The Journal of Supercomputing (2022) 78:10521–10541
https://doi.org/10.1007/s11227-021-04298-y

Analyzing body changes of high-level dance movements through biological image visualization technology by convolutional neural network

Ruizhi Zhang[1]

Accepted: 24 December 2021 / Published online: 24 January 2022

Abstract

A research is designed to comprehensively study the dynamic trends and characteristics of dance movements recognition approaches, providing a valuable reference for the scientific research and development of dance art in China. Based on previous studies, a dance movements recognition model is proposed based on DL biological image visualization technology against the current problems in high-level dance movements. This model consists of the convolutional neural network (CNN)-based dance movements recognition algorithm that can recognize body movements, the dance movements recognition algorithm involving videos in the database, and the algorithm that analyzes image feature similarities of the dance movements. The Open Pose algorithm is adopted for human posture recognition. A high-precision human body pose movement state is finally obtained to judge the body changes of high-level dance movements through the division of human bones. Students with excellent dancing skills are invited to participate in the experiment, for whom specific dance movements are designed. The movement characteristics of bodies' joint points are extracted and compared with standard movements demonstrated by teachers to obtain the differences between students' movements and standard movements. The results review that the proposed dance movements recognition model based on biological video images has high accuracy. Within 6–8 s, the swing amplitude of the subjects' left arm is quite different from the standard dance movements. The movements of the left arms, right arms, left legs, and right legs of experimental objects are quite different from the standard dance movements, which proves the model's effectiveness. The results can provide a valuable reference for the research and development of dance in China and a practical basis for teaching dance movements.

✉ Ruizhi Zhang
13622746@qq.com

Extended author information available on the last page of the article

Keywords Deep learning · Convolution neural network · High-level channel movement recognition · Biological image visualization · Human body gesture recognition

1 Introduction

Dance is a standardized and procedural sporting event completed by a single person or more persons, within a defined range of music and rhythm, where performers need to correctly display poses and use body techniques and skills, combined with artistic expression [1, 2]. As dance event is promoted globally by international dance organizations, the trend of dance globalization has gradually increased. Furthermore, the dance market has transformed its attention from training to competitions, whose international influence has gradually expanded. Dance has attracted many organizations' attention with its unique artistic and competitive qualities [3]. As significant varieties of shows are promoted, dance activities in China have been rapidly improved. With China's culture healthily developing towards the world, the Chinese traditional dance has witnessed a new era, showing a unique style in the global dance culture exchanges [4]. Currently, there are more than 60 colleges and universities providing dance courses in China. Moreover, dance enthusiasts in China have exceeded 30 million, and dance sports have also played an active role in national fitness [5]. However, the quality and strength of research on dance in China lag far behind that of developed countries. Nowadays, human movement recognition mainly refers to recognizing basic movements in daily life. Whereas few researchers have analyzed human postures in the field of high-level dances [6]. High-level dance movements are usually complicated, with a wide variety of movements and particular complexity in expression. Therefore, recognizing human postures in dance movements still focuses on studying individual movements [7].

With the advent of the big data era, methods for human posture recognition include DL and feature-based movement pose analysis [8]. At present, there have been many reports in this field. Kamala and Mary (2015) used distributed cloud computing to build a multi-platform human body pose recognition system, which could effectively process human posture data [9]. Yan et al. (2016) proposed a new system that utilized CNN for automatically learning and predicting the predefined driving poses; the central idea was detecting the position of the drivers' hands by extracting recognition information, thereby predicting the correct/wrong driving poses; the overall accuracy of the proposed system reached 99.78% [10]. Nath et al. (2017) designed a detection algorithm based on the human body poses on a bicycle, which effectively analyzed the characteristics of human force when participants of the experiment were riding a bicycle. It provided a theoretical reference for correct riding [11].

Gladence et al. (2017) proposed a new movement recognition algorithm based on human posture, which could effectively solve human feature detection's fundamental problems with high recognition accuracy [12]. Varol et al. (2018) utilized long-term temporal convolutions (LTC) for analyzing human movement videos; they proved that increasing the time range of the LTC model could improve the accuracy

of movement recognition [13]. Lv (2020) designed a human video detection algorithm based on big data according to the characteristics of the Internet of Things (IoT). This method could upload data in real-time, effectively solving the problems encountered in the network transmission of human posture recognition [14]. Apparently, most of the current algorithms for human posture estimation utilize video data for training, modelling, and analysis. However, few scholars have applied human posture estimation to analyze and research dance poses.

Therefore, DL technology and biological image visualization are the foundations. A CNN-based dance movements recognition algorithm is proposed to solve the difficulty in recognizing body changes in dance movements. In the second section, a dance action recognition algorithm is proposed based on CNN, which solves the difficulty of body change recognition in dance action. Moreover, based on the original CNN, the optical flow information is used to characterize the state change of time-domain motions, and the bidirectional convolution network is constructed to improve the recognition accuracy of dance movements. The video-oriented dance movements recognition method is used to solve the problem of no data input in the modelling process. Finally, with the image feature similarity evaluation of human posture evaluation, a comprehensive analysis model of high-level dance movements is established. In the third section, experiments are conducted on the created data set to verify the effectiveness of the proposed method. The purposes are to fully grasp the dynamic trends and characteristics of dance research and development, provide a valuable reference for the scientific research and development of dance in China, and offer theoretical guidance for dance scientific fitness and improvement of athletic level, thereby consolidating the foundation of dance discipline and promoting the scientific development of dance movements. Based on the features of the human skeleton and the efficient recognition algorithm proposed here, the accuracy of symbols of dance notation generated by the system has been further improved, which is more in line with practical application. It is of great significance for the recording of the activities of cultural heritage such as professional correction of dance movements and national dance.

2 Method

2.1 DL and CNN

The deep neural learning network is currently a popular machine learning algorithm. It combines all individual units through weights and uses datasets for model training and learning. The most common DL algorithm is CNN, which simulates the human brain inputting signals through the nervous system, processes the data, and generates human posture by the central nervous system. It has strong learning ability and adaptability, which is generally applied in image recognition, behaviour prediction, and model control [15]. The primary purpose is to identify the human posture of dance through images. Therefore, the general processes are: (1) inputting the video data into the deep CNN, (2) using the forward propagation method and the

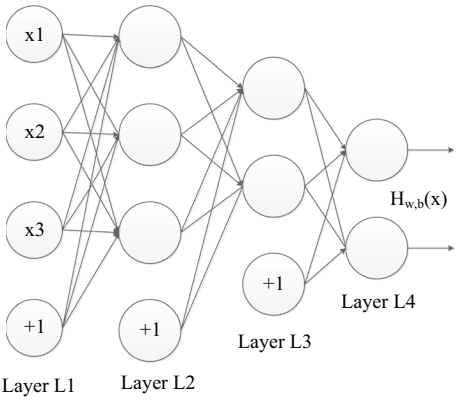

Fig. 1 Structure of a forward neural network

backpropagation method to optimize the objective function, (3) continuously changing the weight by inputting data, and (4) outputting the data when all the weights reach the expected goals [16, 17]. Figure 1 presents the structure of the forward neural network.

The neuron consists of three inputs (x_1, x_2, x_3) and one output $h_{w,b}(x)$. Equation (1) expresses the calculation of the output $h_{w,b}(x)$.

$$h_{w,b}(x) = f\left(W^T x\right) = f\left(\sum_{i=1}^{n} w_i x_i + b\right) \tag{1}$$

In Eq. (1), x represents the input vector, W denotes the input weight vector, b indicates the bias, and f represents the activation function. This method realizes the learning function by continuously adjusting the weight and bias parameters of neurons. Due to the local perceptual region's characteristics and weight sharing, CNN can simulate the human eyes' visual mechanism and directly use the image as the network input.

2.2 CNN-based recognition algorithm for dance movements

Human 3D posture perception is also known as human motion capture. According to different motion capture devices, it can be roughly divided into optical unmarked movement capture device, optical marked movement capture device, mechanical movement capture device, magnetic sensor movement capture device, and inertial sensor movement capture device. The research focuses on the human posture perception of unmarked monocular cameras. With a focus on the pose perception based on the parametric human model, the movement capture results obtained by the ipiSoft series software with moderate cost performance are selected as the reference data to verify the effectiveness of the proposed 3D pose restoration method. The dance movements recognition algorithm based on CNN generates feature images through weights and biases. This method can obtain movement information of multiple consecutive frames. Equation (2) illustrates the detailed calculation process.

$$v_{ij}^{xyz} = \tanh\left(b_{i,j} + \sum_{m}\sum_{p=0}^{P_i-1}\sum_{q=0}^{Q_i-1}\sum_{r=0}^{R_i-1} w_{ijm}^{pqr} v_{(i-1)m}^{(x+p)(y+q)(z+r)}\right) \tag{2}$$

In Eq. (2), *tanh ()* represents the hyperbolic tangent function, b_{ij} refers to the bias of the feature map, R_i stands for the size of the 3D convolution kernel in the time dimension, and w_{ijm}^{pqr} accords to the value of the convolution kernel connected to the m_{th} feature map in the previous layer at the point *(p, q, r)*. When 3D CNN is used to process videos, the dimension of the down-sampled area is usually extended to 3D, which reduces the calculation amount of subsequent operations and enhances the network's invariance. Equation (3) illustrates the maximum down-sampling of the 3D overlapped in the $S_1 \times S_2 \times S_3$ region is:

$$y_{mnl} = \max_{0\le i\le S_1, 0\le j\le S_2, 0\le k\le S_3} \left(x_{m\times s+i, n\times t+j, l\times r+k}\right) \tag{3}$$

In Eq. (3), $x_{m\times s+i,n\times t+j,l\times r+k}$ represents the value of the 3D input at point $(m\times s+i, n\times t+j, l\times r+k)$, y_{mnl} denotes the output after 3D sampling at the point *(m, m, l)*, and s, t, and r are the sliding steps in three directions, respectively. At the beginning, the grayscale image is generated into an H1 layer composed of five feature channels and 33 feature maps. Then, a $7\times7\times3$ convolution kernel performs the 3D convolution operations on each of the five channels. The network uses two different convolution kernels to perform three convolutions in the time dimension. Thus, there are 23 feature maps in each group. A down-sampling layer S5 follows the convolutional layer. Each feature map in the C4 layer is 3×3 spatially down-sampled. The fully connected layer C6 uses a 7×4 convolution kernel to perform the convolution operation only in the space dimension. Eventually, the number of units in the output layer accords to that of behaviour categories. Meanwhile, each unit is connected to the 128 1×1 feature map in the previous layer. All trainable parameters in the network are initialized randomly. The activation function uses the hyperbolic tangent function *tanh ()* [18]. For a sample set with m samples, h_i denotes the output corresponding to the input x_i, and the half-variance cost function for a single sample *(x, y)* is shown in Eq. (4).

$$J(W, b; x, y) = \frac{1}{2}\left\|h_{W,b}(x) - y\right\|_2^2 \tag{4}$$

In Eq. (4), *(W, b)* is the parameters to be adjusted in CNN. The cost function for all m samples is:

$$J(W, b) = \frac{1}{m}\sum_{i=1}^{m}\left(\frac{1}{2}\left\|h_{W,b}(x_u) - y_i\right\|_2^2\right) + \frac{\lambda}{2}\sum_{l=1}^{n_l-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_l+1}(W_{ji}^{l})^2 \tag{5}$$

In Eq. (5), the first term represents the mean square error (MSE), and the second term represents the regularization term, which aims to adjust the weight, thereby avoiding overfitting.

If the sigmoid function is employed as the activation function of the neurons in the model, the cross-entropy cost function is utilized for training the weight of the model and expressed as Eq. (6).

$$J(W, b; x, y) = -\frac{1}{m}\sum_{i=1}^{m}[y_i \ln h_i + (1 - y_i)\ln(1 - h_i)] \tag{6}$$

The cross-entropy cost function has non-negative outputs. If the final output is close to the expected output, the output value will approximate 0. The backpropagation algorithm is applied to determine the gradient of weight adjustment. According to the output equation of the neuron, Eq. (7) signifies the relationship between the input of the neuron and the gradient.

$$\delta^{l-1} = \delta^l \frac{\partial z^l}{\partial z^{l-1}} = (W^l)^T \delta^l \odot f'(z^{l-1}) \tag{7}$$

In Eq. (7), z^l represents the neuron's input, and δ^l points to the layer's parameters' gradient.

After the gradient between CNN layers is determined, the parameters are updated by the gradient descent algorithm [19]. Primarily, the gradient descent algorithm initializes the parameters. Then, the gradient value is calculated and judged. Finally, the previous steps are continuously repeated until the function's minimum value is obtained [20, 21]. The trained CNN is applied for direct feature extraction. In image retrieval based on CNN, the fully connected feature map is usually utilized as the feature vector for retrieval [22, 23].

2.3 Dual-channel CNN based on time–space

The feature extraction method in the space domain is consistent with the extraction method of image information, and the features in time dimension are identified by optical flow. Figure 2 displays the recurrent neural network (RNN). Optical flow reflects the change trajectory of pixels after the movement state of objects in the space is changed, which is widely used in motion detection. The acquisition method is as follows:

$I(x, y, t)$ represents the pixel brightness of the coordinate position point $O(x, y)$ in time t. $(x + dx, y + dy)$ refers to the movement of the point to the new position during time dt. At this point, due to the very short time, the brightness of the point has the following relationship:

$$I(x, y, t) = I(x + dx, y + dy, t + dt) \tag{8}$$

It has a Taylor expansion as:

$$\frac{\partial I}{\partial x}\frac{\Delta x}{Vt} + \frac{\partial I}{\partial y}\frac{\Delta y}{Vt} + \frac{\partial I}{\partial x}\frac{\Delta t}{Vt} = 0 \tag{9}$$

At this time, Eq. (10) refers to the optical flow of the point.

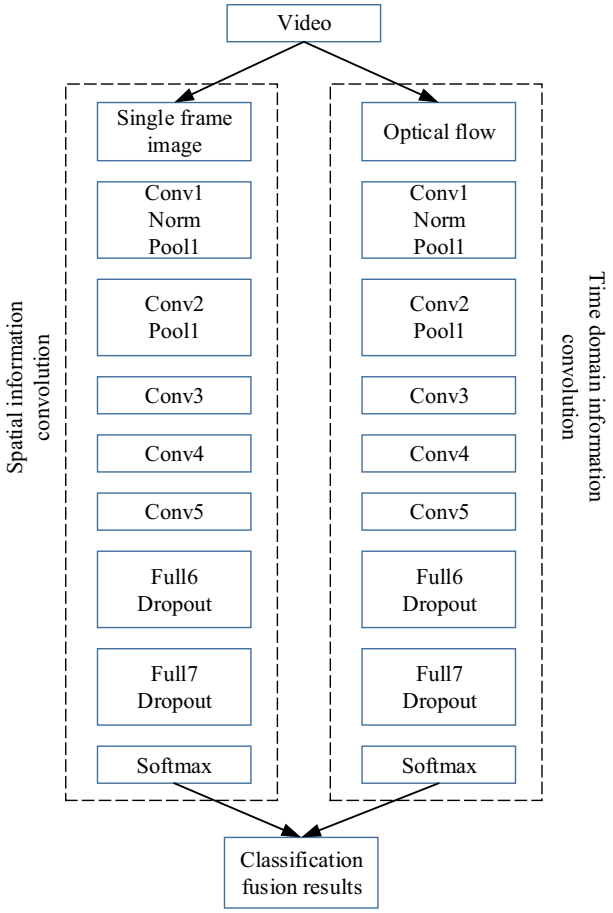

Fig. 2 RNN

$$\frac{\partial I}{\partial x}V_x + \frac{\partial I}{\partial y}V_y + \frac{\partial I}{\partial x} = 0 \tag{10}$$

In Eq. (10), V_x and V_y mean optical flow vectors. The pyramid algorithm is introduced to solve the optical flow vectors of the differential equation.

For 3×3 pixels' region, there are 9 optical flow trajectories, expressed as Eq. (11):

$$Av = b \tag{11}$$

The variables are calculated as Eqs. (12) and (13).

$$A = \begin{bmatrix} I_x(q_1)I_y(q_1) \\ I_x(q_2)I_y(q_2) \\ \cdots \\ I_x(q_9)I_y(q_9) \end{bmatrix}, v = \begin{bmatrix} V_x \\ V_y \end{bmatrix} \tag{12}$$

$$b = \begin{bmatrix} -I_t(q_1) \\ -I_t(q_2) \\ \cdots \\ -I_t(q_9) \end{bmatrix} \tag{13}$$

Through Eqs. (12) and (13), solvents can be obtained.

$$\begin{aligned} A^T Av &= A^T b \\ v &= \left(A^T A\right)^{-1} A^T b \end{aligned} \tag{14}$$

2.4 Video-oriented dance movements recognition

Eight parameters, expressed as $(u, v, \gamma, h, \dot{x}, \dot{y}, \dot{\gamma}, \dot{h})$, are employed to describe the movement state, where *(u, v)* is the centre coordinate of the bounding box, r is the aspect ratio, and h is the height. The remaining four variables represent the corresponding speed information in the image coordinate system. A standard Kalman filter based on a constant velocity model and a linear observation model is utilized for predicting the movement state of the target, and the prediction result is *(u, v, r, h)*. The Mahalanobis distance between the Kalman prediction result and the current moving target's movement state's detection result is adopted to associate the running information, expressed as Eq. (15).

$$d^{(1)}(i,j) = (d_j - y_i)^T S_i^{-1} (d_j - y_i) \tag{15}$$

In Eq. (15), d_j represents the j_{th} detection frame's position, y_i stands for the predicted position of the i_{th} tracker to the target, and S_i means the covariance matrix between the detection position and the average tracking position. The second measurement method calculates the minimum cosine distance between the latest 100 successfully associated feature sets of the i_{th} tracker and the feature vector of the current frame's j_{th} detection result, as Eq. (16).

$$d^{(2)}(i,j) = \min\{1 - r_j^T r_j^{(i)} | r_j^{(i)} \in R_i\} \tag{16}$$

The movement process of the human body is very complicated. Without considering the muscles and nervous system, the movement of the human body is abstracted into a simple chain system movement connected by parts of a rigid body.

The OpenPose algorithm is adopted for human posture recognition. This algorithm divides the human body into 18 bone key points: nose-0, neck-1, right shoulder-2, right elbow-3, right wrist-4, left shoulder-5, left elbow-6, left wrist-7, right hip-8, right knee-9, right ankle-10, left hip-11, left knee-12, left ankle-13, right eye-14, left eye-15, right ear-16, left ear-17, and back-18. This algorithm is suitable for estimating single and multiple persons with excellent robustness [24–26]. Open Pose is the world's first real-time multi-person 2D pose estimation application based on DL.

A video clip is input. The features are extracted through CNN to obtain a set of feature maps. Then, these feature maps are divided into two branches. CNN is adopted to extract part confidence maps and part affinity fields, respectively. After obtaining the above information, the bipartite matching in graph theory is utilized for finding the part association. The joints of the same person are connected. Finally, a person's overall frame is formed via merging the points as shown in Fig. 3.

Open Pose is a bottom-up algorithm to identify the human body, that is, to find the feature parts and then combine the human body. It inputs trichromatic images and outputs 2D position coordinates of each anatomical key point in the image, which can identify the key points of the human body. This algorithm is connected by two CNNs to predict the confidence and affinity vectors of each key point. The two main steps are as follows. (1) Output confidence map: the original image is processed by the first CNN to generate the confidence map set. CNN uses convolution kernel to perform convolution operation on the matrix window taken from the original image. Each confidence map contains an image feature. The confidence map output by the first CNN and the original map are introduced into the next network for calculation, where the confidence level represents the probability of the actual value falling in an interval. (2) Output skeleton diagram: the steps also include two CNNs, which connect the confidence diagram output by the first convolution network and repeat the steps of the first step.

The system processes the image through two branches of multi-stage CNN. The first branch is the prediction confidence map for each stage, and the second branch is the prediction partial affinity domain for each stage. The feedforward network

Fig. 3 The overall framework of the human body

infers a set of 2D confidence maps of human body parts and a set of 2D vector fields reflecting the local relational degree and encodes the correlation degree between joints. Finally, by inferring and analyzing the confidence mapping affinity domain, Open Pose outputs all two-dimensional key points in the image.

In this section, a dance action recognition algorithm is proposed based on CNN, which solves the difficulty of body change recognition in dance action. Moreover, based on the original CNN, the optical flow information is used to characterize the state change of time-domain motion, and a bidirectional convolution network is constructed to improve the recognition accuracy of dance moves. Video-oriented dance action recognition is used to solve the problem of no data input in the modelling process. Finally, with the similarity evaluation method of image features in human posture evaluation, what is established is a comprehensive analysis model of human posture changes in high-level dance movements.

2.5 Evaluation criteria for similarity of image features

Calculating the similarity of image feature affects the accuracy of subsequent image retrieval, and using different similarity measurement functions will result in entirely different image retrieval results. Traditional similarity measurement functions include: in this section, the image feature similarity evaluation is used in human posture evaluation to establish a comprehensive analysis model of human changes in high-level dance movements.

(1) Euclidean distance: It is the most common distance measure, which measures the absolute distance between points in multidimensional space [27]. In calculating the similarity scene, Euclidean distance is a similarity algorithm that is more intuitive and common. The smaller the Euclidean distance is, the greater the similarity is, and vice versa.

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \tag{17}$$

In Eq. (17), x_i and y_i, respectively, represent the feature vector of image x and image y, and n represents the dimension of the image feature vector.

(2) Cosine measurement: Cosine similarity uses the cosine value of the angle between two vectors in the vector space as a measure of the difference between the two individuals. The closer the cosine value is to 1, the closer the angle is to 0 degrees; that is, the two vectors are similar.

$$d(x, y) = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i} \sqrt{\sum_{i=1}^{n} y_i}} \tag{18}$$

The cosine measurement is utilized for calculating the similarity in the proposed model.

2.6 Simulation experiments

In the generation method of dance spectrum of lower limb action based on CNN proposed here, the network structure consists of two convolutional layers. The convolution layer is followed by a pooling layer, then followed by three fully connected layers. Among them, the activation function of the first two fully connected layers uses the ReLU function, and the activation function of the last fully connected layer is the softmax function, which plays the role as a classifier. In the experiment, to prevent overfitting, a dropout layer is connected behind each full connection layer. In the whole training process of CNN, dropout value is 0.25, learning rate is 0.01, the number of convolutions in the first layer is 32, the number of convolutions in the second layer is 64, the size of the convolution kernel is 2×2, and the number of nodes in the middle full connection layer is 128. The number of nodes in the last full connection layer is consistent with the number of action categories.

This simulation experiment uses MATLAB as the development environment. All experiments are based on the Visual Studio 2010 development platform and OpenCV programming environment and are executed on a computer equipped with Intel Core i7-4790 CPU and 16 GB RAM. The operating system is Windows 10. A movement database is created, which contains 18 sets of dance movements fragments. Each set of dance movements contains about 1200 frames. All the videos have a refreshing rate as 25fps, with a resolution of 480×360, and their lengths of time are between 2.31 and 67.24 s. The experimental subjects are students with good dancing skills in school. The subjects are required to imitate the movements taught by the dance teacher. After the corresponding dance movements are made, the movement characteristics of the joint points of the subjects are extracted compared with the standard movements made by the teacher to observe the difference between subjects' movements and standard movements.

Each dance movements image needs to be preprocessed to meet the input requirements of the model. The contour of the human body is normalized to a size of 128×88. Most of the dance videos and images are in colour, which will greatly increase the computing power. Therefore, the 3D colour space can be converted into 1D via dimensionality reduction and image grayscale, which reduces the calculation amounts. Next, the data amount of the video images is compressed by image thresholding, and the contour of the moving image is extracted.

Table 1 manifests the modulated parameters and their value ranges

Modulated parameters	Data range
Convolution kernel size	3×3, 5×5, 7×7, 9×9, 11×11
The layer number of convolutions	2, 3, 4, 5
Batch size	5, 10, 20, 30

3 Results

3.1 Influence of fibre type on pore pressure development

Some network parameters need modulating to optimize the network model; that is, to obtain the highest recognition rate. The parameters to be modulated contain the size of the convolution kernel, the number of convolution layers, and the batch size. The method used is to fix the remaining variables and modulate the single variable until the recognition rate is optimal (Table 1).

The inadequate convolution layers will lead to insufficient expression of action feature information, while the redundant convolution layers will lead to that networks are too complex and prone to overfitting, thus reducing the accuracy of action recognition. The effect of different convolution kernel sizes on the recognition rate is obtained when the batch size is 10, and the network has 3 convolutional layers and down-sampling layers. At this time, a convolution kernel with a size of 5×5 can obtain the optimal accuracy rate. Also, since the larger convolution kernel indicates more massive calculation amounts and lower efficiency, 5×5 is taken as the optimal size of the convolution kernel. Then, the optimal size convolution kernel is set to 5×5, and the batch size is set to 10. Thus, the number of convolution layers is modulated. The impact of the batch size value on the recognition rate cannot be ignored. When its value is small, the randomness of learning is large, and it is difficult for the network to achieve convergence. When its value increases, this part of the sample can better represent the overall samples; thus, the direction of the extreme value can be more accurate. However, the excessively large value not only requires strong hardware support but also slows down the update speed of the weights. Therefore, the number sees an according to increase in iterations required to achieve the same accuracy. Thus, this parameter needs to be weighed and calculated based on factors such as the size of the dataset in the actual applications and the processing power of the GPU.

3.2 Comparison of subjects' movements and standard movements

Figure 4 presents the experimental subjects' movements and the standard movement. Compared to the results obtained after difference comparison, the experimental subjects have a higher degree of leg bending within the time interval of 4–6 s, which is different from the standard movement. Moreover, within the time interval of 6–8 s,

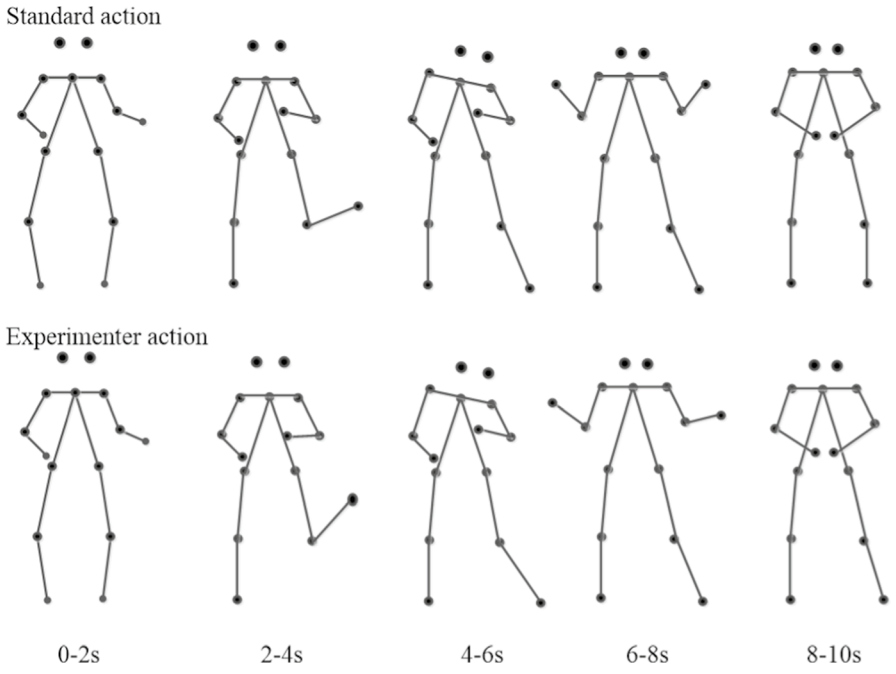

Fig. 4 Comparison of subjects' movements and standard movements

the swing amplitude of the left arm of the experimental subjects is quite different from the standard movement, and results show an obvious difference between the two movements.

3.3 Difference between subjects' left arm movement and standard movement

The difference between the left arm movement of the subjects and the standard movement can be indicated by using the quantitative comparison of image similarity as shown in Fig. 5.

Figure 5 betokens that within 1–5 s, the subjects' movements and the standard movement are highly consistent, with few differences. However, at 6 s, the subjects' movements are different from the standard movement. Within 6–8 s, the direction of the experimental subjects' movements is smaller than that of the standard movement. Within the time interval of 8–10 s, the magnitude of the movement direction becomes more massive. There is great difference between the movement of the subjects' left arm joint and the standard movement. The subjects' left-arm angle is smooth, while that of the standard fluctuates wildly. However, the experimental subjects have failed to achieve the same left-arm angle shown in the standard movement. There is not great difference between the experimental subjects' movement and the standard movement in terms of the angle between the left arm and the trunk;

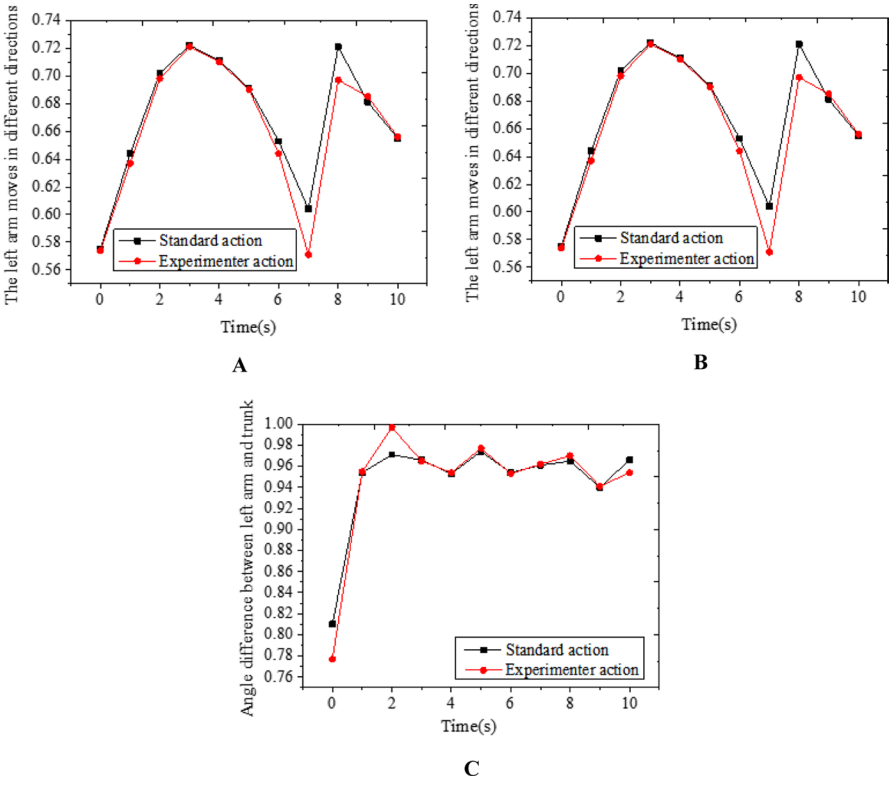

Fig. 5 The difference between the experimental subjects' left arm movement and the standard left arm movement (**A**: the difference in the movement direction of the left arm; **B**: the difference in the movement of the joint angle of the left arm; **C**: the difference in the angle between the left arm and the trunk)

the former and the latter are the same, which implies that the dance skill of the experimental subjects is high.

3.4 Difference between subjects' right arm movement and standard movement

Figure 6 reveals the difference between the subjects' right arm movement and the standard movement.

Figure 6 signals that the difference in the direction of the right arm between the subjects' movement and the standard movement is large at 3–6 s, while at other times, the subjects' movement is consistent with the standard movement. In terms of the difference in the angle movement of the right arm joint, the movement amplitude of the experimental subjects is small, and the movement amplitude is not as large as the standard movement. Regarding the difference in the angle between the right arm and the trunk, during 0–2 s, the angle of the experimental subjects is small; during 3–6 s, the magnitude of angle change is also small.

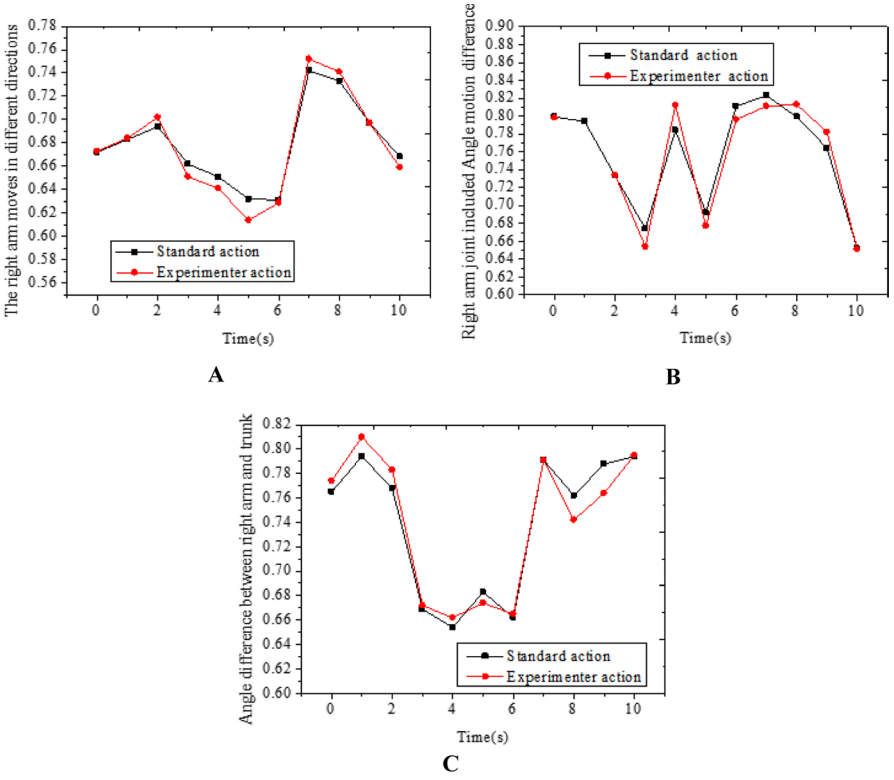

Fig. 6 The difference between the experimental subjects' right arm movement and the standard right arm movement (**A**: the difference in the movement direction of the right arm; **B**: the difference in the movement of the joint angle of the right arm; **C**: the difference in the angle between the right arm and the trunk)

3.5 Difference between subjects' left leg movement and standard movement

Figure 7 bespeaks the difference between the subjects' left leg movement and the standard movement.

Figure 7 hints that the performance of the experimental subjects is inadequate in terms of the difference in the direction of left leg movement. Within 0–5 s, the difference in the direction of the left leg movement differs greatly from the standard movement. For the experimental subjects' movement, the angle of the movement direction is not accurate enough, indicating that the subjects need more training. In terms of the movement of the left leg joint angle, differences appear at 4–6 s. This is caused by the experimental subjects' lack of familiarity with the movements. In terms of the angle between the left leg and the trunk, the differences between the subjects' movement and the standard movement appear at 5–8 s. Currently, the variation of the subjects' movement is excessive.

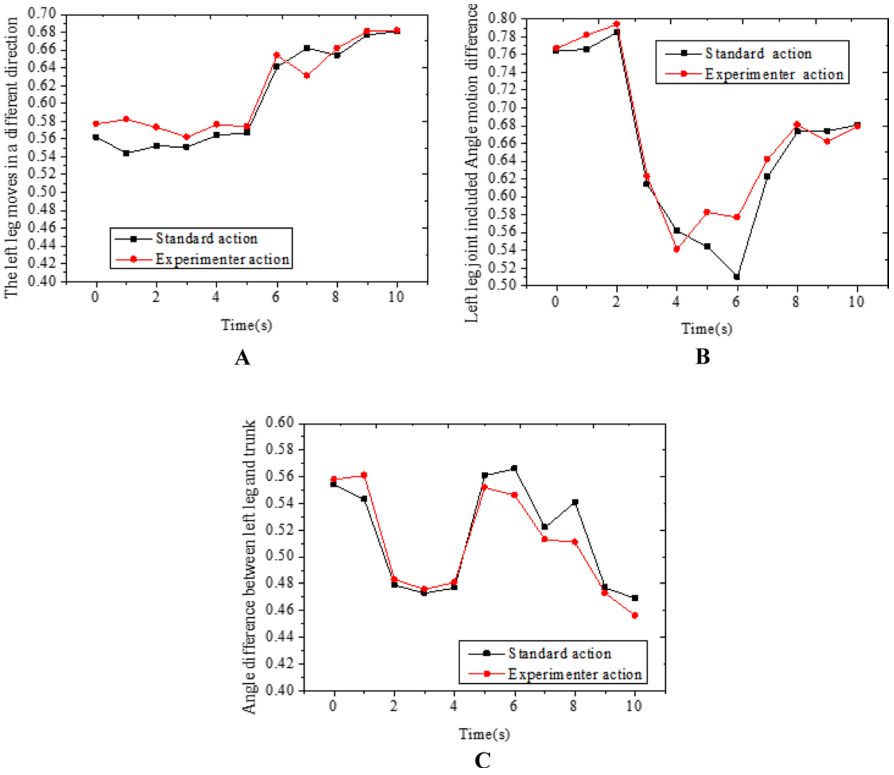

Fig. 7 The difference between the experimental subjects' left leg movement and the standard left leg movement (**A**: the difference in the movement direction of the left leg; **B**: the difference in the movement of the joint angle of the left leg; **C**: the difference in the angle between the left leg and the trunk)

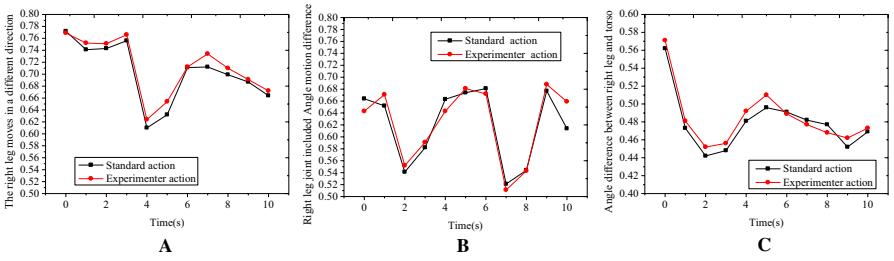

Fig. 8 The difference between the experimental subjects' left leg movement and the standard left leg movement (**A**: the difference in the movement direction of the right leg; **B**: the difference in the movement of the joint angle of the right leg; **C**: the difference in the angle between the right leg and the trunk)

3.6 Difference between subjects' right leg movement and standard movement

Figure 8 demonstrates the difference between the subjects' right leg movement and

the standard movement.

Figure 8 illustrates that differences in the direction of right leg movement are principally the same between subjects' movement and standard movement, except for some difference at a particular time point. Nevertheless, experimental subjects perform better. In terms of the differences in the angle movement of the right leg joint, the experimental subjects also perform well but fail to master the range of movement within 4–6 s. In terms of the difference between the angle between the right leg and the trunk, the experimental subjects' movements are consistent with the standard movement only within 0–1 s, while differences appear at other times. Hence, the experimental subjects need more practice in the leg-trunk angle.

3.7 Effect of CNN-based movement recognition model

The traditional manual features and machine learning (ML) modelling and recognition methods are selected and compared to further verify the effect of the proposed human movement recognition model based on CNN, which are support vector machines (SVM) and artificial neural network (ANN). A dance student's action is taken as an example, and Fig. 9 presents the ROC curves of the three methods. The average recognition rate of CNN- based human movement recognition model on public data is $90.47\% \pm 1.89\%$, which is better than that of traditional SVM ($83.44\% \pm 1.57\%$) and ANN ($79.36\% \pm 2.02\%$).

4 Discussion

Dance belongs to the art. In the short 30-year development period, dance has been affected by the unique mechanism of China's dance development and the selected training models of professional art colleges. The rapid increase in its degree of professionalization and scale expansion has attracted international organizations' attention [28]. However, at the top competitive level, China still lags behind of European

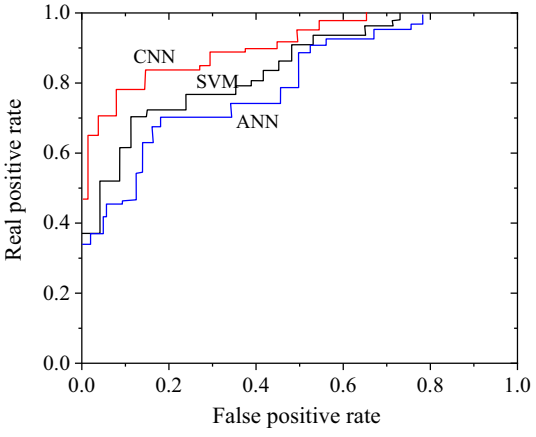

Fig. 9 Comparison of ROC curves of different movement recognition models

and American countries. This inferior position is attributed to the fact that China has concentrated too much energy on pure technology introduction and imitation, but lacking in-depth scientific understanding of the dance event and rapid refining of China's successful experience, which requires strong research evidences to assist the development of high-level dance movements [29]. From the perspective of global dance research, Western countries have researched dance movements earlier. Countries like the UK, the USA, and Russia have long-established research institutions for professional research on dance, and their research often focuses on dance competitions and technical movements. The research division between these institutions and universities is relatively straightforward, and there is mature dance specialization and popularization development system. There are many achievements in dance research in China; however, the research institutions and researchers are relatively scattered, the cooperation is not right, and the research system is not mature enough [30].

In China, research is biassed towards studies on dance value and artistic expression. Although there are also studies describing the current progress of teaching and training of dance art, the exploration of scientific training is only based on the analysis and discussion of textual data, and a few qualitative studies have not followed the scientific research paradigm. In comparison, the quality of dance research in developed countries is significantly higher, and the research literacy of Chinese researchers and the scientific nature of research methods urgently need to be improved. Motion recognition refers to recognizing and analyzing the moving behaviours in a video or image sequence by using the computer. This method of extracting information from image data and adopting appropriate model modelling has been reported in many studies [31–33]. As one of the major development directions in the field of artificial intelligence, behaviour recognition technology has broad application prospects and significant theoretical research significance [34]. However, traditional methods based on feature extraction usually require more prior knowledge; hence, the workload is enormous, and the robustness is lacking [35, 36]. Research on human posture recognition can provide a significant foundation development of AI technology and human–computer interaction. Vecchio et al. (2017) inferred users' poses based on the distance between the devices worn by the users and measured the distance via an ultra-wideband transceiver with two-way ranging mode; then, the measurement results were input into the classifier of current poses; the results found that the accuracy of this method was 98.2% [37]. Due to the inconvenience of wearing extracorporeal devices, the above methods have not been researched. Therefore, in the future, extracorporeal devices can be improved to enrich recognition methods.

The proposed dance movements recognition model can provide high accuracy and strong practicability in multiple application fields, such as competition refereeing, dance movements teaching, and dance movements correction. It has strong guiding significance for the design and performance improvement of computer vision systems in China. Using CNN to recognize dance movements has made some progress; however, there are some limitations. For example, in the case of complex backgrounds and shadows in the video, the progress of the extraction algorithm for human contours needs to be improved. Therefore, constructing a robust human contour extraction algorithm can improve the recognition accuracy and find the

optimal movement feature. In the future, more effective feature combinations will be explored to optimize the network structure, improve recognition accuracy, and reduce the training time.

5 Conclusion

A DL-based model is proposed based on microscopic visual morphology regarding the current problems in recognizing high-level dance movements. This model includes a human dance movements recognition algorithm based on CNN and a dance movements classification image feature similarity analysis algorithm. On public data, the average recognition rate of the human dance movements recognition algorithm based on CNN is $90.47\% \pm 1.89\%$, significantly better than the traditional SVM. In the future, CNN will still be a key technology in movements recognition. Applying CNN to truly realize intellectualization is one of the significant challenges in the current situation. CNN is utilized for researching the recognition technology of dance movements. The innovation lies in introducing CNN into action modelling and recognition for serial data. Aiming at the serial data such as human 3D posture data and EEG data, the corresponding deep CNN structure is optimized to achieve the optimal recognition effect on the corresponding data sets. The proposed method has good performance on dance video action recognition tasks. This shows that the research is of great value for improving the recognition accuracy of dance movements and also has certain reference significance for action recognition in other fields.

Acknowledgements This research was supported by the project of 2020 Teaching Reform Research Project of Colleges and Universities in Hunan Province “Innovative Research on Training Mode of Outstanding Dance Talents in Application-Oriented Colleges and Universities under the Background of Professional Certification” (No.HNJG-2020-0789). It was also supported by the project of Key project of Yiyang Social Science Project in 2020 "Research on Nanxian Dihuagu Dance from the Perspective of Semiotics" (No. 2020YS092).

References

1. Yahya N, Musa H, Ong ZY et al (2019) Classification of motor functions from electroencephalogram (EEG) signals based on an integrated method comprised of common spatial pattern and wavelet transform framework. Sensors 19(22):4878–4883
2. Zbontar J, LeCun Y (2016) Stereo matching by training a CNN to compare image patches. J Mach Learn Res 17(1–32):2–14
3. Jin KH, McCann MT, Froustey E et al (2017) Deep CNN for inverse problems in imaging. IEEE Trans Image Process 26(9):4509–4522
4. Ovtcharov K, Ruwase O, Kim JY et al (2015) Accelerating deep CNNs using specialized hardware. Microsoft Research Whitepaper 2(11):1–4
5. Xiong H, Bairner A, Tang Z (2020) Embracing city life: physical activities and the social integration of the new generation of female migrant workers in urban China. Leis Stud 39(6):782–796
6. Poria S, Cambria E, Gelbukh A (2016) Aspect extraction for opinion mining with a deep CNN. Knowl-Based Syst 108:42–49

7. Chen J, Qiu J, Ahn CR et al (2017) Construction worker's awkward posture recognition through supervised motion tensor decomposition. Autom Constr 77:67–81
8. Huang J, Yu X, Wang Y et al (2016) An integrated wireless wearable sensor system for posture recognition and indoor localization. Sensors 16(11):1825–1836
9. Kamala V R, MaryGladence L (2015) An optimal approach for social data analysis in Big Data. International Conference on Computation of Power, Energy, Information and Communication (ICCPEIC). IEEE, pp. 0205–0208.
10. Yan C, Coenen F, Zhang B et al (2016) Driving posture recognition by CNNs. IET Comput Vision 10(2):103–114
11. Nath S, Sinha S, Gladence L M, et al (2017) Health analysis of bicycle rider and security of bicycle using IoT. International Conference on Communication and Signal Processing (ICCSP). IEEE, pp. 0802–0806.
12. Gladence L M, Sivakumar H H, Venkatesan G, et al (2017) Home and office automation system using human activity recognition. International Conference on Communication and Signal Processing (ICCSP). IEEE, pp. 0758–0762.
13. Varol G, Laptev I, Schmid C et al (2018) Long-term temporal convolutions for action recognition. IEEE Trans Pattern Anal Mach Intell 40(6):1510–1517
14. Lv Z (2020) Virtual reality in the context of Internet of Things. Neural Comput Appl 32(13):9593–9602
15. Moeskops P, Viergever MA, Mendrik AM et al (2016) Automatic segmentation of MR brain images with a CNN. IEEE Trans Med Imaging 35(5):1252–1261
16. Anthimopoulos M, Christodoulidis S, Ebner L et al (2016) Lung pattern classification for interstitial lung diseases using a deep CNN. IEEE Trans Med Imaging 35(5):1207–1216
17. Shafiee A, Nag A, Muralimanohar N et al (2016) ISAAC: A CNN accelerator with in-situ analog arithmetic in crossbars. ACM SIGARCH Comput Arch News 44(3):14–26
18. Brumancia E, Samuel SJ, Gladence LM et al (2019) Hybrid data fusion model for restricted information using Dempster-Shafer and adaptive neuro-fuzzy inference (DSANFI) system. Soft Comput 23(8):2637–2644
19. Senduran C, Gunes K, Topaloglu D et al (2018) Mitigation and treatment of pollutants from railway and highway runoff by pocket wetland system. A case study Chemosphere 204:335–343
20. Yang B, Duan H, Wu S et al (2019) Damage tolerance assessment of a brake unit bracket for high-speed railway welded bogie frames. Chin J Mech Eng 32(1):58–64
21. Brisset S, Ogier M (2019) Collaborative and multilevel optimizations of a hybrid railway power substation. Int J Numer Model Electron Networks Devices Fields 32(4):e2289–e2296
22. Koziarski M, Cyganek B (2017) Image recognition with deep neural networks in presence of noise – Dealing with and taking advantage of distortions. Int Comput-aided Eng 24(4):337–349
23. Lin Z, Mu S, Shi A et al (2018) A novel method of maize leaf disease image identification based on a multichannel CNN. Trans ASABE 61(5):1461–1474
24. Okugawa Y, Kubo M, Sato H et al (2019) Evaluation for the synchronization of the parade with openpose. J Robotics, Netw Artif Life 24:162–166
25. Yoo HR, Lee BH (2019) An openpose-based child abuse decision system using surveillance video. J Korea Inst Inf Commun Eng 23(3):282–290
26. Tsai YS, Hsu LH, Hsieh YZ et al (2020) The real-time depth estimation for an occluded person based on a single image and openpose method. Mathematics 8(8):1333–1343
27. Draisma J, Horobeţ E, Ottaviani G et al (2016) The euclidean distance degree of an algebraic variety. Found Comput Math 16(1):99–149
28. Naevdal G (2019) Positive bases with maximal cosine measure. Optimization Lett 13(6):1381–1388
29. Davies P, Rowe M, Brown DM et al (2020) Understanding the status of evidence in policing research: reflections from a study of policing domestic abuse[J]. Policing Soc 31(6):1–15
30. Lepore W, Hall BL, Tandon R (2020) The Knowledge for Change Consortium: a decolonising approach to international collaboration in capacity-building in community-based participatory research. Can J Dev Studies/Revue canadienne d'études du développement. 42(3):1–24
31. Coelho YL, Salomao JM, Kulitz HR (2017) Intelligent hand posture recognition system integrated to process control. IEEE Lat Am Trans 15(6):1144–1153
32. Brünger J, Gentz M, Traulsen I et al (2020) Panoptic segmentation of individual pigs for posture recognition. Sensors 20(13):3710–3718
33. Wu Z, Zhang J, Chen K et al (2019) Yoga posture recognition and quantitative evaluation with wearable sensors based on two-stage classifier and prior Bayesian network. Sensors 19(23):5129–5136

34. Cai X, Gao Y, Li M et al (2016) Infrared human posture recognition method for monitoring in smart homes based on hidden Markov model. Sustainability 8(9):892–905
35. Fang L, Liang N, Kang W et al (2020) Real-time hand posture recognition using hand geometric features and Fisher Vector. Signal Proc: Image Commun 82:115729–115737
36. Cheng K, Ye N, Malekian R et al (2019) In-air gesture interaction: Real time hand posture recognition using passive RFID tags. IEEE Access 7:94460–94472
37. Vecchio A, Mulas F, Cola G (2017) Posture recognition using the interdistances between wearable devices. IEEE Sens lett 1(4):1–4

Authors and Affiliations

Ruizhi Zhang[1]

[1] College of Music and Dance, Hunan First Normal University, Changsha 413000, China