



Research article

Facial feature point recognition method for human motion image using GNN

Qingwei Wang¹, Xiaolong Zhang² and Xiaofeng Li^{3,*}

¹ Harbin Huade University, Harbin 150025, China

² Northeast Forestry University, Harbin 150040, China

³ Department of Information Engineering, Heilongjiang International University, Harbin 150025, China

* **Correspondence:** Email: Lixiaofeng@hiu.net.cn.

Abstract: To address the problems of facial feature point recognition clarity and recognition efficiency in different human motion conditions, a facial feature point recognition method using Genetic Neural Network (GNN) algorithm was proposed. As the technical platform, we'll be using the Hikey960 development board. The optimized BP neural network algorithm is used to collect and classify human motion facial images, and the genetic algorithm is introduced into neural network algorithm to train human motion facial images. Combined with the improved GNN algorithm, the facial feature points are detected by the dynamic transplantation of facial feature points, and the detected facial feature points are transferred to the face alignment algorithm to realize facial feature point recognition. The results show that the efficiency and accuracy of facial feature point recognition in different human motion images are higher than 85% and the performance of anti-noise is good, the average recall rate is about 90% and the time-consuming is short. It shows that the proposed method has a certain reference value in the field of human motion image recognition.

Keywords: Genetic Neural Network; human motion; facial feature points; facial detection; BP neural network

1. Introduction

Accurate acquisition of effective facial feature points (including eyes, nose, etc.) is a prerequisite for correct facial recognition. Therefore, the more important process in facial feature point recognition

is the location of basic feature points. Among them, facial feature point recognition in human motion images is a type of facial feature point recognition. Facial landmark localization aims to detect the predefined points of human faces, and the topic has been rapidly improved with the recent development of neural network methods. However, it remains a challenging task when dealing with faces in unconstrained scenarios, especially with large pose variations. In this paper, a novel split-and-aggregate strategy is proposed for large-pose faces. By introducing an anchor-based design, our proposed approach can simplify the regression problem by splitting the search space. Moreover, aggregating the prediction results contributes to reducing uncertainty and improving the localization performance [1,2]. There are various forms of facial feature point recognition. In Literature [3], based on the geometric features of the face and the local descriptors, Three-dimension (3D) face recognition is performed, and the key points of the obtained facial features are matched with the descriptors on the face of the library. In addition, the covariance matrix descriptor is extracted to measure the matching degree of the face and improve the efficiency of the recognition algorithm. In Literature [4], the facial image Gaussian pyramid was established by multi-resolution analysis, and the Gabor feature spectrum of each layer of the image in the pyramid was extracted. Aiming at the problem that Gabor features lack a global description of facial image, the local and global feature information of facial image are captured to realize face classification and recognition, and improve facial image recognition performance in complex and changeable environments. In Literature [5], the RANSAC algorithm is used to extract facial image feature points to eliminate unstable matching points, and to perform accurate 3D face reconstruction, and use singular value decomposition singular value decomposition (SVD) to solve the coarse registration transformation matrix, which makes 3D face recognition less computation and higher real-time performance. In Literature [6], in order to improve the efficiency of face recognition, keep as much of the original face information as possible, use principal component analysis to eliminate the correlation and noise between image features, and further project transformation to reduce the data dimension, and its recognition rate is optimized. Literature [7] describes the variation of facial expressions, uses dynamic optical flow features to improve the recognition rate of facial expressions, and proposes a new method of facial expression recognition, which extends the traditional linear judgment analysis method, in JAFFE and CK facial expression database classifies and recognizes facial expressions.

However, the above methods have disadvantages such as long training time, slow convergence speed and easy trapping in local extreme points. Therefore, this paper proposed a method for facial feature point recognition in human motion images using on GNN. The genetic operator and face detection algorithm are combined and imported into the training process of GNN algorithm for many times as a face classifier, which has better anti-interference and makes up for the deficiencies of the neural network. Dynamic transplantation of face feature points is carried out with the local features of face feature points, and face feature points are repeatedly detected to improve the convergence speed of the whole process. The detected face feature points are transferred to face alignment to effectively improve the training speed and accuracy of GNN algorithm, and realizes high-quality recognition of face feature points in human motion images. The main contributions of this paper are as follows:

- 1) In many development platforms, Hikey960 development board is selected as hardware according to the actual image in this paper, which lays the basic conditions of hardware.
- 2) In human motion image collection, for the problem that the classification training speed limit is prone to errors, an error analysis was made to enhance the accuracy of image information collection.
- 3) Combining face detection algorithm and genetic operator, the GNN algorithm is improved to avoid

the problem of inaccurate target representation in GNN algorithm, and the conversion characteristics of GNN algorithm are strengthened, which lays a foundation for high precision recognition of face feature points in human motion images.

2. Related work

There are many intelligent research methods for human motion image facial feature point recognition. In Literature [8], a weakly supervised learning method is proposed, which can learn convolutional neural network (CNN) from unlabeled RGBD video with few annotations, more importantly, this paper introduces a new data set, Birmingham nuclear waste simulation data set, and evaluates the proposed method to solve this new industrial target recognition challenge. The weak supervision method has proved to be very effective in solving a new application of RGBD object detection and recognition; Literature [9] studied the neural network tracking control of under-actuated systems with unknown parameters, matching and mismatching disturbances, and proposed a new adaptive control scheme using multilayer neural networks, adaptive control and variable structure strategies. In order to cope with the uncertainty of approach error, unknown datum parameters, time-varying matching and mismatching external disturbances, new auxiliary control variables are designed to establish the controllability of the non-configurable subset of the under-actuated system. Through the design of an appropriate robust compensator, it effectively cancels the external interference of approximation errors and matching and mismatching. Literature [10] proposed a two-step single-trial classification method to identify the three movements of the left and right arms (fist, extension and elbow flexion), and distinguish the left and right arms by decoding event-related synchronization. The motion is characterized by cortical coherence, and the specific motion of the arm is recognized. Research shows that proposed method is effective for the classification of different types of single-arm motions. Literature [11] proposed an improved general spatial pattern feature extraction method. Firstly, for different objects, the Bhattacharyya distance method is used to select the best frequency band of each electrode. Then the optimal frequency band signal is decomposed into spatial patterns, and features that can describe the largest difference are extracted from the EEG data, and the classification effect is better.

However, the above methods all have the problems of low recognition efficiency, low recognition accuracy and high noise in feature point recognition. For this reason, this paper proposes to recognize facial feature points in human motion images based on GNN algorithm. The results show that the proposed method has high recognition efficiency and recognition accuracy, high signal-to-noise ratio and recall rate, and short time-consuming.

3. Recognition of facial feature points in human motion image

Through the selection and design of the software and hardware development platform, the improvement of the GNN design concept, and the matching and transplantation of facial feature points, the recognition of facial feature points in human motion images is completed.

3.1. Development platform design

The main goal of hardware selection is to shorten the development schedule, and carefully

evaluate the selected target platform. It involves the availability of convenient development environment and technical support. Moreover, the mobile SoC platform released by Huawei has the highest performance. Applying it to the Android system can achieve the goal of application development with strong practicality [12,13]. Therefore, the HiKey960 development board is used as the hardware and configured in the Android Open-Source Project Android (AOSP) hardware opensource system. In order to facilitate the initial debugging of the development, several special pins, such as the Joint Test Action Group (JTAG) interface, will be pulled out for the external debugging module [14].

USB camera is used as video input facility, and a driver is written to call this facility. By obtaining the relevant information about the universal serial bus camera video input drive, and using the opensource library drive facility. The driving process includes:

Step One: Obtain the libusb library, and there is no application program interface that restricts access to USB facilities;

Step Two: Obtain the libuvc library based on the libusb library, call the USB video facility, implement fine-grained control of the facility, and obtain the video stream;

Step Three: The compression library libjpeg-turbo, which obtains the facial image of the human motion in Android, is also a motion facial image codec, and processes the video stream information, and uses single instruction multiple data streams to encode and decode the human motion facial image.

Android 8.0 system is selected as the software application. Its advantage is that it is smooth and stable, and can efficiently collect and recognize facial feature points in human motion image.

3.2. Improved GNN

3.2.1 Human motion image acquisition based on optimized BP neural network

BP neural network belongs to the forward network, which is composed of input layer, middle layer and output layer. The middle layer includes many layers to facilitate the judgment of the interaction between each factor. And each layer is composed of several neurons, and each neuron of the two similar layers is connected by weight, and the connection strength between the two neurons is judged according to the size of the weight. Through the single-phase operation from the input layer to the middle layer and then to the output layer, the calculation process of the entire network is realized, and the learning algorithm is introduced in this process. The structure diagram of the BP neural network used in this paper is shown in Figure 1.

The output sigmoid function of the neurons in each layer of the BP neural network is expressed as:

$$A_{ij} = 1 / (1 + d^{-Z_j}) \quad Z_j = \sum_{I=0}^B C_{ij} T_i \quad (1)$$

where d is node interval, Z_j is node mapping function.

The node weight from the middle layer (input layer) to the output layer (or middle layer) is C_{ij} , T_i is the mapping interval, and the number of node input values is B . $B = 7$ and $B = 12$ refer to the number of nodes from the input layer to the middle layer and the number of nodes from the middle layer to the output layer. Since the number of nodes from the input layer to the middle layer cannot be accurately determined and the classification effect cannot be guaranteed, the revised node weight is

c_{ij} . The premise for the revised weight value as c_{ij} is that the original value of c_{ij} ($i = 1, 2, \dots, 7; j = 1 \dots \dots 12$) is not within the range. And it satisfies the information formula that the intermediate layer accepts the reverse transmission of the output layer:

$$C_{ij}(t+1) = C_{ij}(t) + \alpha \chi_j A_i + \theta [C_{ij}(t) - C_{ij}(t-1)] \quad (2)$$

where $C_{ij}(t)$ refers to the connection weight value at moment t from the layer's neutral node i (the middle layer or input layer) to the upper layer's neutral node j (the input layer or middle layer). The actual output value of neuron i in this layer at time t is A_i ; α , θ and χ_j refer to step size adjustment factor, smoothing factor and error weight adjustment factor, and $\alpha \in (0, 1)$, and $\theta \in (0, 1)$. The intermediate node and the output node are represented by Eqs (3) and (4), respectively.

$$A_{l0} = E_j(1, E_j) \sum_{L=0}^n \chi_l C_{lj} \quad (3)$$

$$A_{l0} = E_j(1, E_j) \sum_{L=0}^n \chi_l C_{lj} \quad (4)$$

where E_j and t_j represent the actual output value and output target value of the intermediate node j . Limited by the classification training speed, it is easy to produce output errors [15, 16]. Therefore, error analysis is performed. In the case that the intermediate layer accepts the information formula passed by the output layer backward, the relative error function is:

$$D_l = \frac{(A_{l0} - A_l)}{A_{l0}} \quad (5)$$

where A_{l0} and A_l refer to the network's actual output value and predicted output value. When there is no error in the calculation of the model, the error D_l shows less than the error tolerance of the network; conversely, if the error D_l exceeds the network error tolerance, return to the second step to adjust the weight and continue to calculate until the error tolerance requirement is met [17, 18].

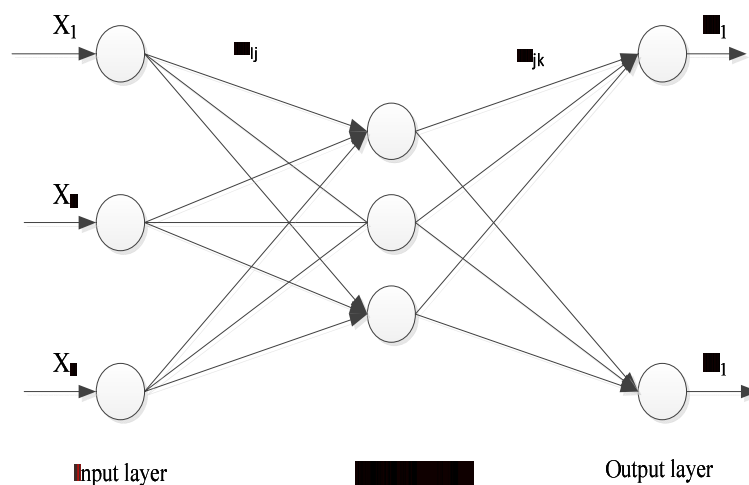


Figure 1. BP neural network structure.

3.2.2 Improve GNN with genetic algorithm

1) Fixed network structure

The genetic algorithm combines Mendel's survival of the fittest and Darwin's law of evolution and is completed by a cumbersome problem coding. Four steps such as genetic manipulation, selection, crossover and mutation are used for generational evolution, and finally the optimal solution and suboptimal solution of the problem are obtained. This study used real-valued coding to fix the network structure [19]. Real-valued coding not only has high calculation accuracy and simple calculation, it can realize the output of neural network through the form of coding. Equation (6) is the coding criterion.

$$B = [\log_2 A] + 1 \quad (6)$$

where A represents the node distribution function. In order to reduce the length of the chromosome of the genetic algorithm and simplify the calculation amount of the genetic operation [20], the coding problem in Eq (6) is solved after $[\log_2 A]$ is rounded and decoded. The fitness function needs to be satisfied during the rounding process:

$$f = \sum_{i=1}^{\text{popsize}} (\sum_{l=1}^{\text{amout}} (t_{il} - y_l)^2) \quad (7)$$

where t_{il} refers to non-negative complexity value, and y_l refers to continuous differentiable conditions. Complete decoding and simplified calculations.

2) Improve GNN network structure with genetic operator

Combined with the genetic algorithm after real value coding, the GNN algorithm is improved to recognize the facial feature points of human motion image.

Firstly, the human motion image collected by the optimized BP neural network is preprocessed to form the morphological filtering of human motion face image:

$$X' = \left(X \left\langle \frac{1}{2} \right\rangle G \right) \quad (8)$$

where X' represents the morphological filter value. X represents the original image shape value. G is a square lattice.

After morphological filtering, in order to ensure the quality of human motion image and improve the image SNR, judge whether the current image pixel is impact noise pixel, as shown in Eq (9).

$$N_{(i,j)} = \begin{cases} 1 & \text{if } x_{(i,j)} > T \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

Where $x_{(i,j)}$ represents the image pixel value, i and j are pixel points, and T represents the threshold. If the judgment result of the impact noise pixel is 1, the pixel is a noise pixel.

The variable threshold of image morphological filtering is binarized, and the eye position in human motion face image is determined by intersection and separation transform, and the specific morphological unit is extracted from human motion face image. The equation is:

$$X' = X \langle * \rangle \kappa T, \kappa \in Z \quad (10)$$

where κ represents the threshold of a specific shape. T represents image morphological space. Z

represents the maximum threshold. The eyeball is the closest circle in the human motion face image. Therefore, in this formula, the circular structural unit is G . Then the unit excluding G in the square range of G is G^* .

The largest point H obtained from the detected morphological transformation is regarded as the judgment point and eyeball point.

3) Train GNN network to clarify the distribution area of each organ

According to the GNN logic, it is determined that the point of the eyeball is clear, because the value of the longitudinal coordinate of the two eyeballs in the front view of the face is small.

Set the distance of the eyeball is r , (x_1, y_1) and (x_2, y_2) are the coordinates of the left and right eyes, the following clear organ regions are:

Left eye: $(x_1 - s_1, y_1 - s_2, x_1 + s_3, y_1 + s_4)$

Right eye: $(x_2 - s_1, y_2 - s_2, x_2 + s_3, y_2 + s_4)$

Taking the displacement: $s_1 = s_3 = 0.4r$, $s_2 = s_4 = 0.5r$

Nose: $\left(\frac{x_1+x_2}{2} - u_1, \frac{y_1+y_2}{2} + h_1r - u_2, \frac{x_1+x_2}{2} + u_3, \frac{y_1+y_2}{2} + h_1r + u_4\right)$

Scale factor: $h_1 = 0.9$

Take the displacement: $v_1 = v_2 = 0.5r, u_2 = u_4 = 0.5r$

Mouth: $\left(\frac{x_1+x_2}{2} - v_1, \frac{y_1+y_2}{2} + h_2r - v_2, \frac{x_1+x_2}{2} + v_3, \frac{y_1+y_2}{2} + h_2r + v_4\right)$

Scale factor: $h_2 = 1.5$

Take the displacement: $u_1 = u_3 = 0.7r, v_2 = v_4 = 0.4r$

4) Identify feature points based on the area point projection map

There are two methods to clarify the location of feature points:

i) Project the human movement face image according to the area points to clarify the location of each organ area. If the projected human motion facial image of this area does not achieve the desired effect, you can adjust the area position and the size conforms to the features of the point projection of the specified area.

ii) The position of feature points is determined by projecting the human motion face image according to the clear regional points.

3.3. Recognition of human facial feature points

The obtained human motion facial image is transferred to the face alignment method to detect the facial feature points. The detailed process is shown in Figure 2.

The face detection method process, as shown in Figure 2, is written using the computer vision library OpenCV, CAFFE, and the programming language Python. This effectively detects the face's location as well as the five feature points. The source code of the face detection algorithm is compiled using MATLAB and CAFFE, and the finished face detection program is compiled into the dynamic link library of the Android application using NDK and CAFFE; face alignment is realized using the regression tree algorithm, and a cascade regression tree is built to restore the actual shape of the face. Gradient Boosting Decision Tree (GBDT) is used in the alignment process, and each GBDT tree is serially connected. A Dlib library, which is an opensource library, is obtained on the basis of building facial recognition programs on multiple platforms. This opensource library's transplantation process is divided into four parts:

- 1) The operation of facial image and files for human motion is realized by compile code, such as human motion facial image format conversion and other operations, and assign them to the project for calling.
- 2) The human motion facial image information obtained by the face detection algorithm is passed to the face alignment method to complete the face alignment. The NDK command combined with the computer vision library is used to compile the face alignment information into a dynamic link library.
- 3) The functions such as calling the dynamic link library and loading the trained model are applied to the main project.
- 4) Rewrite configuration files such as build.gradle, use the gradlew command set to complete the creation of all projects, and generate an Android installation package, use the adb command set to assemble it on the development board, and pass the verification. The program can accurately and effectively obtain the position information of the facial image feature points of 50 human motions, and can display the detection results on the display.

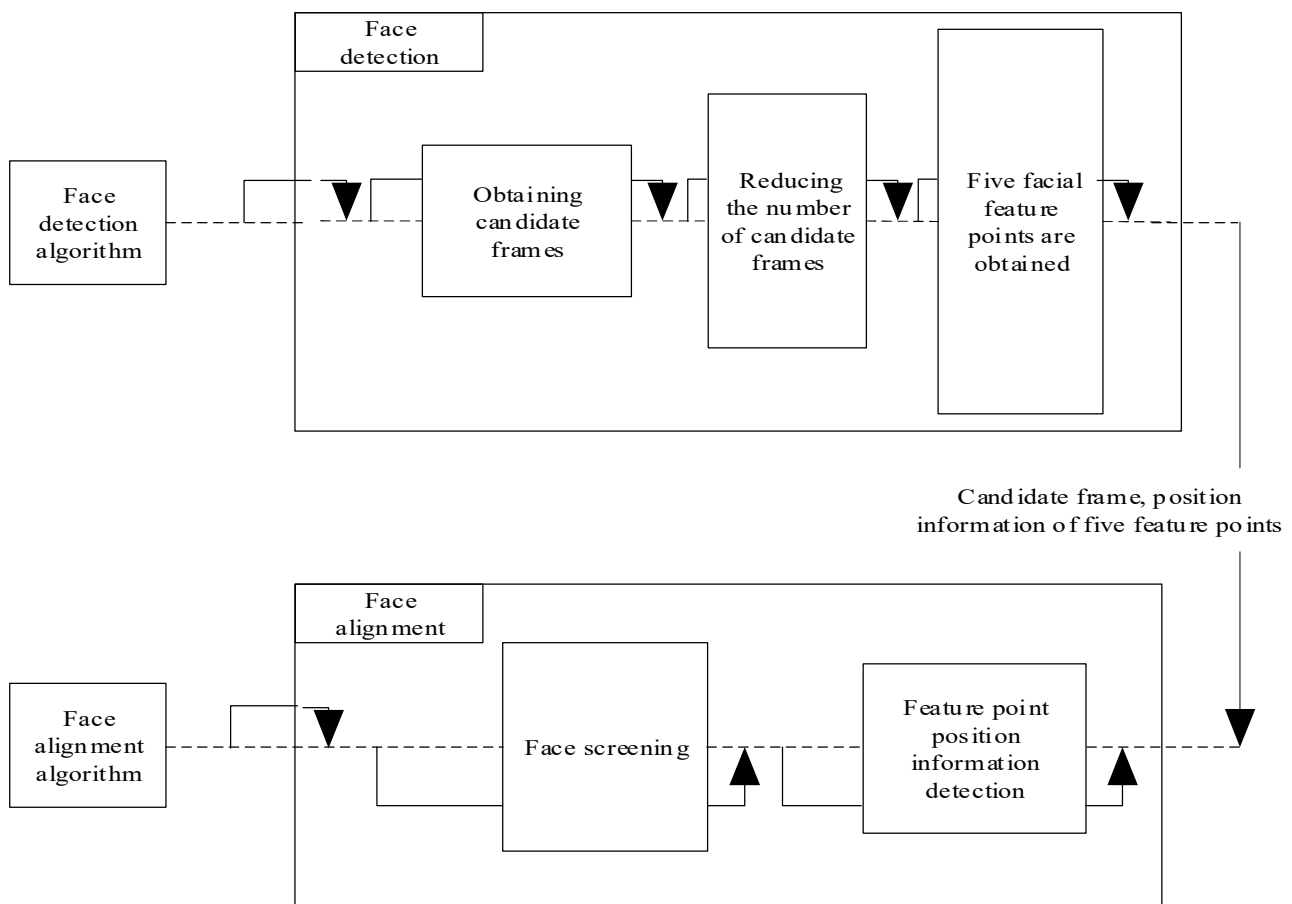


Figure 2. Facial alignment method process.

According to the above process, the clear distribution area of the eyes, nose and mouth of each organ completes the realization of the facial feature point detection algorithm. Complete the effective recognition of facial feature points in human motion images.

3.4. The proposed method

In this study, genetic operators are used to improve network structure, train in the GNN network, and meet specific error requirements, and implement accurate GNN optimization. The optimization process is shown in Figure 3.

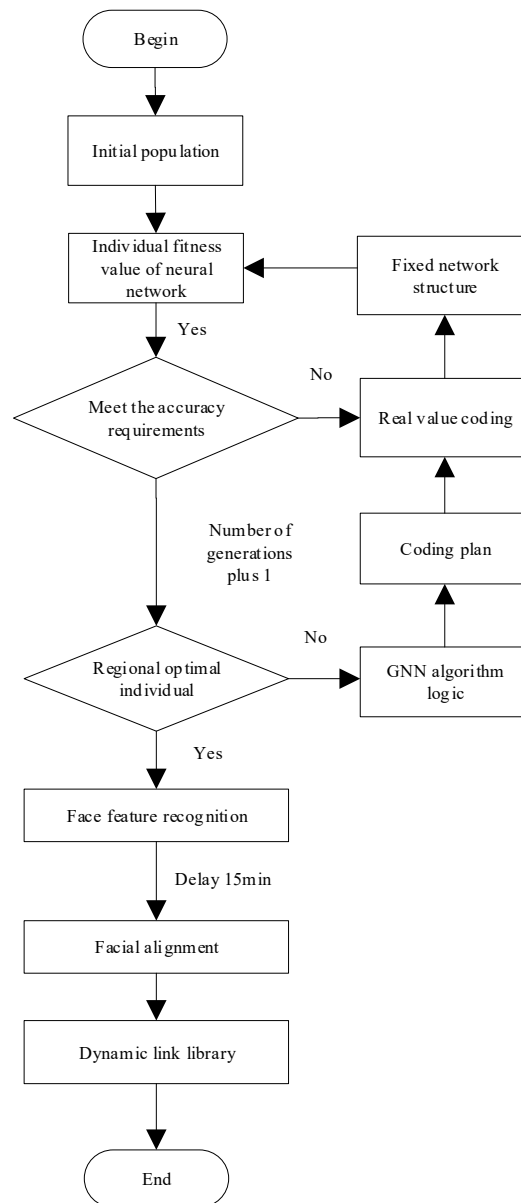


Figure 3. Facial feature point recognition method for human motion image.

- 1) Define the coding plan and generate a set of initial populations;
- 2) Create a neural network structure by the initial population;
- 3) Calculate the fitness value of each individual through the designed fitness function;
- 4) Through the size of fitness, the implementation of genetic operations is completed;
- 5) Calculate the individual fitness value, if the accuracy requirements are met, terminate the calculation, if not, continue the genetic operation according to step 6);

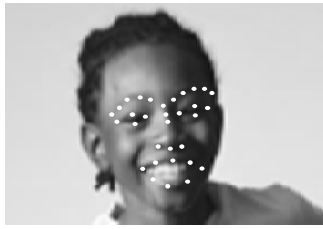
- 6) Add 1 to the number of generations. If the maximum algebra of the designed evolutionary algebra is satisfied, the genetic operation will be terminated. If not, skip to step 2), and use real-valued coding to avoid interference with the design process through the coding plan;
- 7) According to the GNN logic, the optimal individual in the recognition area is obtained, and the GNN network training is performed according to the decryption result of the genetic algorithm;
- 8) The improved GNN is used to perform precise optimization on the network and clarify the location of facial feature points;
- 9) Realize face alignment algorithm;
- 10) Write the USB camera video input driver in turn;
- 11) The realized face detection algorithm is written into a dynamic link library, and the running process is completed through the Android platform.

4. Experimental analysis and results

4.1. Data set

The operating system used in this experiment is window10, which runs the algorithm based on Caffe framework and python. In order to verify the improvement effect of the proposed method, AFW data set [21] and WFLW data set [22] are selected as data sources. AFW data set: this dataset is mainly applicable to face image recognition, including 473 face markers. Each face image is set as a rectangular bounding box style, and each image contains 6 landmark information. WFLW data set: this data set selects 98 key points, takes 10,000 faces, has a large number of images and diverse environments, including occlusion, illumination, expression and other attribute information. The face images of running, playing table tennis and playing basketball are randomly selected from the above two data sets to form two data sets, which are detected by the face detector to form countless overlapping blocks. Each data set contains 30,000 data and 11,000 videos, and 78 typical data points are marked, which are converted into about 200,000 images. Each image is annotated with 68 labels. In this experiment, 200,000 images are selected for data training, and the remaining half are used for experimental test and analysis, which are tested under 40,000, 60,000 and 100,000 data respectively.

After the face image is obtained by setting the boundary box of the face detector, the specific position of the face is analyzed according to the input coordinate data and the face detection data, the specific position of the feature points is determined, and the training test image is checked repeatedly with the feature point pixel data. The internal parameter matrix is obtained by chessboard calibration method, and then the existing model is used to identify facial feature points. According to the coordinates of the detected feature points, the homography matrix of the plane on the positive plane between the corners of eyes and mouth is calculated, and then the face angle of each image is solved combined with the constraint of feature points. According to the average distance between the normalized predicted coordinates and the real coordinates, the face angle is calibrated to ensure the practicability and applicability of each image. The partial sample collection results are shown in Figure 4.



(a) Human facial features when running



(b) Human facial features when playing table tennis



(c) Human facial features when playing basketball

Figure 4. Facial feature sample images under different motions.

The proposed method is compared with [3,8,9] and CNN methods for facial feature point recognition. Figure 4 is analyzed to verify the application effect of the proposed method.

4.2. Evaluation criteria

Facial recognition point mining: In the case of two kinds of noise, the proposed method verifies the mining results of face recognition points of different types of human motion images.

SNR: In the experiment, 20% salt and pepper noise and 50% salt and pepper noise are added to verify the SNR of the proposed method. The calculation formula of the SNR is:

$$SNR=10 \cdot \lg^{(P_s/P_n)} \quad (11)$$

where, p_s and p_n refer to the effective power of the signal and noise, which can also be converted into the ratio of voltage amplitude.

Recognition accuracy/recognition efficiency: when 20% salt and pepper noise and 50% salt and pepper noise are added, the recognition accuracy and recognition efficiency of the proposed method are verified. The verification equation is:

Recognition accuracy:

$$K = \left(1 - \frac{h}{T_{S+1}}\right) \cdot 100\% \quad (12)$$

Recognition efficiency:

$$C = \frac{1}{T} [v(t) - 1] \cdot 100\% \quad (13)$$

where T_s is total recognition feature points, h is the number of facial feature points extracted, T is the adjustment coefficient, and $v(t)$ is the efficiency function.

In the process of recognition accuracy analysis, since each image is annotated with 68 labels. Define the feature points according to the regional point projection to ensure that the face feature points optimized by the adjustment coefficient correspond to the extracted feature points one by one. It is transferred to the dynamic link library after the application of face alignment algorithm to complete face alignment and ensure that h and T_s are recognizable feature points with the same location information.

Recognition time-consuming: Time-consuming is an important index for judging the performance of the method. The recognition time-consuming of proposed method, [3,8,9] and CNN method is compared.

Recall rate: In order to further verify the effectiveness of the proposed method, the recall rate is used for verification and analysis.

4.3. Results and discussion

Figure 5 is the comparison structure of face recognition point mining effect when different methods recognize different motion images.

According to Figure 5, when other literature methods recognize human facial feature points in different motion images, the other methods recognize the features that are rough in the missing part, resulting in poor pixel mining effect.

The CNN method is under 29 recognition points, the [3] method is under 27 recognition points, and the [8,9] have fewer methods, and it does not reach 24 recognition points. The proposed method has a high degree of recognition of feature points, which is more than 32, which can meet the needs of facial feature recognition in motion.

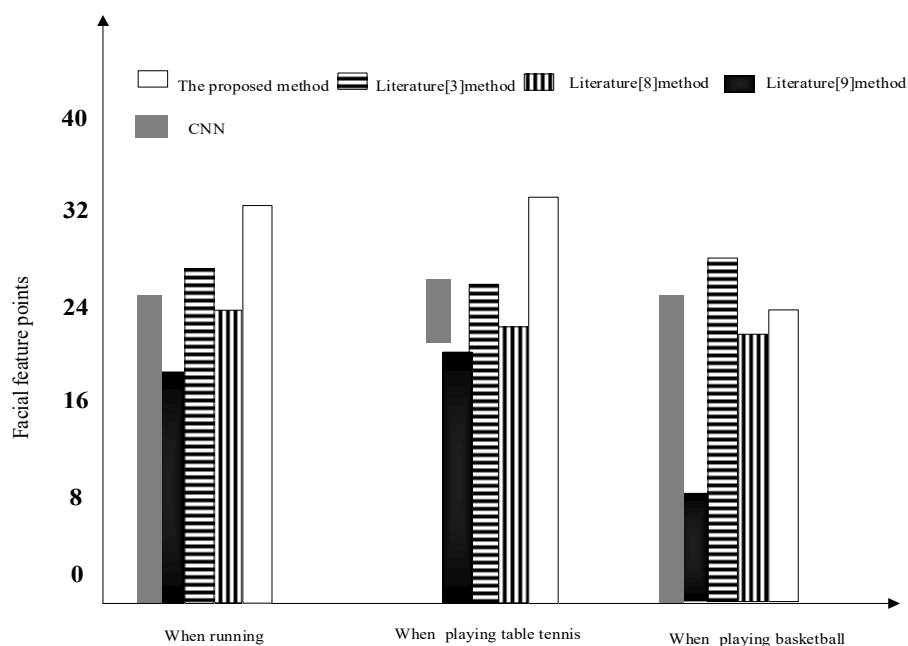


Figure 5. Face recognition point mining results of different motion recognition methods.

After adding 20 and 50% salt and pepper noise, the comparison results of the SNR, recognition accuracy and recognition efficiency of the four methods are shown in Tables 1 and 2. In order to improve the quality of experimental data, the data in Tables 1 and 2 are the average values obtained from multiple experiments.

According to Tables 1 and 2, for different human motion images, when 20% salt and pepper noise is added, the proposed method is higher than 85% recognition efficiency and accuracy of facial feature points in human motion images. when 50% salt and pepper noise is added, the proposed method is still higher than 85% recognition efficiency and accuracy of facial feature points in human motion images, which is not affected by environmental noise. However, the methods of [3,8,9] have obvious differences under the two salt and pepper noises, and the overall performance of 20% salt and pepper noise is relatively better.

According to statistics, under 20% salt and pepper noise, the recognition efficiency and accuracy of the method in [3] is no more than 50%. The recognition efficiency and accuracy of the method in [8] is no more than 60%. The recognition efficiency and accuracy of the [9] method does not exceed 70 and 50%, and the CNN method does not exceed 60%, which are far lower than the proposed method. In summary, when adding different salt and pepper noises, the test results of proposed method have obvious advantages in SNR, recognition accuracy and recognition efficiency. This is because the proposed method combines the face detection algorithm with genetic operators to synchronize and improve the GNN algorithm. The dynamic transplantation results of face feature points are combined with the projection results of regional points to minimize the impact of face corner or noise on the recognition results when ensuring human motion or in the presence of noise. That is to improve the matching degree of feature points, and then improve the recognition accuracy and efficiency.

Table 1. Comparison results of SNR, recognition accuracy and recognition efficiency of four methods under 20% salt and pepper noise.

| Methods | Running | | | Table Tennis | | | Basketball | | |
|-------------------|--------------------------------------|-----------------------------------|------------|-------------------------------------|-----------------------------------|------------|-------------------------------------|-----------------------------------|------------|
| | Recog nition efficie ncy /% | Recogniti on accuracy /% | SNR/d B | Recogniti on efficiency /% | Recogniti on accuracy /% | SNR/d B | Recogniti on efficiency /% | Recogniti on accuracy /% | SNR/d B |
| The proposed | 90 | 85 | 35 | 85 | 90 | 36 | 90 | 90 | 36 |
| Literature [3] | 50 | 40 | 26 | 20 | 40 | 24 | 30 | 50 | 24 |
| Literature [8] | 40 | 50 | 25 | 60 | 60 | 27 | 30 | 60 | 23 |
| Literature [9] | 40 | 40 | 27 | 10 | 50 | 25 | 70 | 50 | 26 |
| CNN | 40 | 50 | 26 | 60 | 60 | 27 | 50 | 60 | 25 |

According to the SNR, it can be known that under 20% salt and pepper noise and 50% salt and pepper noise, the SNR range of the proposed method is between [33dB, 36dB]. The method in [3] is

between [20dB, 26dB]. The methods of [8,9] are between [20dB, 27dB]. The CNN method is between [21dB, 27dB], and the four comparison methods do not exceed 27dB, which is significantly lower than the proposed method.

In view of the result that the image signal noise of the proposed method is higher than that of the traditional method, the main reason is that this paper uses the double compound morphological filter to preprocess the human motion images collected by the optimized BP neural network, and in order to ensure the quality of the human motion images, Improve image SNR. Whether the image pixels are impact noise pixels is further judged, which provides a basis for improving the SNR of human motion images. However, the methods in [3,8,9] are not enough to deal with the image noise, which is lower than the image SNR of the proposed method.

The time-consuming comparison results of different methods are shown in Table 3.

Table 2. Comparison results of SNR, recognition accuracy and recognition efficiency of four methods under 50% salt and pepper noise.

| Methods | Running | | | Table Tennis | | | Basketball | | |
|----------------|---------------------------|-------------------------|---------|---------------------------|-------------------------|---------|---------------------------|-------------------------|---------|
| | Recognition efficiency /% | Recognition accuracy /% | SNR /dB | Recognition efficiency /% | Recognition accuracy /% | SNR /dB | Recognition efficiency /% | Recognition accuracy /% | SNR /dB |
| The proposed | 90 | 85 | 33 | 85 | 90 | 34 | 90 | 85 | 35 |
| Literature [3] | 50 | 40 | 20 | 20 | 20 | 20 | 30 | 40 | 20 |
| Literature [8] | 20 | 30 | 25 | 50 | 60 | 27 | 30 | 70 | 20 |
| Literature [9] | 30 | 40 | 26 | 10 | 50 | 20 | 60 | 50 | 26 |
| CNN | 40 | 50 | 21 | 60 | 50 | 25 | 50 | 50 | 23 |

Table 3. Comparison of time-consuming identification of different methods (s).

| Data size/piece | 40,000 | 60,000 | 100,000 |
|-----------------------|--------|--------|---------|
| The proposed method | 5 | 6 | 7 |
| Literature [3] method | 13 | 20 | 23 |
| Literature [8] method | 20 | 26 | 28 |
| Literature [9] method | 25 | 29 | 30 |
| CNN | 15 | 16 | 17 |

According to Table 3, it can be seen that the identification time of proposed method is significantly lower than other methods. Under different data size conditions, the identification of proposed method does not exceed 7 s. In other methods. The recognition in [9] takes a long time. When the amount of data is 100,000, the recognition time is up to 30 s, the maximum recognition time in [3] is 23 s, and

the maximum recognition time in [8] is 28 s, which is multiple higher than the method in this paper. Therefore, we can show the advantages of the proposed method. The comparison result of the recall rate of different methods is shown in Figure 6.

According to Figure 6, it can be seen that the recall curve of the proposed method changes around 90%, recall rate is high. While the recall curve of other methods is lower than the proposed method, and there are significant differences. [3,8] have a large change in the recall rate curve, the highest is about 80%, the lowest is only about 50%, and the stability is poor. The CNN method varies from 60 to 75%, the overall recall rate of the method in the [9] is the lowest, less than 70%. It can be seen that the proposed method has better recognition.

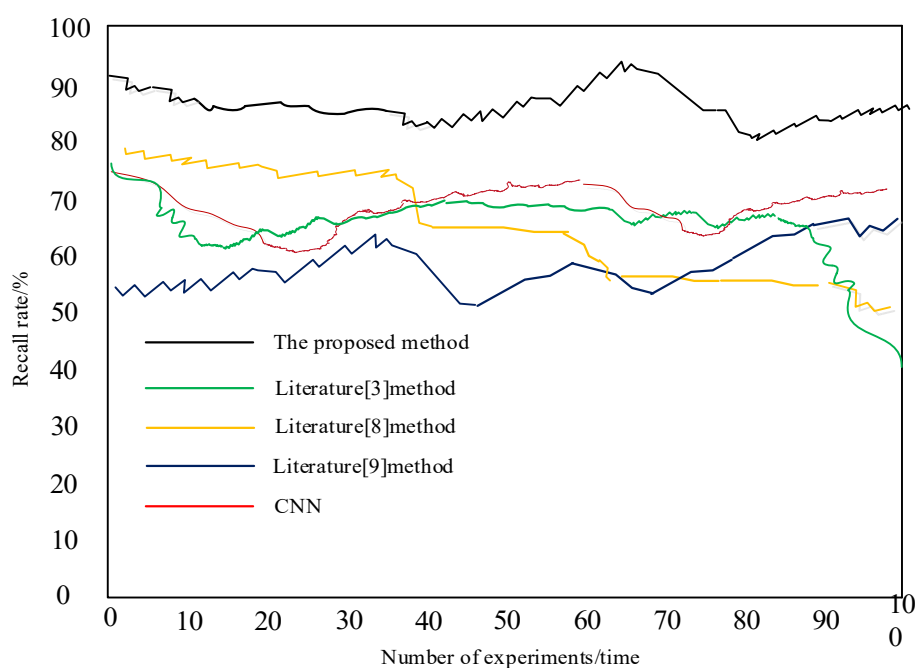


Figure 6. Comparison of recall rates.

5. Conclusions and future works

In the paper, we proposed facial feature point recognition method of human motion image using GNN, based on the HiKey960 development board, and MATLAB and CAFFE are used to write the source code of the facial detection algorithm. Through the transplantation of facial feature point recognition algorithm, facial detection and face alignment, the recognition of facial feature points in human motion images is realized. Experiments show that the recognition efficiency and recognition accuracy of face feature points in different human motion images are high, and the signal-to-noise ratio and recall rate are high, and the time-consuming is short. The proposed method can effectively recognize the facial feature points of different human motion images accurately and stably, and proposed method can be applied to intelligent transportation aerial photography. In future works, we need to increase investment in experimental platform and experimental data to provide theoretical reference for the research of neural network and related fields.

Acknowledgments

This work was supported by Natural Science Foundation of Heilongjiang Province of China under grant number LH2021F040.

Conflict of interest

The authors declare that they have no conflicts of interest.

References

1. Z. Xu, B. Li, M. Geng, Y. Yuan, AnchorFace: An anchor-based facial landmark detector across large poses, preprint, arXiv:2007.03221.
2. P. Gao, K. Lu, J. Xue, J. Lyu, L. Shao, A facial landmark detection method based on deep knowledge transfer, *IEEE Trans. Neural Networks Learn. Syst.*, 2021. <https://doi.org/10.1109/TNNLS.2021.3105247>
3. B. Guo, F. Da, Expression-invariant 3D face recognition based on local descriptors, *J. Comput. - Aided Des. Comput. Graphics*, **31** (2019), 1086–1094. <https://doi.org/10.3724/SP.J.1089.2019.17433>
4. D. Wu, X. Jing, L. Zhang, W. Wang, Face recognition with Gabor feature based on Laplacian Pyramid, *J. Comput. Appl.*, **2** (2017), 63–66.
5. Y. Guo, E. She, Q. Wang, Z. Li, Face point cloud registration based on improved surf algorithm, *Opt. Technol.*, **44** (2018), 333–338. <https://doi.org/10.13741/j.cnki.11-1879/o4.2018.03.014>
6. J. Xu, Z. Wu, Y. Xu, J. Zeng, Face recognition based on PCA, LDA and SVM, *Comput. Eng. Appl.*, **55** (2019), 34–37. <https://doi.org/10.3778/j.issn.1002-8331.1903-0286>
7. T. Liu, X. Zhou, X. Yan, LDA facial expression recognition algorithm combining optical flow characteristics with Gaussian, *Comput. Sci.*, **45** (2018), 286–290.
8. L. Sun, C. Zhao, Z. Yan, P. Liu, T. Duckett, R. Stolkin, A novel weakly-supervised approach for RGB-D-based nuclear waste object detection, *IEEE Sens. J.*, **19** (2019), 3487–3500. <https://doi.org/10.1109/JSEN.2018.2888815>
9. P. Liu, H. Yu, S. Cang, Adaptive neural network tracking control for underactuated systems with matched and mismatched disturbances, *Nonlinear Dyn.*, **98** (2019), 1447–1464. <https://doi.org/10.1007/s11071-019-05170-8>
10. Z. Tang, H. Yu, C. Lu, P. Liu, X. Jin, Single-trial classification of different movements on one arm based on ERD/ERS and corticomuscular coherence, *IEEE Access*, **7** (2019), 128185–128197. <https://doi.org/10.1109/ACCESS.2019.2940034>
11. Z. Tang, C. Li, J. Wu, P. Liu, S. Cheng, Classification of EEG-based single-trial motor imagery tasks using a B-CSP method for BCI, *Front. Inf. Technol. Electronic Eng.*, **20** (2019), 1087–1098. <https://doi.org/10.1631/FITEE.1800083>
12. H. Xiong, C. Jin, M. Alazab, K. Yeh, H. Wang, T. R. R. Gadekallu, et al., On the design of blockchain-based ECDSA with fault-tolerant batch verification protocol for blockchain-enabled IoMT, *IEEE J. Biomed. Health Inf.*, 2021. <https://doi.org/10.1109/JBHI.2021.3112693>
13. W. Wang, C. Qiu, Z. Yin, G. Srivastava, T. R. R. Gadekallu, F. Alsolami, et al., Blockchain and PUF-based lightweight authentication protocol for wireless medical sensor networks, *IEEE Internet Things J.*, 2021. <https://doi.org/10.1109/JIOT.2021.3117762>

14. Z. Xia, J. Xing, C. Wang, X. Li, Gesture recognition algorithm of human motion target based on deep neural network, *Mobile Inf. Syst.*, **2021** (2021),1–12. <https://doi.org/10.1155/2021/2621691>
15. G. Sang, Y. Chao, R. Zhu, Expression-insensitive three-dimensional face recognition algorithm based on multi-region fusion, *J. Comput. Appl.*, **39** (2019), 1685–1689. <https://doi.org/10.11772/j.issn.1001-9081.2018112301>
16. X. Zhou, J. Zhou, R. Xu, New algorithm for face recognition based on the combination of multi-sample conventional collaborative and inverse linear regression, *J. Electron. Meas. Instrum.*, **32** (2018), 96–101. <https://doi.org/10.13382/j.jemi.2018.06.014>
17. F. Wang, Y. Zhang, D. Zhang, H. Shao, C. Cheng, Research on application of convolutional neural networks in face recognition based on shortcut connection, *J. Electron. Meas. Instrum.*, **32** (2018), 80–86. <https://doi.org/10.13382/j.jemi.2018.04.012>
18. X. Ma, X. Li, Dynamic gesture contour feature extraction method using residual network transfer learning, *Wireless Commun. Mobile Comput.*, **2021** (2021). <https://doi.org/10.1155/2021/1503325>
19. Y. Kim, K. Lee, A novel approach to predict ingress/egress discomfort based on human motion and biomechanical analysis, *Appl. Ergon.*, **75** (2019), 263–271. <https://doi.org/10.1016/j.apergo.2018.11.003>
20. L. Wang, Z. Ding, Y. Fu, Low-rank transfer human motion segmentation, *IEEE Trans. Image Process.*, **28** (2019),1023–1034. <https://doi.org/10.1109/TIP.2018.2870945>
21. M. Kostinger, P. Wohlhart, P. M. Roth, H. Bischof, Annotated facial landmarks in the wild: A largescale, real-world database for facial landmark localization, in *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, (2011), 2144–2151. <https://doi.org/10.1109/ICCVW.2011.6130513>
22. W. Wu, C. Qian, S. Yang, Q. Wang, Y. Cai, Q. Zhou, Look at boundary: A boundary-aware face alignment algorithm, in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2018), 2129–2138. <https://doi.org/10.1109/CVPR.2018.00227>



AIMS Press

©2022 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)