Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 7217383, 11 pages
https://doi.org/10.1155/2021/7217383

Research Article

An Empirical Study on Sports Combination Training Action Recognition Based on SMO Algorithm Optimization Model and Artificial Intelligence

Hecai Jiang [1,2] **and Sang-Bing Tsai** [3]

[1]*Xiangsihu College of Guangxi University for Nationalities, Guangxi 530008, China*
[2]*Guangxi University for Nationalities, Guangxi 530008, China*
[3]*Regional Green Economy Development Research Center, School of Business, Wuyi University, Nanping, China*

Correspondence should be addressed to Hecai Jiang; 594316523@qq.com and Sang-Bing Tsai; sangbing@hotmail.com

Received 1 June 2021; Revised 16 June 2021; Accepted 2 July 2021; Published 12 July 2021

Academic Editor: Chenxi Huang

In order to improve the accuracy of sports combination training action recognition, a sports combination training action recognition model based on SMO algorithm optimization model and artificial intelligence is proposed. In this paper, by expanding the standard action data, the standard database of score comparison is established, and the system architecture and the key acquisition module design based on 3D data are given. In this paper, the background subtraction method is used to process the sports video image to obtain the sports action contour and realize the sports action segmentation and feature extraction, and the artificial intelligence neural network is used to train the feature vector to establish the sports action recognition classifier. This paper mainly uses a three-stream CNN artificial intelligence deep learning framework based on convolutional neural network and uses a soft Vlad representation algorithm based on data decoding to learn the action features. Through the data enhancement of the existing action database, it uses support vector machine to achieve high-precision action classification. The test results show that the model improves the recognition rate of sports action and reduces the error recognition rate, which can meet the online recognition requirements of sports action.

1. Introduction

Sports action recognition is a process of recording human movement by tracking the movement of some key points in the time domain, and converting it into a mathematical way to express the movement, which is of great significance for competitive training and national fitness [1]. Traditional motion recognition technologies include mechanical, acoustic, electromagnetic, and optical. Mechanical technology uses external sensors and rigid supports, which will affect limb movement. However, acoustic and electromagnetic technologies are vulnerable to external environment interference, large time delay, and low test accuracy [2, 3]. The traditional optical technology is more accurate, but it also has the disadvantages of high price and long time-consuming data processing. In recent years, inertial measurement technology based on accelerometer, gyroscope, and magnetometer and wearable technology based on EMG have developed rapidly, but there are also some limitations; that is, equipment may affect sports performance and cannot be used in the competition environment [4].

Computer vision uses camera and computer to capture, track, and measure the target and realizes automatic action recognition through artificial intelligence algorithm, which breaks through many limitations of traditional action recognition technology [5, 6]. In 2019, Australian scholar Cust and others summarized the application of machine learning in action recognition and systematically summarized the application of support vector machine (SVM), convolutional neural network (CNN), and other algorithms in computer vision. It is noteworthy that some new attitude estimation

algorithms are emerging. Openpose is one of the most commonly used multiperson pose estimation algorithms [7]. It uses the bottom-up method to detect the key points of all people in the image and then assign the detected key points to each corresponding person. RMPE is a top-down attitude estimation algorithm, which extracts high-quality single person regions from inaccurate candidate frames by using SSTN [8, 9]. As a popular semantic and instance segmentation architecture, mask RCNN can simultaneously predict the position of candidate frames of multiple objects in the image and segment the mask of their semantic information, so as to determine the position of each person, and then recognize the human motion posture through the position information and feature point set [10].

In order to solve the shortcomings of current sports action recognition methods and obtain better sports action recognition effect, a sports action recognition method based on feature reduction and Gaussian mixture model is proposed. Firstly, the feature vector of sports action is extracted, and then the dimension of the feature vector is reduced by using the random projection algorithm. Finally, the Gaussian mixture model is used to learn the reduced training samples, and the sports action recognition model is constructed. Based on SMO algorithm optimization model and artificial intelligence, this paper proposes a three-stream CNN deep learning framework based on convolutional neural network and uses a soft Vlad representation algorithm based on data decoding to represent the learned action features. Through the data enhancement of the existing action database, the support vector machine is used to achieve high-precision action classification. The test results show that this method speeds up the training speed of the classifier and improves the recognition accuracy of sports actions.

2. Action Recognition System of Sports Combination Training

2.1. System Architecture. The system architecture is mainly divided into application layer, data processing layer, communication layer, data acquisition layer, and hardware layer, as shown in Figure 1. The hardware layer includes physical power supply, sensor, and depth camera for 3D data acquisition. The data acquisition layer is mainly to classify, integrate, and analyze the data collected in the physical layer. The communication layer uses the protocol to upload the data at the bottom layer, while the data processing layer at the top layer mainly guarantees the order and preservation of the data and carries out some preprocessing to filter out the useless data. The application layer is the core of the system, which mainly calculates and processes the data, uses the database for action recognition, and assists in action scoring. Among them, responsible for data acquisition and action recognition, these two layers are the focus of this paper; the next two layers are introduced.

2.2. Collection of Action Data. In order to meet the needs of action recognition, people put forward many methods and collect data based on different kinds of sensor systems: marker based system, laser range finder, structured light, Microsoft Kinect sensor, multicamera system, and so on [11]. In order to better capture human bone and joint data, this paper uses Microsoft Kinect sensor for motion data acquisition. The acquisition process can be divided into three parts: depth image, human body part, and 3D joint modeling.

With the launch of Kinect, depth imaging technology has made significant progress. In this paper, the Kinect camera is used to capture 640×480 images at 30 frames per second, and the depth resolution is only a few centimeters. Compared with traditional intensity sensors, depth cameras work at low light levels, providing calibrated scale estimation (color and texture invariance) and resolving contour blur in position pose. The most important thing is that it can directly synthesize the real human depth image, so as to build a large training data set cheaply [12, 13].

In this paper, several local body part tags are defined, which are densely covered on the body. Among them, some parts are defined as directly locating the specific bone joints of interest. The other parts are used to fill in the blank or combine to predict other joints. In this experiment, 31 body parts were used: Lu/Ru/LW/RW head, neck, L/R shoulder, Lu/Ru/ LW/RW arm, L/R elbow, L/R wrist, L/r hand, Lu/Ru/ LW/RW trunk, Lu/Ru/LW/RW leg, L/R knee, L/R ankle, and L/R foot (left and right, up and down).

A simple operation in 3D modeling is to use the known calibration depth for modeling. However, edge pixels seriously degrade the quality of estimation. Therefore, this paper adopts a local pattern finding method based on weighted Gaussian kernel. The density estimates for each body part are defined as

$$g(y) = \sum_{k=1}^{S} \exp \frac{y - y_k}{w_k}, \tag{1}$$

where y represents the three-dimensional coordinates, and s is the pixel. y_k is the projection of the pixel, BC is the learning bandwidth, and w_k is the pixel weight, which is expressed as the product of the square of the given depth and the posterior probability. The final confidence estimation is the sum of the pixel weights of each mode, which is more reliable than the modal density estimation. The detected modes are located on the surface of the object, resulting in a 3D joint model.

2.3. Feature Extraction of Action Data. This section will introduce an aerobics action recognition method based on bone joint and depth features. Among them, the feature of bone and joint is extremely important to distinguish aerobics action, and it requires the robustness of perspective. Therefore, the position feature is selected as the bone joint feature, and the specific calculation formula is as follows:

$$f_{j,s}^{k} = \left(P_j^k + P_s^k\right)\left(P_j^k - P_s^k\right), \tag{2}$$

where P_j^k represents the coordinates of the ith joint in the t-th stream data, and JTK is the same.

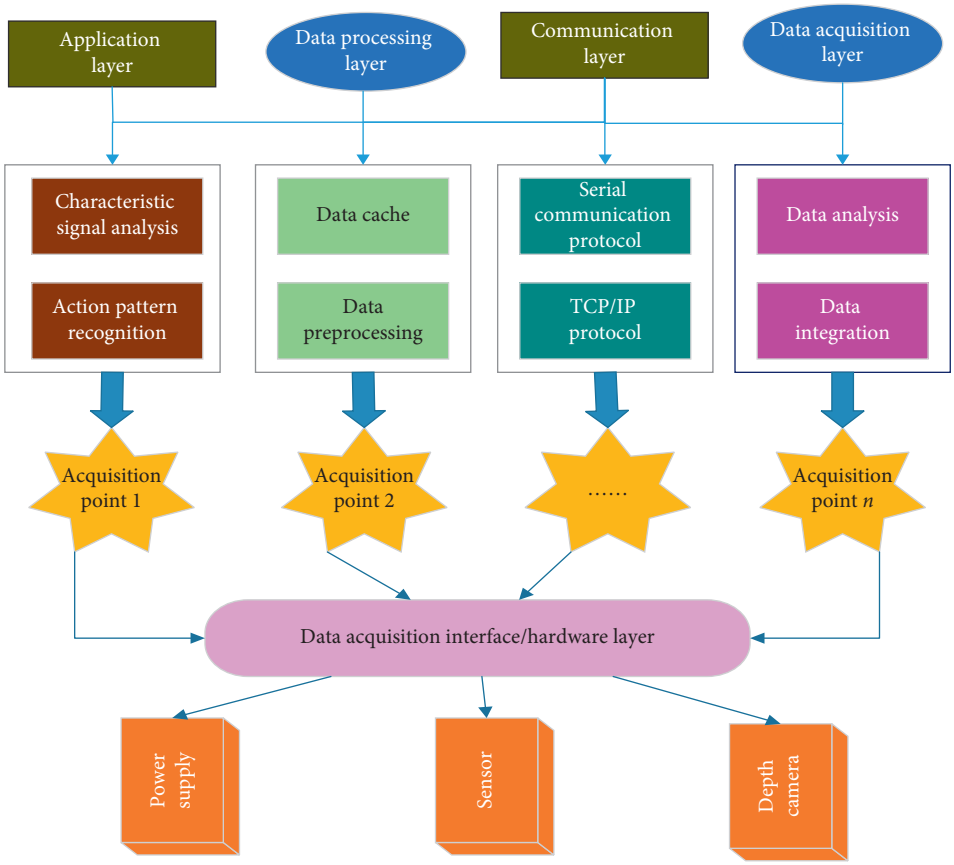

Figure 1: System architecture diagram.

For local pattern feature extraction, the depth image is mapped to 4D space. If there is data in the space, it is 1; otherwise, it is 0. Generally, the number of 1 is much less than that of 0, so the local pattern features can be considered as sparse. The specific operation is to segment the depth image at each joint into a unit cube and calculate the number of pixels falling into each cube. Then, the characteristics of the joint region are calculated by using a standardized function (for example, sigmoid function)

$$\lambda = \frac{1}{1 + e^{-\delta \sum \Re}}, \tag{3}$$

where λ is the local pattern eigenvalue, δ is a constant, and UI indicates whether the ith pixel has data in 4D space. It exists as 1; otherwise, it is 0.

Because the data collected based on Kinect will be more or less interfered into by external factors, in order to more accurately identify Aerobics movements, this paper uses three-layer Fourier space-time pyramid technology to eliminate these interferences. The first layer performs fast Fourier transform on the data frame to obtain the low-frequency coefficients of the data frame. In the middle layer, the half frame is transformed by fast Fourier transform to get the low-frequency coefficients. In the top layer, half of the half frame is transformed by fast Fourier transform to get the low-frequency coefficients of smaller data segments, and these coefficients are connected to get the final feature vector for feature fusion.

In this paper, canonical correlation analysis is used to fuse joint features and local pattern features, so as to reduce the number of unknown variables and retain the required information. Let $u = ax$ be the feature matrix of bone joint position, and let $V = \text{by}$ be the feature matrix of depth local pattern

$$\text{corr}(x, y) = \text{cov}\frac{(ax, by)}{\left(\delta_x \delta_y\right)}. \tag{4}$$

Therefore, we need to get a set of x, y to maximize the correlation coefficient, so as to get the optimal coefficients a, b.

The covariance matrix of the combination vector of X and Y is defined as Π. Then, it is expressed as

$$\Theta = \begin{bmatrix} P \\ Q \end{bmatrix} = \begin{bmatrix} \Theta_{11} & \Theta_{12} \\ \Theta_{21} & \Theta_{22} \end{bmatrix}. \tag{5}$$

Among them Θ_{11}, Θ_{22} are the auto covariance matrix, and Θ_{12}, Θ_{21} are the covariance matrix. The feature matrix is used for feature fusion:

$$U = \begin{bmatrix} a \\ b \end{bmatrix} \begin{bmatrix} \Theta_{11} & \Theta_{12} \\ \Theta_{21} & \Theta_{22} \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix}. \tag{6}$$

Finally, this paper uses support vector machine to classify and compare the fused motion features. Let the samples provided to SVM be represented as $D_i = (x_i, y_i)$, where x and y represent feature vectors and action tags, respectively. Then, the distance between the sample and the hyperplane is expressed as

$$\delta_i = (wx_i + b)y_i + \gamma_i. \tag{7}$$

In order to minimize the number of misclassification, it is necessary to maximize the geometric interval, so it can be minimized to maximize the geometric interval. In order to solve the influence of a small amount of noise on the performance of support vector machine, a relaxation variable is introduced to minimize the $\|w_i\|^2$ ensure the accuracy. The final optimization problem is expressed as

$$\min \frac{1}{2}\|w\|^2 + \mu \sum \delta_i, \tag{8}$$

where I is the number of samples, and μ is the penalty factor, which is used to characterize the importance of outliers. The larger the value is, the greater the loss is. For relaxation variables, only outliers are set here. The larger the outlier, the farther the outlier. After feature classification and recognition by using support vector machine, the difficulty and completion degree of Aerobics corresponding actions are scored by comparing with the corresponding database. Therefore, the development of classifiers more suitable for complex signal classification is an important research direction in the field of action recognition and artificial intelligence in the future.

3. Action Recognition Algorithm Based on Combination of SMO and Artificial Intelligence

3.1. Optimization Model of SMO Algorithm. Suppose that there is a training set X with L samples in total. In this way, to solve a binary classification problem is to find a function that can make all points on X be divided into two parts.

Today's classification problems can be divided into three categories: linear hard separable, linear soft separable, and nonlinear soft separable. As the name suggests, linear hard separability is to accurately find a classification hyperplane, which can completely separate data sets. The linear soft separability is a little more complex than the linear hard separability, which refers to the following situation: a given set of data sets to be classified originally corresponds to the linear hard separable data model, but because, in the process of talent cultivation, the data samples may be affected by some inevitable noise or other interference factors [14, 15]. Therefore, there is no way to find a classification hyperplane that can make the empirical risk zero, or even if it is found, the real classification hyperplane corresponding to the inherent mechanism of the data model deviates seriously. However, because the data model corresponding to the data set can be linearly separable in essence, we should be able to find some kind of machine learning mechanism in theory and get some approximation of the real classification hyperplane [16].

The most common classification problem is "Nonlinear Soft separable problem"; that is, the data cannot be classified in low dimensional space. We need to use some method to transform it into a high-dimensional space, which is transformed into a linear soft separable problem in high-dimensional space. Then, according to the machine learning mechanism to solve the linear soft separable problem, we can get some approximation of a real classification hyperplane. Figure 2 shows a schematic of a nonlinear soft separable dataset.

Since the emergence of support vector machine, its solving problem has been a hot topic of the majority of scholars, and many solving methods have emerged, among which SMO algorithm is one of the best solving methods. It decomposes the quadratic programming problem in the process of SVM solution into a series of small quadratic programming problems with only two variables [17–23]. These problems can be solved analytically, thus avoiding the solution of quadratic programming and the accuracy of numerical value.

SMO algorithm is mainly divided into two parts: the first step is to solve the quadratic optimization problem with only two variables analytically. The second step is to select two Lagrange multipliers to be solved. If we want to improve the efficiency of the whole SVM, we need to improve in these two aspects. SMO algorithm can determine the best threshold and connection weight of BPNN, obtain higher recognition rate than other sports action recognition types, shorten the execution time, and accelerate the speed of sports action recognition.

Wss-k is used to select two Lagrange multipliers to be optimized. The quadratic optimization problem with two variables is as follows:

$$\min K(\alpha, \beta) = -\left(m\alpha^2 + n\beta^2 + p\alpha\beta\right). \tag{9}$$

Using the relationship between α and β and mathematical knowledge, we can get the optimized α and β and then use the constraint conditions to constrain them to complete the whole optimization process. Every iteration in the training process will make the objective function value decrease until the objective function no longer decreases and stop the whole training process.

3.2. Construction of Classifier for Sports Action. With the improvement of computer computing ability, deep learning framework has been widely used in the field of machine vision in recent years. Among the existing deep learning frameworks, convolutional neural networks (CNNs) are extremely suitable for feature extraction and classification of images and videos because of its convolution and pooling operation on each layer of neurons. Its application in the field of action recognition has been paid more and more attention by researchers at home and abroad (see Figure 3).

In this paper, a convolutional neural network (three-stream CNN) algorithm including three channels of space, local time domain, and global time domain is proposed to deeply characterize and recognize human actions. The flow chart is shown in Figure 3. Firstly, the feature map of action

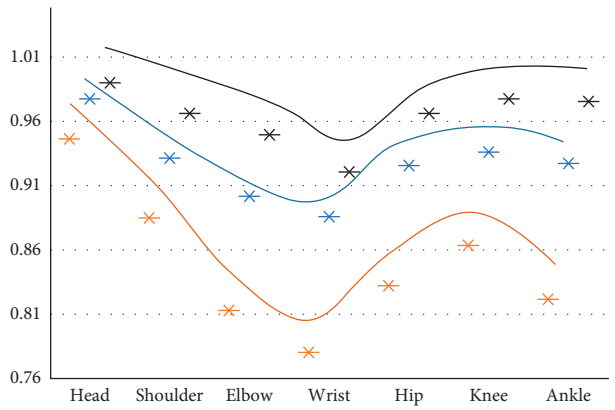

Figure 2: Schematic diagram of nonlinear separable data.

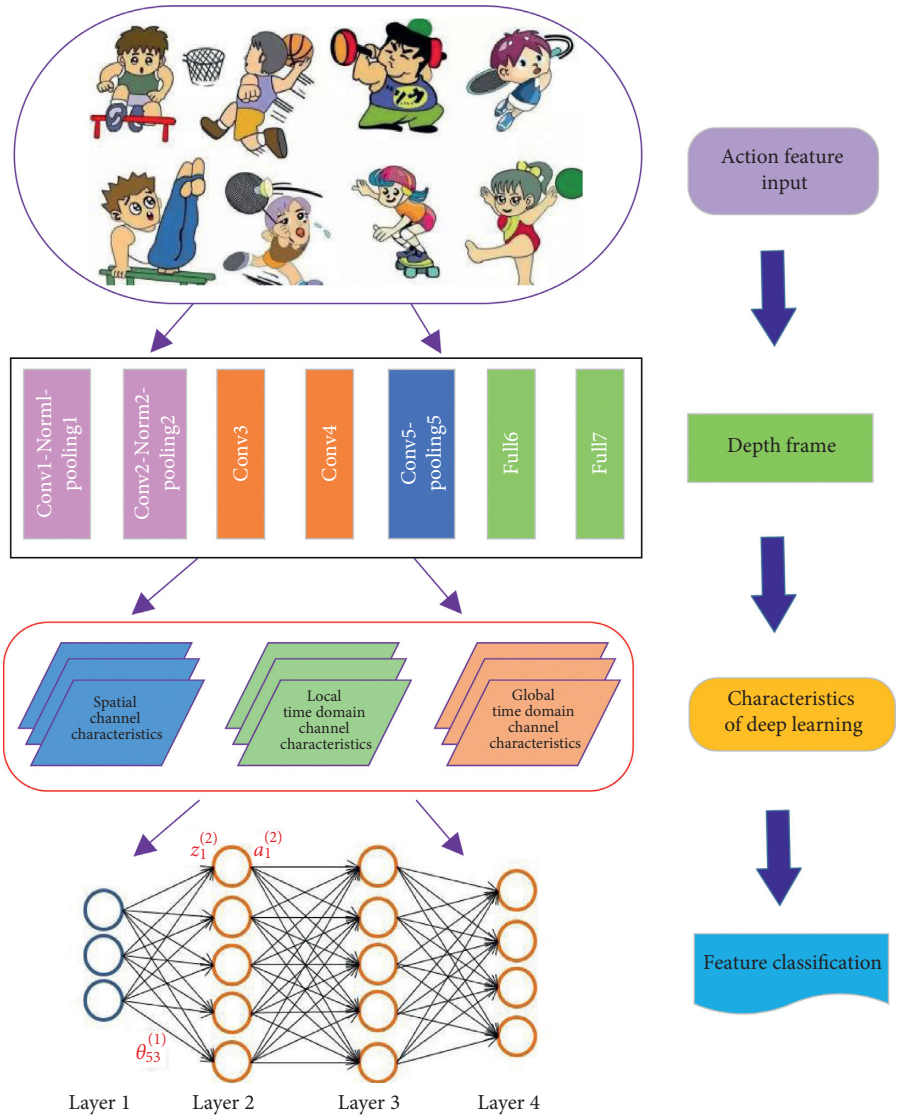

Figure 3: Action recognition flow chart based on three channel convolutional neural network.

image sequence is constructed as the input of three-stream CNN framework. The three-stream CNN framework consists of five convolution layers (conv1–5) as shown in Figure 3. The two convolution layers are normalized (norm1 and norm2) and connected to three pooling layers (pooling1, pooling2, and pooling5), and the last pooling layer (pooling5) is connected to two fully connected layers (full6 and full7). Three channels (spatial channel, local time-domain channel, and global time-domain channel) are used to obtain depth features after deep learning. Among them, the spatial channel CNN performs deep learning on the motion image, the local time-domain channel CNN performs deep learning on the optical flow features, and the global time-domain channel CNN performs deep learning on the motion differential image product (MSDI) proposed in this paper.

Most of the methods based on CNN directly segment the action video into independent images, then construct the spatial CNN framework of the image, and complete the classification of the whole video by fusing the classification results of the action image. However, action is a three-dimensional space-time quantity, and ignoring the time-domain of action is simple and convenient, but it cannot achieve good results in theory and practical application. Another action recognition strategy based on CNN is to use 3D convolution operator to convolute and pool the action video, but the accuracy still can not achieve the expected results. In order to overcome the shortcomings of traditional two-stream CNN in the representation of temporal and spatial relationship of actions, this paper proposes a global feature extraction method of action depth: three-stream CNN. At the same time, in order to reduce the over fitting error caused by small data samples, this paper proposes a data enhancement algorithm based on human action database.

Based on the action decomposition hypothesis of two-stream CNN, three-stream CNNs further decompose the action into three channels: spatial, local, and global. The input of space channel adopts action static image. The local time-domain channel adopts the optical flow characteristics. The difference is that the optical flow of three-stream CNNs adopts the optical flow algorithm and calculates the optical flow in RGB color space. The global time-domain channel input uses a motion stacked difference image (MSDI) based on the action history image. The frame structure of three-stream CNN is shown in Figure 4 below, in which the input layers are motion image frame, optical flow feature map, and MSDI feature map, respectively. The input layer signal is first connected to a convolution layer, and the convolution kernel size is set to 7×7, the step size is set to 2, and convolution is performed on 96 channels. The first convolution layer is connected with a pooling layer, the pooling core window size is set to 3×3, and the step size is set to 2. After that, four convolutions and two pooling processes are repeated, and the corresponding parameters are given in Figure 4. Finally, the last pooling layer is connected to two fully connected layers with 4096 and 2048 neurons (see Figure 3).

Given a video sequence V, the image frame is fully sampled and adjusted to 224×224 as the input of space channel CNN. After that, the adjusted image matrix signal is transmitted to the convolution layer. When convoluting an image, if the feature (brightness or optical flow) of an image is expressed as $g(x, y)$, when the kernel operator is $h(s, t)$, the convolution result is as follows:

$$f(x, y) = \sum g(x - s, y - t)h(s, t). \tag{10}$$

In three-stream CNNs, the convolution kernel operator is chosen as 7×7, 5×5 and 3×3. The step size of the first and second convolution layers is set to 2, and the step size of other convolution layers is set to 1. When convoluting different image feature matrices, the step size and convolution kernel size have a direct impact on the convolution results: when the step size is 1, the convolution kernel is 3×3 size matrix. Assuming that the image is a two-dimensional matrix shown on the upper part in Figure 5 and the kernel convolution is a two-dimensional matrix shown on the right, the convolution process and results are shown in Figure 5. The image convolution operation process is the product and accumulation process of the convolution kernel and the corresponding module of the image, and the result of the convolution operator is determined by the image quality, the step size, and the size of the convolution kernel operator, so the execution efficiency is high.

When the pooling layer is operated, the pooling operator is defined as 3×3, the first two pooling layers' step size is set to 2, and that of the rest is set to 1. The maximum pooling operation is selected for the three-stream CNN pooling operation; that is, the maximum value is selected within the size range of the pooling operator, as shown in Figure 5. It can be seen that the size of pooling operator directly determines the result of pooling process, and its selection plays an important role in the whole deep learning process. In addition, mean pooling is also widely used in deep learning framework.

In the local time domain channel, the optical flow characteristic diagram is used as the input, and the deep learning is carried out according to the deep learning framework. When constructing the optical flow feature map, for the adjacent frames I1 and I2 in the given image sequence, all the imaging values of the image in RGB space are reserved; that is, D is taken as 3. At the same time, x is used to represent the imaging value of a pixel in the image space, and w is used to represent its optical flow. By constraining in color, gradient, and velocity space, the following constraint equation can be constructed:

$$CT = C_{\text{col}} + \gamma C_{\text{gra}} + \beta C_{\text{sm}}. \tag{11}$$

When calculating the input characteristic (MSDI) of the global time-domain channel, $\varphi(x, y, t)$ is used to represent the brightness value of (x, y) pixels in the time t image. First, the difference is made between the images, and its absolute value $\phi(x, y, t)$ is taken:

$$\phi(x, y, t) = |\varphi(x, y, t) - \varphi(x, y, t - 1)|. \tag{12}$$

Then, the global feature $E(x, y, k)$ is characterized as

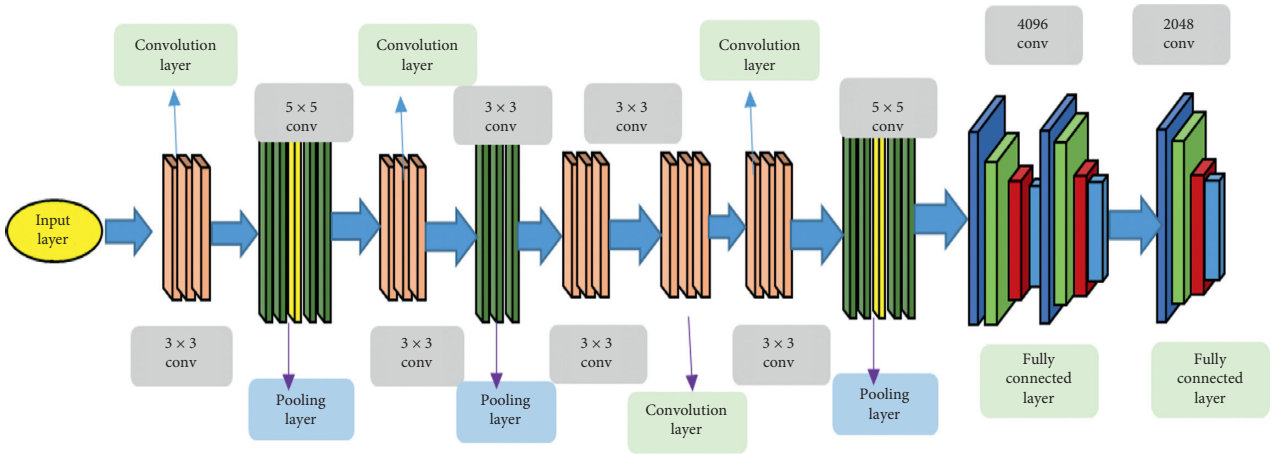

Figure 4: Three-stream CNN framework.

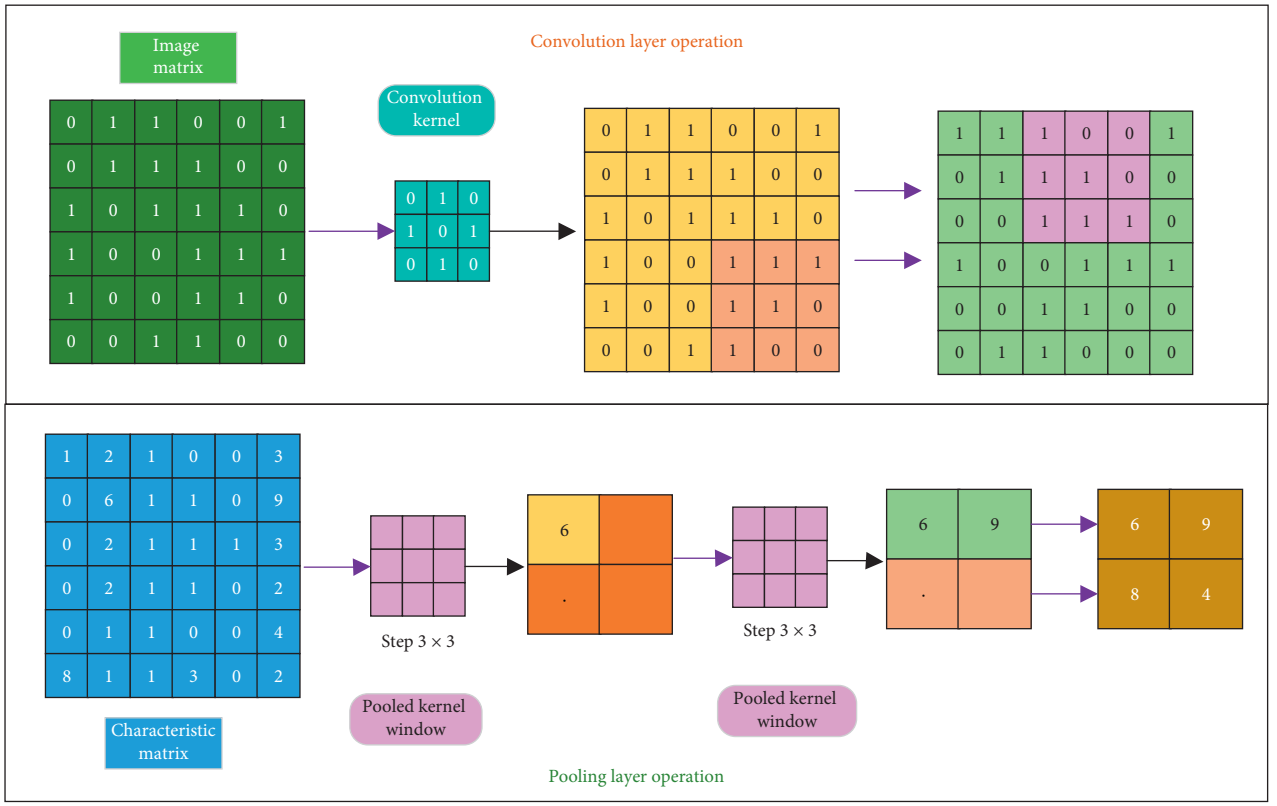

Figure 5: Operation diagram of convolution layer and pooling layer.

$$E(x, y, k) = U\phi(x, y, k - i). \tag{13}$$

After the action depth features are obtained by three-stream CNN framework, the traditional action recognition methods based on CNNs often output them to a classification layer, that is, directly using softmax regression to classify each feature, and finally calculate the mean value to get the action type. The classification layer is ignored, and the action depth feature is regarded as the extracted feature. A soft Vlad algorithm based on Vlad is used to characterize it, and the final action classification is realized by SVM. Based on the three-channel convolutional neural network, the expansibility of its framework is explored. In this framework, the learned action features are represented by a soft Vlad algorithm based on data decoding. Through the data enhancement of the existing action database, the support vector machine is used to achieve high-precision action classification.

Vlad is a hard coding algorithm based on the distance between feature vectors and their nearest cluster centers, while soft coding is more suitable for the representation of action features, because it considers the relationship between

feature vectors and multiple potential decoding points. In addition, the *K*-means clustering algorithm in Vlad is also a hard allocation strategy. It only estimates the nearest center and ignores the relative position of the original data relative to other centers, which is easily affected by random errors. In this paper, the Vlad is improved. The output of full connection layer is regarded as the characterization feature of action depth, and a soft Vlad is proposed to characterize it. Firstly, PCA whiting technique is used to denoise the depth feature operator.

Step 1: discrete Fourier transform (DFT) is used to extract sports action features and normalize them

Step 2: KPCA is used to process the original features of sports actions and select the features that have important contributions to the recognition results

Step 3: simplify the training set and test samples according to the selection features

Step 4: input the simplified training sample set into BP neural network for learning, and determine the threshold and connection weight of BP neural network through particle swarm optimization algorithm

Step 5: according to the optimal threshold and connection weight, the BP neural network sports action recognition classifier is established

Step 6: input the simplified test sample set into the established sports action recognition classifier for testing, and output the recognition results

4. Experiment and Analysis

4.1. Analysis on the Correct Rate of Sports Training Action Recognition. UCF50 data set is selected as the experimental object, which includes 50 sports movements, mainly including playing basketball, diving, weightlifting, horizontal bar, and horse riding. The background is complex, and the visual angle is extremely different. There are 6618 samples in total. 4000 samples are selected as the training set, and other samples as the test set. The average recognition rate is used as the measurement standard of sports action recognition results, and the random projection dimension reduction feature algorithm is used as the comparison experiment.

The results of this method and the comparison method are shown in Table 1. From the experimental results in Table 1 and Figure 6, it can be seen that the method in this paper comprehensively utilizes the advantages of spatial aggregation, and the recognition accuracy of sports action is significantly better than that of the comparison method. Due to the fact that the stochastic projection algorithm reduces the dimension of sports action features according to the maximum contribution, it needs a large number of training samples (see Figure 6). Moreover, it is necessary to reduce the dimension of all the training feature samples of sports movements, which is easy to destroy the internal relationship of important features of sports movements and has high redundancy of feature information. In this paper, the motion features are randomly projected into a low dimensional subspace by using random projection algorithm, which can effectively ensure the reliability of motion recognition.

At the same time, it can be seen from the experimental results that, for all sports movements, the recognition results of the two methods are not ideal; that is, the recognition accuracy of playing basketball is relatively low. The main reason for this problem is that the background of sports action is complex. In the process of moving the target, the camera is disturbed to some extent, which affects the feature extraction of sports action and reduces the recognition accuracy of sports action.

4.2. Analysis on the Efficiency of Sports Training Action Recognition. On the platform of MATLAB r2014b, the recognition efficiency of the two methods is tested. The running time is used to evaluate the recognition efficiency, and the calculation time of dimension reduction of different features (unit: s). The statistics of the experimental results are shown in Table 2 and Figure 7. As can be seen from Table 2, compared with the comparison method, this method significantly improves the efficiency of sports action recognition. This is mainly because the contrast method uses random projection algorithm to reduce dimension, and the matrix feature decomposition is needed, which makes the time complexity high. With the increase of feature dimension, the dimension reduction time increases rapidly, and the random projection algorithm only needs simple matrix operation, which greatly improves the efficiency of feature dimension reduction (see Figure 7).

4.3. Performance Analysis of Sports Training Action Recognition. Select 10 athletes, they demonstrate a variety of simple sports movements, get a total of 600 data, randomly select 400 data to build a training set, the rest of the data is to build a test set, and the basic movements are shown in Tables 3 and 4 and Figure 8. Compared with the sports action recognition model, KPCA select features, BPNN initial threshold and connection value are randomly determined (KPCA**-**BPNN). The initial threshold and connection value (BPNN) of BPNN are optimized by particle swarm optimization algorithm. The recognition rate of sports action and the average recognition time (s) of an action are used as performance evaluation indexes (see Figure 8).

Using training samples to build sports action recognition model, and then using test samples to test, their recognition rate is shown in Figure 9. We can get the following conclusions:

(1) Compared with KPCA-BPNN, PSO-BPNN has higher recognition rate of sports action, which effectively reduces the error recognition rate of sports action. This shows that KPCA-BPNN determines the initial threshold and connection value of BPNN in a random way. The BP neural network with optimal structure cannot be constructed, so the sports action classifier does not reach the optimal, and it is difficult to obtain the ideal sports action recognition results,

Table 1: Correct rate of sports action recognition.

Action type	30 dimensions		60 dimensions	
	Comparison method	Method of this paper	Comparison method	Method of this paper
Play basketball	67.28	71.65	68.21	71.57
Weightlifting	82.91	92.35	84.49	92.86
Golf	87.59	88.93	87.16	89.72
Diving	81.51	88.57	81.56	88.91
Vaulting horse	79.94	83.96	80.68	84.81
Horizontal bar	84.88	90.78	83.59	91.88
Average	80.69	86.04	80.95	86.63

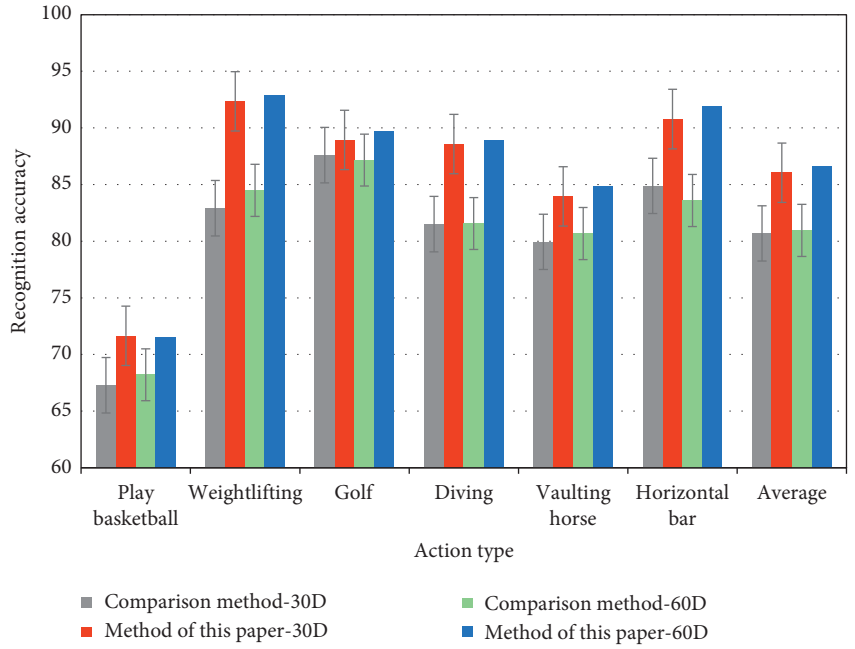

Figure 6: Comparison of correct rate of sports action recognition.

Table 2: Dimension reduction time of sports action method.

Action type	30 dimensions		60 dimensions	
	Comparison method	Method of this paper	Comparison method	Method of this paper
Play basketball	5.22	4.61	22.48	8.63
Weightlifting	5.35	3.58	24.34	8.89
Golf	6.29	4.89	22.17	9.56
Diving	5.92	3.96	20.12	8.84
Vaulting horse	6.74	3.15	23.86	8.93
Horizontal bar	5.14	4.68	23.14	9.92
Average	5.78	4.15	22.69	9.13

which verifies the effectiveness of PSO algorithm to optimize BP neural network.

(2) Compared with BPNN, PSO-BPNN improves the recognition rate of sports actions, which indicates that there are some repetitive features and useless features in the original features of sports actions, which will adversely affect the construction of sports classifier. The PSO-BPNN uses KPCA to select some important features and solves the problem of feature selection and classifier parameter optimization, which makes the recognition result more reliable. In order to speed up the running speed of the algorithm, on the one hand, we should vigorously develop hardware technology, and, on the other hand, we should make in-depth research on acceleration and parallel computing technology.

It is often necessary to carry out online recognition of sports video actions, so test experiments are used to analyze the recognition speed of sports actions. The average recognition time of PSO-BPNN and other models is shown in

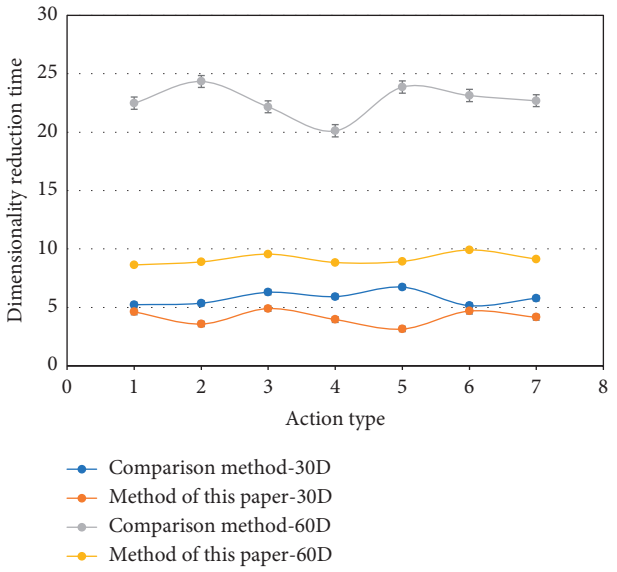

Figure 7: Comparison of dimension reduction time of sports action methods.

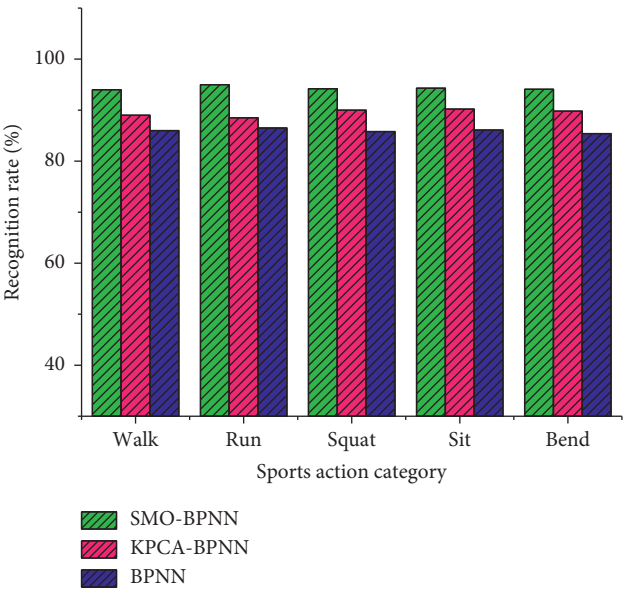

Figure 8: Recognition rate of sports action.

Table 3: Recognition rate of sports action.

Action type	SMO-BPNN	KPCA-BPNN	BPNN
Walk	94	89	86
Run	95	88.5	86.5
Squat	94.2	90	85.8
Sit	94.3	90.2	86.1
Bend	94.1	89.8	85.4

Table 4: Recognition time of sports action.

Action type	SMO-BPNN	KPCA-BPNN	BPNN
Walk	0.3	0.31	0.36
Run	0.28	0.32	0.37
Squat	0.27	0.3	0.39
Sit	0.28	0.29	0.4
Bend	0.28	0.3	0.36

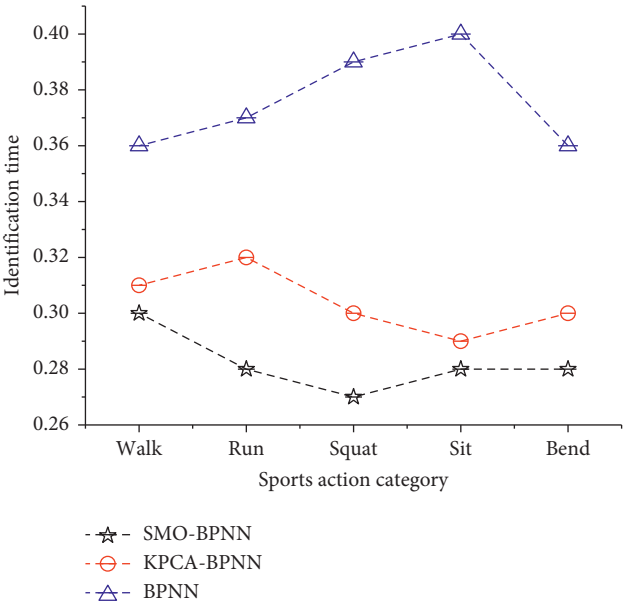

Figure 9: Comparison of recognition time of sports action.

Table 4. According to the average recognition time of sports action in Table 4, the average recognition time of PSO-BPNN is less than KPCA-BPNN and BPNN. This is because PSO-BPNN uses KPCA to select important features, reduces the input dimension of sports action classifier, and speeds up the modeling speed of sports action recognition. At the same time, the PSO algorithm is used to determine the threshold and connection weight of BP neural network, which speeds up the convergence speed of BP neural network, improves the recognition efficiency of sports action, and better meets the requirements of practical application. By using GMM model, the Vlad algorithm, which accords with the principle of hard allocation, is improved to the mechanism of soft allocation, which improves the accuracy of action representation.

5. Conclusion

The realization of action recognition system provides a scientific and intelligent training system for athletes and students, so that users are not limited by time and space control. It is also more strict with the movement standard. Action recognition system is a high-precision training system, which can make training plan according to the user's own training situation and help them better complete the standardization of action. With the rapid development of modern information technology, the emergence of action recognition system updates the traditional concept of sports teaching and training, brings better services for participants and teachers, and also makes contributions to the cultivation and selection

of professional priority talents in the field of national sports. Sports action recognition is a multiclassification pattern recognition problem, including two key issues: sports action characteristics and sports action classification. There are two types of sports action features: silhouette and contour. The high dimension of silhouette makes the number of input vectors of sports action classifier too large, and the computational complexity is too long, which cannot meet the requirements of online sports action recognition. The classifier of sports action recognition mainly adopts neural network design; especially, the artificial intelligence neural network has the best classification performance and is widely used. In the process of sports action classification, the initial threshold and connection weight of artificial intelligence neural network affect the recognition rate. At present, the initial threshold and connection are mainly set according to experience, so it is difficult to obtain the optimal structure of artificial intelligence neural network. In order to obtain more ideal results of sports action recognition, this paper proposes a sports combination training action recognition method based on SMO algorithm optimization model and artificial intelligence and tests the advantages and disadvantages of sports action recognition results through specific experiments. Although this paper has achieved some research results in the above aspects, the research of motion recognition based on vision is an extremely challenging topic, which needs further research and exploration in the future.

Data Availability

All the information is within the paper.

Conflicts of Interest

No competing interests exist concerning this study.

References

[1] B. Ma, S. Nie, M. Ji, J. Song, and W. Wang, "Research and analysis of sports training real-time monitoring system based on mobile artificial intelligence terminal," *Wireless Communications and Mobile Computing*, vol. 2020, no. 6, pp. 1–10, 2020.

[2] J. Wang and H. Qu, "Analysis of regression prediction model of competitive sports based on SVM and artificial intelligence," *Journal of Intelligent & Fuzzy Systems*, vol. 39, no. 4, pp. 5859–5869, 2020.

[3] W. Liu, H. Ma, and A. Walsh, "Advance in photonic crystal solar cells," *Renewable and Sustainable Energy Reviews*, vol. 116, Article ID 109436, 2019.

[4] X. Zhang, C. Zang, H. Ma, and Z. Wang, "Study on removing calcium carbonate plug from near wellbore by high-power ultrasonic treatment," *Ultrasonics Sonochemistry*, vol. 62, p. 104515, 2020.

[5] H. Ma, X. Zhang, F. Ju, and S.-B. Tsai, "A study on curing kinetics of nano-phase modified epoxy resin," *Scientific Reports*, vol. 8, no. 1, p. 3045, 2018.

[6] D. Gao, Y. Liu, Z. Guo et al., "A study on optimization of CBM water drainage by well-test deconvolution in the early development stage," *Water*, vol. 10, no. 7, 2018.

[7] D. Xu and H. Ma, "Degradation of rhodamine B in water by ultrasound-assisted TiO_2 photocatalysis," *Journal of Cleaner Production*, vol. 313, Article ID 127758, 2021.

[8] M. Ling, M. J. Esfahani, H. Akbari, and A. Foroughi, "Effects of residence time and heating rate on gasification of petroleum residue," *Petroleum Science and Technology*, vol. 34, no. 22, pp. 1837–1840, 2016.

[9] H. Ma and S.-B. Tsai, "Design of research on performance of a new iridium coordination compound for the detection of Hg^{2+}," *International Journal of Environmental Research and Public Health*, vol. 14, no. 10, p. 1232, 2017.

[10] L. Mo, W. Sun, S. Jiang et al., "Removal of colloidal precipitation plugging with high-power ultrasound," *Ultrasonics Sonochemistry*, vol. 69, Article ID 105259, 2020.

[11] S. Abe, "Fusing sequential minimal optimization and Newton's method for support vector training," *International Journal of Machine Learning and Cybernetics*, vol. 7, no. 3, pp. 345–364, 2016.

[12] P. R. Singh, M. A. Elaziz, and S. Xiong, "Modified spider monkey optimization based on Nelder-Mead method for global optimization," *Expert Systems with Applications*, vol. 110, pp. 264–289, 2018.

[13] N. A. Omid and R. Modjtaba, "A new fuzzy membership assignment and model selection approach based on dynamic class centers for fuzzy SVM family using the firefly algorithm," *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 24, pp. 1797–1814, 2016.

[14] W. C.-C. Chu, C. Shih, W.-Y. Chou, S. I. Ahamed, and P.-A. Hsiung, "Artificial intelligence of things in sports science: weight training as an example," *Computer*, vol. 52, no. 11, pp. 52–61, 2019.

[15] S.-B. Tsai and H. Ma, "A research on preparation and application of the monolithic catalyst with interconnecting pore structure," *Scientific Reports*, vol. 8, no. 1, 2018.

[16] X. Leng, H. Jiang, X. Zou et al., "Motion feature quantization of athletic sports training based on fuzzy neural network theory," *Cluster Computing*, vol. 22, no. 2, pp. 4631–4638, 2019.

[17] B. Li, C. Jiang, G. Zhang, and Y. He, "Management and performance evaluation of DSR aggregator based on a bi-level optimization model," *IEEJ Transactions on Electrical and Electronic Engineering*, vol. 13, no. 3, pp. 432–441, 2018.

[18] H. Xu and R. Yan, "Research on sports action recognition system based on cluster regression and improved ISA deep network," *Journal of Intelligent & Fuzzy Systems*, vol. 39, no. 4, pp. 5871–5881, 2020.

[19] C. Sun and D. Ma, "SVM-based global vision system of sports competition and action recognition," *Journal of Intelligent & Fuzzy Systems*, vol. 40, no. 2, pp. 2265–2276, 2021.

[20] F. Ai, X. Yin, R. Hu, H. Ma, and W. Liu, "Research into the super-absorbent polymers on agricultural water," *Agricultural Water Management*, vol. 245, Article ID 106513, 2021.

[21] S. Nazir, M. H. Yousaf, J.-C. Nebel, and S. A. Velastin, "A bag of expression framework for improved human action recognition," *Pattern Recognition Letters*, vol. 103, pp. 39–45, 2018.

[22] J. F. Yang, M. Zheng, and X. Mei, "Multiscale spatial position coding under locality constraint for action recognition," *Journal of Electrical Engineering & Technology*, vol. 10, no. 4, pp. 1852–1864, 2015.

[23] X. Li and S. Geng, "Research on sports retrieval recognition of action based on feature extraction and SVM classification algorithm," *Journal of Intelligent & Fuzzy Systems*, vol. 39, no. 4, pp. 5797–5808, 2020.