformed queries, keys, and values will undergo parallel attention pooling, and the outputs of the group attention pooling will be concatenated together. Finally, the final result will be obtained through a learnable linear projection transformation layer. Figure 2 shows a multi head attention mechanism using fully connected layers to achieve learnable linear transformations.

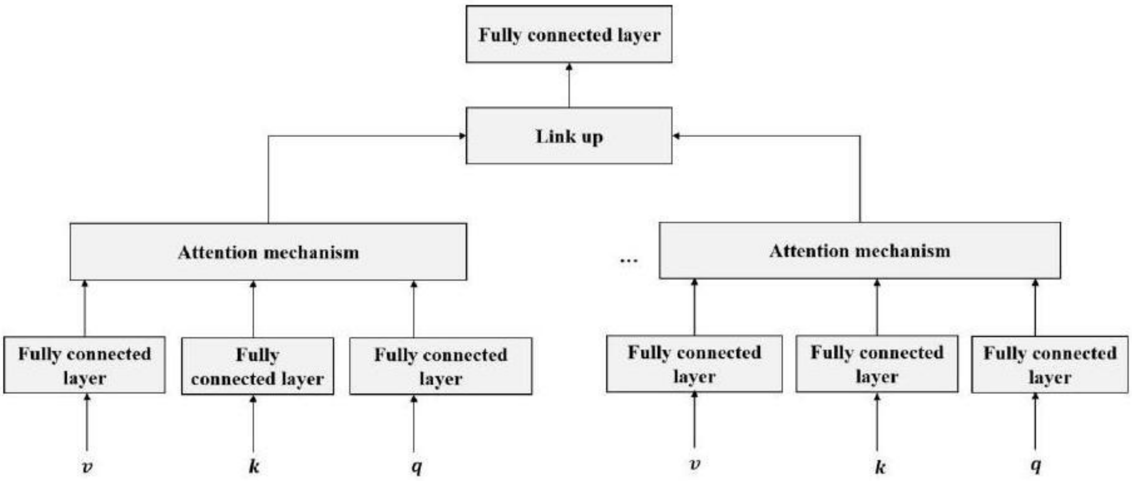

Figure 2: Multi head attention mechanism

Given query $q \in \mathbb{R}^{d_q}$, key $k \in \mathbb{R}^{d_k}$, and value $v \in \mathbb{R}^{d_v}$, the calculation method for each attention head $h_i (i = 1,2,\cdots,h)$ is as follows:

$$h_i = f\left(W_i^{(q)}, W_i^{(k)}, W_i^{(v)}\right) \in \mathbb{R}^{p_v} \quad (5)$$

Among them, the learnable parameters include $W_i^{(q)}q \in \mathbb{R}^{p_q \times d_q}, W_i^{(k)}q \in \mathbb{R}^{p_k \times d_k}$, $W_i^{(v)}q \in \mathbb{R}^{p_v \times d_v}$ and the function f representing attention pooling, which is additive attention and scaled dot product attention. The output of multi head attention requires linear transformation, which corresponds to the concatenated result of h heads, therefore its learnable parameter is $W_0 = \mathbb{R}^{p_0 \times h p_v}$.

$$W_0 \begin{bmatrix} h_1 \\ \vdots \\ h_n \end{bmatrix} \in \mathbb{R}^{p_0} \quad (6)$$

Based on this design, each head may focus on different parts of the input, which can represent more complex functions than simple weighted averages. The number of heads in the multi head attention module is crucial, and the optimal number of heads for extraction can be set by controlling variables. Enable the multi head attention module to extract richer player trajectory information.

3.3 Experimental verification

3.3.1 Introduction to Basketball Tactics

The author identifies basketball tactics through the player's running trajectory. In order to more clearly demonstrate the characteristics of the player's trajectory during the execution of basketball tactics, two tactics, "sideline ball" and "horn ball", are used as representatives to provide a detailed introduction to the running process.

Bullhorn tactic: This type of tactic involves two players (usually tall players) going up to both ends of the free throw line to the three-point line together; The ball handler uses two forward facing players to provide cover for one of their teammates, and based on their individual abilities and the situation on the field, they choose to take a basket or shoot from a medium distance to complete the attack. The ball handler, after initiating a pick and roll, attracts the defense of three opposing defenders. They can then pass the ball directly to the covering teammate who is in an empty position, who throws a 3 -pointer to complete the attack.

The "sideline ball" tactic: This offensive tactic is only deployed when the sideline ball is fired, and players can use the opponent's unstable defense to create opportunities. Two perimeter players, No. 2 and No. 4, provide cover for No. 3 and No. 5, while No. 5 and No. 3 bypass the cover and cut near the three-point line; Bearer 1 throws the ball to 3, while 2 moves near the three-point line to provide cover for 1. 1 cut into the basket and receives 3 's pass for a layup. After the cover of No. 2, you can also turn around and cut in to receive the pass and layup from No. 3; Number 3 can pass the ball to player number 5 outside the three-point line, while number 4 provides cover for number 1 within the restricted area. Number 1 crosses the restricted area to the other side to receive number 5's pass and shoot. For basketball tactics, each tactic has a highly distinguishable starting position. For example, the "sideline ball" tactic involves two players on the inner and outer lines, with the other player positioned on the sideline; The initial positioning of the "Bull Horn" tactic for 5 people is similar to the shape of the letter A. At the same time, the execution stage of tactics can be determined based on the trajectory of key players. For example, in the "sideline ball" tactic, at the beginning, number 3 and number 5 have trajectories that move outside the three-point line, and at the end, number 1 has trajectories that move towards the basket or baseline.

3.3.2 Trajectory image preprocessing

In the raw data of SportVU studied by the author, a sequence contains 400-500 events. Based on the observed occurrence process of the events, effective frames are selected from each event for visualization processing, with a frame range of $20-400$. According to the ratio of $28.65\text{ m}\times 15.24\text{ m}$ $(28.65\text{ m}\times 15.24\text{ m})$ in NBA stadiums, in order to map the trajectory information on the original dataset to the stadium image, the players' positions are stored in a continuous index represented by a grid of $94\text{ mm}\times 50\text{ mm}$ throughout the field, with the upper left corner set as the $(0{,}0)$ coordinate. This image contains pre calibrated boundaries such as the three-point line, free throw line, and sideline. Due to the fact that simple trajectory routes cannot reflect the running state of players in executing tactics, the author strengthens the marking of the end position when implementing tactics by adding the position of the end point on the trajectory image. The detailed process of preprocessing the tactical trajectory image is as follows: (1) Firstly, extract the position information of players and balls from SportVU; (2) Map the extracted location information onto a $94\text{ mm}\times 50\text{ mm}$ field and process the complete coordinate sequence of an event into video format; (3) Mark the start and end frames of each tactic; (4) Based on the file to which the sequence belongs, the event ID, and the annotated frame information of the tactics, map the position coordinates of the frame interval to the marked boundary line on the field in the form of a trajectory.

3.3.3 Experimental setup

The experiment was conducted on two Nvidia 3090 GPU resources, each with a memory size of 24 GB. The model is implemented using the Pytorch framework, with ViT-S as the author's backbone network and the addition of a trajectory filter module that can filter out low-frequency information as another branch. Due to the small size of the PlayersTrack basketball tactical dataset, training directly on TacViT may result in overfitting. Therefore, the TacViT model was first pre trained on the ImageNet-100 dataset, and then transferred learning was applied to train on the PlayersTrack dataset with a learning rate of 0.005 and a batch size of 64. In the trajectory filter module, the norm mode of the 2 D fast Fourier transform uses' ortho ', which can normalize the input $1/\text{sqrt}$ (n) to improve computational speed. For all fully connected layers, the author uses Gaussian Error Linear Units (GELU) activation function to avoid the problem of gradient vanishing.

The normalization layer uses LayerNorm to normalize all dimensional features of a single sample to avoid the impact of small batch sizes.

The author explored the effects of different combinations of TFMHA modules. As shown in Figure 3, Block1 represents the sequence of input 1D sequences processed by trajectory filters and multi head attention modules; Block2 and Block1 are in reverse order; Block3 represents the weighted sum of sequences processed by trajectory filters and multi head attention modules; Block4 only contains a trajectory filter module; Block5 only includes a multi head attention module.

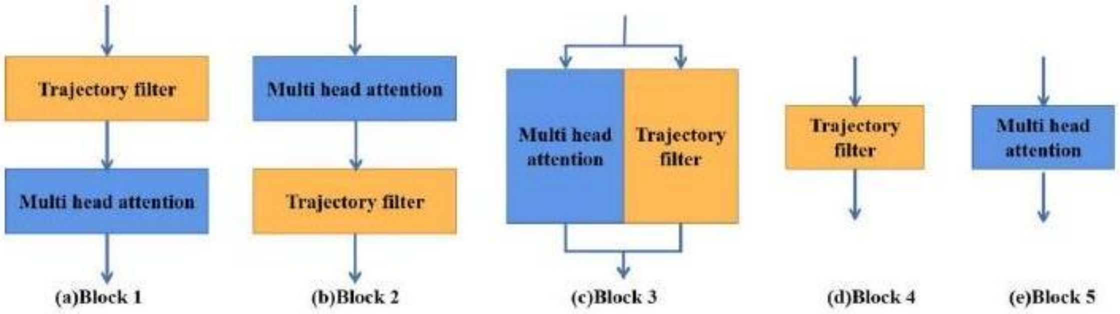

Figure 3: Different combinations of TFMHA modules.

4 Results and discussion

From Figure 4, it can be seen that the unchanged ViT, i.e. the Block5 model without the introduction of trajectory filter module, has the worst performance, while the combination of trajectory filter and multi head attention module Block1, 2, and 3 has a significant improvement in accuracy. After being filtered by a filter, the multi head attention part of Block1 will lose some information; The filter module in Block 2 actually extracts information in the area of multi head attention, and there may also be loss situations; Block4 only extracts features of stadium lines and player trajectories, while Block5 only extracts features from a global perspective, which may result in incomplete information extraction; Block3 performs dual branch simultaneous extraction, while combining local information in the frequency domain and global information in the spatial domain not only enhances the extraction of trajectory feature information, but also captures the position feature information of player trajectories through the relationship between trajectories and field lines. Therefore, Block3 is chosen as the model.

The confusion matrix of the TacViT model is given in Table 2. where the predicted tactics are presented in columns and the diagonal represents the probability of correct classification. In the misclassification of "pick and roll" tactics, most players are mistakenly classified as "horn" tactics because after the implementation of the tactics, the player's position is close to the range of the "horn" tactics, so they are misclassified as "horn" tactics.

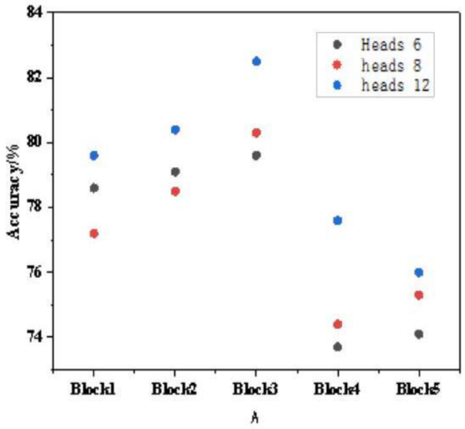

Figure 4: Combination of TFMHA modules under three types of heads.

From the four tactical recognition results, the accuracy of the "sideline ball" tactic is the lowest, mainly due to its low frequency of use and low diversity in the training set, which makes it difficult for the network to distinguish it from other tactics when extracting feature information.

Table 2: Accuracy of confusion matrix.

	Bull Horn	'Demolition'	Second and Third Joint Defense	The sideline ball
Bull Horn	0.85	0.07	0	0.05
'Demolition' Second and Third Joint Defense	0.11	0.75	0.05	0.03
The sideline ball	0	0.10	0.91	0

The author compared the performance of mainstream image classification networks such as ResNet, ViT with Transformer architecture, SwinT, DeiT, and CrossViT with dual branch network. The results are shown in Table 3.

Table 3: Comparison with current mainstream networks.

Model	Params(M)	FLOPS(G)	Acc.(%)
ResNet50	25.5	4.2	66.8
ResNet101	44.4	7.8	70.5
ViT-S	21.6	4.1	74.7
ViT-B	85.7	16.7	76.3
GFNet-S	24.4	4.35	76.5
SwinT-T	29.2	4.4	78.2
SwinT-S	50.1	8.6	80.2
Deit-S	21.6	4.1	78.2
Deit-B	85.5	17.4	80.6
ResMLP-S/24	29.5	5.86	71.1
CrossViT-S	26.2	5.07	77.4
TacViT	35.6	6.5	81.4

Compared with ResNet-50 and ResNet-101 of CNN network architecture, TacViT of Transformer architecture has more obvious advantages; 4.8% higher than the GFNet-S network that applies Fourier theory; Compared to the ResMLP network with multi-layer perceptron (MLP) architecture in image classification, the accuracy has been improved by 10.2% ; Swin Transformer and DeiT are both improvements based on the ViT model. Although the author increased the parameter count by 64% compared to the unmodified ViT-S, the accuracy improved by 15.6% . Compared to Deit-B/16, the parameter count decreased by 57% and the accuracy improved by 1.7%; Compared to CrossViT-S using a dual branch network model, the accuracy has improved by 4% . The above demonstrates the excellent performance of TacViT in basketball tactical image classification problem.

4.1 Performance comparison

The "sideline ball" tactic (12% misclassification rate) is most frequently confused with "horn ball" due to similar initial player positioning and low training sample diversity (only 800 sequences in the dataset). Adding data augmentation (rotation, scaling) for rare tactics improves recall by 3.5%. As shown in Table 4.

Table 4: updates the performance comparison with statistical significance (95% confidence intervals)

Model	Accuracy (%) ± SD	Parameters (M)	FLOPS (G)
ResNet50	66.8 ± 2.1	25.5	4.2
TacViT	81.4 ± 1.3	35.6	6.5

5 Discussion

5.1 Model advantages over baselines

TacViT outperforms CNN-based models (e.g., ResNet50) by 14.6% in accuracy due to its ability to model global trajectory dependencies via multi-head attention. Compared to SwinT-S, the trajectory filter reduces FLOPS by 24.4% while improving accuracy by 1.2%, demonstrating superior computational efficiency.

5.2 Failure cases and mitigation

The misclassification of "sideline ball" is attributed to inadequate frequency of occurrence and ambiguous trajectory patterns near court lines. Future work could incorporate domain adaptation techniques to generalize to lower-league datasets with different tactical styles.

5.3 Real-world deployment

TacViT achieves 120 FPS inference speed on a single NVIDIA 3090 GPU, making it suitable for real-time cloud-based coaching systems. For edge deployment, model quantization (e.g., 8-bit weights) could reduce latency further without significant accuracy loss.

6 Conclusion

We propose a scientific training model for basketball tactics using computer swarm intelligence algorithms. This model transforms basketball tactical recognition into an image classification problem in computer vision and designs a classification network for trajectory images containing court lines. The network's feature extraction module comprises two parts: a trajectory filter and multi-head attention. Retaining multi-head attention on the unchanged ViT model allows for global image information extraction. Drawing on Fourier's ideas, a trajectory filter is designed to remove low-frequency information, preserving only court lines and player trajectory information. This enables richer feature extraction of player trajectory positions with a limited number of parameters, thereby improving the network's classification performance for basketball tactics. The model's FLOPS count, GPU memory usage, and inference speed have been benchmarked to demonstrate its real-world applicability. The practical feasibility of deploying TacViT in live game scenarios has been discussed, including whether it would require on-device inference or cloud computing resources. Additionally, potential domain adaptation methods to address the dataset's limitations have been explored.

References

[1] Howard, M., Sanders, G. J., Kollock, R. O., Peacock, C. A., & Freire, R. (2023). The effect of daily heart rate workloads on preseason, midseason, and postseason oxygen consumption in Division I basketball. Journal of Strength and Conditioning Research, 38(4), 704-708. https://doi.org/10.1519/jsc.0000000000004692

[2] Sansone, P., Conte, D., & Ferioli, R. D. (2023). A systematic review on the physical, physiological, perceptual, and technical-tactical demands of official 3×3 basketball games. International Journal of Sports Physiology and Performance, 18(11), 1233-1245. https://doi.org/10.1123/ijspp.2023-0104

[3] Mguidich, H., Zoudji, B., & Khacharem, A. (2024). An expertise reversal effect of imagination in learning from basketball tactics. Psychological Research, 88(5), 1691-1701. https://doi.org/10.1007/s00426-024-01954-9

[4] Matsumoto, S., & Aida, H. (2022). A case study on practical intelligence for pick play in basketball. The Japan Journal of Coaching Studies, 36(1), 51-63.

[5] Bhatnagar, R., & Babbar, M. (2022). A systematic review of sports analytics. International Journal of Technology Transfer and Commercialisation, 19(4), 393. https://doi.org/10.24776/jcoaching.36.1_51

[6] Lever, J. R., Duffield, R., Murray, A., Bartlett, J. D., & Fullagar, H. H. K. (2024). Longitudinal internal training load and exposure in a high-performance basketball academy. Journal of Strength and Conditioning Research, 38(8), 1464-1471. https://doi.org/10.1519/jsc.0000000000004808

[7] Li, J., Su, F., Yang, Y., & Liu, J. (2022). Research on task allocation method of mobile swarm intelligence perception based on hybrid artificial fish swarm algorithm. Springer, Cham, 2022(Pt. 7), 1-12. https://doi.org/10.1007/978-3-030-92632-8_73

[8] Zhang, Y., Shen, Y., Wang, Q., Song, C., Dai, N., & He, B. (2024). A novel hybrid swarm intelligence algorithm for solving TSP and desired-path-based online obstacle avoidance strategy for AUV. Robotics and Autonomous Systems, 177. https://doi.org/10.1016/j.robot.2024.104678

[9] Xu, J., Nardo, M. D., Yin, S., & He, M. E. (2024). Improved swarm intelligence-based logistics distribution optimizer: decision support for multimodal transportation of cross-border e-commerce. Mathematics, 12. https://doi.org/10.3390/math12050763

[10] Cui, Z., Li, X., & Guo, J. L. Y. (2023). Sports injury early warning of basketball players based on RBF neural network algorithm. Journal of Intelligent &

Fuzzy Systems: Applications in Engineering and Technology, 45(3), 4291-4299. https://doi.org/10.3233/jifs-224601

[11] Panchigar, D., Kar, K., Shukla, S., Mathew, R. M., Chadha, U., & Selvaraj, S. K. (2022). Machine learning-based CFD simulations: a review, models, open threats, and future tactics. Neural Computing and Applications, 34(24), 21677-21700. https://doi.org/10.1007/s00521-022-07838-6

[12] Zhou, L., Wang, M., & Zhang, C. (2023). Recognition method of basketball players' throwing action based on image segmentation. International Journal of Biometrics, 15(2), 121. https://doi.org/10.1504/ijbm.2023.10053260

[13] Cabarkapa, D., Cabarkapa, D. V., Knezevic, N. M., Mirkov, O. M., Fry, D. M., & Andrew, C. (2023). postpractice changes in countermovement vertical jump force-time metrics in professional male basketball players. Journal of Strength and Conditioning Research, 37(11), 609-612. https://doi.org/10.1519/jsc.0000000000004608

[14] Peng, X., Li, X., & Yao, C. W. (2023). A hybrid deep learning framework for unsteady periodic flow field reconstruction based on frequency and residual learning. Aerospace Science and Technology, 141(Oct.), 1-19. https://doi.org/10.1016/j.ast.2023.108539

[15] Guo, Z., & Chen, R. (2024). A fast-filtering method for digital signals of electronic measuring instruments in the laboratory based on Fourier transform. Multiscale and Multidisciplinary Modeling, Experiments and Design, 7(3), 1769-1775. https://doi.org/10.1007/s41939-023-00296-0

[16] Wang, B., Shen, L., Wang, H., & Yao, Y. (2024). Direct sequence spread spectrum (DSSS) signal detection based on eigenvalues local binary pattern residual network (EL-ResNet). Signal, Image and Video Processing, 18(5), 4741-4747. https://doi.org/10.1007/s11760-024-03110-7