# Research on Action Recognition Method of Traditional National Physical Education Based on Deep Convolution Neural Network

Liuyu Bai[1], Wenbao Xu[1](✉), Zhi Xie[1], and Yanuo Hu[2]

[1] Air Force Engineering University, Xi'an 710051, China
rendanjie@126.com
[2] Xianyang City Qindu District Teachers Training School, Xianyang 712046, China

**Abstract.** With the continuous development of machine vision and image processing technology, more and more attention has been paid to human action recognition in physical education teaching. In order to improve the performance of action recognition in traditional P. E. teaching, the method of action recognition based on deep convolution neural network is proposed. The length of elbow joint and shoulder joint was calculated by using the distance between the camera and the action image. According to the range characteristics of sports teaching action, monitoring sports teaching action. Deep convolution neural network was introduced to predict the state variables of PE teaching action, and the coordinate data information of all related nodes was obtained. Based on the theory of Deep Convolution Neural Network, this paper transforms and deals with the action posture of traditional national sports teaching in colleges and universities. Through the probabilistic value of the motion image pixel of the traditional PE teaching in colleges and universities, the motion characteristics of PE teaching are extracted. Through detecting the extreme point of PE teaching action in the scale space, locate the extreme point of PE teaching action range. Using the probability of the range of sports teaching action outside the exercise area, we can identify the traditional sports teaching action in colleges and universities. The experimental results show that the method can successfully identify the traditional national sports teaching behavior. This method has good performance in the precision of motion feature extraction, recognition rate and recognition speed. It perfects the problem of low precision in sports action recognition.

**Keywords:** Deep Convolutional Neural Network · Physical Education · Action Recognition · Feature Extraction · Action Monitoring

## 1 Introduction

With the rapid development of artificial intelligence technology in computer vision, mobile Internet, big data analysis and other fields. Combined with the new characteristics of deep learning and cross-border integration, it has gradually become a new focus of


B. Wang et al. (Eds.): ICMTEL 2023, LNICST 533, pp. 239–253, 2024.
https://doi.org/10.1007/978-3-031-50574-4_17

international competition. At the same time, artificial intelligence technology is used in computer vision to segment, classify and recognize the objects. And its application in virtual reality and human-computer interaction, especially in traditional national physical education in colleges and universities, has become a hot topic in industry and academia [1]. Through the traditional sports teaching in colleges and universities, the students' sports training and competition video can be analyzed and evaluated. Through the analysis of the skills of students of sports normative movement, physical training and other aspects of targeted. At the same time, in the sports meeting, the students' movements and positions are detected, tracked and analyzed, which promotes the improvement of technical level.

Basketball is the most basic sports item in the traditional sports teaching in colleges and universities. In basketball movements, the basic movements include dribbling, shooting and lay-up. Among them dribble is the most basic action in basketball, shooting is the key to score in the whole game. The accuracy of the basic movements has a great influence on the score of the match [2]. With the development of basketball sports, the algorithm of human posture estimation and the algorithm of action recognition are combined. It plays a vital role in helping to improve the scoring rate. Human posture estimation is to detect and estimate the position, direction and scale information of the target human body from the image, which needs to be translated into digital form. And output the current human posture movement. However, motion recognition is based on the results of attitude estimation as input object to judge whether a person's motion is standardized and how to improve the standardization.

In the domestic research, He Bingqian et al. [3] proposed the network structure based on batch normalization transform and Goog Le Net network model to solve the problems of complex feature extraction and low recognition rate. The normalization of image classification is applied to the field of motion recognition to improve the training algorithm and realize the network input of video action training samples. The process of normalization was carried out. The method takes RGB image as input of space network and optical flow field as input of time network. Then the final recognition result is obtained by fusing the spatiotemporal network. Experiments on UCF101 and HMDB51 datasets show 93.50% and 68.32% accuracy, respectively. Experimental results show that the improved network architecture has a high recognition accuracy in video human motion recognition. Liu Guoping et al. [4] In order to realize the target of dumbbell action classification recognition, an inertial sensor module is added to dumbbell to collect the motion signals during dumbbell training. After signal normalization, filtering and segmentation based on initial static vector period, 5 kinds of dumbbell action characteristic vectors are extracted. An improved Relief F feature selection algorithm is used to select the optimal feature vector. Support vector machine based on balanced decision tree is used to recognize different dumbbells. Through the laboratory independently developed dumbbell action recognition system for testing. The results show that the system can recognize dumbbells in a single dumbbell action cycle, and the recognition rate can reach 90%. Provides the more individualized dumbbell movement instruction to lay the foundation.

In foreign research, Ding W et al. [5] proposed a square grid based on skeleton. Data used to transform dynamic skeletons into three-dimensional mesh structures so that CNN can be applied to these data. In order to enhance the ability of depth feature to capture the correlation between 3D mesh data, a joint based square mesh and a rigid body based square mesh sequence are used to construct a dual flow 3D CNN. Three data sets, NTU RGB D, Kinetics Motion and SBU Kinect interactive dataset, are used to verify the effectiveness of the model in motion recognition. Experimental results show the effectiveness of the proposed method and its superiority over the existing methods. Gao P et al. [6] In order to improve the accuracy of small-scale human motion recognition and the computational efficiency of large-scale data sets in video, a multi-dimensional data model based on deep learning framework for video image motion recognition and motion capture is proposed. Firstly, the moving foreground of the target is extracted by Gaussian mixture model and the human body is recognized by gradient histogram. In the second layer, according to the integration of global coding algorithm and convolution neural network, dense trajectory feature and deep learning feature are fused. In the deep learning feature, the fusion of the video feature and the video RGB feature is the feature of the deep learning. Simulation results based on large scale real dataset and small scale gesture dataset show that the algorithm has high recognition accuracy for large scale dataset and small scale gesture. In addition, the average classification accuracy is 85.79% when the human behavior dataset is used in computer vision and learning laboratory. The algorithm can run at about 20 frames per second.

At home, with the rapid improvement of computing technology, action recognition in traditional sports teaching has gradually become a new subject in the field of computer simulation. It has the extremely vital research value in the university nationality traditional sports teaching domain. At present, many research institutions in developed countries are carrying out relevant research and development. It is mainly committed to the development of boxing, table tennis, skiing and other sports movement amplitude larger. Moreover the university national tradition sports teaching movement recognition, may bring the enormous convenience to student's sports training. For example, students can observe the execution details of standard movements from different angles in the process of sports training. At the same time, the differences between actual athletes and standard movements were compared to better assess the effect of student completion. However, in the actual teaching of traditional sports in colleges and universities in the process of movement recognition, one of the key issues is how to effectively identify the key points of sports movement. The optimal tracking method of 3D visual motion amplitude in sports is used to obtain the time series model of motion amplitude, and the probability of 3D visual motion amplitude is given. Based on this, the action recognition in the teaching of traditional national sports in colleges and universities is the fundamental way to solve the above problems, which arouses the attention of many experts and scholars.

Based on the above research background, in order to improve the accuracy of sports action feature recognition, this paper designs an action recognition method for traditional physical education teaching in traditional colleges and universities by applying deep convolutional neural network. According to the length and distance of the elbow joint and the shoulder joint, combined with the characteristics of physical education teaching actions, the teaching action monitoring is carried out. Predicting Action State Variables Using Deep Convolutional Neural Networks. Actions in the physical education teaching process are identified based on the acquired joint coordinate data. Thereby promoting the development of computer vision recognition sports.

# 2 Design of Action Recognition Method in National Traditional Physical Education Teaching in Colleges and Universities

## 2.1 Monitoring Physical Education Actions

Before recognizing the teaching action of college traditional national sports, we should monitor the teaching action of college traditional national sports to obtain the process of the teaching action. According to the two-dimensional data of students' motion joint characteristic points in traditional P. E teaching, the plane distance between camera and P. E teaching action image is obtained. The calculation formula is:

$$u_{ij} = \frac{\psi \times f^*}{dis \cdot X^*}\sqrt{\left(X_i - X_j\right)^2 + \left(Y_i - Y_j\right)^2 + \left(Z_i - Z_j\right)^2} \tag{1}$$

$\psi$ represents the scale factor. $dis$ indicates the distance between the center of the motion image and the center of the camera. $X^*$ stands for camera coordinates. $(X_i, Y_i, Z_i)$ represents the coordinates of the action image for physical education. $\left(X_j, Y_j, Z_j\right)$ represents the coordinates of the action image after the transformation. $f^*$ represents the focal length of the camera. The formula is:

$$f^* = \frac{d_0}{S^* \times d_{ij}}\sqrt{\left(X_i - X_j\right) + \left(Y_i - Y_j\right)} \tag{2}$$

Here, $d_0$ represents the camera lens diameter. $d_{ij}$ indicates the spacing between monitoring points. $\left(X_i - X_j\right)$ and $\left(Y_i - Y_j\right)$ represent the image coordinates of the monitoring point, and $S^*$ represents the area of the camera lens section.

Suppose $\lambda_{\max}$ represents the maximum threshold of the monitoring point of physical education. $\lambda_{\min}$ represents the minimum threshold of movement monitoring points in physical education. The following formula can be used for binary processing of physical education teaching action, expressed as:

$$\lambda_0 = \frac{(\lambda_{\max} - \lambda_{\min}) * N}{N^*} + \Lambda(y) \tag{3}$$

Among them, $N$ represents the limit threshold of monitoring points in physical education. $N^*$ represents the number of sports teaching actions. $\Lambda(y)$ stands for Monitoring Point of Physical Education.

When monitoring the movement of PE teaching, the result of binarization should be considered. Use the following formula to calculate the length between the elbow and shoulder. The formula is:

$$L^* = \sqrt{f(X_1 - X_2) + (Y_1 - Y_2) + (Z_1 - Z_2)} \tag{4}$$

$(X_1, Y_1, Z_1)$ represents the elbow coordinates of the student's motion. $(X_2, Y_2, Z_2)$ indicates the coordinates of the shoulder joint when the student is exercising.

Suppose $\Omega_k$ represents the range of action in physical education. If the movement range $J_n$ of PE teaching is detected in the range of display, the predictive error $\left(\Delta_p, \Delta_v, \Delta_a\right)$ of the state variable of movement range can be calculated. Predicts the range of physical education teaching action, namely:

$$\Omega_k\left(\Delta_p, \Delta_v, \Delta_a\right) = \frac{\partial(p) \cdot J_n}{\{R_k\}\left(\Delta_p, \Delta_v, \Delta_a\right)} \tag{5}$$

Among them, $\{R_k\}$ represents the three-dimensional movement data of the $k$-movement. $\partial(p)$ stands for Candidate Monitoring Points for Physical Education Teaching Movements.

After obtaining the display scope of the physical education teaching movement, the range characteristics of the physical education teaching movement are extracted in the area, which are expressed as follows:

$$F(i,j) = \sum_{i=1}^{n} \Phi_i \times \Omega_k \tag{6}$$

Among them, $\Omega_k$ represents the range of movement of physical education production area.

If the student sports discrete space is expressed as $RGB$. Then we can use the following formula to classify the color in the sports teaching action image, namely:

$$P(y/x) = \exp\left[\frac{e_i(x, y)}{RGB \cdot h(e)}\right] \tag{7}$$

Among them, $e_i(x, y)$ represents the color distribution of the students' skin at the current time. $h(e)$ represents the color distribution function of action images in physical education.

According to the result of color classification of sports teaching action image, monitoring sports teaching action, expressed as:

$$\sigma(x, y) = \frac{G^* \cdot P(\frac{y}{x})}{G_0(x, y)} \tag{8}$$

Among them, $G^*$ means the observation value of the movement characteristics of physical education. $G_0(x, y)$ indicates the normalized observation error between movement features when the movement parameter of PE teaching is $x$.

The length of elbow joint and shoulder joint was calculated by using the distance between the camera and the action image. Through the binarization of PE teaching action, the range of PE teaching action is obtained. According to the range characteristics of sports teaching action, monitoring sports teaching action.

## 2.2 Constructing the Action Model of Physical Education Teaching

According to the monitored physical education actions, the data of physical education actions are collected, which are expressed as:

$$\hat{\Omega}_{k+1} = \frac{1}{\phi}\Omega_k \tag{9}$$

In the above formula, $\phi$ is the structural parameter of physical education teaching action. $\Omega_k$ is the covariance of physical education action in frame $k$.

Following the acquisition of movement data for physical education, a deep convolution neural network [7] was introduced and expressed as follows:

$$B_j = k_y\left(\sum_{i=1}^{j=1} x_i \tau_{hij} + \zeta_j\right) \tag{10}$$

In the equation, $B_j$ represents the output state vector of the hidden layer of the deep convolution neural network. $x_i$ represents the input state vector of the deep convolution neural network. $\tau_{hij}$ represents the connection value from the input layer of the deep convolution neural network to the hidden layer. $\zeta_j$ represents the threshold of hidden layer neurons in deep convolution neural networks. $k_y$ stands for the excitation function. Output state vector $Z_k$ from $B_j$:

$$Z_k = k_y\left(\sum_{j=1}^{j=J} B_j \tau_{ojk} + \phi_k\right) \tag{11}$$

In the equation, $\tau_{ojk}$ represents the connection weights from the input layer to the output layer of the deep convolution neural network. $\phi_k$ represents the threshold value of the output layer of the deep convolution neural network.

Using the structure of deep convolution neural network, the state variables of physical education action are predicted and expressed as follows:

$$Z_k = \frac{V_k}{J} g_{k-1} + \varsigma^* M_k \tag{12}$$

In formula, $Z_k$ is the dynamic information of PE teaching action. $g_{k-1}$ is the dynamic vector of physical education. $J$ is the transformation matrix. $\varsigma^*$ stands for transfer matrix. $M_k$ represents the identification factor. $V_k$ The error value predicted for the state variable.

It is assumed that all the gray values of the action image satisfy the expectation value of $\mu$ and the variance of the gray distribution is $\varepsilon^*$ normal distribution. The Gaussian distribution of motion image pixels is a single variable. Then in the action image frame of PE teaching, the target and background probability models composed of all pixels can be expressed as follows:

$$p(X|s_i) = \frac{1}{\sqrt{2\pi\varepsilon^*}} \exp\left(\frac{X-\mu}{\varepsilon^*}\right) \tag{13}$$

$p(X|s_i)$ is a univariate normal distribution. $s_i$ is the eigenvalue of the gray information of the action image in physical education.

In the action image frame of PE teaching, all pixels are used to form the target and background probability models with normal distribution. Updating the mean and variance of the target and background of the action image in physical education [8], the formula can be expressed as follows:

$$\mu_{k+1}^1 = \frac{k-1}{k}\mu_k^1 + \frac{1}{k}\mu_k^1 \tag{14}$$

$$\varepsilon_{k+1}^2 = \frac{k-1}{k}\varepsilon_k^2 + \frac{1}{k}\varepsilon_k^2 \tag{15}$$

According to the actions in the traditional P. E teaching in colleges and universities, the action characteristics of P. E teaching are calculated. The formula is as follows:

$$\mu_{k+1} = (1-\vartheta)\mu_k^* + \vartheta\mu_k \tag{16}$$

$$\varepsilon_{k+1} = (1-\vartheta)\varepsilon_k^* + \vartheta\varepsilon_k \tag{17}$$

Among them, $\vartheta$ is the covariance of movement characteristics in physical education. $\mu_k$ is the $k$ frame PE teaching action image. $\mu_k^*$ represents the movement characteristics of physical education. $\varepsilon_k$ represents the characteristic vector of physical education. $\varepsilon_k^*$ indicates the range characteristics of PE teaching movements.

According to the calculation results of the movement characteristics of physical education, the coordinate data information of all related nodes of physical education is obtained. Build a physical education action model, expressed as:

$$Z_i(k+1) = \frac{1}{\vartheta}\Theta(k+1) \tag{18}$$

Among them, $\Theta(k+1)$ represents the covariance matrix of movement recognition in physical education.

The state variables of physical education action are predicted by collecting physical education action and introducing deep convolution neural network. Using the target and background probability models composed of all pixels, the coordinate data information of all related nodes in physical education action is obtained. Constructs the sports teaching movement model.

### 2.3 Extracting Movement Characteristics of Physical Education Based on Deep Convolution Neural Network

This paper makes use of the theory of deep convolution neural network to extract the action features in the traditional national physical education teaching. The extracted action feature points are classified and processed, combined with the convolution layer of deep convolution neural network [9]. This paper transforms the posture of traditional sports teaching in colleges and universities as follows:

$$U_{(i,j)}(g,h) = |V(i,j) \times W(g,h)| \tag{19}$$

In the above formula, $U_{(i,j)}(g,h)$ represents the filtered result after processing the action image of traditional national physical education teaching in colleges and universities. $V(i,j)$ represents the action image of traditional national physical education in colleges and universities. $W(g,h)$ stands for filter combination result.

In the process of extracting the movement features of physical education, the depth convolution neural network is used to calculate the characteristic parameters of each movement posture of students in physical education. The formula is as follows:

$$Z_{ij} = \frac{\left|Z_{ij} - \min\{Z_{ij}\}\right|}{\Phi^*} \tag{20}$$

Among them, $\Phi^*$ represents the movement image of physical education. $\min\{Z_{ij}\}$ represents the characteristic parameters of each student's movements in the traditional sports teaching in colleges and universities.

Using the characteristic parameters of each action posture of students in physical education. This paper defines the $m+n$-rank matrix of the teaching action image $I(x, y)$ in the traditional national sports teaching in colleges and universities, which is expressed as:

$$\chi_{mn} = \sum_x I(x, y) \sum_y (x - \overline{x})^m (y - \overline{y})^n \tag{21}$$

Among them, $(\overline{x}, \overline{y})$ represents the center coordinate of the action image of college traditional national physical education.

According to the order matrix of the teaching action image, $m+n$ acquires the pixel information of the teaching action image of the traditional national physical education. The pixels of the resulting image are:

$$f(d_B) = \frac{\gamma_a}{8} \tag{22}$$

$$f(d_F) = 1 - f(d_B) \tag{23}$$

Among them, $f(d_B)$ represents the characteristic distribution function of the pixel value HB of the action image of the traditional national physical education. $\gamma_a$ represents the pixel information of the action image in physical education. $f(d_F)$ represents the characteristic distribution function of the pixel value HF in the action image of the traditional P. E. When $f(d_B) \geq f(d_F)$, the motion image features of PE teaching are regarded as foreground pixels. When $f(d_B) < f(d_F)$, the action image of PE teaching is used as background pixel.

The characteristic point $(i, j)$ in the coordinate system of the action image of the traditional P. E teaching is obtained by iterative processing [10]. The pixel probability of the feature point in time $t$ is:

$$f\left(a_{ij}^t\right) = \sum_{s=g,h} f_s\left(a_{ij}^t | \alpha_{ij,s}^t\right) \tag{24}$$

In the above formula, $a_{ij}^t$ represents the pixel value of the action image of traditional national physical education teaching in colleges and universities at time $t$. $f_s\left(a_{ij}^t | \alpha_{ij,s}^t\right)$ represents the actual pixel probability of the action image of national traditional physical education teaching in colleges and universities at any time $t$.

At time $t$, the mixed model of the action feature pixel $(i, j)$ in the national traditional physical education teaching in colleges and universities is expressed as:

$$f_b\left(a_{ij}^t\right) = \frac{1}{\varpi_{ij,k}^t} \sum_{k \in M} \eta\left(a_{ij}^t, \delta_{ij,k}^t, R_{ij,k}^t\right) \tag{25}$$

In the above formula, $k$ represents the number of recognition models of traditional national physical education teaching images. $\varpi_{ij,k}^t$ represents the weight of characteristic vectors of traditional sports teaching in universities. $\delta_{ij,k}^{t}$ represents the characteristic vector value of the traditional sports teaching action. $R_{ij,k}^{t}$ represents the covariance matrix of the movement characteristics of traditional national physical education in colleges and universities.

Through the weight analysis of the mixed model of action feature pixel $(i, j)$, the fitness value of the action image of traditional P. E. Looking for the suitable movement characteristic distribution model in the university national traditional sports teaching. Then in any time $t$, the probability value of action pixel $(i, j)$ of traditional sports teaching in colleges and universities is expressed as follows:

$$Q(y, e, \kappa) = \frac{1}{\sqrt{|\kappa|}} \exp\left(-\frac{1}{2}(y - e)_{T}\right) \tag{26}$$

Among them, $y$ represents the characteristic vector of the coordinates $(i, j)$ of the national traditional physical education teaching. $e$ represents the probability value of the coordinates. $\kappa$ represents the vector matrix of teaching actions.

After obtaining the probability value of pixel $(i, j)$, the motion image of traditional P. E teaching is compared with the standard motion image. The characteristics of traditional sports teaching in colleges and universities can be expressed as follows:

$$f_u(y) = \ell_h \sum_{i=1}^{n} Q(y, e, \kappa)\delta[\varepsilon(x_i) - \xi] \tag{27}$$

Among them, $\ell_h$ is the normalized processing result of action image of traditional national physical education in colleges and universities. $\xi$ is the characteristic value of the teaching action image. $\varepsilon(x_i)$ represents the central pixel of the action image of the traditional national physical education.

Based on the theory of Deep Convolution Neural Network, this paper transforms and deals with the action posture of traditional national sports teaching in colleges and universities. According to the characteristic parameters of each action posture in physical education, the pixel information of action image in traditional physical education is obtained. Through the probabilistic value of the motion image pixel of the traditional PE teaching in colleges and universities, the motion characteristics of PE teaching are extracted.

### 2.4 Designing Recognition Algorithm of Physical Education Teaching

According to the movement characteristics of PE teaching, the extremum of the movement in the scale space is detected, that is:

$$W(x, y, \varphi) = I(x, y)\frac{\sqrt{G(x, y, \varphi) - (x, y)}}{L(x, y, \varphi)} \tag{28}$$

Among them, $I(x, y)$ represents the initial coordinates of the action image in physical education. $L(x, y, \varphi)$ represents the convolution of sports teaching action. $G(x, y, \varphi)$ represents the variable Gaussian function of physical education movement in the scale space. $(x, y)$ represents the spatial coordinates of physical education. $\varphi$ represents the scale coordinates of physical education.

In the scale space, according to the extreme point of the physical education teaching action, the extreme point of the range of the physical education teaching action can be located and expressed as:

$$\mu \leq (\kappa^*) = \frac{d_i \cdot C(w)}{D_{a,b} \times \kappa^*} \times d_{\min} \tag{29}$$

$D_{a,b}$ represents the Euclidean distance between $a$ and $b$, the characteristic point of movement amplitude. $\kappa^*$ represents the distance threshold between the range features of physical education movement. $d_i$ represents the subproximity of the range of motion in physical education. $d_{\min}$ represents the closest distance to the range of movement in physical education. $C(w)$ represents the scale of movement in physical education.

After locating the extreme point of the range of movement of PE teaching, the probability of the range of movement of PE teaching outside the display area is calculated. The formula is as follows:

$$E\left(\Omega_x, \Omega_y\right) = \frac{\left(v_x, y_y\right)}{\left(\Omega_x, \Omega_y\right)} \cdot \chi(P) \times \Gamma_{\left(v_x, y_y\right)} \tag{30}$$

Among them, $\Omega_x$ and $\Omega_y$ represent the registration area of physical education. $\left(v_x, y_y\right)$ represents the speed of movement in physical education. $\chi(P)$ represents the number of iterations of the PE teaching action. $\Gamma_{\left(v_x, y_y\right)}$ represents the exact position of the physical education movement.

According to the above process, identify the traditional national sports teaching actions in colleges and universities, namely:

$$\sigma\left(\Omega_x, \Omega_y\right) = \frac{\Sigma_\alpha}{g(t)} \times \Gamma_{\left(v_x, y_y\right)} \tag{31}$$

Among them, $g(t)$ represents the time series of the action state of physical education. $\Sigma_\alpha$ represents the weighted average processing weight of movement recognition in physical education.

To sum up, through the detection of PE teaching action in the scale space of the extreme point, positioning to the extreme point of PE teaching action range. Using the probability of the range of sports teaching action outside the exercise area, we can identify the traditional sports teaching action in colleges and universities.

## 3 Experimental Analysis

### 3.1 Experimental Data Set

This article selects the UTD-MHAD data set and the MSR Daily Activity 3D data set as the experimental data set. The UTD-MHAD data set contains depth information, skeletal joint position data, RGB video sequence and inertia data. The data set includes 27 different actions, such as right-arm sliding to the left, right-arm sliding to the right,

right-hand waving, two-handed forehand racket, right-arm throwing, cross-arm in chest, basketball shooting, right-hand drawing x, right-hand drawing circle (clockwise), right-hand drawing circle (counterclockwise), drawing triangle, bowling (right hand), fore-boxing, right-hand baseball swing, tennis right-hand forehand swing, arm curling (both arms), tennis service, two-handed push, right-hand knocking door, right-hand grasping an object, right-hand picking up and throwing, jogging in place, walking in place, sitting, standing, forward sprint (left foot forward), squatting (two arms extended). Each action was performed four times by eight people, of which three corrupted sequences were removed, and a total of 861 action sequences were performed, making the dataset challenging because 27 actions were performed along the line of sight (perpendicular to the video plane).

The MSR Daily Activity 3D dataset contains RGB video, depth information, and skeletal joint position data. The dataset contains 16 types of actions: drinking, eating, reading, talking on a cell phone, writing, using a laptop, using a vacuum cleaner, cheering, sitting still, throwing paper, playing games, lying on the sofa, walking, playing the guitar, sitting up, sitting down. Each type of action was performed by 10 people, each person carried out two activities, one standing mode, one sitting mode, a total of 960 documents. So strictly speaking, there are 17 action categories for this dataset. Because meditation in the execution of the two types of action, meditation and standing. The dataset was photographed in a real environment with background objects, and the subject's distance from the camera was not fixed.

### 3.2 Position Motion Recognition Sensor

Selects 8 sensors to monitor the university nationality traditional sports teaching movement. Eight nano CMOS image sensors are attached to eight parts of the body, as shown in Fig. 1.


**Fig. 1.** Sensor placement

### 3.3 Result Analysis

In order to highlight the performance of this method, the action recognition method based on neural network and the action recognition method based on improved ReliefF algorithm are compared. Get the following results.

**Recognition of Physical Education**

Get the standing long jump in the UTD-MHAD dataset and the running in the MSR Daily Activity 3D dataset. According to the sensor arrangement of 2.2, the sensor is tied to the human body to recognize the action of different positions. The results are shown in Figs. 2 and 3.

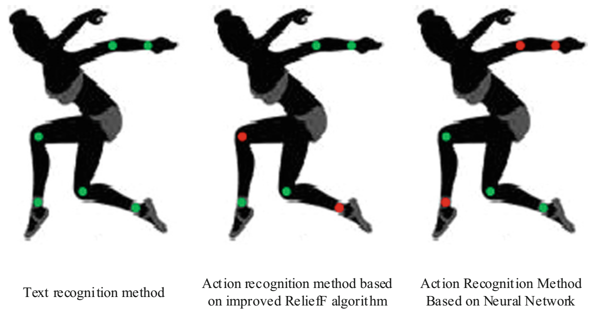


**Fig. 2.** Action recognition results on the UTD-MHAD dataset

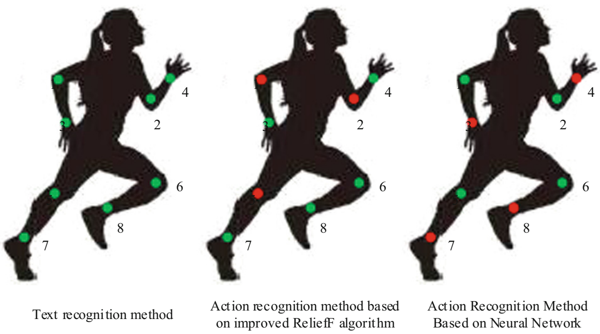


**Fig. 3.** Action recognition results on the MSR Daily Activity 3D dataset

In these results, the green circle represents success and the red circle represents failure. The results in Figs. 2 and 3 show that only the method in this paper can recognize the physical education actions of all monitoring points in two data sets. Action recognition methods based on neural network and improved ReliefF algorithm have two or three cases of failure. Because the method in this paper can predict the state variables of sports teaching actions through deep neural networks. Thus, the coordinate data information of all relevant nodes is obtained. Based on these coordinate data, physical education teaching actions can be accurately identified.

**Performance Contrast**

In the MSR Daily Activity 3D dataset, a group of running movements were selected as experimental subjects. According to the sensor arrangement in Fig. 1, the performance of 8 positions is monitored. The method is compared with other two methods. The accuracy, recognition rate and recognition speed of movement feature extraction in PE teaching were tested, and the results were as follows.

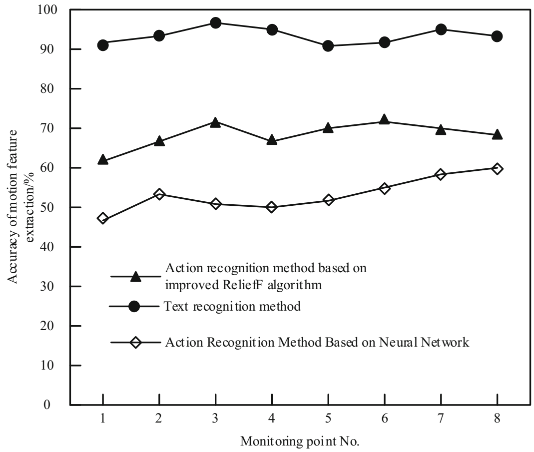


**Fig. 4.** The accuracy of action feature extraction in physical education

From the results in Fig. 4, it can be seen that the accuracy of motion feature extraction can be controlled above 90% before the method in this paper recognizes sports. The feature extraction accuracy of the action recognition method based on neural network and improved ReliefF algorithm is lower than 80%.

According to the results in Fig. 5, it is compared with the neural network-based action recognition method and the improved ReliefF algorithm. The action recognition rate of the two methods compared is lower than 71%, while the recognition rate of the method in this paper is higher than 90%, which can provide higher technical support for the analysis of physical education teaching.

The result of Fig. 6 shows that the recognition speed can be raised to more than 180 frames per second. The higher recognition speed ensures the quality of action images in physical education.

In summary, the method in this paper has certain advantages in the accuracy of motion feature extraction, motion recognition rate and motion speed recognition. Because the method in this paper monitors the physical education teaching action according to the joint distance and the range characteristics of the physical education teaching action. The motion characteristics of physical education teaching are extracted by taking the probability value of the moving image pixels in traditional physical education teaching in colleges and universities. The proposed method has better performance in recognizing actions.

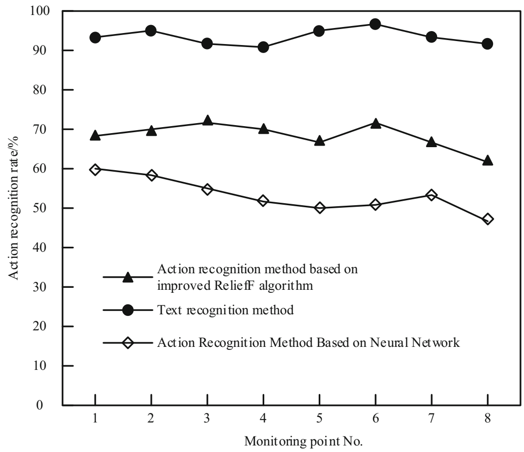


**Fig. 5.** Physical education action recognition rate

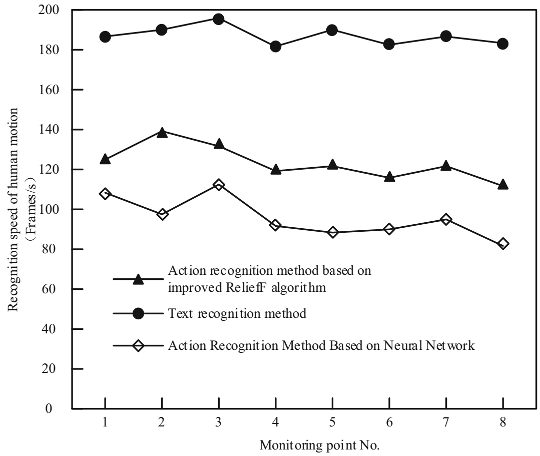


**Fig. 6.** Speed of movement recognition in physical education

## 4 Conclusion

In this paper, the deep convolution neural network is applied to the recognition of the traditional national physical education in colleges and universities. The performance of this method is better in recognizing the teaching action of traditional national physical education. But there are still a lot of deficiencies in this study, in the future research, we hope to be able to analyze the joints in advance. According to the motion direction of the joint, identify the action.

## References

1. Basak, H., Kundu, R., Singh, P.K., et al.: A union of deep learning and swarm-based optimization for 3D human action recognition. Sci. Rep. **12**(1), 5494 (2022)
2. Ma, C., Fan, J., Yao, J., et al.: NPU RGB+ D dataset and a feature-enhanced LSTM-DGCN method for action recognition of basketball players. Appl. Sci. **11**(10), 4426 (2021)
3. He, B., Wei, W., Zhang, B., et al.: Improved deep convolutional neural network for human action recognition. Appl. Res. Comput. **36**(10), 3107–3111 (2019)
4. Liu, G., Wang, N., Zhou, Y., et al.: Dumbbell motion recognition based on improved ReliefF algorithm. Sci. Technol. Eng. **19**(32), 219–224 (2019)
5. Ding, W., Ding, C., Li, G., Liu, K.: Skeleton-based square grid for human action recognition with 3D convolutional neural network. IEEE Access **9**, 54078–54089 (2021)
6. Liu, W.: Simulation of human body local feature points recognition based on machine learning. Comput. Simul. **38**(6), 387–390+395 (2021)
7. Tsai, M.F., Chen, C.H.: Spatial temporal variation graph convolutional networks (STV-GCN) for skeleton-based emotional action recognition. IEEE Access **9**, 13870–13877 (2021)
8. Yuan, Y., Yu, B., Wang, W., et al.: Multi-filter dynamic graph convolutional networks for skeleton-based action recognition. Procedia Comput. Sci. **183**, 572–578 (2021)
9. Xie, J., Xin, W., Liu, R., Sheng, L., Liu, X., Gao, X., Miao, Q.: Cross-channel graph convolutional networks for skeleton-based action recognition. IEEE Access **9**, 9055–9065 (2021)
10. Liu, S., Liu, D., Muhammad, K., Ding, W.: Effective template update mechanism in visual tracking with background clutter. Neurocomputing **458**, 615–625 (2021)