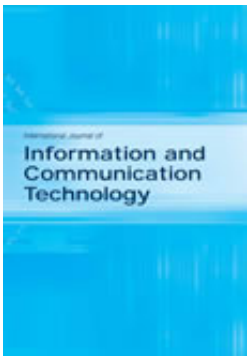
INDERSCIENCE PUBLISHERS
Linking academia, business and industry through research



International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Basketball player action optimisation based on deep reinforcement learning: multimodal biomechanical modelling

Tao Huang

DOI: 10.1504/IJICT.2025.10075267

Article History:

Received:	17 July 2025
Last revised:	05 September 2025
Accepted:	08 September 2025
Published online:	12 January 2026

Basketball player action optimisation based on deep reinforcement learning: multimodal biomechanical modelling

Tao Huang

School of Physical Education,
Huanggang Normal University,
Huanggang, 438000, China
Email: huangtao2004@hotmail.com

Abstract: Basketball player action optimisation is key to improving competitive performance. To address the issue of insufficient mining of biomechanical features and poor action optimisation effects in current research, this paper first conducts biomechanical data analysis of basketball players and constructs a motion state equation for capturing and analysing key data. Then, it integrates image and biomechanical motion data to provide comprehensive multimodal perception information. A multimodal feature extraction and fusion module based on self-attention mechanism is designed. Secondly, the pose action decision-making task of the athlete is modelled as a deep reinforcement learning (DRL) problem. Finally, a hybrid reward function is designed to achieve efficient training of the model and action strategy optimisation. Experimental outcome indicates that the high model improves the action optimisation success rate by at least 5% compared to the baseline model, demonstrating good action optimisation effects.

Keywords: basketball action optimisation; deep reinforcement learning; biomechanical modelling; multimodal feature fusion; attention mechanism.

Reference to this paper should be made as follows: Huang, T. (2025) 'Basketball player action optimisation based on deep reinforcement learning: multimodal biomechanical modelling', *Int. J. Information and Communication Technology*, Vol. 26, No. 51, pp.1–17.

Biographical notes: Tao Huang obtained his Master's degree from Beijing Sport University in 2010. He is currently a Lecturer at the School of Physical Education, Huanggang Normal University. His research interests include digital and intelligent sports, basketball training and physical conditioning.

1 Introduction

In the intense confrontation and precise competition of basketball, every detail of an athlete's movement may turn into a crucial element in deciding the game's result (Ren and Wang, 2021). Traditional methods of improving movements based on experience accumulation have become tough to meet the requirements of high-level training. Biomechanics provides a scientific framework for analysing the internal mechanisms of basketball movements, and by collecting kinematic data of athletes, researchers can quantitatively evaluate the biomechanical characteristics of movements and identify

potential unreasonable force patterns (Lam et al., 2022). However, single-modal data often only reflects the local characteristics of movements, but fails to explain the causal relationship between the force value and movement stability (Wang et al., 2024a). The fusion analysis of multimodal biomechanical data can build a more comprehensive movement evaluation model, but how to mine effective information from massive heterogeneous data for movement optimisation remains a current research challenge (Cheng and Cheng, 2024). DRL learns optimal decision-making strategies through continuous interaction between the agent and the environment, and its strong nonlinear fitting ability and sequence decision-making advantages are exactly suitable for the needs of basketball movement optimisation (Zhang and Tao, 2023). Therefore, how to combine biomechanical modelling data and reinforcement learning (RL) theory to develop efficient methods for optimising basketball athletes' movements has important application value.

Traditional methods for optimising basketball athletes' movements are based on machine learning methods. Typical statistical learning methods include SVM (Pradhan, 2012), AdaBoost (Wang and Sun, 2021), etc. Liu and Wang (2023) used the relative position and velocity of athletes as inputs to train a support vector machine (SVM) for optimising athlete movement decisions. Zhu (2022) adopted a sports movement optimisation method based on particle swarm optimisation SVM, using Bayesian parameter optimisation to better determine parameters. In addition to the SVM method, Xie and Wu (2024) applied AdaBoost to athlete movement optimisation, using a series of athlete movements as inputs and selecting the optimal strategy as the output, but the practicality was weak. Li (2025) defined an appropriate objective function based on the biomechanical principles of athlete movements, transforming the generation of movements into a specific offline spatiotemporal optimisation problem to calculate joint torques, and obtained the optimised movement results through decision trees. Wang (2023) adopted a low-dimensional physical model method, simplifying complex athletes into a low-dimensional model to simulate athlete movements, and used random forests as classifiers to predict the optimal movement output.

Deep learning approaches share similarities with conventional machine learning techniques. The key distinction is that deep learning employs neural networks to automatically capture data characteristics, thereby improving the performance of optimised design. By harnessing deep learning's capabilities in image processing, a sports movement enhancement system was created. In most cases, sensor-recorded movement data is processed by a trained deep learning model to yield optimised outputs. Javadpour et al. (2022) employed monocular front-view motion sequences as input data, utilising a convolutional neural network for end-to-end policy learning to derive optimised movement strategies. Yoon et al. (2019) developed an attention branch network (ABN) architecture for sports motion optimisation decision-making. First, the original visual image was input, and the network architecture incorporated an attention mechanism that produced a characteristic attention map at its intermediate level. Finally, the attention map of the original image was combined with the convolutional features of the self-expected action to generate the optimised sports movement. Yan (2024) combined motion matching with deep learning, and proposed learned motion matching, which reduces the storage space required by motion matching through a learning approach, while ensuring that the output movements can be mapped to the motion library. Xiao (2024) extracted features from motion images and motion videos using a multi-scale

CNN, and used an attention mechanism to integrate multi-modal characteristics, and output the results of sports movement prediction.

RL enables agents to progressively learn optimal policies through trial-and-error interactions with the environment, maximising cumulative rewards. This characteristic holds significant potential for optimising athletes' movements. The agent's kinematics, athletic scenarios, and biomechanical constraints are encoded into the environmental state space. Deep neural networks are employed to construct policy functions, taking the current state as input and generating action parameters as output. Through end-to-end learning, states are directly mapped to optimal actions, circumventing the limitations of manually designed features in traditional approaches. The current model undergoes cyclical refinement through iterative updates incorporating both historical data and newly acquired environmental exploration data, achieving the optimisation of athletes' movements. Chen et al. (2021) used an adversarial learning framework to train a policy network as a generator to generate motion sequences, and used an adversarial discriminator to the kinematic plausibility of synthesised motions. Liu and Hodgins (2018) utilised the reward mechanism and strong decision-making ability of DRL to quickly calculate the joint torque of athletes, effectively overcoming the problems of poor generality and real-time performance in traditional methods. Wang et al. (2024b) used the proximal policy optimisation algorithm combined with motion trajectory optimisation to train a virtual character to learn basketball dribbling actions, making it closer to the real actions of athletes. Arumugam (2025) designed a basketball player's action optimisation method based on an improved deep Q network, achieving a high level of optimisation accuracy. Chang et al. (2025) used transformer to extract and fuse the action features of different athletes, simulating the athletes' movement environment, and using DRL to make decisions on the optimal athlete action strategy, which has good robustness.

From the above in-depth analysis of action optimisation methods, it can be seen that traditional action analysis methods rely on expert experience and limited biomechanical indicators, making it difficult to achieve personalised and real-time action optimisation. Existing research has problems such as insufficient mining of biomechanical features and unsatisfactory action optimisation effects. To address these issues, this article puts forward a basketball player action optimisation model based on DRL and multimodal biomechanical modelling. First, biomechanical data analysis and interpolation reconstruction of basketball players are performed to build a motion state equation for capturing and analysing key data for basketball movement modelling. Then, a basketball player action optimisation model is designed. The model contains a policy network and a Q-value network, whose core is the multimodal feature extraction and fusion module. The policy network is composed of a multimodal characteristic extraction and integration module, a fully linked network, and a fully connected layer, which is used to generate the action probability distribution in the action space. The Q-value network adopts a completely consistent network structure, composed of a multimodal feature extraction and fusion module and a fully linked network. The parameters of these target Q-value networks are dynamically updated from the corresponding Q-value networks through a soft update mechanism, which is adopted to collaboratively complete the model's optimisation and decision-making tasks. Through the collaborative effect of the policy network and the Q-value network, the training efficiency and decision-making performance are significantly improved, ensuring the efficiency and stability of the model in the basketball movement action optimisation task. Experimental outcome indicates that

the action optimisation success rate and average reward function of the proposed model are better than those of the baseline model, significantly improving the efficiency of action optimisation.

2 Relevant technologies

2.1 Deep reinforcement learning algorithm

Traditional RL is prone to falling into the myopia trap when handling long-term sequential decision making. For instance, conventional policy gradient methods update policies solely based on the 'current step reward', failing to account for the cumulative benefits of future steps and resulting in locally optimal policies. DRL overcomes the limitations of traditional RL in handling high-dimensional state spaces and complex decision problems by integrating deep neural networks with RL. It achieves a direct mapping from raw inputs to optimal actions through a single neural network. This end-to-end optimisation reduces information loss and enhances policy robustness. The fundamental architecture of DRL comprises an intelligent agent interacting with its environment. The agent engages in dynamic exchanges with its environment. Environmental states evolve as a direct consequence of the agent's interventions. Environmental states may evolve through intrinsic dynamics without agent intervention. The agent obtains a reward signal from the environment, which conveys to the agent an assessment of whether the current environmental state is favourable or unfavourable (Wang et al., 2022). During the entire interaction process, the agent's goal is to maximise the cumulative reward obtained. Next, several concepts in RL will be introduced.

1 State and observation: a state is generally denoted by s, and the state space is generally denoted by S, which is used to describe the state information of the environment. Observation $_o$ pertains to the environmental state information that is perceptible to the agent. Usually, a portion of the environmental state information remains concealed.

2 Policy: the policy of the agent is divided into deterministic policy and stochastic policy. A deterministic policy is generally denoted by μ, as shown below:

$$a_t = \mu(s_t) \tag{1}$$

3 Trajectory: a trajectory τ is a sequence of states and actions in the environment, $\tau = (s_0, a_0, s_1, a_1, \ldots)$. The initial state of the environment ρ_0 is randomly selected from the initial state distribution S_0.

4 Reward and return: the reward function R holds significant importance in the domain of reinforcement learning. Its value is contingent upon the present state of the environment, the action that has just been executed, and the subsequent state of the environment, as shown below:

$$r_t = R(s_t, a_t, s_{t+1}) \tag{2}$$

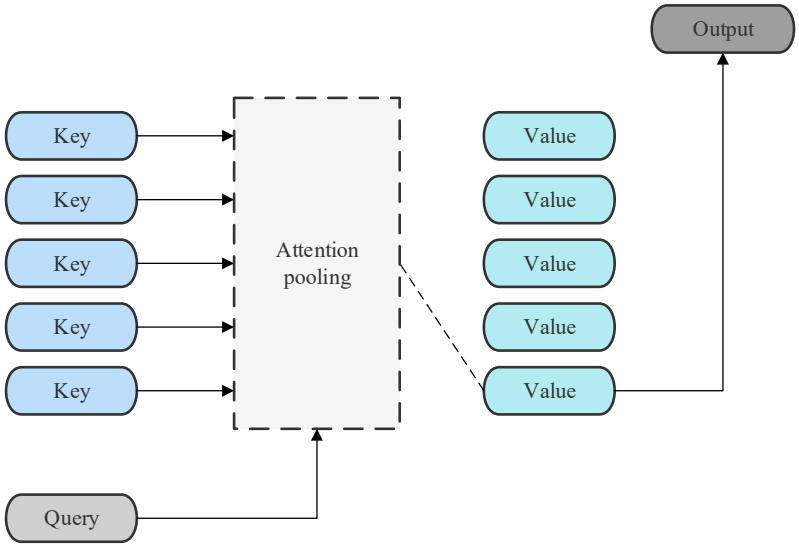
DRL algorithm can be classified into model-based approaches and model-free methods according to whether the agent can access the environment model. Model-based methods

allow the agent to think and plan in advance, and the agent can distil the results of the pre-planned strategy into a learning strategy.

2.2 Attention mechanism

The attention mechanism is a neural network component that dynamically weights input features, mimicking human cognitive attention patterns. We do not pay attention to all information at the same time, but focus on some key information according to the needs of the task. The integration of attention mechanisms enables neural networks to learn dynamic weight distributions across input features, enhancing their capacity for complex pattern recognition (Brauwers and Frasincar, 2021). The fundamental concept underlying the attention mechanism is the dynamic allocation of resources, enabling the model to concentrate on the most pertinent segments of the input. The structure of the attention mechanism is implied in Figure 1.

Figure 1 The structure of the attention mechanism (see online version for colours)



In its most basic form, the attention mechanism can be viewed as a mapping function of queries, keys, and values. For a given query, the similarity or matching degree between the query and a set of keys is calculated, and then these similarity scores are used to weight the corresponding values, and finally the sum of the weighted values is output. This process can be summarised as below:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{3}$$

where Q is the query matrix, K is the key matrix, and V is the value matrix. d_k is the dimension of the key vector, used to scale the dot product result, preventing large dot product values from causing the softmax function to be in the saturation region, thus affecting the propagation of gradients. The softmax function serves the purpose of

transforming the outcome of the dot-product operation into a probability distribution, which delineates the weight assigned to each individual value.

Attention mechanisms enhance models' ability to process long sequences and high-dimensional data in deep learning tasks by dynamically allocating weights to focus on key information. Based on the continuity and determinism of weight distribution, attention mechanisms can be categorised into soft attention and hard attention. Soft attention adjusts feature weights across channels to improve image classification performance. Hard attention reduces noise by ignoring irrelevant information but may result in the loss of global context.

3 Analysis of biomechanical data for basketball athletes

3.1 Collection of biomechanical data for basketball athletes

Common methods for collecting biomechanical data include Acclaim skeleton file/Acclaim motion capture data (ASF/AMC) and Biovision Hierarchy (Scibek and Carcia, 2013). Given the nonlinearity and stochastic nature of basketball biomechanical data, this study employs the ASF/AMC skeletal animation format for standardised motion parameter representation. The biosensor's kinematic state output at discrete timestep k is denoted as $\omega_k = [\omega_k \quad \omega_y \quad \omega_z]^T$, the output of the posture information of the basketball athlete calculated by the accelerometer and magnetometer is $b_k^a = [a_x \quad a_y \quad a_z]^T$, and the output of the motion attitude angle is $b_k^m = [m_x \quad m_y \quad m_z]^T$. Assuming the reference coordinate system is the geodetic coordinate system, under the influence of the gravity vector and the geomagnetic vector, in the high-dimensional space of the basketball athlete's biomechanical movement, the mechanical measurement data of the basketball athlete's movement are obtained through accurate posture estimation as $r_a = [0 \quad 0 \quad -g]^T$ and $[h\cos\alpha \quad 0 \quad -h\sin\alpha]^T$, where g is the absolute value of gravitational acceleration, h and α are the tracking errors of the sensor data with the geomagnetic dip angle.

When the basketball athlete is performing movements such as walking, jumping, and handstands, a closed set of spatial motion equations for the basketball athlete's biomechanical data is obtained as follows:

$$m\frac{\mathrm{d}V}{\mathrm{d}t} = P\cos\alpha - X - mg\sin\theta \tag{4}$$

$$mV\frac{\mathrm{d}\theta}{\mathrm{d}t} = P\sin\alpha + Y - mg\cos\theta \tag{5}$$

$$J_z\frac{\mathrm{d}\omega_z}{\mathrm{d}t} + \left(J_y - J_x\right)\omega_y\omega_x + J_{xy}\left(\omega_y^2 - \omega_x^2\right) = M_z \tag{6}$$

$$\frac{\mathrm{d}x}{\mathrm{d}t} = V\cos\theta \tag{7}$$

$$\frac{\mathrm{d}y}{\mathrm{d}t} = V\sin\theta \tag{8}$$

$$\frac{\mathrm{d}\vartheta}{\mathrm{d}t}=\omega_z \tag{9}$$

$$\alpha=\vartheta-\theta \tag{10}$$

$$\delta_z=f(e_1) \tag{11}$$

where θ is the jumping tilt angle of the movement, ϑ is the pitch forward tilt angle of the athlete when running, α quantifies the superior-inferior body displacement in the sagittal plane during the sideways handstand position, x, y are horizontal and vertical positions of the posture during movement, w_x, w_y are the torques on the coordinate system Ox_1, Oy_1 axes when in non-accelerated motion state, δ_z is the body angle deviation at any posture, e_1 is the control error of longitudinal motion; m is the mass of the athlete; X, Y are the air resistance, lift, and lateral force of the human body in running and jumping movements, M_z is the pitch torque, J_z is the moment of inertia of the human body during movement with coordinate system transformation; J_{xy} is the moment of inertia of the human body's motion space model with respect to the velocity coordinate system Oz_1.

3.2 Interpolation reconstruction of basketball athlete movement

According to the above motion equations and data capture results, perform motion interpolation reconstruction, and obtain the following biomechanical data observation equation for basketball movement under global search.

$$\begin{cases} q_{k+1}=\Phi_k q_k-\dfrac{\Delta t}{2}I_k\varepsilon_k \\ 0=H_{k+1}q_{k+1}-\dfrac{1}{2}I_k\delta b_{k+1}^{x} \end{cases} \tag{12}$$

where k indexes the discrete-time sampling instants in the system, q_k is the unit quaternion encoding the orientation state of the basketball motion capture system in the local carrier coordinates at sample k, Φ_k maps vectors from body coordinates to the navigation frame, and the motion data for two adjacent key frames is determined through the application of approach w_k, H_{k+1} is the observation ε_k and δb_{k+1}^{x} are the observation disturbances during the basketball motion modelling process; I_k is the disturbance coefficient matrix, and the Newtonian mechanics coefficient q_k is obtained by processing the raw motion capture data and solving the inverse kinematics equations.

The kinematic state equations for basketball motion reconstruction incorporate nonlinear attitude representations derived through rigid-body transformation methods, expressed as below:

$$\begin{cases} \begin{bmatrix}\dot{\omega}\\ \dot{q}\end{bmatrix}=\begin{bmatrix}-\dfrac{1}{\tau}\omega\\ \dfrac{1}{2}q\otimes\bar{\omega}\end{bmatrix}+\begin{bmatrix}-\dfrac{1}{\tau}\varepsilon\\ 0\end{bmatrix} \\ \begin{bmatrix}\hat{\omega}\\ \hat{q}\end{bmatrix}=\begin{bmatrix}\omega\\ q\end{bmatrix}+\begin{bmatrix}\xi\\ \eta\end{bmatrix} \end{cases} \tag{13}$$

where $\bar{\omega}=[0 \quad \omega]^T$ serves as the quaternion formulation of the output vector; ε, ξ and η are the attitude information output by the basketball motion attitude equation respectively.

To ensure the linearity of the observation equation, the captured data of basketball motion biomechanics b_{k+1}^a and b_{k+1}^m is obtained through second-order filtering, and is calculated by the Gauss-Newton iterative algorithm. Read the sensor data worn on the athlete for error analysis, and through the interpolation reconstruction method, the stochastic filtering equations for basketball trajectory features in perceptual space yield the following Kalman representation.

$$\begin{cases} q_{k+1}=\Phi_k q_k-\dfrac{\Delta t}{2}I_k\varepsilon_k \\ \begin{bmatrix} b_{k+1}^a \\ b_{k+1}^m \end{bmatrix}=\begin{bmatrix} C_r^b(q_{k+1}) & 0 \\ 0 & C_r^b(q_{k+1}) \end{bmatrix}\begin{bmatrix} r_a \\ r_m \end{bmatrix}+\begin{bmatrix} v_{k+1}^a \\ v_{k+1}^m \end{bmatrix} \end{cases} \tag{14}$$

where $C_r^b(q_{k+1})$ constitutes the state estimation pertaining to the attitude information within the reference coordinate system, v_{k+1}^a, v_{k+1}^m denote the errors in attitude alteration that occur during the basketball's motion trajectory under the influence of the gravitational acceleration vector.

4 Basketball athlete action optimisation based on deep reinforcement learning and multimodal biomechanics

4.1 Feature extraction of multimodal biomechanical information

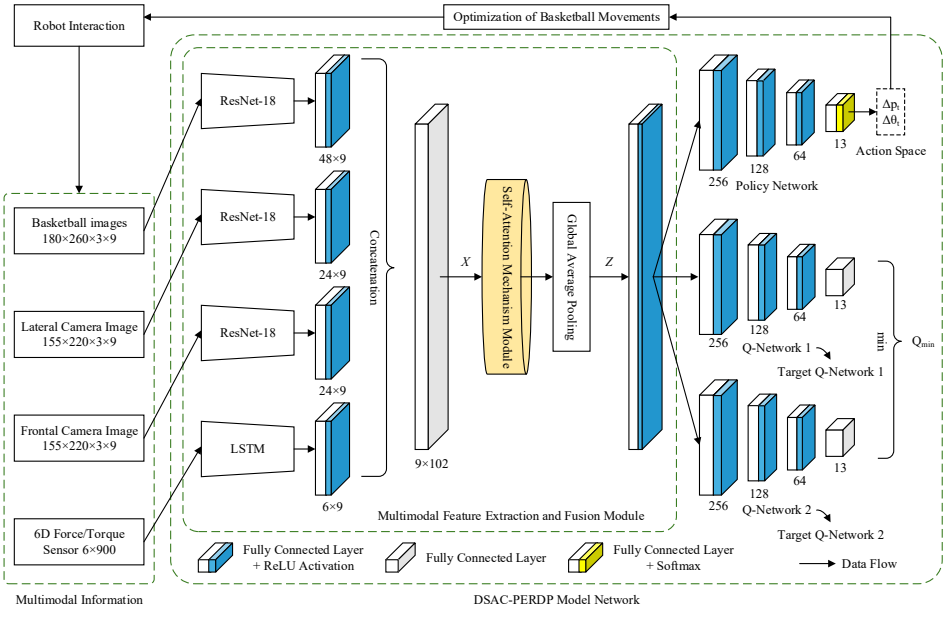
After modelling the basketball athlete's biomechanical data, to solve the problem that existing research does not fully extract the action features of the basketball athlete, leading to low efficiency in action optimisation, this paper proposes a basketball athlete action optimisation model based on DRL and multimodal biomechanics. The model aims to efficiently integrate multimodal biomechanical information, optimise the execution of autonomous sensor perception tasks and intelligent decision-making capabilities. As shown in Figure 2, the model architecture includes a policy network and a Q-value network, whose core is the multimodal feature extraction and fusion module. The neural network sizes contained in each network structure are detailed in Figure 2. The policy network consists of a multimodal feature extraction and fusion module, a fully linked network, and a softmax layer, used to generate the action probability distribution in the action space. The Q-value network adopts a completely consistent network structure, consisting of a multimodal feature extraction and fusion module and a fully linked network. To simplify the diagram, the target Q-value network is not shown in the figure. The parameters of these target Q-value networks are dynamically updated from the corresponding Q-value networks through a soft update mechanism, used to collaboratively complete the model's optimisation and decision-making tasks.

The multimodal biomechanical information feature extraction and fusion module extracts spatial and temporal features from multimodal information through the ResNet-18 (Thongpance et al., 2023) module and the long short-term memory (LSTM) (Li et al.,

2024) module, and completes feature fusion using the self-attention mechanism and global average pooling with a fully connected layer, providing a unified feature representation for each network. The following details the working principle of this module from two aspects: feature extraction and feature fusion.

The system input includes camera images, camera images, and biomechanical sensor data. Image data processing uses multimodal data to construct an image time series, and the ResNet-18 module is used for feature extraction. The residual connections in ResNet-18 mitigate gradient vanishing through identity mapping pathways, enabling stable backpropagation in deep architectures, enhancing the feature learning ability. For different modal data, this paper customises the ResNet-18: the image input is 180 × 260 × 3 × 9, and the last layer is replaced with a fully connected layer, with the output being F_{us}.

Figure 2 Basketball player movement optimisation framework (see online version for colours)



The data acquisition frequency of the biomechanical sensor is 1,000 Hz, and the LSTM network is used to process the six-dimensional tactile data. The LSTM network is widely used in sequence data processing due to its advantages in time series modelling and handling long-term dependencies. The recent sensor sequence 6 × 900 is used as input, and the last layer of the LSTM network is connected to a fully linked level, with the output being $F_{tactile}$, ensuring the full capture of the temporal features of the tactile data.

4.2 Multimodal feature fusion based on self-attention mechanism

The self-attention mechanism (SAM) can capture the mechanism of mutual relationships between elements in the input sequence (Zhang et al., 2025). The features of each modality are concatenated by time steps to form a unified input feature matrix $X = [F_{US}; F_{cam1}; F_{cam2}; F_{tactile}]^T$.

In this paper, given an input sequence X, SAM calculates the weighted representation of the input features by introducing three matrices: query Q, key K and value V. The input feature matrix is transformed linearly to generate the query, key, and value matrices.

$$Q = XW_Q \, , K = XW_K \, , V = XW_V \tag{15}$$

where W_Q, W_K, W_V are studied linear transformation weight matrices, d_Q, d_K and d_V stand for the feature dimensions of the query, key, and value, respectively. This paper sets $d_Q = d_K = d_V = d$ to simplify the calculation.

The correlation between the query and the key is calculated by the dot product, and the outcome is normalised by the Softmax function to generate the attention weights. The specific equation is as follows.

$$A(Q, K, V) = \text{softmax}\left(\frac{QK^{\mathrm{T}}}{\sqrt{d}}\right)V \tag{16}$$

where QK^T is the dot product similarity matrix between the query and the key, d is the scaling factor, $A(Q, K, V)$ is the fused feature representation, with a dimension of $R^{n\times d}$ (where n is the sequence length and d is the characteristic dimension), which contains the spatiotemporal correlation information of the multimodal features.

The generated feature representation is processed by global average pooling, as implied in the following equation:

$$Z = (1/n)\sum_{i=1}^{n} A(Q, K, V)_i \tag{17}$$

Then, it is mapped to the target feature dimension through softmax, as shown below:

$$F = \sigma\left(W_f Z + b_f\right) \tag{18}$$

The dimension of W_f is 128 × 102, b_f is a 128-dimensional bias, $\sigma(\cdot)$ s the ReLU activation operation; the final characteristic vector F is the final output of the multimodal feature extraction and fusion module.

4.3 Establishment of the reinforcement learning model

The task of optimising the basketball player's movement is modelled as a Markov decision process under the reinforcement learning framework. This framework defines the state space S, the action space A, the reward function $R(s_t, a_t)$, and the termination condition, comprehensively describing the decision-making process in the task of optimising the basketball player's movement. The agent interacts with the environment through selecting an action a_t in light of the current state $s_t \in S$, and continuously optimises the strategy according to the feedback reward signal, ultimately achieving precise optimisation of the target movement.

a State space S: S integrates multimodal information, including the front camera image $C_{1,t}$, the side camera image $C_{2,t}$, and the force sensor data F_t. The state space is defined as $s_t = \{C_{1,t}, C_{2,t}, F_t\}$.

b Action space A: basketball actions involve continuous control (such as movement and dribbling strength) and discrete decisions (such as when to shoot or pass), so a

hybrid action space design is usually adopted. Basketball actions can be decomposed into dribbling height, strength, direction, and whether to switch the ball-hand, and the action space is represented as $A = \{a_1, a_2, \ldots, a_t\}$.

The position action is denoted as increments along the x, y and z axes, with each action corresponding to a small displacement. The position of the basketball movement is represented by accumulating the displacement as follows:

$$p_{t+1} = p_t \pm \Delta p_t \tag{19}$$

where $p_t = (x_t, y_t, z_t)$ stands for the position of the probe at time step t, and $\Delta p_t = (\Delta x_t, \Delta y_t, \Delta z_t)$ stands for the displacement offset at the current time step. The displacement distance decreases linearly with the number of steps. The posture action is achieved by accumulating the rotation matrix, and the posture is updated based on rotations around the x, y, and z axes, as shown below:

$$R_{t+1} = R_t \cdot R_x\left(2 \pm \theta_{x_t}\right) \cdot R_y\left(2 + \theta_{y_t}\right) \cdot R_z\left(2 + \theta_{z_t}\right) \tag{20}$$

where R_t is the rotation matrix of the motion posture at time step t, R_x, R_y and R_z are the rotation matrices around the x, y, and z axes, respectively; θ_{x_t}, θ_{y_t} and θ_{z_t} are the corresponding rotation angles. The stop command in the action space is used to indicate the completion of the task, ensuring that the action remains stable after reaching the optimal policy.

c Design of the hybrid reward function and termination conditions: to imitate the actions of a basketball player, the hybrid reward function takes into account factors such as the current position, posture, image quality, applied force, and task completion, encouraging the agent to minimise position error, maintain a reasonable posture, achieve high-quality imaging, and safe interaction. The hybrid reward function is defined as follows:

$$R\left(s_t, a_t\right) = w_p R_{position} + w_o R_{orientation} + w_s R_{SSIM} + w_f R_{force} + w_c R_{completion} \tag{21}$$

where $R_{position}$, $R_{orientation}$, R_{SSIM}, R_{force} and $R_{completion}$ are the reward functions set considering the action position, probe posture, collected image quality, interaction force in the action space, and the completion of the action optimisation task, respectively, w_p, w_o, w_s, w_f and w_c are the weights of each reward component.

The position reward is based on the change in the Euclidean distance among the athlete's current position and the target position. Let d_t represent the Euclidean distance between the athlete's current position and the target position at time step t. The reward is defined as below:

$$R_{position} = \begin{cases} +0.2 & \text{if } d_t < d_{t-1} \\ -0.4 & \text{if } d_t \geqslant d_{t-1} \\ -1.0 & \text{if } d_t > 30\,\text{cm} \end{cases} \tag{22}$$

The posture reward needs to guide the basketball player to move within a reasonable range of motion postures. Let $\Delta\theta_{x_t}, \Delta\theta_{y_t}, \Delta\theta_{z_t}$ represent the differences between the

current posture angles around the x, y and z axes at time step t and their target posture angles. The posture reward is defined as follows.

This reward penalises postures that exceed the reasonable range, encouraging the ultrasound probe to remain within a safe and reasonable angle range.

$$R_{orientation} = \begin{cases} 0 & \text{if } |\Delta\theta_{x_t}|, |\Delta\theta_{y_t}|, |\Delta\theta_{z_t}| \leqslant 15^\circ \\ -1.0 & \text{otherwise} \end{cases} \tag{23}$$

The image quality reward aims to guide the probe to adjust its pose to maximise the ultrasound image quality. The structural similarity index (SSIM) is adopted to assess the similarity between the currently acquired ultrasound image I and the target image T, as an image quality metric, with the following calculation equation:

$$SSIM(I,T) = \frac{(2\mu_I\mu_T + c_1)(2\sigma_{IT} + c_2)}{(\mu_I^2 + \mu_T^2 + c_1)(\sigma_I^2 + \sigma_T^2 + c_2)} \tag{24}$$

where μ_I and μ_T stand for the means of the current picture and the target image, respectively; σ_I^2 and σ_T^2 are the variances; σ_{IT} is the covariance of the two images; c_1 and c_2 are the stability constants. The SSIM-based reward function is defined as follows:

$$R_{SSIM} = \begin{cases} +0.1 & \text{if } SSIM_t > SSIM_{t-1} \\ -0.1 & \text{if } SSIM_t \leqslant SSIM_{t-1} \end{cases} \tag{25}$$

During the optimisation process, behaviours that improve image quality are rewarded, while appropriate penalties are given for the opposite. The biomechanical reward function ensures that the force applied by the probe to the tissue membrane remains within a safe and effective range. Let F_z represent the force along the z-axis (vertical pressure), F_x and F_y represent the forces along the x-axis and y-axis, respectively; τ_x, τ_y and τ_z represent the torques around these axes

$$R_{forcez} = \begin{cases} 0 & \text{if } 2\ N \leqslant |F_z| \leqslant 15\,\text{N} \\ -1.0 & \text{if } F_z > 15\,\text{N} \\ -0.5 & \text{if } F_z < 2\,\text{N} \end{cases} \tag{26}$$

$$R_{force_{others}} = \begin{cases} 0, & \text{if } |F_x|, |F_y| \leqslant 15\,\text{N} \\ -0.5, & \text{otherwise} \end{cases} \tag{27}$$

The force reward function is the sum of the z-axis force reward and other force rewards, as shown below:

$$R_{force} = R_{force} - R_{force} + R_{force_{others}} \tag{28}$$

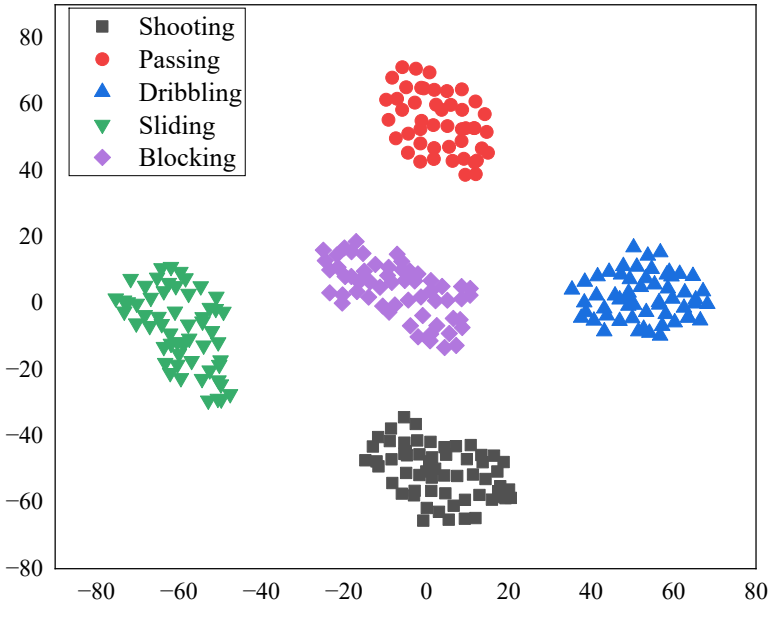
Experience is dynamically prioritised based on the TD error and immediate reward that are updated in real time. The immediate reward reflects the learning value of the experience and can measure the contribution of the experience to the current learning stage, thereby enabling the basketball player to focus on experiences that are more

critical for policy optimisation, accelerating the learning process, and improving the stability of the action optimisation strategy.

5 Experimental results and analyses

This paper uses the BallPlay dataset, which contains 52,973 kinematic data (position, velocity, angle), dynamic data, and contact graphs (CG) of basketball skills, covering full-body actions such as shooting and passing. The dataset is divided into training set, test set, and validation set in a ratio of 6:3:1. The experimental platform uses an Intel R Core TM i9-12900K processor and an NVIDIA RTX 3080 Ti GPU, Python 3.8, 16 GB of running memory, and the PyTorch framework for deep learning calculations, with Adam as the optimiser. In the experiment, the training batch size is set to 64, the capacity of the dynamic priority experience replay pool is 30,000, $\beta = 1$, and the number of training and validation rounds for all models is 300. The discount factor is 0.99 for all models, and the network learning rate is set to 3×10^{-4}.

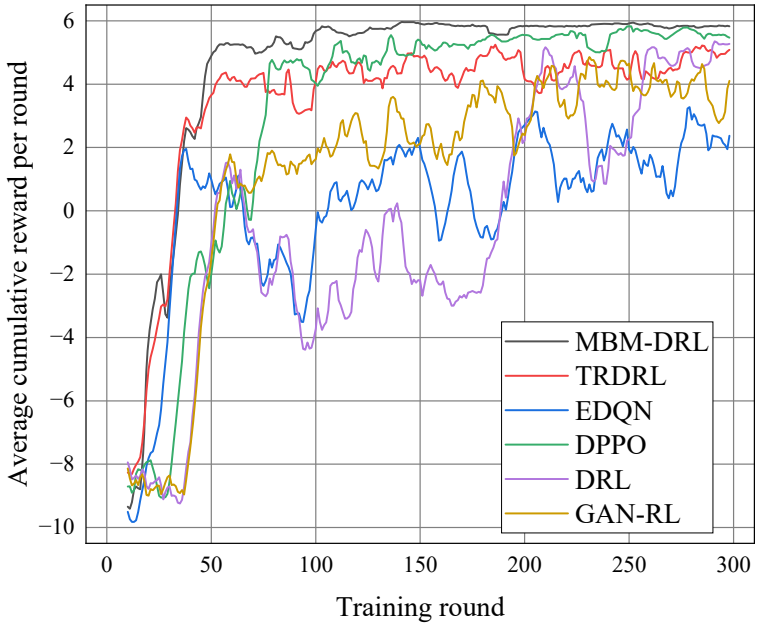
Figure 3 MBM-DRL strategy network feature visualisation (see online version for colours)



The feature visualisation results of various basketball movement actions optimised in the proposed MBM-DRL model are shown in Figure 3. The feature visualisation of the fully connected layer of the MBM-DRL policy network is processed using the T-SNE algorithm, as shown in Figure 3. This figure shows the feature information extracted by the model from the data, where each point represents a sample, and different colors represent different sample labels. By observing, it is found that samples of different types are scattered in relatively independent local areas, and there are clear boundaries between these areas. This result indicates that the basketball action optimisation method based on

MBM-DRL can effectively extract action features and accurately distinguish different types of actions.

Figure 4 Average cumulative rewards per round for different models (see online version for colours)



This paper selects GAN-RL (Chen et al., 2021), DRL (Liu and Hodgins, 2018), DPPO (Wang et al., 2024b), EDQN (Arumugam, 2025), TRDRL (Chang et al., 2025) as comparison models, and analyses the average accumulated reward of different models, as shown in Figure 4. Improving the accumulated reward per round is the goal of model training and also a direct reflection of whether the model converges. The round reward curve of D stabilises after about the 180th round, while the round reward curve of MBM-DRL stabilizes at the 50th round. DRL stabilises at the 251st round, showing poor performance. EDQN stabilises after the 90th round, with performance lower than MBM-DRL but higher than DPPO. These results indicate that compared to the literature GAN-RL, DRL, DPPO, EDQN, TRDRL, MBM-DRL not only has stable and excellent learning ability, but the introduced dynamic priority experience replay mechanism significantly improves the training speed and model stability.

The training time, average reward, success rate, and average steps of different models are compared in Table 1. The proposed model has the shortest training time, the highest success rate, and the fewest execution steps. The action optimisation success rate of MBM-DRL is 97%, which is 24%, 17%, 12%, 8% and 5% higher than the other five models, respectively. The average reward value of MBM-DRL is also 0.5, 1.68, 1.04, 0.89 and 0.17 higher than GAN-RL, DRL, DPPO, EDQN, TRDRL, respectively.

Although GAN-RL embeds a deep learning framework into DRL, the model does not consider the multimodal biomechanical characteristics of basketball players. DRL models the biomechanics of athletes through DRL, which can quickly calculate the joint torque of athletes, but the extraction of multimodal features is insufficient, so its action

optimisation effect is worse than that of MBM-DRL. DPPO optimises the movement trajectory of athletes through the proximal policy optimisation algorithm, but does not introduce a relevant mechanism to optimise the stability of the model, resulting in poor action optimisation effect. EDQN designs an action optimisation method for basketball players based on the improved deep Q network, but does not extract the multimodal biomechanical characteristics of athletes. TRDRL extracts and fuses the action features of different athletes through the transformer, but does not consider biomechanical characteristics, so its action optimisation effect is worse than that of MBM-DRL. Comprehensive analysis above, the MBM-DRL model achieves effective feature extraction and fusion of multimodal biomechanical information, improving the effect of basketball movement action optimisation.

Table 1 Rationality indices for layout optimisation

Model	*Training cycle*	*Average reward*	*Success rate*
GAN-RL	960	5.41	73%
DRL	820	4.23	80%
DPPO	600	4.87	85%
EDQN	450	5.02	89%
TRDRL	90	5.74	92%
MBM-DRL	50	5.91	97%

6 Conclusions

Intending to the issues of insufficient mining of biomechanical characteristics and poor effect of action optimisation in existing approaches for optimising basketball movements, this paper proposes a basketball player action optimisation model based on DRL and multimodal biomechanical modelling. First, biomechanical data analysis and interpolation reconstruction of basketball players are conducted to build a motion state equation for capturing and analysing key data for basketball movement modelling. Then, an action optimisation model for basketball players is designed. The model includes a policy network and a Q-value network, both of which are based on a multimodal feature extraction and fusion module. The policy network is composed of a multimodal feature extraction and fusion module and a fully linked layer, adopted to generate the action probability distribution in the action space. The Q-value network adopts an identical network structure, consisting of a multimodal feature extraction and fusion module and a fully connected network. To accurately capture the spatiotemporal information in multimodal data and achieve efficient fusion of multimodal features, a multimodal feature extraction and fusion module in light of SAM is designed. Through the collaborative effect of the policy network and the Q-value network, the training efficiency and decision-making performance are significantly improved, ensuring the efficiency and stability of the model in the task of optimising basketball movement actions. Experimental outcome indicates that the success rate of the proposed model's action optimisation is 97%, demonstrating the best action optimisation efficiency.

Declarations

The author declares that he has no conflicts of interest.

References

Arumugam, M. (2025) 'Multi-agent deep q-learning for intelligent cricket decision-making: batting and bowling strategy optimization', *Authorea Preprints*, Vol. 4, pp.21–28.

Brauwers, G. and Frasincar, F. (2021) 'A general survey on attention mechanisms in deep learning', *IEEE Transactions on Knowledge and Data Engineering*, Vol. 35, No. 4, pp.3279–3298.

Chang, L., Rani, S. and Akbar, M.A. (2025) 'ChampionNet: a transformer-enhanced neural architecture search framework for athletic performance prediction and training optimization', *Discover Computing*, Vol. 28, No. 1, pp.63–71.

Chen, W-C., Tsai, W-L., Chang, H-H., Hu, M-C. and Chu, W-T. (2021) 'Instant basketball defensive trajectory generation', *ACM Transactions on Intelligent Systems and Technology (TIST)*, Vol. 13, No. 1, pp.1–20.

Cheng, W. and Cheng, W. (2024) 'Optimization research on biomechanical characteristics and motion detection technology of lower limbs in basketball sports', *Molecular & Cellular Biomechanics*, Vol. 21, No. 3, pp.488–488.

Javadpour, L., Blakeslee, J., Khazaeli, M. and Schroeder, P. (2022) 'Optimizing the best play in basketball using deep learning', *Journal of Sports Analytics*, Vol. 8, No. 1, pp.1–7.

Lam, W.K., Kan, W.H., Chia, J.S. and Kong, P.W. (2022) 'Effect of shoe modifications on biomechanical changes in basketball: a systematic review', *Sports Biomechanics*, Vol. 21, No. 5, pp.577–603.

Li, J. (2025) 'Machine learning-based analysis of defensive strategies in basketball using player movement data', *Scientific Reports*, Vol. 15, No. 1, pp.13–27.

Li, X., Luo, R. and Islam, F.U. (2024) 'Tracking and detection of basketball movements using multi-feature data fusion and hybrid YOLO-T2LSTM network', *Soft Computing*, Vol. 28, No. 2, pp.1653–1667.

Liu, L. and Hodgins, J. (2018) 'Learning basketball dribbling skills using trajectory optimization and deep reinforcement learning', *ACM Transactions on Graphics*, Vol. 37, No. 4, pp.1–14.

Liu, Z. and Wang, X. (2023) 'Action recognition for sports combined training based on wearable sensor technology and SVM prediction', *Preventive Medicine*, Vol. 173, pp.10–22.

Pradhan, A. (2012) 'Support vector machine – a survey', *International Journal of Emerging Technology and Advanced Engineering*, Vol. 2, No. 8, pp.82–85.

Ren, H. and Wang, X. (2021) 'Application of wearable inertial sensor in optimization of basketball player's human motion tracking method', *Journal of Ambient Intelligence and Humanized Computing*, Vol. 4, pp.1–15.

Scibek, J.S. and Carcia, C.R. (2013) 'Validation and repeatability of a shoulder biomechanics data collection methodology and instrumentation', *Journal of Applied Biomechanics*, Vol. 29, No. 5, pp.609–615.

Thongpance, N., Dangyai, P., Roongprasert, K., Wongkamhang, A., Saosuwan, R., Chotikunnan, R., Imura, P., Nirapai, A., Chotikunnan, P. and Sangworasil, M. (2023) 'Exploring ResNet-18 estimation design through multiple implementation iterations and techniques in legacy databases', *Journal of Robotics and Control (JRC)*, Vol. 4, No. 5, pp.650–661.

Wang, C. (2023) 'Optimization of sports effect evaluation technology from random forest algorithm and elastic network algorithm', *Plos One*, Vol. 18, No. 10, pp.17–28.

Wang, J., Li, C. and Zhou, X. (2024a) 'Decoding the court: insights into basketball training and performance optimization through time-motion analysis', *Education and Information Technologies*, Vol. 29, No. 18, pp.24459–24488.

Wang, J., Zuo, L. and Martínez, C.C. (2024b) 'Basketball technique action recognition using 3D convolutional neural networks', *Scientific Reports*, Vol. 14, No. 1, pp.13–26.

Wang, W. and Sun, D. (2021) 'The improved AdaBoost algorithms for imbalanced data classification', *Information Sciences*, Vol. 563, pp.358–374.

Wang, X., Wang, S., Liang, X., Zhao, D., Huang, J., Xu, X., Dai, B. and Miao, Q. (2022) 'Deep reinforcement learning: a survey', *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 35, No. 4, pp.5064–5078.

Xiao, M. (2024) 'The best angle correction of basketball shooting based on the fusion of time series features and dual CNN', *Egyptian Informatics Journal*, Vol. 28, pp.10–25.

Xie, Z. and Wu, G. (2024) 'Optimization method of basketball match evaluation based on computer vision and image processing', *Informatica*, Vol. 48, No. 23, pp.1–12.

Yan, X. (2024) 'Effects of deep learning network optimized by introducing attention mechanism on basketball players' action recognition', *Informatica*, Vol. 48, No. 19, pp.45–51.

Yoon, Y., Hwang, H., Choi, Y., Joo, M., Oh, H., Park, I., Lee, K-H. and Hwang, J-H. (2019) 'Analyzing basketball movements and pass relationships using realtime object tracking techniques based on deep learning', *IEEE Access*, Vol. 7, pp.56564–56576.

Zhang, J. and Tao, D. (2023) 'Research on deep reinforcement learning basketball robot shooting skills improvement based on end to end architecture and multi-modal perception', *Frontiers in Neurorobotics*, Vol. 17, pp.12–23.

Zhang, Z., Li, B., Yan, C., Furuichi, K. and Todo, Y. (2025) 'Double attention: an optimization method for the self-attention mechanism based on human attention', *Biomimetics*, Vol. 10, No. 1, pp.23–34.

Zhu, Q. (2022) 'Classification and optimization of basketball players' training effect based on particle swarm optimization', *Journal of Healthcare Engineering*, Vol. 20, No. 1, pp.21–26.