ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

# Design of Chinese dance Performance Movement Optimization and teaching Feedback System based on motion capture model

*Ying Tang*[1,*]

[1] School of Dance, Anhui Art Vocational College, Hefei, Anhui 230001, China

**SUMMARY:** *Aiming at the problems that movement evaluation in Chinese dance teaching relies on empirical observation, feedback lags and is difficult to quantify, this paper designs a performance movement optimization and teaching feedback system based on motion capture model. The system takes multi-view video as input, combines two-dimensional pose estimation, three-dimensional skeleton recovery, spatio-temporal feature modeling and deviation semantic mapping, and realizes the integrated processing of action recognition, quality assessment and correction suggestion generation. The experimental results show that the recognition accuracy of the proposed method on the self-built dataset reaches 94.82%, and Macro-F1 reaches 94.17%. After the system assisted training, the joint Angle error, trajectory deviation, rhythm deviation and center of gravity stability deviation are significantly decreased. This method can improve the refinement level of Chinese dance performance movement recognition, deviation diagnosis and classroom feedback, and provide an implemensible technical path for Chinese dance digital teaching.*

***KEYWORDS:*** *motion capture; Chinese dance teaching; Pose estimation; Teaching Feedback System*

## 1 Introduction

As a comprehensive performing art with program specification, body control and aesthetic expression, Chinese dance has formed a distinctive movement vocabulary, body rhyme structure and teaching inheritance system in the long-term development process. As digital technology continues to enter the scene of art education and stage training, traditional teaching methods that rely on teachers' oral and personal teaching, mirror demonstration and after-class experience correction have been difficult to fully meet the practical needs of high-frequency training, fine evaluation and individual differentiated feedback. On the one hand, Chinese dance movements often contain multi-dimensional collaborative features such as torso extension, joint opening and closing, center of gravity transfer and rhythm conversion. It is easy to miss subtle Angle deviation, timing dislocation and force imbalance by simply relying on naked eye observation. On the other hand, performance training and classroom teaching are also restricted by factors such as site conditions, repeated demonstration costs and students 'immediate understanding ability, which makes the feedback lag and evaluation standard not uniform enough in the process of action optimization. In this context, the introduction of motion capture model into Chinese dance performance movement analysis and teaching feedback has gradually become an important direction in the research of digital dance.

*18605511610@163.com
https://doi.org/10.65102/is2026326
Published: 30 April, 2026

The development of computational methods such as motion capture, pose estimation and spatio-temporal feature modeling provides a feasible path for the structured expression of dance movements. With the help of human keypoint localization, joint trajectory reconstruction and action sequence encoding, the body posture, motion amplitude and rhythm changes of dancers during training can be transformed into computable and comparable digital information. This not only helps to improve the objectivity of action recognition, but also provides a data basis for action quality assessment, error localization, and instructional feedback generation. At the same time, although the existing research has made some progress in dance tracking, motion scoring and real-time posture recognition, the system design for Chinese dance teaching scenarios still has shortcomings. Some methods focus on general motion recognition, and do not describe the synergistic relationship of "form, spirit, strength and law" in Chinese dance. Although some studies can output recognition results, it is difficult to further form interpretable and operable feedback suggestions, resulting in a significant gap between model results and teaching applications. Therefore, how to construct a teaching support system that integrates motion capture, optimization analysis and feedback output around the characteristics of Chinese dance movements is still worthy of in-depth discussion.

Based on this, this paper designs a Chinese dance performance movement optimization and teaching feedback system based on motion capture model. The system takes Chinese dance training video and key point sequence as input, extracts action Angle, trajectory shape, rhythm synchronization and stability indicators through posture feature modeling, and combines the reference action template and deviation analysis mechanism to realize the automatic recognition, quality evaluation and optimization suggestions generation of performance movements. On this basis, the system further constructs a feedback module for teaching scenarios, and outputs the action errors, problem parts, adjustment directions and training suggestions, so as to enhance the pertinence and continuity of classroom training. The research goal of this paper is not only to improve the accuracy of motion recognition, but also to establish a set of computerized analysis framework that can serve the practice of Chinese dance training, so that motion data can truly participate in the teaching decision-making process. The innovation of this paper lies in the following aspects: combining motion capture with Chinese dance posture feature modeling, constructing an action recognition and optimization framework for classroom training; On this basis, the interpretable feedback generation mechanism was introduced to form a closed-loop system of action acquisition, deviation diagnosis and teaching feedback linkage.

This paper is divided into six parts. The first part introduces the research background, problem sources and technical ideas. The second part reviews the related research progress. The third part describes the system construction method. The fourth part is experimental verification. Section 5 opens the discussion; The conclusion is given in Section 6.

# 2 Related Research

In recent years, the research on digital analysis and teaching assistance of dance movements has continued to increase, and the related work mainly focuses on motion capture, pose estimation, motion quality assessment and interactive feedback applications. In terms of dance action recognition, researchers have begun to try to combine convolutional neural networks, keypoint detection and sensor acquisition technologies to improve the computable expression ability of complex action sequences. Mu et al. proposed a dance tracking system assisted by pose estimation, and proved that human keypoint information has obvious value for dance trajectory analysis [1]. Li and Huang combined beat features with key frame

acquisition mechanism for intelligent evaluation of dance movements, and achieved good results in movement timing alignment [2]. Li M. further applied motion capture sensors and machine learning methods to national dance action recognition, indicating that structured action data has positive significance for improving the stability of recognition [3]. Li N et al., Wang et al also discussed the problem of dance action recognition from the perspective of video pose estimation and computer vision detection [4, 5]. This kind of research provides a basis for the digital modeling of dance movements, but most of the work is still focused on the level of "identification", and less involved in the fine-grained diagnosis of torso traction, hand-eye-body movement coordination and movement completion quality common in Chinese dance, and there is also a lack of a close enough mapping relationship between system output and teaching language.

Another category of research focuses more on the construction of action feedback and training support systems. With the development of mixed reality, real-time pose estimation and 3D human modeling technology, researchers begin to try to convert human action recognition results into training feedback directly. Treffer et al. designed a mixed reality mirror system for dance teaching, emphasizing the role of immersive environment in promoting movement imitation and classroom participation [6]. Zhou et al. proved through the study of improvisational dance interaction that the enhanced mirror interface can improve learners' immediate perception of movement changes to a certain extent [7]. Liu et al. proposed a real-time attitude estimation and tracking method for motion performance, which has outstanding performance in inference speed and attitude continuity [8]. Tharatipyakul et al. and Roggio et al. systematically reviewed the research on human pose feedback based on deep learning and pointed out that there are still obvious differences in the feedback interpretability, cross-scene robustness and real-time deployment of existing methods [9, 10]. Dibenedetto et al. used human pose estimation for interpretable corrective feedback generation, indicating that the integration of "recognization-diagnosis-recommendation" is an important development direction of action training system [11]. Ye et al. proposed an action analysis algorithm based on human pose estimation for teacher behavior analysis, which also shows from the side that posture features can not only be used for result discrimination, but also serve for procedural teaching evaluation [12].

At the same time, the development of 3D action recovery and human motion representation learning provides more detailed technical support for the quality optimization of dance movements. Hamilton et al. compared the performance of various computational pose estimation models in joint Angle analysis, and pointed out that the two-dimensional model may still introduce errors under complex occlusion or large body rotation conditions [13]. Choi et al. conducted research on 3D joint Angle estimation, which provided a more reliable quantitative basis for action measurement in the real environment [14]. Dang et al. proposed a video-oriented human kinematics modeling network, which helps to extract stable motion features from long-term actions [15]. Zheng et al. review further shows that deep learning human pose estimation has gradually moved from static detection to spatio-temporal joint modeling [16]. Lu et al. proposed RTMO real-time multi-person pose estimation method, which achieved a good balance between speed and accuracy [17]. MotionBERT and PoseFormerV2 improve the robustness of 3D pose estimation from the perspective of unified action representation learning and frequency domain modeling, respectively [18, 19]. In the direction of action quality evaluation, datasets and models such as FineDiving, FineParser and LucidAction have further promoted action scoring, spatio-temporal parsing and hierarchical quality evaluation to the fine-grained level [20-22]. However, these studies mostly serve for sports movement scoring, general behavior analysis or standardized movement quality

comparison, and the research directly aiming at "how to explain movement deviation, how to correct errors, and how to embed feedback into the classroom process" in the context of Chinese dance teaching is still insufficient. The comparison of related methods is shown in Table 1.

*Table 1: Comparison results of different methods in the Chinese dance action recognition task*

| Authors | Year | Methodologies | Key results and findings | Limitations |
|---|---|---|---|---|
| Mu et al. [1] | 2022 | Pose estimation-assisted dance tracking with CNN | Improved dance motion tracking capability and keypoint representation | Focused more on tracking than teaching feedback |
| Li H et al. [2] | 2024 | Keyframe acquisition with beat-aware motion evaluation | Enhanced temporal alignment and motion quality scoring | Limited interpretability for corrective guidance |
| Li M. [3] | 2024 | Motion capture sensor with machine learning for ethnic dance recognition | Improved recognition stability of dance movements | Weak support for fine-grained action correction |
| Li N et al. [4] | 2023 | Dance image motion recognition based on attitude estimation | Verified the feasibility of pose-based dance image analysis | Relatively insufficient temporal modeling |
| Wang et al. [5] | 2023 | Dance motion detection algorithm based on computer vision | Improved movement detection efficiency in dance videos | Limited capacity for structural posture diagnosis |
| Treffer et al. [6] | 2024 | Mixed reality mirror for dance teaching | Strengthened interactive learning experience | Feedback logic depended heavily on interface presentation |
| Zhou et al. [7] | 2023 | Improvisational dance generation with mixed reality mirror | Improved real-time interaction in dance creation | Not designed for standardized instructional correction |
| Liu et al. [8] | 2024 | Real-time pose estimation and motion tracking with deep learning | Achieved better real-time motion capture performance | More suitable for tracking than pedagogical evaluation |
| Dibenedetto et al. [11] | 2024 | Explainable corrective feedback based on pose estimation | Demonstrated the value of interpretable motion correction | Application scenario was not dance-specific |
| Hamilton et al. [13] | 2024 | Comparative study of computational pose models and 3D motion capture | Provided basis for joint-angle reliability analysis | Focused on measurement comparison rather than system design |
| Xu et al. [21] | 2024 | Fine-grained spatio-temporal action parser | Improved action quality parsing ability | Mainly oriented to general action quality assessment |
| This paper | – | Motion capture-based Chinese dance movement optimization and teaching feedback system | Integrates pose modeling, movement optimization and teaching feedback into one closed-loop framework | – |

In general, existing studies have proved that motion capture, pose estimation and deep learning models can effectively support the recognition, tracking and preliminary evaluation of dance movements, and also provide a method reserve for the construction of digital

teaching systems. However, from the perspective of Chinese dance teaching needs, there are still three shortcomings in related research. First, the expression of the structural characteristics of Chinese dance movements is not sufficient, and the general human posture model is difficult to completely describe the relationship between body rhyme, route, rhythm and center of gravity transition. Second, most methods stop at recognition or scoring, and lack an interpretable feedback generation mechanism for teachers and learners. Thirdly, the closed-loop design at the system level is relatively weak, and the action collection, action optimization and teaching feedback are often dispersed, which is difficult to form a continuous application process. Based on this, this paper intends to construct a Chinese dance performance movement optimization and teaching feedback system based on motion capture model, which realizes the unified organization of deviation recognition, motion optimization analysis and teaching feedback output on the basis of motion pose modeling.

# 3 Chinese dance motion optimization and teaching feedback system construction based on motion capture model

The digital modeling of Chinese dance performance movement not only divides the human movement process into a number of discrete coordinate points, but also identifies the internal relationship between body extension, center of gravity migration, limb opening and closing, movement route and rhythm organization in continuous time series. In traditional classroom, teachers mainly rely on empirical observation and on-site correction to judge the quality of movement, which has distinct characteristics of artistic training, but is also easily affected by observation Angle, movement speed and group teaching environment. When the training object enters the stage of jumping, turning, turning, tilting, and twisting, the local joint deviation is often transmitted to the trunk instability, route deviation or rhythm break in a very short time, and it is difficult to complete the quantitative recording stably by the naked eye. Based on this problem, this paper constructs a motion capture and teaching feedback system for Chinese dance performance training, which makes motion acquisition, posture analysis, feature modeling, motion optimization judgment and feedback output form a unified computational closed loop.

The system is composed of action acquisition layer, pose reconstruction layer, feature encoding layer and feedback mapping layer. The acquisition layer is responsible for obtaining the multi-view video stream of dancers, and synchronizing, cropping and suppressing the noise of the frame sequence. The pose reconstruction layer completes key point detection, timing repair and 3D skeleton recovery. The feature coding layer constructs a multi-scale representation around the sense of line, openness, stability, rhythm consistency and movement completion in Chinese dance training. The feedback mapping layer generates interpretable teaching suggestions according to the reference template and deviation rules. Different from the general human pose analysis, the skeleton modeling in this paper does not stop at the general joint Angle recognition, but further highlights the core variables frequently appearing in Chinese dance teaching, such as shoulder-hip relationship, torso axis, arm extension path, toe orientation and beat alignment, so that the system output can be connected with the classroom language, instead of just remaining at the abstract numerical level. Figure 1 shows the overall process of the system.

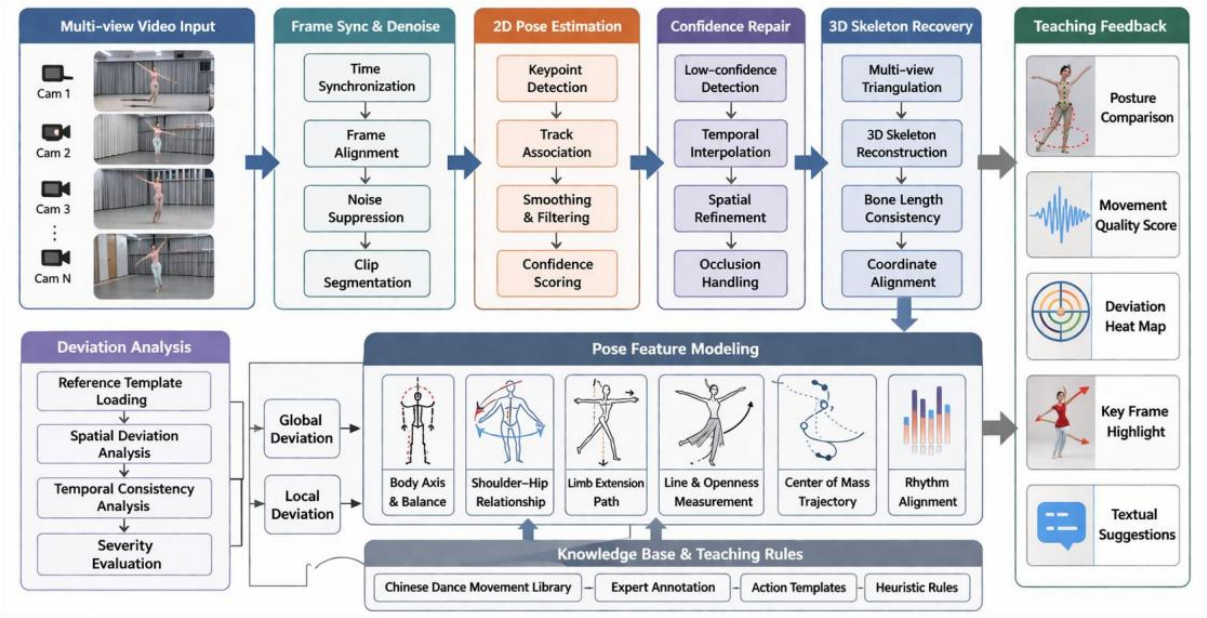


*Figure 1: The overall flow chart of Chinese dance motion capture and teaching feedback system*

## 3.1 Motion capture and posture feature modeling of Chinese dance performance

The core goal of Chinese dance motion capture and pose feature modeling is to transform the continuous, delicate body movement process with strong style attributes into a computable, comparable, and traceable spatio-temporal representation. Considering the actual deployment requirements in dance teaching, this paper takes multi-view video acquisition as the basic input, represents the action sequence as a set of frames of length T, and extracts human key point information in each frame. Let the input video sequence be V, the single frame image be $I_t$, and the set of human keypoints in frame t be $\mathcal{P}_t$, then:

$$\mathcal{V} = \{I_t\}_{t=1}^{T}, I_t \in \mathbb{R}^{H\times W\times 3}, \mathcal{P}_t = \{p_{t,j} \mid j = 1,2,\dots,J\} \tag{1}$$

where, H and W represent the height and width of the image respectively, and J represents the number of keypoints. In order to adapt to the common limb extension and toe control in Chinese dance movements, in addition to the conventional key points such as shoulder, elbow, wrist, hip, knee and ankle, this paper introduces the sternal reference point, pelvic center point and left and right toe derived point, so that the skeleton expression is closer to the context of dance training.

In the 2D posture detection stage, the system uses the key point heat map and confidence joint output to locate the human joints. For the JTH keypoint in frame t, its two-dimensional position and confidence can be obtained from the heat map peak:

$$p_{t,j} = \arg\max_{(x,y)} H_{t,j}(x,y), c_{t,j} = \max_{(x,y)} H_{t,j}(x,y) \tag{2}$$

where $H_{t,j}(x,y)$ is the response intensity of the JTH keypoint at pixel coordinates (x,y), and $c_{t,j}$ reflects the credibility of the current detection. Due to large arm swing, body rotation and occlusion in Chinese dance movements, it is inevitable to produce key point beat and short time miss in single frame detection. In order to reduce the localization noise caused by fast

motion and clothing occlusion, this paper introduces a local temporal smoothing strategy based on confidence to modify the weighted keypoints in adjacent Windows:

$$\tilde{p}_{t,j} = \frac{\sum_{k=-K}^{K} c_{t+k,j}\ p_{t+k,j}}{\sum_{k=-K}^{K} c_{t+k,j} + \varepsilon} \tag{3}$$

where $\tilde{p}_{t,j}$ are the smoothed 2D coordinates, K is the time window radius, and ε is the minimal constant to prevent the denominator from being zero. After this step, the jitter amplitude of the key point sequence will decrease significantly, especially for the action segments with fine hand trajectories such as Lanhua-finger, Yun-hand, and slew-throwing.

Considering that the classroom scene does not always provide high-precision optical capture equipment, this paper adopts the modeling strategy of "multi-view 2D estimation - 3D skeleton recovery" to seek a balance between cost and accuracy. For the same keypoint observation from M views, the system recovers its 3D position $q_{t,j}$ by minimizing the reprojection error:

$$q_{t,j} = \arg\min_{q} \sum_{m=1}^{M} \left\| \pi_m(q) - \tilde{p}_{t,j}^{(m)} \right\|_2^2 \tag{4}$$

Here, $\pi_m$ denotes the projection function of the MTH camera, and $\tilde{p}_{t,j}^{(m)}$ are the corrected 2D keypoints in that view. The 3D skeletons obtained in this way, although not pursuing lab-level absolute millimeter accuracy, are sufficient to support pose comparison, Angle calculation, and motion deviation analysis in classroom training.

After the recovery of the 3D skeleton, it is also necessary to normalize the action sequences under different height, body shape and stance conditions, otherwise the scale differences between different learners will interfere with the action evaluation. In this paper, the center point of the pelvis is taken as the root node, the shoulder width and hip width are combined to construct the human scale factor, and the skeleton coordinates are normalized by translation and scaling:

$$\bar{q}_{t,j} = \frac{q_{t,j} - q_{t,r}}{s_t}, s_t = \frac{1}{2}\left(\|q_{t,ls} - q_{t,rs}\|_2 + \|q_{t,lh} - q_{t,rh}\|_2\right) \tag{5}$$

where, $q_{t,r}$ represent the position of the root node in frame t, ls,rs,lh, represent the left shoulder, right shoulder, left hip and right hip, respectively. After this normalization operation is completed, the action data is more suitable for cross-individual comparison, and it is also convenient for subsequent template matching.

In Chinese dance training, only three-dimensional coordinates are still not enough to fully reflect the quality of movement, because many teaching judgments are not directly based on the point position itself, but on the joint opening, torso direction and line direction. Therefore, this paper further constructs multi-layer features around joint angles, motion speed, stability, and rhythmic consistency. For any central joint j, if its adjacent bone endpoints are a and b, the corresponding joint Angle can be expressed as follows.

$$\theta_{t,j} = \arccos\left(\frac{(q_{t,a} - q_{t,j}) \cdot (q_{t,b} - q_{t,j})}{\|q_{t,a} - q_{t,j}\|_2 \cdot \|q_{t,b} - q_{t,j}\|_2}\right) \tag{6}$$

The Angle feature can be directly used to analyze indicators such as arm opening degree,

knee extension degree, and the Angle relationship between torso and lower limbs. This feature is especially critical for action clips such as "side leg", "sea exploration" and "big tuck step", because action aesthetics are often highly correlated with Angle control accuracy.

In addition to the static Angle, the velocity and acceleration in the dynamic action also determine whether the action is clean and whether it is dragging. In this paper, the velocity and acceleration amplitude of each joint are calculated based on the normalized 3D skeleton to characterize the force generation process and conversion rhythm of the action:

$$v_{t,j} = \frac{\|\bar{q}_{t,j} - \bar{q}_{t-1,j}\|_2}{\Delta t}, a_{t,j} = \frac{|v_{t,j} - v_{t-1,j}|}{\Delta t} \tag{7}$$

where, $\Delta t$ is the adjacent frame time interval. For Chinese dance, the speed feature can reflect whether the movement cohesion is smooth, and the acceleration feature is helpful to identify the problems such as excessive switching of the center of gravity and insufficient fluctuation control. If a learner shows a significant speed spike and subsequent trajectory drift when turning and taking the turn, the system can catch the situation that the learner is too hard or unstable through this index.

In order to make the skeleton expression more in line with the aesthetic logic of Chinese dance, in addition to the universal joint features, this paper also constructs the subregional feature structure, as shown in Figure 2. In this structure, the human body is divided into the trunk axis area, the upper limb line area, the lower limb support area and the extension endpoint area, and the key indicators are calculated respectively. The reason for such processing is that "standing upright", "walking round arms" and "having roots under feet" in Chinese dance are not requirements at the same level. If all features are simply spliced, the model is easy to ignore the difference in the importance of local areas.

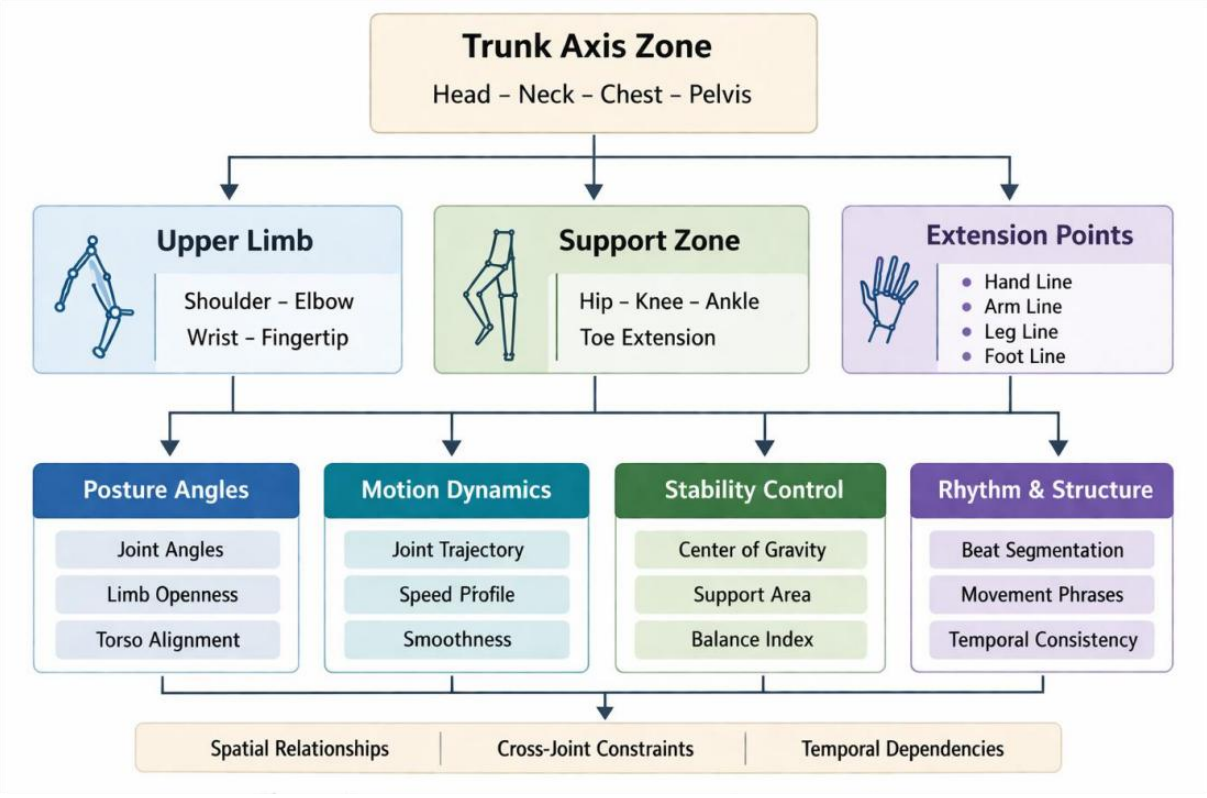


*Figure 2: Schematic diagram of key joint topology and hierarchical features of Chinese dance*

Based on this, this paper introduces the rhythm alignment index to measure the degree of synchronization between action completion and music beat point. Let the critical moment of the reference action or music beat be $\tau_i$, and the action peak moment actually detected by the learner be $\hat{\tau}_i$. Then the rhythm consistency can be expressed as follows.

$$R = 1 - \frac{1}{N_b}\sum_{i=1}^{N_b}\frac{|\hat{\tau}_i - \tau_i|}{\delta} \tag{8}$$

Here, $N_b$ is the number of beat points and $\delta$ is the maximum time deviation allowed. The closer this index is to 1, the more consistent the action changes are with the beat organization. For the Chinese dance classroom, the rhythm problem is not only the speed of the music is not accurate, but more reflected in the body's beginning and ending and the rhythm of the phrase. With R, the system can say not just "did you get it right" but "did you get it right?"

In order to further describe the overall structure of action progression over time, this paper organizes the continuous skeleton sequence into a spatio-temporal graph, and learns the adjacency relations and cross-frame dependencies between joints in a graph convolution manner. Let the feature of the JTH node in the TTH frame in the LTH layer be $h_{t,j}^{(l)}$, the adjacency matrix be A, and the node degree be $d_j$. The graph convolution update process can be written as follows.

$$h_{t,j}^{(l+1)} = \sigma\left(\sum_{k\in\mathcal{N}(j)}\frac{A_{jk}}{\sqrt{d_j d_k}}W^{(l)}h_{t,k}^{(l)} + U^{(l)}h_{t-1,j}^{(l)}\right) \tag{9}$$

where N(j) represents the neighborhood set of node j, $W^{(l)}$ and $U^{(l)}$ are the spatial and temporal weight matrices, and $\sigma(\cdot)$ is the nonlinear activation function. This modeling method can preserve the skeleton structure of single frame and the movement evolution law across frames at the same time, and has a good expression ability for the sequential characteristics of Chinese dance such as continuous turn, wavy arm advance and center of gravity transition.

Figure 3 shows the idea of spatio-temporal skeleton modeling and template alignment adopted in this paper. The system does not directly compare each frame of action in isolation, but maps the learner sequence and the standard sequence into the same action representation space, and then comprehensively analyzes the Angle deviation, route deviation and rhythm deviation. After this processing, the output result of the system is no longer just "high and low similarity", but can locate the time period and body parts where the problem occurs, and provide a clear entrance for the subsequent teaching feedback module.

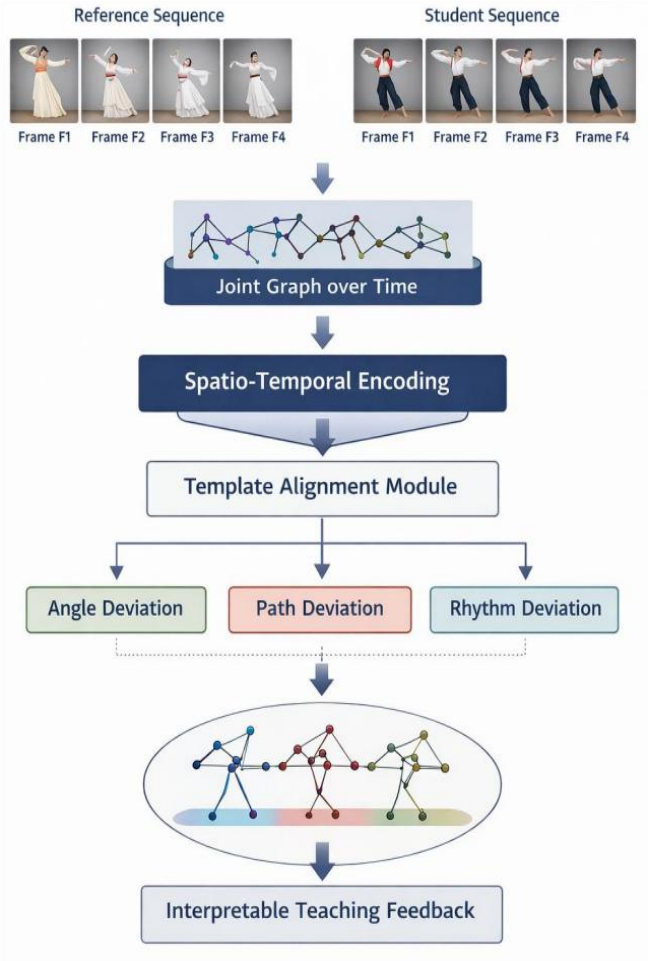


*Figure 3: Schematic diagram of spatio-temporal skeleton modeling and template alignment*

In summary, this paper completes three levels of processing in the stage of motion capture and pose feature modeling. First, a stable skeleton representation of Chinese dance is established through multi-view 2D key point detection and 3D recovery. Secondly, the motion quality is characterized by Angle, velocity, acceleration and rhythm. Thirdly, the overall organization law of action sequences is learned through spatio-temporal graph modeling. After this layer of modeling, the body language originally attached to the performance moment is transformed into computable structural features, which lays a unified data foundation for subsequent action optimization judgment and teaching feedback generation.

## 3.2 Chinese dance movement Optimization and teaching feedback System Design

After capturing the movements of the Chinese dance performance, recovering the 3D skeleton and modeling the posture features, the task of the system is no longer just to identify "what movements have been done", but to further judge "how well the movements have been done, where the problems have occurred, and how to correct them". This step determines whether motion capture technology can really enter the teaching scene. If the system can only output category labels or some abstract scores, its help for Chinese dance classroom is still limited, because teachers really care about whether the students' arm lines are in place, whether the trunk axis is stable, whether the movement route is deviated, and whether the beat drop point is accurate, and how these problems should be corrected. Based on this requirement, this

paper further designs a Chinese dance movement optimization and teaching feedback system based on the above posture feature modeling, so that the movement sequence is transformed from "identifiable" to a teaching object that is "evaluable, interpretable, and interventionable".

The system consists of an action template library, a sequence alignment module, a deviation assessment module, a feedback generation module and an interaction presentation module. The template library is used to store standard action segments and their segment labels. The sequence alignment module solves the problems of different learners 'different speeds and inconsistent starting times. The deviation evaluation module constructed a comprehensive score around Angle, route, rhythm, center of gravity and stability. The feedback generation module is responsible for mapping the numerical results into the teaching language. The interactive presentation module outputs the correction information to the teacher end and the student end synchronously, forming a training closed loop. Figure 4 shows the overall structure of the system.

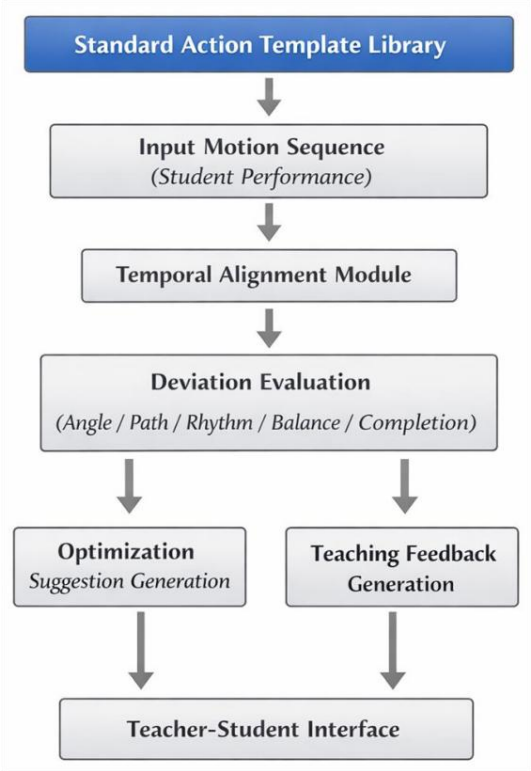


*Figure 4: Structure diagram of Chinese dance movement optimization and teaching feedback system*

In order to enable the system to deal with multiple types of movements in Chinese dance training, this paper first constructs a standard movement template library. Let the type c Chinese dance be taken as $A_c$, which contains $N_c$ standard demonstration samples, then the template library can be expressed as follows.

$$\mathcal{X} = \{S_{c,n} \mid c = 1,2,\dots,C;\ n = 1,2,\dots,N_c\} \tag{10}$$

where, C denotes the total number of action categories and $S_{c,n}$ denotes the NTH standard sequence of the CTH action category. Different from the common action classification library, the template library in this paper does not only store the skeleton sequence itself, but also appends structural labels such as "beginning, unfolding, transformation, and wrapping" to

each action, so that the system can determine the stage position of the deviation in a more fine-grained way.

For any standard sample $S_{c,n}$, in this paper, it is represented as a time sequence fragment composed of consecutive feature frames:

$$S_{c,n} = \{z_1, z_2, \dots, z_L\}, z_t \in \mathbb{R}^d \tag{11}$$

Here, L is the action sequence length, d is the feature dimension of a single frame, and $z_t$ contains the information of joint angles, velocity, acceleration, axis offset, rhythm markers, and local stability obtained in Section 3.1. The significance of this treatment is that the system is no longer comparing simple coordinate points, but composite representations directly related to the quality of Chinese dance movements.

Due to the obvious differences in the learners' movement speed, starting frame and transition duration, the rhythm change will be misjudged as movement error if it is directly compared frame by frame. Based on this, the system introduces a dynamic temporal alignment mechanism to align the learner sequence $Q = \{q_1, q_2, \dots, q_T\}$ and the standard sequence $S_{c,n}$ for elastic matching. Its cumulative alignment cost is defined as follows.

$$D(i,j) = \|q_i - z_j\|_2 + \min\{D(i-1,j), D(i,j-1), D(i-1,j-1)\} \tag{12}$$

where, D(i,j) represents the optimal cumulative distance between frame i of the learner and frame j of the template. Through this formula, the global minimum cost path can be obtained, so that the action segments with different speeds but similar structures can be mapped into the same comparison coordinate system. This treatment is particularly necessary for adagio openings, alleyway connections, and rhythmic appearances in Chinese dance, otherwise differences in movement styles can be mistaken for technical errors.

After completing the timing alignment, the system enters the bias evaluation phase. Considering that the quality of Chinese dance movements cannot be summarized by a single index, this paper divides the deviation into four levels: Angle deviation, route deviation, rhythm deviation and center of gravity stability deviation. For the JTH key joint, the Angle error between the learner and the template in the ith frame after alignment is defined as follows.

$$e_{i,j}^{\theta} = \left|\theta_{i,j}^{(q)} - \theta_{i,j}^{(s)}\right| \tag{13}$$

Here, $\theta_{i,j}^{(q)}$ and $\theta_{i,j}^{(s)}$ denote the corresponding joint angles of the learner and the standard template, respectively. This index is suitable for analyzing problems such as insufficient arm opening, insufficient knee flexion and extension, and excessive trunk tilting back or forward. For the fixed-point movements in the Chinese dance classroom, the Angle error often directly affects the action lines, so its weight cannot be too low.

In addition to joint angles, whether the movement route is smooth or not significantly affects performance completion. In this paper, the high discernibility endpoints such as wrist, toe and head are used as observation objects to construct the trajectory deviation index. Let the sequence of trajectories of the u-th endpoint in an action segment be $\Gamma_u^{(q)}$ and $\Gamma_u^{(s)}$, respectively, then its path error is defined as follows.

$$e_u^p = \frac{1}{T_a}\sum_{i=1}^{T_a}\left\|\gamma_{u,i}^{(q)} - \gamma_{u,i}^{(s)}\right\|_2 \tag{14}$$

where, $T_a$ is the length of the aligned fragments. This index can better reflect the problems such as "whether the arm is round", "whether the foot route is floating outside" and "whether the sleeve swing arc is complete". For performance forms such as Chinese dance that attach great importance to trajectory, it is far from enough to compare static stopping points. Path continuity also needs to be included in the evaluation.

Rhythm consistency is another key dimension in instructional feedback. Many movements seem to be similar in shape, but the actual problem is in the timing of entry and the end of the movement. To this end, this paper combines music beat points and action energy peaks to construct rhythm deviation scores. Let the template beat point be $\tau_k$, and the learner's corresponding action peak time be $\hat{\tau}_k$. Then the rhythm error is denoted as follows.

$$e^r = \frac{1}{K}\sum_{k=1}^{K}\frac{|\hat{\tau}_k - \tau_k|}{\Delta} \tag{15}$$

Here, K represents the number of key rhythm points and $\Delta$ is the normalized time scale. The smaller this formula is, the better the rhythm synchrony is. If the learner has the phenomenon of snap or drag at the connection position of "start-accept-transfer-receive", the system can identify it through this index.

Chinese dance training also highly emphasizes "stand, steady and control", so the deviation of center of gravity stability cannot be ignored. In this paper, the offset between the projection of the center of the pelvis and the center of the support domain is used to measure the steady-state level of the action. Let the barycentric projection point of frame i be $g_i$ and the center of the support region be $b_i$, then the stable deviation is as follows.

$$e^b = \frac{1}{T_a}\sum_{i=1}^{T_a}\| g_i - b_i\|_2 \tag{16}$$

When the value is too large, it often means insufficient lower limb support, unstable retracting after turning, or insufficient trunk control during high movements. For teaching scenarios, this index can help teachers quickly determine whether the problem originates from the upper limb circuit or from the lower disk support and core control. In order to integrate the deviation of different dimensions, this paper constructs a comprehensive scoring function of action quality:

$$\text{Score} = 100 - \left(\alpha\bar{E}^{\theta} + \beta\bar{E}^{p} + \gamma E^{r} + \eta E^{b}\right) \tag{17}$$

where $\bar{E}^{\theta}$ represents the average Angle deviation of the whole joint, $\bar{E}^{p}$ represents the average path deviation of the multi-endpoint, $E^r$ and $E^b$ represent the rhythm deviation and stability deviation respectively, $\alpha,\beta,\gamma,\eta$ are the weight coefficients, and $\alpha+\beta+\gamma+\eta=1$ is satisfied. This score does not seek to completely reduce artistic performance to a single numerical value, but rather provides a comparable overall scale that enables the system to support hierarchical training, stage control, and class management.

Grading alone is still not enough to support instructional feedback, because students need to know "what went wrong" and teachers need to know "what kinds of problems are most worthy of priority correction". Therefore, the system further constructs the feedback priority index. For the m-th problem corresponding to the JTH part, its priority is defined as follows.

$$P_{j,m} = \lambda_1 \cdot \bar{e}_{j,m} + \lambda_2 \cdot \omega_j + \lambda_3 \cdot \rho_m \tag{18}$$

where $\bar{e}_{j,m}$ represents the average error of the problem in the current action, $\omega_j$ represents the importance weight of the part, and $\rho_m$ represents the influence coefficient of this kind of problem on the overall action completion. For example, if "trunk axis tilt" and "insufficient fingertip extension" are present at the same time, the former should usually be given priority feedback because it further affects movement stability and breath organization.

Based on this, the system maps numerical deviations into instructional semantics. In order to avoid the feedback text being mechanically single, this paper uses the generation method of "rule base + template variant" to combine the problem location, deviation direction, severity and adjustment suggestions. For example, when the arm trajectory deviation is large and the rhythm deviation is small, the system gives the prompt of "the arm is not open enough, the arc route is narrowed in advance, it is recommended to extend the elbow drive time and keep the wrist extension". When the deviation of the center of gravity continues to be too large, the feedback of "the support foot bearing is unstable, the trunk is not lifted enough, it is recommended to reduce the speed before turning and strengthen the core control" is generated. The feedback mapping relationship is shown in Figure 5.

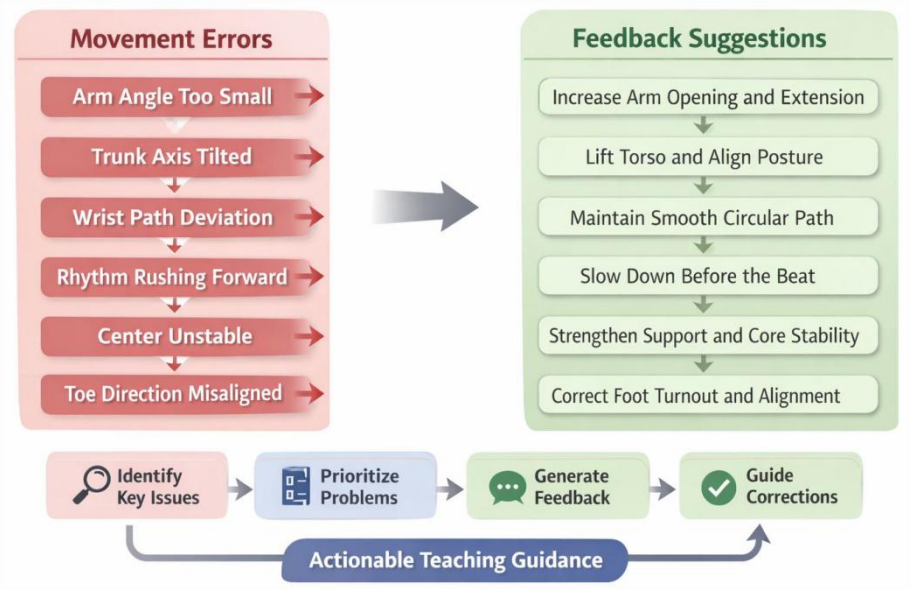


*Figure 5: Semantic mapping diagram between action deviation and teaching feedback*

In order to adapt the system to classroom scenarios instead of only serving offline analysis, this paper also designs a dual-channel output mechanism on the teacher end and the student end. The student side is more emphasis on instant, concise and executable, and mainly displays the total score, hot spot of problem parts, key frame comparison and one or two core suggestions. The teacher side retains more detailed data, including action segmentation scores, deviation curves for each joint, rhythm synchronization curves, and class comparison results. The reason for this distinction is that the more information the more effective the teaching feedback is, if the students present too much data at the same time, it may weaken the pertinence of correction. The teacher side needs more complete details to support hierarchical teaching and long-term tracking.

The closed-loop of the classroom operation of the system is shown in Figure 6. After the dancer completed the movement, the platform extracted the posture features in real time and aligned the template. Then, the comprehensive score, problem location and feedback suggestions were output. Teachers can manually confirm or revise according to the system results. The revised feedback is again written into the student training profile and becomes the

reference for the next round of action optimization. In this way, the system is not to replace the teacher's judgment, but to provide more detailed, stable and traceability data support for action analysis under the premise of maintaining the subjectivity of teaching.

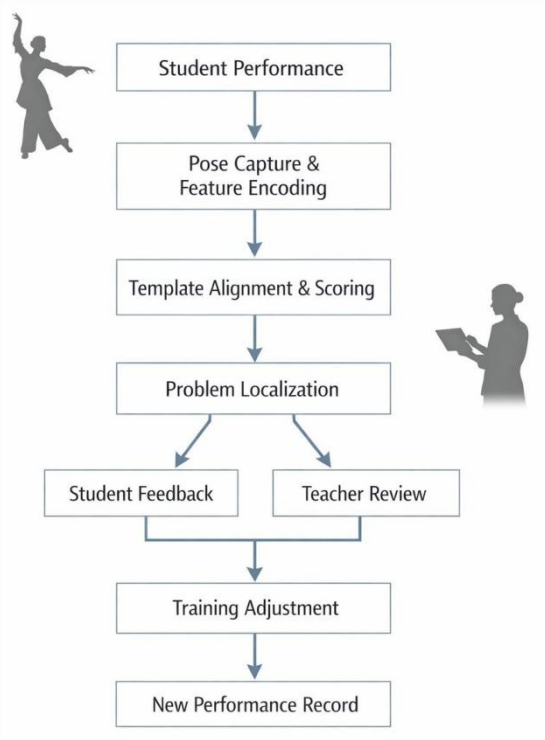


*Figure 6: Flow chart of closed-loop application of feedback in Chinese dance teaching*

# 4 Experimental verification of Chinese dance motion optimization and teaching feedback system based on motion capture model

## 4.1 Chinese dance action recognition and optimization effect experiment

In order to verify the effectiveness of the Chinese dance motion optimization and teaching feedback system based on motion capture model constructed in this paper at the two levels of motion recognition and motion optimization, the experiment was carried out around "whether the recognition is accurate" and "whether the optimization is effective". The experimental platform is deployed on Ubuntu 22.04 environment with Intel Core i7-12700 processor, 32 GB RAM and NVIDIA RTX 4080 16 GB GPU. The development framework is Python 3.11, PyTorch 2.2 and OpenCV 4.8. The two-dimensional key point extraction module was initialized by RTMO pre-training weights, and the three-dimensional pose recovery and spatio-temporal graph modeling module was fine-tuned and trained on a self-built Chinese dance movement dataset. The training batch size is set to 32, the initial learning rate is $2\times10^{-4}$, the optimizer is AdamW, the maximum training round is 80, and the validation set is stopped early if there is no boost for 8 consecutive rounds.

The experimental data consists of a self-built Chinese dance movement dataset. The data collection objects were 48 dance learners and 12 professional demonstrators. A total of 3,960 action clips were recorded, covering 12 types of common Chinese dance training movements, such as cloud hand, mountain shoulder, step over, diving into the sea, shooting the yan, side

leg, lying fish in front of the catch, and round field step. The duration of each video was controlled at 6-12 s, the frame rate was 30 fps, and the action category, key stage boundary and action completion quality level were annotated synchronously. All samples are divided into training set, validation set and test set according to 8:1:1. Considering that pure category recognition is not enough to reflect the value of the system, we set up two sets of experiments. One is to compare the classification performance of different models in action recognition tasks. The other set examines whether the action optimization results given by the system can significantly reduce the critical attitude error.

In the action recognition experiment, four groups of models including 3D-CNN, two-stream LSTM, ST-GCN and the proposed method are selected as comparison models. The evaluation metrics include Accuracy, Macro-F1, and average inference delay. Table 2 presents the overall results. It can be seen that the proposed method maintains a more balanced performance in the three indicators, especially in the recognition accuracy and time series stability. Compared with ST-GCN, the Accuracy of the proposed method is improved from 91.36% to 94.82%, and Macro-F1 is improved from 90.91% to 94.17%. This indicates that in tasks such as Chinese dance that are sensitive to movement details and stage transitions, it is still not enough to rely on general skeleton spatio-temporal modeling. If rhythm alignment, route deviation and local structure constraints are further introduced, the model's ability to distinguish similar movements will be stronger. In terms of inference delay, the average single segment of the proposed method is 38 ms, which is slightly higher than that of the pure lightweight classification model, but it can still meet the requirements of quasi-real-time feedback in classroom training.

*Table 2: Comparison results of different methods in the Chinese dance action recognition task*

| Model | Accuracy / % | Macro-F1 / % | Precision / % | Recall / % | Avg. Latency / ms |
|---|---|---|---|---|---|
| 3D-CNN | 86.47 | 85.91 | 86.22 | 85.63 | 41 |
| Dual-stream LSTM | 88.94 | 88.27 | 88.65 | 87.96 | 46 |
| ST-GCN | 91.36 | 90.91 | 91.14 | 90.73 | 35 |
| Proposed method | 94.82 | 94.17 | 94.39 | 93.98 | 38 |

In order to further observe the recognition differences of the model on different action categories, this paper counts the classification accuracy of six representative actions, and the results are shown in Figure 7. It can be seen from the figure that the proposed method is more obvious in the actions with high requirements for route continuity or center of gravity control, such as "cloud hand", "sea exploration", "swallow shooting" and "step over". Among them, the accuracy of "exploring the sea" action recognition reached 95.8%, which was 4.7 percentage points higher than ST-GCN. "Step over" reaches 93.9%, which is 6.4 percentage points higher than the two-stream LSTM. The reason for this difference is that these types of actions not only contain static modeling differences, but also contain obvious coupling relationships between timing advancement and local posture. In the feature modeling stage, the proposed method has incorporated joint angles, endpoint trajectories and rhythmic drop points into the unified representation, so it is easier to identify the structural differences between similar actions.

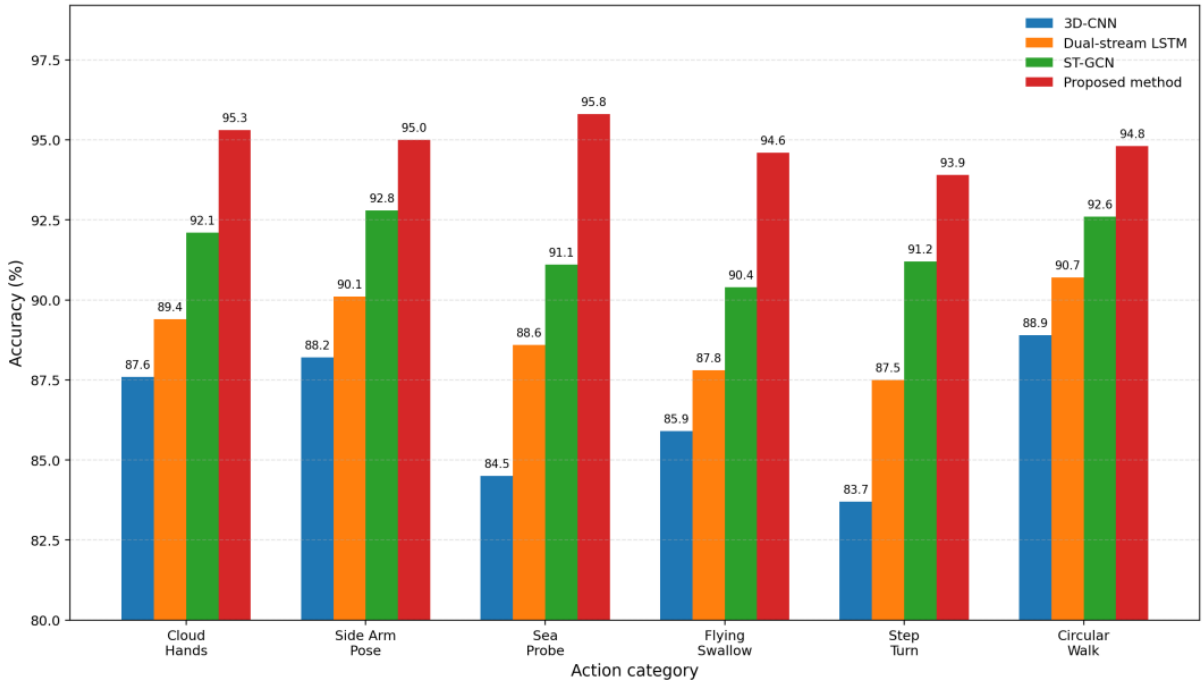


*Figure 7: Comparison of recognition accuracy of different models on representative Chinese dance movements*

In the action optimization effect experiment, this paper does not investigate the category discrimination, but directly tests whether the optimization suggestions output by the system can reduce the key posture error. The specific methods were as follows: 288 action clips of 36 learners were extracted from the test set, and the posture indicators of the learners when they completed the action for the first time were recorded. Then, two rounds of targeted exercises were carried out according to the optimization suggestions given by the system, and the action sequence was collected again and the error change was calculated. The evaluation metrics include the average joint Angle error, endpoint trajectory deviation, rhythm deviation, and center of gravity stability deviation. The results show that the four indicators are significantly improved after the system-aided optimization. As shown in Figure 8, all key indicators showed a consistent downward trend before and after action optimization, and the average error of joint Angle, endpoint trajectory deviation, rhythm deviation and center of gravity stability deviation were significantly reduced, indicating that the correction suggestions generated by the system could effectively improve the quality of action completion in short cycle training. The average joint Angle error is reduced from 9.84 to 6.21, the endpoint trajectory deviation is reduced from 0.143 to 0.091, the rhythm deviation is reduced from 0.118 to 0.072, and the center of gravity stability deviation is reduced from 0.087 to 0.053.

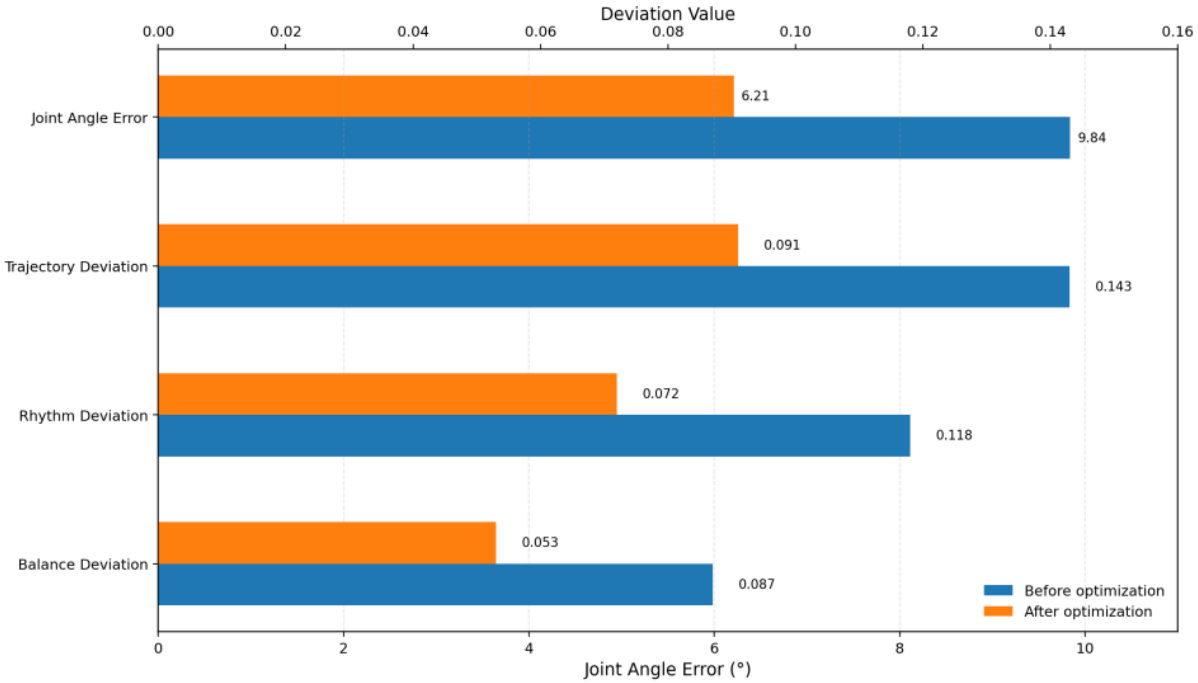


*Figure 8: Change results of key indicators before and after action optimization*

## 4.2 Application experiment of teaching feedback system

In order to test the application effect of the teaching feedback system constructed in this paper in the real Chinese dance training scene, the experiment was carried out in the 8-week teaching process of a higher vocational dance course. The participants included 42 Chinese dance learners and 6 teachers. The training content covered the high-frequency movements in the classroom, such as cloud hand, mountain bang, exploring the sea, circular field step and stepping over. The system was deployed in Ubuntu 22.04 environment. The teacher side viewed the segmented scores, hot spots of problem parts and feedback statements through the Web interface, and the student side received key frame comparison, error tips and brief correction suggestions through the tablet terminal. The experiment no longer examined the action category discrimination itself, but focused on the feedback trigger accuracy, system response time, teacher adoption and user experience.

In order to ensure the comparability of the evaluation, the feedback output by the system is divided into four categories: Angle correction, route correction, rhythm correction and stability correction. During the 8-week teaching process, 840 decidable feedback trigger events were recorded by the system, which were manually reviewed by two dance teachers with the title of associate senior or above. The statistical results show that the system can stably identify the main problems and output the corresponding suggestions in the classroom scene. As shown in Figure 9, the trigger accuracy of the four types of feedback are all higher than 85%, among which the accuracy of Angle correction is the highest, 92.8%, stability correction is 90.4%, route correction is 88.9%, and rhythm correction is relatively slightly lower, 85.7%. The corresponding teacher adoption rates were 90.5%, 88.6%, 86.8% and 83.9%, respectively. This shows that the feedback generated by the system can be consistent with the teacher's judgment in most cases, especially in the structural problems such as joint opening, trunk axis and support stability. The feedback of rhythm class is slightly lower, which is mainly related to the strong action stop and connection processing of some students and incomplete rules of music syntactic boundaries.

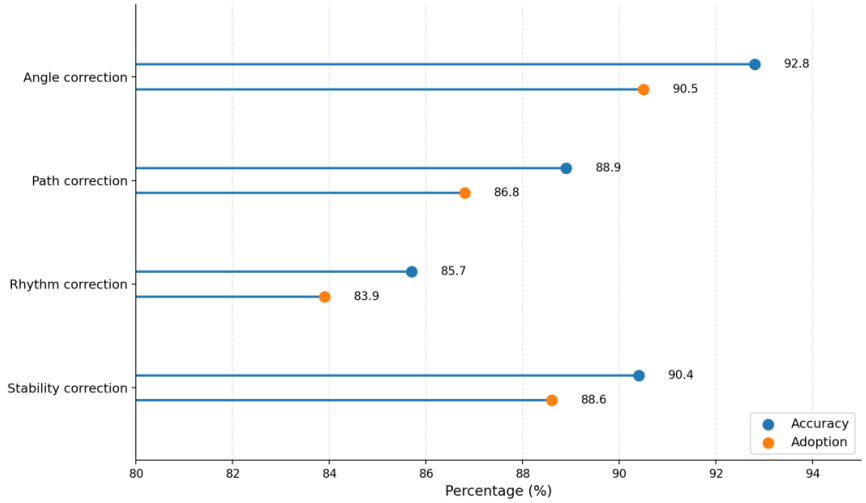


*Figure 9: Comparison of trigger accuracy and teacher adoption rates for different feedback types*

In addition to the accuracy of the feedback, whether the system can quickly return results within the pace of the class also determines the practical value. Therefore, this paper compares the "system automatic feedback mode" with the "traditional teacher oral comment mode after watching the video", calculates the average time required from the completion of a single action to the presentation of feedback, and records the error reduction rate of students in the next round of practice. Figure 10 shows that the average response time of the automatic feedback mode on the five types of actions remains between 1.21 and 1.48 s, which is significantly lower than that of the traditional mode of 6.72 to 8.15 s. After two consecutive rounds of practice, the error correction rate of the system-assisted group was steadily higher than that of the manual review group, reaching an average of 31.6%, while that of the traditional mode was 18.9%. This result indicates that the short feedback link helps students to complete the immediate correction before the action memory has not decay, so as to improve the efficiency of classroom correction. For training programs that rely on muscle sensation and body coherence, such as Chinese dance, the closer the feedback is to the moment of action, the higher the intervention value is usually.

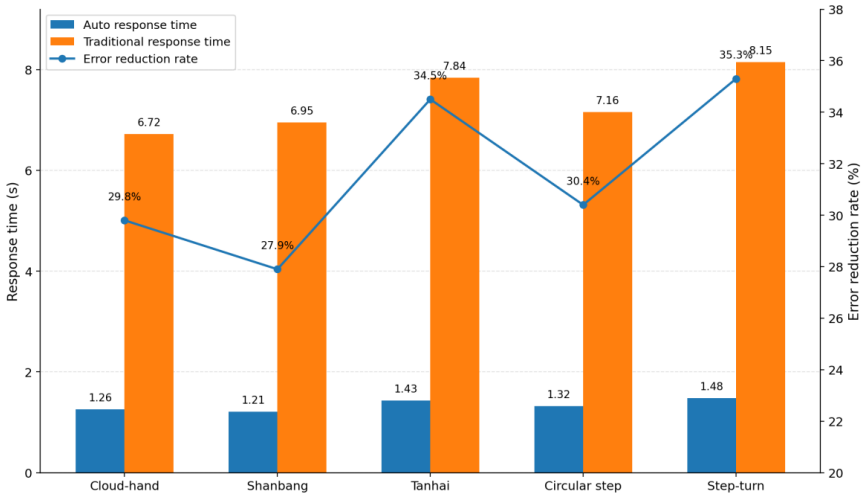


*Figure 10: Comparison of average response time and error fallback rate under different feedback modes*

After completing the statistics of process indicators, this paper further collects user experience through questionnaires and interviews. The questionnaire was evaluated from the four dimensions of feedback clarity, classroom auxiliary value, operation convenience and continued use intention using the hundred-mark system. The results are shown in Table 3. The scores of students on feedback clarity and classroom auxiliary value were 91 and 93, respectively, indicating that the prompt language generated by the system could be understood and directly used for action adjustment. The teacher group rated the value of classroom assistance as 95, indicating that the system has high practical help in problem localization and hierarchical guidance. In contrast, the score of operation convenience is slightly lower, 84 for the teacher side and 87 for the student side. The interview shows that the reasons mainly focus on the number of action playback switching and the local hot spot labeling can still be further simplified. The willingness to continue to use both groups were more than 90 points, indicating that the system had a good teaching acceptance.

*Table 3: User experience evaluation results of the teaching feedback system*

| Group | Feedback Clarity | Teaching Support Value | Ease of Use | Continued Use Intention | Main suggestion |
|---|---|---|---|---|---|
| Students (n=42) | 91 | 93 | 87 | 92 | Add more segmented demonstration clips and slower replay options |
| Teachers (n=6) | 89 | 95 | 84 | 94 | Refine rhythm feedback wording and strengthen class-level statistics |

# 5 Discussion

The Chinese dance performance movement optimization and teaching feedback system based on motion capture model proposed in this paper has achieved relatively stable results in terms of movement recognition accuracy, movement optimization effect and classroom feedback efficiency. Combined with the experiments in Chapter 4, it can be seen that the recognition accuracy of the method in this paper on the self-built Chinese dance movement dataset reaches 94.82%, and Macro-F1 reaches 94.17%. At the same time, the joint Angle error, trajectory deviation, rhythm deviation and center of gravity stability deviation are significantly improved. This result indicates that the integration of multi-view pose estimation, 3D skeleton recovery, spatio-temporal graph modeling and feedback semantic mapping into the same computational framework can more effectively capture the intrinsic relationship between line control, rhythm organization and body stability in Chinese dance movements. Compared with the general general action recognition model, the proposed method is more adaptable to the Chinese dance training scene, because the system not only focuses on "whether the action is recognized", but also further analyzes "where the action deviation occurs and how it affects the completion quality". However, some limitations are still exposed in the experiment. For action clips with small amplitude, insignificant direction change or strong partial occlusion, the stability of system recognition and feedback will still decrease. In terms of rhythm feedback, due to differences in the processing of music syntax and action stop-connect among different learners, some prompts still need to be revised twice by teachers. In addition, although the current system can meet the quasi-real-time application in the classroom, there is still room for further optimization of response efficiency under the

conditions of complex combination of actions and multi-person training at the same time. Subsequent research can further train the model on a larger scale of Chinese dance data sets, enhance the recognition ability of fine-grained movement and style expression, and improve the feedback text generation mechanism and multi-terminal interaction mode, so that the system can play a more stable supporting role in dance teaching, performance training and digital curriculum construction.

# 6 Conclusions

Aiming at the problems that action recognition in Chinese dance teaching relies on experience judgment, feedback lags and is difficult to quantify, this paper designs a motion optimization and teaching feedback system based on motion capture model. Based on multi-view video acquisition, two-dimensional pose estimation and three-dimensional skeleton recovery, this study further combines the features of joint angles, endpoint trajectories, rhythm consistency and center of gravity stability, and constructs a framework for action recognition, deviation analysis and feedback output for Chinese dance training scenes. On this basis, the system integrates action template matching, timing alignment and feedback semantic mapping, so that action data can enter the teaching layer from the recognition layer, forming a relatively complete technical closed loop. The experimental results show that the proposed method achieves 94.82%recognition accuracy and 94.17%Macro-F1 on the self-built Chinese dance movement dataset, which significantly reduces the joint Angle error, trajectory deviation, rhythm deviation and center of gravity stability deviation in the movement optimization experiment. In the teaching application experiment, the overall feedback trigger accuracy of the system is more than 85%, the teacher adoption rate remains at a high level, and the average response time of the automatic feedback mode is significantly shorter than that of the traditional manual comment mode after watching the video. The above results show that the motion capture and spatio-temporal modeling technology can not only improve the accuracy of Chinese dance motion recognition, but also provide more direct data support for classroom correction, stage training and teaching decision-making. The significance of this paper is that the original Chinese dance teaching process, which is mainly based on subjective observation, is transformed into an analysis process with both computational expression and teaching interpretability, so that the movement optimization no longer stops at general judgment, but can be implemented to specific parts, specific stages and specific adjustment directions. At the same time, there is still room for further improvement in the current system under the conditions of subtle motion capture, complex combined motion processing and multi-person simultaneous training. The follow-up research can continue to expand the Chinese dance special data set, optimize the fine-grained posture recovery and feedback generation strategy, and strengthen the linkage between the system and the digital curriculum platform, so as to promote the development of Chinese dance teaching to a more refined, continuous and intelligent direction.

# Funding

Fund Project: Research on Key Technologies of Digital Asset Modeling and AI Creation for Intangible Cultural Heritage Huagu Deng Dance, supported by Anhui Provincial University Scientific Research Project (Natural Science), Project No. 2025AHGXZK31190.

# References

[1] Mu J. Pose Estimation-Assisted Dance Tracking System Based on Convolutional Neural Network[J]. Computational Intelligence and Neuroscience, 2022, 2022(1): 2301395.

[2] Li H, Huang X. Intelligent dance motion evaluation: an evaluation method based on keyframe acquisition according to musical beat features[J]. Sensors, 2024, 24(19): 6278.

[3] Li M. Ethnic dance movement recognition based on motion capture sensor and machine learning[J]. International Journal of Information and Communication Technology, 2024, 25(8): 81-96.

[4] Li N, Boers S. Human motion recognition in dance video images based on attitude estimation[J]. Wireless Communications and Mobile Computing, 2023, 2023(1): 4687465.

[5] Wang Y, Wu Z. Dance motion detection algorithm based on computer vision[J]. International Journal of Advanced Computer Science and Applications, 2023, 14(10).

[6] Treffer A, Clark A, Lukosch S. Teaching dance with mixed reality mirrors[C]//2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). IEEE, 2024: 971-980.

[7] Zhou Q, Grebel L, Irlitti A, et al. Here and now: Creating improvisational dance movements with a mixed reality mirror[C]//Proceedings of the 2023 CHI conference on human factors in computing systems. 2023: 1-16.

[8] Liu L, Dai Y, Liu Z. Real-time pose estimation and motion tracking for motion performance using deep learning models[J]. Journal of Intelligent Systems, 2024, 33(1): 20230288.

[9] Tharatipyakul A, Srikaewsiew T, Pongnumkul S. Deep learning-based human body pose estimation in providing feedback for physical movement: A review[J]. Heliyon, 2024, 10(17).

[10] Roggio F, Trovato B, Sortino M, et al. A comprehensive analysis of the machine learning pose estimation models used in human movement and posture analyses: A narrative review[J]. Heliyon, 2024, 10(21).

[11] Dibenedetto G, Polignano M, Lops P, et al. Human pose estimation for explainable corrective feedbacks in office spaces[C]//Adjunct Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization. 2024: 264-275.

[12] Ye Y, Wang J, He P, et al. An action analysis algorithm for teachers based on human pose estimation[J]. Computers and Electrical Engineering, 2023, 111: 108915.

[13] Hamilton R I, Glavcheva-Laleva Z, Milon M I H, et al. Comparison of computational pose estimation models for joint angles with 3D motion capture[J]. Journal of Bodywork and Movement Therapies, 2024, 40: 315-319.

[14] Choi J Y, Ha E, Son M, et al. Human joint angle estimation using deep learning-based three-dimensional human pose estimation for application in a real environment[J]. Sensors, 2024, 24(12): 3823.

[15] Dang Y, Yin J, Zhang S, et al. Kinematics modeling network for video-based human pose estimation[J]. Pattern Recognition, 2024, 150: 110287.

[16] Zheng C, Wu W, Chen C, et al. Deep learning-based human pose estimation: A survey[J]. ACM computing surveys, 2023, 56(1): 1-37.

[17] Lu P, Jiang T, Li Y, et al. Rtmo: Towards high-performance one-stage real-time multi-person pose estimation[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2024: 1491-1500.

[18] Zhu W, Ma X, Liu Z, et al. Motionbert: A unified perspective on learning human motion representations[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2023: 15085-15099.

[19] Zhao Q, Zheng C, Liu M, et al. Poseformerv2: Exploring frequency domain for efficient and robust 3d human pose estimation[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023: 8877-8886.

[20] Xu J, Rao Y, Yu X, et al. Finediving: A fine-grained dataset for procedure-aware action quality assessment[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 2949-2958.

[21] Xu J, Yin S, Zhao G, et al. Fineparser: A fine-grained spatio-temporal action parser for human-centric action quality assessment[C]//Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition. 2024: 14628-14637.

[22] Dong L, Wang W, Qiao Y, et al. Lucidaction: A hierarchical and multi-model dataset for comprehensive action quality assessment[J]. Advances in neural information processing systems, 2024, 37: 96468-96482.