浙江工业大学

本科毕业论文（设计）

文献综述

论文题目：Design and Implementation of Lab Attendance System based on Face Recognition

学　　院: 计算机科学与技术学院、软件学院

专　　业: 软件工程 (中外合作办学)

班　　级: 20192019 软件工程（中外合作办学）01

学　　号: 201906150117

学生姓名: 沈浪

指导老师: 李小薪

提交日期: 2023 年 02 月

Abstract: This paper is a literature review on the design and implementation of face recognition attendance system based on MTCNN and FaceNet models. It first introduces the project and its research meaning, and then introduces the research status and key issues of the project.Then this paper briefly describes the MTCNN model and FaceNet model, and then introduces the main process of face detection and recognition system.

Keywords: MTCNN, FaceNet, Face recognition, Attendance system

1 Introduction

Face recognition using camera equipment to collect face image is a technology for identification according to the facial features of the target to be detected.With the popularization and application of computer network technology, face recognition technology is widely used in gate access control, attendance management, face payment and many other fields.Traditional biometrics, such as fingerprint recognition, voice recognition and iris recognition, are either seriously affected by internal and external factors, so the recognition efficiency is low, or have high requirements for software and hardware facilities, so it is difficult to popularize and apply.Compared with the traditional biometrics technology, face recognition technology can be more intuitive and convenient through video surveillance equipment to check personnel identity information, has the characteristics of simplicity, efficiency, economy and scalability, can be applied to security verification, video surveillance, personnel control and many other aspects.Traditional face recognition methods mainly use manual facial feature detector design, in unconstrained environment, the feature information is easy to be affected by external factors, resulting in reduced robustness of face recognition algorithm.At present, traditional face recognition methods have been gradually replaced by deep learning methods based on convolutional neural networks. The advantage of deep learning methods lies in the use of a large number of data sets for feature training, so as to learn the best features of the deep level of the target to be detected[1]。

2 Research Significance

The traditional laboratory attendance system requires laboratory staff to sign in and sign out according to a specific Angle, which has the following two disadvantages : 1) once the location of the face is not accurate, the accuracy of the system recognition will be very low. 2) Staff need to remember to clock in every day, once forgotten, a day's work will be wasted.The laboratory attendance system designed by this project is a face recognition system based on deep learning model MTCNN and Facenet.It mainly

uses MTCNN model for face detection and image extraction, uses Facenet model for face feature extraction, and completes real-time recognition of face image by building local face feature database. Laboratory staff in and out of the laboratory every day, the system will real-time face recognition and automatic attendance, improve the accuracy and convenience of recognition.

3 Research Status and Key Problems

3.1 Methods of Face Recognition

Since the 1960s, scholars began to study face recognition. Through decades of development, face recognition has been applied and developed in many fields, and the research at home and abroad is also very active.Some developed countries and some developing countries have set up research institutions and teams for face recognition.

Face recognition research was first carried out in the United States, and its research technology is higher than other countries. In 1993, The U.S. Department of Defense's Advanced Research Projects Agency and the U.S. Army Research Laboratory established Feret(Face Recognition) Technology project team, established the FERET face database, used to evaluate the performance of face recognition algorithms.[2].In recent years, Japan has been accelerating the research of intelligent video analysis technology, in 2015, a Japanese company launched a video surveillance face recognition technology scanning speed has been able to reach 36 million images per second. In the same year, the backbone airports in Japan introduced face recognition systems to provide a more convenient way for inbound and outbound tourists to enter and exit the country[3].

Compared with foreign countries, China's face recognition algorithm research is relatively lagging behind, the earliest began in the 1990s. At the beginning of the transformation from artificial to computer intelligent recognition in China, the biometric technology used is fingerprint recognition. However, with the development of The Times, people's needs have also changed, therefore, more accurate technology is needed to meet the new needs of practical applications, so face recognition came into being. Face recognition was first used in the field of security. In 2001, the public security department began to use face recognition technology to prevent and combat major criminal crimes, and won the support of the state. Li Ziqing, a doctor of automation from Chinese Academy of Sciences, led a team to develop "Zhongke Oxen", which was applied in the 2008 Beijing Olympic Games and the 2010 World Expo to improve the safety index of the venue and ensure the smooth implementation of activities[4].In 2010, the Shanghai World Expo, the

technology has been more widely used, while major companies compete to join the camp of this technology, to achieve the large-scale application of face recognition in China.

From 1960s to 1990s, the research on face recognition mainly used the geometric features of facial features for matching. Firstly, the position of facial features was located, and then the relationship between their shape and position was analyzed. Finally, the pattern matching was carried out by measuring Euclidean distance and classifier. This method is easy to lose face information, recognition effect is not ideal, temporarily stay in the theoretical stage, can not be used in the actual scene.

In 1987, Sirovich and Kireby proposed a principal component analysis-based feature face dimensionality reduction method[5] for PCA(Principle Componet Analysis) based face representation and recognition. With this approach, Turk and Pentland convert the entire face image into a vector and compute the feature face with a set of samples. PCA is able to obtain data representing a face at an optimal level with the data obtained from the image. Different faces and light levels of the same person are considered as the weakness of PCA.

In 1997, Yale University Belhhumur improved on Fisherfaces and introduced linear discriminant analysis[6]. This method uses supervised training to map faces onto the feature space, so that the feature information of the same person can be as close as possible and the feature information points of different people can be as far away as possible to reduce the impact of facial expression, posture, and lighting.Improve the identification efficiency.France and Herault Jutten put forward the method of independent component analysis[7], will get the face feature vectors independent component of the linear combination, as PCA algorithm.Hong Ziquan proposed a dimensionality reduction method based on singular value decomposition. First, singular value decomposition is used as the feature vector of face image, and finally, high-dimensional information is compressed into a low-dimensional subspace for classification.

F. maria uses the hidden Markov model of five states[8], and uses a group of unobstructible state sequence parameters to represent the relationship between various organs of the face, and puts forward a pseudo-two-dimensional hidden Markov model.Applied to face recognition, this model can well reflect the correlation between organs, is not sensitive to the change of expression, and has good robustness.

Researchers combined with the learning ability and association ability of neural network, neural network applied to face recognition, neural network can learn and extract face features through the learning ability of the data set, do not need researchers to spend time and energy to design face feature extraction algorithm, at the same time, neural net-

work has good classification ability, especially in the field of face recognition.Kohenen proposed that the self-organizing mapping network[9] could reproduce faces well.Lin applies probabilistic neural network to face recognition, repeatedly trains samples, and improves the learning efficiency and convergence speed of neural network through modular structure.Many traditional neural network models have been proposed.

In 2006, Galundor University GeofferyHinton et al.[10] proposed the concept of deep learning and the method of greedy layer by layer pre-training for the single problem of neural network single hidden layer structure. The paper proposed that the multi-hidden layer can better extract features without the need of artificial design algorithm, thus improving the accuracy of classification.Unsupervised learning is used for layer by layer pre-training to solve the problem of too many network parameters.Convolutional neural network is one of the classical algorithms in deep learning at present. Due to its simple structure and few training parameters, it has become a widely used model in deep learning and a mainstream method for face recognition, playing an extremely important role.

In 2014, Tang Xiaoou team of the Chinese University of Hong Kong[11] proposed DeepID algorithm based on convolutional neural network, which fused deep features of different regions of face image, verified faces by using Bayesian method, and tested on open face data set LFW[12], achieving 97.45% accuracy.The framework is mainly used for face verification, that is, comparing whether two faces are the same person, and finally, face recognition is carried out through softmax regression.FaceBook proposed DeepFace[13] algorithm to triangulate faces, reconstruct faces through key feature information, and finally, extract and classify features to realize face recognition, achieving a recognition rate of 97.35%.In 2015, FaceNet[14] system developed by Google in the United States introduced the TripletLoss function to map face features to Euclidean space, so that the spatial distance of images matching samples is closer, while that of unmatched samples is farther.The accuracy of LFW data set was 99.63%.In 2016, IacopoMasi proposed an unconstrained face recognition method[15]to study the recognition problem caused by changes in face posture.

With the development of Convolutional neural network, in recent years, more and more companies at home and abroad, such as Google, Facebook, and SenseTime, have applied face recognition to the security check of public places, the place where the company checks in for attendance and signs in for meetings. However, there are still some problems and challenges in the actual application. Although many methods have achieved high accuracy in LFW data sets, even exceeding the ability of human

eye recognition However, these deep models need to be trained through a large number of samples, which is difficult to achieve for universities or small research institutions. Therefore, how to train a relatively simple model with good performance, which can obtain more distinctive facial features and match faces well is an urgent problem that we need to solve.

3.2 Development of Attendance System

The attendance system was first born in the 1970s[16] and has undergone the following evolution:

The first generation of attendance is a paper card system, replacing the traditional manual attendance system.The time is printed on the card through the micro print head to realize attendance.Advantages are simple operation, the machine is not complex;The disadvantage is that it can not effectively identify the identity of the attendance person, it is easy to take the exam on behalf of the situation, and there is a lot of statistical work in the later period, the paper consumption is large.

The second generation of attendance is the bar code attendance system.It is mainly to record the situation of mine workers in the well[17]. The bar code is projected far away from the optics, and the bar code is scanned by the mining lamp irradiated by the camera, so as to realize the attendance check.The advantages are high accuracy, low cost and fast speed, but the disadvantages are that the bar code is easy to be damaged and forged.

The third generation of attendance is magnetic card type attendance system.The attendance test is now widely used in major enterprises and institutions with remarkable results.The advantage is high attendance efficiency, the disadvantage is that the magnetic card is easy to lose, easy to take the test.

The fourth generation of attendance is biological information identification attendance system.First input fingerprint, iris[18], face information, and then compare the information for recognition.The advantage is that there is no need to carry other proof, identity identification is unique, there will not be a proxy test.The disadvantage is that it is easy to be affected by external factors, the fingerprint needs to be clean and not damaged, the face needs no external occlusion, the stability and accuracy of the verification is low, the cost is high.

3.3 Key Problems

- Illumination problem: Illumination is always a very important and difficult problem in face detection and recognition. At present, many recognition methods have varying degrees of dependence on illumination conditions. The existence of over-

light, over-dark or polarized light may lead to a sharp decline in recognition rate. Although some specific solutions have been proposed, in general, the illumination problem is relatively few research, lack of efficient and practical algorithms.At present, the solutions to illumination problems include: first, seeking the underlying visual features that are insensitive to illumination changes to improve the recognition performance; second, establishing a illumination model to carry out targeted illumination compensation to eliminate the influence caused by non-uniform front illumination; third, using any illumination image generation algorithm to generate training samples under different illumination conditions.The face recognition method with good learning ability is used for recognition.In addition, it is also worth trying to solve the illumination problem by integrating various methods.But in general, the existing methods are not enough to solve the problem of the influence of light on face recognition.Further research is mainly focused on how to establish a general illumination model, extract the underlying visual features that are not sensitive to illumination changes, such as transverse features, wavelet transform coefficients, and establish a light model discrimination method based on learning strategies.

- Attitude estimation and matching: Due to the diversity of face posture, the face image obtained in the natural state is not always positive, positive face image is only an ideal recognition state at the same time, due to the deflection or pitch of the face will cause partial loss of facial information, which causes a certain degree of difficulty for the accurate face features. In addition, the estimation of deflection Angle during recognition,It is also an important factor affecting accurate matching.Therefore, in face recognition, we must consider the influence of attitude change on it, attitude problem is inevitable. At present, there are many researches on face recognition based on attitude changes, and a large number of research results have been produced. Typical solutions include the establishment of multi-pose face database and recognition through multi-text learning methods. Second, methods based on the invariant features of posture are sought, such as the recognition method based on elastic map matching and skin color model mentioned above. Automatic generation algorithm is used to automatically generate multi-angle view based on single view for recognition. With the development of face recognition technology, the application field of face recognition has been expanding in recent years. Many occasions require identity identification based on a single face

image.At present, the research on this aspect has just started, and there are few references. The main research includes a method based on three-dimensional reconstruction, such as using the method of inducing virtual view point, combining two-dimensional image and three-dimensional face model, to recognize the multi-pose face image.However, the 3D model must be accurate enough and the calculation is very complicated, so its effect is not ideal in practical application.The other is based on two-dimensional image generation, it does not need complex three-dimensional reconstruction, simple and feasible.At present, there are many researches based on this kind of method in China. For example, Chen Xilin, Gao Wen et al. proposed the multi-candidate class weighting method to realize multi-attitude recognition, and Liang Luhong et al. proposed the multi-attitude recognition method based on affine template matching.In general, multi-pose face recognition based on single view or small sample is a key problem urgently needed to be solved, which directly affects the practical process of face recognition.

4 Model Introduction

4.1 MTCNN Model

MTCNN(Multi-task convolutional neural Network) is a network proposed in Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks[19]. By using the convolutional neural network (CNN) structure, MTCNN can complete the two tasks of face detection and face alignment simultaneously through multi-task learning, and output the coordinate of the center point, the scale and the position of the feature point of the face.MTCNN adopts image pyramid + 3-stage cascade CNN (P-Net, R-Net, O-Net) for face detection, and its detection framework is shown in Figure 1：

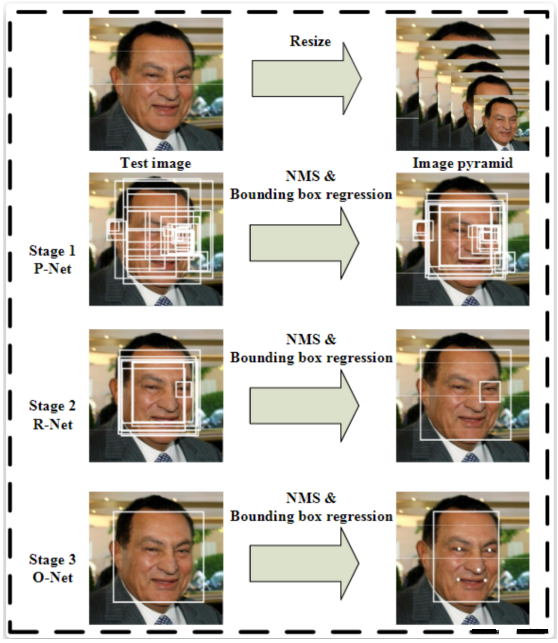

Figure 1: Pipeline of cascaded framework that includes three-stage multi-task deep convolutional networks.

Image pyramid transforms the image scale to detect faces on different scales.coarse to fine was tested by CNN in a series. The output of coarse to fine is the input of the latter, and the areas that are not faces were removed quickly. The areas that may contain faces were sent to a more complex network for further screening using more information. Figure 2 shows the three intermediate connected neural networks (P-Net, R-Net, O-Net) of MTCNN. The number of network layers of each layer of the network is gradually deepened, the receptive field of the input image is gradually enlarged, and the final output feature dimension is also increased, which means that more and more information is used.

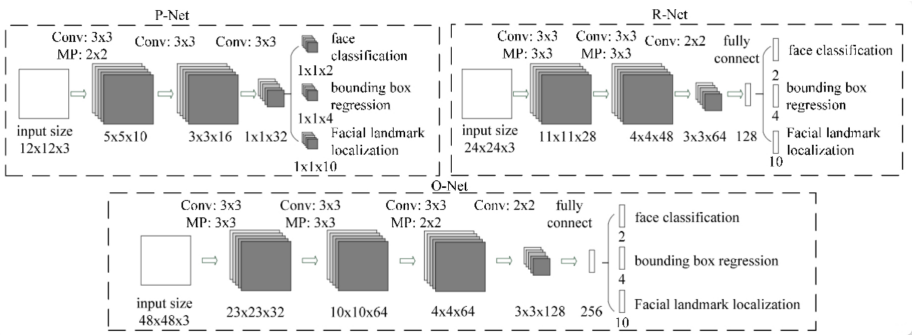

Figure 2: The architectures of P-Net, R-Net, and O-Net, where “MP” means max pooling and “Conv”means convolution. The step size in convolution and pooling is 1 and 2, respectively.

The three - tier network is explained as follows：

- P-Net: In fact, it is a full convolutional neural network (FCN), and the feature map obtained by forward propagation is a 32-dimensional feature vector at each position, which is used to judge whether the region of about 12×12 size at each position contains a face. If it contains a face, the Bounding Box of the face will be returned.Bounding Box corresponding to the region in the original image can be obtained. Non-maximum suppression (NMS) can be used to reserve Bounding boxes with the highest score and remove Bounding boxes with too large overlapping regions.

- R-Net: Is a simple convolutional neural network (CNN), the input is refined selection, and most of the wrong input is eliminated, and the border regression and facial feature point locator are used again to carry out the border regression and feature point location of the face region, and finally the more reliable face region will be output for O-Net.In contrast to P-Net, which uses 1132 features of full convolutional output, R-Net uses a 128 fully connected layer after the last convolutional layer, which retains more image features and has better accuracy performance than P-Net.

- O-Net: It is also a pure convolutional neural network (CNN). Bounding Box bilinear interpolation to 48×48, which is considered by R-Net to contain human face, is inputted to O-Net for face detection and feature point extraction.

In the training stage, the three networks will take the location of the key point as the supervision signal to guide the network learning, but in the prediction stage, P-Net and R-Net only detect the face and do not output the location of the key point, and the location of the key point is only output in O-Net.The output of each level of neural network is explained below:

- face classification uses softmax to judge whether it is a face, so the output is two-dimensional

- bounding box regression outputs offsets for the upper left and lower right corners, so the output is four-dimensional

- Faciallandmarklocalization positioning the left eye, right eye, nose, mouth left and right corners of the mouth, a total of five point position, so the output is ten-dimensional

4.2 FaceNet Model

FaceNet is a unified solution framework proposed by Google on FaceNet: A Unified Embedding for Face Recognition and Clustering to identify[20] (who is this), verify (whether it is the same person), and cluster (find the same person in multiple faces). FaceNet believes that all the above problems can be uniformly handled in feature space, and the difficulty lies in how to better map faces to feature space.Its essence is to learn the mapping of face image to 128-dimensional Euclidean space through convolutional neural network. This mapping maps face image to 128-dimensional feature vector, and uses the reciprocal of the distance between feature vectors to represent the similarity between face images.For different images of the same individual, the distance between the feature vectors is small, and for images of different individuals, the distance between the feature vectors is large.Finally, based on the similarity between feature vectors, face image recognition, verification and clustering are solved. The main process of FaceNet algorithm is as follows:

- The image is mapped to the 128-dimensional feature space (Euclidean space) by deep convolutional neural network, and the corresponding 128-dimensional feature vector is obtained
- The feature vector was regularized by L2 to screen out the effective features
- Using regularized feature vectors, Triplets Loss was calculated

Triplets means triplets, and unlike neural networks' two-parameter calculations (prediction tags and real tags), TripletLoss is calculated with three parameters. Triplet specifically refers to anchor,positive and negative parts, all of which are feature vectors after L2 regularization.Specifically, anchor and positive refer to the two matching thumbnail images of faces, in which anchor is the benchmark image in model training, positive refers to the image of an individual with the same personality as anchor, and negative refers to the image of an individual with different personality as anchor.

FaceNet uses deep convolutional neural network to learn mapping and further designs TripletsLoss to train the network. The reason why it is called triplet is that the loss function contains two matched face thumbnails and one unmatched face thumbnail. Its

goal is to distinguish the positive and negative classes in the sample by distance boundary.The face thumbnail is a cropped face image with no 2D or 3D alignment except for zooming and panning.

The purpose of FaceNet is to embed images of faces into a 128-dimensional Euclidean space.In this vector space, the distance positive between anchor and other feature vectors of the anchor was small, while the distance positive between Anchor and other feature vectors of the Anchor was large, as shown in Figure 3:

Figure 3: The Triplet Loss minimizes the distance between an anchor and a positive, both of which have the same identity, and maximizes the distance between the anchor and a negative of a different identity.

Using mathematical language, it is described as:

$$\|f(x_i^a) - f(x_i^a)\|_2^2 + \alpha < \|f(x_i^a) - f(x_i^n)\|_2^2, \forall (x_i^a, x_i^p, x_i^n) \in \mathbb{R}^{128}$$

That is, there is a boundary value such that the distance between all eigenvectors of any individual is always less than the distance between any eigenvectors of that individual and other individual eigenvectors. Further, Triplets Loss can be defined:

$$L_{loss} = \sum_i^N [\|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha]_+, \text{and } [x]_+ = \max\{0, x\}$$

The selection of Triplets is important for model convergence. For x_i^a,In practice, need to select different pictures of the same individual x_i^p,it satisfies $\arg\max_{x_i^p} \|f(x_i^a) - f(x_i^p)\|_2^2$,Also select pictures of different individuals x_i^n,it satisfies $\arg\max_{x_i^p} \|f(x_i^a) - f(x_i^p)\|_2^2$.

In actual training, it is not realistic to calculate the above two items across all training samples, and it will also cause difficulty in training convergence due to wrong labeled images.Therefore, two methods are commonly used to filter:

- Every n steps, x_i^p and x_i^n of the subsets are evaluated.
- Triplets are generated online, that is, positive and negative samples are selected in each mini-batch.

5 Summary

In recent years, OCR technology based on deep learning has significantly improved the accuracy and efficiency of face recognition.The current mainstream face recognition is mainly divided into two parts, namely face detection and face recognition. 1) In terms of face detection, there are relatively mature research results in traditional single-dimension face detection.Texture feature extraction algorithm LBP also has a good effect on multidimensional face detection. 2) In terms of face recognition, the mainstream models are CNN and transformer.In addition, some teams still apply the Attention mechanism to face recognition, which strengthens the ability of network mining characteristic information.At present, the accuracy of face recognition is still not high. Based on the existing deep learning model, this project will use the combination of MCTNN+facenet to achieve better recognition effect.

References

[1] 李志华, 张见雨, 魏忠诚. 基于 MTCNN 和 Facenet 的人脸识别系统设计 [J]. 现代电子技术, 2022(004) : 045.

[2] 史涛, 秦琴, 任红格. 基于区域分割 Haar-SIFT DBN 的人脸识别 [J]. 计算机仿真, 2019, 36(3) : 6.

[3] 刘俊, 王岩, 韩为选. 基于视频图像的人脸识别与跟踪 [J]. 电子技术与软件工程, 2019(11) : 1.

[4] ANON. 中科院自动化所" 面向公共安全的视觉物联网关键智能技术与产品" 项目通过验收 [J]. 信息网络安全, 2014(2) : 1.

[5] CRESCENZI M, GIULIANI A. The main biological determinants of tumor line taxonomy elucidated by a principal component analysis of microarray data[J]. FEBS Letters, 2001, 507.

[6] IBRAHIM W, ABADEH M S. Protein fold recognition using Deep Kernelized Extreme Learning Machine and linear discriminant analysis[J]. Neural Computing and Applications, 2019, 31 : 4201 – 4214.

[7] COMON P. Independent component analysis, A new concept?[J]. Signal Process., 1994, 36 : 287 – 314.

[8] SAMARIA F. Face recognition using Hidden Markov Models[C] // . 1995.

[9] KOHONEN T. Self-organization and associative memory: 3rd edition[C] // . 1989.

[10] SISODIYA A S. Reducing Dimensionality of Data Using Neural Networks[J], .

[11] SUN Y, WANG X, TANG X. Deep Learning Face Representation from Predicting 10,000 Classes[J]. 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014 : 1891 – 1898.

[12] HUANG G B, MATTAR M A, BERG T L, et al. Labeled Faces in the Wild: A Database forStudying Face Recognition in Unconstrained Environments[C] // . 2008.

[13] TAIGMAN Y, YANG M, RANZATO M, et al. DeepFace: Closing the Gap to Human-Level Performance in Face Verification[J]. 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014 : 1701 – 1708.

[14] SCHROFF F, KALENICHENKO D, PHILBIN J. FaceNet: A unified embedding for face recognition and clustering[J]. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015 : 815 – 823.

[15] MASI I, RAWLS S, MEDIONI G G, et al. Pose-Aware Face Recognition in the Wild[J]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016 : 4838 – 4846.

[16] 周克辉, 罗玮, 陈泰峰. 高校学生考勤管理系统发展现状和需求分析研究 [J]. 电子测试, 2019(22) : 3.

[17] 曲爱玲, 马长路, 刘红梅, et al. 基于 ARM 与条码技术的学生考勤系统设计 [J]. 北京农业职业学院学报, 2017, 31(5) : 5.

[18] 胡韶山, 刘培进, 聂兵荣. 一种基于虹膜识别的小区用人员实名制信息管理系统 [EB]. .

[19] ZHANG K, ZHANG Z, LI Z, et al. Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks[J]. IEEE Signal Processing Letters, 2016, 23(10) : 1499 – 1503.

[20] SCHROFF F, KALENICHENKO D, PHILBIN J. FaceNet: A Unified Embedding for Face Recognition and Clustering[J]. IEEE, 2015.