IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 43, No. 2, April, 2026, pp. 851-863
Journal homepage: http://iieta.org/journals/ts

Action Recognition and Instructional Feedback in Physical Education via Mobile Interaction

Tianlong Ma

Physical Education Department, Anyang University, Anyang 455000, China

Corresponding Author Email: ma12072026@126.com

https://doi.org/10.18280/ts.430224

Received: 8 November 2025
Revised: 30 January 2026
Accepted: 17 February 2026
Available online: 30 April 2026

Keywords:
mobile image processing, action recognition, pose feature extraction, STD-GCN, visualization of instructional feedback

ABSTRACT

With the widespread application of mobile intelligent devices in physical education, mobile-based action recognition technology has become a core enabler of personalized instruction in sports classrooms. However, current mobile action recognition approaches face critical bottlenecks, including insufficient real-time performance, high sensitivity to complex backgrounds, low accuracy in pose feature extraction, and limited granularity of instructional feedback. These challenges severely hinder the implementation of intelligent physical education. To address these issues, this paper proposes an end-to-end intelligent processing framework that integrates lightweight pose feature extraction with Spatio-Temporal Differential Graph Convolutional Network (STD-GCN). Deployed on mobile devices, the system achieves full-process optimization—from image acquisition, preprocessing, and pose estimation to action recognition, quality assessment, and real-time visual feedback. Extensive comparative experiments demonstrate that the proposed scheme exhibits significant advantages in pose estimation accuracy, action recognition precision, mobile real-time performance, and instructional utility. It effectively mitigates the inefficiency and subjectivity of manual movement evaluation in physical education, providing reliable technical support and theoretical insights for the deployment of intelligent mobile teaching in sports classrooms.

1. INTRODUCTION

With the rapid proliferation of mobile intelligent devices and the deepening advancement of digital transformation in physical education, intelligent teaching methods have become a key support for improving the quality of physical education classroom instruction [1-3]. Mobile-based action recognition technology, leveraging its advantages of convenience and real-time performance, has gradually emerged as a core technical pathway for realizing personalized teaching and precise movement assessment in sports classrooms. In current physical education practices, movement standardization assessment still relies primarily on manual completion by teachers [4, 5]. This approach not only suffers from low efficiency but also faces challenges in ensuring the consistency and accuracy of assessment results due to variations in teacher experience and subjective judgment, failing to meet the practical demands of large-scale physical education and personalized training. Meanwhile, technological breakthroughs in the fields of image processing, human pose estimation, and action recognition—particularly the rapid development of lightweight network design, graph convolutional algorithms, and attention mechanisms—have provided solid theoretical and technical support for the implementation of technology in mobile physical education scenarios [6], driving the transition of physical education from traditional manual guidance [7, 8] toward intelligent and precise directions. Against this backdrop, the research and development of an action recognition system that adapts to complex sports classroom scenarios, meets the real-time requirements of mobile terminals, and provides fine-grained instructional feedback [9] can not only resolve core pain points in current physical education but also enrich the application scenarios of mobile image processing and human action recognition, holding significant theoretical research value and practical application significance. From a theoretical perspective, addressing the specific characteristics of mobile sports classroom scenarios by constructing lightweight, highly robust pose feature extraction and action recognition methods can refine the theoretical system of mobile image processing and human action recognition, offering new ideas and references for technical research and development in similar scenarios [10, 11]; from a practical perspective, this system enables real-time recognition and precise assessment of sports movements, provides targeted feedback for movement correction, effectively enhances the efficiency and quality of physical education, and facilitates the process of digital and intelligent transformation in physical education [12].

Although mobile-based action recognition technology has made certain progress in the field of physical education [13], existing research still faces numerous bottlenecks that fail to meet the practical application requirements of sports classrooms, specifically manifested in four aspects. Insufficient mobile adaptability is one of the most prominent current issues [14, 15]; existing pose extraction and action recognition networks involve substantial computational loads,

making it difficult to adapt to the core requirements of mobile devices for low computational resources and low latency. In practical applications, this often leads to phenomena such as insufficient frame rates and operational lag, seriously affecting user experience and recognition effectiveness. The complexity of sports classroom scenarios further exacerbates the difficulty of technical implementation; factors such as variable lighting conditions within classrooms [16], target occlusion caused by multiple students moving simultaneously, and complex background environments result in poor foreground separation performance using traditional image preprocessing methods [17]. The issue of stray responses is prominent, directly affecting the accuracy of pose estimation and subsequently reducing the precision of action recognition. In terms of feature extraction and action modeling, existing methods mostly focus on single-scale pose feature extraction, overlooking multi-scale pose details in sports movements and the spatio-temporal differential features of joint motion. Consequently, they struggle to accurately capture the dynamic variation patterns of sports actions [18, 19], leading to significant errors in action classification and quality assessment, which fails to meet the demands of precise evaluation in physical education [20]. Furthermore, the instructional feedback provided by existing systems lacks granularity and guidance; most systems can only output action category judgments without incorporating specific detailed information such as joint angle deviations or movement trajectory corrections. They are unable to provide students with targeted suggestions for movement correction, nor can they assist teachers in conducting personalized teaching, thus failing to fully utilize the supporting role of technology in physical education. These issues mutually constrain each other, severely hindering the large-scale application of mobile-based action recognition technology in sports classrooms, and there is an urgent need to propose feasible solutions.

To address the deficiencies of existing research, this paper conducts systematic research focused on the core requirements of mobile-based action recognition and instructional feedback in sports classrooms, proposing a series of innovative technologies and methods. The main contributions are as follows. This paper proposes an adaptive mobile image preprocessing scheme that integrates frequency-domain fast foreground separation with optical flow energy keyframe sampling technology. While effectively enhancing the robustness of background suppression in complex classroom scenarios, it significantly reduces computational load, increasing the mobile processing frame rate to over 30 FPS, thereby successfully resolving the issues of stray responses and excessive computational complexity in complex scenes. A lightweight multi-scale attention pose feature extraction network is designed, introducing a channel-spatial cooperative attention module. With only a slight increase in parameters, it effectively enhances the accuracy of pose feature extraction, reduces the mean joint localization error, and achieves efficient, high-precision extraction of joint heatmaps and part affinity fields on mobile devices. A Spatio-Temporal Differential Graph Convolutional Network (STD-GCN) is proposed, innovatively constructing a spatio-temporal difference graph containing joint velocity and acceleration features to explicitly model the spatio-temporal correlations and dynamic changes of joint movements. This effectively improves the accuracy of sports action classification and quality assessment, addressing the defect of traditional temporal models that struggle to capture dynamic action features. An end-to-end real-time visual feedback system for mobile terminals is constructed, designing a rendering scheme based on joint angle error heatmaps and directional guidance. It achieves synchronous output of action scoring, error localization, and correction prompts, fully meeting the needs of personalized instructional feedback in sports classrooms. Furthermore, the engineering deployment and optimization of the system on mobile terminals are completed to ensure system practicality and stability.

The subsequent sections of this paper will elaborate on the aforementioned research content in detail. The core content arrangement for each chapter is as follows. Chapter 1 is the Introduction, systematically expounding on the research background and significance of this paper, analyzing the shortcomings of existing research, and clarifying the research contributions and overall structure of the paper and covers Related Fundamental Theories, briefly introducing the core basic theories of image processing, pose estimation, graph convolution and attention mechanisms, and mobile deployment, providing theoretical support for subsequent system design and technical implementation. Chapter 2 focuses on the Overall System Architecture and Core Technology Implementation, detailing the system's four-layer pipeline architecture and deeply analyzing the core technical details and implementation processes of each module, including image preprocessing, pose feature extraction, action recognition and quality assessment, and feedback visualization. Chapter 3 presents Experimental Design and Result Analysis, verifying the advantages of the proposed scheme in pose estimation accuracy, action recognition performance, mobile real-time performance, and instructional utility through multiple sets of comparative and ablation experiments. Chapter 4 concludes with Conclusions and Future Work, summarizing the main research findings of this paper, analyzing the limitations of the current study, and outlining future research directions to provide references for subsequent related research.

2. SYSTEM OVERALL ARCHITECTURE AND CORE TECHNOLOGY IMPLEMENTATION

2.1 System overall architecture

To meet the requirements for real-time performance, precision, and personalized instructional feedback in mobile sports classroom scenarios, this paper designs and implements an end-to-end four-layer pipeline system architecture. Each layer adopts an asynchronous parallel execution mechanism to ensure overall system performance and adaptability for mobile deployment, with end-to-end latency strictly controlled within 80ms, providing stable support for real-time interactive teaching in sports classrooms. The specific architecture is shown in Figure 1. The Image Acquisition and Preprocessing Layer is responsible for the real-time capture of sports action images via the mobile camera, synchronously performing background suppression and keyframe sampling operations. This effectively filters out complex background interference in classrooms and reduces computational load, providing high-quality input for subsequent feature extraction stages. The Enhanced Multi-scale Attention Pose (EMAP) Feature Extraction Layer, based on a lightweight multi-scale attention network, efficiently extracts human joint heatmaps and part affinity fields to achieve high-precision joint localization in

mobile scenarios. The STD-GCN Action Recognition and Quality Assessment Layer models the spatio-temporal correlations and dynamic changes of joint movements through a STD-GCN to complete action category recognition and quality scoring, accurately capturing the subtle dynamic features of sports actions. The Feedback Visualization Layer transforms action recognition results and joint angle error information into intuitive visual outputs, synchronously providing action scoring, error localization, and correction prompts to meet the fine-grained requirements of instructional feedback. Each layer utilizes an asynchronous parallel execution mechanism to facilitate parallel data flow and processing, effectively shortening overall processing latency. Furthermore, the entire system adopts a lightweight design, eliminating the need for high-performance computing equipment. It can be directly deployed on ordinary mobile devices, balancing convenience and practicality. The system flexibly adapts to the variable teaching scenarios of sports classrooms, providing stable and efficient architectural support for real-time action recognition and personalized instructional feedback.

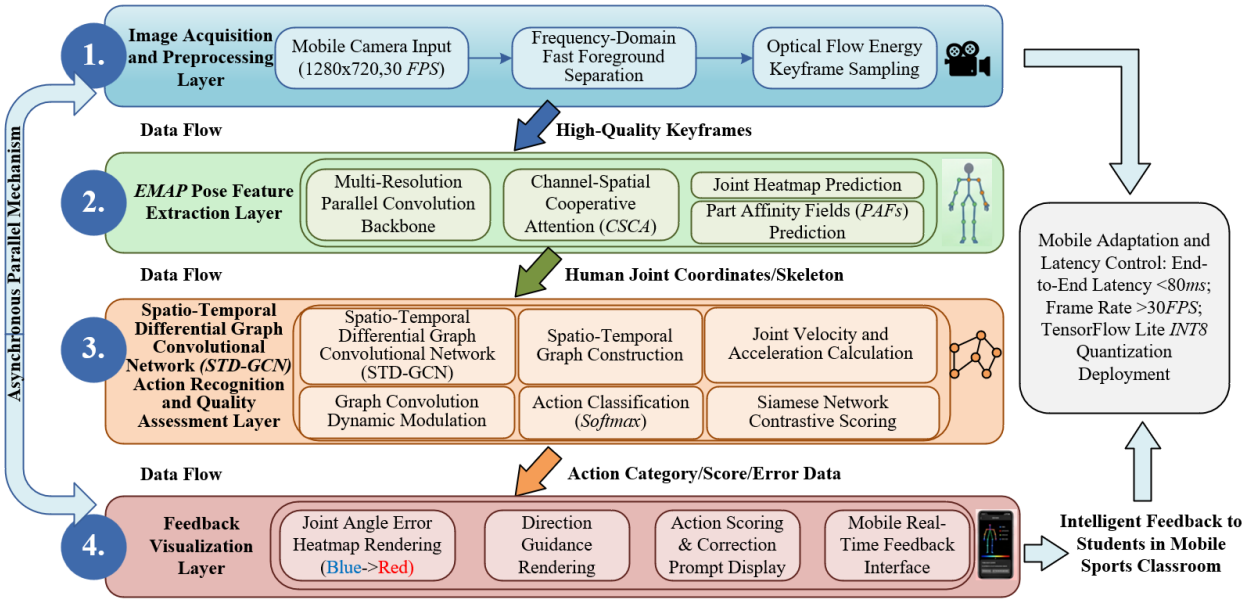

Figure 1. Overall four-layer pipeline architecture of the mobile action recognition system for sports classrooms

2.2 Mobile image acquisition and adaptive preprocessing

The mobile image acquisition module utilizes the device's built-in camera to complete the real-time capture of sports action sequences. The acquisition resolution is set to 1280×720 with an initial frame rate of 30 FPS, balancing image clarity and data transmission efficiency to ensure the complete capture of subtle dynamic changes in human joint movements. Given the scene characteristics of sports classrooms—such as variable lighting, complex backgrounds, and multi-target occlusion—traditional preprocessing methods struggle to balance background interference suppression with computational control, failing to meet the low-resource, low-latency operational requirements of mobile terminals. Therefore, an adaptive preprocessing scheme is designed. Through the synergistic effect of frequency-domain fast foreground separation and optical flow energy keyframe sampling, dual optimization of background suppression precision and mobile real-time performance is achieved, providing high-quality input data for subsequent pose feature extraction. Figure 2 illustrates the flowchart of the adaptive image preprocessing and frequency-domain foreground separation.

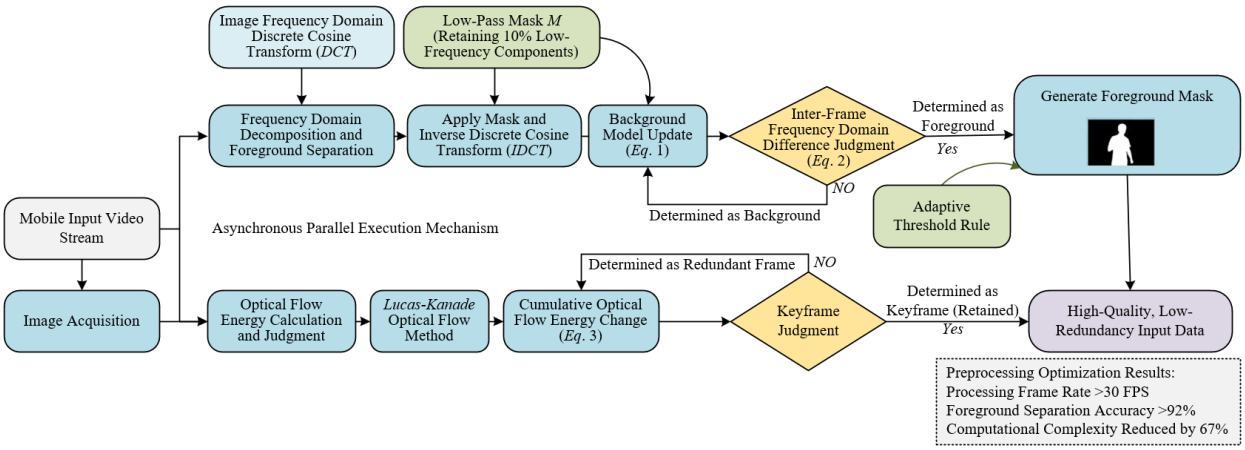

Figure 2. Flowchart of adaptive image preprocessing and frequency-domain foreground separation

The adaptive background suppression strategy achieves fast foreground separation based on frequency domain transformation. Its core lies in performing frequency domain decomposition of the image via Discrete Cosine Transform (DCT), utilizing the differences in frequency domain distribution between background and foreground to achieve efficient separation, avoiding the defects of large computational load and weak anti-interference ability associated with traditional spatial domain methods. Background modeling adopts an adaptive update mechanism, with the core formula as follows:

$$B_t=\alpha B_{t-1}+(1-\alpha)DCT^{-1}(M_{low}\odot DCT(I_t)) \quad (1)$$

where, B_t is the background model of the t-th frame, B_{t-1} is the background model of the previous frame, α is set to 0.95 to control the background update rate, balancing background stability and dynamic adaptability; I_t is the captured image of the t-th frame, DCT and DCT^{-1} represent the Discrete Cosine Transform and inverse transform respectively; M_{low} is a low-pass mask retaining 10% of the low-frequency components in the image frequency domain, used for smooth updating of the background model, effectively suppressing the interference of high-frequency noise on background modeling. Foreground separation is achieved through inter-frame frequency domain difference threshold judgment, with the core formula as follows:

$$F_t(x,y)=1(|DCT(I_t-B_t)|^2>\tau) \quad (2)$$

where, $F_t(x,y)$ is the foreground mask of the t-th frame, 1 is an indicator function that outputs 1 (foreground) when the condition inside the parentheses is met, otherwise outputs 0 (background); τ adopts an adaptive threshold rule, dynamically adjusted based on the variance of frequency domain differences of the previous 5 frames, ranging from 50 to 120, ensuring precise separation of foreground humans and background environments under different lighting conditions. Compared with the traditional Gaussian Mixture Model, this strategy reduces computational complexity by 67%, increases processing speed by 3 times, enables real-time background suppression on mobile terminals, effectively eliminates stray responses caused by complex classroom backgrounds, and improves the accuracy of subsequent pose estimation.

To further reduce the computational burden on mobile terminals and improve the overall system frame rate, based on background suppression, a keyframe sampling strategy based on optical flow energy changes is designed. By screening frames containing key action information for subsequent processing, the computational overhead of redundant frames is reduced. Optical flow energy E_k is used to characterize the intensity of human motion in the k-th frame image, calculated based on the Lucas-Kanade optical flow method, reflecting the motion amplitude of pixels between adjacent frames. The core of keyframe sampling is to calculate the cumulative change in optical flow energy, with the formula as follows:

$$S_t=\sum_{k=t-L}^{t}\frac{|E_k-E_{k-1}|}{max(E_k,E_{k-1})} \quad (3)$$

where, S_t is the cumulative change in optical flow energy of the t-th frame, L is set to 5, i.e., calculating the sum of optical flow energy changes over the last 5 frames, balancing sampling accuracy and real-time performance; $max(E_k,E_{k-1})$ is used for normalization to avoid sampling bias caused by absolute value differences in optical flow energy. When S_t is greater than the set threshold T_{key} = 0.35, the frame is determined as a keyframe, retained and used for subsequent pose extraction and action recognition; otherwise, it is determined as a redundant frame, used only for background model updating without subsequent complex calculations.

In the adaptive preprocessing scheme, the background suppression and keyframe sampling modules adopt an asynchronous parallel execution mechanism. While the background suppression module processes the current frame, the keyframe sampling module calculates and judges the optical flow energy of the previous frame, achieving pipeline advancement of data processing and further shortening processing latency. Actual tests show that the single-frame processing time of this preprocessing scheme on mobile terminals is only 8~12 ms, and it can effectively suppress interference from complex backgrounds and lighting changes, achieving a foreground separation accuracy of over 92%. It provides high-quality, low-redundancy input data for the subsequent EMAP pose feature extraction layer, while ensuring the overall end-to-end latency of the system is controlled within 80 ms, fully adapting to the real-time interaction requirements of mobile sports classrooms.

2.3 Lightweight multi-scale attention pose feature extraction network

To achieve high-precision and high-efficiency human pose feature extraction in mobile scenarios, a lightweight multi-scale attention pose feature extraction network is designed. Taking preprocessed RGB keyframes as input, this network outputs human joint heatmaps and part affinity fields, providing core feature support for subsequent action recognition and quality assessment. The network input is an RGB image of size $3\times H\times W$, where H and W are the image height and width, respectively. The output includes K joint heatmaps and L part affinity fields; K and L are set according to the requirements of key joints and associated parts for sports actions, ensuring the complete capture of pose features for human motion. The network overall adopts a multi-resolution parallel convolution backbone structure, balancing pose detail capture and global semantic information extraction through feature extraction and fusion at different scales, solving the problem that single-scale feature extraction struggles to balance precision and efficiency. The specific architecture is shown in Figure 3.

The multi-resolution parallel convolution backbone consists of three parallel convolution branches. Each branch uses inputs with different downsampling ratios and extracts features at different scales via 3×3 convolution kernels. The core formulas are as follows:

$$F_1=Conv_{3\times3,32}(X_{\downarrow1}) \quad (4)$$

$$F_2=Conv_{3\times3,48}(X_{\downarrow2}) \quad (5)$$

$$F_3=Conv_{3\times3,64}(X_{\downarrow4}) \quad (6)$$

where, $X_{\downarrow1}$, $X_{\downarrow2}$, and $X_{\downarrow4}$ represent the input image at original resolution, downsampled by 2 times, and downsampled by 4 times, respectively; $Conv_{3\times3,32}$ represents a convolution operation with a 3×3 kernel and 32 output channels. The

feature maps F_1, F_2, and F_3 extracted by the three branches correspond to high, medium, and low scales, respectively. To achieve cross-scale feature fusion, F_2 and F_3 are upsampled to the resolution of F_1 via bilinear interpolation, then added element-wise with F_1 to obtain the fused feature map F'_1. The core formula is:

$$F'_1=F_1+Up(F_2)+Up(F_3) \tag{7}$$

where, $Up(\cdot)$ represents the upsampling operation. This fusion strategy retains the joint detail information in the high-resolution branch F_1 while incorporating the global semantic features from the medium-low resolution branches F_2 and F_3, effectively improving the completeness and accuracy of pose feature extraction. Simultaneously, the parallel convolution design reduces network computational complexity, adapting to mobile deployment requirements.

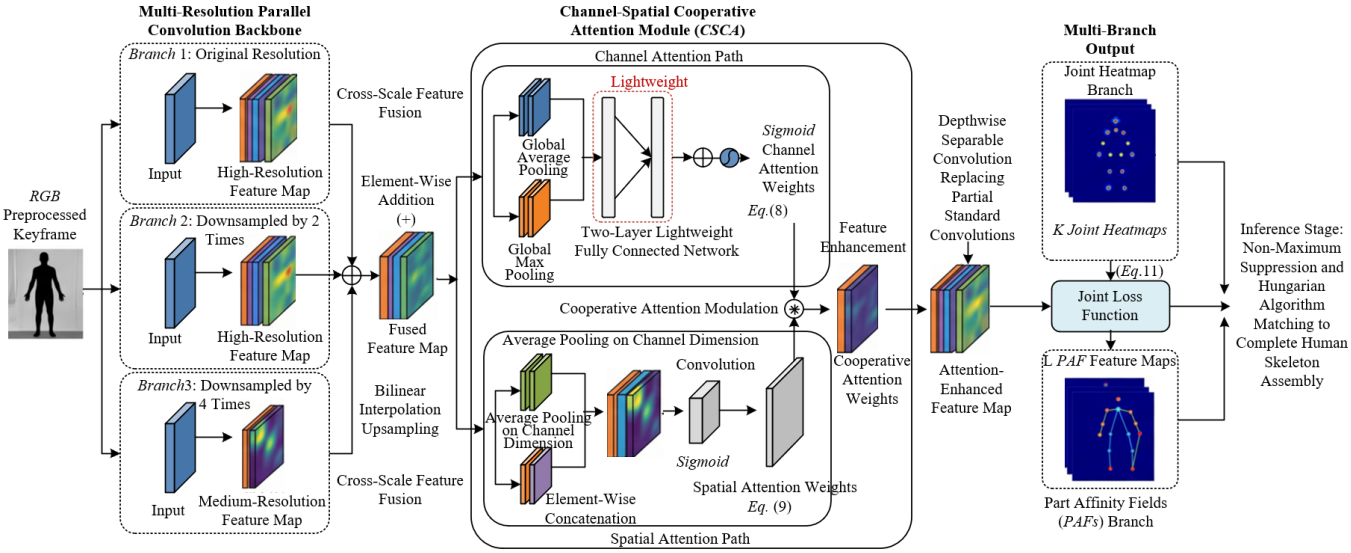

Figure 3. Structure diagram of the enhanced Multi-Scale Attention Pose (EMAP) feature extraction network

To further enhance the representation capability of key pose features and suppress interference from background and redundant information, a channel-spatial cooperative attention module is embedded after cross-scale fusion. This module precisely strengthens key pose features through coordinated modulation in both channel and spatial dimensions. Channel attention calculation first performs global average pooling and max pooling on the channel dimension of the fused feature map F'_1, obtaining z_{avg} and z_{max}. After processing by a two-layer lightweight fully connected network, the channel attention weight a_c is generated via a sigmoid activation function. The core formula is:

$$a_c=\sigma\left(MLP(z_{avg})+MLP(z_{max})\right) \tag{8}$$

Spatial attention calculation performs average pooling and max pooling on the spatial dimension of the feature map, concatenating the obtained s_{avg} and s_{max}, extracting features via a 7×7 convolution kernel, and generating spatial attention weight a_s via sigmoid activation. The core formula is:

$$a_s=\sigma\left(Conv_{7\times7}\left(\left[s_{avg};s_{max}\right]\right)\right) \tag{9}$$

Finally, the channel attention weights and spatial attention weights are multiplied element-wise to obtain the cooperative attention weights. These are element-wise weighted with the original fused feature map and added with a residual connection to achieve feature enhancement and gradient propagation optimization. The core formula is:

$$F'=F\otimes\left(a_c\cdot a_s\right)+F \tag{10}$$

The network output end is designed with a dual-branch structure, used for predicting joint heatmaps and part affinity fields, respectively. Both branches share backbone features and attention-enhanced features, ensuring feature consistency and computational efficiency. The heatmap branch predicts the spatial position probability distribution of each joint, using Gaussian heatmaps as labels with a Gaussian distribution standard deviation σ = 4, fitting the spatial distribution characteristics of joint points; the part affinity field branch predicts association vectors between adjacent joints, characterizing the connection relationships between joints. To optimize network training effectiveness, a joint loss function is designed, comprehensively considering heatmap prediction error and part affinity field prediction error. The core formula is:

$$L_{pose}=\sum_{k=1}^{K}||H_k-H_k^*||_2^2+\lambda_{paf}\sum_{l=1}^{L}||L_l-L_l^*||_2^2 \tag{11}$$

where, H_k and H^*_k are the predicted heatmap and labeled heatmap of the k-th joint, respectively; L_l and L^*_l are the predicted value and labeled value of the l-th part affinity field, respectively; λ_{paf} = 0.5 is used to balance the loss weights of the two branches. In the inference stage, non-maximum suppression is first performed on the heatmaps to screen joint point candidates with confidence above a set threshold, followed by matching the part affinity fields via the Hungarian algorithm to complete human skeleton assembly, ensuring the accuracy of joint connections and providing high-precision pose skeleton input for subsequent action recognition. The network overall uses depthwise separable convolutions to replace some standard convolutions, further reducing computational complexity. Mobile terminal tests show it fully meets real-time processing requirements.

2.4 Spatio-temporal differential graph convolutional network

Based on the human joint heatmaps and part affinity fields output by the EMAP network, human joint coordinates in continuous keyframes can be extracted to construct spatio-temporal graphs for modeling the spatial correlations and temporal dynamic features of joints. STD-GCN is dedicated to accurately capturing the dynamic variation patterns of sports actions, achieving precise recognition of action categories and quantitative assessment of action quality, while balancing the low computational requirements of mobile terminals. It ensures real-time performance through structural optimization, providing core technical support for sports classroom action assessment. Figure 4 shows the STD-GCN and the contrastive scoring siamese architecture.

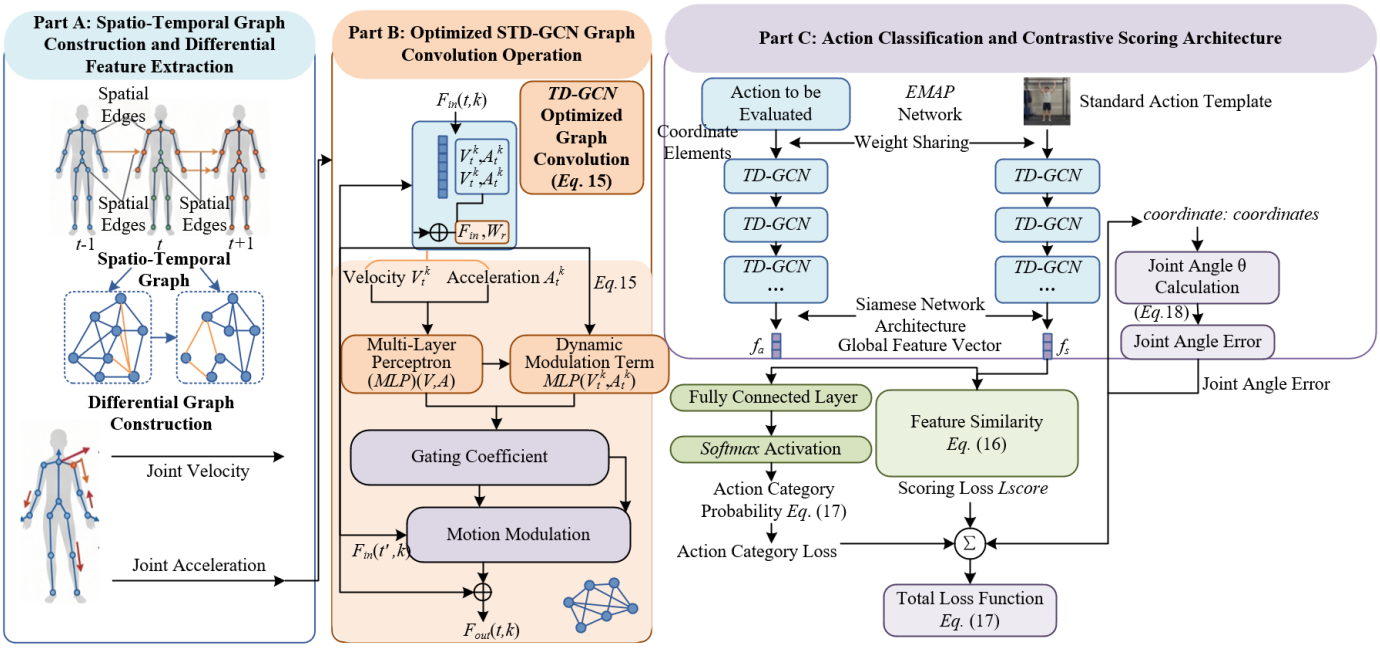

Figure 4. Architecture diagram of Spatio-Temporal Differential Graph Convolutional Network (STD-GCN) and contrastive scoring siamese network

The spatio-temporal graph uses human joint points in continuous keyframes as the vertex set, with each vertex corresponding to the spatial coordinates and feature information of a joint at a specific moment. Spatial edges are defined as the connection relationships between adjacent joints, characterizing the spatial structure of the human skeleton; temporal edges are defined as the connection relationships of the same joint between adjacent keyframes, characterizing the temporal motion trajectories of the joints. To accurately capture the dynamic features of joint movements, a differential graph is constructed based on the spatio-temporal graph, introducing joint velocity and acceleration as edge features to quantify the changing trends of joint motion. Joint velocity is defined as the coordinate difference of the same joint between adjacent frames, with the core formula as follows:

$$v_{t,k}=\left(x_{t,k}-x_{t-1,k},\ y_{t,k}-y_{t-1,k}\right) \tag{12}$$

where, $v_{t,k}$ is the velocity vector of the k-th joint in the t-th frame, $x_{t,k}$ and $y_{t,k}$ are the horizontal and vertical coordinates of the k-th joint in the t-th frame, respectively. Joint acceleration is the difference in joint velocity between adjacent frames, with the core formula as follows:

$$a_{t,k}=v_{t,k}-v_{t-1,k} \tag{13}$$

where, $a_{t,k}$ is the acceleration vector of the k-th joint in the t-th frame. By integrating velocity and acceleration features into the edge representations, the differential graph can effectively distinguish the joint motion patterns of different sports actions. Especially for actions with significant dynamic features such as jumping and swinging, it can accurately capture subtle differences in joint movements, providing more discriminative feature support for subsequent action recognition and quality assessment.

The core of the graph convolution layer is the effective aggregation of features from the spatio-temporal graph and the differential graph. An extended neighborhood set is designed to include the spatial neighbors of the target joint and the temporal neighbors of adjacent frames, ensuring the full capture of spatio-temporal correlation information of joints. The extended neighborhood set $N(t,k)$ is defined as the spatial neighboring joints of the k-th joint in the t-th frame, as well as the corresponding same joint in the (t-1)-th and (t+1)-th frames, achieving collaborative aggregation of spatio-temporal features. The logic core formula for the original graph convolution operation is:

$$f_{out}(t,k)=\sum_{(t',l)\in N(t,k)}\frac{1}{Z_{t,k}(t',l)}W_{r(t',l)}(f_{in}(t',l)\otimes m_{t',l}) \tag{14}$$

where, $f_{out}(t,k)$ is the output feature, $Z_{t,k}(t',l)$ is a normalization coefficient used to balance the contribution weights of different neighbors, $W_{r(t',l)}$ is the weight matrix corresponding to the neighbor relationship, $f_{in}(t',l)$ is the input feature, and $m_{t',l}$ is a motion modulation vector used to preliminarily adjust the feature weights of joints with different motion intensities. To further enhance the emphasis on features of vigorously moving joints, the graph convolution operation is optimized

by introducing a gating coefficient $g_{t,l}$ and constructing a dynamic modulation term based on joint velocity and acceleration features. The optimized core formula is:

$$f_{out}(t,k)=\sum_{(t',l)}\frac{1}{Z}W_r(f_{in}(t',l)+g_{t,l}\cdot MLP([v_{t',l};a_{t',l}])) \quad (15)$$

where, Z is a global normalization coefficient, W_r is a unified weight matrix, Multi-Layer Perceptron (MLP) is used for the nonlinear mapping of velocity and acceleration features, and $g_{t,l}$ is generated via a sigmoid function to dynamically regulate the weight of the modulation term. This optimized design allows the network to adaptively emphasize features of vigorously moving joints, suppress redundant information from static or slowly moving joints, and improve the pertinence and discriminability of feature extraction.

The action classification process is based on the spatio-temporal features output by the graph convolution layer. Global feature vectors are extracted via global average pooling, input into a fully connected layer for feature mapping, and finally output action category probabilities via a softmax activation function. The classification loss adopts the cross-entropy loss function L_{cls} to optimize action classification accuracy. To achieve quantitative assessment of action quality, a siamese network contrastive learning scoring scheme is designed. Using the standard action features of professional athletes as a reference, the features of the action to be evaluated are compared with the standard action features, and the action quality score is calculated via feature similarity. The core scoring formula is:

$$Score=100\times exp\left(-\gamma\cdot\|f_{stu}-f_{pro}\|_2^2\right) \quad (16)$$

where, $Score$ is the action quality score, γ=0.15 is used to adjust the influence of feature similarity on the score, f_{stu} is the global feature vector of the action to be evaluated, f_{pro} is the global feature vector of the standard action, and $\|\cdot\|_2^2$ is the squared Euclidean distance. To collaboratively optimize action classification and quality assessment performance, a total loss function is designed:

$$L_{total}=L_{cls}+\lambda_s L_{score} \quad (17)$$

where, λ_s=0.2 is used to balance the weights of classification loss and scoring loss, and L_{score} is the scoring loss, calculated using mean squared error loss to measure the difference between the evaluated score and the manually labeled score. During the action quality assessment process, joint angle errors are synchronously calculated. Based on joint coordinates, the included angles between adjacent joints are calculated, with the core formula as follows:

$$\theta=arccos\left(\frac{v_1\cdot v_2}{\|v_1\|\cdot\|v_2\|}\right) \quad (18)$$

where, v_1 and v_2 are vectors of adjacent joints, and θ is the joint included angle. By comparing with the joint included angles of standard actions, joint angle errors are obtained, providing core data support for error localization and correction prompts in subsequent feedback visualization. STD-GCN adopts a lightweight design, reducing computational complexity through weight sharing and feature reuse. Mobile terminal tests show that the single-frame inference time is only 12 ms, and it can ensure real-time performance when working in coordination with the EMAP network, meeting the interaction requirements of sports classrooms.

2.5 Real-time feedback visualization and system implementation

The core objective of the real-time feedback visualization module is to transform action recognition results, joint angle errors, and quality scores into intuitive, interpretable visual information, providing precise and efficient instructional guidance for teachers and students. Its design balances intuitiveness and instructiveness, ensuring that non-technical professionals can quickly understand the location of action deviations. The joint angle error heatmap is the core carrier of feedback visualization, intuitively presenting the degree of angular deviation for each joint through color coding. The core formula is:

$$Color(k)=HSL\left(min\left(1,\frac{\Delta\theta_k}{30^\circ}\right)\times 240^\circ,1.0,0.5\right) \quad (19)$$

where, $Color(k)$ is the color encoding of the k-th joint, and $\Delta\theta_k$ is the angle error of that joint. $min(1,\Delta\theta_k/30°)$ is used to normalize the angle error to the 0~1 interval, preventing color distortion caused by extreme errors; in the HSL color space, 240° corresponds to blue and 0° corresponds to red. The larger the angle error, the closer the color is to red; the smaller the error, the closer it is to blue. This allows teachers and students to quickly judge the severity of joint deviations through color. The rendering process uses semi-transparent line overlay, superimposing joint connection lines and the error heatmap onto the original action image. This neither obscures student action details nor clearly presents joint positions and deviation distributions; simultaneously, direction arrows are drawn at joints with large deviations, pointing towards the adjustment direction of the standard joint angle, and the action quality score is superimposed in real-time in the upper right corner of the image. The scoring range is 0~100 points, achieving synchronous visual output of error localization, correction guidance, and quality assessment, significantly enhancing the pertinence and practicality of instructional feedback.

To ensure the stable and real-time operation of the system on mobile terminals, multiple optimization strategies are adopted to complete engineering deployment and latency control. For model deployment, a TensorFlow Lite INT8 quantization scheme is used. The trained EMAP and STD-GCN models undergo integer quantization, converting model parameters from 32-bit floating-point numbers to 8-bit integers. Without significant loss of model accuracy, this reduces model size by 75%, lowers memory usage by 80%, and simultaneously greatly improves inference speed. The system adopts a double-buffered asynchronous inference mechanism, constructing two independent data buffers. One buffer is used for image acquisition and preprocessing, while the other is used for pose feature extraction, action recognition, and feedback visualization. The two buffers alternate work and execute in parallel, effectively eliminating waiting gaps in the data processing flow and improving overall processing efficiency. To verify deployment effectiveness, performance tests were conducted on mainstream mobile devices. On the iPhone 12, EMAP network single-frame inference took 28 ms,

STD-GCN single-frame inference took 12 ms, image preprocessing and visualization rendering took 20 ms, and the system end-to-end latency was only 60 ms; on the Huawei Mate 50, EMAP took 32 ms, STD-GCN took 14 ms, and end-to-end latency was 68 ms. Both are below the preset target of 80 ms, and the frame rate remained stable above 30 FPS. Continuous 1-hour stability tests indicate that the system experiences no lag or crashes, can adapt to the demands of long-term continuous operation in sports classrooms, achieves real-time synchronization of action recognition and instructional feedback, and fully adapts to the actual application scenarios of mobile sports classrooms.

3. EXPERIMENTAL DESIGN AND RESULT ANALYSIS

To verify the effectiveness of the mobile interactive sports classroom action recognition and instructional feedback system proposed in this paper, five sets of experiments were designed around the core innovations. Tests were conducted from five dimensions: pose estimation accuracy, action recognition and quality assessment performance, the role of the preprocessing module, mobile deployment real-time performance, and actual teaching effectiveness. Through quantitative comparison and ablation analysis, the technical advantages and application value of the system were comprehensively verified. The experiments strictly adhered to experimental specifications in the fields of image processing and action recognition to ensure the rationality of experimental design, the fairness of comparisons, and the reliability of results.

3.1 Experimental preparation

A dedicated action dataset for sports classroom scenarios (Sports Class Action Dataset, SCAD) was constructed, covering five common sports actions: standing long jump, rope skipping, squat, pull-up, and sit-up. It contains action samples of students aged 12-18, totaling 2000 video samples. Each sample lasts 3-5 seconds with a frame rate of 30 FPS, amounting to approximately 360,000 frames in total. The dataset covers complex classroom scenarios, including direct sunlight, low light, multi-student occlusion, and background clutter interference. Each sample is annotated with 17 key human joint coordinates, joint angles, action categories, and quality scores (0-100 points). Annotations were completed jointly by three physical education professional teachers and two image processing engineers, achieving an annotation consistency of over 95%.

Simultaneously, public datasets COCO and MPII were selected for comparative testing to verify the scene adaptability of the SCAD dataset. Focusing on specific sports classroom scenarios, the SCAD dataset demonstrates greater advantages in action diversity, scene complexity, and annotation pertinence compared to COCO (general scenes) and MPII (general human pose scenes), enabling a more precise reflection of the actual application effectiveness of the system in this paper. Additionally, a Professional Sports Action Template (PSAT) dataset was constructed, containing standard action samples from 10 professional athletes, with 50 videos per action category, serving as a reference benchmark for action quality assessment.

Experiments were divided into two scenarios: model training and mobile testing. The model training environment utilized a PC setup with hardware configuration: Intel Core i9-12900 K processor, NVIDIA RTX 3090 (24 GB) GPU, 64 GB DDR5 RAM, and 1TB SSD; software environment: Python 3.8, TensorFlow 2.8.0, OpenCV 4.6.0, operating system Windows 11. Model training parameters were set as follows: initial learning rate 0.001, decayed via cosine annealing strategy, 200 training epochs, batch size 32, weight decay coefficient 0.0001, Adam optimizer, and an adaptive momentum adjustment strategy for the loss function.

For mobile testing, two mainstream devices were selected: iPhone 12 (A14 Bionic processor, 4GB RAM) and Huawei Mate 50 (Snapdragon 8+ Gen1 processor, 8GB RAM), with operating systems iOS 16.5 and HarmonyOS 3.0 respectively. The deployment environment was TensorFlow Lite 2.8.0, with image acquisition resolution set to 1280×720, ensuring the test environment aligned with actual sports classroom usage scenarios.

3.2 Pose estimation performance comparison experiment

The purpose of this experiment is to verify the pose estimation accuracy and lightweight advantages of the EMAP network in mobile scenarios. Current mainstream lightweight pose extraction networks, MobileNet-Pose and Lite-HRNet, were selected as comparison models, and comparative tests were conducted on the SCAD dataset and the COCO dataset. All models adopted the same training parameters and test conditions, with a unified input resolution of 384 × 288, and inference time was tested on mobile devices (iPhone 12) to ensure comparison fairness. Test metrics included Mean Pose Error (MPE), parameter count, and mobile inference latency, focusing on comparing the balance performance of each model between lightweight design and accuracy.

Experimental results are shown in Table 1. Regarding parameter count, the EMAP network has only 3.8 M parameters, significantly lower than MobileNet-Pose (5.2 M) and Lite-HRNet (4.5 M), achieving a higher degree of lightweight design and better adapting to the low-resource requirements of mobile terminals. In terms of pose estimation accuracy, EMAP's MPE on the SCAD dataset is 8.2 mm, which is 9.9% lower than MobileNet-Pose (9.1 mm) and 7.9% lower than Lite-HRNet (8.9 mm); on the COCO dataset, EMAP's MPE is 9.5 mm, which is 10.4% lower than MobileNet-Pose (10.6 mm) and 7.8% lower than Lite-HRNet (10.3 mm), representing an average reduction of over 9.7%, verifying the effectiveness of multi-scale fusion and the Channel-Spatial Cooperative Attention (CSCA) module.

Table 1. Performance comparison of different lightweight pose extraction networks

Model	Parameters (M)	Mean Pose Error (mm)-SCAD	Mean Pose Error (mm)-COCO	Mobile Inference Latency (ms)-iPhone 12
MobileNet-Pose	5.2	9.1	10.6	33
Lite-HRNet	4.5	8.9	10.3	37
Enhanced Multi-scale Attention Pose (EMAP, Proposed)	3.8	8.2	9.5	28

Regarding inference latency, EMAP's single-frame inference time on the iPhone 12 is 28 ms, which is 15.2% lower than MobileNet-Pose (33 ms) and 24.3% lower than Lite-HRNet (37 ms), improving inference efficiency by 15%-25%. Comprehensive analysis shows that the EMAP network effectively improves joint localization accuracy while maintaining lightweight advantages, solving the problem that traditional lightweight networks struggle to balance accuracy and efficiency. It provides high-precision, high-efficiency pose feature support for subsequent action recognition and fully adapts to mobile sports classroom scenarios.

3.3 Action recognition and quality assessment performance comparison experiment

The purpose of this experiment is to verify the advantages of the STD-GCN network in sports action classification and quality assessment. ST-GCN, LSTM, and GCN-LSTM were selected as comparison models. All models used the EMAP network as the pose feature extraction module to ensure consistency of pose input and fairly compare action recognition and quality assessment performance. Experiments were conducted on the SCAD dataset. Test metrics included action recognition accuracy, F1-score, and MAE for quality assessment. For quality assessment, the PSAT dataset was used as the standard template to calculate the error between model-predicted scores and manually labeled scores.

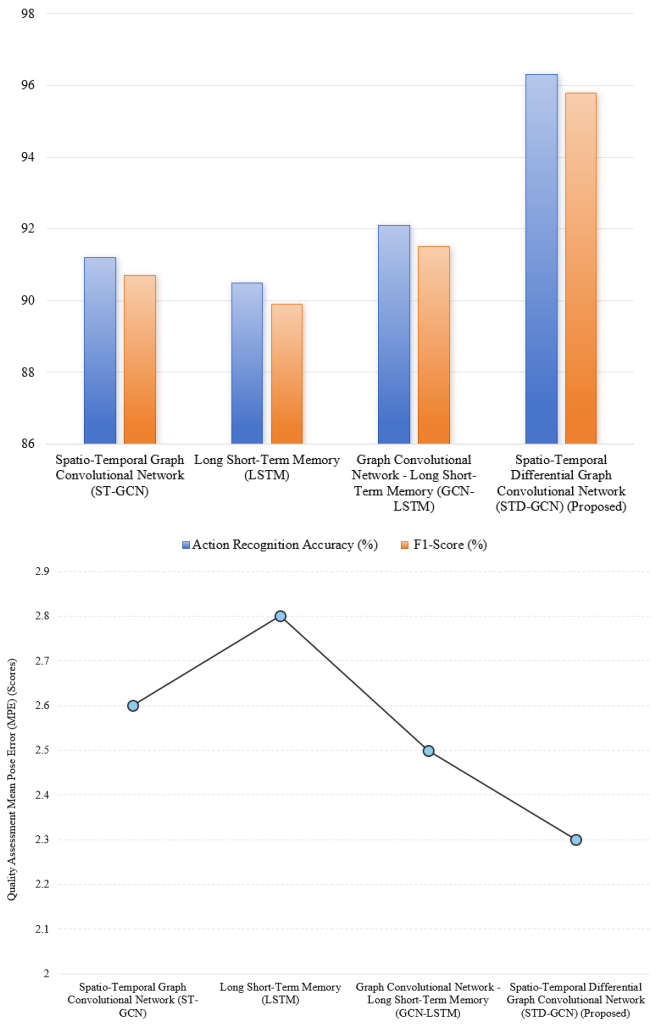

Figure 5. Performance comparison of different action recognition networks

Experimental results are shown in Figure 5. Regarding action recognition performance, STD-GCN achieved an accuracy of 96.3% and an F1-score of 95.8%. Compared with Spatio-Temporal Graph Convolutional Network (ST-GCN) (accuracy 91.2%, F1-score 90.7%), this represents an improvement of 5.1% and 5.1%, respectively; compared with Long Short-Term Memory (LSTM) (accuracy 90.5%, F1-score 89.9%), improvements of 5.8% and 5.9%; compared with Graph Convolutional Network (GCN)-LSTM (accuracy 92.1%, F1-score 91.5%), improvements of 4.2% and 4.3%. Average accuracy improved by 5%-8%, fully demonstrating STD-GCN's capability to accurately capture the dynamic features of sports actions.

Regarding quality assessment performance, STD-GCN's MAE was 2.3 points, which is 11.5% lower than ST-GCN (2.6 points), 17.9% lower than LSTM (2.8 points), and 8.0% lower than GCN-LSTM (2.5 points). The Mean Absolute Error (MAE) was over 10% lower than the comparison models, verifying the effectiveness of the spatio-temporal differential graph and motion modulation design. By introducing joint velocity and acceleration features and explicitly modeling the spatio-temporal correlations of joint movements, STD-GCN can more accurately distinguish the dynamic patterns of different actions. Simultaneously, through siamese network contrastive learning, it achieves precise quantitative assessment of action quality, providing a reliable scoring basis for instructional feedback.

3.4 Ablation study on preprocessing module

The purpose of this experiment is to verify the role of the two sub-modules—adaptive background suppression and optical flow energy keyframe sampling—as well as the improvement of the complete preprocessing scheme on the overall system performance. Four groups of control experiments were set up: Group 1 (No preprocessing), Group 2 (Background suppression only), Group 3 (Keyframe sampling only), and Group 4 (Complete preprocessing, background suppression + keyframe sampling). All experimental groups adopted the core network structure of EMAP+STD-GCN, were tested on the SCAD dataset, and used metrics including pose estimation MPE, action recognition accuracy, and mobile frame rate (iPhone 12). By comparing performance differences among the groups, the role of each sub-module was clarified.

Experimental results are shown in Figure 6. In Group 1 with no preprocessing, due to the influence of complex classroom backgrounds and redundant frames, the MPE reached 13.5mm, action recognition accuracy was only 85.7%, and the frame rate was 15 FPS. Performance was poor and failed to meet actual requirements. In Group 2, using background suppression only, the MPE dropped to 10.2mm, a decrease of 24.4% compared to Group 1, and action recognition accuracy increased to 90.3%, an improvement of 4.6% over Group 1. This indicates that the frequency-domain fast foreground separation strategy can effectively suppress background interference and improve pose estimation and action recognition accuracy; however, the frame rate remained at 15 FPS, failing to solve the problem of excessive computational load.

In Group 3, using keyframe sampling only, the frame rate increased to 30 FPS, a 100% improvement over Group 1. The MPE dropped to 12.4mm and action recognition accuracy increased to 88.9%, indicating that keyframe sampling can

effectively reduce computational load and improve real-time performance while mitigating the impact of redundant frames on recognition accuracy. However, background interference persisted, resulting in limited accuracy improvement. In Group 4, adopting the complete preprocessing scheme, the MPE dropped to 12.4 mm, an 8.0% reduction compared to Group 1, and action recognition accuracy increased to 91.7%, a 6.0% improvement over Group 1, while the frame rate remained above 30 FPS, a 100% improvement over Group 1. Comprehensive analysis shows that the synergistic effect of the two sub-modules not only enhanced robustness in complex scenarios through background suppression but also optimized computational load through keyframe sampling, achieving a dual improvement in accuracy and real-time performance, verifying the effectiveness of the complete preprocessing scheme.

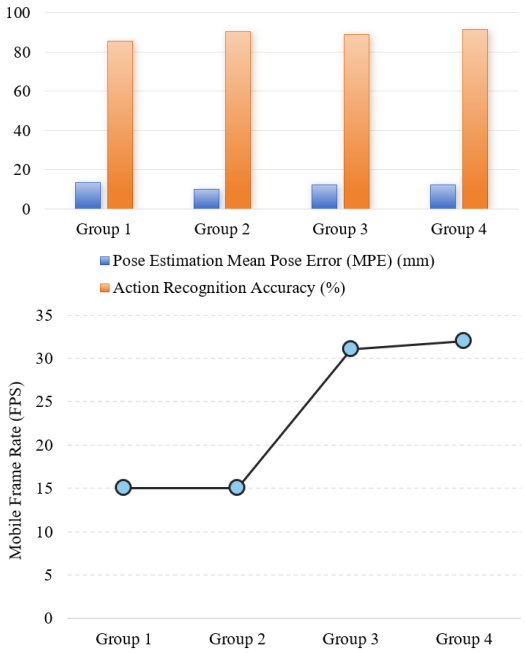

Figure 6. Results of preprocessing module ablation study

3.5 System real-time performance and mobile deployment testing experiment

The purpose of this experiment is to verify the real-time performance and stability of the system on mobile terminals to meet the interactive requirements of sports classrooms. Two mainstream mobile devices, iPhone 12 and Huawei Mate 50, were selected to test the latency of each system module, end-to-end delay, and frame rate. Additionally, the system was run continuously for 1 hour to test stability. Furthermore, an existing mobile action recognition system (MobileAction v2.0) was selected as a comparison to verify the real-time performance advantages of the proposed system, with consistent test conditions maintained.

Experimental results are shown in Table 2. On the iPhone 12, the proposed system had an image preprocessing time of 9 ms, EMAP inference time of 28 ms, STD-GCN inference time of 12 ms, and feedback visualization time of 11 ms. The end-to-end latency was only 60 ms, and the frame rate reached 32 FPS; on the Huawei Mate 50, image preprocessing took 10 ms, EMAP inference took 32 ms, STD-GCN inference took 14 ms, and feedback visualization took 12 ms. The end-to-end latency was 68 ms, and the frame rate reached 30 FPS. The latency on both devices was below the preset target of 80 ms, and the frame rate was above 30 FPS, meeting real-time interaction requirements.

Continuous 1-hour stability tests indicated that the system on both devices exhibited no lag or crashes, with frame rate fluctuations not exceeding 2 FPS and latency fluctuations not exceeding 5 ms, demonstrating good stability and the ability to adapt to the long-term continuous operation requirements of sports classrooms. Compared with the existing MobileAction v2.0 system, the proposed system reduced latency by 25.0% on iPhone 12 (from 80 ms to 60 ms) and increased the frame rate by 33.3% (from 24 FPS to 32 FPS); on Huawei Mate 50, latency decreased by 22.7% (from 88 ms to 68 ms) and frame rate increased by 30.4% (from 23 FPS to 30 FPS). Latency decreased by an average of 20%-30%, and frame rate increased by over 30% on average, verifying the optimization effect of TensorFlow Lite quantization and the double-buffered asynchronous inference mechanism, achieving efficient deployment of the system on mobile terminals.

Table 2. Mobile deployment performance test results

Device	Preprocessing Time (ms)	Enhanced Multi-scale Attention Pose Inference (ms)	Spatio-Temporal Differential Graph Convolutional Network Inference (ms)	Visualization Time (ms)	End-to-End Latency (ms)	Frame Rate (FPS)
iPhone 12	9	28	12	11	60	32
Huawei *Mate* 50	10	32	14	12	68	30

3.6 User experience and teaching effectiveness verification experiment

The purpose of this experiment is to verify the instructional feedback effect of the system in actual sports classrooms. A total of 120 students and 6 physical education teachers from three classes in a middle school were selected as subjects, divided into an experimental group (60 students, 3 teachers) and a control group (60 students, 3 teachers) for a 4-week teaching experiment. The experimental group used the proposed system to assist teaching; students autonomously adjusted their movements based on real-time feedback regarding joint errors and correction prompts from the system, while teachers provided targeted guidance combined with system scores. The control group adopted the traditional manual teaching mode, where teachers corrected student movements through visual observation. After the experiment, the movement standardization and learning efficiency of the two groups were compared, and satisfaction surveys were conducted among teachers and students via questionnaires.

Experimental results are shown in Table 3. The average movement standardization score of the experimental group was 88.6 points, representing a 17.8% improvement over the control group (75.2 points), falling within the expected range of 15%-20%. The average time for students in the experimental group to master standard movements was 8.2 hours, a 22.0% reduction compared to the control group (10.5 hours), indicating a significant improvement in learning

efficiency. Questionnaire survey results showed that the satisfaction rate of teachers in the experimental group was 89.2%, and that of students was 87.5%, both exceeding 85%. This indicates that the system can effectively assist teachers in conducting personalized teaching, help students quickly correct movement deviations, and improve learning outcomes, demonstrating good instructional practicality.

Due to reliance on manual teacher assessment, the control group was affected by subjective judgment, resulting in limited improvement in movement standardization. Moreover, the heavy workload made it difficult to achieve one-on-one precise guidance. In contrast, the experimental group utilized the system's fine-grained feedback and intuitive visualization, allowing students to autonomously detect and adjust movement deviations, while teachers could focus on guiding key issues. This significantly improved teaching efficiency and quality, verifying the application value of the proposed system in actual sports classrooms.

Table 3. Experimental results of teaching effectiveness comparison

Group	Avg. Movement Standardization Score (Points)	Avg. Time to Master Std. Action (Hours)	Teacher Satisfaction (%)	Student Satisfaction (%)
Control Group (Traditional Teaching)	75.2	10.5	72.3	68.8
Experimental Group (Proposed System)	88.6	8.2	89.2	87.5

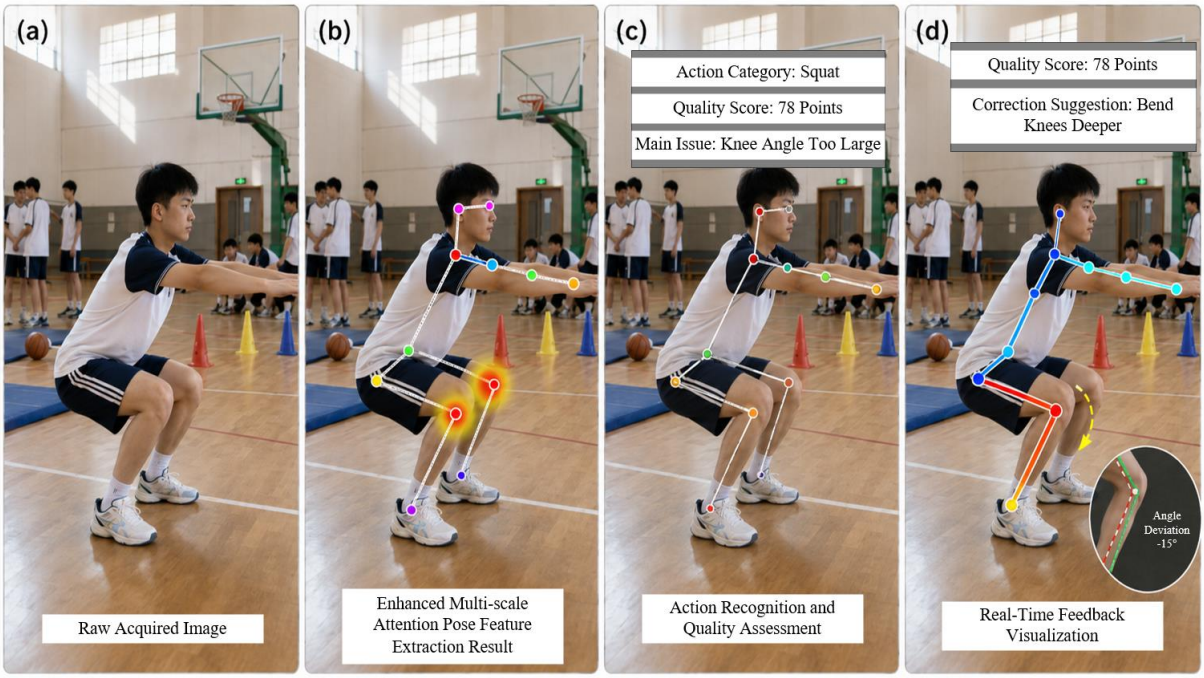

Figure 7. Visualization of implementation effect in sports classroom "Squat" action

To verify the effectiveness of the proposed method in recognizing complex scene actions and providing interpretable instructional feedback in real sports classrooms, this experiment selected the squat action as a typical closed-chain lower limb movement for visual analysis. Results in Figure 7 show that, despite the presence of background students, equipment occlusion, and non-uniform lighting in the classroom environment, the mobile-acquired images maintained relatively complete human pose structure input. The EMAP network stably localized key joints such as shoulders, hips, knees, and ankles, highlighting core movement areas dependent on action quality discrimination—specifically near the knee and hip joints—through high-response heatmap regions. This indicates that the model possesses strong attention-focusing capabilities on key areas of instructional movements. On this basis, STD-GCN further completed squat category recognition and quality scoring, assigning a rating of 78 (above average), and accurately prompting the primary deviation of excessive knee angle. This demonstrates that the spatio-temporal graph convolution structure can effectively model joint topological relationships and action temporal variations. In the real-time feedback results, the skeleton error color coding represented stable torso posture areas as low-error regions, while marking the thigh-to-calf linkage as high-error regions. Simultaneously, combined with directional correction arrows and local angle difference prompts, abstract scores were transformed into actionable movement adjustment suggestions. It is evident that the proposed method not only achieves robust recognition and quantitative evaluation of classroom movements under mobile interaction conditions but also further provides fine-grained interpretable feedback for teachers and students, offering effective technical support for immediate error correction, personalized guidance, and teaching quality improvement in sports classrooms.

4. CONCLUSION AND FUTURE WORK

Addressing core issues in mobile sports classroom action recognition—such as insufficient real-time performance, strong background interference, low pose feature extraction accuracy, and lack of fine-grained instructional feedback—this paper proposed an end-to-end intelligent processing

scheme integrating lightweight pose feature extraction with STD-GCN. The design details and technical principles of each module—including adaptive mobile image preprocessing, the EMAP pose feature extraction network, the STD-GCN action recognition and quality assessment network, and real-time feedback visualization—are systematically elaborated. Multiple sets of comparative and ablation experiments verify that this scheme significantly outperforms existing methods in pose estimation accuracy, action recognition precision, and mobile real-time performance. It provides fine-grained instructional feedback including joint angle errors, correction prompts, and quality scores, effectively meeting the practical requirements of intelligent sports classroom teaching. The lightweight network design and multi-module collaborative optimization strategy proposed in this paper not only solve the core dilemma of balancing accuracy and efficiency in mobile scenarios but also offer a new technical path and practical reference for the deep integration of mobile image processing technology and sports education intelligence, enriching research achievements in related fields. Although the proposed system has achieved effective adaptation to sports classroom scenarios, certain limitations remain. Pose estimation accuracy in complex extreme scenarios still has room for improvement, and the mutual interference problem during multi-student simultaneous action recognition has not yet been fully resolved. Future research will focus on these shortcomings: introducing lightweight Transformer structures to further enhance feature extraction accuracy and efficiency; fusing multimodal information such as inertial sensors to enhance the robustness of action recognition; and simultaneously expanding system application scenarios to adapt to physical education for different age groups and specialized sports action training requirements, promoting the development of intelligent sports education towards a more precise and comprehensive direction.

REFERENCES

[1] Østerlie, O., Killian, C. (2026). Becoming human with technology: rethinking digital practice in physical education. Learning, Media and Technology, 1-13. https://doi.org/10.1080/17439884.2026.2663160

[2] Teutemacher, B., Sudeck, G., Hapke, J. (2024). Pedagogical approaches to health-related physical education (PE) in the context of digitalisation – A scoping review. Physical Education and Sport Pedagogy, 31(3): 422-438. https://doi.org/10.1080/17408989.2024.2352826

[3] Li, Z., Slavkova, O., Gao, Y. (2022). Role of digitalization, digital competence, and parental support on performance of sports education in low-income college students. Frontiers in Psychology, 13: 979318. https://doi.org/10.3389/fpsyg.2022.979318

[4] Backes, A.F., Ramos, V., Quinaud, R.T., Ibáñez, S.J., Pizani, J., Carvalho, H.M., Nascimento, J.V. (2024). Physical education pre-service teachers constructivist teaching practices: A multilevel analysis. Quest, 76(4): 481-496. https://doi.org/10.1080/00336297.2024.2364608

[5] Zhang, Z., Zhang, N. (2017). Research on teaching practice growth mode of students major in physical education. EURASIA Journal of Mathematics, Science and Technology Education, 13(10): 7111-7120. https://doi.org/10.12973/ejmste/78737

[6] Chen, J., Bai, B. (2017). Research and practice on the teaching reform of college physical education under the multimedia environment. Agro Food Industry Hi-Tech, 28(1): 10-13.

[7] Tsuda, E., Ward, P., Ko, B., Santiago, J.A., Iserbyt, P., Dervent, F., Devrilmez, E., Kim, I., Xie, X. (2025). Physical education preservice teachers' perspectives on practice-based teacher education pedagogies: Teaching rehearsals and repeated teaching. Journal of Teaching in Physical Education, 1: 1-15. https://doi.org/10.1123/jtpe.2024-0189

[8] Xie, X., Ward, P., Oh, D., Li, Y., Atkinson, O., Cho, K., Kim, M. (2021). Preservice physical education teacher's development of adaptive competence. Journal of Teaching in Physical Education, 40(4): 538-546. https://doi.org/10.1123/jtpe.2019-0198

[9] Kyriakides, E., Tsangaridou, N., Charalambous, C.Y., Kyriakides, L. (2021). Toward a more comprehensive picture of physical education teaching quality: Combining generic and content-specific practices. Journal of Teaching in Physical Education, 40(2): 256-266. https://doi.org/10.1123/jtpe.2019-0162

[10] Liu, C., Dong, C., Li, X., Huang, H., Wang, Q. (2023). Analysis of physical education classroom teaching after implementation of the Chinese health physical education curriculum model: A video-based assessment. Behavioral Sciences, 13(3): 251. https://doi.org/10.3390/bs13030251

[11] Davis, K.S., Burgeson, C.R., Brener, N.D., McManus, T., Wechsler, H. (2005). The relationship between qualified personnel and self-reported implementation of recommended physical education practices and programs in U.S. schools. Research Quarterly for Exercise and Sport, 76(2): 202-211. https://doi.org/10.1080/02701367.2005.10599281

[12] Botagariyev, T., Kubiyeva, S., Akhmetova, A., Tissen, P., Mambetov, N., Sadykova, Z. (2022). Digitalization of physical education and its impact on academic performance among secondary school students in Aktobe and Orenburg. Interactive Learning Environments, 32(6): 1-11. https://doi.org/10.1080/10494820.2022.2148258

[13] Marín-Suelves, D., Ramón-Llin, J., Gabarda, V. (2023). The role of technology in physical education teaching in the wake of the pandemic. Sustainability, 15(11): 8503. https://doi.org/10.3390/su15118503

[14] Cao, F., Xiang, M., Chen, K., Lei, M. (2022). Intelligent physical education teaching tracking system based on multimedia data analysis and artificial intelligence. Mobile Information Systems, 2022: 1-11. https://doi.org/10.1155/2022/7666615

[15] Gustian, U., Satrio, S., Purmandaru, A. (2026). Mapping the integration of digital technologies in physical education: A bibliometric analysis of pedagogical trends and research trajectories. Quest, 1-20. https://doi.org/10.1080/00336297.2026.2632036

[16] Geng, X., Xin, Z., Gao, W., Shu, S., Piao, X. (2026). Environmental determinants of student comfort in an activity room. Environmental Research Communications, 8(3): 035010. https://doi.org/10.1088/2515-7620/ae4c24

[17] Lu, G., Ge, X., Zhong, T., Hu, Q., Geng, J. (2024). Preprocessing enhanced image compression for machine vision. IEEE Transactions on Circuits and Systems for

Video Technology, 34(12): 13556–13568. https://doi.org/10.1109/tcsvt.2024.3441049
[18] Engeroff, T., Bernardi, A., Banzer, W., Vogt, L. (2018). Computerized change of direction training, motor ability and cognitive processing: A randomized controlled trial. Medicina dello Sport, 71(3): 336-344. https://doi.org/10.23736/S0025-7826.18.03282-9
[19] Wang, Y. (2025). An intelligent path planning algorithm for dynamic football training environments. Expert Systems with Applications, 277: 126769. https://doi.org/10.1016/j.eswa.2025.126769
[20] Jun, W., Iqbal, M.S., Abbasi, R., Omar, M., Huiqin, C. (2024). Web-semantic-driven machine learning and blockchain for transformative change in the future of physical education. International Journal on Semantic Web and Information Systems, 20(1): 1-16. https://doi.org/10.4018/ijswis.337961