Received March 2, 2020, accepted March 16, 2020, date of publication March 23, 2020, date of current version April 15, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2982779

Face Detection in Security Monitoring Based on Artificial Intelligence Video Retrieval Technology

ZUOLIN DONG[1], JIAHONG WEI[2], XIAOYU CHEN[1], AND PENGFEI ZHENG[3]

[1]College of Electronic and Information Engineering, Henan Institute of Technology, Xinxiang 453000, China
[2]College of Electrical Engineering and Automation, Henan Institute of Technology, Xinxiang 453000, China
[3]College of Electronic and Electrical Engineering, Henan Normal University, Xinxiang 453007, China

Corresponding author: Jiahong Wei (jhwei1213@163.com)

This work was supported in part by the Key Scientific Research Project of Henan Universities in 2020 Research and Design of Rocker Submerged Arc Welding Workstation Based on PLC Project under Grant 20B470002, in part by the Ph.D. Scientific Research Initial Foundation of Henan Institute of Technology, and in part by the Research on Online Monitoring and Damage Identification of Structures Based on Internet of Things under Grant KQ1911.

ABSTRACT With the rapid development of video monitoring, the massive information of the monitoring image has far exceeded the effective processing range of human resources. Intelligent video retrieval technology has become an increasingly indispensable part of video monitoring system. Intelligent video retrieval technology integrates video processing, computer vision and artificial intelligence, which greatly improves the efficiency of monitoring and the accuracy and linkage of monitoring system. Face recognition and other emerging technologies continue to rise and apply to the security monitoring system. Based on deep learning theory and face detection neural network, this paper proposes a video oriented cascaded intelligent face detection algorithm, which builds deep learning network by cascading multiple features, from edge features, contour features, local features to semantic features, and advances layer by layer. According to the last semantic features, the information of the input data is obtained to accurately realize the face detection under the non ideal condition. Simulation results show that the algorithm has good detection performance for single face and multi face images, and has strong robustness for rotating face. At the same time, the algorithm is fast and can basically meet the requirements of real-time face detection.

INDEX TERMS Artificial intelligence, video retrieval, security monitoring, face detection.

I. INTRODUCTION

In recent years, with the rapid development of artificial intelligence, deep learning and other technologies, face recognition also ushered in the outbreak period [1]. Face recognition technology has been an important subject in the field of computer vision for a long time, and it was mainly used in the field of public security in the early stage. With the popularity of face attendance, face passing and other applications in recent years, face recognition is widely used in intelligent transportation, intelligent medical treatment, building intercom, financial education, safe city construction and other fields; it can also unlock mobile phones and log in payment accounts by face. It can be seen that face recognition has been widely used in our life. When artificial intelligence ushers in the explosive year, massive data becomes an important material for artificial intelligence learning, and security becomes one of the huge application fields of artificial intelligence [2]–[6]. Because of its intuitive, accurate, timely and rich content, it is widely used in many occasions, but the huge video data brings the difficulty of video analysis which is hard to overcome by manpower. With the rapid development of video monitoring, the massive information of the monitoring image has far exceeded the effective processing range of human resources [7]–[9]. Intelligent video retrieval technology has become an increasingly indispensable part of video monitoring system.

The international anti-terrorism situation is becoming more and more serious, and the social demand for security is even stronger. Airport, customs, border defense and other places, which are necessary for entry and exit due to their wide distribution and large flow of people, have always been the areas with frequent emergencies, which determines the requirements of high security level control at checkpoints [10]–[14]. However, at present, the security system of each checkpoint is still in the stage of traditional monitoring, which has been difficult to meet the actual demand for security. Due

The associate editor coordinating the review of this manuscript and approving it for publication was Honghao Gao.

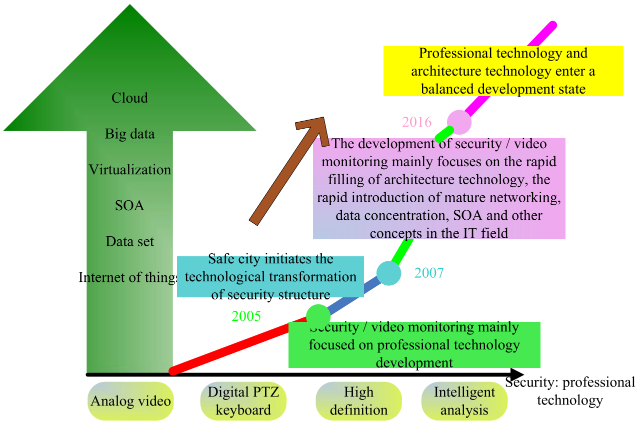

Cloud
Big data
Virtualization
SOA
Data set
Internet of things
Professional technology and architecture technology enter a balanced development state
2016
The development of security / video monitoring mainly focuses on the rapid filling of architecture technology, the rapid introduction of mature networking, data concentration, SOA and other concepts in the IT field
Safe city initiates the technological transformation of security structure
2007
2005
Security / video monitoring mainly focused on professional technology development
Security: professional technology
Analog video
Digital PTZ keyboard
High definition
Intelligent analysis

FIGURE 1. **Integration of security monitoring technology and artificial intelligence technology.**

to the lack of intelligent video analysis, it is still in a passive state that can only "monitor" but not "control", most of which only have the function of obtaining evidence after the event, and can not play a role of prevention and early warning for suspicious and abnormal behaviors [15], [16]. Using face recognition, we can take the initiative to extract and record the faces in the surveillance video in real time, to give an alarm in time and use the stored information to effectively retrieve the video data, to effectively assist the security personnel in dealing with the crisis, to minimize the false alarm and missing alarm, and to effectively help the checkpoint deal with the attack and emergencies [17]. The appearance and maturity of face recognition technology provide a good technical means to solve this problem. Security monitoring obtains big video data. It can use advanced artificial intelligence technology of video analysis to detect faces, so as to screen out suspects.

After years of development, video surveillance has experienced the development process of introduction, imitation, digestion and absorption, and independent innovation. From the introduction of foreign products and technologies to leading the development direction of global video monitoring, domestic video monitoring technology has experienced four different stages of development: analog monitoring, modular combined monitoring, network monitoring and intelligent monitoring. In recent years, with the basic completion of infrastructure construction in domestic safe cities, the security video monitoring system is rapidly entering the information application construction stage with data as the core and information driven, as shown in Figure 1. At this stage, how to collect, analyze and use value data more efficiently is the focus. In short, security video monitoring began to change from "visible" (standard definition, Networking), "visible" (high definition) to "understandable" (intelligent analysis) [18]–[22].

Face detection is to determine whether there is a face in a static or dynamic image after certain processing and analysis. If there is a face, the location and size information of each face is recorded and identified in the image. Although the research of face detection has been decades, and has made some achievements, but at present, the more mature face detection technology is basically for the detection of some ideal situations, and still faces many challenges in practical applications [23]–[26]. Although the existing algorithms have been improved for some of these situations and achieved some results, when there are many of the above situations, these algorithms still can not bring good detection results. Therefore, it is a research focus to develop a video face detection technology which can adapt to the above-mentioned non ideal situations [27]–[30]. It not only has a high academic value, but also will bring great contribution to national security, social stability and so on.

Thanks to the development of national economic construction and the support of policies, people's travel and long-distance mobility are more frequent, convenient and free, and the public security situation is becoming increasingly complex. A large number of population movements have led to various illegal and criminal activities in different places, not only the deployment of criminal plans is secret, but also the identity of the perpetrators is difficult to locate quickly. Among them, face recognition technology is a very effective one. This technology can be integrated with the existing public security business system to quickly extract, analyze and recognize personnel information, provide functions such as face retrieval, face comparison, ID card duplicate check, etc., to meet the application requirements of efficient management and control of various personnel in the field of public security video surveillance. People have higher and higher requirements for video analysis, and face recognition is one of the most important problems to be solved. In this paper, we use the deep learning framework and image processing technology to propose an algorithm solution for face detection and face recognition in natural scenes, and compare it with the face recognition algorithm based on multi-scale and high-dimensional features in this paper. Experiments show that the face recognition algorithm based on deep learning in this paper can better deal with the face recognition problems in natural scenes, which is of great significance to broaden the application field of face recognition and deepen the application level of face recognition.

II. INTELLIGENT VIDEO RETRIEVAL AND ITS APPLICATION IN FACE DETECTION

A. SECURITY VIDEO MONITORING AND RETRIEVAL TECHNOLOGY

There are three characteristics of video monitoring system data: massive, unstructured and low value density. The traditional analysis method is manual real-time monitoring and video query and playback. With the explosive growth of video monitoring data, the current video monitoring system, data management mode, data analysis application and other technologies have not met the requirements due to the number and ability of personnel (prone to fatigue for a long time, wrong or missed views, too late to see), limited display equipment (the number of front-end cameras is far greater than the number

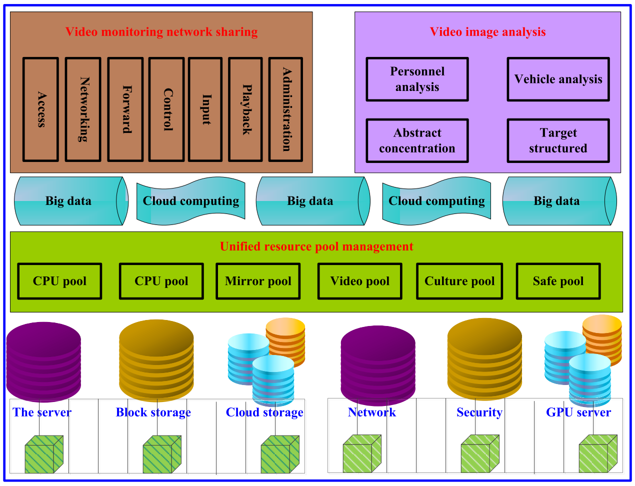


FIGURE 2. Intelligent video monitoring architecture.

of display screens) and other factors [31]–[35]. Specifically speaking, the current system architecture can not guarantee that all scenes can be monitored accurately and efficiently in 24 hours. At the same time, it is very difficult to find and trace the video surveillance image afterwards, and the human and material resources investment has exceeded the acceptable range. For example, in the process of detecting a major case, in order to find the criminal suspect in the video surveillance video, a local public security organ has used about 2000 police forces to conduct video playback and manual search for more than ten hours every day [36]–[39]. The total video viewing volume is equivalent to 830000 films, which consumes a lot of manpower and material resources.

In this context, the IT cutting-edge technology represented by GPU, Internet of things, big data, cloud computing, big data, AI and other technologies has been introduced into the field of security monitoring, promoting the evolution of security monitoring system architecture to the direction of intelligent monitoring around AI, cloud computing, big data as the core of a new generation of intelligent video monitoring architecture - video cloud came into being. Video cloud, a new generation of intelligent video monitoring architecture centered on AI, cloud computing and big data, emerges as the times require, as shown in Figure 2.

The main purpose of video surveillance retrieval is to locate the cause of an event and the development process of its association. The key information data of an event includes: time, place, person or object of the leading event, image and sound information. The richer the information transferred by the search conditions, the more accurate the location, and the simpler the search algorithm; on the contrary, the simpler the information transferred by the search conditions, the more fuzzy the location, and the more difficult the search algorithm is to locate accurately. Generally speaking, users expect that the retrieval conditions are simple and the location is accurate at the same time.

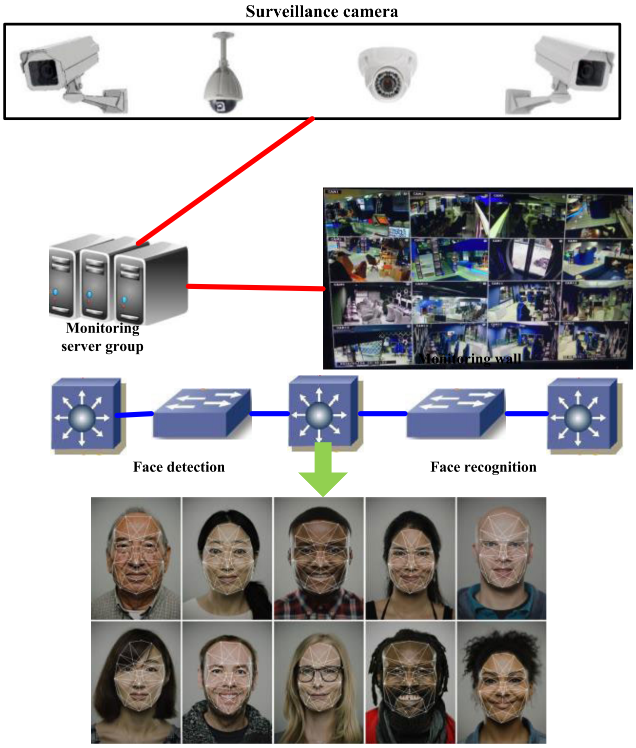


FIGURE 3. Working principle of security monitoring.

The emergence of intelligent video retrieval technology liberates people from monotonous and onerous tasks. It uses video segmentation, automatic digitization, voice recognition, shot detection, key frame extraction, automatic content association, video structure and other technologies, based on the knowledge of image processing, pattern recognition, computer vision, image understanding and other fields [40]–[42]. Through automatic intelligent analysis and preprocessing, the disordered and illogical monitoring video content (moving objects, pedestrians, vehicles) is combed, and the key information of events and targets in the video is automatically obtained, and the video content and index are generated based on these information. In order to improve the computing speed, the cluster mode is adopted at present, which can provide more than tens of times of real-time fast analysis ability, and can be expanded according to the application needs, improve the computing ability, and save the case handling time.

Fundamentally, the application of video retrieval technology in security monitoring is based on intelligent video analysis technology. Intelligent video analysis technology refers to the use of computer image visual analysis technology, by separating the background and target in the scene and then analyzing and tracking the target in the camera scene. In recent years, the term "big data" has been mentioned and used more and more, involving various industries. People use it to describe and define the massive data produced in the era of information explosion. The introduction of intelligent video analysis technology may greatly improve the retrieval efficiency and hit rate of the original mass monitoring video storage system, as shown in Figure 3.

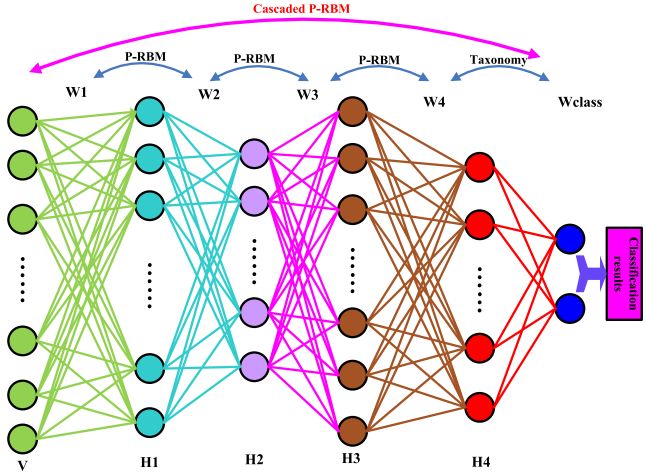


FIGURE 4. **Cascaded P-RBM deep learning detection network.**

At present, the technology is mainly through intelligent analysis preprocessing and face detection algorithm to collect the faces in the surveillance video, obtain the relevant information of interested objects in the video, and generate an index based on these face information. By viewing the face image, the relevant personnel can view all the targets contained in the video for several hours in a few minutes, determine the suspect target, and also watch the existing clips and motion tracks of the target in the whole video. In the system, input the face photos to be queried, select the faces to be retrieved, set parameters such as similarity, and then start to retrieve. Finally, the results of similar faces retrieved will be displayed on the interface.

B. DEEP LEARNING FACE DETECTION BASED ON VIDEO

According to the human vision system, the image information is processed from edge feature to contour feature, then to local feature and finally to global feature to generate intelligent semantics. This paper proposes a cascaded P-RBM (Probability state Restricted Boltzmann Machine) deep learning detection network with P-RBM as the core, which consists of one layer of visual layer, four layers of hidden layer and one layer of classification layer. The visual layer and four layers of hidden layer in the network constitute four P-RBM, corresponding to the four-layer processing process of image information by the above-mentioned visual system. In addition, in order to speed up the detection speed, the candidate region generated by skin color detection is used as the input of the detection network in the preprocessing layer, and then the face is selected in the image according to the output of the detection network to complete the detection. Face detection is to determine whether there is a face in a static or dynamic image after certain processing and analysis. If there is a face, the location and size information of each face is recorded and identified in the image.

The detection network model is shown in Figure 4. V represents the visible layer, $H_i = (i = 1, 2, 3, 4)$ represents the I hidden layer, and $W_i = (i = 1, 2, 3, 4, class)$ represents the connection weight between adjacent layers. In the figure, each layer is arranged longitudinally, the length of which corresponds to the number of neurons in the corresponding layer, and there is no connection between neurons in the layer. The number of neurons in hidden layer H1 is usually larger than that in visible layer V, which is conducive to fully extracting the lowest level features of input data; the number of neurons in each hidden layer (H_2, H_3, H_4) decreases layer by layer to remove redundant information and reduce the network size to improve the detection speed. The visible layer, the first hidden layer and the four P-RBMs of the adjacent hidden layer form a cascade P-RBM. The purpose of this structure design is not only to enable the neurons to extract enough bottom information to generate high-level semantics, but also to constrain the size of neurons in the network to reduce the local optimal problem, and to improve the calculation efficiency. Finally, the semantic information obtained by the classification layer is judged to output the detection results. In addition, the candidate area data is transformed into a column of one-dimensional data for input, which is convenient for calculation.

Because the direct training of multilayer neural network needs to transfer the error from the classification layer to the visualization layer, the multilayer transfer makes the error received by the visualization layer very small, which is not conducive to adjusting the network parameters, but also easy to cause the neural network to fall into the local optimum. Therefore, the training of detection network is divided into three stages: cascade P-RBM training, classification layer training and overall optimization. Cascaded P-RBM training uses unsupervised learning method to make the detection network fully learn the characteristics of input data at all levels, which constitutes the overall optimization of pre training process. Based on the results of cascaded P-RBM training, supervised learning is used to ensure the accuracy of training according to the output of H4. In the overall optimization, the results of the first two stages are taken as the initial value to adjust the detection network.

a: CASCADE P-RBM (PROBABILITY STATE-RESTRICTED BOLTZMANN MACHINE) TRAINING

Cascade P-RBM training uses greedy learning method to train multi-layer P-RBM layer by layer to avoid multi-layer transmission of training error. The purpose of P-RBM training is to learn the probability distribution model of input data by adjusting parameters. The probability density function of the input layer (i.e. the viewing layer) is:

$$p(V) = \sum_{H} \frac{e^{C^T V + B^T V + V^T WH}}{\sum_{V,H} e^{C^T V + B^T V + V^T WH}} \tag{1}$$

where C represents the offset of the visible layer, B represents the offset of the hidden layer, and W represents the connection weight between the visible layer and the hidden layer. Through the derivation of the log likelihood function of the probability density function of the visible layer, we can get:

$$\frac{\partial \ln p(V)}{\partial w_{ij}} = p\left(h_j = 1|V\right) p\left(v_i = 1|H\right) - \sum_V p(V)\, p\left(h_j = 1|V\right) p\left(v_i = 1|H\right) \quad (2)$$

$$\frac{\partial \ln p(V)}{\partial b_j} = p\left(h_j = 1|V\right) - \sum_V p(V)\, p\left(h_j = 1|V\right) \quad (3)$$

$$\frac{\partial \ln p(V)}{\partial c_i} = p\left(v_i = 1|H\right) - \sum_V p(V)\, p\left(v_i = 1|H\right) \quad (4)$$

The items in equation (4) $\sum_V p(V)$ need to be obtained by summing the corresponding probability values of the input data under the condition of obtaining the probability density function of the input data through multiple Gibbs sampling, but the calculation amount of this method is large. Here, we use the contrast difference algorithm (CD) to reduce the amount of computation. The definition of k-order contrast difference is as follows:

$$CD - k = P_0||P_\infty - P_k||P_\infty \quad (5)$$

where P_0 is the probability distribution when initializing network parameters; P_∞ is the exact probability distribution of input data; P_K is the probability distribution after k-times Gibbs sampling; $P_i||P_\infty$ is the KL distance between PI and P_∞. Since the probability distribution obtained after each Gibbs sampling is closer to the exact probability distribution of the input data than the previous one, therefore:

$$P_{k-1}||P_\infty \geq P_k||P_\infty \quad (6)$$

The equal sign only holds when P_{k-1} and P_k are the same probability distribution, and the probability distribution at this time is the exact probability distribution of input data. Therefore, k = 1 is generally set to obtain the same probability distribution of P_0 and P_1 at its minimum value (CD-1 = 0). At the same time, only one Gibbs sampling can greatly simplify the complexity of P-RBM training and accelerate the training speed.

A complete Gibbs sampling process includes the following two steps:

Step 1: The probability $p\left(h_j|H_{j'}, V\right)$ of each neuron HJ in the hidden layer is obtained when the state of the visible layer and other neurons in the hidden layer are known, and $H_{j'}$ is the other neurons in the hidden layer except H_j. Because the neurons in the hidden layer of P-RBM are conditionally independent from each other, so:

$$p\left(h_j|H_{j'}, V\right) = p\left(h_j|V\right) \quad (7)$$

Step 2: According to the hidden layer state obtained in step 1, the probability state $p\left(v_i|V_{i'}, H\right)$ of each neuron in the visible layer is reconstructed by sampling. Because each neuron in the visible layer in P-RBM is also conditionally independent from each other, therefore:

$$p\left(v_i|V_{i'}, H\right) = p\left(v_i|H\right) \quad (8)$$

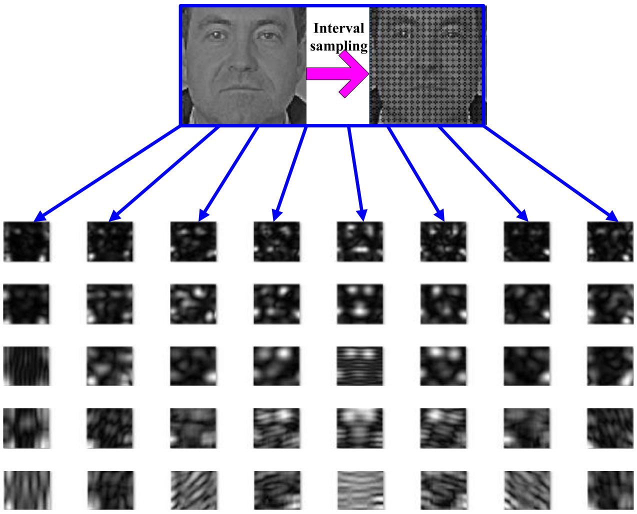


FIGURE 5. Interval sampling and filtering.

From a complete Gibbs sampling, the probability of each neuron in the visible layer can be obtained, then the probability distribution P_1 of the visible layer after Gibbs sampling can be obtained through probability modeling, and then the partial derivative of CD-1 under each parameter can be obtained as follows:

$$\frac{\partial}{\partial w_{ij}}(CD-1) \leq P\left(v_i = 1|H\right) P\left(h_j = 1|V\right) > P_0 - < P\left(v_i = 1|H\right) P\left(h_j = 1|V\right) > P_1 \quad (9)$$

$$\frac{\partial}{\partial b_j}(CD-1) \leq P\left(h_j = 1|V\right) > P_0 - < P\left(h_j = 1|V\right) > P_1 \quad (10)$$

$$\frac{\partial}{\partial c_i}(CD-1) \leq P\left(v_i = 1|H\right) > P_0 - < P\left(v_i = 1|H\right) > P_1 \quad (11)$$

When the parameter value of CD-1 = 0 is obtained, the P-RBM training is completed. The data of this P-RBM hidden layer is used as the initial data of the next P-RBM visible layer and the above training steps are repeated until all P-RBM training is completed. Compared with the traditional method of randomly initializing multi-layer network parameters and then using error back propagation to optimize the parameters, the multi-layer back propagation of errors is avoided, and the problem that multi-layer network training is easy to fall into local optimum is solved.

By setting a neuron in the hidden layer to be activated, the rest neurons are not activated, and then the neuron state in the hidden layer is reversed to the visible layer. Finally, the information obtained by the visible layer can be displayed through the gray-scale image to obtain the characteristics of the neuron. . When setting the activation state of neurons in Hi, the role of H_j (j > i) layer is not considered.The schematic diagram of interval sampling and filtering is shown in Figure 5.

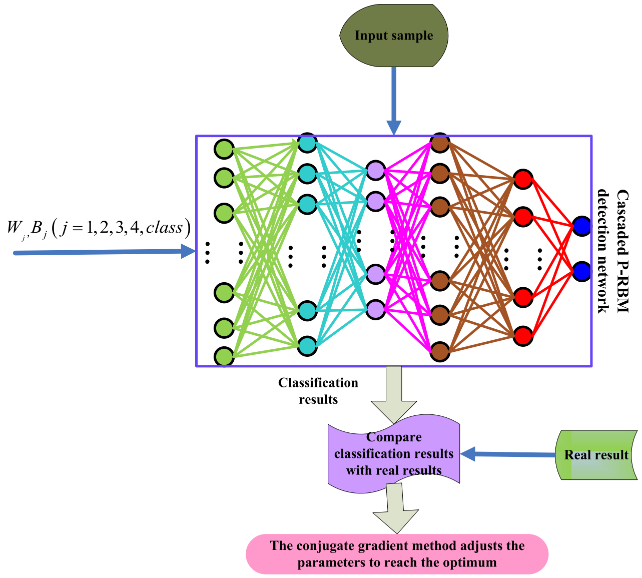

Input sample
$W_j, B_j\left(j=1,2,3,4,class\right)$
Cascaded P-RBM detection network
Classification results
Compare classification results with real results
Real result
The conjugate gradient method adjusts the parameters to reach the optimum

FIGURE 6. Overall optimization process.

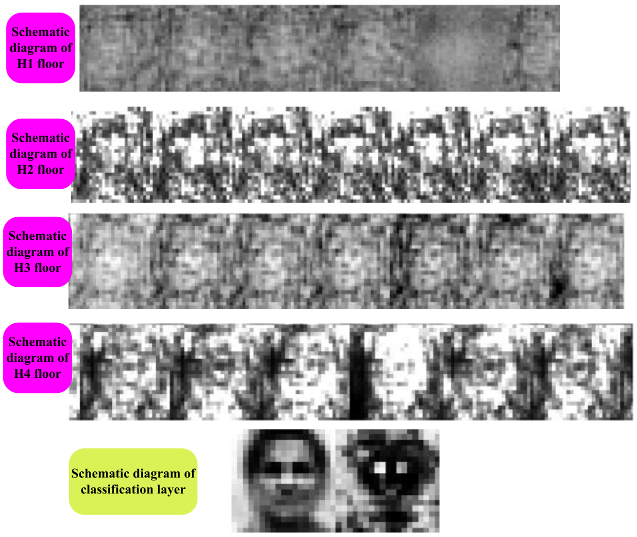

Schematic diagram of H1 floor
Schematic diagram of H2 floor
Schematic diagram of H3 floor
Schematic diagram of H4 floor
Schematic diagram of classification layer

FIGURE 7. Features of hidden layer and classification layer after optimization.

1) OVERALL OPTIMIZATION AND TESTING

Because cascaded P-RBM training and classification layer training are limited to training a single sub network, the parameters obtained from training are only the optimal parameters of the sub network, not the optimal parameters of the whole deep learning detection network. In this paper, a whole optimization process is used to train all the sub networks, so as to obtain the optimal parameters of the whole learning network. The previous cascaded P-RBM training forms the pre training of the whole optimization, which makes the whole optimization have a better initialization parameter, and solves the problem that the whole training of multi hidden layer learning network is easy to fall into the local optimization. The overall optimization process is shown in Figure 6:

The whole optimization process is very similar to the classification layer training, the difference is only that the parameters in the classification layer are adjusted in front, and the parameters of the whole detection network are adjusted here. Because the data in the detection network is transmitted from left to right, that is, from the visible layer to the hidden layer, in order to simplify the training process, the bias of the visible layer is no longer optimized.

After the overall optimization, the activation state of neurons in each hidden layer and classification layer is set, and then the information obtained from the visible layer is used to display the characteristics of the activated neurons. Some schematic diagrams are shown in Figure 7. Compared with Figure 6 and Figure 7, it can be found that after overall optimization, the facial features learned by neurons in each layer are more clear. As the number of layers increases, the face shape becomes more and more obvious. Especially, the activation state of neurons in the classification layer can be set to obtain nearly standard face templates.

According to the output of H4, the classification layer obtains the semantic information of the input data and realizes the face and non face classification of the input data. In order to make the detection network unified, the conditional probability distribution of a neuron being activated in the classification layer is the same as that of a neuron being activated in P-RBM, that is:

$$p\left(h_j=1|V\right)=\frac{1}{1+e^{-\left(b_j+V^T w_j\right)}} \tag{12}$$

The trained cascaded P-RBM deep learning detection network realizes face detection in a single video frame by classifying the candidate areas generated by skin color detection. Firstly, the video frame image is read, and the illumination compensation is carried out to reduce the influence of illumination on the detection effect. At the same time, the image scale is adjusted to reduce the too large image so that the subsequent detection can be faster. Then, in the YCbCr color space, we use the skin color ellipse fitting algorithm to detect the skin color region. Then we transform the candidate region into one-dimensional column vector and input the training depth learning network one by one. According to the classification results, we delete the non face region. Finally, we calibrate the face in the image to complete the face detection.

The face detection process of deep learning network is as follows:

1) The gray value of input data is normalized and assigned to each neuron in visual layer V as the activation state of the neuron.

2) According to equation (12), the data is transferred layer by layer to H4, in which the former hidden layer is the input layer of the latter.

3) According to equation (12), the probability of each neuron in the classification layer is obtained, and then the probability value of each neuron in the classification layer is compared. The largest one is the classification result of

the input data. If the classification result is face, the region corresponding to the input data will be determined as the result of face detection in the image; if the classification result is non face, the region will be skipped.

III. Face RECOGNITION BASED ON INTELLIGENT VIDEO RETRIEVAL TECHNOLOGY

Face recognition is a kind of biometric technology based on the facial feature information. Using camera or camera to collect the image or video stream containing face, and automatically detect and track the face in the image, and then carry out a series of face related technologies for the detected face [43]. The process mainly includes face image collection and detection, face image preprocessing, face image feature extraction, matching and recognition.

For the collection and monitoring of face image, in the dynamic monitoring environment of the airport, the face recognition technology must be able to meet the dynamic recognition, because in the real environment, except for the mandatory stay in the security inspection, in other cases, people are basically mobile. Therefore, in order not to miss the face capture, the algorithm requirements for face recognition are correspondingly high to ensure the operation speed and detection and capture speed of the face recognition system [44]–[46]. Therefore, for many companies that provide face recognition technology, whether they have advanced algorithms and belong to their own intellectual property rights is an important criterion to distinguish the technical strength of the company. The system operation flow chart is shown in the figure below:

In the process of intelligent video retrieval, users can get the key information they want from massive videos by intelligent video analysis technology according to their own retrieval conditions. What is the function of face recognition? In fact, face recognition is an advanced biometric technology, and it is also an advanced technology based on the analysis of face features to automatically identify people. Face recognition design to digital image, video processing, pattern recognition, computer vision algorithm and other aspects of professional technology.

A. CONVOLUTIONAL NEURAL NETWORK MODEL

The face recognition process based on convolutional neural network is mainly divided into training stage and testing stage. In the training stage, firstly, a large number of face sample data are collected to establish the training database, and the training samples are preprocessed. Preprocessing is usually divided into two parts. First, in order to remove the noise of training data, normalization is needed. Secondly, in order to make the distribution of training samples cover the illumination, scale and other situations as much as possible, we need to enhance the training samples to improve the generalization ability of the model. Then, the training of multi-layer convolution network model is divided into forward propagation and back propagation. Forward propagation refers to that the training samples are input into the network model by batch for training, and the output eigenvectors are extracted through the feature extraction part. After the prediction through the classification layer, the predicted probabilities are obtained. In the loss layer, the loss amount is calculated by using the loss function by comparing the predicted results with the given label errors. Back propagation refers to the use of convergence function (such as gradient descent function) for the loss of forward propagation, which transfers the loss layer back to the feature extraction part, and updates the parameters of multilayer network layer by layer. The purpose of this function is to update the parameters of the network to their respective gradients, so as to reduce the loss gradually. That is to say, the training goal of the whole network model is to make the prediction results of the forward propagation of the model consistent with the given label results. After the training of network model, the feature vector of hidden layer has the feature of distinguishing face. Finally, face matching and recognition are carried out by measuring the feature vector. In the test phase, compared with the training phase, it is relatively simple. First of all, a face test database needs to be established. However, each sample in the test database usually appears in pairs. After the test sample is passed through the network feature extraction part, the feature vectors of a pair of faces in the test sample are obtained. Then, whether the pair of faces are the same person is determined by measurement and threshold. The whole process is shown in Figure 8.

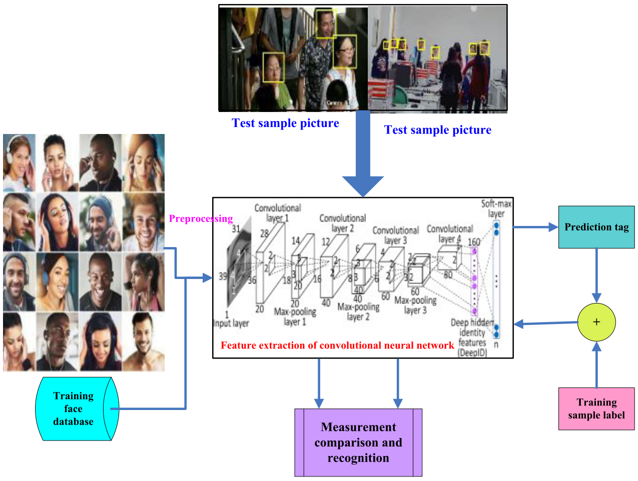


FIGURE 8. Recognition process based on deep convolution neural network.

1) ESTABLISH TRAINING SET AND VERIFICATION SET

The face training set used here is CASIA web face face database, which has 10575 types of people, covering people of different races and ages. The number of samples of each type varies from several to hundreds. In order to ensure the balance of the number of samples in each class, this paper selects more than 100 people in each class, and selects 915 people in total. 100 samples were randomly selected from each person, 80 of which were used as training samples, and the remaining 20 were used as verification samples. In this way, the training set and verification set were established. All samples were from natural scenes, including yellow people, white people, men and women, and different age groups.

2) FACE PREPROCESSING

The method of face preprocessing used in this paper is divided into three steps: face correction, data enhancement and normalization.

a: FACE CORRECTION

The distorted face pose in the original sample database is corrected by affine transformation. First of all, through the face calibration of Dlib open source database, the coordinates of the eye center of the face are recorded as left_eye_pos, right_eye_pos, as shown in the following formula:

$$left\_eye\_pos = (x_0, y_0) \tag{13}$$

$$right\_eye\_pos = (x_1, y_1) \tag{14}$$

According to the binocular coordinates, calculate the center of rotation (POS) and the rotation angle, as shown in formula (15) and (16):

$$center\_pos = ((x_0 + x_1)/2, (y_0 + y_1)/2) \tag{15}$$

$$angle = \arctan\left(\frac{y_1 - y_0}{x_1 - x_0}\right) \tag{16}$$

Then, the rotation matrix is calculated according to the rotation center and rotation angle, as shown in formula (17) :

$$rotation = \begin{bmatrix} \alpha, \beta, (1-\alpha)\cdot center\_pos.x - \beta \cdot center\_pos.y \\ -\beta, \alpha, \beta \cdot center\_pos.x + (1-\alpha)\cdot center\_pos.y \end{bmatrix} \tag{17}$$

The relationship between the corrected image coordinates (x ', y') and the original image coordinates (x, y) is shown in equation (18):

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} rotation_{(0,0)}, rotation_{(0,1)}, rotation_{(0,2)} \\ rotation_{(1,0)}, rotation_{(1,1)}, rotation_{(1,2)} \\ 0, \quad 0, \quad 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \tag{18}$$

The green points in the figure are facial feature points, and the effect of facial correction is shown in Figure 9. It can be seen that the experimental results are good.

b: DATA ENHANCEMENT

In order to improve the generalization ability of the model to deal with the scale, the up sampling and down sampling operations are respectively carried out on the scale of the original image. The up sampling operation means that the length and width of the original image are respectively enlarged to twice the length and width of the original image, and the down sampling operation means that the length and width of the original image are reduced to 1 / 2 of the length and width of the original image. In addition, in order to expand the sample, we turn the original data horizontally, as shown in formula (19):

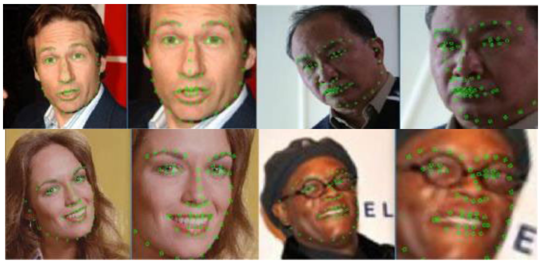


FIGURE 9. Face correction effect.

$$\mathrm{Im}\,age\,(x, y) = \mathrm{Im}\,age\,(w - 1 - x, y) \tag{19}$$

where Im $age\,(x, y)$ represents the pixel value corresponding to the original image coordinate (x, y), and w is the width of the original image size. The top row represents the effect of different scales, and the bottom row represents the effect of horizontal flipping.

c: NORMALIZATION

The normalization method adopted here is image size normalization and pixel value normalization. The operation of dimension normalization will normalize all samples after (1) (2) operation to 55 * 47. The operation of pixel value normalization linearly compresses all sample pixel values to [− 1,1].

3) FEATURE EXTRACTION OF CONVOLUTIONAL NEURAL NETWORK

Through the work in the previous section, the training set and verification set are sorted out, with a scale ratio of 4:1, 915 categories in total. Then, the network model is constructed to enter the training stage. In this paper, based on deepID1 classification network model, two improvements are made, and dropout operation is added in the last hidden feature layer. For the whole face training, the network with the whole face feature perception ability is obtained, and the training is carried out for the left eye, right eye, nose, left mouth corner, right mouth corner and other five parts. The network with local feature perception ability is obtained, and then the whole and local features are combined for recognition.

In general, the structure of multilayer convolution neural network includes basic elements such as convolution layer, pooling layer, full connection layer and loss layer. We will explain the principle of each basic element one by one later. The structure of the improved deep Id1 classification network model is as follows: 4 convolution layers, 3 Maximum pooling layers, 1 hidden feature layer, 1 dropout operation, 1 softmax divided into probability layer. In addition, the cross entropy loss function is used here in the loss function. The specific network structure is shown in Figure 10.

It should be noted that in the training stage, each training sample gets a probability vector after passing through the final softmax layer of the network. Each value of this

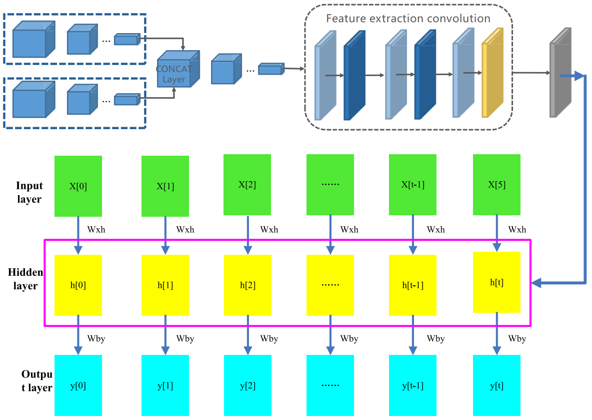


FIGURE 10. Schematic diagram of improved network model.

probability vector represents the probability value that the sample is classified into a certain category. In this paper, it is 915 people, so this probability vector should be 915 dimensions. At the end of the training stage, the feature vector formed in the hidden feature layer is the feature vector after feature extraction. The eigenvector of this layer is expanded into multi-dimensional vector by multi-dimensional matrix of the third largest pooling layer and multi-dimensional matrix of the fourth convolution layer. Then, the feature vector of the third largest pool layer is j, which is $5 \times 4 \times 60 = 1200$ dimension, and that of the fourth convolution layer is k, which is $3 \times 4 \times 80 = 960$ dimension. The calculation process is as follows (20):

$$Y = f\left(J \cdot W_1 + K \cdot W_2 + B\right) \tag{20}$$

Y is the eigenvector of the hidden layer, W_1 is the weight matrix of the fully connected layer and J, behavior 240, column 60. W_2 is the weight matrix of the full connection layer and K, with behavior 960 and column 60; B is the 160 dimensional vector offset by the full connection layer, and F is the activation function used by the full connection layer.

B. FACE RECOGNITION BASED ON GLOBAL AND LOCAL FEATURES

For making the whole face feature sample set, and taking five key parts of the face, namely, the left eye, the right eye, the nose tip, the left mouth corner and the right mouth corner, to make the local feature sample set. Two network models are trained separately, and then the feature vectors extracted from the two network models are used for face matching and recognition.

1) GLOBAL AND LOCAL FEATURES OF HUMAN FACE

Considering that in real life, the whole features of the face, such as the skin color distribution and contour distribution of the face, and the local features of the face, such as the ears, eyes, scars, etc., the recognition process of the person does not rely solely on the whole features or a certain type of local features for recognition, but needs to combine the two types of features for comprehensive judgment. Based on the above considerations, it is proposed to combine the whole feature network with the local feature network, which is more in line with the human physiological recognition process. In this paper, 915 types of people selected from webface are enhanced, and the whole face of the sample set is taken as the overall feature of the face. In addition, with the left eye, right eye, nose tip, left mouth corner and right mouth corner of the whole face as the center, the neighborhood is taken as the local feature of the face, as shown in Figure 11.

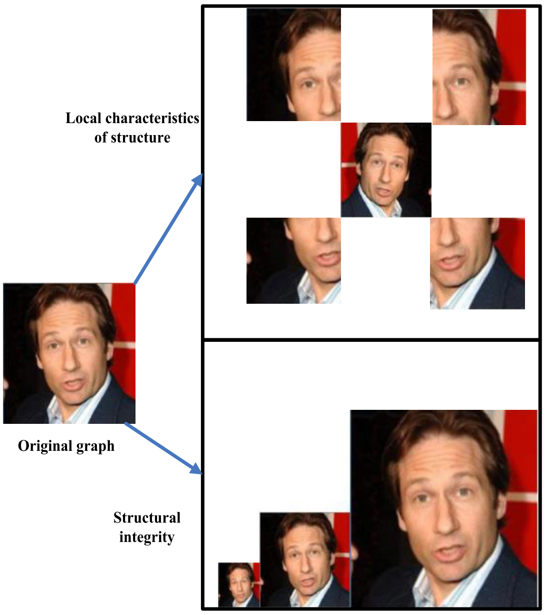


FIGURE 11. Global and local global features of human face.

In Figure 11, the left side represents the original face image, the upper right side represents the local feature image block set of the face, and the upper left corner represents the block taking operation centered on the left eye of the person. The upper right corner represents the operation of taking the right eye as the center, the middle part represents the operation of taking the nose tip as the center, and the lower left and lower right are the image blocks taking the left and right corners of the mouth as the center, respectively. These five parts make up different local features of human face; the lower part on the right represents the whole feature image set of human face, from left to right, which is the down sampling operation, the horizontal turning operation and the up sampling operation.

C. ASSOCIATED RECOGNITION

In the last section, we get the whole feature set and the local feature set. The whole feature set has 915 kinds of people. Each class of 100 * 4 = 400 samples is divided into training set and verification set, with the proportion of 4:1. For different local features, the local feature set is divided into five parts of the sub set, each sub set has 915 classes, each class of 100 * 4 = 400 samples, and the proportion of training set and verification set is also 4:1. After that, the improved deep id classification network model is used for different set training, and the network model with different perception ability is obtained. The overall network model is g, and the local characteristic network model of different parts is P_1, P_2, P_3, P_4, P_5.

The hidden layer eigenvectors of G and Pi $(i = 1, 2, 3, 4, 5)$ are the final face eigenvectors of this algorithm, which are used in face matching and recognition. For example, for face a and face B, the eigenvectors obtained from the two models are $G(A), Pi(A), G(B), Pi(B)$. Here, cosine measurement is used for comparison, so the similarity between a and B is as shown in formula (21):

$$similar = \gamma \cos(G(A), G(B)) + (1-\gamma) \times \left(\frac{\sum_{i=1}^{5} \cos(Pi(A), Pi(B))}{5} \right) \tag{21}$$

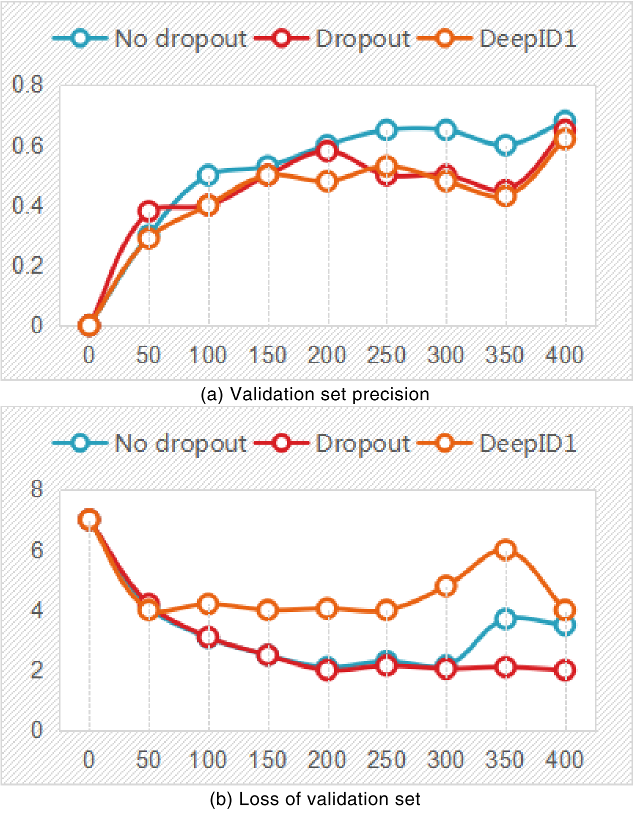
γ is the weight factor, the range is [0,1], the left part is the overall feature similarity, the right part is the local feature similarity, and the threshold is the similarity discrimination threshold. When the similar is greater than the threshold, a and B are the same person; otherwise, different people are considered.

IV. EXPERIMENTS AND RESULTS

A. COMPARISON EXPERIMENT OF DROPOUT OPERATION

It is determined that the activation function is tanh function and two super parameters batch_size = 16, learning_rate = 0.0001. The data set is consistent with the model in the previous section. The purpose of this section is to verify the feasibility of this paper. After hiding the feature layer, add dropout operation to prevent over fitting. The experimental results in the model training phase are shown in Figure 12.

Among them, the blue curve represents the operation of adding dropout, the red curve represents the operation of not adding dropout, and the orange represents the prototype of deepID1. It can be seen from Figure 12(a) that the three kinds of curves can converge, making the training accuracy close to 100%, but they can not be selected by the fitting effect of the training set alone. Because it is impossible to judge whether there is over fitting by relying on the training curve alone, it needs to judge from the verification set when there is little difference in the curve performance of the training set. Combined with the validation set curve, the effect of blue curve is better than that of red curve, which is about 10% higher than that of "one small grid". From the loss curve in Figure 12 (b), when the number of iterations is about 150k, the red curve slowly has an upward trend, indicating that there is an over fitting at this time, while the blue curve appears after about 250K. In conclusion, the addition of dropout operation can alleviate the over fitting phenomenon, and improve the accuracy of the verification set by about 10%, which also verifies the effectiveness of adding dropout to improve the network model.



(a) Validation set precision

(b) Loss of validation set

FIGURE 12. Dropout comparison curve or not.

B. COMPARISON ON TEST SET

This paper uses LFW (Labled Faces in the Wild) test library, which contains 3000 pairs of same person test samples and 3000 pairs of different person test samples, respectively, to add dropout improved model. There is no dropout model, deep ID1 model (in which the activation function of the first two models is tanh function, while deep Id1 uses relu function). Take the models whose basic convergence has not been fitted yet (i.e. the number of iterations of blue curve is about 250K, the number of iterations of red curve is about 200K, and the number of iterations of orange curve is about 150k) for comparison. Take the true positive rate (the probability of correctly identifying the same person between the same person) as the ordinate value, and the false positive rate (the probability of incorrectly identifying the same person between different people) as the abscissa value. The ROC curve experimental results are shown in Figure 13.

The values of each point of ROC curve represent the values of true and false positive rates of the model under different discrimination thresholds. For example, the coordinates (0,0) in the figure indicate that when the discrimination threshold is 1, the true and false positive rates of the model are both 0. As the discrimination threshold falls from 1 to 0, the true and false positive rates will increase, and the curve trend will rise. The larger the area enclosed by the curve and the x-axis, the better the representation performance. From Figure 13, it can be seen that the effect of adding dropout model is

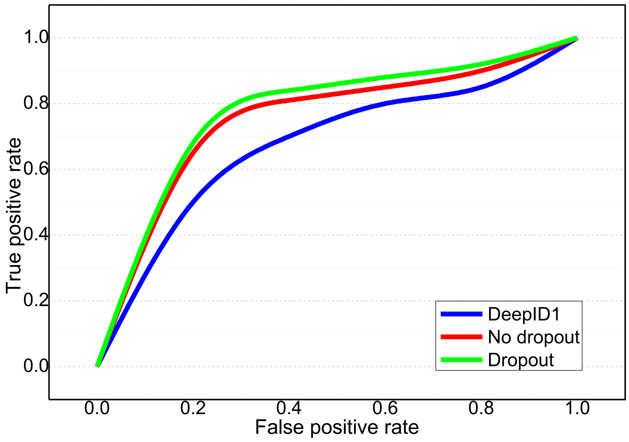


FIGURE 13. ROC curve diagram.

TABLE 1. Performance comparison of algorithm under threshold 0.6.

Single network	True positive rate	False positive rate
DeepID1	98.8%	87.84%
Dropout	72.38%	26.46%
No dropout	74.34%	25.89%

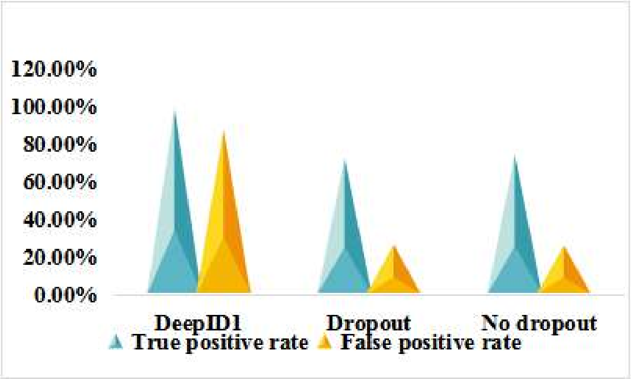


FIGURE 14. Performance comparison of algorithm under threshold 0.6.

the best of the three, which can prove the effectiveness of this approach. In addition, to compare the performance more intuitively, when the discrimination threshold is 0.6 from the above figure, the performance of the three is shown in Table 1.

It can be seen from Figure 14 that although the true positive rate of deepID1 classification network is higher than that of the improved model in this paper, the false recognition rate is very high, which indicates that the differentiation between the same person and the different person in deepID1 model is not enough. Because the similarity between the same person and the different person is very high, the true and false positive rate will be very high. Although the true positive rate of dropout is not as high as that of deepID1, the false recognition rate is far lower than that of deepID1 classification model. In contrast, the improved network homo and non Homo in this paper is significantly better than that of deepID1 classification model.

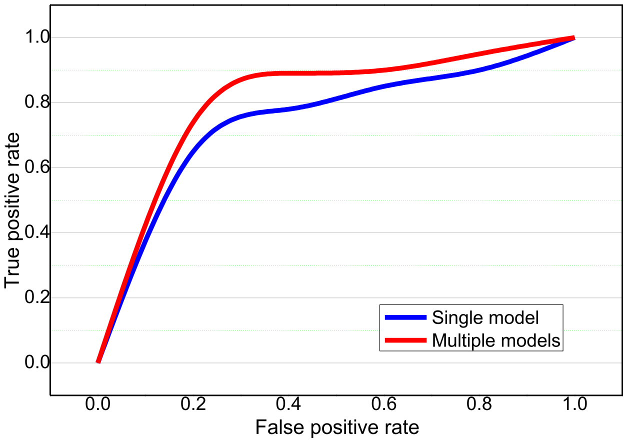


FIGURE 15. ROC curve comparison between single model and multi model.

TABLE 2. Performance comparison of algorithm under threshold 0.6.

Single network	True positive rate	False positive rate
Single network model	74.44%	25.89%
Multi network model	77.34%	18.64%

C. COMPARISON EXPERIMENT OF SINGLE MODEL AND MULTI MODEL

This paper selects the best super parameters and activation functions, and verifies one of the improvement points in this paper: adding dropout operation in the hidden feature layer. The purpose of this section is to verify another improvement point in this paper: joint global feature network model and local feature network model. In this paper, we use the improved network model to train the global and local features of human face, and get the global feature model G, and five local feature models P1, P2, P3, P4, P5. The weight of the global network model is r. In this section, the experiment takes r = 0.5, the test set data used is LFW, including 6000 pairs of test samples (3000 pairs of test samples for the same person, 3000 pairs of test samples for the different person). The experimental results of ROC curve are shown in Figure 15 for true positive rate (the same person is correctly identified as the same person) and false positive rate (the different people are wrongly identified as the same person).

It can be seen from Figure 15 that the effect of combining multiple models is better than using a single model. In addition, when the threshold value is 0.6 from the curve, the performance indexes of single model and multiple models are shown in the Table 2 and Figure 16 below.

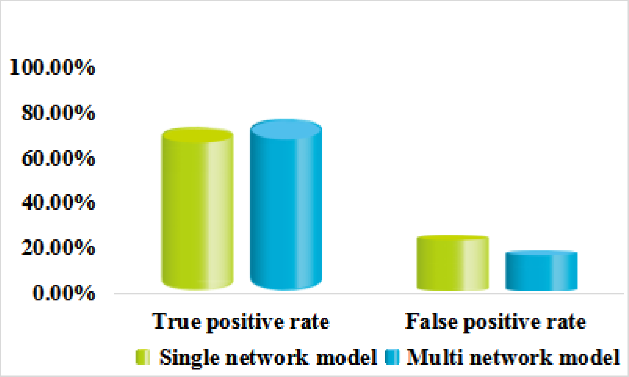


FIGURE 16. Performance comparison of algorithm under threshold 0.6.

To sum up, from the true positive rate and false positive rate, combined with multi network model joint recognition is slightly better than single network recognition.

V. CONCLUSION

With the continuous development of social economy and the demand of market economy, more and more enterprises are gradually aiming at face recognition products of surveillance video, and face recognition products can be seen everywhere in the current traffic and housing construction. At the same time, the video monitoring system in the field of public security mainly focuses on the pre-warning analysis of the collected video images, but the post video analysis wastes a lot of manpower and energy. Therefore, face recognition technology is used in the field of public security video monitoring, which will reduce the probability of illegal crime and maintain the stability of the society and the country. According to the continuous distributed activation state of human brain neurons, this paper proposes a cascaded deep learning detection network for video single frame face detection. By reducing the number of neurons in each hidden layer layer layer by layer to restrict the size of neurons in the network and improve the calculation efficiency, and using skin color detection to generate candidate areas in the preprocessing layer, the robustness of the neural network is improved while reducing the number of neurons in each layer of the network. The detection network initializes the network parameters through pre training, and uses greedy learning to avoid multi-layer network training. Because of the multi-layer error transfer, the error obtained at the last layer is too small to adjust the network parameters well, which further improves the robustness while maintaining the global optimization of the learning network. Face recognition is the key of biometric recognition. The most popular research of artificial intelligence is face recognition technology. With the development of science and technology, the application of face recognition technology in intelligent video surveillance system has broad application prospects and great practical significance. In this paper, a face recognition algorithm based on multi model combination of deep learning is proposed. The weight factor used in the whole feature network combined with the local feature network is a fixed value. The following idea is to take the feature vectors extracted from these two kinds of network models as input and output them as face comparison results (0 for different people, 1 for the same people), and then train a face comparison network model, which can distinguish by comparing the network rather than setting a fixed weight factor to calculate the similarity.

REFERENCES

[1] P. Glauner, J. A. Meira, P. Valtchev, R. State, and F. Bettinger, "The challenge of non-technical loss detection using artificial intelligence: A survey," *Int. J. Comput. Intell. Syst.*, vol. 10, no. 1, pp. 760–775, 2016.

[2] M. A. Tebbi and B. Haddad, "Artificial intelligence systems for rainy areas detection and convective cells' delineation for the south shore of Mediterranean Sea during day and nighttime using MSG satellite images," *Atmos. Res.*, vols. 178–179, pp. 380–392, Sep. 2016.

[3] P. Dai, X. Wang, W. Zhang, P. Zhang, and W. You, "Implicit relative attribute enabled cross-modality hashing for face image-video retrieval," *Multimedia Tools Appl.*, vol. 77, no. 18, pp. 23547–23577, Sep. 2018.

[4] Q. W. Wang and Z. L. Ying, "A face detection algorithm based on Haar-like t features," *Pattern Recognit. Artif. Intell.*, vol. 28, no. 1, pp. 35–41, 2015.

[5] M. Sepandi, M. Taghdir, and A. Rezaianzadeh, "Assessing breast cancer risk with an artificial neural network," *Asian Pacific J. Cancer Prevention*, vol. 19, no. 4, pp. 1017–1019, 2018.

[6] Y.-H. Lai and C.-K. Yang, "Video object retrieval by trajectory and appearance," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 6, pp. 1026–1037, Jun. 2015.

[7] A. K. G. Worner, "Realtime quality monitoring of compressed video signals," *SMPTE J.*, vol. 111, no. 9, pp. 373–377, Sep. 2002.

[8] D. Saravanan and S. Srinivasan, "Video data mining information retrieval using BIRCH clustering technique," in *Advances in Intelligent Systems and Computing*, vol. 325. New Delhi, India: Springer, 2015, pp. 583–594.

[9] M. Okabe, Y. Dobashi, and K. Anjyo, "Animating pictures of water scenes using video retrieval," *Vis. Comput.*, vol. 34, no. 3, pp. 347–358, Mar. 2018.

[10] Y. Li, R. Wang, Z. Cui, S. Shan, and X. Chen, "Spatial pyramid covariance-based compact video code for robust face retrieval in TV-series," *IEEE Trans. Image Process.*, vol. 25, no. 12, pp. 5905–5919, Dec. 2016.

[11] S. Jairath, S. Bharadwaj, and M. Vatsa, "Adaptive skin color model to improve video face detection," in *Advances in Intelligent Systems and Computing*, vol. 390. New Delhi, India: Springer, 2016, pp. 131–142.

[12] M. Chouchene, F. E. Sayadi, H. Bahri, J. Dubois, J. Miteran, and M. Atri, "Optimized parallel implementation of face detection based on GPU component," *Microprocessors Microsyst.*, vol. 39, no. 6, pp. 393–404, Aug. 2015.

[13] O. Dospinescu and I. Popa, "Face detection and face recognition in Android mobile applications," *Inf. Economica*, vol. 20, no. 1, pp. 20–28, 2016.

[14] A. E. Sergeev, A. S. Konushin, and V. S. Konushin, "Reducing background false positives for face detection in surveillance feeds," *Comput. Opt.*, vol. 40, no. 6, pp. 958–967, 2016.

[15] Q. Chen, L. Yang, D. Zhang, Y. Shen, and S. Huang, "Face deduplication in video surveillance," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 32, no. 03, Mar. 2018, Art. no. 1856001.

[16] A. Alotaibi and A. Mahmood, "Deep face liveness detection based on nonlinear diffusion using convolution neural network," *Signal, Image Video Process.*, vol. 11, no. 4, pp. 713–720, May 2017.

[17] K. Patel, H. Han, and A. K. Jain, "Secure face unlock: Spoof detection on smartphones," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 10, pp. 2268–2283, Oct. 2016.

[18] M. A. Haque, K. Nasrollahi, T. B. Moeslund, and R. Irani, "Facial video-based detection of physical fatigue for maximal muscle activity," *IET Comput. Vis.*, vol. 10, no. 4, pp. 323–330, Jun. 2016.

[19] S. Merat and W. Almuhtadi, "Artificial intelligence application for improving cyber-security acquirement," in *Proc. IEEE 28th Can. Conf. Electr. Comput. Eng. (CCECE)*, May 2015, pp. 1445–1450.

[20] M. arković and Z. Stojković, "Analysis of artificial intelligence expert systems for power transformer condition monitoring and diagnostics," *Electr. Power Syst. Res.*, vol. 149, pp. 125–136, Aug. 2017.

[21] C. Li, J. Valente de Oliveira, M. Cerrada, F. Pacheco, D. Cabrera, V. Sanchez, and G. Zurita, "Observer-biased bearing condition monitoring: From fault detection to multi-fault classification," *Eng. Appl. Artif. Intell.*, vol. 50, pp. 287–301, Apr. 2016.

[22] J. A. Quinn, C. K. I. Williams, and N. McIntosh, "Factorial switching linear dynamical systems applied to physiological condition monitoring," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 9, pp. 1537–1551, Sep. 2009.

[23] Y. Azan Basallo, V. Estrada Senti, and N. Martinez Sanchez, "Artificial intelligence techniques for information security risk assessment," *IEEE Latin Amer. Trans.*, vol. 16, no. 3, pp. 897–901, Mar. 2018.

[24] B. Wang, C. Liu, Y. Xiao, J. Zhong, W. Li, Y. Cheng, B. Hu, L. Huang, and J. Zhou, "Ultrasensitive cellular fluorocarbon piezoelectret pressure sensor for self-powered human physiological monitoring," *Nano Energy*, vol. 32, pp. 42–49, Feb. 2017.

[25] H. Taheri Shahraiyni, S. Sodoudi, A. Kerschbaumer, and U. Cubasch, "A new structure identification scheme for ANFIS and its application for the simulation of virtual air pollution monitoring stations in urban areas," *Eng. Appl. Artif. Intell.*, vol. 41, pp. 175–182, May 2015.

[26] J. Liu, J. Chen, and J. Zhang, "Geographic information dynamic monitoring in ntelligent Era," *Wuhan Daxue Xuebao (Xinxi Kexue Ban)/Geomatics Inf. Sci. Wuhan Univ.*, vol. 44, no. 1, pp. 92–96, 2019.

[27] J. Xu and K. Wu, "Living with artificial intelligence: A paradigm shift toward future network traffic control," *IEEE Netw.*, vol. 32, no. 6, pp. 92–99, Nov. 2018.

[28] X. Liu, X. Zhang, and V. M. Venkatasubramanian, "Distributed voltage security monitoring in large power systems using synchrophasors," *IEEE Trans. Smart Grid*, vol. 7, no. 2, pp. 982–991, Mar. 2016.

[29] T. Nguyen, H.-L. Mai, G. Doyen, R. Cogranne, W. Mallouli, E. M. D. Oca, and O. Festor, "A security monitoring plane for named data networking deployment," *IEEE Commun. Mag.*, vol. 56, no. 11, pp. 88–94, Nov. 2018.

[30] W. Li, *Security Monitoring Technology of Smart Grids* (Lecture Notes in Electrical Engineering), vol. 334. Cham, Switzerland: Springer, 2015, pp. 187–196.

[31] S. Scardapane, M. Scarpiniti, and M. Bucciarelli, "Microphone array based classification for security monitoring in unstructured environments," *AEU-Int. J. Electron. Commun.*, vol. 69, no. 11, pp. 1715–1723, 2015.

[32] C. F. He, L. Ruan, and H. Feng, "Security monitoring system based on a line structure Sagnac interferometer with 3×3 coupler," *Guangdianzi Jiguang/J. Optoelectron. Laser*, vol. 26, no. 6, pp. 1125–1131, 2015.

[33] R. Rada, "Artificial intelligence," *Artif. Intell.*, vol. 28, no. 1, pp. 119–121, Feb. 1986.

[34] A. Ligęza and T. Potempa, "Artificial intelligence for knowledge management with BPMN and rules," *IFIP Advances in Information communication Technology*, vol. 422. Berlin, Germany: Springer, 2016, pp. 19–37.

[35] W. Glauser, "Artificial intelligence, automation and the future of nursing," *Can. Nurse*, vol. 113, no. 3, pp. 24–26, 2017.

[36] B. Fan and R. Zhang, "Unmanned aircraft system and artificial intelligence," *Wuhan Daxue Xuebao (Xinxi Kexue Ban)/Geomatics Inf. Sci. Wuhan Univ.*, vol. 42, no. 11, pp. 1523–1529, 2017.

[37] A. Kasthuri, A. Suruliandi, and S. P. Raja, "Gabor-oriented local order feature-based deep learning for face annotation," *Int. J. Wavelets, Multiresolution Inf. Process.*, vol. 17, no. 05, Sep. 2019, Art. no. 1950032.

[38] A. Dhomne, R. Kumar, and V. Bhan, "Gender recognition through face using deep learning," *Procedia Comput. Sci.*, vol. 132, pp. 2–10, 2018.

[39] D. Luo, G. Wen, D. Li, Y. Hu, and E. Huan, "Deep-learning-based face detection using iterative bounding-box regression," *Multimedia Tools Appl.*, vol. 77, no. 19, pp. 24663–24680, Oct. 2018.

[40] Y. Zhao, F. Tang, W. Dong, F. Huang, and X. Zhang, "Joint face alignment and segmentation via deep multi-task learning," *Multimedia Tools Appl.*, vol. 78, no. 10, pp. 13131–13148, May 2019.

[41] S. Reardon, J. Tollefson, A. Witze, and E. Ross, "US science agencies face deep cuts in trump budget," *Nature*, vol. 543, no. 7646, pp. 471–472, Mar. 2017.

[42] Y. Sun, X. Wang, and X. Tang, "Hybrid deep learning for face verification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 10, pp. 1997–2009, Oct. 2016.

[43] X. Liu, M. Kan, W. Wu, S. Shan, and X. Chen, "VIPLFaceNet: An open source deep face recognition SDK," *Frontiers Comput. Sci.*, vol. 11, no. 2, pp. 208–218, Apr. 2017.

[44] W. Naiguo, Z. Sitao, and W. Huitao, "Connected effect between surface subsidence and rock burst in fully mechanized caving face under deep alluvium," *J. Mining Saf. Eng.*, vol. 33, no. 2, pp. 214–219, 2016.

[45] G. Mai, K. Cao, P. C. Yuen, and A. K. Jain, "On the reconstruction of face images from deep face templates," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 5, pp. 1188–1202, May 2019.

[46] S. Chokkadi and I. NMAMIT, "A study on various state of the art of the art face recognition system using deep learning techniques," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 4, pp. 1590–1600, Aug. 2019.

ZUOLIN DONG was born in Nanyang, Henan, in February 1970. He received the bachelor's degree in engineering from the Beijing University of Technology and the master's degree in engineering from Xi'an Jiaotong University. He is currently an Associate Professor and the Deputy Director of the Instruction Center of Innovation and Staring a Business in the Henan Institute of Technology. He is also an Outstanding Young Expert in Science and Technology in Xinxiang, and a Senior Assessor of National Vocational Skills Appraisal. He has presided over and participated in three provincial and ministerial level projects, four prefecture-level projects and seven school level projects, and published four teaching materials and more than ten academic articles in journals at all levels. He is a member of the Communist Party of China.

JIAHONG WEI was born in Jinzhou, Liaoning, in December 1983. She received the doctor's degree from the Dalian University of Technology. She is currently a Lecturer with the Henan Institute of Technology. She has participated in three national level projects, two provincial level projects and one school level project, and published seven articles in important journals at home and abroad, including two articles cited by SCI and EI, and five articles belonging to Chinese core journal. She is a member of the Communist Party of China.

XIAOYU CHEN was born in Xinxiang, Henan, in December 1983. He received the bachelor's degree in engineering from the PLA University of Information Engineering and the master's degree in engineering from the Wuhan University of Technology. He is currently a Teacher with the Henan Institute of Technology. He has presided over one prefecture-level project, participated in five provincial and ministerial level projects and two school level projects, and published two teaching materials and more than ten academic articles in journals at all levels. He is a member of the Communist Party of China.

PENGFEI ZHENG was born in Sanmenxia, Henan, China, in November 1994. He received the bachelor's degree from Henan Normal University, where he is currently pursuing the master's degree with the College of Electronic and Electrical Engineering. He has participated in five fund projects.