Chen,w.; Hou,Y; (2022) Accurate calibration of physical education actions based on artificial intelligence deep learning technology. Revista Internacional de Medicina y Ciencias de la Actividad Física y el Deporte vol. 22 (85) pp. 267-282
DOI: https://doi.org/10.15366/rimcafd2022.85.017

ORIGINAL

Accurate calibration of physical education actions based on artificial intelligence deep learning technology

Wenta Chen[1] , Yanwu Hou[2] *
[1] Department of Physical Education, Nanchang Institute of Technology, Nanchang, Jiangxi, 330099, China
[2] Sports Department, Zhongnan University of Economics and Law, Wuhan, 430073, China
Email: 18971639876@163.com

UNESCO Code / UNESCO Code:
Council of Europe classification / Council of Europe classification:

Recibido 09 de enero de 2021 **Received** January 09, 2021
Aceptado 09 de enero de 2022 **Accepted** January 09, 2022

Abstract

Allowing computers to recognize human actions and behaviors based on video sequences captured by cameras is an important branch of computer vision and AI. At present, human action recognition has been widely used in security monitoring, interactive entertainment, smart home and other aspects. It "aims at cultivating students' ability", which is an educational mode of online education for students with the help of network information platform before class, focusing on collaborative inquiry, interactive communication, promoting knowledge internalization and consolidating knowledge after class. Inspired by the visual mechanism of human brain, the proposal of deep learning makes a breakthrough in machine learning, and also brings a new direction to the research of human motion recognition. Deep learning is based on a series of algorithms, and unsupervised high-level abstraction of data is obtained through hierarchical nonlinear transformation. This paper focuses on human motion recognition in complex scenes and the extraction of spatio-temporal features in motion video, so as to overcome the difficulties caused by environmental differences and time changes. Based on the study of CNN and deep confidence network, an innovative human motion recognition model is proposed in this paper.

Keywords. AI, PE teaching action, Deep learning

1. INTRODUCTION

The Human action recognition has always been a research hotspot and difficulty in computer vision, and it also has important application value in real life. Human action recognition is widely used in video surveillance, virtual

reality, human-computer intelligent interaction and other fields (Chu, Shih, Chou, Ahamed, & Hsiung, 2019). Most of the current research starts from three aspects: moving target detection, action feature extraction and action feature understanding, and basically solves the problem of human action recognition in simple scenes (Guang, Jing, & Fu, 2011). Today, surveillance cameras are all over the corners of our lives, and the role of cameras is not only monitoring, but also helping us obtain areas of interest and targets, helping humans to better complete the expected work (Guo & Niu, 2021). In the field of machine vision, target detection, recognition, location, tracking and navigation play a very important role. School sports is the foundation of national sports, and the perfection of school sports directly leads to the physical condition of students (L, 2018). Therefore, we should take school sports as the foundation, cultivate students' concepts and habits of lifelong sports, and lay a good foundation for students' lifelong sports.

At the present stage of PE teaching, many teachers neglect the teaching of students' necessary motor skills, and make specific research and analysis on the accuracy of teachers' teaching actions in PE classes. PE class only studies vividly, and does not ask students how to master skills; Just go through the motions, not improve your skills; Only seeking pleasure, not seeking the actual effect of physical exercise, neglecting the teaching of sports skills and losing the foundation of PE (Lanius, Kobayashi, Ouchi, & Aoki, 2018). From this, we can know that it is one of the ultimate goals of quality education to make students master sports skills and cultivate lifelong sports awareness (Li H, 2017). Mastering sports skills and getting fun from sports is a necessary condition for cultivating lifelong sports. The study of action elements will have a great impact on future PE teaching. Therefore, the application of action elements in PE teaching should be emphasized, teachers' strict requirements and students' earnest study, so as to solve the problem of teachers' accuracy in imparting technical actions in PE teaching, and thus make due contributions to the development of PE teaching (Li J, 2020).

Deep learning is developed on the basis of machine learning. It adopts hierarchical processing mechanism and can automatically learn advanced features layer by layer from input data (Libing J, 2013). Therefore, through unsupervised or semi supervised feature learning and hierarchical feature extraction, deep learning has successfully replaced the manual feature extraction method. Specifically, relevant technologies will be used in traffic scheduling, urban public security, gymnastics rehearsal, stage scene analysis, etc. many scenes can be detected and positioned through target detection and action recognition, which can greatly improve the efficiency of work and reduce the consumption of human and material resources. In school education, the process of PE is the longest, from kindergarten to university to graduate students. The development of school PE is closely related to the development of social history and the inheritance of traditional culture. Therefore, the current development of school PE in our country must also be adapted to our country's national conditions and social needs.

With the proposal of the concept of deep learning, a large number of

scholars have invested in deep learning and made great progress and innovation. This paper mainly studies the real-time detection and location of human moving objects and action recognition based on deep learning. First, the traditional moving object detection and location methods and some action recognition methods are studied and summarized, and the advantages and disadvantages of these methods are analyzed. Combined with deep learning, a reasonable scheme is designed for the stage experiment scene in this paper to achieve the detection and location of human bodies in the scene and action analysis.

2. Related Work

At present, the mainstream processing methods of human motion recognition can be divided into two categories: global feature method and local feature method. The global feature method uses the background difference method or the tracking method to locate the entire human body of interest, and uses the edge, silhouette outline, optical flow and other information of the human body to describe the human body (Yang, Jiang, Rong, & Xu, 2020). Li J et al., Representation and recognition of human motion in the form of temporal templates. Two templates are proposed: motion energy map and motion history image (Zhong H, 2014; Zhu, 2021). The motion energy map is to differentiate and binarize the adjacent images, and then obtain the motion area image by superimposing, and obtain the historical motion image after weighting according to the time sequence of the action, and finally obtain the feature representation based on the moment (Liu H B. [J]. 2018). Ma C proposes a body tracking architecture for periodic repetitive motion. Fourier transform was used to find the period of motion and meshed with the human body as the center, and the motion features were matched with the labeled 2D motion template (Liu, Dong, & Wang, 2021). Wang M proposed a target detection method based on Deformable Part Model (DPM), which is a contour-based star structure. It consists of root node filter, part detector, considered part detector and global detector. Calculate the dot product of the filter weights and the sub-window feature vector to obtain the values of the global detector and each part detector, and finally calculate the DPM value [13]. Chu C C et al. put forward R-CNN (region-based CNNs), which is the first time to apply deep learning to target detection. It is a target detection method that combines CNNs with Regionproposal prediction [14]. Guo W clarified the different definitions of teaching methods and modes, and based on further perfecting the traditional teaching theory of complete and decomposed teaching mode. This paper mainly introduces the application and successful experience of systematic method and program teaching mode in track and field teaching, and points out the problems to be solved and the basic ways to develop track and field teaching theory. Teachers should teach students correct body posture and clear movement track, and make good movement speed with effective movement time, so as to achieve ideal movement speed and strength. And rhythm control the whole process of learning and competition (M, 2021). Zhong h pointed out that the scientific analysis of the formation process of PE teachers' teaching ability found that PE teachers have the laws of gradual, complete, practice, guidance and integration, and then put forward suggestions to improve PE teachers'

teaching ability according to the influencing factors of PE teachers' teaching ability (Ma & Yang, 2021). Lanius C put forward the concept of optical flow: optical flow reflects the motion speed of real objects in space according to the instantaneous speed of pixels on the imaging plane (Xiao-Mei L V, 2016). It is a method of calculating object motion information by using the temporal changes of pixels in the video sequence and the correlation between adjacent images to find the corresponding relationship (N, 2015). Libing J et al. obtained motion energy map (MEI) and motion history map (MHI) by superimposing the motion silhouette map of the human body, and matched these two feature maps with templates to realize action recognition (Noussia, Nedeva, Aakaansha, Wang, & Glynou, 2022). Bu G proposed a bottom-up method to associate partial detection candidates with a single human body, however, the final detection time is relatively long (Paek & Kim, 2021). On the basis of [19], Fan L et al. combined the detection method of image pairing score with the ResNet network, which greatly improved the computational efficiency, but it took several minutes to detect each image, and it was still impossible to achieve real-time detection (U, 2016).

This paper focuses on the recognition of human actions in complex scenes and the extraction of temporal and spatial features in sports videos, so as to overcome the difficulties caused by environmental differences and time changes.

3. Methodology

3.1 Deep learning combined with NN technology is used to analyze PE teaching actions

The description of motion trajectory is of great significance in the fields of motion analysis, behavior detection, video description and so on. Body posture refers to the state of the body and all parts of the body in each phase of the exercise process. In the practice of movement speed, the special requirements are different, and the tasks and contents of movement speed practice are also different. Therefore, movement speed is closely related to the perfection of movement technique. To a great extent, the improvement of movement speed depends on the perfect movement technique, because the movement amplitude, working distance, working time, movement direction, angle and position are all closely related to the movement speed. Secondly, in technical practice, the coordination of human body will be improved accordingly.

CNN (CNN) is an ANN, which is also the first successful algorithm for effective training and learning of multilayer network structure. CNN is proposed to minimize data preprocessing and directly extract the most expressive features from the original data input, rather than manually specifying features. CNN is also a multilayer feedforward NN, and it is also a supervised learning. Generally, the gradient descent method is used for supervised learning and training. Generally, it includes three structures: convolution layer, pooling layer and full connection layer; The pooling layer is usually connected after the convolution layer, alternating with the convolution layer, and finally connected to the full connection layer.During the convolution

process, convolution kernels of different sizes can be used to enhance the generalization performance of the entire network; at the end of the entire network, a classifier, usually SoftMax function or SVM, will be connected to the final input data. Classification: (1) Convolution layer: Convolution layer is the most important part of CNN, which is mainly used for image feature extraction; (2) Pooling layer: Pooling layer is mainly based on the principle of local correlation of image space, that is, according to the high similarity between each pixel in the image and the surrounding points, by aggregating and counting the features of local areas of the image, it can effectively reduce the dimension of the feature graph, and keep the effective information in the graph that is useful for training tasks, so the trained model has certain robustness, that is, it has certain fault tolerance for small translation and scaling of the input graph;

3.2 Optimal Design of PE Teaching Action Based on BP NN

Human behavior recognition aims to identify the behaviors that occur in an image sequence or video data. Since 1980s, this research has attracted the attention of computer science. At present, in the field of motion recognition research in the industry, when the input video sequence is a skeleton sequence, the skeleton distribution is often shown in Figure 1.

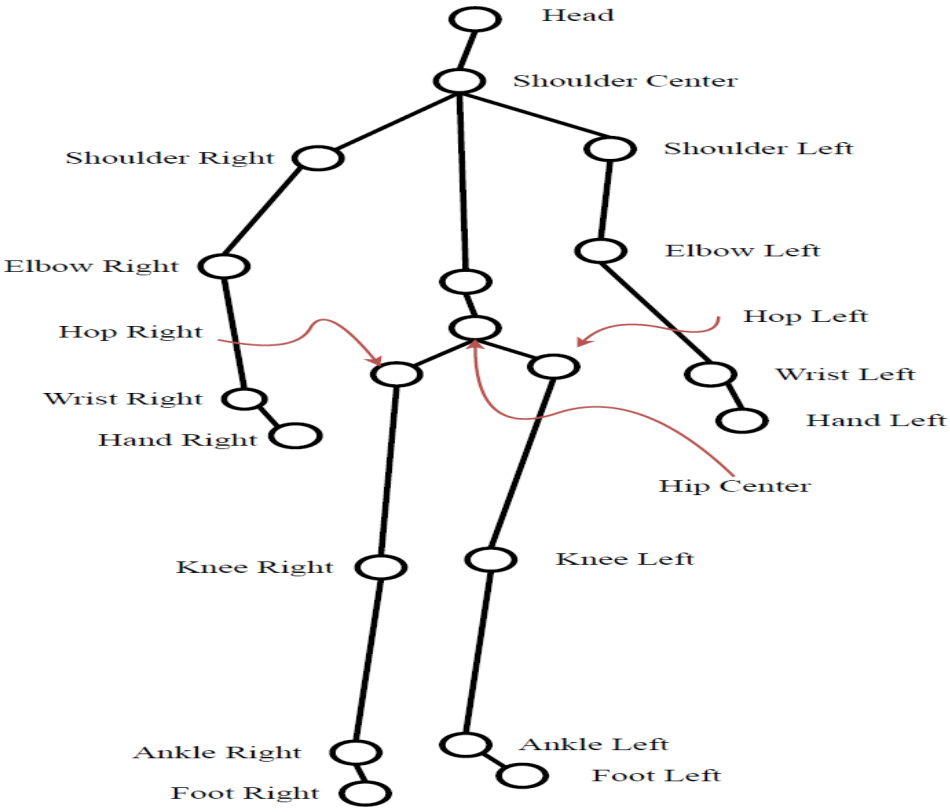


Figure 1. Kinect parses the skeleton nodes according to the depth map

For each three-dimensional bone node, its coordinates can be projected on three planes of Cartesian orthogonal system. For a certain bone node position, the position at the next moment and the bone displacement between two adjacent frames are shown in Figure 2.

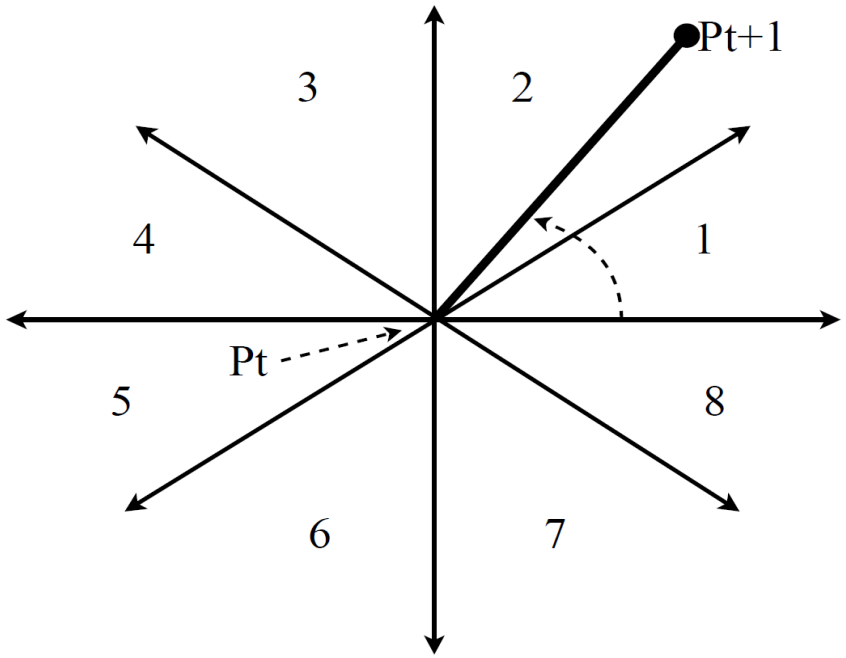


Figure 2. Coordinate representation of skeletal nodes in two adjacent frames

According to figure 2, first introduce the hod calculation method on a two-dimensional projection plane: the displacement between these two moments can be divided into different areas according to the direction. As shown in the figure, the plane of each projection is evenly divided into 8 regions. For a bone node, a histogram with a dimension of 8 can be obtained by counting the number of times each displacement falls in different areas in the whole video range. The physical meaning of each dimension is the frequency that the displacement of the bone falls in this direction in the whole video sequence.

An important issue in the design of computer vision algorithms or computer vision systems is the selection of feature representations. In many cases, the high-level details in the feature description are the key to solving the problem. The purpose of deep learning is to learn the hierarchical relationship between features from low-level features, so as to obtain high-level features. Deep learning is a branch of machine learning, also known as deep structured learning, hierarchical learning or deep machine learning. Based on a series of algorithms, it obtains high-level abstract models in data through complex structures of multi-processing layers or multi-level nonlinear transformation operations. The motivation of deep learning is to build a model that simulates the human brain for analytical learning, which imitates the visual mechanism of the human brain to explain data, such as images, sounds and texts. This model simulates the thinking process of the brain by combining the mathematical model of logical threshold with some mathematical algorithms.

In 1960, BP model (backpropagationmodel) was established by henryj Kelley proposed that BP model is one of the most widely used algorithms in NNs until now because of its simple structure, complete theoretical knowledge

and learning mechanism. The basic unit of ANN is neuron, which is also the most basic unit of this algorithm model to process all information. By simulating the process of human neuron processing external stimuli, the neuron appears as a many-to-one nonlinear mapping unit. Its abstract mathematical model is shown in Figure 3.

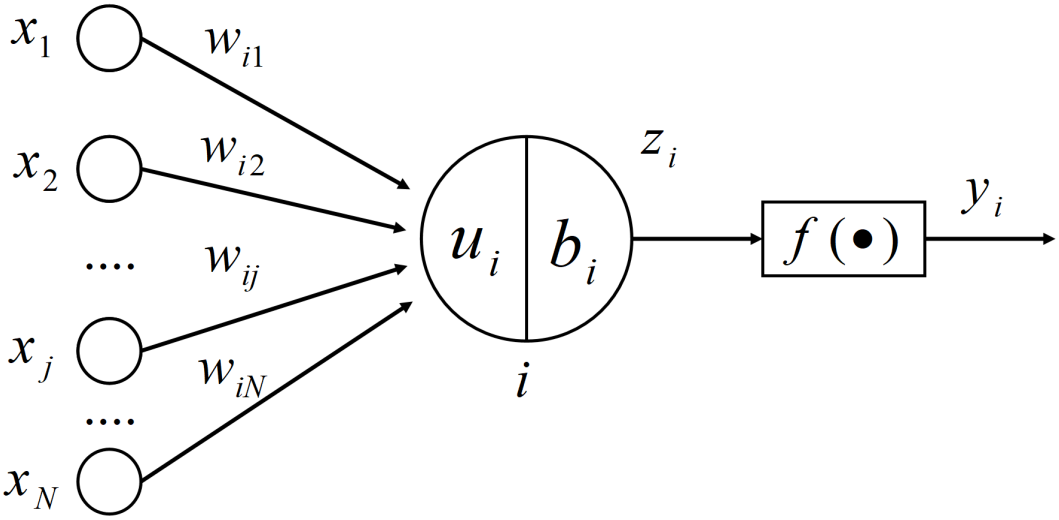


Figure 3. Artificial neuron model

I neurons can have multiple inputs, but only one output, that is, one to many. The neuron I receives the input signal $x_j(j-1,2,\ldots,N)$ from the neuron J, and the connection weight between two adjacent neurons is w_{ij}. u_i is the linearly combined value of each input signal, that is, the weighted sum of the input values, $u_i = \sum_{j=1}^{N} w_{ij} x_j$; b_i is the bias, and z_i is the final input value of neuron I, namely $z_i = u_i + b_i$; F () is the activation function, and z_i can get the output y_i after the activation function. The function of activation is to introduce nonlinear factors, so that all neurons can produce nonlinear output.

3.3 Motion recognition model based on CNN

In machine learning, CNN is a kind of backward propagation ANN, in which the neurons of each layer are connected with the local visual field area overlapped by the previous layer. Inspired by biological processes, convolutional network is a multi-layer perceptron that tries to use the least number of preprocessing changes. The whole model consists of three basic building blocks stacked, which are convolution layer, maximum pooling layer and classification layer. The image data is directly used as the input of the network without too much preprocessing. After 1 to 3 feature extraction stages, the classifier is connected to obtain the output. One of the most common pool layer settings is shown in Figure 4. A filter with a size of 2×2 performs the maximum pool operation on the image in steps of 2, and the size of the feature map is reduced to the original size of . As shown in Figure 4.

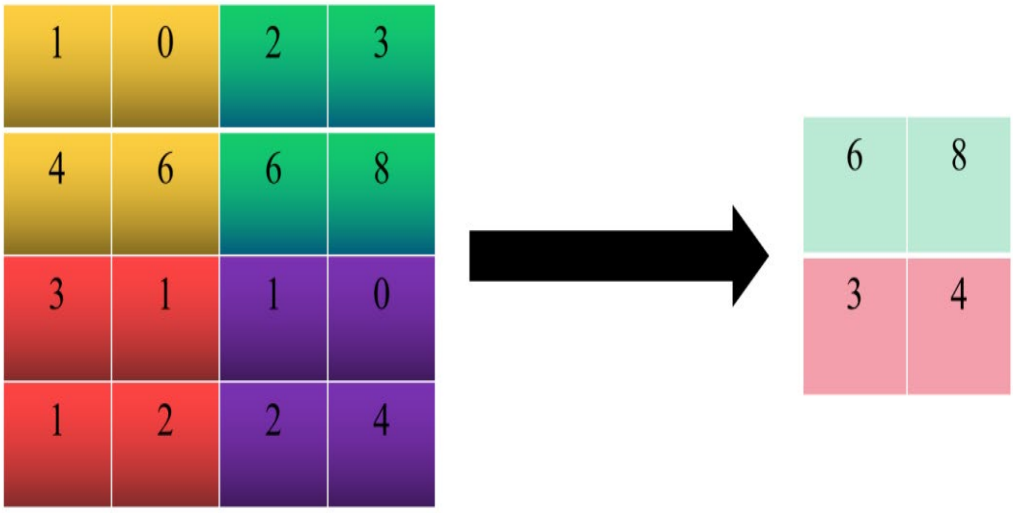


Figure 4. Maximum pooling operation diagram

Convolution network is supervised training, and its training sample set is composed of vector pairs (input vector, ideal output vector) collected in the actual running system. Before training, use some different small random numbers to initialize all weights, that is, pre training. Among them, the guaranteed weight of "small random number" is too large to exceed the capacity limit of the model, resulting in training failure. "Different" ensures that the network can learn parameters normally, because the same number of initialization weights will make the network unable to learn.

The training algorithm of CNN is similar to BP algorithm, which is also divided into successive stages: The first stage, forward propagation:

(1) Take a sample (X, Y_p) from the sample set and input X into the network; (2) Calculate the corresponding actual output O_p .

Among them, the sample information is input from the input layer and reaches the output layer after multiple transformations. At this stage of the network, the following transformations are performed:

$$O_p = F_n(\ldots(F_2(F_1(X\ {}_pW(1)\ )W(2)\ )\ldots)W(n)\ ) \qquad (1)$$

The second stage, backward propagation:

(1) the difference between the actual output O_p and the corresponding ideal output Y_p; (2) Adjust the weight matrix by the method of minimum haul error.

The backward propagation stage corresponds to the normal propagation of the previously input sample information. When adjusting the connection weights of neurons.

The error between the output and the ideal input is calculated first, and then the errors of other layers are pushed back layer by layer by using this

error, so it is also called the error propagation stage.

$$E_p = \frac{1}{2}\sum_{j=1}^{m}(y_{pj} - o_{pj})^2 \quad (2)$$

In order to explain the specific training process of CNN more clearly, the following assumptions are made: the number of units in input layer, intermediate layer and output layer are n, l and m respectively.

The input vector of the network is and the output vector of the middle layer is $X = (x_0, x_1, \ldots, x_N)$. The connection weight from the output vector unit I to the hidden unit J is V_{ij}, and the connection weight from the hidden unit J to the output vector unit is W_{jk}. The thresholds of the two units are θ_k and φ_j respectively.

Then, the output of the middle layer neural unit is:

$$h_j = f(\sum_{i=0}^{N-1} V_{ij} x_i + \phi_j) \quad (3)$$

The output of each unit of the output layer is:

$$y_k = f\sum_{j=0}^{L-1} W_u h_j + \theta_k \quad (4)$$

Here, the excitation function uses the sigmoid function:

$$f(x) = \frac{1}{1+e^{-kx}} \quad (5)$$

By calculating formula 4, an intermediate layer output vector h is obtained. Then the actual output vector y of the network is obtained by calculating formula 5.

Compute M output errors, where and represent the components in the output and target vectors, respectively.

$$\delta_k = (d_k - y_k) y_k (1 - y_k) \quad (6)$$

Similarly, calculate the L intermediate layer errors:

$$\delta_j = h_j (1 - h_j) \sum_{K=0}^{M-1} \delta_k W_{jk} \quad (7)$$

The adjustment formulas for calculating the weights and thresholds respectively are as follows:

$$\Delta W_{jk}(n) = (a/(1+L)) * (\Delta W_{jk}(n-1) * \delta_k * h_j \quad (8)$$

$$\Delta V_{ij}(n) = (a/(1+N) * (\Delta V_{ij}(n-1)+1) * \delta_k * h_j \quad (9)$$

$$\Delta \phi_j(n) = (a/(1+L)) * (\Delta \phi_j(n-1)+1) * \delta_j \quad (10)$$

When k is from 1 to m, judge whether the index meets the precision requirement: $E \leq \varepsilon$, where e is the total error function, and $E = \frac{1}{2}\sum_{k=0}^{M-1}(d_k - y_k)^2$. If not, go back to (3) and continue the iteration. If satisfied, go to the next step.

4. Result Analysis and Discussion

This paper uses kth data set for experiments, as shown in the following table. The data set includes six kinds of human behaviors: walking, jumping, running, boxing, waving, clapping (walking, jogging, running, boxing, handwaiting, handclapping), which are performed by 25 different people in four scenarios respectively. S1: outdoor, S2: outdoor with angle changes, S3: different outdoor clothes, S4: indoor. At present, there are 2391 video sequences in the data set. The background is relatively static, except for the close and far of the lens, the motion of the camera is relatively slight. The sequences have been downsampled to a spatial resolution of 120160 pixels and an average length of 4 seconds. In this experiment, all sequences are divided into a training set (8 individuals), a validation set (8 individuals) and a testing set (9 individuals). The classifier is trained on the training set, and the validation set is used to optimize the model parameters, and finally the recognition results are obtained through the test set.The network input is a series of video sequence frames of size, which is reduced by four times from the KTH data set described above. In the human motion recognition model based on CNN, the parameters of each layer of CNN are shown in Table 1.

Table 1. Specified parameters of multilayer structure in 1CNN structure

layer	Type	Characteristic diagram	Nuclear size
0	Input	1	-
1	Convolution	5	3*4
2	Pooling	6	6*6
3	Convolution	11	2*2
4	Pooling	12	3*3
5	Full connection	-	-
6	Classification	-	-

The network has 6 hidden layers: (a) a convolutional layer with 6 convolutional filters of size ; (b) a max-pooling layer with a scaling factor of ; (c) a convolutional layer with 12 convolutional filters of size ; (d) max pooling layer with scaling factor; (e) fully connected layer; (f) classification. The softmax regression classifier in the last layer of CNN is trained with the feature vectors of high-order training samples, and each query image can be classified according to its feature vector.

At point P, there is a symmetrical curve for different real points x, as shown in Figure 5, and the maximum value operation is performed on:

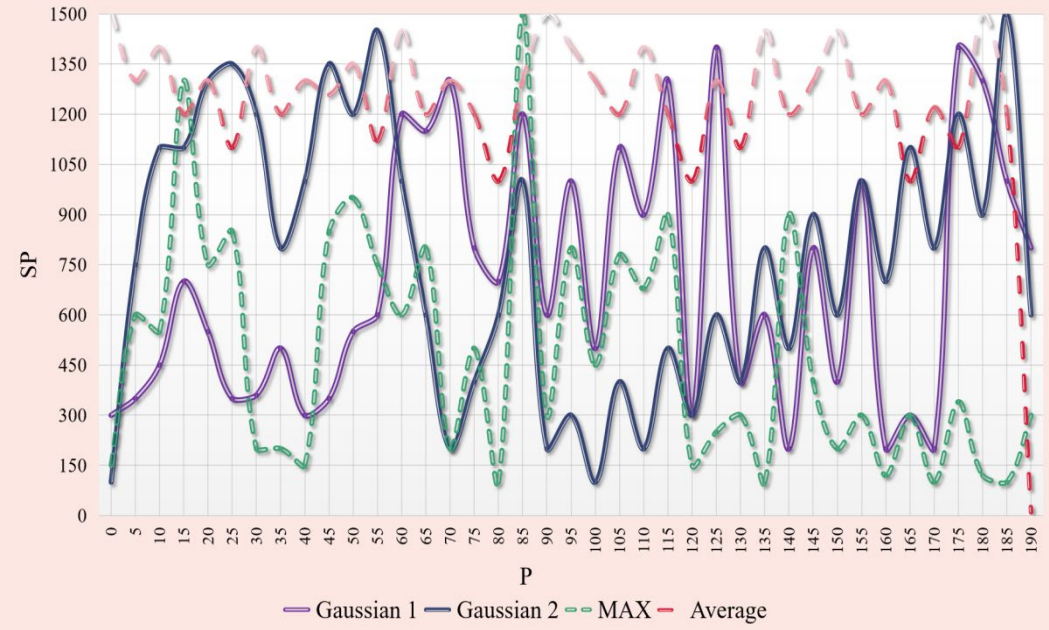
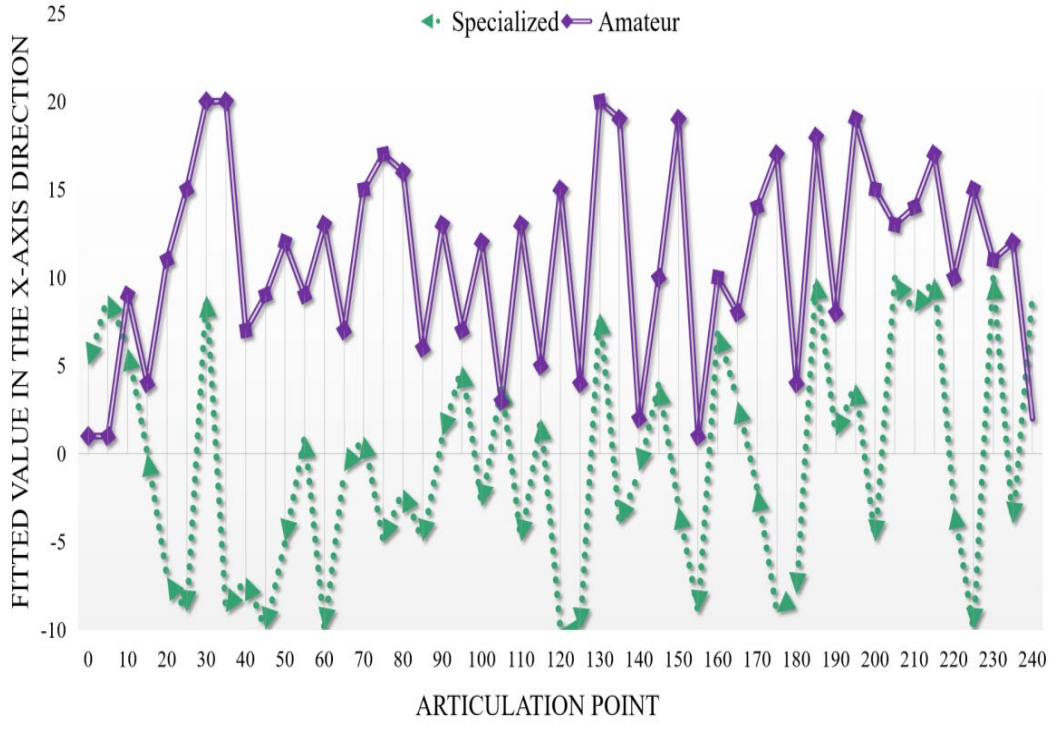
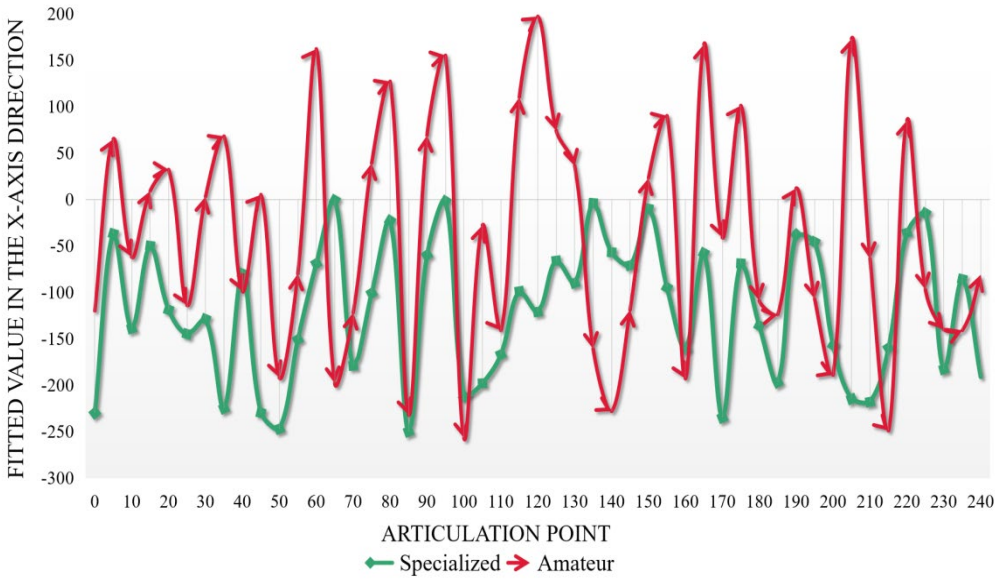


Figure 5. Curve representation of joint points

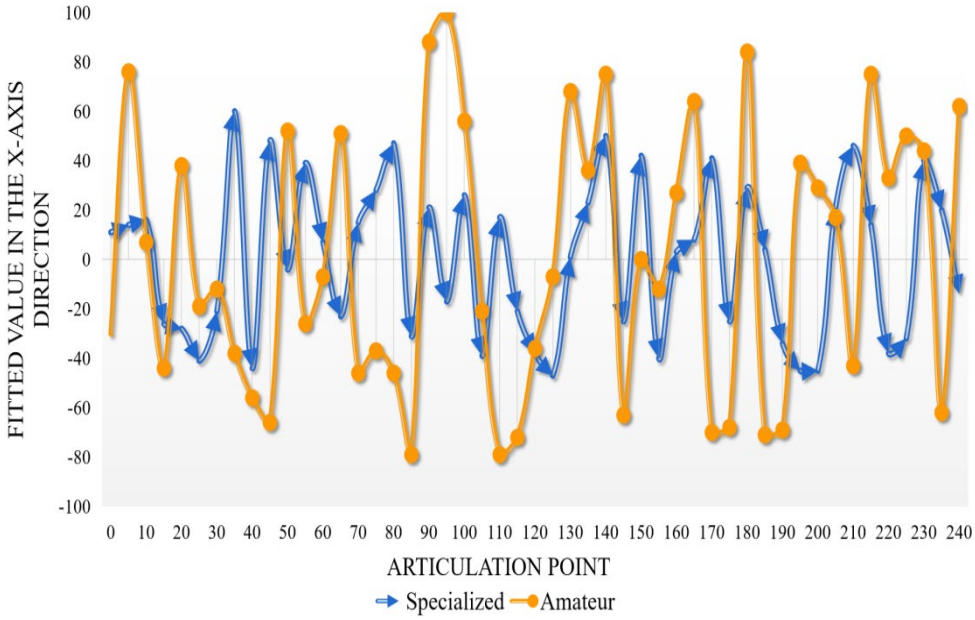
In the polynomial fitting of this paper, in order to set the origin of coordinates as the midpoint of two hipbones, it is assumed that all coordinates are on the x-y plane. As mentioned in this paper, the joints of human body are divided into three parts: BB, UB and LB, and each part is fitted by polynomial from the direction of X axis. As shown in Figure 6.



(a) partial fitting of x-axis BB



(b) Partial fitting of x axis UB



(C) x-axis LB partial fitting

Figure 6. Polynomial fitting diagram in the X-axis direction

Figure 6 is a schematic diagram of polynomial fitting of randomly selected data of a professional model and an amateur model, which is an effect diagram for measuring polynomial fitting.

In this paper, the 10 fold cross validation method is used to test all training samples, and the average accuracy is 71.9%, which preliminarily realizes the professional evaluation of the model walking on the platform. In order to compare the results of different models' motion recognition, this paper will test each model one by one. Randomly select all key frame data of 5 professional models and 5 non professional models in the test set for testing,

The results are shown in Table 2.

Table 2. Individual test results

Detection result	Major 1	Major 2	Major 3	Major 4	Major 5	Amateur 1	Amateur 2	Amateur 3	Amateur 4	Amateur 5
Correct quantity	36	36	50	48.23	38	38	40	45	76	58
Number of errors	10	12	3	6	5	12	10	11	13	16
Accuracy /%	75.2	70.3	90.3	85.2	80.2	76.3	72.8	80.2	81.77	70.63

Physics defines velocity as the magnitude of velocity. To improve the speed quality of athletes through certain methods and means has positive significance for the development of athletes' rapid movement ability. Since the speed quality includes three aspects: reaction speed, action speed, and displacement speed, and these three aspects are both related and different, the way to improve speed capability also has many characteristics. Among the action elements, the action speed is the starting point The main role, action time, and action trajectory also show a related practice relationship. This concept is not easy for students to understand. How to explain and explain is the most important thing in teaching. ways to make this concept known and understood by students.

Kinetosensory Exercise: Kinetosensory exercise is a method of combining physical training with mental training. In the process of human reaction, improve the time perception of minute time discrimination, so as to develop the reaction speed. This exercise has certain practical significance for sports practice. Kinesthetic exercises generally go through three stages.

The operator repeats the action instructions within the range of 5-90m from the UAV, and makes all five control instructions 20 times at intervals of 5m, that is, a total of 100 effective control instructions. During this period, he will walk around in different distance ranges and perform a series of interference actions (such as kicking, turning, jumping, etc.). As shown in Figure 7.

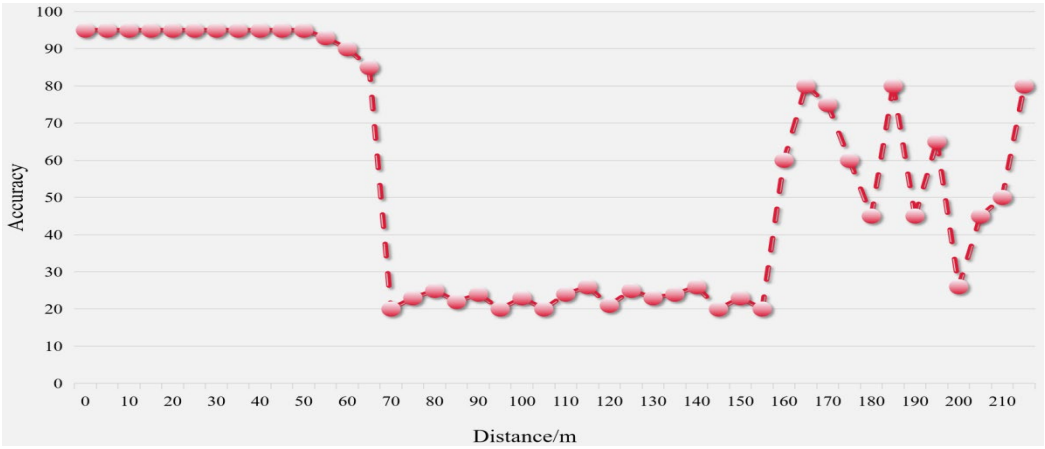


Figure 7. Relationship between system recognition distance and recognition accuracy

Figure 7 shows the corresponding relationship between recognition distance and recognition accuracy. It can be seen from the figure that the recognition accuracy is high within the range of 5-60m from the operator to the UAV. In this range, a small number of actions are confused with "fly up" commands, and the error recognition here is mainly due to the fact that non control commands contain more similar texture maps. However, after the distance exceeds 60m, the recognition accuracy decreases significantly, which is mainly due to the low resolution of the operator in the picture, so that it is unable to effectively calculate the optical flow characteristics.

The processing steps of training data and test data are slightly different. The training data should be calculated by the mean value, that is, the sample mean value. Each training data should be subtracted from the sample mean value. Then the covariance matrix is calculated, and the eigenvalues, eigenvectors and dimension reduction matrix are solved. The dimension reduction matrix is determined according to the contribution rate of the calculated eigenvectors. The final training data can be obtained by multiplying the de-averaged data of the training samples by the dimensionality reduction matrix. SVM maps the sample space to a high-level feature space (Hilbert space) through a nonlinear mapping, that is, a kernel function, so that the nonlinearly separable problem in the original space is transformed into a linearly separable problem in the feature space. In this paper, the Gaussian kernel function is used to transform the feature space, and finally the action recognition is realized through the SVM training model. The number of hidden units has a great influence on the distribution of DBN, and if this parameter is set incorrectly, it will damage the representation ability. Some theories suggest that the optimal number of hidden units can represent any data distribution.

Recently, experiments show that when the number of hidden units is very small, increasing the hidden units of RBM will have a strong impact in the classification stage. On the other hand, the improvement of performance can't prove that the increase of hidden cells is correct, and the amount of computation will increase. Because of a large number of parameters and their possible configurations, the grid search method cannot be carried out on this depth model. For this reason, this chapter focuses on a hierarchical network. Because in order to better describe different actions, the feature space is divided at the lower level to increase the number of regions. Due to the correlation between regions, the dimension of feature vector and the number of regions are reduced hierarchically. In addition, since the quality of data essentially depends on the length of records, it is found that the quality of data also has a great impact on the effectiveness by observing the performance of records with different lengths. In fact, a higher quality of data can provide more input data for the classification of actions. Of course, as the input of the deep network, the increase in the number of unlabeled data can greatly improve the performance of the method in this chapter, making it more suitable for practical applications that cannot provide a large number of labeled data.

5. Conclusions

Based on the CNN in deep learning, this paper studies the detection

and recognition algorithms of human actions for different video sequence formats (normal RGB video, deep video sequence, skeleton sequence). From the composition of action elements to guide the learning of motor skills, it means that beginners can actively analyze, comprehend, and understand the characteristics of the action elements of the motor skills at a specific learning stage, and use this to determine how to learn actions and how to practice them. A learning method that ultimately masters the skill. The results show that the use of imagery training in closed sports skills in middle schools can significantly improve the achievement of fixed-point shooting and quick jump and technical evaluation, and the effect is remarkable. The physiological essence of learning and mastering motor skills is to establish complex and interlocking proprioception's motor conditioned reflex. According to the reflex arc principle, the receptor receives various changes (and stimuli) inside and outside the body, and converts the stimulus energy into nerve impulses, which are then transmitted to the central nervous system, such as nerves, and the information is transmitted to the effector through the efferent nerves through the central analysis and comprehensive action. Aiming at the research scene of sports, we conduct in-depth research on the detection and positioning of human body and action recognition in the scene, constantly improve the performance of optimization algorithm, and finally realize the real-time and accurate detection and positioning of human body in the stage scene, as well as an objective evaluation of human action.

References

Chu, W. C.-C., Shih, C., Chou, W.-Y., Ahamed, S. I., & Hsiung, P.-A. (2019). Artificial intelligence of things in sports science: weight training as an example. *Computer, 52*(11), 52-61.

Guang, B., Jing, L., & Fu, Z. (2011). *The application of the multimedia technology in the dance teaching of college sports.* Paper presented at the 2011 IEEE International Symposium on IT in Medicine and Education.

Guo, W., & Niu, Y. (2021). Interactive teaching system of basketball action in college sports based on online to offline mixed teaching mode. *Mobile Information Systems, 2021*, 1-10.

L, F. (2018). A Preliminary Study on Learning Evaluation Design of Students in PE Specialized Teaching Based on "Sports Skill Fitness". *Bulletin of Sport Science & Technology, 58*(6), 24-62.

Lanius, C., Kobayashi, D., Ouchi, K., & Aoki, Y. (2018). *Single Image, Context Aware Action Estimation in Sports.* Paper presented at the 2018 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS).

Li H, Z. M. (2017). Research on the teaching design and experiment of sports micro course based on the fusion of motion capture technology. *Journal of Liaoning Normal University(Natural Science Edition), 47*(62), 64-36.

Li J, L. S. (2020). The Integration of Reading and Writing from the Perspective of Deep Learning:Theories and Teaching Cores. *Curriculum,Teaching Material and Method, 678*(27), 658-687.

Libing J, D. C., Jing Z. (2013). Development and Application of Three-

dimensional Teaching Resources of Modern Sports Educational Technology Based on B-Learning. *Journal of Wuhan Institute of PE, 48*(68), 6-8.

Liu H B. [J]., 248(76):48-763. (2018). Making good use of Mobile Phone Image Records to Improve the Effectiveness of Action Teaching. *Fujian Sports Science and Technology, 248*(76), 748-763.

Liu, Y., Dong, H., & Wang, L. (2021). Trampoline motion decomposition method based on deep learning image recognition. *Scientific Programming, 2021*, 1-8.

M, W. (2021). Design of College PE Teaching System Based on AI Technology. *JPhCS, 13*(6), 73-56.

Ma, C., & Yang, P. (2021). *Research on classroom teaching behavior analysis and evaluation system based on deep learning face recognition technology.* Paper presented at the Journal of Physics: Conference Series.

N, Z. (2015). Application of Action Comparative Evaluation Software in the Teaching Process in PE. *Contemporary Sports Technology, 48*(623), 567-579.

Noussia, K., Nedeva, S., Aakaansha, A., Wang, C., & Glynou, M. (2022). Bias of arbitrators: a critical analysis on the law post-Halliburton v. Chubb and a comparative approach. *Journal of International Trade and Arbitration, 2*(1), 31-92.

Paek, S., & Kim, N. (2021). Analysis of worldwide research trends on the impact of artificial intelligence in education. *Sustainability, 13*(14), 7941.

U, J. L. (2016). Research on Sports Teaching Practice Based on the Technology of VR——Taking Golf as an Example. *China School PE(Higher Education), 25*(5), 53-750.

Xiao-Mei L V, J. D. U. J. (2016). The Application of Sports Technique Diagnosis to Sports Teaching Practice. *Theory and Practice of Education, 509*(7), 673-675.

Yang, X., Jiang, X., Rong, L., & Xu, Z. (2020). A sports teaching mode based on social networking service teaching resources. *International Journal of Emerging Technologies in Learning (iJET), 15*(8), 180-194.

Zhong H, D. Z. (2014). Basic action teaching methods in Basketball Teaching. *Contemporary Sports Technology, 562*(56), 49-24.

Zhu, L. (2021). Computer vision-driven evaluation system for assisted decision-making in sports training. *Wireless Communications and Mobile Computing, 2021*, 1-7.