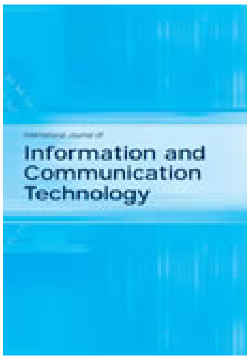


International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

An improved deep neural network method for an athlete's human motion posture recognition

Zhe Dong, Xiongying Wang

DOI: 10.1504/IJICT.2023.10052030

Article History:

Received:	02 November 2020
Accepted:	05 January 2021
Published online:	14 December 2022

An improved deep neural network method for an athlete's human motion posture recognition

Zhe Dong and Xiongying Wang*

College of Physical Education,
Hunan City University,
Yiyang 413000, China
Email: 85466731@qq.com
Email: xiongying33@mls.sinanet.com
*Corresponding author

Abstract: Aiming at improving the accuracy of motion gesture recognition and reducing the time-consuming recognition, this paper proposes an athlete's body motion gesture recognition method based on improved deep neural network. According to the movement characteristics and basic structural characteristics of the human body, the movement posture recognition standard is designed and set in the model. Collect athletes' human motion images, and perform motion image extraction, residual compensation, filtering, normalisation, and morphological processing. On this basis, the features of human motion images are extracted and fused. The recognition classifier is constructed based on the improved deep neural network, the fused feature vector is input into the classifier, and the recognition result is output. The results of the comparative experiment show that the proposed method has a high recognition rate and short overhead time only at 1.2 s.

Keywords: improved depth neural network; athlete; human motion; posture recognition; residual compensation; filtering; normalisation.

Reference to this paper should be made as follows: Dong, Z. and Wang, X. (2023) 'An improved deep neural network method for an athlete's human motion posture recognition', *Int. J. Information and Communication Technology*, Vol. 22, No. 1, pp.45–59.

Biographical notes: Zhe Dong received his Master's in Physical Education and Sports Training at North West Normal University in 2006. Currently, he is a Lecturer in College of Physical Education at Hu Nan City University. His research interests include physical education and sports training, sports facilities managing.

Xiongying Wang received her Master's in Developmental and Educational Psychology at Central China Normal University in 2012. Currently, she is a Lecturer in College of Physical Education at Hu Nan City University. His research interests include education psychology, education managing.

1 Introduction

Human behaviour recognition has a wide range of applications in real life, such as intelligent video surveillance, customer type, shopping behaviour analysis and so on. However, due to the influence of cluttered background, occlusion and line of sight changes, the accurate recognition of human behaviour and posture is limited. Due to the wide application of surveillance video in daily life, effective video understanding technology is particularly important, especially the sports posture recognition of athletes. Athletes' motion posture recognition method can be applied to athletes' daily training and related competitive events. In daily training, through the recognition results of sports posture, we can analyse athletes' exertion mode, movement track and other behaviours, avoid the repeated use of wrong sports actions, and minimise the possibility of athletes' injury in the process of training to avoid the use of illegal actions in practice (Zhu et al., 2018). In the competitive events of athletes, it is more convincing to use the method of human body's motion posture recognition as a reference for judging whether the athletes violate the rules. The research of athlete body's motion posture recognition method started in 1970, and has made great progress in the following 20 years. Many effective methods of computer vision information processing have emerged. After 21 years of practice, the research on the analysis of athlete body's motion posture and the detection of abnormal behaviour time have made rapid development (Fei et al., 2018).

At this stage, the research on the method of athlete's human motion posture recognition has made certain progress. For example, Liu (2019) proposed a method of athlete's human motion posture recognition based on multi feature fusion. This paper uses optical image collector to collect athletes' body motion posture images, and carries out gray level transformation on the collected images, and uses shadow elimination technology and frame difference method to obtain body contour and movement posture area. Based on radon transform and discrete wavelet transform, the motion pose region and body contour are extracted, and the final pose recognition is realised by fusing the two complementary features. However, this method is complex and time-consuming. In Li et al. (2018), a method of athlete's posture recognition based on MEMS sensor is proposed. In this method, MEMS sensor is used to collect the change signal of athletes' body posture, and the method of smoothing filtering and searching for the peak, trough and zero point of the signal is used to segment the signal, and each individual action data is extracted, and the feature value of each movement data is extracted, and the collected training samples are trained by using BP neural network algorithm. BP neural network is used to output the recognition results of athletes' body movement posture. However, this method has the problem of low correct recognition rate, which is difficult to achieve the ideal application effect. Cheng and Zhou (2020) proposed a method of human motion posture recognition based on joint points. This method combines the background difference method and the inter frame difference method to detect the foreground of athletes' image, and filter the segmentation results by morphological processing. Eight neighbourhood algorithm is used to skeletonise human silhouette, and a multi feature fusion method is proposed to extract human joint points. Then, Lucas Kanade optical flow algorithm and Kalman filter are used to track and predict the motion of joint points, and the results of human motion pose recognition are output. However, this method can only deal with human movement in limited environment such as fixed view angle and static background, so the correct recognition rate is not high.

In order to solve a series of problems in the traditional human motion gesture recognition method, the concept of an improved deep neural network is introduced, and a motion gesture recognition method based on the improved deep neural network is proposed. The specific implementation plan is as follows:

1 According to the sports characteristics of athletes and the basic structural characteristics of human body, the human motion model is built, and the recognition standard of motion posture is designed and set in the model.

2 The human motion image of athletes is collected, and the image pre-processing is realised through the steps of motion image extraction, residual compensation, filtering, normalisation and morphological processing. On this basis, human motion image features are extracted and fused.

3 Depth is improved neural network architecture constructed based on motion recognition classifier, fusion moving image processed as an input variable, input to the motion classification, the input is the result of the exercise posture recognition result.

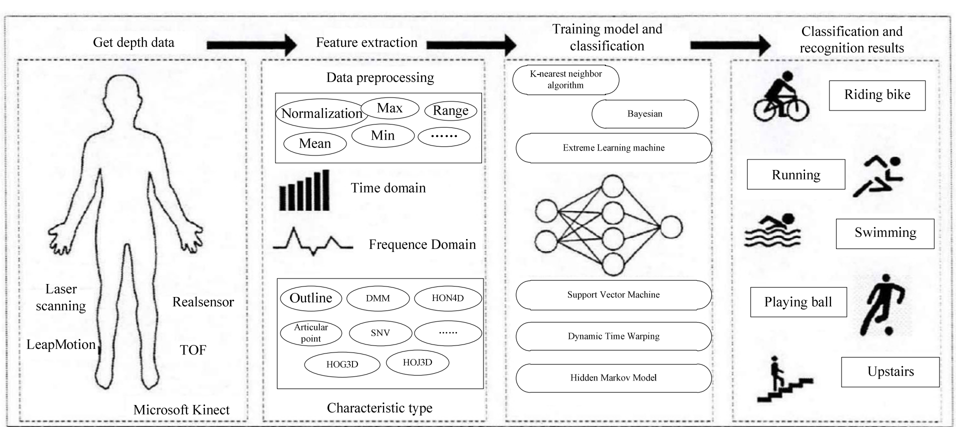
4 The experiment compares the correct recognition rate and time cost of different methods to test the application performance of different methods.

5 It summarises the full text and draws a conclusion.

2 The design of recognition method for athlete body's motion posture

The ultimate goal of human motion posture recognition is to identify the main body of the movement, that is, the athlete's movement behaviour from an image sequence or video data. The process of recognition can be divided into low-level data perception, data pre-processing, motion feature extraction, motion feature selection and classifier for specific classification. In the application of the improved deep neural network algorithm, the basic implementation process of the optimisation method for athlete body's motion posture recognition is shown in Figure 1.

Figure 1 Flow chart of athlete's posture recognition



According to the process of motion posture recognition shown in Figure 1, firstly, the required human motion information is extracted from image sequence or video data. After pre-processing, motion detection, feature extraction and neural network training, the obtained data can be classified into a certain class of classifier track to realise the classification of motion posture, so as to realise the recognition of motion posture and action.

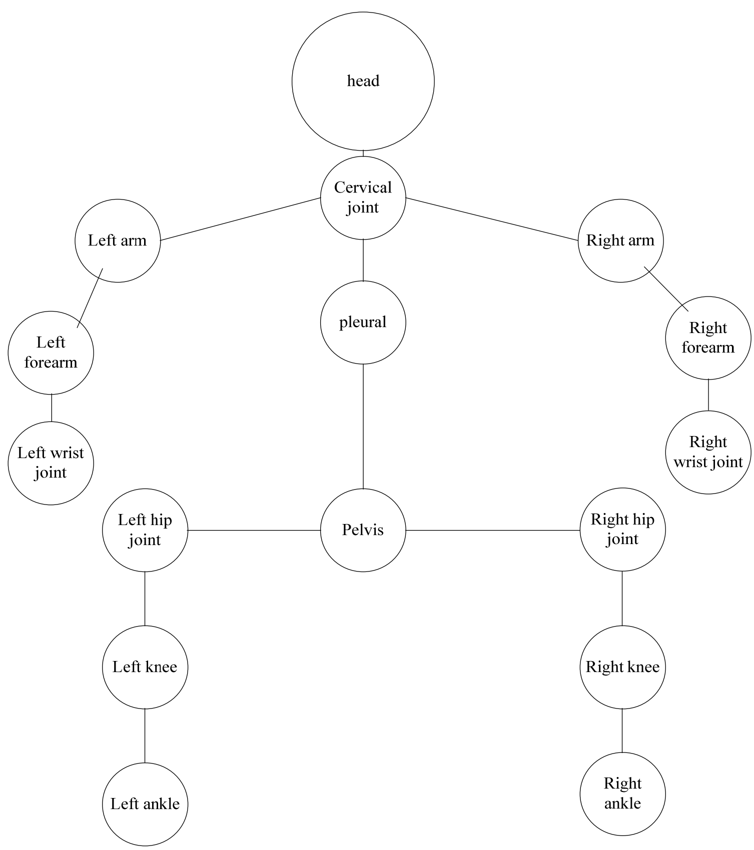
2.1 Setting the recognition standard of motion attitude

In order to ensure the accurate recognition of athletes' body posture, it is necessary to construct the human body database and set the corresponding recognition standards before starting the recognition task. Firstly, the mathematical model of human body movement of athletes is constructed. It can be expressed as:

$$m(t) = \left(p(t), q^1(t), \ldots q^i(t), \ldots, q^n(t) \right) \tag{1}$$

where $p(t)$ and $q^i(t)$ represent the position and orientation information of human head and joint centres respectively. The corresponding structure of the build model is shown in Figure 2.

Figure 2 Mathematical model of human body movement of athletes



In the mathematical model shown in Figure 2, the relationship between a point (X, Y, Z) in space and the projection point (u, v) on the image can be shown by equation (2).

$$\begin{pmatrix} u \\ z \end{pmatrix} = s \begin{pmatrix} 1\,0\,0 \\ 0\,1\,0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \tag{2}$$

In formula (2) s is the unknown size, and the calculation equation of s can be expressed as follows:

$$s = \frac{z}{f} \tag{3}$$

In the above formula, z represents the coordinate point of z axis in space, and f is the focal length of moving video or image capturing equipment. From this we can get the specific parameters in the mathematical model of human body movement. Different athletes are different in height and body shape, so the corresponding parameters in the mathematical model of human body movement are different. From Figure 2, we can see the corresponding position coordinates of each joint node. Combined with the limb characteristics of athletes, it can get the approximate range of motion of each joint point in different motion posture, and make it into the relevant database of human body movement of athletes, and take it as the standard of motion posture recognition.

2.2 Acquisition and processing of motion video data

In the method of athlete's posture recognition, real-time motion video data is captured by relevant hardware equipment, and image processing and analysis are carried out based on this. The high-speed camera with 500 frames per second is used to collect video data, which ensures that the difference between adjacent images is small and can reflect the corresponding motion information of each pixel. However, in the actual video capture process, the camera equipment cannot be used alone, so in addition to the high-frequency camera, we also need to install computer equipment, two video capture cards and a synchronous signal generator, in which the memory capacity of the computer needs to be more than 20 G. Two video capture cards are connected to the computer through PCI-E interface, then the high-frequency camera is connected with the acquisition card through the data line, and finally the acquisition card is connected to a synchronous signal generator. In the actual motion signal acquisition process, the synchronous signal generator drives the acquisition card to drive the camera to collect relevant data, and the real-time motion data acquisition results are written into the memory.

In order to obtain more accurate data information, before feature extraction, it is necessary to carry out a series of pre-processing operations on the initial captured motion video image data, so as to improve the image clarity. Generally, the pre-processing process of video sequence includes spatial domain filtering, binarisation, etc. For some motion joint and bone data, it is necessary to filter the initial image data to remove some video sequences with missing data.

From the video image sequence of athletes' body movement recorded by the camera, we can analyse the periodic swing of athletes' limbs when they walk, so as to determine whether there is movement behaviour in the video picture, solve the motion cycle and extract the corresponding key frames. In the captured video picture, W is set as the width of each frame, *H* as the length of the image, and the test area is select as the part from $\frac{3}{4}H$ to *H* below the image (Guanm, 2018). The video picture is divided into three sub regions horizontally, and the corresponding contour points in each area are respectively recorded as *numl* (*i*), *numc* (*i*) and *numr* (*i*). Then, the following equation can be used to calculate the athlete's body motion signal in the video picture.

$$X(i) = \frac{numl(i) + numr(i) - numc(i)}{\max\left(numl(i) + numr(i), numc(i) + \left|numl(i) - numr(i)\right|\right)} \tag{4}$$

A video image sequence is selected to test, then it can get the function image corresponding to equation (4). From the curve fluctuation in the image, we can get the athlete's exercise cycle.

Then, environment modelling and background subtraction are carried out for the key frames with athletes' motion information. Environment modelling is the basis of athlete body's motion analysis based on video sequence. The camera used is edited to establish background model, which lays a good foundation for motion segmentation (Zhang et al., 2020). Suppose that the extracted video sequence consists of n frames, and the pixel value of each frame at position (*x*, *y*) is marked as P_j (*x*, *y*), where *j* represents the number of frames. Then, the pixel value *B*(*x*, *y*) of the video sequence at position (*x*, *y*) can be expressed as:

$$B(x, y) = \underset{j=1,2,\ldots,n}{Median}\left(P_j(x, y)\right) \tag{5}$$

Formula (5) represents the modelling result of athletes' motion background. Combining with the background modelling results to segment the video image, the background image in the video can be subtracted. The foreground image can be expressed as follows:

$$d_i(x, y) = \left|f_i(x, y) - B(x, y)\right| \tag{6}$$

In the formula, $f_i(x, y)$ represents the previous frame image in the video. By substituting the calculation result of formula (5) into formula (6), the extraction result of moving object part can be obtained.

Because the athletes are always in motion, the camera needs to follow the movement of the athletes in the process of video image capture. In the process of the camera moving, there may be no focus, resulting in blurred motion picture. Therefore, it is necessary to compensate the motion residual and blur in video (Kim et al., 2019). Combined with the athlete's speed, it can be divided into whole motion with constant speed and accelerated whole motion, and motion compensation has only one limited unified motion vector for each block. Therefore, under the condition of motion compensation, the whole motion at constant speed can be expressed as follows:

$$F^{CBMC}(u, v, k) = F\left(u + u_{mv}, v + v_{mv}, k - 1\right) \tag{7}$$

where *u*, *v* and *k* represent the coordinate components in the *s*, *y* and *z* directions respectively, while *k* and v_{mv} correspond to the horizontal and vertical components of the motion vector obtained by block matching (Mengu et al., 2019), and *F* represents the motion compensation coefficient. According to the criterion of block matching, the mismatch of motion compensation is derived as follows.

$$SAD(CBMC) = \sum_{u,v\in D} \left| F(u, v, k) - F^{CBMC}(u, v, k) \right| \tag{8}$$

where D is the size of the compensation block. Formula (8) calculates the amount of compensation needed for motion. If the direction of motion compensation is determined, the fussy compensation and sharpening can be realised.

The purpose of image filtering is to reduce the noise points in the image, so as to improve the proportion of effective information in the image. According to the spatial probability density distribution characteristics of noise, noise can be divided into Gaussian noise, impulse noise, exponential distribution noise and other types. The main types of noise in the process of video image acquisition and processing are impulse noise and Gaussian noise (Kong, 2019). The source of Gaussian noise is electronic circuit noise and hardware equipment noise caused by low illumination or high temperature. The probability density function corresponding to Gaussian noise can be expressed as follows:

$$p(v) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(v-\mu)^2}{2\sigma^2}} \tag{9}$$

In, the formula, v represents the grey value of the video image after gray processing, and μ and σ represent the average value and standard deviation of v respectively. Similarly, the function expression corresponding to impulse noise can be obtained. Specific noise reduction process:

$$\hat{f}\left(i, j, k\right) = \sum_{i=-N}^{M} w(l) g(i, j, l-1) \tag{10}$$

where $g(i, j, k–1)$ is the grey value of the image of the k–1^th^ frame in the athlete video sequence at the position of pixel (i, j), and $w(l)$ is the weight of each frame in the video (Wu, 2018). The parameters M and N are the upper and lower bounds of the time domain window, respectively. In addition, Gaussian noise in video sequence images is processed by smooth linear filtering method. The results of denoising are as follows:

$$\hat{f}(i, j) = \frac{1}{m * n} \sum_{(x,y)\in N} f(x, y) \tag{11}$$

$f(x, y)$ is the athlete body’s motion video image before noise reduction, and m and n are the length and width of window template processed by smooth filtering.

Image normalisation is divided into two steps. One is to realise image normalisation by using linear function. After normalisation, the original image is converted into standard form to prevent the influence of affine change and geometric change, so as to accelerate the convergence speed of improved depth neural network (Zhao and Chen, 2020). In the process of normalisation, the data is converted to the range of [0, 1], and the original data is scaled equivalently. The normalised data can be expressed as follows:

$$norm = \frac{x - Min}{Max - Min} \tag{12}$$

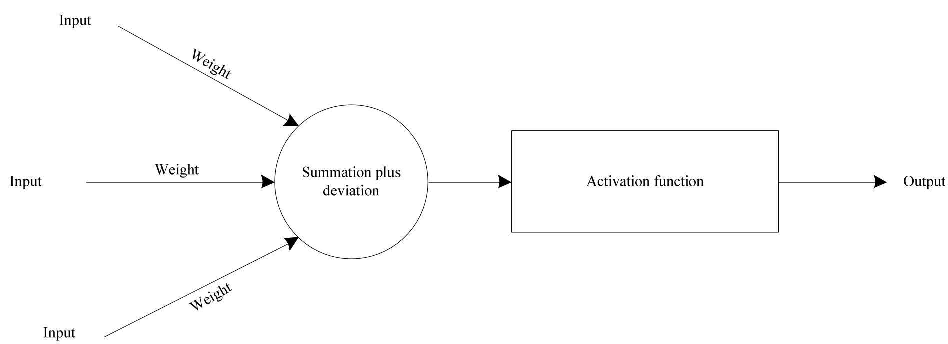
In the formula, x represents the pixel of the original image, and Max and Min represent the maximum and minimum of the motion video image pixels.

2.3 Improved motion gesture recognition under deep neural network

According to the above feature extraction results, by changing the connection form, the depth neural network is optimised and improved to improve the extraction ability of local features, effectively prevent over fitting phenomenon, and effectively improve the recognition accuracy while reducing the time cost.

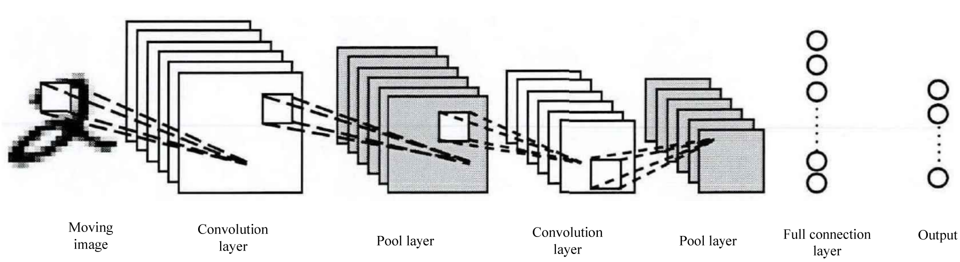
An improved deep neural network is a network composed of some form of connection, and the basic unit of the network is neuron, as shown in Figure 3.

Figure 3 Neuron model



The deep neural network is a multi-level back propagation artificial neural network. The basic structure of deep neural network includes input layer, hidden layer and output layer. The basic structure of improved deep neural network is shown in Figure 4.

Figure 4 Structure of improved depth neural network



Before using the improved depth neural network to construct the recognition classifier, it is necessary to extract the pose features, motion features and offset features of each video frame, and then get the extraction results of limb angle features after dimension reduction (Bulbul et al., 2019). The extraction and calculation process of each feature component can be expressed as follows:

$$\begin{cases} f_{cc} = \left\{ x_i - x_j \left| i, j = 1, 2, \cdots, N; i \neq j \right. \right\} \\ f_{cp} = \left\{ x_i^c - x_j^p \left| x_i^c \in X_c; x_j^p \in X_p \right. \right\} \\ f_{ci} = \left\{ x_i^c - x_j^i \left| x_i^c \in X_c; x_j^i \in X_i \right. \right\} \end{cases} \tag{13}$$

The contour features of human body can reflect the basic types of human motion of athletes. The extraction of contour features can be divided into two steps: contour point extraction and contour vectorisation. Based on the pre-processed moving video image, each element in the video image is the position of the point on the closed curve by default. Then, by calling the opencv function, it can find the contour from the binary moving image and extract some contour nodes (Gao et al., 2020). Starting from any pixel in the extracted contour node, the boundary pixels are connected along one direction, and the direction is recorded with code until it returns to the starting point. To repeat the above operation, the athlete body's motion video image will be fitted into multiple lines, and the endpoint of the line is the extracted contour feature inflection point.

The coordinate system is established on the video image, and the centroid position of the athlete body's motion posture is calculated. The calculation formula is.

$$\begin{cases} X = \dfrac{\sum U(x, y) \times x}{\sum U(x, y)} \\ Y = \dfrac{\sum U(x, y) \times y}{\sum U(x, y)} \end{cases} \tag{14}$$

where $U(x, y)$ is the pixel value corresponding to any point (x, y) in the video image. The region is segmented in the established coordinate system, and the real-time changes of contour feature extraction results are tracked. Taking the divided area of image region as the research object, the extraction result of area feature is obtained through accumulation calculation (Zhenchuan et al., 2018).

Because different types of features can reflect different amount of information of human motion posture (Elza et al., 2019), in the process of fusion, it is necessary to assign different weights for different features, and then conduct fusion processing. The calculation process of characteristic weight value is as follows:

$$\frac{\left(\dfrac{1}{\sum_{k=1}^{N} \left(\dfrac{1}{e_k} \right)} \right)}{e_k} \tag{15}$$

in the formula, k represents the number of feature categories, and N represents the number of features extracted from the video of athletes' body motion posture. When R^* is defined as the quantised value of feature matching after normalisation, the weighted fusion process of features can be expressed as follows:

$$g_i = \sum_{k=1}^{N} \omega_k R^* \tag{16}$$

where ω_k is the characteristic parameter. Through the feature fusion processing, the extraction results of the athletes' body motion posture features are output in the form of vector (Liu et al, 2019). In order to realise the accurate recognition of athletes' body posture types, an improved depth neural network is used to construct a classifier. The input of the classifier is the feature vector completed by fusion processing, and the output result is the motion type (Cho et al., 2019).

The activation function of the neural network is hyperbolic tangent and the tanh function is as follows:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{17}$$

Assuming, that the input layer of the improved depth neural network is a matrix M of $m \times m$, the a convolution kernel is a matrix N of $n \times n$, and the bias is b_1, then the generated convolution characteristic matrix F can be expressed as follows:

$$F_{ij} = S\left(\sum_{i=1}^{n}\sum_{i=1}^{n}\left(M_{ij}C_{ij}\right) + b_1\right) \tag{18}$$

The neurons in convolution layer are connected to the local region of the upper feature surface by a set of weights, and then the local weights are transferred to a nonlinear function to obtain the output. In the same way, the number of neurons and training parameters of convolution layer, and the basic operation principle of pooling layer, full connection layer and output layer can be obtained from the structure diagram of improved deep neural network (Shao et al., 2020). On the basis of the above principles, a feature classifier of human body movement is constructed, and the output of feature classification is obtained through network training. The data size of network output layer is shown in Table 1.

Table 1 Data size of output layer of improved deep neural network

Network layer name	*Input size*	*Number of filters*	*Nuclear size*	*Output size*
Data	3*16*240*320			3*16*112*112
Conv1a+relu1a	64*16*112*112	64	3*3*3	64*16*112*112
Pool1	64*16*112*112		2*2*1	64*16*56*56
Conv2a+relu2a	64*16*56*56	128	3*3*3	128*16*56*56
Pool5	512*2*7*7		2*2*2	512*1*4*4
Fc6+relu6	512*1*4*4	4,096		4,096

The output of the improved depth neural network classifier is compared with the set recognition standard of motion attitude, and the similarity between them is calculated (Ullah et al., 2018). Setting the similarity threshold value of motion posture to η, if the calculated similarity is greater than the set η, the compared pose can be directly output as the recognition result. Otherwise, another standard feature of human motion posture needs to be selected for similarity calculation until the recognition result meeting the similarity requirements is obtained.

3 Experimental analysis of recognition effect test

The overall experimental scheme is designed as follows:

1 Experimental environment and parameters

The experimental data is divided into simulation data and actual shooting data, so the experimental environment is also divided into two parts. Among them, the simulation data is to use the relevant software to generate an athlete's animation of different movements in different environments. Virtual cameras are installed in the simulation environment and placed in different angles of the athletes to generate multiple 300 frames of animation. The resolution of the image is 240*320. In addition, we need to prepare a large number of simulation data, that is, the comparison standard data of experimental data, to provide data basis for the operation of the experiment. Three high-frequency cameras are used to capture the athletes' body movement from three directions. The results of the motion shooting are written with Avisynth 2.5 software and imported into the experimental environment. The related experimental data are shown in Table 2.

Table 2 Sample data

Dataset name	*Action*	*Number of topics*	*Frequency*	*Number of samples*	*Data type*
MSR Action 3D	20	10	2/3	567	D+S
MSR daily activity 3D	16	10	2	320	C+D+S
UT kineetaction	10	10	2	200	C+D+S
RGBD-HuDaAct	12	30	2/4	1,189	C+D
CAD-60	12	2+2	-	60	C+D+S
MSRC-12 kineet gesture	12	30	-	594	S
CAD-120	10	2+2	3	120	C+D+S
MSR action pairs	6	10	3	180	C+D+S
G3D	20	10	¾	659	C+D+S
Osaka	10	8	1	60	C+D+S
Workoutsu-10 gesture	10	12	10	1,200	C+D+S
Morning routine	7	1	8	56	S

2 Experimental methods

Because the improved deep neural network is applied in the designed method of human body's motion posture recognition, it is necessary to provide the operation environment for the training of deep neural network based on the basic experimental environment. When the recognition samples are trained, the size of the data will change according to the characteristics of the neural network layer in different depths. In order to form experimental comparison, in addition to the designed recognition method, the traditional method of athlete body's motion posture recognition based on multi feature fusion in Liu (2019) and the athlete body's motion posture recognition method based on MEMS sensor in Li et al. (2018) are set as experimental comparison methods.

3 Evaluation index

In the experiment, the evaluation index of recognition effect is the recognition efficiency, that is, the number of correct recognition posture of athletes in unit time, or the recognition time consumed by different recognition methods for fixed recognition tasks. It can be seen that the efficiency index of this experiment consists of two parts: the correct rate of recognition and the time cost of recognition operation. The correct rate of recognition is the ratio between the correct recognition samples and the total samples, and the time cost data can be obtained by calling the background running data in the host computer.

The initial recognition samples need to be reset before the recognition method is run, and the final recognition result is obtained through the operation of the recognition method.

3.1 Comparison of correct recognition rate

The recognition results of the three methods are compared with the results of the athlete body's motion posture, and the corresponding statistical results of correct recognition rate are obtained, as shown in Table 3.

Table 3 Comparison results of correct recognition rate

Type of motion posture	*Number of samples/piece*	*Method in Liu (2019)/%*	*Method in Li et al. (2018)/%*	*The proposed method/%*
Walking	224	215	218	222
Running	305	297	301	303
Rope skipping	117	103	111	114
Play basketball	256	246	252	254
Play badminton	311	299	302	308
Play table tennis	278	268	271	274
Swimming	98	88	91	95
Aerobics	125	115	119	122
Pull up	110	100	102	105
Abdominal curl	252	244	248	250
Play football	312	304	307	311

It can be seen from Table 3 that the correct recognition rate of traditional identification method and artificial neural network identification method is 95.4% and 97.2% respectively, while the average correct recognition rate of design method is 98.2%. Because the method in this paper collects the athlete's body image and carries out a series of pre-processing, the recognition result is more accurate.

3.2 Comparison of recognition time cost

The input time of each sample and the output time of recognition results are respectively called, and the running time cost of recognition method is obtained by subtracting. The results are shown in Table 4.

Table 4 Comparison of recognition time cost

Type of motion posture	*Number of samples*	*Method in Liu (2019)/s*	*Method in Li et al. (2018)/s*	*The proposed method/s*
Walking	224	12.5	10.2	1.2
Running	305	11.7	9.5	1.8
Rope skipping	117	16.5	9.6	1.3
Play basketball	256	13.8	8.9	1.4
Play badminton	311	14.7	9.2	1.2
Play table tennis	278	15.9	8.7	1.4
Swimming	98	17.5	9.5	1.6
Aerobics	125	16.3	9.7	1.3
Pull up	110	12.8	8.6	1.2
Abdominal curl	252	11.6	7.2	1.6
Play football	312	13.5	7.6	1.4
Mean value	–	14.3	9.0	1.4

Through the calculation of the time cost, it is known that the average time cost of the two methods is 14.3 s and 9.0 s, respectively, while the average time cost of this method is 1.4 s. It can be seen that for the same experimental sample, the correct recognition rate of athletes' body movement posture recognition method based on improved depth neural network is higher, and the time is shorter, that is to say, the recognition efficiency is higher. The reason for the above experimental results is that the proposed method improves the depth neural network and constructs a recognition classifier. The pre-processed athlete's body image is input into the classifier to finish the classification accurately.

4 Conclusions

1 In the background of the explosive growth of video image data, the rapid improvement of computer computing ability has also promoted the development and innovation of improved deep neural network in the application fields of natural language processing and computer vision.

2 In this paper, we design a recognition method of athlete body's motion posture based on the improved deep neural network. The recognition classifier is constructed based on the improved deep neural network. The fused feature vectors are input into the classifier and the recognition results are output.

3 Experimental results show that the proposed method has high recognition rate and short time cost.

3 By improving the application of deep neural network, the problem of low accuracy in the recognition of athlete body's motion posture is effectively solved, which is of great significance for many athletes' competitive events and daily training.

References

Bulbul, M.F., Islam, S. and Ali, H. (2019) '3D human action analysis and recognition through GLAC descriptor on 2D motion and static posture images', *Multimedia Tools and Applications*, Vol. 78, No. 15, pp.21085–21111.

Cheng, S. and Zhou, G. (2020) 'A method of human motion posture recognition based on joint points', *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 34, No. 7, pp.265–283.

Cho, S.G., Yoshikawa, M., Ding, M., Takamatsu, J. and Ogasawara, T. (2019) 'Machine-learning-based hand motion recognition system by measuring forearm deformation with a distance sensor array', *International Journal of Intelligent Robotics and Applications*, Vo. 3, No. 4, pp.418–429.

Elza, F.D.G.A., Chaves, T.D.M., Fallavollita, P., Figueiredo, L.S. and Teichrieb, V. (2019) 'Rehabilitation motion recognition based on the international biomechanical standards', *Expert Systems with Application*, Vol. 116, No. 5, pp.396–409.

Fei, G., Teng, H., Jinping, S. and Wang, J. (2018) 'A new algorithm of SAR image target recognition based on improved deep convolutional neural network', *Cognitive Computation*, Vol. 12, No. 2, pp.1–16.

Gao, L., Zhang, G., Yu, B., Qiao, Z. and Wang, J. (2020) 'Wearable human motion posture capture and medical health monitoring based on wireless sensor networks', *Measurement*, Vol. 166, No. 2, pp.108–252.

Guanm, W. (2018) 'Performance optimization of speech recognition system with deep neural network model', *Optical Memory and Neural Networks*, Vol. 27, No. 4, pp.272–282.

Kim, H., Lee, W.S., Yoo, J., Park, M. and Kang, H. (2019) 'Origin of the higher difficulty in the recognition of vowels compared to handwritten digits in deep neural networks', *Journal – Korean Physical Society*, Vol. 74, No. 1, pp.12–18.

Kong, F. (2019) 'Facial expression recognition method based on deep convolutional neural network combined with improved LBP features', *Personal and Ubiquitous Computing*, Vol. 23, Nos. 3–4, pp.1–9.

Li, Y.L., Shi, Z.Q., Li, S.H., Chen, F., Wang, R. and Ding, H. (2018) 'Human motion recognition method based on MEMS sensor', *Computer Applications and Software*, Vol. 35, No. 8, pp.243–248+285.

Liu, S. (2019) 'Athlete's posture recognition algorithm based on multi feature fusion', *Information Technology*, Vol. 25, No. 8, pp.17–19+23.

Liu, T., Wang, J., Hutchinson, S. and Q.H. (2019) Skeleton-based human action recognition by pose specificity and weighted voting', *International Journal of Social Robotics*, Vol. 11, No. 2, pp.219–234.

Mengu, D., Luo, Y., Rivenson, Y. and Ozcan, A. (2019) 'Analysis of diffractive optical neural networks and their integration with electronic neural networks', *IEEE Journal of Selected Topics in Quantum Electronics*, Vol. 26, No. 1, pp.1–14.

Shao, J., Niu, Y., Xue, C. and Wu, Q. (2020) 'Single-channel SEMG using wavelet deep belief networks for upper limb motion recognition', *International Journal of Industrial Ergonomics*, Vol. 76, No. 1, pp.102–109.

Ullah, A., Ahmad, J., Muhammad, K., Sajjad, M. and Baik, S.W. (2018) 'Action recognition in video sequences using deep bi-directional LSTM with CNN features', *IEEE Access*, Vol. 6, No. 99, 1155–1166.

Wu, D. (2018) 'Online position recognition and correction method for sports athletes', *Cognitive Systems Research*, Vol. 52, No. 16, pp. 174–181.

Zhang, A., He, J., Lin, Y., Li, Q. and Qu, G. (2020) 'Recognition of partial discharge of cable accessories based on convolutional neural network with small data set', *COMPEL-The International Journal for Computation and Mathematics in Electrical and Electronic Engineering*, Vol. 39, No. 2, pp.431–446.

Zhao, L. and Chen, W. (2020) 'Detection and recognition of human body posture in motion based on sensor technology', *IEEJ Transactions on Electrical and Electronic Engineering*, Vol. 15, No. 5, pp.766–770.

Zhenchuan, Z., Yingchun,, Y., Zhaohui, W. and Qian, M. (2018) 'A posture recognition system for rat cyborg automated navigation', *Chinese Journal of Electronics*, Vol. 27, No. 4, pp.25–31.

Zhu, X., Zhu, M. and Ren, H. (2018) 'Method of plant leaf recognition based on improved deep convolutional neural network', *Cognitive Systems Research*, Vol. 52, No. 2, pp.223–233.