Computer Modeling in Engineering & Sciences
DOI:10.32604/cmes.2020.011046
Article

Tech Science Press

A Multi-View Gait Recognition Method Using Deep Convolutional Neural Network and Channel Attention Mechanism

Jiabin Wang* **and Kai Peng**

College of Engineering, Huaqiao University, Quanzhou, 362021, China
*Corresponding Author: Jiabin Wang. Email: wjb202004@163.com
Received: 16 April 2020; Accepted: 20 July 2020

Abstract: In many existing multi-view gait recognition methods based on images or video sequences, gait sequences are usually used to superimpose and synthesize images and construct energy-like template. However, information may be lost during the process of compositing image and capture EMG signals. Errors and the recognition accuracy may be introduced and affected respectively by some factors such as period detection. To better solve the problems, a multi-view gait recognition method using deep convolutional neural network and channel attention mechanism is proposed. Firstly, the sliding time window method is used to capture EMG signals. Then, the back-propagation learning algorithm is used to train each layer of convolution, which improves the learning ability of the convolutional neural network. Finally, the channel attention mechanism is integrated into the neural network, which will improve the ability of expressing gait features. And a classifier is used to classify gait. As can be shown from experimental results on two public datasets, OULP and CASIA-B, the recognition rate of the proposed method can be achieved at 88.44% and 97.25% respectively. As can be shown from the comparative experimental results, the proposed method has better recognition effect than several other newer convolutional neural network methods. Therefore, the combination of convolutional neural network and channel attention mechanism is of great value for gait recognition.

Keywords: EMG signal capture; channel attention mechanism; convolutional neural network; multi-view; gait recognition; gait characteristics; back-propagation

1 Introduction

Human interaction with the external environment is multimodal. With the enhancement of computer's ultra-large-scale parallel computing capabilities and the development of various sensor technologies, it has provided new research ideas for analyzing and understanding human behavior and intentions. For example, the combination of deep learning [1–3] and voice signals, the combination of deep learning and vision-based human motion signals, the combination of deep learning and signals based on human motion intentions that indicate electromyography, the combination of deep learning and motion intent signals based on electroencephalogram, and even the combination of deep learning and human emotion signals based on electroencephalogram. As one of the human body motion recognition [4], gait recognition is used to

authenticate or recognize the identity by walking posture or footprints and is one of the most effective methods in long-distance identity recognition. As one of human body motion recognition, gait recognition uses the myoelectric signal of human walking to authenticate or recognize the identity. It is considered as one of the most effective methods in long-distance identity recognition. Therefore, it is very important to design an efficient gait recognition method.

Gait recognition faces many difficulties in practical applications. The main manifestation is that pedestrians are affected by the external environment and their own factors during walking, which leads to strong intra-class changes in extracted gait characteristics. The perspective factor is one of the most important factors affecting the recognition performance of systems. When a pedestrian's walking direction changes, or when the pedestrian moves from one surveillance area to another surveillance area with different setting, the perspective changes occur. The gait images in different perspectives have great differences and the gait contours inside perspectives [5,6] contain more valuable information. Thus, feature extraction is mostly based on side contours. However, the traditional single-view gait recognition has a significant decrease in recognition performance when the perspective changes.

Aiming at multi-view gait recognition, this paper proposes a new method combining deep Convolutional Neural Network [7–10] and Channel Attention Mechanism (CAMCNN). The main innovations of this paper are summarized as follows:

1. A new convolutional neural network architecture is proposed. Each layer is trained by back propagation learning, which enhances the learning ability of convolutional neural network.
2. A Convolutional Neural Network (ACNN) combined with channel Attention mechanism is proposed. Compared with the original Convolutional Neural Network (CNN), the convolution layer is further strengthened due to adding channel attention mechanism. It can better express gait characteristics, which helps to improve the accuracy of recognition.

The overall structure of this article is as follows: Section 2 introduces the related work for multi-view gait recognition and motivation. Section 3 introduces the architecture and detailed process of the proposed method. Section 4 introduces experimental verification on gait recognition datasets and comparative analysis with other existing methods. Section 5 introduces conclusions and future work.

2 Related Work

Scholars have proposed lots of methods for the problem of gait recognition. For example, Ghaeminia et al. [11] proposed gait saliency images and classified templates by applying appropriate spatiotemporal filters. However, when the amount of data is large, the effect of this model will be affected greatly. Xu et al. [12] proposed a subspace-based multi-view maximum edge subspace learning method. This method simultaneously minimizes intra-class variation and maximizes local inter-class variation from low-dimensional embedding between views and within views. The data from different views are mapped to the projection matrix of common subspaces to improve model recognition rate. However, this method is prone to overfitting. Shaikh et al. [13] proposed a gait recognition method based on partial contours, which improved the recognition accuracy using a multi-modal system. However, when there are lots of image noises, its effect will be affected greatly and needs to be further improved. More et al. [14] proposed a multi-view gait recognition system that combines two features. It extracts dynamic features by cross wavelets and extracts static features by a bipartite graph model, which improves the accuracy of this system after fusion. However, this method is prone to overfitting. Verlekar et al. [15] proposed a gait recognition method based on four-dimensional space, which identified users' walking direction by fitting the direction of lines. However, the parameters of this model are more difficult to adjust. Xing et al. [16] proposed an invariant gait recognition method based on 3D convolutional neural network, which extracts spatial and temporal information by learning viewpoint-invariant features to improve model performance.

However, more image noises will reduce the recognition accuracy. Zhang et al. [17] proposed a local discriminative gait recognition method, which improves the robustness by extracting robust local weighted histogram feature vectors for training. However, this method ignores some edge features. Wang et al. [18] proposed a generalized LDA gait recognition method based on Euclidean norm. It uses discrete matrices to separate adjacent samples and improves the accuracy of this model. However, when the amount of data is large, the classification effect is not good. Portillo-Portillo et al. [19] proposed a vision-invariant gait recognition algorithm based on joint direct linear discriminant analysis. It improves the stability of this model through the dimensionality reduction feature provided by direct linear discriminant analysis. However, its features extracted by this method need to be further improved. Chhatrala et al. [20] proposed a gait recognition algorithm based on hidden Markov model. It uses moving average filter model to denies the gait data, which improves the accuracy of this model. However, the model takes longer to run. Yu et al. [21] proposed a gait recognition method based on curve transformation and PCANet. It extracts differentiated robust features by non-linear and irreversible reversal of PCANet, which improves the effectiveness of this model. However, some edge features are difficult to extract.

Gadaleta et al. [22] proposed the extraction of invariant gait based on the depth model of auto encoder. It realizes the stepwise synthesis of gait characteristics by multi-layer self-encoding, which improves the accuracy of this model. However, the model is prone to overfitting. Rashwan et al. [23] proposed a gait recognition method based on histogram. This method uses a two-dimensional array of histograms to encode the dynamics of gait cycle and improves the model's accuracy. However, when the amount of data is large, the recognition effect of this method is not good. Sun et al. [24] proposed a vision-invariant gait recognition method based on Kinect skeleton features, which improves the accuracy by fusing static and dynamic features. However, this method is prone to overfitting. Wu et al. [25] proposed a feedback weighted convolutional neural network, which extracts features by controlling weights and improves the model recognition rate. However, the parameters of this model are more difficult to adjust.

Based on the above analysis, it can be known that deep learning has good modeling and processing capabilities for massive data. Most existing multi-view gait recognition methods based on images or video sequences use gait sequences to superimpose and synthesize images for constructing energy map-like templates. However, information may be lost during image synthesis process, and the influence of factors such as period detection may introduce errors and affect the accuracy of recognition. In order to better solve these problems, this paper continues to study multi-view gait recognition based on deep convolutional neural networks. It mainly includes two parts: improving the learning ability of CNN by strengthening the training of each layer, and gait characteristics can be better expressed by strengthening the convolutional layer.

3 Proposed Method

The general process of gait recognition based on two-dimensional visual perception is: data preprocessing, gait contour extraction, gait cycle calculation, gait feature extraction, similarity calculation and gait classification. According to this process, we design a new gait recognition method.

3.1 Overall Architecture of the Proposed Method

The flow chart of convolutional neural network combined with channel attention mechanism is shown in Fig. 1. *Acc* represents acceleration, *Gry* represents angular velocity, *Con* represents convolution kernel, *Pool* represents pooling operation, *Attention* represents attention mechanism module and *Output* represents final classification features.

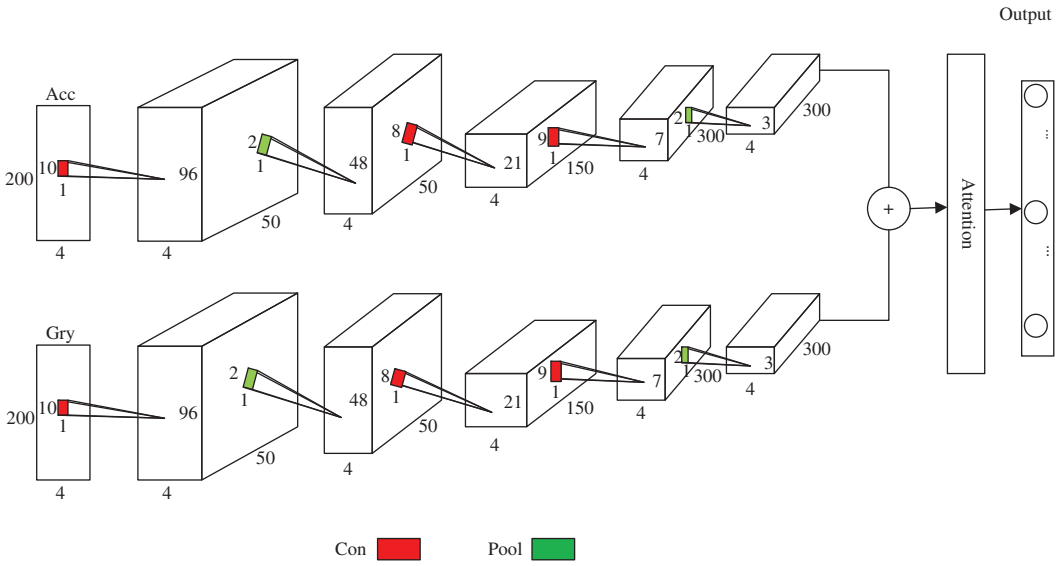

Figure 1: Flow chart of the proposed method

3.2 Lower Limb Electromyography Signal Capture

The lower limb EMG signal is a statistical waveform signal with zero mean. Observed from the perspective of mathematical, the signal can be considered as a system described by differential equations. Therefore, let the amplitude of the EMG signal be the variable x and the derivative of x with respect to time is the variable y. Therefore, (x, y) can be considered as a coordinate of a state point and the phase diagram of the EMG signal segment can be drawn in the $x - y$ phase plane. Fig. 2 shows the original EMG signal with a sampling frequency of 2000 Hz.

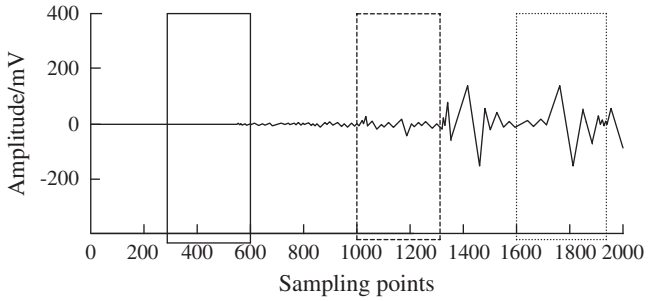

Figure 2: Original EMG signals of lower limbs

Take three rectangular windows on the signal, assuming a simple harmonic oscillator with a mass of m and a spring rate of k. Similarly, the amplitude and speed are expressed by variables x and y. The state of the vibrator can be expressed by the following dynamic system.

$$\begin{cases} x=y \\ y=-\dfrac{k}{m}x \end{cases} \tag{1}$$

x and y represent the amplitude and velocity of the vibrator respectively. And for a simple harmonic oscillator, x^2 is proportional to the potential energy of the vibrator and y^2 is proportional to its kinetic energy. Therefore, the total energy E of the vibrator is:

$$E = \frac{1}{2}kx^2 + \frac{1}{2}my^2 \tag{2}$$

Convert Eq. (2) to the following format:

$$\frac{x^2}{2E/k} + \frac{y^2}{2E/m} = 1 \tag{3}$$

The energy core S can be expressed as Eq. (3)

$$S = \frac{2\pi}{\sqrt{km}}E \tag{4}$$

As can be seen from Eq. (4), the energy core is proportional to the energy of the EMG signal oscillator.

Besides, the energy of the EMG signal is the sum of the energy of each harmonic, that is, the energy of the action potential of the motor unit determines the energy of the EMG vibrator. Therefore, k and m reflect the inherent physical characteristics of the action potential conduction medium of the motor unit. For a list of harmonics, the average energy density is:

$$\overline{E} = \frac{1}{2}\rho A^2\omega^2 \tag{5}$$

ρ is the mass density of the medium. A is the amplitude, ω is the angular frequency of the amplitude, And, the frequency of the motor unit action potential corresponding to the action potential of the motor neuron and the dominant frequency component is recorded as ω_F. Then, Eq. (5) can be rewritten as:

$$\overline{E} \cong \frac{1}{2}\rho(A_i^2)\omega_F^2 \tag{6}$$

A_i represents the amplitude of the i component. It can be seen from Eq. (6) that the square root of E is directly proportional to ω_F and signal strength. Therefore, there is a linear relationship between the square root of E and the muscles.

Through the above analysis, when estimating the EMG signal, the sliding time window method can be used to calculate the energy core S in each window combining with Eq. (4). Then, the EMG signal is characterized by $\sqrt{S}$.

3.3 Data Preprocessing and Gait Cycle Calculation

3.3.1 Data Interpolation and Demising

Due to the inaccuracy of software clock, the smartphone's sampling [26] of gait data is uneven, and the obtained data sampling intervals are inconsistent. For the convenience of processing, the data is firstly spline interpolated three times to achieve one data every 5 milliseconds. A complete gait cycle takes about 1 second, so a gait cycle has 200 data points approximately. Then a low-pass Finite Impulse Response (FIR) filter is used to denies the data after interpolation and reduce motion artifacts that may occur at higher frequencies. Generally, the cut-off frequency is $f = 40$ Hz and the window length is set to 1 second.

3.3.2 Normalization of Contours

In the acquired image data, target contours occupy only a small part of original images. Besides, since the camera has a fixed angle when shooting, there is a change in the distance between pedestrian and cameras, which directly affects the size of silhouette contours. In this paper, contour images are cropped to obtain the target silhouette image, and then normalized to scale it into a fixed-size template. The bilinear interpolation

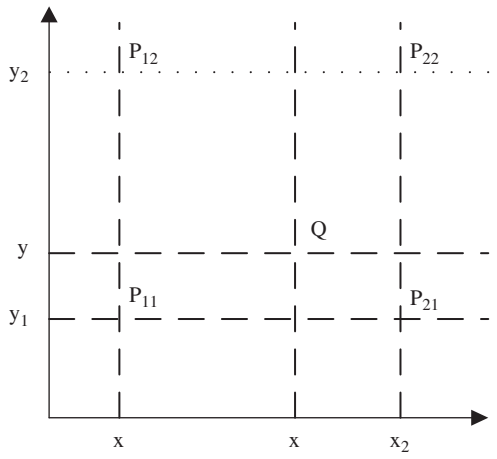

Figure 3: Bilinear interpolation method

method is mainly used and Eq. (1) is used to transform the image scale [27]. As shown in Fig. 3, $P_{11} = (x_1, y_1)$, $P_{12} = (x_1, y_2)$, $P_{21} = (x_2, y_1)$, and $P_{22} = (x_2, y_2)$ are the positions of 4 pixels, $Q = (x, y)$, $f(P)$ are the pixel values of points.

$$f(x,y) = \frac{1}{(x_2 - x_1)(y_2 - y_1)} \times [x_2 - xx - x_1] \begin{bmatrix} f(P_{11}) f(P_{12}) \\ f(P_{21}) f(P_{22}) \end{bmatrix} \begin{bmatrix} y_2 - y \\ y - y_1 \end{bmatrix} \tag{7}$$

3.3.3 Gait Cycle Extraction

A complete gait cycle [28] refers to the process from one side heel to the ground to this heel to the ground again when walking. It contains the single-step feature of gait and is an important basis for gait recognition. In general, gait characteristics of the same person are stable and unique. Thus, continuous gait cycles should be highly correlated. In order to detect a dynamically changing gait cycle, firstly a gait cycle that is easier to distinguish is identified, and the gait signal of this cycle is used as a template. Then use the template matching to find signal segment with the greatest correlation as the next gait cycle. At the same time, iterative update of the template is performed, so that the detection of the next cycle is more accurate. During the gait cycle detection process, gait signals used is mainly the amplitude signal of total acceleration. For each sample $i (i = 1, 2, \ldots)$, the total acceleration amplitude signal is calculated as follows:

$$a_{mag(i)} = (a_x^2 + a_y^2 + a_z^2)^{1/2} \tag{8}$$

Let the amplitude value $a_{mag}(i^{'})$ of $i^{'}$ be a minimum value at the beginning of gait signals. Using $i^{'}$ as the center, we extract 200 acceleration data sets, the formula is as follows:

$$Z = (a_{mag}(i^{'} - 99), \ldots, a_{mag}(i^{'}), \ldots, a_{mag}(i^{'} + 100)) \tag{9}$$

where Z represents acceleration template in the first cycle. Let $C(i)$ be the next continuous data segment starting with point i, and its length is $N = 200$, that is:

$$C(i) = (a_{mag}(i), \cdots, a_{mag}(i + N - 1)) \tag{10}$$

The correlation distance $V(i)$ between $C(i)$ and the template is calculated as follows:

$$\overline{Z} = \frac{\sum_{i=1}^{N} Z_i}{N} \tag{11}$$

$$\overline{C} = \frac{\sum_{i=1}^{N} C_i}{N} \tag{12}$$

$$V(i) = \frac{(Z - \overline{Z}1_N) \times (C - \overline{C}1_N)}{||Z - \overline{Z}_N||||C - \overline{C}_N||} \tag{13}$$

Here, $\overline{Z}$ and $\overline{C}$ represent the mean of each element in vectors Z and C respectively. 1_N is a vector of all 1s of length N, and $||.||$ is the second norm of the vector.

The correspondence between two maximums is a gait cycle. The location of a large value can be located by a simple threshold method. A threshold of 0.5 is sufficient. Use the first template to find the next gait cycle, which is the second cycle $Z^{'}$. Its update formula is as follows:

$$Z = 0.9Z + 0.1Z \tag{14}$$

It can be seen that the new template is a weighted average of old templates Z and $Z^{'}$. The above process continues until the last gait cycle of the data. In this way, a relatively accurate template can be obtained in each new cycle.

3.3.4 Directional Projection

Since the data is collected by mobile phone in pants pocket, and the position of mobile phone is not fixed, the acceleration and angular velocity will be slightly shifted in direction. To this end, a new direction-invariant coordinate system needs to be established for the collected data [29]. The three orthogonal coordinate axes in the new coordinate system are independent of the direction of smartphone and aligned with the direction of gravity and motion. Let a gait cycle sample length be N_1, the acceleration and angular velocity of each sample are expressed as follows:

$$A = [a_x, a_y, a_z] \tag{15}$$

$$K = [k_x, k_y, k_z] \tag{16}$$

where x represents the direction facing vertical screen of the phone, y represents the direction from left to right of the phone, and z represents the direction from bottom to the top of phone. A represents acceleration and K represents angular velocity. a_x, a_y and a_z represent acceleration vectors in x, y and z directions. k_x, k_y and k_z represent the angular velocity vectors in x, y and z directions.

The acceleration in the direction of gravity is main low frequency component in accelerometer data. However, since the position of smart phone changes during walking, the acceleration in the direction of gravity is not a constant vector in (x, y, z) coordinate system. To this end, the mean acceleration vector in a gait cycle is used to estimate the gravity acceleration vector, which is expressed as follows:

$$\overline{(G)} = (\overline{a}_x; \overline{a}_y; \overline{a}_z) \tag{17}$$

Here, $\overline{a}_x$, $\overline{a}_y$ and $\overline{a}_z$ represent the mean vector of accelerations in x, y and z directions in a gait cycle. Then, the first coordinate axis direction of the new coordinate system is calculated as follows:

$$f_1 = \frac{\overline{G}}{||\overline{G}||} \tag{18}$$

To find the second direction, the original acceleration is projected onto a horizontal plane orthogonal to f_1. Let $A_1 = [a_x^1; a_y^1; a_z^1]$ be the acceleration data in horizontal plane, where a_x^1, a_y^1 and a_z^1 are the projected components (assuming that their lengths are N_1). The calculation formula is as follows:

$$A_1 = A - f_1 \times (A^T f_1)^T \tag{19}$$

In the horizontal direction, we set the direction in which the acceleration data changes the most (that is, the direction of travel with the largest variance) as the second coordinate axis of the new coordinate system. For this reason, Principal Component Analysis (PCA) is used to calculate the direction in which the data variance is greatest. Firstly, the covariance matrix is calculated, the formula is as follows:

$$H_1 = \frac{(A_1 - \sum_{\mathrm{i}=1}^{N_1} A_1 1_{N1}) \times (A_1 - \sum_{i=1}^{N_1} A_1 1_{N1})^T}{N_1 - 1} \tag{20}$$

where 1_{N1} is an all 1 vector of length N_1 The eigenvector. H_1 corresponding to the maximum eigenvalue of h_1 is the direction of maximum variance. In this way, the direction of the second coordinate system is calculated as follows:

$$f_2 = \frac{h_1}{|h_1|} \tag{21}$$

Since the above two directions are orthogonal, the third direction can be obtained by cross product:

$$f_3 = f_1 \times f_2 \tag{22}$$

Original acceleration and angular velocity data are projected to the new coordinate space. Each component is calculated as follows:

$$a_{f1} = A^T f_1; a_{f2} = A^T f_2; a_{f3} = A^T f_3 \tag{23}$$

$$k_{f1} = K^T f_1; k_{f2} = K^T f_2; k_{f3} = K^T f_3 \tag{24}$$

Then get the new gait data after coordinate transformation:

$$A^{'} = (a_{f1}, a_{f2}, a_{f3}); K^{'} = (k_{f1}, k_{f2}, k_{f3}) \tag{25}$$

3.3.5 Data Normalization

Due to the changes in walking speed and stride, each gait cycle has a different duration, which causes the data length of each gait cycle to be inconsistent. Deep learning models usually need to keep the length of input data consistent. According to the characteristics of gait data in this paper, the data length is unified to 200 by interpolation and extraction in the experiment. In order to obtain better training and classification performance, the data is amplitude normalized to obtain a vector with zero mean and unit variance. Since acceleration and angular velocity each have data in three directions (x, y, z), plus the calculated total acceleration and total angular velocity, there are eight vectors of length 200 for each gait cycle. They together constitute the input signals in the experiments.

3.4 Gait Feature Extraction

For gait feature extraction, we propose a deep convolutional neural network architecture, which is suitable for gait recognition. The architecture consists of 8 layers, including 4 convolutional layers and 4 sampling layers. Besides, there are 8 feature maps in each layer. In each convolutional layer, 8 convolution filters are used for initialization, and there are 8 sampling maps in each sampling layer. The architecture trains these layers using a back propagation [30] learning algorithm. It also uses the root mean square propagation of stochastic gradient descent with an adaptive learning rate to optimize the algorithm, thereby minimizing the cost function [31].

3.4.1 Convolution Method

Use *Xavier* uniform variance scaling method to initialize the weights of convolution filter:

$$Var(W) = \sqrt{(6/(Fan^{in} + Fan^{out}))} \tag{26}$$

A convolution filter is applied with a step size of 1 and a size of 5×5.

The output feature map is added to the bias term and the result is transformed by a non-linear activation function. Each feature map in convolutional layer is calculated as follows:

$$FM^i = \text{Tanh}(W^i \otimes FM^{i-1} + \beta^i) \tag{27}$$

Here, $\otimes$ represents convolution operation, and FM^{i-1} is the feature map of previous layer. In the first layer, FM^{i-1} represents the original pixels of GEI. Each feature map has a bias term β, which is initialized to zero. The activation function is:

$$\text{Tanh}(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{28}$$

Here, x is the result of convolution operation. This operation is added to the bias term of the feature map, as shown in Eq. (21). In the first convolution layer, each unit outputs a feature map of 136×136. In layer 3, each unit outputs 64×64 feature maps. In layer 5, each unit outputs 28×28 feature maps. In the final convolutional layer, each of 8 filters produces a 10×10 output feature map. The output of the final convolutional layer is directly fed to 8 subsampling units in the pooling layer.

3.4.2 Pooling Method

In the proposed deep convolutional neural network architecture, each pooling layer outputs 8 pooled feature maps and summarizes the output values of the adjacent neuron groups mapped by each kernel. At the same time, the pooling layer also helps to reduce spectral changes in the input data and produce translation-invariant features. Because the body shape in gait recognition is a non-rigid shape that can experience many fluctuations, this advantage is very valuable in gait recognition. In the model of this paper, the pooling unit performs maximum pooling, where pooling factor $C = 2$. The data is down sampled using a maximum pooling filter with a pooling unit size of 2×2 in steps of 2. The pool windows in the model do not overlap and the specific operations are defined as follows:

$$FM^i = MaxP(FM^{i-1}) \tag{29}$$

Here, $MaxP$ is the maximum pool operation. In the first sub-sampling layer, each of the 8 merge filters produces 68×68 outputs. In the fourth layer, each pooling filter produces 32×32 outputs. In the sixth layer, each layer produces 14×14 outputs. In the last pooling layer, each of the 8 pooling filters produces 5×5 outputs. In the fully connected part, there are only two layers (input layer and output layer), where *soft* max is the classifier of this paper.

The proposed architecture does not have any hidden layers. The input layer has 200 neurons, which are mainly from the last pooling layer ($5 \times 5 \times 8$).

3.4.3 Layer Connection

The original deep convolutional neural network architecture consists of millions of parameters and is trained on large data sets [32]. However, the data set is relatively small and cannot train all of these parameters in gait recognition. Therefore, overfitting problems may occur.

In the proposed deep convolutional neural network architecture, each feature map FM^i in l layer is connected to only one feature map FM^i from the previous $l-1$ layer. It greatly reduces the computational cost, speeds up training time and reduces the number of parameters. Fig. 4 shows an example of a one-to-one connection or a single connection between three layers of cores.

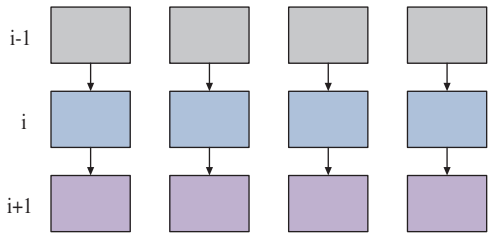

Figure 4: Schematic diagram of one-to-one connection between cores

3.5 Gait Classification

As an auxiliary method, Attention Mechanism (AM) [33] is increasingly being introduced into deep networks to optimize network structure. It is more like the mode in which human eyes observe things, making the network more focused on learning, thereby improving network's learning ability. Generally, Channel Attention (CA) [34] selects and optimizes different channels of the same feature map to obtain re-adjusted channel information. An improved CA is proposed for image processing and it is shown in Fig. 5. For a given intermediate feature map X ($L \times H \times W \times C$; L, H, W represents the spatial dimension of feature map, and C represents the number of channels). The principle is as follows:

$$C_M = FC(\delta(FC(MaxPooling(X)))) \tag{30}$$

$$C_A = FC(\delta(FC(AvgPooling(X)))) \tag{31}$$

$$X_C = \sigma(C_M + C_A)X \tag{32}$$

MaxPooling and AvgPooling represent the global maximum pooling and global average pooling in spatial direction respectively. And RELU is activation function, σ represents the sigmoid activation function, FC is a fully connected layer.

In the classification stage, the fully connected layer is used to compress high-dimensional feature ε to a lower dimension equal to the number of categories. And then the probability of corresponding category is calculated by classifier. The formula is as follows:

$$p = soft\max(W_{2\varepsilon} + b_2) \tag{33}$$

where W_2 and b_2 are weight matrices. The loss function of the entire network is classification cross entropy function, which is defined as:

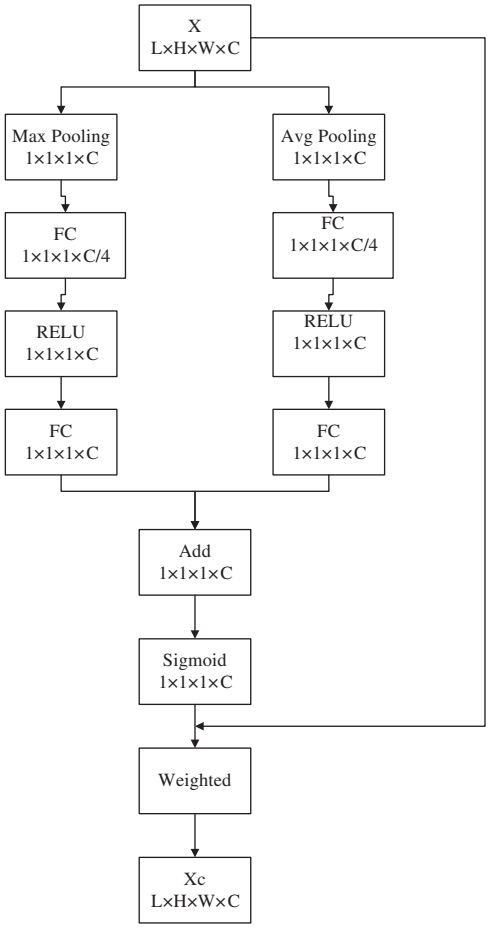

Figure 5: Schematic diagram of the channel mechanism

$$loss = -\sum_{j=1}^{k} y_j \times \log p \tag{34}$$

In the formula: k represents the number of training samples. During the training process, a random gradient descent algorithm is used to optimize and update all parameters in convolutional neural network combined with channel attention mechanism.

4 Experiments and Results

In order to verify the effectiveness of our proposed convolutional neural network combined with channel attention mechanism, sufficient experimental evaluations were performed on OULP dataset and CASIA-B dataset. Based on 3-D Convolutional Neural Network (3DCNN) proposed by [16], Deep Convolutional Location Weight Descriptor (DLWD) proposed by [25] and CAMCNN, we compared them by experiments. These methods are implemented in Python 3.0 using image processing toolbox.

4.1 Experimental Datasets

OULP dataset contains more than 4,000 experimenters, and each sample performs 2 normal walks. Each type of video clip is recorded by a camera with 4 angles (55°, 65°, 75° and 85°).

Figure 6: Sample images of OULP dataset

Figure 7: Sample images of the CASIA-B dataset

CASIA-B dataset contains 124 experimenters, recorded by cameras at 11 angles (0°−180°, 18° apart). Six normal walks are performed on each sample (only angle factors were considered). Figs. 6 and 7 show sample images from OULP and CASIA-B respectively.

4.2 Comparison of Recognition Results under Different Classifiers

The algorithm model is used to model the angle variables, and the most classic gait feature GEI is used as input. Support Vector Machine (SVM), Random Forest (RF) and Gradient Boosting Decision Tree (GBDT) were used on both OULP dataset and CASIA-B dataset. We performed two sets of experiments separately, and the experimental results are shown in Figs. 8 and 9.

As can be seen from Figs. 8 and 9, CAMCNN can better learn the features because of channel attention mechanism using the same classifier and gait characteristics. Therefore, the recognition results obtained by CAMCNN are superior to several other comparison methods. The reason for poor performance of SVM is that, since its calculation time increases sharply with the increase of training samples, it is not suitable for

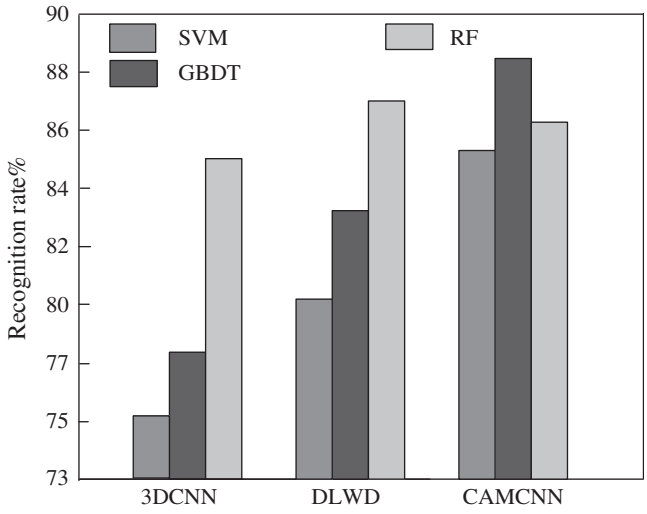

Figure 8: Recognition rates of OULP data sets on different neural networks based on different classifiers

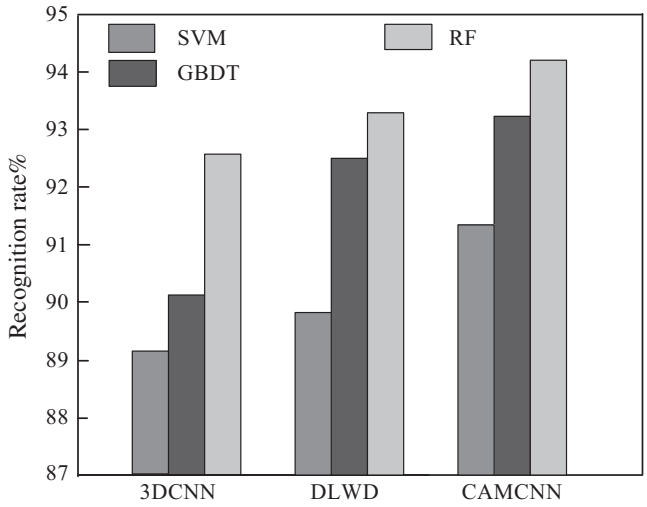

Figure 9: Recognition rates of CASIA-B data sets on different neural networks based on different classifiers

samples with large datasets. When the dataset is large, RF can process a large amount of input data, hence RF works best.

4.3 Comparison of Recognition Results with Advanced Methods

The simplest nearest neighbor classifier (K-Nearest Neighbor, KNN) [35] is used, and only the scenes with no angular difference in the state of cooperation are considered in the experiment. Since the experiment does not require a training set, the samples at each angle are used for testing. First test the performance of the original dimension on OULP dataset. Then, by the experiments, it is obtained that the computer performance and recognition accuracy have achieved a good balance of feature dimensions. Finally, the reference dataset and the probe data set are exchanged and then the recognition rate is obtained. Similarly, CASIA-B dataset also calculates the recognition rate at 11 angles in the same way.

The purpose of this experiment is to verify the difference in recognition rate under traditional manual features on different neural networks. It includes Gait energy image (GEI), Masked GEI based on GEnI

Table 1: Gait recognition rate of different neural networks based on different feature extraction methods on OULP dataset (%)

Comparative methods	Angle	GEI	MGEI	GEnI	GFI
3DCNN [16]	55	84.34	78.46	73.76	67.28
	65	84.65	82.31	75.61	68.95
	75	85.27	83.02	73.26	70.21
	85	84.78	81.65	71.49	70.64
	Average	84.76	81.36	73.53	69.27
DLWD [25]	55	86.21	82.31	77.22	71.02
	65	85.49	85.25	78.21	72.51
	75	87.27	86.03	77.41	74.47
	85	86.79	85.69	74.09	75.24
	Average	86.44	84.82	76.71	73.31
CAMCNN	55	89.63	88.54	84.31	76.32
	65	90.22	89.09	83.88	77.48
	75	90.69	90.54	85.63	79.16
	85	91.22	90.75	85.42	80.56
	Average	90.44	89.73	84.81	78.38

(MGEI), Gait entropy image (GEnI) and Gait optical flow diagram (Gait flow image, GFI). The multi-view gait recognition results obtained on OULP and CASIA-B datasets are shown in Tabs. 1 and 2.

As can be seen from Tabs. 1 and 2, the proposed CAMCNN can achieve a higher recognition rate on gait discrimination task using only the simplest classifier compared with 3DCNN and DLWD which shows that CAMCNN can learn features well. At the same time, it can be found that the recognition rate of this model is not significantly different under different angles. Therefore, the proposed CAMCNN has a certain degree of angle invariance.

4.4 CMC and ROC Curves

To verify the superiority of the proposed multi-view gait recognition method using deep convolutional neural network and channel attention mechanism, experiments were conducted on OULP and CASIA-B datasets using different convolutional neural networks. To ensure comparability, the feature extraction methods all use multi-scale audio difference normalization algorithm. The experimental results are shown in Figs. 10–13. Since the part is just to verify the quality of the model, to facilitate comparison with similar studies, feature extraction method uses GEI.

As can be seen from Figs. 10–13, compared with the other two existing methods, the recognition rate of the proposed method is higher. As can be seen from the analysis of reasons, the proposed method improves the learning ability of the network by optimizing the cost function. In addition, the proposed method incorporates the attention mechanism, which improves the ability of expressing gait features of the network and suppresses recognition errors effectively.

Table 2: Gait recognition rate of different neural networks based on different feature extraction methods on CASIA-B dataset (%)

Comparative methods	Angle	GEI	MGEI	GEnI	GFI
3DCNN [16]	0	83.44	79.09	87.19	82.54
	18	83.58	68.33	85.14	72.06
	36	78.21	72.14	80.32	71.25
	54	83.51	80.22	81.03	82.06
	72	87.14	84.61	87.05	85.49
	90	86.62	85.25	85.34	85.26
	108	88.21	83.42	88.41	82.43
	126	89.31	82.04	87.25	85.09
	144	92.47	81.48	88.47	88.69
	162	92.33	83.41	89.61	89.35
	180	93.06	88.59	89.05	90.21
	Average	87.08	80.78	86.26	83.13
DLWD [25]	0	84.95	82.31	91.31	84.19
	18	85.67	70.14	87.23	75.34
	36	80.47	74.24	83.13	73.51
	54	85.12	83.05	85.26	83.59
	72	90.34	87.23	89.03	87.84
	90	91.02	88.13	87.54	88.21
	108	90.47	86.14	91.27	85.29
	126	92.36	86.24	90.39	87.72
	144	93.01	85.41	90.21	91.34
	162	92.87	87.26	93.05	93.21
	Average	95.14	92.31	91.69	94.22
CAMCNN	0	92.29	85.49	92.49	86.48
	18	87.32	73.69	90.48	78.86
	36	82.51	76.31	85.52	76.53
	54	88.91	85.62	87.68	85.03
	72	92.38	89.53	91.47	90.48
	90	93.26	91.22	92.42	91.54
	108	92.34	89.16	94.21	88.53
	126	92.57	89.54	92.16	90.47
	144	94.32	87.64	92.54	94.49
	162	93.33	92.31	95.41	95.55
	180	96.15	95.28	93.33	96.42
	Average	91.58	86.89	91.61	88.58

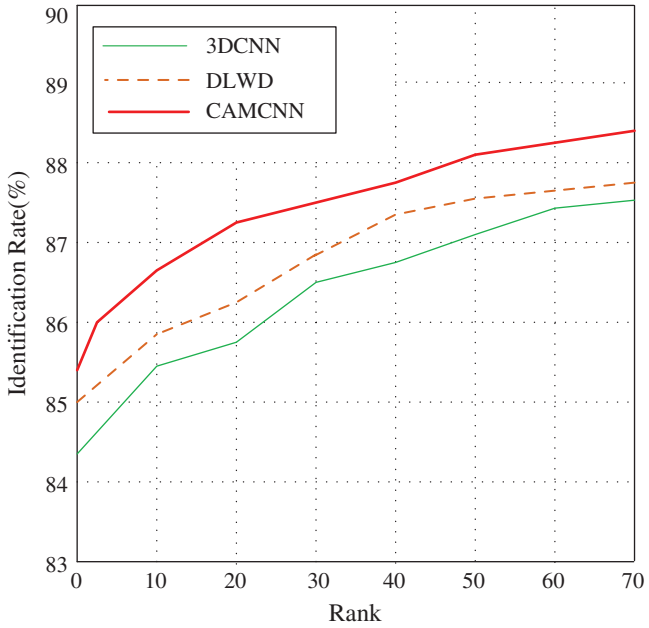

Figure 10: CMC curve of OULP dataset under non-cooperative state

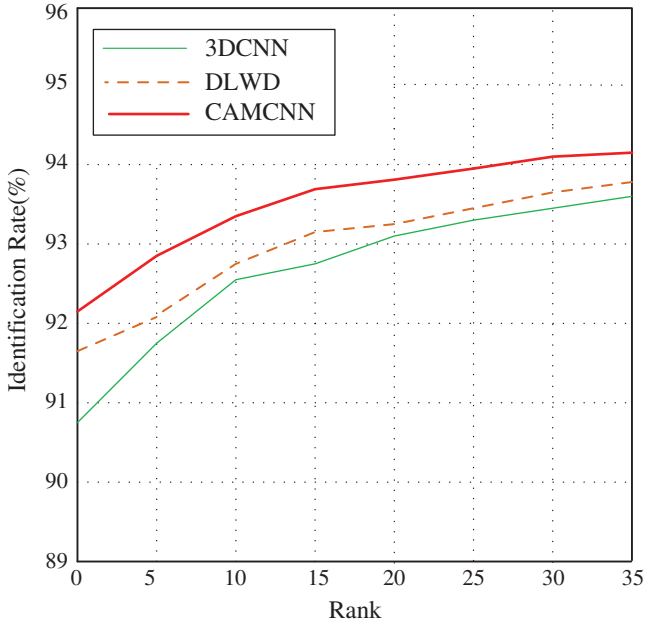

Figure 11: CMC curve of CASIA-B dataset under non-cooperative state

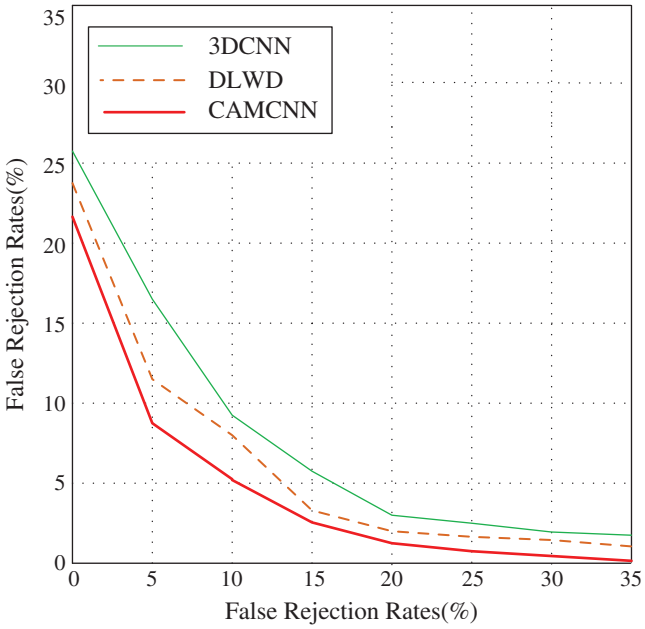

Figure 12: ROC curve of OULP dataset under non-cooperative state

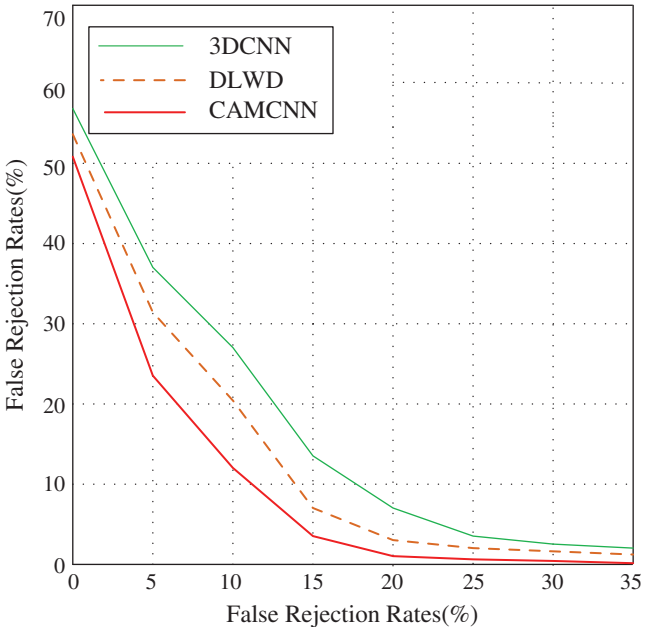

Figure 13: ROC curve of CASIA-B dataset under non-cooperative state

5 Conclusions and Future Works

In this paper, a new multi-view gait recognition algorithm combining channel attention mechanism and convolutional neural network is proposed. The method trains the convolution layer through back propagation learning and uses the root mean square propagation of stochastic gradient descent to optimize the cost function, which improves the learning ability of the convolutional neural network. In addition, the channel attention mechanism is integrated, which makes the network more focused on learning and

further improves the ability of expressing gait features. In addition, two sets of comprehensive experiments are discussed, which uses different feature extraction and classification methods for multi-view gait recognition. The experiment results show that multi-view gait recognition performs better under our proposed CAMCNN neural network training and the same feature extraction methods or classification algorithms.

In the future human body gait recognition for serialized signals, in addition to considering the angle factor, other variables (such as clothing, backpacks and walking speed) must be comprehensively considered. In addition, in the actual landing process of the algorithm, the integration of long-distance face features is considered to achieve multi-modal body identity verification and authentication. Based on this, let the algorithm land as quickly as possible in digital security, digital entertainment and other scenarios that require identity recognition.

Funding Statement: This work was supported by the Natural Science Foundation of China (No. 61902133), Fujian natural science foundation project (No. 2018J05106) Xiamen Collaborative Innovation projects of Produces study grinds (3502Z20173046).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Zhang, Y. D., Govindaraj, V. V., Tang, C., Zhu, W., Sun, J. (2019). High performance multiple sclerosis classification by data augmentation and AlexNet transfer learning model. *Journal of Medical Imaging and Health Informatics, 9(9),* 2012–2021. DOI 10.1166/jmihi.2019.2692.
2. Xia, K., Yin, H., Qian, P., Jiang, Y., Wang, S. (2019). Liver semantic segmentation algorithm based on improved deep adversarial networks in combination of weighted loss function on abdominal CT images. *IEEE Access, 7,* 96349–96358. DOI 10.1109/ACCESS.2019.2929270.
3. Wang, X., Zhang, J. (2020). Gait feature extraction and gait classification using two-branch CNN. *Multimedia Tools and Applications, 79(3),* 2917–2930. DOI 10.1007/s11042-019-08509-w.
4. Wang, S. H., Xie, S., Chen, X., Guttery, D. S., Tang, C. et al. (2019). Alcoholism identification based on an AlexNet transfer learning model. *Frontiers in Psychiatry, 10,* 205. DOI 10.3389/fpsyt.2019.00205.
5. Xu, C., Makihara, Y., Li, X., Yagi, Y., Lu, J. (2019). Speed-invariant gait recognition using single-support gait energy image. *Multimedia Tools and Applications, 78(18),* 26509–26536. DOI 10.1007/s11042-019-7712-3.
6. Anusha, R., Jaidhar, C. D. (2020). Human gait recognition based on histogram of oriented gradients and Haralick texture descriptor. *Multimedia Tools and Applications, 79(11–12),* 8213–8234. DOI 10.1007/s11042-019-08469-1.
7. Wang, S., Tang, C., Sun, J., Zhang, Y. (2019). Cerebral micro-bleeding detection based on densely connected neural network. *Frontiers in Neuroscience, 13,* 422. DOI 10.3389/fnins.2019.00422.
8. Zhang, Y., Wang, S., Sui, Y., Yang, M., Liu, B. et al. (2018). Multivariate approach for Alzheimer's disease detection using stationary wavelet entropy and predator-prey particle swarm optimization. *Journal of Alzheimer's Disease, 65(3),* 855–869. DOI 10.3233/JAD-170069.
9. Kang, C., Yu, X., Wang, S. H., Guttery, D., Pandey, H. et al. (2020). A heuristic neural network structure relying on fuzzy logic for images scoring. *IEEE Transactions on Fuzzy Systems*. DOI 10.1109/TFUZZ.2020.2966163.
10. Wang, S. H., Sun, J., Phillips, P., Zhao, G., Zhang, Y. D. (2018). Polarimetric synthetic aperture radar image segmentation by convolutional neural network using graphical processing units. *Journal of Real-Time Image Processing, 15(3),* 631–642. DOI 10.1007/s11554-017-0717-0.
11. Ghaeminia, M. H., Shokouhi, S. B. (2018). GSI: efficient spatio-temporal template for human gait recognition. *International Journal of Biometrics, 10(1),* 29–51. DOI 10.1504/IJBM.2018.090127.
12. Xu, W., Zhu, C., Wang, Z. (2018). Multiview max-margin subspace learning for cross-view gait recognition. *Pattern Recognition Letters, 107,* 75–82. DOI 10.1016/j.patrec.2017.10.033.
13. Shaikh, S. H., Saeed, K., Chaki, N. (2017). Partial silhouette-based gait recognition. *International Journal of Biometrics, 9(1),* 1–16. DOI 10.1504/IJBM.2017.10005047.

14. More, S. A., Deore, P. J. (2018). Gait recognition by cross wavelet transform and graph model. *IEEE/CAA Journal of Automatica Sinica, 5(3),* 718–726. DOI 10.1109/JAS.2018.7511081.
15. Verlekar, T. T., Soares, L. D., Correia, P. L. (2018). Gait recognition in the wild using shadow silhouettes. *Image and Vision Computing, 76,* 1–13. DOI 10.1016/j.imavis.2018.05.002.
16. Xing, W., Li, Y., Zhang, S. (2018). View-invariant gait recognition method by three-dimensional convolutional neural network. *Journal of Electronic Imaging, 27(1),* 013010. DOI 10.1117/1.JEI.27.1.013010.
17. Zhang, S., Zhang, L. (2018). Combining weighted adaptive CS-LBP and local linear discriminant projection for gait recognition. *Multimedia Tools and Applications, 77(10),* 12331–12347. DOI 10.1007/s11042-017-4884-6.
18. Wang, H., Fan, Y., Fang, B., Dai, S. (2018). Generalized linear discriminant analysis based on euclidean norm for gait recognition. *International Journal of Machine Learning and Cybernetics, 9(4),* 569–576. DOI 10.1007/s13042-016-0540-0.
19. Portillo-Portillo, J., Leyva, R., Sanchez, V., Sanchez-Perez, G., Perez-Meana, H. et al. (2018). A view-invariant gait recognition algorithm based on a joint-direct linear discriminant analysis. *Applied Intelligence, 48(5),* 1200–1217.
20. Chhatrala, R., Jadhav, D. (2017). Gait recognition based on curvelet transform and PCANet. *Pattern Recognition and Image Analysis, 27(3),* 525–531. DOI 10.1134/S1054661817030075.
21. Yu, S., Chen, H., Wang, Q., Shen, L., Huang, Y. (2017). Invariant feature extraction for gait recognition using only one uniform model. *Neurocomputing, 239,* 81–93. DOI 10.1016/j.neucom.2017.02.006.
22. Gadaleta, M., Rossi, M. (2018). IDNet: Smartphone-based gait recognition with convolutional neural networks. *Pattern Recognition, 74,* 25–37. DOI 10.1016/j.patcog.2017.09.005.
23. Rashwan, H. A., García, M. Á., Chambon, S., Puig, D. (2019). Gait representation and recognition from temporal co-occurrence of flow fields. *Machine Vision and Applications, 30(1),* 139–152. DOI 10.1007/s00138-018-0982-3.
24. Sun, J., Wang, Y., Li, J., Wan, W., Cheng, D. et al. (2018). View-invariant gait recognition based on kinect skeleton feature. *Multimedia Tools and Applications, 77(19),* 24909–24935. DOI 10.1007/s11042-018-5722-1.
25. Wu, H., Weng, J., Chen, X., Lu, W. (2018). Feedback weight convolutional neural network for gait recognition. *Journal of Visual Communication and Image Representation, 55,* 424–432. DOI 10.1016/j.jvcir.2018.06.019.
26. Wang, S. H., Zhang, Y., Li, Y. J., Jia, W. J., Liu, F. Y. et al. (2018). Single slice based detection for Alzheimer's disease via wavelet entropy and multilayer perceptron trained by biogeography-based optimization. *Multimedia Tools and Applications, 77(9),* 10393–10417. DOI 10.1007/s11042-016-4222-4.
27. Wang, S. H., Zhang, Y. D., Yang, M., Liu, B., Ramirez, J. et al. (2019). Unilateral sensorineural hearing loss identification based on double-density dual-tree complex wavelet transform and multinomial logistic regression. *Integrated Computer-Aided Engineering, 26(4),* 411–426. DOI 10.3233/ICA-190605.
28. Gale, T., Anderst, W. (2019). Asymmetry in healthy adult knee kinematics revealed through biplane radiography of the full gait cycle. *Journal of Orthopaedic Research, 37(3),* 609–614. DOI 10.1002/jor.24222.
29. Yin, B., Xu, Y., Huang, Y., Lu, Y., Liu, Z. (2017). Direction finding for wideband source signals via steered effective projection. *IEEE Sensors Journal, 18(2),* 741–751. DOI 10.1109/JSEN.2017.2773567.
30. Ramesh, V. P., Baskaran, P., Krishnamoorthy, A., Damodaran, D., Sadasivam, P. (2019). Back propagation neural network based big data analytics for a stock market challenge. *Communications in Statistics-Theory and Methods, 48(14),* 3622–3642. DOI 10.1080/03610926.2018.1478103.
31. Zhou, Y., Chen, N. (2019). The LAP under facility disruptions during early post-earthquake rescue using PSO-GA hybrid algorithm. *Fresenius Environmental Bulletin, 28(12A),* 9906–9914.
32. Bakkouri, I., Afdel, K. (2019). Multi-scale CNN based on region proposals for efficient breast abnormality recognition. *Multimedia Tools and Applications, 78(10),* 12939–12960. DOI 10.1007/s11042-018-6267-z.
33. Sang, H. F., Chen, Z. Z., He, D. K. (2020). Human Motion prediction based on attention mechanism. *Multimedia Tools and Applications, 79(9),* 5529–5544. DOI 10.1007/s11042-019-08269-7.
34. Ge, H., Yan, Z., Yu, W., Sun, L. (2019). An attention mechanism based convolutional LSTM network for video action recognition. *Multimedia Tools and Applications, 78(14),* 20533–20556. DOI 10.1007/s11042-019-7404-z.
35. Zhou, Y., Yu, H., Li, Z., Su, J., Liu, C. (2020). Robust optimization of a distribution network location-routing problem under carbon trading policies. *IEEE Access, 8,* 46288–46306. DOI 10.1109/ACCESS.2020.2979259.