International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18 (2026) pp. 1–17

https://doi.org/10.70917/ijcisim-2026-0236
Article

Exploration of Gymnastic Movement Analysis and Skill Improvement Paths Based on Long and Short Term Memory Networks

Dan Mo[1,*], Yintong Wang[1] and Mengyun Hu[1]

1 College of Sports Arts, Jilin Sport University, Changchun, Jilin, 130022, China
* Correspondence author: momo0909415@163.com

Abstract: For a long time, the diversity and complexity of gymnastics movements have made it more difficult to carry out the analysis of gymnastics movement evaluation, which is an obstacle to the improvement of gymnastics skills. In order to improve the recognition effect of gymnastics movements, this paper combines OpenPose and LSTM to establish a HOPL model for gymnastics movement recognition and evaluation. The model combines the original OpenPose network with a high-resolution network to achieve multi-feature fusion of gymnastic actions, and introduces the LSTM model to recognize or predict complex actions and behaviors by using the temporal information of skeleton sequences. Experiments show that the recognition accuracy and global F1 score of the HOPL model are 90.75±1.76% and 86.72±0.93%, respectively, and the computational consumption is low, so it is feasible to apply it to the evaluation of gymnastic movements. Relying on the application of deep learning technology in the field of gymnastics movement recognition, it can provide optimization strategy support for gymnastics movement skill improvement.

Keywords: OpenPose; LSTM; high-resolution network; HOPL model; gymnastics movement recognition

1. Introduction

Traditional sports training methods often rely on the knowledge, experience and analytical judgment ability of the coach, and although this approach has its own unique features, there are certain limitations in terms of objectivity and accuracy [1]. In the high-intensity, high-frequency modern sports training, how to quickly and accurately recognize the wrong movements of athletes and give timely feedback has become the key to improve the training effect and prevent sports injuries [2-3]. In recent years, with the rapid development of computer vision technology, image processing and deep learning algorithms can provide a brand new solution for the recognition of wrong movements in sports training videos [4-5]. Image processing and deep learning algorithms first extract key information from massive sports training videos, then analyze the details of athletes' movements, and finally identify their erroneous movements in real time [6]. This method can not only greatly improve the accuracy and efficiency of the recognition of erroneous movements, but also provide coaches with richer and more intuitive feedback data for the development of more scientific and personalized training plans [7-8].

Long and short-term memory network is a typical deep learning algorithm, in which the input units are fixed-length vectors [9]. Through self-contained recurrent neurons, the output of the network is not only connected to the inputs of the current layer, but also to the outputs of the previous and next layers [10]. Long and short-term memory networks solve the interference and gradient mutation problems of standard recurrent neural networks by controlling the state of the gate transfer, remembering what needs to be memorized for a long time, and forgetting irrelevant information, which is particularly effective in many batch processing tasks that require “long-term memory” [11-12]. Long and

short-term memory networks have strong temporal processing ability, do not need to manually extract features, can effectively process the time series data in sports training videos, and improve the real-time and accuracy of action analysis [13]. Therefore, by constructing a long short-term memory network sports action recognition model to analyze the action of training videos, and then improve the skill level of athletes [14].

With the development of sensing technology, deep learning and big data, sports action recognition technology plays an increasingly important role in athletic training, competitive analysis, and recreational sports. In recent years, deep learning has made significant progress in the field of sports action recognition, which plays an important role in improving accuracy and robustness [15]. Deyzel et al [16] used graph convolutional neural network to map similar skeletal sports actions to the metric space and then used a single metric learning approach to classify and recognize different sports actions and the hybrid approach performed better on a dataset containing seven sports actions. Pham et al [17] developed a real-time human action recognition application, in which the deep neural network used can accurately recognize the training actions in the dataset, with an accuracy rate and F1 score of more than 90%, and is able to recognize the sports actions in real time and evaluate the quality of the training. Liu et al [18] designed a sports action recognition system based on the optimization model of the Support Vector Machines, in which the system acquires the user's sports data through a wearable The system acquires the user's movement data through wearable sensors, and then analyzes and recognizes the movements using the classifier of the support vector machine algorithm, and this method combined with the human skeleton model can efficiently guide the athletes in sports training. Jiang et al [19] launched a research on sports combination training action recognition with three-stream convolutional neural network as the core framework, and the constructed model has a high recognition rate for sports actions, which can meet the basic requirements of sports training. Kong et al [20] designed a joint framework capable of tracking athletes in sports videos and modeling them based on discriminative temporal cues to the tracking results, which in turn enables sports action recognition, and the experimental results proved the effectiveness of the framework. Yuna et al [21] applied artificial neural networks, Hidden Markov Models, and graph neural networks for intelligent gymnastics teaching to realize the recognition of complex movements in gymnastics, and the accuracy, recall, and F1 scores of the hybrid model on the gymnastics movement dataset were 98.2%, 97.5%, and 97.8%, respectively.

Long and short-term memory networks are weak in capturing information in space, so combining two or more models to provide robust spatio-temporal modeling of human movement using different deep network architectures can improve recognition efficiency [22]. Literature [23] integrated convolutional neural network and long and short-term memory network to extract data features and classify data sequences from gymnastic sports sensors, and the performance test results showed that the combination of sensors and deep learning techniques has good application prospects in fitness sports training. Literature [24] proposed a vision-based approach which creates an application to capture movements and achieves high model recognition accuracy for different yoga asanas using deep learning algorithms such as Convolutional Neural Networks, Long Short-Term Memory Networks, and SoftMax Regression. Khobdeh et al [25] applied the YOLO algorithm to detect the movements of athletes in a sports video in real time, and then classified the detected actions by deep fuzzy long short-term memory network, and the two were nested to obtain a more transparent, explanatory and accurate prediction model for sports actions. Meng et al [26] designed a deep learning network for sports action recognition by integrating the advantages of quaternionic spatio-temporal convolutional neural network and long short-term memory network.The organic combination of multi-models avoids the loss of spatial features, and is able to capture the dependencies between video segments. Fok et al [27] developed a deep learning framework based on recurrent neural networks and long short-term memory to classify human actions by learning the temporal features of video frames, and achieved 92.9% accuracy in recognizing different sports actions. Muhammad et al [28] constructed a deep learning network based on a bidirectional long short-term memory network and expanded convolutional neural network artificial intelligence framework, which can retain more informative features and learn the long-term dependencies of data sequence features, and improve the performance of the proposed method by 1% to 3% compared with the existing state-of-the-art sports recognition methods. Ullah et al [29] utilized the synergy between channel attention and spatial attention mechanisms and bidirectional long short-term memory networks for tennis dataset for action recognition testing and achieved excellent results in terms of accuracy, recall and other performance metrics. Sun et al [30] introduced a hybrid long and short-term memory network structure in a sports action recognition system to achieve comprehensive spatial-temporal feature modeling of data sequences, which promotes synergistic effect of sports action prediction module and recognition module. .

In addition, the optimization algorithm has further improvement effect on the action recognition

performance of the long short-term memory network.The research of Chen et al [31] solved the problem of insufficient processing power of the traditional sport recognition algorithms, and their proposed method based on gradient descent optimization-long short-term memory network had an average accuracy of more than 90% for action recognition of four sports, namely, rope skipping, swimming, ice skating, and shot-put throw, which has a wide range of application value. Chen et al [32] combined long and short-term memory network with bio-inspired optimization algorithm for recognizing athletes' movements.The hybrid algorithm models the discriminative monitoring effect of defined movements and uses spatial pyramid network to obtain the features of tracking frames.This method has high accuracy for athletes' movement assessment, which is helpful for the enhancement of their sports skills.

Aiming at the current problems of chaotic timing information extraction, difficult multi-feature fusion and poor recognition effect in the process of gymnastics action recognition, this paper establishes a HOPL model that combines OpenPose, HRNet and LSTM. In this study, OpenPose's VGG19 network and high-resolution network are used to realize the fusion of multi-feature data of gymnastics movements, and then combined with the LSTM model to obtain the timing data of human skeleton, which provides support for recognizing gymnastics movements. The results show that the HOPL model has a better recognition effect, and the overall feature fusion is better and the computational consumption is lower, which can meet the needs of gymnastic movement evaluation and analysis, and can also lay a reliable theoretical foundation for exploring the path of gymnastic movement improvement.

2. Gymnastic Movement Recognition Technical Basis

Under the support of deep learning, the accurate recognition of gymnastics movements has become an important means to further enhance the performance of gymnastics skills. In order to realize the accurate recognition and analysis of gymnastics movements, it is necessary to base on deep learning technology, supplemented by some human posture recognition algorithms, so as to improve the recognition accuracy of gymnastics movements.

2.1. LSTM and human pose solving

2.1.1. Long and short-term memory neural networks

Long Short-Term Memory (LSTM) network is a recursive neural network structure that is commonly used to process sequence data. With the development of machine learning, it is widely used in natural language processing, speech recognition, image recognition and other fields. Different from the traditional recurrent neural network (RNN), LSTM can effectively avoid the phenomenon of gradient vanishing as well as gradient explosion, and has the ability of long-term dependent modeling [33].

The key to LSTM is the introduction of three gating units (i.e., input gate, forgetting gate, and output gate) to control the input, forgetting, and output of information. Figure 1 shows the structure of the LSTM unit, where the input gate selectively stores new information into the memory cell and transfers it to the next layer, the forgetting gate is able to control which information can be deleted from the long-term state, and the output gate determines which information is output from the long-term state. In this way, the LSTM is able to accurately capture useful information in long sequences and ignore irrelevant information, thus improving the accuracy and generalization of the model.

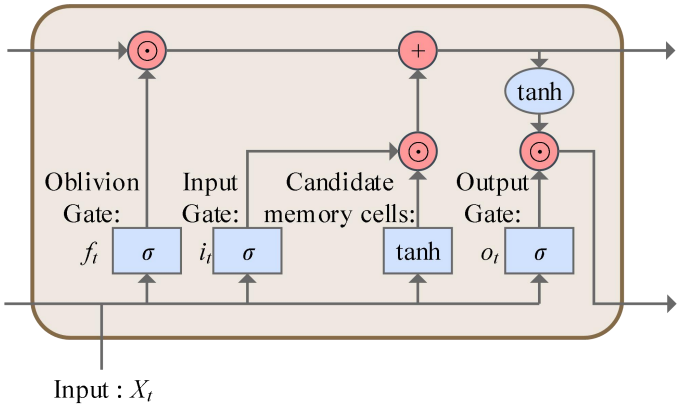


Figure 1. LSTM unit structure

The mathematical representation of LSTM is as follows:

$$f_t = \sigma\left(W_{lf} l_t + W_{mf} m_{t-1} + b_f\right) \tag{1}$$

$$i_t = \sigma\left(W_{li} l_t + W_{mi} m_{t-1} + b_i\right) \tag{2}$$

$$o_t = \sigma\left(W_{lo} l_t + W_{mo} m_{t-1} + b_o\right) \tag{3}$$

$$a_t = \tanh\left(W_{la} l_t + W_{ma} m_{t-1} + b_a\right) \tag{4}$$

$$c_t = c_{t-1} \otimes f_t + i_t \otimes a_t \tag{5}$$

Where σ denotes the sigmoid function, f_t, i_t and o_t represent the forgetting gate, input gate, and output gate vectors, respectively, and c_t and a_t are the memorization unit and hiding vector. $W_{l*} = \{W_{lf}, W_{li}, W_{lo}, W_{la}\}$ and $W_{m*} = \{W_{mf}, W_{mi}, W_{mo}, W_{ma}\}$ are the cyclic weights of the corresponding gates, b_f, b_i, b_o and b_a denote the output bias, and $\otimes$ is denoted as the Adama product.

2.1.2. Algorithm for solving human posture

When accurately measuring human posture, it is necessary to choose a suitable method for human posture solving to make the represented posture unique and increase the reliability of the recognition model. Euler's method, as a posture solving algorithm, is able to represent the posture in an intuitive and concise way with a small amount of computation, which meets the needs of the gymnastic movement recognition model.

The Euler angle method uses three rotation variables to determine the rotation of the object around the coordinate system, where the heading angle β denotes the angle obtained by rotating the target carrier around the OZ_g axis, the roll angle α denotes the angle obtained by rotating the target carrier around the OX_g axis, and the pitch angle θ denotes the angle obtained by rotating the target carrier around the OY_g axis. The initial coordinates of the carrier $OX_1Y_1Z_1$ coincide with the reference geographic coordinate system $OX_gY_gZ_g$, and then rotate around the OX_g axis, OY_g axis, and OZ_g axis, respectively, and ultimately obtain a new coordinate system $OX_3Y_3Z_3$.

According to the coordinate system and the direction cosine matrix theorem, it can be deduced that the attitude matrix obtained from three rotations around the three coordinate axes is:

$$T_\alpha = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & \sin\alpha \\ 0 & -\sin\alpha & \cos\alpha \end{bmatrix} T_\theta = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix} T_\beta = \begin{bmatrix} \cos\beta & -\sin\beta & 0 \\ \sin\beta & \cos\beta & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{6}$$

The three attitude angles represent a matrix of coordinate changes $T = T_\alpha \cdot T_\beta \cdot T_\theta$. Then:

$$T = \begin{bmatrix} \cos\beta\cos\theta & -\sin\beta\cos\alpha - \cos\beta\sin\alpha\sin\theta & -\sin\beta\sin\alpha + \cos\beta\cos\alpha\sin\theta \\ \sin\beta\cos\theta & \cos\beta\cos\alpha - \sin\beta\sin\alpha\sin\alpha & \cos\beta\sin\alpha + \sin\beta\sin\theta\cos\alpha \\ -\sin\theta & -\sin\alpha\cos\theta & \cos\theta\cos\alpha \end{bmatrix} \tag{7}$$

A set of attitude angles $[\alpha, \beta, \theta]$ is obtained according to the above equation, and the coordinate transformation matrix T can be derived from the attitude angles, which determines the attitude of the target carrier in the spatial coordinate system.

2.2. OpenPose pose recognition

2.2.1. OpenPose model

OpenPose is an open source library based on convolutional neural networks and supervised learning methods and developed under the caffe framework, which is capable of estimating a wide range of poses such as body movements, finger movements, and facial expressions for single or multiple people, and is particularly robust in multi-person pose estimation.The two important components of the OpenPose algorithm are the Convolutional Pose Machine (CPM) and local affinity (PAF).

The CPM Convolutional Pose Machine is composed of multiple stages, each of which is tasked with generating a confidence map to localize the various parts of the human body and to determine the location of each skeletal key point. The confidence maps are continuously passed and modified to improve the accuracy of the estimation. In order to enhance the ability to model the spatial uncertainty of the location of each skeletal keypoint, the network architecture of each stage takes into account the role of the confidence maps, and higher detection accuracy is achieved by designing convolutional kernels and pooling layers with larger sensory fields by using the confidence maps generated in the previous stage and the original image features as inputs. And as the stages increase, the output confidence maps become more and more accurate, which can predict the location of key points in the human body more accurately [34].

PAF local affinity is used to measure the affinity of each pixel point in the image, which is used to connect the scattered keypoints to determine whether the output keypoints are from the same person. The PAF corresponding to a pixel point on a limb is denoted as a unit vector, which represents the limb direction. After labeling the limb keypoints, the position and orientation information of the limb can be represented by connecting the keypoints. This connection is called affine field and is composed of 2D vector pointing. Since the PAF represents the position and orientation information between limb parts, it is able to accurately detect the pose of the human body.

Usually OpenPose contains more than five Stages, which are calculated as follows:

$$S^t = \rho^t\left(F, S^{t-1}, L^{t-1}\right), t \geq 2 \tag{8}$$

$$L^t = \varphi^t\left(F, S^{t-1}, L^{t-1}\right), t \geq 2 \tag{9}$$

OpenPose uses a maximal bipartite graph matching algorithm in order to solve the problem of localization and matching of key points on each human body in multi-person pose detection. The algorithm uses the Hungarian algorithm to find the best match between each key point, and finally generates a human skeleton map to realize multi-person pose detection.

2.2.2. Partial affinity algorithm

Local affinity is used to calculate the affinity value of the pixel points in the obtained image [35]. The human body in the image is labeled and the key points of the body are connected to get the position and orientation information between the limbs, which finally forms a 2D vector. Suppose $X_{j1,k}$ and $X_{j2,k}$ are the Ground Truth positions corresponding to body parts j1 and j2 connected to limb stem c of the k rd person, p refers to the current pixel point, and $Lc,k(p)$ is the affinity vector. If the pixel point p exists on the limb stem c, then $L_{c,k(p)}$ is the unit vector v from j1 to j2, and if the pixel point does not exist, then $L_{c,k(p)}$ is 0. It is computed as follows:

$$\begin{cases} L_{c,k(p)} = \begin{cases} v, & p \text{ on limb } c,k \\ 0, & \text{else} \end{cases} \\ v = \dfrac{\left(X_{j1,k} - X_{j2,k}\right)}{X_{j1,k} - X_{j2,k}{}^2} \end{cases} \tag{10}$$

where v is calculated as a single vector of limbs and satisfies the qualification. i.e:

$$\begin{cases} 0 \le v \cdot \left(p - X_{j1,k} \right) \le l_{c,k} \\ \left| v \cdot \left(p - X_{j1,k} \right) \right| \le \sigma_1 \end{cases} \tag{11}$$

Where σ_1 is the width of the limb, the limb length $l_{c,k}$ is calculated as:

$$l_{c,k} = \left\| X_{j1,k} - X_{j2,k} \right\|^2 \tag{12}$$

With the above computation, the network can then predict the connectivity domain of the desired site. Its limb c of the individual included in the image and the prediction computation can be expressed as:

$$L_{c^*,k(p)} = \frac{1}{n_{c(p)}} \sum_k L_{c^*,k(p)} \tag{13}$$

Where $n_{c(p)}$ is the number of non-zero vectors in k objects at point p. The judgment of whether the joints on the joint-point confidence diagram are capable of forming a limb is made by calculating the line integrals corresponding to the PAFs along the positions connecting the candidate keypoints. The judgment process involves comparing the keypoints to be detected d_{j_1} and d_{j_2} and then calculating the consistency of the keypoint vectors with the corresponding line segments in the joint domain of the body part. The confidence level of different region associations is calculated after sampling AF, Lc, and the calculation process is:

$$\begin{cases} E = \int_{u=0}^{u=1} L_c(p(u)) \cdot \dfrac{d_{j_1} - d_{j_2}}{\left\| d_{j_1} - d_{j_2} \right\|^2} du \\ p(u) = (1-u) d_{j1} + u d_{j2} \end{cases} \tag{14}$$

where $p(u)$ is the position of the insertion between the different sites.

2.2.3. Maximum bipartite graph matching

Usually, the distance of the skeletal keypoints of the same person should be relatively close to each other, the simple approach is to store all the keypoint information, and then the distance between the keypoints of the target type and the keypoints that need to be connected is sorted from smallest to largest, and then connect the keypoints in turn to complete the matching of the skeletal keypoints. The logic of this algorithm is simple, but in the case of character occlusion, the method of applying distance for clustering will cause a large error. Therefore, the method of clustering keypoints based on the distance relationship is not able to make an accurate judgment in the case where the characters' limbs are occluded from each other, so OpenPose adopts the Hungarian algorithm to complete the key algorithm.

The joints affinity is obtained through the previous algorithm, and the clustering can be completed by choosing the appropriate constraints to complete the matching of joints, the Hungarian algorithm, which is the algorithm for maximum bipartite graph matching, is used in OpenPose, because the same joint does not allow multiple edges to share a node and satisfy the constraints as follows:

Assuming that there is a possibility of matching between node type 1 and node type 2, the sum of the confidence level of the nth joint of type 1 and all the key parts of type 2 must not exceed 1. Otherwise, it means that the number of connections between joint n and node type 2 is more than 1. Then, this clustering method is violated. Conversely, the same is true for detecting the constraints of shutdown point type 2 with joint type 1.

After that, all the connections that satisfy the constraints are found and their maximum sum of integrals is found, and the connection of node 1 to node 2 is determined by finding the maximum value of the sum of integrals in the connections, and the limb that connects node 1 to node 2 is thus found. Repeat the above steps for all other limbs to complete the human posture detection.

3. Analytical model for gymnastics movement recognition

For a long time, due to the diversity and complexity of gymnastics movements, it is difficult to carry out the work of gymnastics movement identification and evaluation. As we all know, the comprehensive evaluation of gymnastic movements is only completed by subjective feelings, which to a certain extent leads to the evaluation of gymnastic movements detached from the objective facts. Based on this, this paper applies neural networks to gymnastics movement identification and evaluation, aiming to further ensure the accuracy of gymnastics movement identification and improve the objectivity of gymnastics movement evaluation.

3.1. Gymnastic movement data production

3.1.1. Gymnastic movement acquisition

The gymnastic action acquisition apparatus used mainly includes 9 infrared high-speed motion capture systems, 1 high-speed camera, 50 infrared reflective spheres, and other items such as ground coordinates, computers, alcohol, tape, and recording forms. The acquisition process of gymnastics movements is as follows:

(1) Record the subject's basic information, including name, age, height, weight, years of training, athletic level, and any history of injury or disease, and explain the test process to the subject, as well as the matters that need to be paid attention to in the test.

(2) Layout and placement of equipment, connection and debugging. The test site is the middle of the carpet with a length and width of 10m*10m, 9 cameras were placed on the carpet in front of, behind and on both sides of the athlete when he was at rest, and they were connected in series for debugging in order after the cameras were placed. Ensure that in the process of experimental acquisition, the infrared light reflected from a certain infrared reflective ball on the subject can be captured by two cameras in an instant, and obtain the three-dimensional coordinates of the infrared reflective ball.

(3) Proofreading system. The sampling frequency of the infrared high-speed camera is 24Hz, in order to ensure the shooting effect, before the test starts, clear all the idle people and interference points, calibrate the action shooting space range, and complete the system to capture the sign point. In order to reduce the error, the subjects uniformly wear tight-fitting training clothes with gymnastics half-leg shoes.

(4) Warm-up. The athletes were allowed to warm up sufficiently before the start of the test to avoid any accidental event of sports injury during the test. In addition, wipe the paste Mark reflective ball area with alcohol, stick 50 markers for the athletes, and carefully verify the position of each marking point. Let the subjects clear the whole experimental process and familiarize with the experimental environment after and signal the good condition to start the test.

(5) Static data collection. Subjects stand with legs open and arms spread out in an anatomical position for calibration, and static data were collected and saved.

(6) Dynamic data acquisition. After ensuring that everything is normal, remove the medial knee drop marker and perform dynamic data acquisition on the subject, who is located in the test site and in a preparatory position. After hearing the operation command from the control center personnel, the subject started to complete the technical movements as required, while the control center technicians clicked on the camera. During the testing process, the head coach of the subjects watched whether the movements were effective or not, to ensure that the experiment was effective data collection. This gymnastic movement data mainly includes six movements, i.e., stretching, chest expansion, kicking, body side, body turning and whole body movement, i.e., AC1~AC6.

(7) End work. Subjects complete the movement when clicking on the end of the shooting, and check whether there is no drop point situation to save the data.

3.1.2. Data set pre-processing

(1) Data outliers processing

In the process of designing the data acquisition platform, this paper prevents the generation of abnormal data by means of hardware initialization and sensor correction, but the inadvertent jittering of the tester in the process of data acquisition of gymnastic movements also leads to the generation of abnormal data, which tend to be much larger or much smaller than the normal data collected. In order for the abnormal data not to affect the subsequent identification and evaluation of gymnastics movements, it is necessary to identify and remove these abnormal data. For the abnormal data judgment standards, the following two categories of conventional standards are currently available:

One category is that the data far exceeds the standard value of the data, and both the standard value

and the value of deviation are established based on the actual situation of that data source. The other category compares whether the data deviation exceeds three times the standard deviation of the mean. This is also known as the Lajda criterion, which is determined by the following formula:

$$|d_t - d_{mean}| > 3d_{std} \tag{15}$$

Where d_i represents the i th frame of data, d_{mean} represents the mean of the previous $i-1$ frames of data, and d_{std} is the standard deviation of the data.

In this paper, the Lajda criterion is used to deal with the outliers in the data, and the frame data can be deleted because the acquisition platform collects data at 240 Hz acquisition rate.

(2) Data filtering

In the process of sensor acquisition of gymnastics data, it is inevitable that the data acquired will be affected by personnel jitter or the environment, which will have an unpredictable impact on the subsequent recognition and evaluation of gymnastics movements, so data filtering is an indispensable step in data processing. The main purpose of data filtering is to highlight the information characteristics of gymnastics action data, the frequency of human action information is concentrated in the low-frequency region, but the noise comes from various frequencies, so the filtering algorithm is needed to remove the noise data distributed in the high-frequency and low-frequency. In this paper, the first-order hysteresis filtering method is used to filter the data, which is more suitable for the scene of high noise frequency, and the algorithm has excellent anti-interference ability to the periodic noise. The first order lag filtering formula is as follows:

$$Y(n) = \alpha X(n) + (1-\alpha)Y(n-1) \tag{16}$$

Where α is the filter coefficient, $X(n)$ is the sampled value of the frame, $Y(n-1)$ the filtered output value of the previous frame, and $Y(n)$ the filtered value of this frame.

3.2. Gymnastic action posture recognition model

3.2.1. Primitive feature extraction network

OpenPose is a deep learning network that can estimate the pose of multiple objects in the same scene, and the key points of the human body obtained are 2D coordinate information without spatial information. It will first recognize the skeletal points belonging to the person in the picture, and then find the skeletal points belonging to the same person by clustering to connect them to form the skeletal pose map of the person. The whole network structure of OpenPose is built based on a VGG19 convolutional neural network.

In the VGG19 network structure, the size of its convolutional kernel is 3x3, which is different from other common convolutional networks. The 3x3 convolutional kernel ensures that enough feature points can be captured, and at the same time, it also reduces the number of parameters of the model, improves the computing speed, and strengthens the ability of the model's nonlinear expression. Meanwhile, the pooling kernel of the maximum pooling layer of the VGG19 network is 2x2, allowing this model to outperform some other lightweight networks in detail capture as well.

When the OpenPose network performs feature extraction, it first extracts the detailed features through the VGG19 convolutional neural network, and then feeds the feature maps into each stage module, which consists of two serial branches, one serial branch outputs the PCM, and the other serial branch outputs the PAF, and the PCM and the PAF from each stage module are summed up by the Loss solver to form the total Loss. Loss. The purpose of having multiple stage modules in the network is to solve the key points more accurately. When the first stage module produces the output, it has already detected some key points that are easy to detect, however, places like legs and waist, which are difficult to detect, may not be detected.

3.2.2. High Resolution Network HRNet

High Resolution Network (HRNet) is a feature extraction network with excellent performance. Previously, the high-resolution features of ordinary networks could only provide low-level semantic expressions by themselves due to only a small number of convolution operations, or the high-resolution features obtained from low-resolution features through up-sampling, which have good semantic expressions by themselves, but the up-sampling itself cannot make up for the loss of spatial resolution.

Therefore, the spatial sensitivity of the final output high-resolution representation is not high.

HRNet, on the other hand, can always maintain high-resolution features throughout the network by gradually introducing low-resolution convolutions and connecting different resolution convolutions in parallel. At the same time, the expression ability of high-resolution features and low-resolution features is improved by constantly exchanging information between different resolution features. Fig. 2 shows the structure of HRNet network, which is divided into three parts, i.e., parallel convolution module, fusion module, and output module.

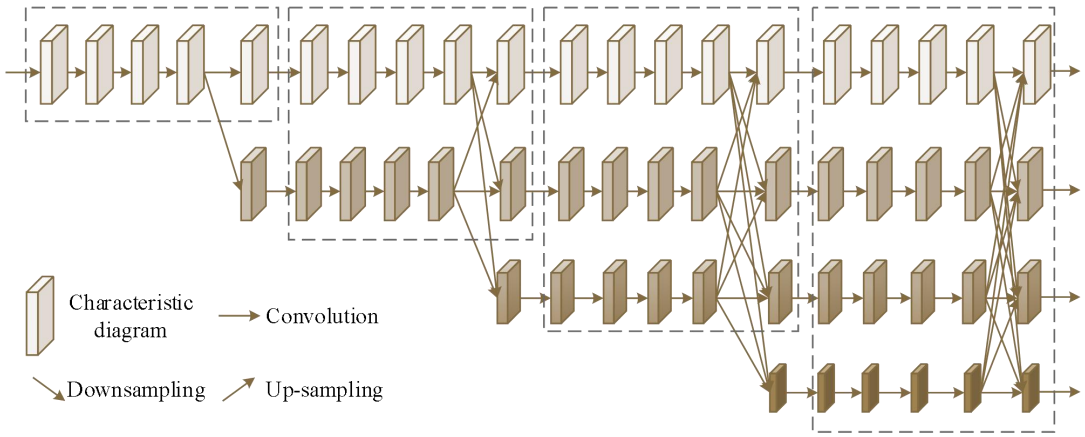


Figure 2. HRNet network structure

(1) The parallel convolution module is the core module of HRNet, which starts from the convolution stream with high-resolution features, and gradually introduces low-resolution convolution streams obtained by downsampling the high-resolution convolution streams in parallel, and finally forms convolution streams with multiple parallel branches.

(2) The role of the feature fusion module is to fuse the features from different parallel branches, thus realizing the exchange of information between multiple features of different resolutions, aiming to make full use of the feature information of each scale, so as to improve the model performance. The high-resolution features are transformed into features of the same size as the low-resolution features by downsampling.

(3) Output module, the high-resolution network can adopt three different output modes according to different task requirements, i.e., HRNetV1, HRNetV2, and HRNetV3. HRNetV1 mode outputs only high-resolution features, which is suitable for some high-precision tasks, while HRNetV2 mode outputs all the resolution features to the same size as the high-resolution ones, and then adjusts them to the same size as the high-resolution ones. HRNetV2 mode is based on HRNetV2, and its output is downsampled to form a multilayer feature representation, which is suitable for more complex deep learning tasks.

3.2.3. HRNet-OpenPose-LSTM Modeling

In the task of gymnastics action recognition, the traditional single-frame detection method suffers from the problem of temporal information fragmentation, which leads to the occurrence of significant keypoint localization jitter (average jitter amplitude up to ±6.2 pixels), sharp fluctuation of keypoint confidence (standard deviation of confidence up to 0.31) due to short-term occlusion, and the lack of physical kinematic constraints for successive actions (the joint angle mutation rate is more than 27%) in a motion blur scenario. In order to solve the above problems, this study proposes to embed the LSTM temporal modeling module in the OpenPose framework, which is constructed between the feature extraction network and the key-point regression layer, forming a three-stage processing architecture of “spatial features - temporal modeling - key-point prediction”.

On the basis of the above structure, in order to better solve the problem of spatial information loss and timing jitter of the OpenPose framework in complex scenes, this study proposes an improved architecture that integrates multi-resolution feature extraction and timing modeling, i.e., the HRNet-OpenPose-LSTM model (HOPL model), which is composed of HRNet, the original OpenPose framework, and the LSTM model. OpenPose framework and LSTM timing optimizer are combined, and its network structure is shown in Fig. 3.

Among them, HRNet is mainly responsible for extracting high-precision key point localization from

the input images and videos to ensure that the OpenPose framework can accurately detect human joints, the OpenPose framework is mainly responsible for calculating the output human skeleton sequences (coordinates of the key points and their connectivity relationships), and the LSTM utilizes the temporal information of the skeleton sequences to recognize or predict complex actions and behaviors.

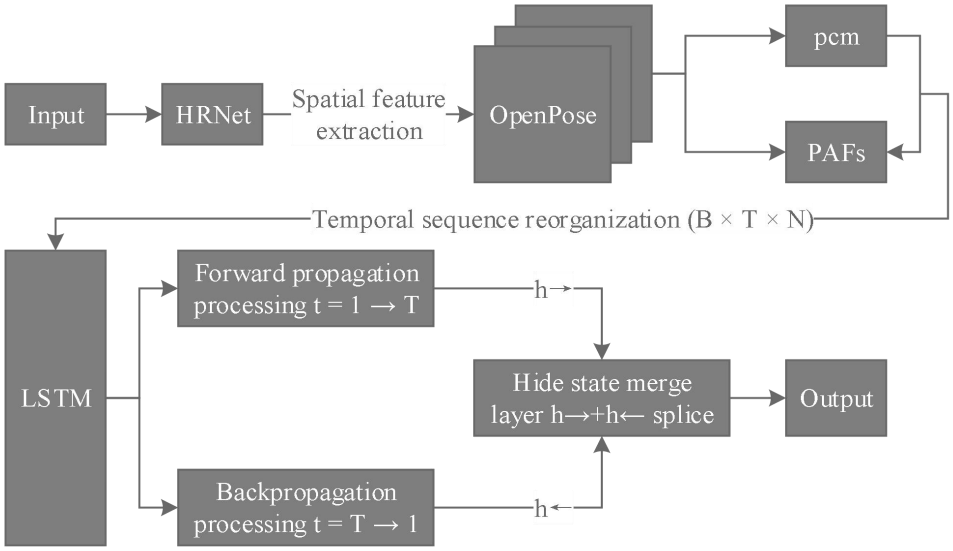


Figure 3. HRNet-OpenPose-LSTM network structure diagram

4. Experimental analysis of gymnastics movement recognition

Aiming at the current problem of not being able to identify gymnastic movements timely, effectively and accurately in gymnastics, this paper proposes a HOPL model that integrates high-resolution network, LSTM and OpenPose model, aiming to effectively enhance the recognition accuracy of gymnastic movements, so as to lay the foundation for the evaluation of gymnastic movements, and also to provide support for optimizing the exercise strategy of gymnastic skills, and to further promote the enhancement of gymnastic skills.

4.1. Gymnastic movement recognition experiment

4.1.1. Analysis of model identification effects

The HOPL model proposed in this paper is built on PyTorch platform (GPU: NVIDIA TITAN XP, RAM: 4×16G, Driver Version: 515.64.02, CUDA Version: 11.64). The Adam optimizer was used in all the models in training, with an L2 regularization factor of size $1*10^{-5}$, a Dropout ratio of 0.15, and cross-entropy for the loss function.

Based on the gymnastics sports action dataset obtained in the previous section, the dataset was divided into training and testing datasets in training, the selection of hyperparameters was partly done using ten-fold cross-validation, the training dataset was divided into 10 subsets, and 7 subsets were taken each time for training the model, and 3 subsets were used for validating the model. The evaluation metrics in the model comparison experiments include classification accuracy, F1 scores for each class, and global F1 scores. The global F1 score represents the overall model score, and all the evaluation metrics are calculated from the confusion matrix.

A total of eight baseline models are selected, of which models 1~4 are data layer fusion models (i.e., single-layer LSTM, double-layer LSTM, BiLSTM and GRU), the number of hidden elements are all set to 16, the learning rate are all set to 0.0001, and the total number of training times is 200. The raw data of the three modalities are directly spliced according to the dimensionality after preprocessing and normalization, and subsequently directly input to the baseline model. The classification results are output after dimensionality reduction by the fully connected layer. Models 5~8 are feature layer fusion models (LSTM-CONCAT, LSTM-ADD, LSTM-TFN, and LSTM-LWF), with a learning rate of 0.0001 and a total number of training times of 400, which are divided into feature extraction and feature fusion, and feature extraction is done by a single-layer long and short-term memory network with the number of hidden elements of 16, and fusion is done by a fully connected layer with dropout strategy. After

fusion, the features are output through the fully connected layer using Dropout strategy.

Table 1 and Table 2 show the model classification effects and F1 scores for different action types, respectively. Combining the data in Tables 1 and 2, it can be seen that comparing with the baseline model, the HOPL model proposed in this paper is able to achieve the highest accuracy and global F1 scores, with values of 90.75±1.76% and 86.72±0.93%, respectively, and also achieves the best F1 scores in classifying each category of gymnastic actions. Meanwhile, the classification effect for the gymnastics action category AC6 (whole-body movement) is especially improved, indicating that the problem of low classification accuracy caused by the difficulty of learning with a small number of samples can be alleviated by extracting deep action data features. The feature fusion model can intuitively compare the effects of the fusion algorithms due to the identical feature extraction part, the addition operator with weights is better than the splicing operator, and the LWF algorithm is better than the TFN algorithm. In addition, due to the limited amount of overall data, serious overfitting occurs when the model is a two-layer stacked LSTM, and the classification effect decreases the accuracy by 3.29% compared with the single-layer LSTM, and the classification F1 scores of AC1, AC2, and AC6 decrease by more than 15%. It shows that elevating the number of model parameters when the number is limited will instead cause the classification effect to decrease, but the HOPL model proposed in this paper uses a 2-layer stacked LSTM, and the theoretical model parameters are much larger than the baseline model, and still achieves the best classification scores, which fully indicates that the HOPL model in this paper possesses structural reasonableness.

Table 1. Model classification effectiveness evaluation

No.	Name	Accuracy /%	Global *F1* score/%
1	LSTM	84.71±8.16	75.01±10.08
2	LSTM (Layer 2)	81.42±6.08	65.63±9.96
3	GRU	85.94±4.23	72.36±7.15
4	BiLSTM	88.79±4.31	78.12±6.84
5	LSTM-CONCAT	83.06±4.72	73.05±8.57
6	LSTM-ADD	83.95±4.46	73.57±5.12
7	LSTM-TFN	84.85±4.47	74.53±10.46
8	LSTM-LWF	84.81±2.64	75.63±7.51
9	HOPL	**90.75±1.76**	**86.72±0.93**

Table 2. *F1*-score of 6 types of actions (%)

No.	AC1	AC2	AC3	AC4	AC5	AC6
1	77.87	80.21	85.25	88.75	86.18	49.51
2	62.31	72.17	85.76	86.83	84.27	31.72
3	71.86	85.45	86.37	89.94	85.63	46.37
4	63.51	86.28	88.42	92.06	90.14	70.26
5	69.86	80.32	84.96	86.53	87.59	44.53
6	67.42	82.94	83.41	87.42	86.32	52.29
7	66.21	81.86	86.18	89.37	88.06	55.68
8	76.83	81.75	86.75	88.41	87.41	48.15
9	**78.85**	**90.24**	**91.56**	**93.47**	**91.27**	**86.58**

4.1.2. Comparison of the performance of different features

In order to further understand the performance differences that exist between LSTM and HOPL models, validation of the performance under different features (pose, skeleton, and fusion of the two) is carried out, which mainly includes recognition accuracy, training and computation time. Table 3 shows the performance comparison results of the models under different features.

The HOPL model using pose information and human skeleton has some improvement in

recognition accuracy as follows:

(1) The HOPL model has higher recognition accuracy than the LSTM model in all cases, with an accuracy improvement of 0.61 percentage points in the recognition of pose information, 0.31 percentage points in the recognition of human skeleton features, and 1.13 percentage points in the fusion recognition of the two.

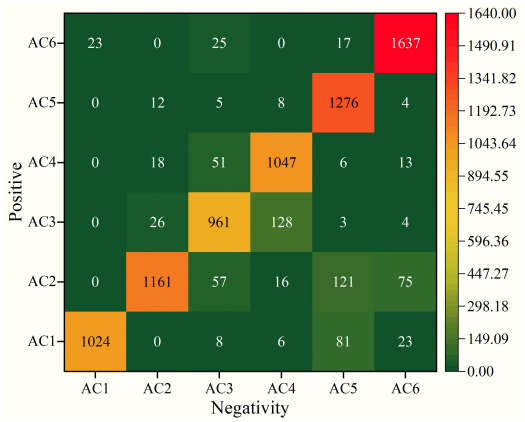
(2) HOPL models that use both pose information and human skeleton fusion features have improved recognition accuracy compared to HOPL models that use only pose information or human skeleton data.

Taken together, the HOPL model using the fusion of both pose information and human skeleton data has a certain advantage in recognition accuracy and processing complex data compared to using only pose information or human skeleton data, but the training and recognition time is slightly extended, thus the HOPL model using pose information and human skeleton data is suitable for scenarios that require high accuracy of action details.

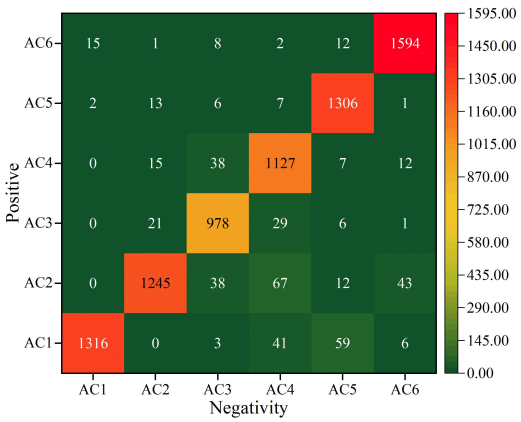
Table 3. Performance comparison results under different characteristics

Model	Feature	Accuracy /%	Training time /s	Recognition time /s
LSTM	Posture information	93.92	2051.42	29.59
	Human skeleton	97.85	36.58	8.64
	Both integration	98.71	77.63	13.23
HOPL	Posture information	94.53	3013.67	35.47
	Human skeleton	98.16	24.59	6.18
	Both integration	99.84	104.53	15.02

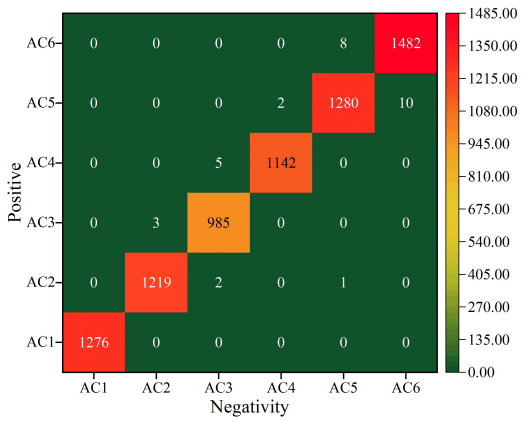
In order to clearly compare the results of the proposed features, the mixing matrices of human skeleton, pose information, and the fusion of human skeleton and pose information are plotted, and the specific results are shown in Fig. 4(a)~(c). It can be seen that compared with single features, the HOPL model has a relatively high recognition accuracy for the six gymnastic movements and a more accurate characterization of human movement features. Combined with the HOPL model for classification recognition, the human action behavior recognition method has stronger generalization ability and higher recognition accuracy. The fusion of human skeleton and pose information features can well characterize human movement intention. The HOPL model proposed in this paper improves the recognition effect by 5.72 percentage points and 2.38 percentage points compared to the pose information and human skeleton features, respectively. In the figure, the error of the fourth category of human skeleton features (body side movement) is 11.47% at the maximum, the error of the fifth category of pose information features (body turn movement) is 15.16% at the maximum, and the error of the third category of fused pose information and human skeleton features (kicking movement) is 0.71% at the maximum. This shows that the prediction is most accurate in the fusion of postural information and human skeleton features.



(a) Posture information



(b) Human skeleton



(c) Both integration

Figure 4. The confusion matrix of the HOPL model

4.1.3. Comparison of modeled computational consumption

Recognition accuracy and computational consumption are often dual challenges, and in addition to considering action recognition accuracy, computer consumption is equally important, which is an important metric affecting action recognition inference time. In order to optimize the time complexity and space complexity of the HOPL model, three groups of experiments were implemented, the first group used the fully connected method, the second group used the three-stream network but all the convolutional kernels were 6*6 and there were two convolutional layers, and the third group, i.e., HOPL model proposed in this paper, had convolutional kernels of (6*6/3*3). The article estimates the computational consumption of the five groups of learning models on the dataset, and their specific results are shown in Table 4.

The data analysis found that the HOPL model forms a better solution between computational consumption, data size and recognition accuracy, compared with the Bi-CNN-LSTM model, not only the action recognition accuracy is slightly higher by 1.69%, the computer consumption is significantly lower, and the time complexity and space complexity only account for 66.48% and 72.69%, respectively. According to the computational consumption calculation method, the convolution kernel size and fully connected layers are crucial to the spatially responsible metric influence, and the number of convolutional layers and the pixel size of the input data are the core elements to determine the time complexity, so the article optimizes the network structure again, shrinks the pixels of the input image, and alters the convolution kernel and the fully connected layers, to reduce the computational consumption under the circumstance of ensuring a small loss of accuracy, thus reducing the inference response latency.

Table 4. Comparison of time and space complexity

Model	Time~O (Flops)	Space~O (MACCs)	Accuracy
Motif-STGCN	2.531bn	3.127bn	84.26%
MTANs	3.815bn	3.956bn	84.75%
Bi-CNN-LSTM	5.487bn	5.693bn	87.13%
TS-CNN-LSTM (6*6)	4.282bn	4.462bn	82.97%
HOPL (6*6/3*3)	**3.716bn**	**4.018bn**	**88.82%**

4.2. Analysis of Gymnastic Movement Assessment

4.2.1. Overall assessment of gymnastic movements

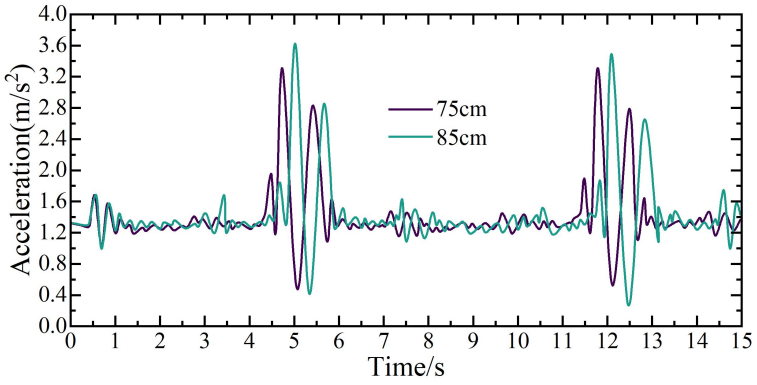
Gymnastic movement assessment is to change the stretching angle and kicking height as reference index.

(1) Height assessment

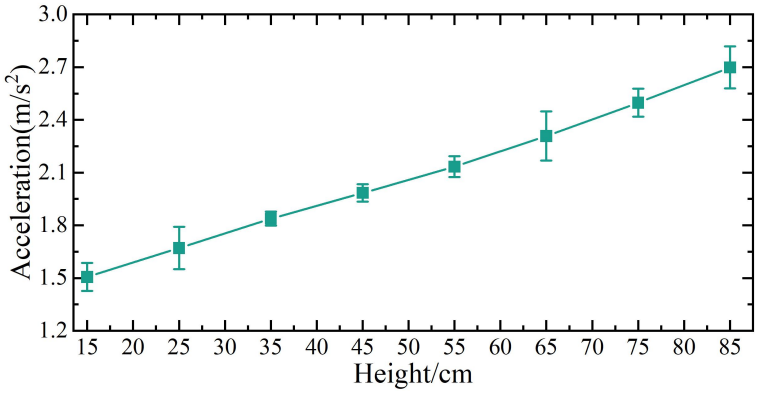
Height assessment is realized by changing the kicking height in the kicking movement to assess the gymnastic movement. Eight heights of 15cm, 25cm, 35cm, 45cm, 55cm, 65cm, 75cm and 85cm were pre-set for the kicking movement experiment. Among them, the acceleration data curve intuitively

reflects the movement state of the human body, and the angular velocity data curve reflects the direction of the human body movement, so the three-axis acceleration curve is generally chosen to further analyze the movement of the human body movement. Since the Y-axis direction of the sensor position node is perpendicular to the ground, the X-axis direction is parallel to the ground, and the Z-axis direction is opposite to the direction of the human body's movement, the obtained Y-axis acceleration curve has the data curve with the most significant cyclic variation rule, so the Y-axis acceleration curve is chosen to evaluate different kicking heights. Figure 5 shows the results of error analysis for different kicking heights, where Figure 5(a)~(b) shows the Y-axis acceleration change curve and height error graph of kicking motion, respectively.

Through the peaks and patterns of the curves, it can be concluded that the higher the kicking height, the higher the acceleration value. Therefore, the kicking heights of 15~35cm, 45~65cm and 75~85cm can be rated as poor, good and excellent, respectively. Combined with the height error graph, it can be seen that the higher the kicking height, the more obvious the acceleration error, which further indicates that the larger the gymnastic movement amplitude, the higher the recognition rate for it.



(a) The Y axis acceleration curve

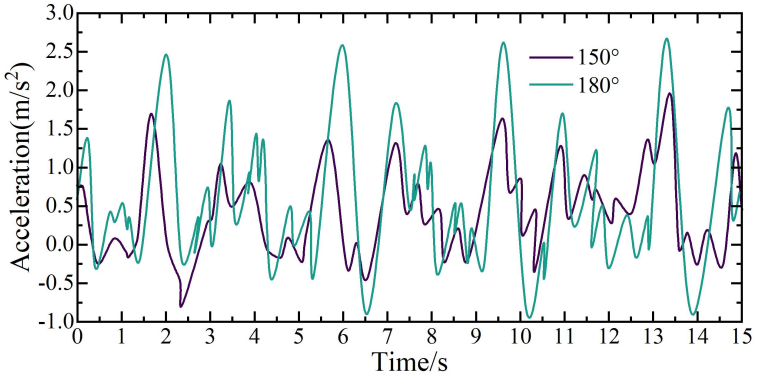


(b) Altitude error diagram

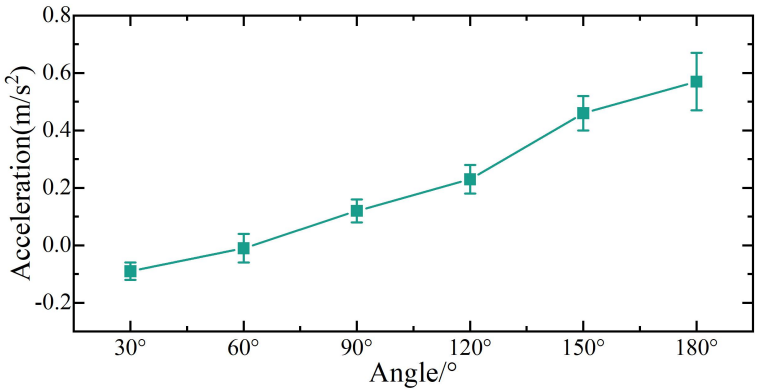
Figure 5. Error analysis of different kicking height

(2) Angle assessment

Angle assessment is realized by changing the stretching angle in the stretching exercise to assess the gymnastic movement. Six angles of 30°, 60°, 90°, 120°, 150° and 180° were pre-set for the experiments with different stretching angles, and the results of error analysis for different stretching angles were obtained as shown in Figure 6. Figure 6(a)~(b) shows the acceleration change curve of Y-axis of stretching motion and the error map of stretching motion angle, respectively. Through the change rule of the peak value of the curve, it can be concluded that the larger the stretching angle is, the higher the acceleration value is. Therefore, the stretching motion angles of 30°~60°, 90°~120° and 150°~180° can be categorized into three grades: poor, good and excellent. Combined with the stretching motion angle error diagram, it can be seen that the larger the stretching angle is, the more obvious the acceleration error is.



(a) The Y axis acceleration curve



(b) Altitude error diagram

Figure 6. Error analysis of different stretching angles

4.2.2. Evaluation of details of gymnastic movements

In order to further carry out the evaluation of gymnastic movements, this paper takes the 8-beat movements of the kicking movement as an example, the 1st to the 8th beat of the kicking movement, including 8 static movements (JT1~JT8) and 8 dynamic movements (DT1~DT8), with a duration of about 8 seconds. The human movement features extracted from the experiment were the 8 joint pinch angles of the skeletal model, which were right shoulder pinch angle, left shoulder pinch angle, right elbow pinch angle, left elbow pinch angle, right hip pinch angle, left hip pinch angle, right knee pinch angle, and left knee pinch angle. After analyzing the action of kicking movement, the eight joint angles selected in this experiment can already fully reflect the action changes of kicking movement.

After the initialization of the parameters in the experiment was completed, the radio gymnastics movement data of three testers (tester S1, tester S2, and tester S3) were then collected for evaluation, in which tester S1 had already proficiently mastered the radio gymnastics movement, and tester S2 and tester S3 had average proficiency procedures. In this study, joint angle similarity was used as an evaluation index to measure the degree of similarity between user movements and standard movements by calculating the difference between each joint angle of the user and each joint angle of the standard movements. Figure 7 shows the results of the calculation of the similarity of the tester's joint angles.

As can be seen from the figure, the joint angle similarity values for each movement of tester S1 were all between 3° and 7°, and the movement was completed better overall compared to testers S2 and S3. Both test subject S2 and test subject S3 have relatively large joint Angle similarity calculation results for certain movements, such as "dynamic movement 3" and "static movement 7" for test subject S2, and "dynamic movement 3", "dynamic movement 7" and "static movement 7" for test subject S3. It indicates that the spatial angle error of some movements of testers S2 and S3 is large, and the quality of movement completion is average. Therefore, after obtaining the gymnastic movement posture data, further comprehensive assessment of gymnastic movements can be carried out to provide reliable data support for the quantitative evaluation of gymnastic movements and decision support for the improvement of gymnastic skills.

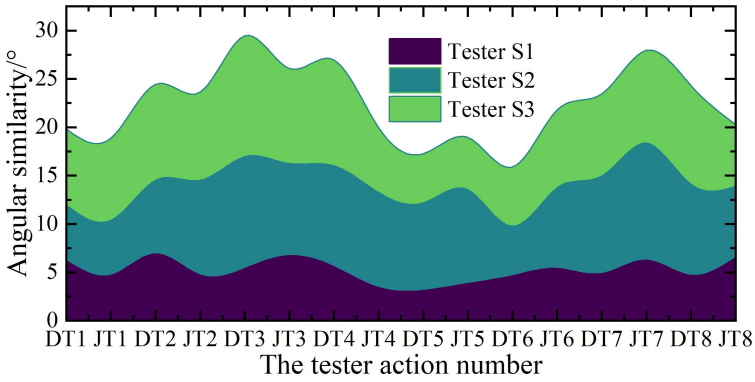


Figure 7. The Joint Angle Similarity of the Testers

5. Conclusion

In order to realize the accurate recognition and evaluation of gymnastic movements, this paper combines LSTM and OpenPose model to establish a HOPL model for the recognition and analysis of gymnastic movements. The model introduces a high-resolution network instead of the original feature extraction network, and combines with the LSTM model to obtain the timing information of gymnastic movements, so as to better realize the accurate recognition of gymnastic movements. It was found that the accuracy and global F1 score of the HOPL model were 90.75±1.76% and 86.72±0.93%, respectively, achieving the best F1 score in the classification of each type of gymnastics action. Further comprehensive assessment of gymnastics movements based on the results of gymnastics movement recognition can provide research-based data for improving the level of gymnastics skills.

Although this study has achieved certain research results, there are still shortcomings. One is that one-way LSTM is unable to comprehensively extract the timing information of human skeleton, which may lead to some timing data being ignored, thus making the results of gymnastics movement recognition inaccurate. Second, the fusion framework of high-resolution network and VGG19 is not further verified for its adaptability, which may cause the problem of inaccurate spatial feature detection to appear. In future research, the fusion of bidirectional LSTM and attention mechanism is further explored so as to enhance the ability of extracting the temporal information of gymnastic actions. The fusion of high resolution network and VGG19 is further improved to make its operation efficiency better when fusing multi-resolution features again.

References

1. Issurin, V. (2008). Block periodization versus traditional training theory: a review. Journal of sports medicine and physical fitness, 48(1), 65.
2. Frangoudes, F., Matsangidou, M., Schiza, E. C., Neokleous, K., & Pattichis, C. S. (2022). Assessing human motion during exercise using machine learning: A literature review. IEEE Access, 10, 86874-86903.
3. Huang, P. C., Liu, K. C., Hsieh, C. Y., & Chan, C. T. (2017, May). Human motion identification for rehabilitation exercise assessment of knee osteoarthritis. In 2017 International Conference on Applied System Innovation (ICASI) (pp. 246-249). IEEE.
4. Al-Faris, M., Chiverton, J., Ndzi, D., & Ahmed, A. I. (2020). A review on computer vision-based methods for human action recognition. Journal of imaging, 6(6), 46.
5. Zhang, H. B., Zhang, Y. X., Zhong, B., Lei, Q., Yang, L., Du, J. X., & Chen, D. S. (2019). A comprehensive survey of vision-based human action recognition methods. Sensors, 19(5), 1005.
6. Yao, G., Lei, T., & Zhong, J. (2019). A review of convolutional-neural-network-based action recognition. Pattern Recognition Letters, 118, 14-22.
7. AlShami, A. K., Rabinowitz, R., Lam, K., Shleibik, Y., Mersha, M., Boult, T., & Kalita, J. (2025). Smart-vision: survey of modern action recognition techniques in vision. Multimedia tools and applications, 84(27), 32705-32776.
8. Karim, M., Khalid, S., Aleryani, A., Khan, J., Ullah, I., & Ali, Z. (2024). Human action recognition systems: A review of the trends and state-of-the-art. IEEE Access, 12, 36372-36390.
9. Jia, Q., Zhu, Y., Xu, R., Zhang, Y., & Zhao, Y. (2022). Making the hospital smart: using a deep long short-term memory model to predict hospital performance metrics. Industrial Management & Data Systems, 122(10), 2151-2174.
10. Wang, Y. B., You, Z. H., Yang, S., Yi, H. C., Chen, Z. H., & Zheng, K. (2020). A deep learning-based method for drug-target interaction prediction based on long short-term memory neural network. BMC medical informatics and decision making, 20(Suppl 2), 49.
11. Gao, R., Tang, Y., Xu, K., Huo, Y., Bao, S., Antic, S. L., ... & Landman, B. A. (2020). Time-distanced gates

in long short-term memory networks. Medical image analysis, 65, 101785.
12. Wen, M., Zhou, Q., Tao, B., Shcherbakov, P., Xu, Y., & Zhang, X. (2023). Short-term and long-term memory self-attention network for segmentation of tumours in 3D medical images. CAAI Transactions on Intelligence Technology, 8(4), 1524-1537.
13. Yu, Z., Yang, K., Luo, Y., & Shang, C. (2020). Spatial-temporal process simulation and prediction of chlorophyll-a concentration in Dianchi Lake based on wavelet analysis and long-short term memory network. Journal of Hydrology, 582, 124488.
14. Wang, P. (2021). Research on sports training action recognition based on deep learning. Scientific Programming, 2021(1), 3396878.
15. Zhang, S., Li, Y., Zhang, S., Shahabi, F., Xia, S., Deng, Y., & Alshurafa, N. (2022). Deep learning in human activity recognition with wearable sensors: A review on advances. Sensors, 22(4), 1476.
16. Deyzel, M., & Theart, R. P. (2023). One-shot skeleton-based action recognition on strength and conditioning exercises. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 5169-5178).
17. Pham, Q. T., Nguyen, D. A., Nguyen, T. T., Nguyen, T. N., Nguyen, D. T., Pham, D. T., ... & Vu, H. (2022, December). A study on skeleton-based action recognition and its application to physical exercise recognition. In Proceedings of the 11th international symposium on information and communication technology (pp. 239-246).
18. Liu, Z., & Wang, X. (2023). Action recognition for sports combined training based on wearable sensor technology and SVM prediction. Preventive Medicine, 173, 107582.
19. Jiang, H., & Tsai, S. B. (2021). An empirical study on sports combination training action recognition based on SMO algorithm optimization model and artificial intelligence. Mathematical Problems in Engineering, 2021(1), 7217383.
20. Kong, L., Huang, D., Qin, J., & Wang, Y. (2019). A joint framework for athlete tracking and action recognition in sports videos. IEEE transactions on circuits and systems for video technology, 30(2), 532-548.
21. Yuna, X., Ruitongb, L., Zhuob, Z., Xueliangb, C., & Xinyub, Z. (2025). Construction Strategies and Experimental Design of the Action Recognition and Optimization System for Gymnasts. International Journal of Frontiers in Engineering Technology, 7(3).
22. Wang, L., Zhao, X., & Liu, Y. (2018). Skeleton feature fusion based on multi-stream LSTM for action recognition. IEEE Access, 6, 50788-50800.
23. Mekruksavanich, S., Tancharoen, D., & Jitpattanakul, A. (2024, July). Gym exercise recognition using deep convolutional and lstm neural network based on imu sensor data. In 2024 International Technical Conference on Circuits/Systems, Computers, and Communications (ITC-CSCC) (pp. 1-5). IEEE.
24. Rishan, F., De Silva, B., Alawathugoda, S., Nijabdeen, S., Rupasinghe, L., & Liyanapathirana, C. (2020, December). Infinity yoga tutor: Yoga posture detection and correction system. In 2020 5th International conference on information technology research (ICITR) (pp. 1-6). IEEE.
25. Khobdeh, S. B., Yamaghani, M. R., & Sareshkeh, S. K. (2024). Basketball action recognition based on the combination of YOLO and a deep fuzzy LSTM network: SB Khobdeh et al. The Journal of Supercomputing, 80(3), 3528-3553.
26. Meng, B., Liu, X., & Wang, X. (2018). Human action recognition based on quaternion spatial-temporal convolutional neural network and LSTM in RGB videos. Multimedia Tools and Applications, 77(20), 26901-26918.
27. Fok, W. W., Chan, L. C., & Chen, C. (2018, November). Artificial intelligence for sport actions and performance analysis using recurrent neural network (RNN) with long short-term memory (LSTM). In Proceedings of the 4th International Conference on Robotics and Artificial Intelligence (pp. 40-44).
28. Muhammad, K., Ullah, A., Imran, A. S., Sajjad, M., Kiran, M. S., Sannino, G., & de Albuquerque, V. H. C. (2021). Human action recognition using attention based LSTM network with dilated CNN features. Future Generation Computer Systems, 125, 820-830.
29. Ullah, M., Yamin, M. M., Mohammed, A., Khan, S. D., Ullah, H., & Cheikh, F. A. (2021). Attention-based LSTM network for action recognition in sports. Electronic Imaging, 33, 1-6.
30. Sun, X., Wang, Y., & Khan, J. (2023). Hybrid LSTM and GAN model for action recognition and prediction of lawn tennis sport activities. Soft Computing, 27(23), 18093-18112.
31. Chen, P., & Peng, J. (2025). Analysis of the sports action recognition model based on the LSTM recurrent neural network. Nonlinear Engineering, 14(1), 20240050.
32. Chen, J., Samuel, R. D. J., & Poovendran, P. (2021). LSTM with bio inspired algorithm for action recognition in sports videos. Image and Vision Computing, 112, 104214.
33. Chaonan Zhang & Shuzhang Liang. (2025). sEMG-based recognition of wrist-hand movements using a composite transformer-LSTM model. The Review of scientific instruments, 96(8), https://doi.org/10.1063/5.0271372.
34. Qi Chen. (2023). Interaction and psychological characteristics of art teaching based on Openpose and Long Short-Term Memory network. PeerJ. Computer science,9,e1285-e1285.https://doi.org/10.7717/PEERJ-CS.1285.
35. Monika Dhiman,Akash Sharma & Sarbjeet Singh. (2022). Recognition and Classification of Human Actions Using 2D Pose Estimation and Machine Learning. Journal of Computing Science and Engineering,16(4), https://doi.org/10.5626/JCSE.2022.16.4.199.