

Retraction

Retracted: An Online Arrangement Method of Difficult Actions in Competitive Aerobics Based on Multimedia Technology

Security and Communication Networks

Received 26 December 2023; Accepted 26 December 2023; Published 29 December 2023

Copyright © 2023 Security and Communication Networks. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi, as publisher, following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of systematic manipulation of the publication and peer-review process. We cannot, therefore, vouch for the reliability or integrity of this article.

Please note that this notice is intended solely to alert readers that the peer-review process of this article has been compromised.

Wiley and Hindawi regret that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] L. Li, "An Online Arrangement Method of Difficult Actions in Competitive Aerobics Based on Multimedia Technology," *Security and Communication Networks*, vol. 2021, Article ID 9968401, 12 pages, 2021.

Research Article

An Online Arrangement Method of Difficult Actions in Competitive Aerobics Based on Multimedia Technology

Li Li 

Yantai Institute of Technology, Yantai 264005, China

Correspondence should be addressed to Li Li; 201001001766@ytu.edu.cn

Received 19 March 2021; Revised 25 April 2021; Accepted 8 May 2021; Published 19 May 2021

Academic Editor: Chi-Hua Chen

Copyright © 2021 Li Li. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In accordance with the development trend of competitive aerobics' arrangement structure, this paper studies the online arrangement method of difficult actions in competitive aerobics based on multimedia technology to improve the arrangement effect. RGB image, optical flow image, and corrected optical flow image are taken as the input modes of difficult action recognition network in competitive aerobics video based on top-down feature fusion. The key frames of input modes in competitive aerobics video are extracted by using the key frame extraction method based on subshot segmentation of a double-threshold sliding window and fully connected graph. Through forward propagation, the score vector of video relative to all categories is obtained, and the probability score of probability distribution is obtained after normalization. The human action recognition in competitive aerobics video is completed, and the online arrangement of difficult action in competitive aerobics is realized based on this. The experimental results show that this method has a high accuracy in identifying difficult actions in competitive aerobics video; the online arrangement of difficult actions in competitive aerobics has obvious advantages, meets the needs of users, and has strong practicability.

1. Introduction

Competitive aerobics is a kind of sport that can perform continuous, complex, and high-intensity complete sets of movements with music accompaniment. It has a history of 20 years and has attracted much attention in international competitive sports. Compared with the ordinary aerobics, the difficult action and its completion effect under the condition of high frequency and large amplitude are the key to the success of competitive aerobics. With the continuous improvement of the competitive level of competitive aerobics, the difficult action is also developing towards the direction of high difficulty. As the main manifestation of the difficulty characteristics of competitive aerobics, the selection, arrangement, and completion quality of difficult action directly affect the quality of competitive aerobics. Therefore, it is particularly important to study the scientific and reliable online arrangement method of difficult actions in competitive aerobics [1, 2]. In the process of online arrangement of difficult actions in competitive aerobics, most of them take

the competition rules as the basis of arrangement or analyze the complete arrangement and advantages of the top few in the competition through video or make technical analysis on a certain difficult action.

With the rapid development of multimedia technology and its own advantages, it has been widely used in various industries. At the same time, video has gradually developed into one of the main data carriers. The application of multimedia technology in online arrangement of difficult actions in competitive aerobics can effectively improve the scientificity, creativity, and humanity of the arrangement [3]. At this stage, people have entered the digital information age. The application of multimedia technology in the online arrangement of difficult actions of competitive aerobics is deepening. It can not only solve the technical problems in the arrangement process but also effectively expand the popularity of research results.

In a network including access points, via which communication units can connect to the network, wherein each of the access points belongs to a zone, a method comprising

the steps of assigning a message type to each of the zones in the network; receiving from the communication unit a query message for each zone, where the query message for each zone is of the message type assigned to that zone; filtering the query messages at an access point, via which the communication unit is connected to the network, so that only the query message of the message type, which is assigned to the zone to which the access point belongs, is allowed to pass through; and determining the zone, to which the access point belongs, depending on the query message allowed to pass through in the step of filtering is used.

The online arrangement of difficult actions in competitive aerobics is a key research topic at present, which attracts the attention of many experts and scholars. Chen et al. improved the arrangement effect of difficult actions by improving the deep convolution neural network and put forward the development strategy of competitive aerobics, which should reasonably match the male and female athletes and strengthen the special training, strength training, and the cultivation of expression. However, the expression ability of video content was not good, and the recognition effect of human action in longer video was not obvious [4]; Zheng et al. recognized human action in video through posture space-time characteristics, to improve the rationality of difficult action arrangement, and they thought that competitive aerobics should learn from the advantages of other competitive sports according to their own characteristics and explore the road suitable for their own development, but this method had high computational complexity and large memory consumption [5]; Johansson completed human action recognition by combining and tracking the key human nodes in the video, which provided a strong basis for difficult action arrangement, but the efficiency of the training dataset was poor.

This paper studies the online arrangement method of difficult actions in competitive aerobics based on multimedia technology. The recognition method of difficult actions in competitive aerobics video based on top-down feature fusion is used to accurately identify the human movements in competitive aerobics video and make online arrangement of difficult actions in competitive aerobics on this basis. The experimental results can verify the effectiveness and rationality of this method, which provides a new research direction to promote the technical level of competitive aerobics in China. The specific contributions of this paper include the following:

- (1) Transplant research results in the field of artificial intelligence into computer-aided online arrangement actions
- (2) The basic concepts of computer-aided arrangement and artificial intelligence are introduced
- (3) The application of artificial intelligence in computer-aided arrangement is discussed

The rest of this paper is organized as follows. Section 2 discusses online arrangement of difficult actions in competitive aerobics based on multimedia technology; the related works and our proposal are introduced and detailed

experimental analysis is given in Section 3. It also shows result analysis and discussion, and Section 4 concludes the paper with summary and future research directions.

2. Research on Online Arrangement of Difficult Actions in Competitive Aerobics Based on Multimedia Technology

2.1. Key Frame Extraction Method Based on Subshot Segmentation of a Double-Threshold Sliding Window and Fully Connected Graph. With the development of multimedia technology, the multimedia information in today's work and life is increasingly rich. How to quickly and effectively retrieve useful information from the massive competitive aerobics video, so as to provide a reliable basis for online arrangement of difficult actions in competitive aerobics, has become an urgent problem to be solved.

A key frame is the image representing the shot, and the feature of the key frame can represent the feature of shot without considering the motion feature. The principle of extracting a key frame is better to make mistakes than to make fewer mistakes, and the basic requirement of extracting a key frame is to represent the main content of shot more completely and accurately [6–8]. In order to realize the key frame extraction of competitive aerobics video, the flow of the key frame extraction method based on subshot segmentation of a double-threshold sliding window and fully connected graph is designed as shown in Figure 1. The flow is divided into three modules from top to bottom: shot segmentation, subshot segmentation based on a sliding double window, and key frame extraction based on a fully connected graph. Feature extraction is the basis of these three modules.

2.1.1. Shot Segmentation. Among the shot segmentation methods, the method based on double threshold is the most widely used. In the method, two thresholds T_b and ($T_b > T_s$) are set, and the adjacent frame difference is calculated from the start frame of the shot. When the adjacent frame difference is greater than T_b , it is considered as the abrupt change boundary of the shot; when the adjacent frame difference is greater than T_s , it is considered as the gradual change boundary of the shot, and the interval frame difference is calculated from this frame. Once the accumulated interval frame difference is greater than T_b , the adjacent frame difference is less than T_s , and it is considered as the gradient boundary of the lens [9, 10].

2.1.2. Subshot Segmentation Method Based on the Sliding Window. Because there are long shots with frequent background changes and complex content coverage, it is easier to extract key frames based on semantics by dividing the shot into subshots with a single background and content.

Subshot segmentation also uses frame difference. Different from shot segmentation calculation and frame difference of the reference frame, subshot segmentation cannot avoid calculating frame difference between all frames, and the time complexity is $O(n^2)$. When the number of frames in

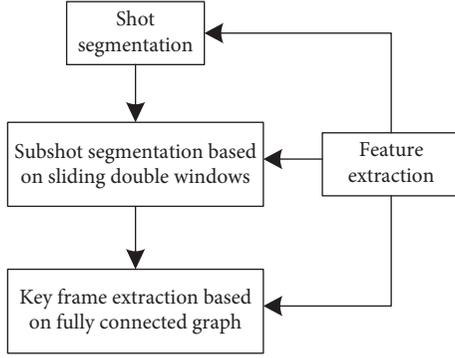


FIGURE 1: Key frame extraction process of competitive aerobics video.

the shot is large, the time complexity of $O(n^2)$ may affect the efficiency of segmentation. Therefore, a sliding window is added to the video frame sequence to find the possible subshot boundary in a small range of the window.

Definition 1. The idea of Fisher linear discriminant is to project the data samples from multidimensional space into one-dimensional space, so that the variance between two classes is as large as possible, while the variance within class is as small as possible.

The two classes are set as C_1 and C_2 , and the sample is set as x . The mean value of the two classes is defined as m_1 and m_2 , and the variances are s_1 and s_2 . The ratio of interclass variance to intraclass variance is J , which is expressed as follows:

$$J = \frac{(m_2 - m_1)^2}{s_1^2 + s_2^2}. \quad (1)$$

The variance between different classes is made as large as possible, and the variance within class is made as small as possible, that is, to make the ratio J of the two largest. When J is maximum, the two classes are optimally divided in one-dimensional space.

The Fisher criterion is the ratio of interclass variance and intraclass variance. The partition that maximizes the Fisher criterion function in the sliding window is the local optimal solution, which can be considered as the subshot boundary [11]. The size of the sliding window should satisfy the assumption that the window does not span three subshots, that is, there is a two-class problem in the window. In this way, a multiclass problem can be transformed into multiple two-class problems, and the complexity and computational complexity can be reduced.

This method does not directly use Fisher linear discriminant to try to find the best partition in the sliding window, but naturally divides the window into two parts from the front. Because the frame difference between consecutive frames in the shot is relatively small, these two parts can be treated as classes. In the ideal state, the Fisher criterion inside the subshots is smaller and approaches zero, while the Fisher criterion at the boundary of the subshots is larger due to the motion of the camera or

the object. In fact, there is a local maximum caused by a small fluctuation in the subshots, which should be regarded as noise. A simple method to solve the noise problem is to select the local maximum higher than the mean value of the Fisher criterion as the subshots boundary (Algorithm 1).

The time complexity analysis of the subshot segmentation method based on sliding window is

$$\begin{aligned} T(N) &= O(N - 2M) * (O(M) + O(M)) + O(N - 2M) \\ &\quad + O(N - 2M), \\ &= O(2MN + 2N - 2M^2 - 4M), \end{aligned} \quad (2)$$

where since $2M \ll N$, the time complexity of the method is $T(N) = O(N)$.

2.1.3. Key Frame Extraction Based on a Fully Connected Graph. On the basis of subshot segmentation, key frames are extracted. The subshot is regarded as a fully connected graph, the frame is regarded as the vertex of the graph, and the frame difference is regarded as the edge of the connected vertex [12, 13]. In this way, the problem of key frame extraction can be transformed into the problem of finding the center of a fully connected graph.

Definition 2. Let G be an undirected connected graph with n vertices. The distance between vertex v_i and its farthest vertex is called the radius of v_i , denoted by $\rho(v_i)$. The following expression can be obtained:

$$r(v_i) = \max_{1 \neq jn} \{d(v_i, v_j)\}. \quad (3)$$

The vertex with the smallest radius is called the center of the graph, denoted as v_G . The radius of the undirected connected graph G is the radius of the graph, denoted as ρ_G . The following expression can be obtained:

$$r_G = \max_{1 \neq jn} \{d(v_G, v_j)\} = \min_{1 \neq jn} \{d(v_i, v_j)\}. \quad (4)$$

The center of the graph is the point with the smallest radius on the adjacency graph, and the corresponding frame is the frame with the smallest maximum frame difference from all other frames in the subshot, which can be considered as representing the content of the subshot and regarded as a key frame (Algorithm 2).

The time complexity analysis of the key frame extraction method based on subshot segmentation is shown as follows:

$$T(N) = O(N) * O(N) + O(N) * O(N) = O(2N^2) = O(N^2). \quad (5)$$

The time complexity of this method is $T(N) = O(N^2)$.

2.2. Recognition Method of Difficult Actions in Competitive Aerobics Video Based on Top-Down Feature Fusion. The recognition of difficult actions in competitive aerobics video is actually the classification of difficult actions in competitive

- (1) Enter the shot boundary until the end of the shot.
- (2) Calculate the average values of frames $i + 1$ to $i + M$ and frames $i + M + 1$ to $i + 2M$.
- (3) Calculate the variance of frames $i + 1$ to $i + M$ and frames $i + M + 1$ to $i + 2M$.
- (4) Use the Fisher criterion to find the mean value of J_i .
- (5) Output the boundary of subshot.

ALGORITHM 1: Subshot segmentation method based on the sliding window.

- (1) Enter the subshot boundary until the end of the shot
- (2) Calculate the frame difference between the frame i and frame j
- (3) Calculate the radius of the vertex i
- (4) Output subshot boundary

ALGORITHM 2: Key frame extraction method based on subshot segmentation.

aerobics video. The degree of feature expression on image or video content directly affects the effect of classification task. When the deep convolution neural network is used to complete the classification task in computer vision, the top-level feature map is usually used for classification [14]. Because the video content is dynamic and has the characteristics of multitarget and multiscale, based on the idea of feature fusion, a top-down-feature-fusion-based video difficult action recognition method of competitive aerobics is proposed to enhance the expression ability of the feature map to video information, so as to improve the accuracy of video difficult action recognition.

2.2.1. Construction of a Feature Pyramid. In the process of forward propagation of deep convolution neural network, each convolution layer outputs a certain number of feature maps, so there are a large number of high-level feature maps and low-level feature maps. In the process of network forward propagation, the scale of the feature graph changes after convolution pooling operation of some layers, while the scale of the feature graph does not change when it passes through other layers. In a feature pyramid network, the layer whose scale does not change is regarded as the same level, and on the contrary, the layer whose scale changes is regarded as different levels.

The deep convolution neural network can be divided into three different levels, and the scale of the characteristic graph is 28×28 , 14×14 , 7×7 . In the process of forward propagation, with the deepening of the network, the receptive field of the convolution layer is increasing, and the output feature map has more semantic information. Therefore, in each level, the high-level feature map has stronger information expression ability. The high-level feature map in each level constitutes the feature pyramid of the deep convolution neural network, and the feature map in the feature pyramid is taken as the feature map to be fused of the competitive aerobics video difficulty action recognition method based on top-down feature fusion [15–17]. The schematic diagram of the feature pyramid in the deep convolution neural network is shown in Figure 2.

Before feature fusion, the feature images to be fused should be unified to the same scale.

2.2.2. Structure Design of Top-Down Feature Fusion. The design idea is to unify the scale of the high-level feature map with that of low-level feature map through upsampling and fuse the upsampled high-level feature map with the corresponding low-level feature map [18]. The top-down feature fusion method is used to fuse the two scale feature maps, and the schematic diagram is shown in Figure 3.

In the figure, the block represents the output characteristic diagram of the corresponding layer. The feature of this structure is that the feature graph is upsampled and the short circuit structure is used. The feature fusion method is pixel-level addition.

(1) Upsampling Mode. The structure uses deconvolution as the upsampling method. The deconvolution operation is the inverse process of the convolution operation. If the convolution operation is regarded as a matrix transformation, the transformation matrix of the deconvolution operation is the inverse matrix of the transformation matrix of the convolution operation.

If the input feature diagram size i , convolution kernel size k , step s , and filling dimension p are set, the output dimension o of the feature graph is set; then, the relationship between them is

$$o = \left\lceil \frac{i + 2p - k}{s} \right\rceil + 1, \quad (6)$$

where $\lceil \cdot \rceil$ is the rounding operation.

The input and output of deconvolution correspond to the output and input of convolution, respectively. Suppose the size of input characteristic graph i' , convolution kernel size k' , step size s' , filling size p' , and output characteristic graph size o' of deconvolution operation; then, the relationship between them is described by

$$o' = (i' - 1) * s' + k' - 2p'. \quad (7)$$

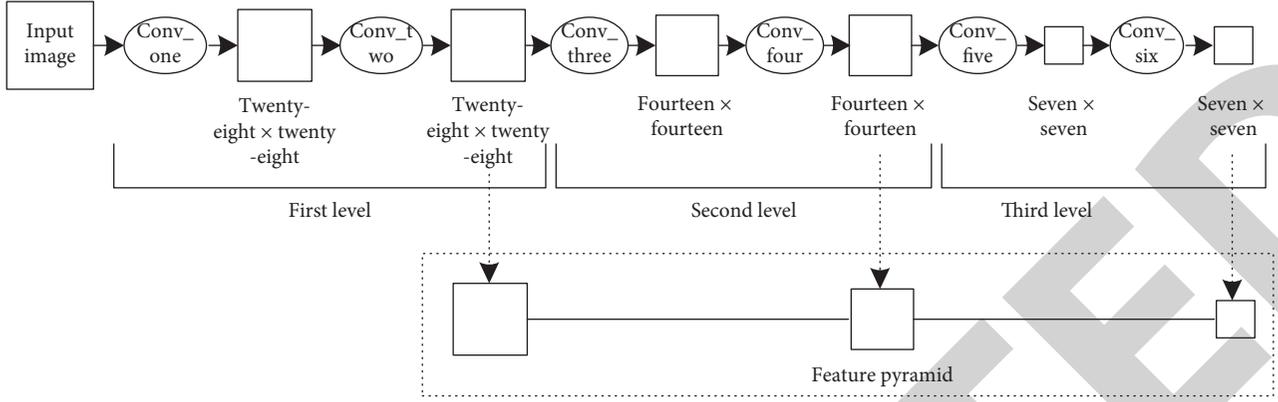


FIGURE 2: Schematic diagram of the feature pyramid in the deep convolution neural network.

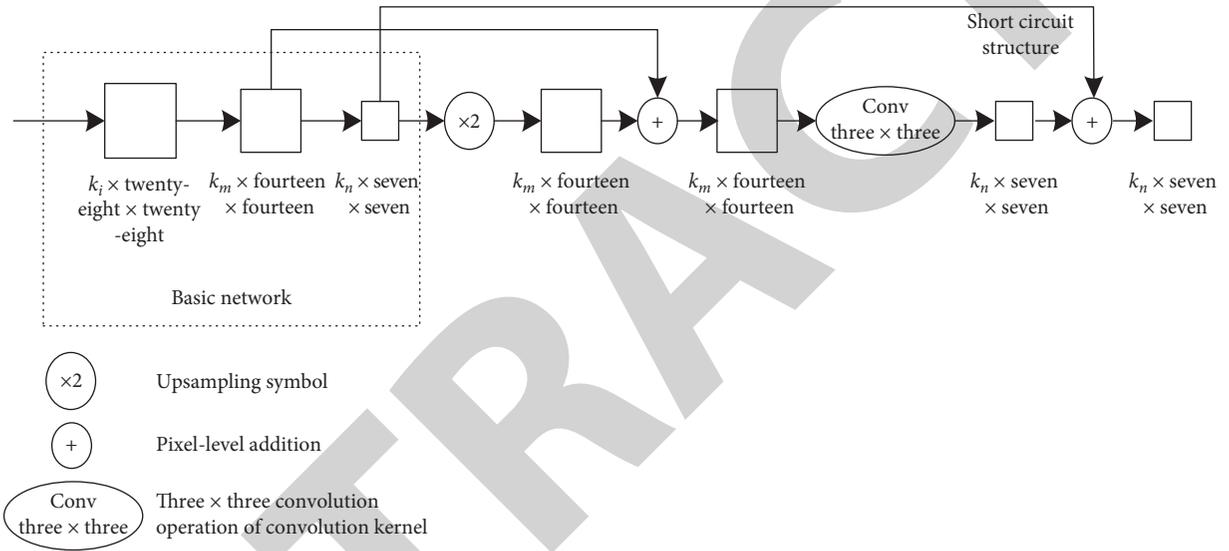


FIGURE 3: Schematic diagram of fusing two scale feature maps using the top-down feature fusion method.

(2) *Short Circuit Structure.* The main idea of the short circuit structure is that the output result is not only the result of forward propagation but also the input itself. The short circuit structure is used in this method, which can speed up the training speed on the one hand and improve the accuracy on the other hand.

The forward process of top-down feature fusion is as follows: let the low-level feature graph to be fused $P_1 \in R^{k_m \times 14 \times 14}$, the high-level feature graph be $P_2 \in R^{k_n \times 7 \times 7}$, P_2' be the feature graph obtained by P_2 deconvolution, and the parameter of deconvolution be W_i . Formula (8) describes the expression of P_2' :

$$P_2' = \text{Deconv}(P_2; W_i), \quad (8)$$

where $\text{Deconv}(\cdot)$ is the deconvolution symbol.

The feature diagram after fusion is set as $P_f \in R^{k_m \times 14 \times 14}$, and its expression is described by

$$P_f = P_1 + P_2'. \quad (9)$$

In order to maintain consistency with the output before fusion, a 3×3 convolution operation is adopted, and a short

circuit structure is used to avoid information loss. The output feature graph of top-down feature fusion is set as $P_f^v \in R^{k_n \times 7 \times 7}$, and its expression is described by

$$P_f^v = \text{conv3}(P_f) + P_2. \quad (10)$$

In the formula, $\text{conv3}(\cdot)$ is the convolution operation of the feature graph and $+$ is the polar addition of pixels [19].

2.2.3. Difficult Action Recognition Method of Competitive Aerobics Video. The difficult action recognition network of competitive aerobics video based on top-down feature fusion adopts a dual-stream network structure, in which an RGB image is the input mode of the space stream network and optical flow image or correction optical flow image is the input mode of the time stream network. The input modes of the time flow network and space flow network are different, but the network structure is the same [20, 21]. The top-down feature fusion part is added to the dual stream network of the video's difficult action recognition, and the difficult action recognition network of

competitive aerobics video based on top-down feature fusion is obtained.

Each video data is divided into three video segments, and the key frames of the RGB image and optical flow image are extracted from each video segment to represent the information of the video.

Each video segment will get the score vector relative to all categories through forward propagation. The score vectors

$$\begin{aligned} \text{Score}(T_1, T_2, \dots, T_k) &= \text{soft max}(G(F(T_1; W), F(T_2; W), \dots, F(T_k; W))), \\ G &= \sum_{k=1}^K \frac{F(T_k; W)}{K}, \end{aligned} \quad (11)$$

where T_k is the video segment sampled from S_k , RGB and optical flow images input to the network are sampled from T_k , $\{T_1, T_2, \dots, T_k\}$ is the sampling segment sequence of a long video; $F(T_k; W)$ represents the input video segment T_k through the function based on the feature fusion network with parameter W and returns a C -dimensional vector, representing the score of the video segment relative to all categories; C is the total number of categories; and $G(\cdot)$ returns a C -dimensional vector representing the final score of the video in all categories.

The loss function adopts cross-entropy loss, shown as follows:

$$L(y, G) = - \sum_{i=1}^C y_i \left(G_i - \log \sum_{j=1}^C \exp G_j \right), \quad (12)$$

where C is the total number of categories of recognition behavior and y_i is the label of the video.

In the process of back propagation, the model parameters can be optimized simultaneously according to multiple video segments. The gradient of loss function is as follows:

$$\frac{\partial L(y, G)}{\partial W} = \frac{\partial L}{\partial G} \sum_{k=1}^K \frac{\partial G}{\partial F(T_k)} \frac{\partial F(T_k)}{\partial W}, \quad (13)$$

where K is the number of video segments and W is the model parameter.

The RGB image and optical flow image are tested, respectively, to extract the key frames of the RGB image and optical flow image in the test video and get their scores relative to all categories through the test network. Also, the test results are averaged to get the score vector of each video. Then, the classification accuracy on the video set D is defined as

$$\text{Acc}(f; D) = \frac{1}{m} \sum_{i=1}^m \Pi(f(X_i) = Y_{\text{label}}), \quad (14)$$

where $\Pi(a = b) = \begin{cases} 1, & a = b, \\ 0, & a \neq b, \end{cases}$, $f(X_i)$ is the video prediction category and m is the total number of videos in video set D .

of three video segments are averaged to get the score vector of a complete video, and then, the score vector is normalized to get the probability score of probability distribution.

Let a video V be divided into K segments $\{S_1, S_2, \dots, S_k\}$ at equal intervals. The network model based on top-down feature fusion is shown as follows:

2.3. Online Arrangement of Difficult Actions in Competitive Aerobics Based on Multimedia Technology. Gait recognition, as a way to identify uncontrolled sexual characteristics, is an emerging biometric identification technology and was able to become one of the biometric identification technologies; From the uniqueness of each gait, from the perspective of gait anatomy, each person's body structure is different, the leg muscle length is also different, and the strength is also different. The height of the center of gravity and the sensitivity of the motor nerve together determine the uniqueness of the gait [22]. Pedestrians weight identification (PersonRe-identification), also calls the pedestrians to identify the abbreviation for ReID, is one of the research directions of pedestrians' intelligence cognition, especially in computer vision industry. ReID is the use of computer vision technology to judge whether there is a specific image or video in a sequence of pedestrians widely considered to be a child image retrieval problem given a monitoring pedestrian image, image retrieval across the equipment under the bank people [23]. Skeleton key point detection and attitude recognition generally identify several key points of the human body, such as head, shoulder, palm, and foot, used in the task of pedestrian pose recognition; these technologies can be applied in interactive entertainment scenes, similar to the human-computer interaction of Kinect, and key point detection technology is very valuable.

Based on the top-down feature fusion of the difficult action recognition method of competitive aerobics video, the online arrangement of difficult action of competitive aerobics based on multimedia technology is realized.

2.3.1. Determining the Purpose and Task. It is necessary to understand and master the requirements of competition rules. In addition, it is necessary to investigate the age, characteristics, physical condition, training level, and occupation of the users, so as to determine the task of online arrangement of difficult actions in competitive aerobics.

2.3.2. Drawing the Layout Plan. In order to understand the development level and trend of competitive aerobics at home

and abroad in time, according to the current popular innovative action, music melody, combination structure, and rule requirements, as well as the level of users, a set of overall scheme of online arrangement of difficult action in competitive aerobics is preliminarily designed, and special attention should be paid to determine their own unique style.

2.3.3. Choreography and Design Movements. Learning from other forms of dance, martial arts, gymnastics, skills, and movements, this paper comprehensively refines and recreates them, determines the main core movements of this set of competitive aerobics, and at the same time, pays attention to the symmetrical and balanced distribution of the whole set of competitive aerobics movements, which not only makes people feel coordinated but also ensures the balanced and comprehensive development of all parts of the body.

2.3.4. Arranging and Organizing the Sequence of the Whole Set of Actions. Generally, the structure of competitive aerobics can be divided into three parts: the first part is warm up, from small joint activity to large joint and large muscle group, from local body activity to whole body activity, from in-situ action to moving exercise, and mainly for stretching joint and spine pulling exercise; the second part is the main body action, generally starting from the far end of human body activity, namely, from the head, neck, hand, wrist, ankle, and knee joint gradually transitioning to the elbow, shoulder, waist, hip, and whole body comprehensive exercise. Competitive aerobics should arrange specific movements in this part to form the climax of the whole set of aerobics; the third part is to sort out, the range of movements in this part gradually becomes smaller and the speed slows down, and consciously arrange the relaxation of the body, so as to gradually reduce the frequency of heart rate and make the body return to normal.

In this paper, the convolutional neural network used for classification is transformed into a convolutional neural network used for coordinate estimation of nodes by changing the error function using the convolutional neural network and cascading ideas. At the same time, the method of cascading is adopted. Firstly, the coordinate of a node is obtained by preliminary calculation. Then, according to this coordinate, the local picture is reobtained from the original picture, and the coordinate calculation of higher progress is carried out with the local picture. The idea of cascading and local high-precision images to obtain higher-precision node coordinates is a place of innovation. However, the method of cascading and local high-precision images proposed in this paper is not applicable to the situation where the original image resolution is relatively small. Meanwhile, since the cascading method is adopted, the coordinate of each node is equivalent to doing multiple convolutional neural networks, and the computational complexity should be very high. Finally, the convolutional neural network used by the author is used for classification.

In the first stage, the first rough estimate of the last part of the attitude of the outline and then, in the next phase, the key position constantly optimizes the location of the key

points. The key point of each step of the way is used as forecast to cut out the neighborhood based on the key points, the image will be used for the rest of the network input, and the rest of the network will see higher resolution image, ultimately achieving better accuracy. Equation (15) is the model loss function.

$$\arg \min_{\theta} \sum_{(x,y) \in D^n} \sum_{i=1}^k \|y - \psi(x; \theta)\|^2. \quad (15)$$

In this paper, we utilize the entropy loss function to build the model for our research problems. It can be defined as follows: $\text{loss}(x, y) = \sum_{i=1}^n -p_i \log(1 - p_i)$, where x and y are represented as the real arts and crafts score and difficulty and y means the predicted score and difficult of our proposal. P_i means the probability of them when they are similar. The bigger the value of the loss, the worse our proposal performed. Also, our proposal is used to train a model that fit the real and predicted arts and crafts, so that the machine can assistant the arts and crafts designed. Compared with the existing methods, our proposal has three main advantages: firstly, it is less time consuming than others, which indicates that our proposal can translate as quick as possible and in time. Secondly, our proposal has a higher accuracy than others, which shows that our proposal can perform well. Thirdly, our proposal can be adapted to any situations with English and business. However, the main limitation of our proposal is that it needs a huge computation space, which indicates that our proposal has a strong requirement for computation and it may not be easy to realize.

3. Result Analysis

Taking different types of competitive aerobics video as the experimental object, this paper tests the effect of online arrangement of difficult actions in competitive aerobics. There are five types of competitive aerobics videos: men's single, women's single, mixed double, three, and six.

Also, we use the confuse matrix to evaluate the model performance, and the matrix can be defined as follows: accuracy means that the ratio of actions that we wanted and what we rejected and total actions are correctly recognized. Recall is the ratio of difficult actions correctly found and all difficult actions (Table 1).

$$\begin{aligned} \text{accuracy} &= \frac{TP + TN}{TP + TN + FP + FN}, \\ \text{recall} &= \frac{TP}{TP + FN}, \\ \text{precision} &= \frac{TP}{TP + FP}. \end{aligned} \quad (16)$$

To test the key frame extraction effect of different types of competitive aerobics video and design a comparative experiment, the human action recognition method based on the improved deep convolution neural network (referred to as the improved deep network recognition method) in [4] and the human behavior recognition method based on the

TABLE 1: Ratio of wanted actions to rejected actions and total actions.

Real Prediction	Positive	Negative
Positive	TP	FN
Negative	FP	TN

space-time characteristics of posture (referred to as the spation temporal characteristics of the posture recognition method) in [5] are selected as the comparison of the method in this paper. A, B, and C are set to represent the number of extracted key frames, the number of redundant key frames, and the accuracy rate, respectively. The key frame extraction results of different types of competitive aerobics video by three methods are described in Table 2.

According to the analysis of Table 2, the highest accuracy rates of the improved deep network recognition method and the attitude's space-time feature recognition method are 95.02% and 93.66%, respectively, the accuracy rate of the proposed method is always higher than 95%, and the highest is 98.21%; the average number of key frames of the improved deep network recognition method is 23, and the average number of key frames of the attitude's space-time feature recognition method is 27, while the average number of key frames of the proposed method is 21, and the number of key frames is the lowest. Compared with these data, it can be shown that the proposed method has better effect of key frame extraction in competitive aerobics video and can well maintain the main content of the video. The content of the key frame is not redundant, and the number is appropriate. It can provide scientific and accurate data support for the subsequent realization of difficult action recognition of competitive aerobics video.

To test the recognition effect of difficult actions in competitive aerobics video, the deep learning framework used in the experiment is the Caffe toolkit, and the dual flow network adopts three input modes, namely, the RGB image, optical flow image, and corrected optical flow image. The input video is divided into three segments, the key frames of the image are extracted, and the output results of all modes are fused to obtain the final recognition results. The recognition accuracy of difficult actions in competitive aerobics video by three methods is described in Figure 4.

According to the analysis of Figure 4, the recognition accuracy of the proposed method is 1.8%, 1.7%, and 0.9% higher than that of the improved deep network recognition method in the RGB image, optical flow image, and corrected optical flow image and 2.7%, 2.1%, and 2.8% higher than that of the attitude's space-time feature recognition method; the recognition accuracy of the proposed method is 2.7%, 2.1%, and 2.8% higher than that of the improved deep network recognition method. The results show that the recognition accuracy of the proposed method is in the highest value. This shows that this method has high accuracy in identifying difficult actions in competitive aerobics video and can better realize online arrangement of difficult actions in competitive aerobics.

In order to more accurately analyze the recognition effect of the difficult action recognition method for competitive

aerobics video based on top-down feature fusion and the other two methods on a single class, 10 kinds of movements are tested, including vertical split, horizontal split, split twist, front control leg balance, side balance, bent jump, split jump for a week, split jump into push up, kick jump, and kicking. The recognition accuracy of the three methods is described in Table 3.

From the analysis of Table 3, it can be concluded that the recognition accuracy of single action of the proposed method is higher than that of the other two methods. It can be seen that the recognition effect of difficult action of competitive aerobics video of the proposed method has significant advantages. Based on it, online arrangement of difficult action of competitive aerobics can greatly improve the reliability and applicability of online arrangement. The recall rate of different recognition methods for difficult action of competitive aerobics video is tested, and the comparison results of the three methods are described in Figure 5.

As can be seen from the experimental results in Figure 5, the recall rate of difficult action recognition in competitive aerobics video of the proposed method is significantly higher than that of the other two methods. Firstly, the recall rate of the proposed method is higher than 98%, and the change is stable; secondly, the recall rate of the improved deep network recognition method is stable in the range of [97%, 98%], and the fluctuation is small; the recall rate of the attitude's space-time feature recognition method is in the range of [94%, 98%], and the fluctuation is large. The comparison results show that the proposed method has higher recall rate and superior recognition performance.

To test the effect of online arrangement of difficult actions in competitive aerobics based on multimedia technology, 200 users of four age groups are randomly selected for satisfaction survey, namely, 50 people aged 20 and below, 50 people aged 21–30, 50 people aged 31–40, and 50 people aged 41 and above. The comparison results of the satisfaction rate of the four types of personnel to the three methods for online arrangement of difficult actions in competitive aerobics are described in Figure 6.

As can be seen from the experimental results in Figure 6, users of different age groups have the highest satisfaction rate with the online arrangement of difficult actions in competitive aerobics, especially for the users under 20 years of age, with the satisfaction rate as high as 97%; and the lowest satisfaction rate is for the attitude's space-time feature recognition method, especially for the users aged 21–30 years, with the satisfaction rate as low as 90%. Compared with these data, we can see that the proposed method has better online arrangement effect of difficult actions in competitive aerobics, more in line with the needs of users, and has strong practicability.

In the last three years, many methods are proposed to handle the online arrangement method of difficult actions; here, we introduced four outstanding methods proposed by Onan and Toçoğlu (Eichner) [24], Yang et al. [25], Zhang and Tao (Sapp) [26], and Hossain and Muhammad (MODEC) [27], which can be used to solve the related works

TABLE 2: Comparison of video key frame extraction results.

Video type		Men’s single sculls	Women’s single sculls	Mixed double	Three people	Six people
Method of this paper	A (frames)	10	18	25	16	36
	B (frames)	4	1	2	3	1
	C (%)	96.23	95.86	97.64	96.88	98.21
Improved deep network recognition method	A (frames)	15	22	31	19	28
	B (frames)	6	8	12	5	9
	C (%)	93.12	94.25	90.89	95.02	91.85
Recognition method of attitude’s space-time feature	A (frames)	21	32	22	15	45
	B (frames)	8	10	16	8	10
	C (%)	92.45	93.66	89.99	93.25	90.38

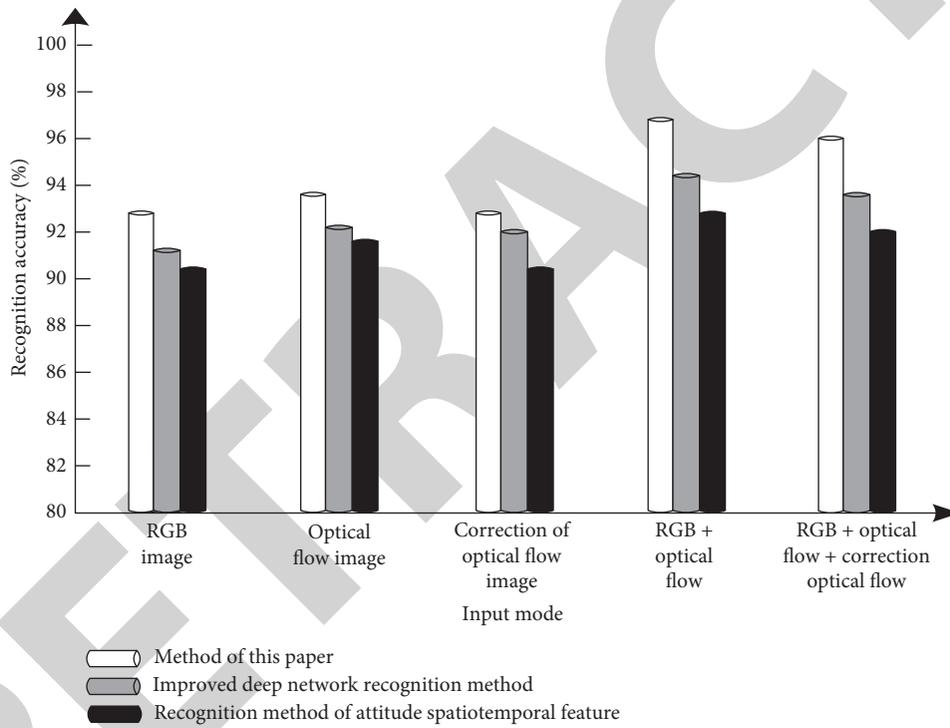


FIGURE 4: Comparison of recognition accuracy of difficult actions in video of mixed-pair competitive aerobics.

TABLE 3: Comparison of recognition accuracy of single action (%).

Single action	Method of this paper	Improved deep network recognition method	Recognition method of attitude’s space-time feature
Vertical split	98.23	97.26	96.36
Recumbent split	98.65	98.23	97.15
Split twist	97.36	97.31	96.33
Sagittal balance free support	98.55	98.32	97.25
Lateral balance	99.03	94.38	91.28
Bend jump	97.95	92.16	90.68
Split jump for one week	96.31	95.66	94.31
Split into push ups	95.98	91.89	90.59
Kick and jump	99.12	96.14	93.27
Pian kick	98.37	96.69	92.97

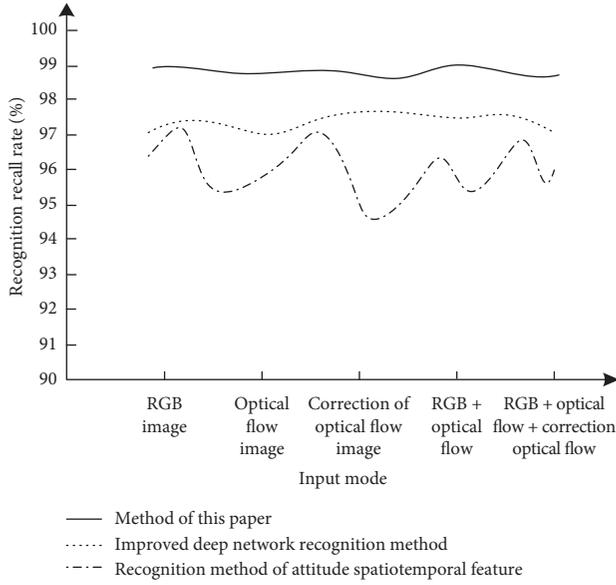


FIGURE 5: Comparison of the recall rate of difficult action recognition in competitive aerobics video.

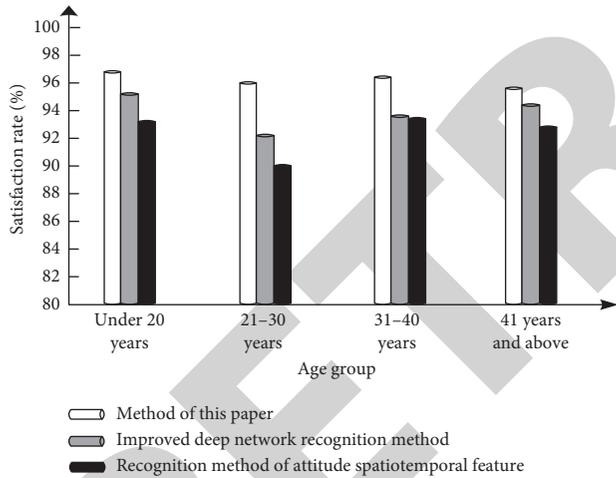


FIGURE 6: Comparison of the satisfaction rate of online arrangement of difficult actions in competitive aerobics.

taking different kinds of network structures. Eichner is a graph-based network that builds connections between different risk nodes. Also, the method of Yang et al. uses a specific loss structure to keep the similarity of real and predicted crafts design. Sapp is the basic model that needs more computation consuming to obtain the desiring performances as MODEC. However, these methods have their disadvantages, respectively. Eichner and the method of Yang et al. are too slow, Sapp is so complicated, and MODEC also needs more spaces.

In order to investigate the effectiveness of our proposal and other methods, here we take the F1-score into account to assess the experiment results, which can be defined as follows:

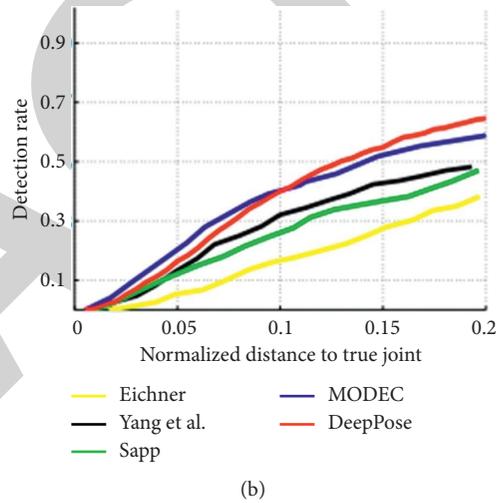
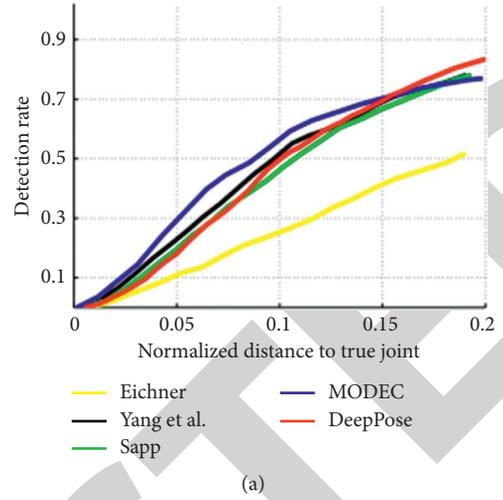


FIGURE 7: Percentage of detected joints (PDJs) on the Buffy dataset for two joints: elbow and wrist. The models have been trained on FLIC. We compare DeepPose after two cascade stages, with four other approaches. (a) Elbows. (b) Wrists.

$$f1 - score = \frac{2 * precise * recall}{precise + recall},$$

$$precise = \frac{TP}{TP + FP}, \quad (17)$$

$$recall = \frac{TP}{TP + FN}.$$

Also, we compared four methods proposed in the last three years with our proposal to investigate the effectiveness of our methods. The results are shown in Figure 7. As shown in Figure 7, the red curve represents our proposal, and on all datasets, we can see that our proposal is better than the others with x increased, while all methods obtain the same results. It indicated that our proposal can perform well than the other three methods. Figures 7 and 8 show the results of our proposal and the latest methods on different action datasets. Figure 7 shows that the detection changes with normalized distance between true joints and our proposal



FIGURE 8: Visualization of pose results on images from LSP. Each pose is represented as a stick figure, inferred from predicted joints. Different limbs in the same image are colored differently, and the same limb across different images has the same color.

(DeePoSE) beats others on wrists and obtains better results on elbows. Figure 8 shows the joint connection clearly.

4. Conclusions

With the establishment of international organizations and the improvement of competitive rules, competitive aerobics has gradually become a popular competitive sport. Difficult actions and difficult combination movements are the key to win in the competition. With the popularity of multimedia technology, video has gradually developed into one of the main data carriers. This paper studies the online arrangement method of difficult actions in competitive aerobics based on multimedia technology, accurately identifies the human movements in competitive aerobics video, and compiles scientific and reasonable difficult actions in competitive aerobics on this basis. The experimental results effectively verify the online arrangement effect of difficult actions in competitive aerobics and provide the reference value for the whole set of movements arrangement and training of coaches in China. Although our online scheduling model can provide the choreography of difficult movements, some combinations with too high difficulty for our athletes are not applicable, so how to find can either pursue high and the actual use of layout, make our future research focused, and make intelligent online scheduling more targeted. In the case of low resolution, how to use the cascade method to obtain higher node coordinate progress

and how to reduce the complexity of the convolutional neural network at the cost of less complexity are all problems that must be solved [1].

Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest

The author declares that there are no conflicts of interest.

References

- [1] C.-T. Liu, P.-C. Shih, Z.-H. Cai et al., "Rotor conductor arrangement designs of high-efficiency direct-online synchronous reluctance motors for metal industry applications," *IEEE Transactions on Industry Applications*, vol. 56, no. 4, pp. 1337–1344, 2020.
- [2] R. Casanova, V. Miller, J. Cheon et al., "Field test results of sex- and gender-specific health multimedia case-based learning modules," *Journal of Women's Health*, vol. 28, no. 12, pp. 1755–1761, 2019.
- [3] D. M. Hull, P. B. Lowry, J. E. Gaskin, and K. Mirkovski, "A storyteller's guide to problem-based learning for information systems management education," *Information Systems Journal*, vol. 29, no. 5, pp. 1040–1057, 2019.

- [4] S. D. Chen, W. Wei, B. Q. He, S. Y. Chen, and J. Y. Liu, "Action recognition based on improved deep convolutional neural network," *Application Research of Computers*, vol. 36, no. 3, pp. 945–949, 2019.
- [5] X. Zheng, X. D. Peng, and J. X. Wang, "Human action recognition based on pose spatio-temporal features," *Journal of Computer-Aided Design & Computer Graphics*, vol. 30, no. 9, pp. 29–38, 2018.
- [6] A.-P. Correia, N. Koehler, A. Thompson, and G. Phye, "The application of phet simulation to teach gas behavior on the submicroscopic level: secondary school students' perceptions," *Research in Science & Technological Education*, vol. 37, no. 2, pp. 193–217, 2019.
- [7] T. H. Wang, C. H. Kao, and Y. L. Dai, "Developing a web-based multimedia assessment system for facilitating science laboratory instruction," *Journal of Computer Assisted Learning*, vol. 35, no. 4, pp. 529–539, 2019.
- [8] E. Alemdag and K. Cagiltay, "A systematic review of eye tracking research on multimedia learning," *Computers & Education*, vol. 125, no. 21, pp. 413–428, 2018.
- [9] Y. Xu, D. Li, Z. Wang, Q. Guo, and W. Xiang, "A deep learning method based on convolutional neural network for automatic modulation classification of wireless signals," *Wireless Networks*, vol. 25, no. 7, pp. 3735–3746, 2019.
- [10] Z. Q. Xu and T. Zhang, "Simulation of accurate classification method for digital multimedia scene images," *Computer Simulation*, vol. 36, no. 7, pp. 385–388, 2019.
- [11] F. Zhang, N. Cai, J. Wu, G. Cen, H. Wang, and X. Chen, "Image denoising method based on a deep convolution neural network," *IET Image Processing*, vol. 12, no. 4, pp. 485–493, 2018.
- [12] G. Jian and Y. Mengfei, "Evolutionary fault tolerance method based on virtual reconfigurable circuit with neural network architecture," *IEEE Transactions on Evolutionary Computation*, vol. 22, no. 6, pp. 949–960, 2018.
- [13] X. Bai, N. Liao, W. Wu, and W. Song, "Color reproduction method based on neural networks and visual matching," *Applied Optics*, vol. 59, no. 35, Article ID 11027, 2020.
- [14] M. Zheng and S. Kahrizi, "Protein molecular defect detection method based on a neural network algorithm," *Cellular and Molecular Biology*, vol. 66, no. 7, pp. 76–83, 2020.
- [15] R. Cui, A. Zhu, J. Wu, and G. Hua, "Skeleton-based attention-aware spatial-temporal model for action detection and recognition," *IET Computer Vision*, vol. 14, no. 5, pp. 177–184, 2020.
- [16] C. Calmels, M. Elipot, and L. Naccache, "Probing representations of gymnastics movements: a visual priming study," *Cognitive Science*, vol. 42, no. 5, pp. 1529–1551, 2018.
- [17] X. W. Chen, M. X. Yuan, H. Guo, and Z. Zhu, "Hexagonal arrangement of phospholipids in bilayer membranes," *Chinese Physics B*, vol. 29, no. 3, pp. 166–170, 2020.
- [18] M. Mack, "Exploring cognitive and perceptual judgment processes in gymnastics using essential kinematics information," *Advances in Cognitive Psychology*, vol. 16, no. 1, pp. 34–44, 2020.
- [19] Y. F. Zhao, K. Xie, Z. Z. Zou et al., "Intelligent recognition of fatigue and sleepiness based on InceptionV3-LSTM via multi-feature fusion," *IEEE Access*, vol. 8, Article ID 144205, 2020.
- [20] H. Root, A. N. Marshall, A. Thatcher, A. R. S. Valier, T. C. Valovich McLeod, and R. C. Bay, "Sport specialization and fitness and functional task performance among youth competitive gymnasts," *Journal of Athletic Training*, vol. 54, no. 10, pp. 1095–1104, 2019.
- [21] A. Banerjee and A. Roychoudhury, "The gymnastics of epigenomics in rice," *Plant Cell Reports*, vol. 37, no. 1, pp. 25–49, 2018.
- [22] J. Li, F. Ren, S. Nishide et al., "Character action recognition based on deep convolutional neural network and action sequence," in *Proceedings of the 2019 IEEE 6th International Conference on Cloud Computing and Intelligence Systems (CCIS)*, IEEE, Singapore, April 2019.
- [23] M. A. Cui-Hong, Z. Q. Mao, J. L. Cui et al., "Action recognition based on deep LSTM and two-stream fusion network," *Computer Engineering and Design*, vol. 8, 2019.
- [24] A. Onan, "A term weighted neural language model and stacked bidirectional lstm based framework for sarcasm identification," *IEEE Access*, vol. 9, pp. 7701–7722.
- [25] S. Yang, Z. Gong, K. Ye, Y. Wei, Z. Huang, and Z. Huang, "EdgeRNN: a compact speech recognition network with spatio-temporal features for edge computing," *IEEE Access*, vol. 8, Article ID 81468, 2020.
- [26] J. Zhang and D. Tao, "Empowering things with intelligence: a survey of the progress, challenges, and opportunities in artificial intelligence of things," *IEEE Internet of Things Journal*, vol. 8, no. 10, pp. 7789–7817, 2021.
- [27] M. S. Hossain and G. Muhammad, "An audio-visual emotion recognition system using deep learning fusion for a cognitive wireless framework," *IEEE Wireless Communications*, vol. 26, no. 3, pp. 62–68, 2019.