Journal of Computers Vol. 34 No. 2, April 2023, pp. 29-39
doi: 10.53106/199115992023043402003

# Unrestricted Face Recognition Algorithm Based on Improved Residual Network IR-ResNet-SE

Hong-Rong Jing[1], Guo-Jun Lin[2*], Tian-Tian Chen[1], Hong-Jie Zhang[1], Long Zhang[3], Shun-Yong Zhou[2]

[1] Department of Automation and Information Engineering, Sichuan University of Science & Engineering,
Yibin 633000, Sichuan, China
{2366464126, 1450935593, 236909586}@qq.com

[2] Artificial Intelligence Key Laboratory of Sichuan Province Sichuan University of Science & Engineering,
Yibin 633000, Sichuan, China
{386988463,14409186}@qq.com

[3] Chengdu University of Information Technology,
Chengdu 610103, Sichuan, China
1744811259@qq.com

Received 5 May 2022; Revised 25 August 2022; Accepted 19 September 2022

**Abstract.** To solve the problem of poor face recognition performance in unrestricted environments. A face recognition algorithm based on improved residual IR-ResNet-SE is designed. Firstly, the IR structure is added to the 34-layer residual network to reduce the variability of different features; Secondly, we add the channel attention module to increase the weight of important channel features; Finally, the Arcface loss function is used to improve the classification ability of the model. The LFW, AgeDB, and AR datasets reflect unrestricted factors such as pose, age, expression, occlusion, and illumination. The algorithm proposed in this paper is experimented on these three datasets. The experimental results show that the IR-ResNet-SE algorithm proposed in this paper can achieve 99.74% accuracy in the dataset LFW. And it has excellent robustness in face recognition under unrestricted conditions.

**Keywords:** residual network, SE attention, face recognition, Arcface

## 1 Introduction

Face recognition has many potential applications [1]. Faces are widely used in security、finance、research and other fields. Face recognition has made an enormous leap in the past five years [2]. Previously, face recognition used classical methods to extract features and classification, such as LBP [3] and SVM [4]. These methods can achieve good results when the number of samples was small. However, with the increase of face data and the emergence of unrestricted face datasets, classical methods can no longer meet our requirements. Convolutional neural network (CNN) is a powerful classification method commonly used for image recognition and verification [5]. Face recognition technology [6] is developing fast and it has gradually come out of the laboratory and gradually entered our life. Compared with the traditional face recognition methods, face recognition based on deep learning has a tremendous advantage in recognition accuracy. This shows the importance of deep learning for face recognition [7]. And in the case of difficulties in further upgrading and optimization of network structure, researchers gradually turned their attention to the field of loss function and attention network [8]. In 2014, the proposal of DeepFace [9] network made significant progress in the direction of face recognition. Because It was the first time that deep learning was applied to face recognition, the accuracy was 97.35% in the public dataset LFW. In the same year, the University of Hong Kong developed the DeepID [10] algorithm, which has a simple structure and achieved an accuracy of 98.60% on the LFW face recognition dataset. Then, based on the original algorithm, DeepID2 [11], DeepID2+ [12], and DeepID3 [13] algorithms were developed, and their accuracy rates in LFW were 98.95%, 97.70%, and 99.47%. Although these algorithms improve recognition accuracy to a certain degree, the robustness of algorithms are not ideal enough for unrestricted datasets. 2015, Google created the FaceNet [14] model. The model was trained using Triplet Loss and achieved an accuracy of 98.70% on LFW. However, the triplet of the model is difficult to construct, and the calculation is large. In addition to proposing a

* Corresponding Author

generic network structure, a series of loss functions have been created. The center loss [15] only constrains the intra-class distance, not the inter-class distance, and is usually used together with Softmax loss, achieving 99.05% accuracy on LFW. From 2017 onwards, researchers have created Margin-based classification loss functions, such as SphereFace [16], which normalizes only the features. The advantage of this algorithm is that the face features are distributed between hyperspheres with high cohesion and low coupling, and the disadvantage is that the loss function is difficult to optimize. The algorithm has an accuracy of 99.42% on the LFW dataset. In 2018, CosFace [17] performed normalization operations on features and weights, but the disadvantage is that it needs to segment the optimization function and the training is unstable, and the accuracy rate in the LFW dataset is 99.33%. In 2019, ArcFace has the advantage of easy classification using normalized hyperspherical optimized embedding features, and the accuracy in LFW dataset is 99.53%. In 2020, Circle Loss [18] is able to learn good training distributions, but requires high GPU performance and has a latest accuracy of 99.68% in the public dataset LFW. In 2021, MagFace [19] is a generic representation for face recognition and quality assessment. Recently,In order to further improve the performance of face recognition, many attention modules have been gradually applied to the face recognition field. In general, face recognition is divided into two types: under restricted conditions and unrestricted conditions. In restricted conditions, face data are mostly recognized as a single gray headshot in a single background with a frontal pose. Under unrestricted conditions, face data are mostly collected from the proper environment and integrated with various influencing factors such as face pose, lighting, expression, occlusion, etc., which have more practical application value. However, in practical application, face recognition becomes particularly difficult due to the complex interference of illumination, expression, pose and occlusion in the face images under unrestricted conditions.

In this paper, the algorithm incorporates IR structure in the residual block to reduce the difference of face features. To improve the recognition accuracy, a channel (SE) attention mechanism is inserted in IR-ResNet. The Softmax function is replaced by the Arcface function. The IR-ResNet-SE network based on 34 layers is designed in this paper, and experiments are conducted under the same conditions. Excellent results are achieved in unrestricted datasets such as LFW, AgeDB, AR (occlusion), AR (illumination, expression). It is a suitable solution to the problem of unrestricted datasets, poor performance of face recognition.

The paper is structured as follows. The first section is an introduction. The second Section is the related work. This section includes three parts. The first part is the principle of residual network, the second part is the introduction of batch normalization, and the third part is the introduction of SE attention. The third section is the introduction of loss function. This section includes two parts: Softmax and Arcface. the fourth section is the IR-ResNet-SE network model. The fifth section is the experiment. The sixth section is the conclusion and the last section is the references.

## 2 Related Work

### 2.1 The Residual Network

ResNet is a network model proposed by Kai-ming He [20] in 2015. When the depth of the network is directly increased in order to improve the accuracy. It causes 2 problems. problem 1 is the diffusion or explosion of the network gradient, and problem 2 is the decrease of the accuracy. Problem 2 is not caused by over-fitting, but by saturation or even a decrease in accuracy. For problem 1, its solution uses BatchNorm [21]. For problem 2, to solve the performance degradation caused by increasing the number of convolution layers, a residual learning mechanism is used. H(x) is used to represent the actual mapping that is expected to be obtained. A stacked multilayer nonlinear network is used to represent the fitting of the mapping relationship. Then the multilayer network can gradually approximate some complex function. Suppose it is equivalent to approximating the residual function F(x) , where x is the input of the first of these layers and F(x) denotes the residual function. Then the actual mapping relationship can be expressed as:

$$H(x) = F(x) + x \; . \tag{1}$$

The advantage of using residual blocks is that there are no redundant parameters and no increase in computational complexity. The following figure shows the most basic residual mechanism. The idea is to assume that the models are mapped to each other constantly, and only the constant mapping function needs to be solved.

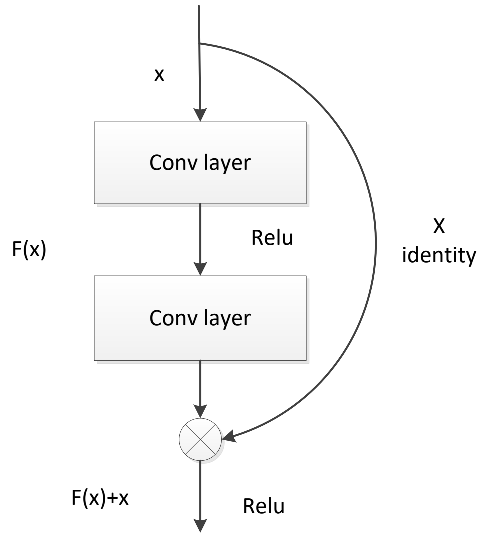



**Fig. 1.** Residual block

It is difficult to find H(x) directly, so the residual unit is used to solve it by short-cut connections. We can achieve this by adding short-cut connections to the forward neural network. The output of each layer is not a mapping of the traditional neural network input, but a superposition of the mapping and the input. The residual network consists of a superposition of residual blocks, and the general construction of the residual blocks is shown in Fig. 1. The residual learning mechanism has been shown to be used in ResNet networks to train ultra-deep convolutional neural networks. The accuracy is better than other convolutional neural networks, and the deeper the network model level, the better the accuracy.

### 2.2 Batch Normalization

Batch Normalization can effectively solve the problem that the convergence speed decreases rapidly, and the gradient disappears or explodes as the number of layers of the network deepens. And also can significantly improve the training speed. BN layers are a way to normalize the activation of a neural network, which can easily change at each layer due to the increase in layers and nonlinear units. It keeps the activation values of each layer equally distributed while keeping the network capacity constant. This effectively increases the training speed and solves the gradient disappearance or explosion problem. Batch Normalization (BN layer for short) is a data pre-processing method. The BN layer has the advantage of eliminating variability, reducing the interference of useless data, and accelerating the convergence of the network. By adjusting the distribution of inputs to each layer, the stability of the learning process in the deep network is increased and smoother optimization surfaces are generated.The BN algorithm is formulated as follows. $x_i$ is the i-th sample in the mini-batch, and N is the size of the mini-batch, $\gamma$, $\beta$ is the hyper-parameter that can recover the normalized data.

Mini-batch mean:

$$\mu_\beta \leftarrow \frac{1}{N}\sum_{i=1}^{N} x_i \quad . \tag{2}$$

Mini-batch variance:

$$\sigma_\beta^2 \leftarrow \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \mu_\beta\right)^2 \quad . \tag{3}$$

Normalize:

$$\hat{x}_i \leftarrow \frac{x_i - \mu_\beta}{\sqrt{\sigma_\beta^2 + \in}} \ . \tag{4}$$

Scale and shift:

$$y_i \leftarrow \gamma \hat{x} + \beta \equiv BN_{\gamma,\beta}(x_i) \ . \tag{5}$$

### 2.3 SE Attention Mechanism

CNN are commonly used in fields such as face recognition, but they have a great disadvantage. CNN only focuses on local information and ignores global information. For example, they ignore spatial and channel information. So, changing this drawback, the attention mechanism is used. In CVPR 2018 conference, there is a paper from Jie Hu [22], a senior R&D engineer, talks in detail about the features of SENet network. The SE (Squeeze-and-Excitation) module, the primary function of this module is to assign weights to each channel, which functions the same as the attention mechanism [23], helps the network to bring important feature information to be learned. The SE block inside the paper is not a complete network structure, but a substructure that can be combined with other classification models. The authors inserted the SENet block into several classification networks and all achieved excellent results. The core idea of SENet is to learn the feature weights according to the loss of the network. So that the effective feature map weight is high and the invalid or small effect feature map weight is reduced to train the model in a way that can achieve better results. Of course, the SE block inserted into some classification networks inevitably increases some parameters and computation. But the improved performance is excellent and can accept the network inserted into the SENet.

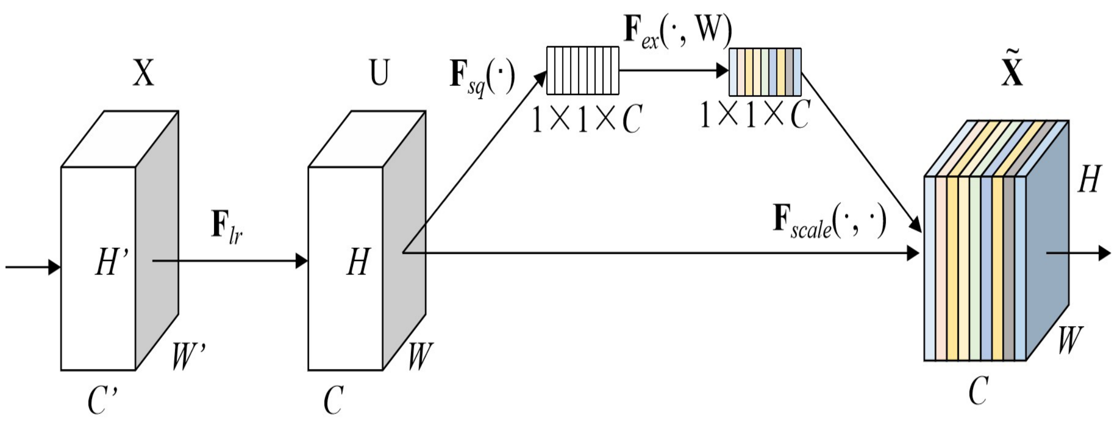



**Fig. 2.** SE attention mechanism

SENet has three operations:

The first is the Squeeze operation, which compresses the features along the spatial dimension, and the output dimension matches the number of channels of the input features. Squeeze operation is to compress the image with the input WxHxC dimensions into 1x1xC dimensions using the maximum average pooling operation, which is then used as the input for the Excitation operation. C is the number of channels.

Next is the Excitation operation, which is a mechanism similar to the gates in recurrent neural networks. The

Excitation operation is intended to take the output of the Squeeze operation as an input. The first fully connected layer acts as a dimensionality reduction with an attenuation factor of 16, and then is raised to 1x1xC by another fully connected layer. The Sigmoid activation function is used to compress the weights between 0 and 1. The effect of the two fully connected layers is to increase the nonlinearity of the network and to reduce the computational effort and the number of parameters.

The last operation is Reweight, which takes the weights of the output of Excitation as the importance of each feature channel after feature selection. Then re-calibrates the original features in the channel dimension by multiplying the weights channel by channel to the previous features (Fig. 2). The algorithm designed in this paper consists of the residual block in Fig. 3. The residual block of Fig. 3 incorporates the IR structure for the feature normalization operation. This helps to reduce the variability of unrestricted face features. After the normalization operation for the last convolutional layer, SE attention is inserted. Fig. 3 shows the residual block with the SE attention mechanism inserted into the IR-Resnet.

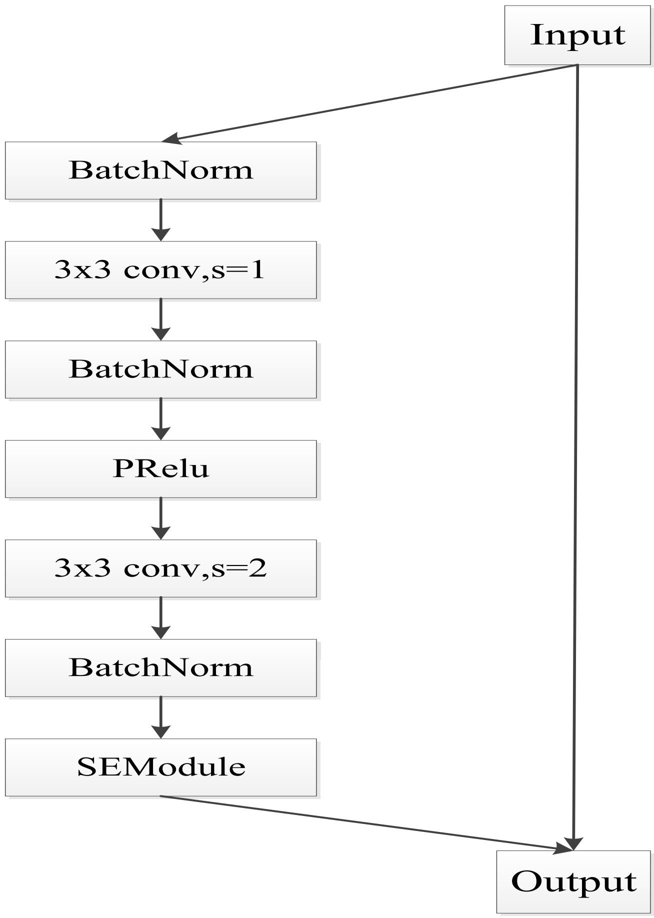



**Fig. 3.** Block of IR-ResNet inserted into SE

## 3 Loss Functions

### 3.1 Softmax Function

We often used Softmax function as the final classifier in convolutional neural networks for multi-classification tasks. Softmax belongs to supervised learning and can be combined with deep learning algorithms to achieve better results. Softmax maps the outputs of multiple neurons to the (0, 1) interval and all output values sum to

1. Softmax is a monotonic increasing function with the computation is simple and the effect is significant. The Softmax loss function ensures good separability between classes, but the intra-class distance of features is scattered over a large range. The disadvantage is that the intra-class features are not compact enough, and even the intra-class features are more spaced from each other than from each other. Its functional form is as follows:

$$\text{Loss} = -\frac{1}{\text{N}}\sum_{i=1}^{N}\log\left(\frac{e^{w_{y_i}^{T}x_i+b_{y_i}}}{\sum_{j=1}^{n}e^{w_{y_j}^{T}x_i+b_j}}\right). \tag{6}$$

The symbol $w_{y_j}^{T}x_i+b_j$ is the output of the fully connected layer. For loss to fall, the value of $w_{y_j}^{T}x_i+b_j$ must be increased. So that all faces belonging to this class of samples are inside this type of decision boundary. Softmax mainly considers whether it can classify correctly and lacks constraints on intra- and inter-class distances. The ability to classify similar, not similar faces is particularly poor. The experimental results of this paper [24] show that the traditional Softmax still has a large intra-class distance. By improving the loss function, the constraint on the intra-class distance can be increased and can improve the performance of the network.

### 3.2 Arcface Function

Softmax mainly considers whether the samples can be classified correctly or not, and its drawback is that it does not restrict the intra-class and inter-class distances. By observing the relationship between weights and class centers, this epoch-making paper—Sphereface is proposed. The important concept of angle margin is introduced, but it requires some approximate computation, thus leading to unstable training. Therefore, an additive angular margin loss-named Arcface [25]-is proposed, which allows the intra-class distance to be reduced and the inter-class distance to be increased. Thus improving the recognition ability of the trained model. The dot product between the features extracted by the convolutional neural network and the last fully connected layer is equal to the cosine distance after normalization of the features and weights. We use the arc-cosine function to calculate the angle between the current features and the weights. An additional angle margin m is added to the target angle. To facilitate the calculation, so let the bias $b_j$ equal to 0. The inner product of the weights and the input features are expressed as follows:

$$W_j^T x_i = \| W_j \| \| x_i \| \cos \quad _j . \tag{7}$$

Regularize the weights with the features L2 processing, $\|W_j\|=1$ and $\|x_i\|=1$, $\theta_j$ is the angle of the weights $W_j$ with the features $x_i$. Thus, the learned to embed features are distributed on a hypersphere of radius s. Since the embedding features are distributed around the center of each feature on the hypersphere. We add an additional angle m between the weights and the features. The advantage of this is that it corresponds to simultaneously enhancing the intra-class tightness and expanding the inter-class variance. The focus is on maximizing the classification bounds directly in the angular space. The loss function of Arcface is (s = 64, m = 0.5):

$$\text{Loss} = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{s(\cos(\theta_{y_i}+m))}}{e^{s(\cos(\theta_{y_i}+m))}+\sum_{j=1,j\neq y_i}^{n}e^{s\cos\theta_j}}. \tag{8}$$

N: batch size
n: number of categories
s: radius of hypersphere
m: angle margin

## 4 Improved Residual IR-ResNet-SE Algorithm

Due to the complexity of the dataset in this paper, it contains many faces under unrestricted conditions. Such as expression changes, local occlusion, and lighting changes, a model with stronger classification ability and superior performance is required. The classical 34-layer residual network is chosen. Meanwhile, the model capacity can produce serious overfitting phenomenon, and the BN layer with Dropout layer is used to optimize the model. To improve the recognition accuracy, SE attention is inserted. The output of the Average Pool layer is taken as the depth feature and the output is used as the input of Arcface loss function for classification. Table 1 First Conv operation followed by Max Pool operation is used. [Block1] X3 is having three such Blocks1, and the structure of each Block is shown in Fig. 3. The Batch Normalization operation is used before the convolution operation inside the Block. The Batch Normalization operation is added after the convolution operation to improve the training speed and solve the problem of gradient explosion. The activation function uses PReLU function to improve the recognition accuracy, and the FC layer is preceded by Average Pool and Dropout operations. Finally, Arcface function will be used instead of Softmax to improve the classification accuracy.

**Table 1.** IR-ResNet-SE structure

| Layer | Kernel size | Stride | Output size |
|---|---|---|---|
| Conv, Max Pool | 7 x 7 | 2 | 112x122x64 |
| [Block1 ]X3 | 3 x 3 | 1 | 56x56x64 |
| | 3 x 3 | 2 | 56x56x64 |
| [Block2 ]X4 | 3 x 3 | 1 | 28x28x128 |
| | 3 x 3 | 2 | 28x28x128 |
| [Block3 ]X6 | 3 x 3 | 1 | 14x14x256 |
| | 3 x 3 | 2 | 14x14x256 |
| [Block4 ]X3 | 3 x 3 | 1 | 7x7x512 |
| | 3 x 3 | 2 | 7x7x512 |
| Average Pool, Dropout | | | 1x1x512 |
| FC, Arcface | | | 1x1x512 |

## 5 Experimental Results and Analysis

### 5.1 Introduction to The Dataset

**Fig. 4.** Partial face dataset

The LFW dataset is widely used for face recognition. It includes about five thousand individuals with a total of 13,000 face images. The CASIA-WebFace dataset is a large-scale face dataset. It contains 494,414 images of 10,575 individuals. CASIA-WebFace is used as the training set for the residual network in the experiments of this paper. The AR face database has four thousand face images of 126 individuals. The database contains not only 4 basic expressions of natural, joy, annoyance and surprise, and face images with different lighting conditions, but also partially obscured face images with sunglasses or scarves. The AgeDB dataset is a dataset of the same person's age change. There are about 16,000 face images. The age range of the same person is between the teens and the eighties, and the average age inside each class is 30 years (Fig. 4).

### 5.2 Experimental Setup

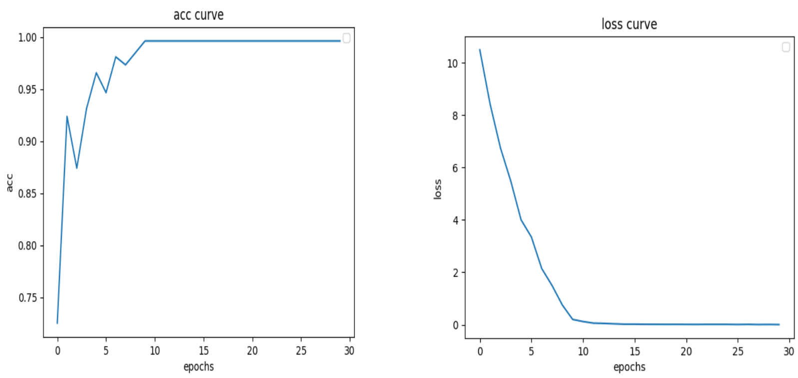
This paper is based on a deep learning pytorch framework, python 3.7, running in windows environment. When the network starts training and testing set the BS value to size 256; learning rate 1/e; Dropout=0.6. The optimizer uses stochastic gradient descent (SGD), and the training loss function uses focal_loss. The CASIA-WebFace dataset is used as the training set for LFW, and the training and testing of AgeDB and AR are divided according to 8: 2 ratio. To test the effectiveness of the algorithm, they did the following comparison experiments in this paper. Table 2 shows the experimental details.

**Table 2. Experimental detail settings**

| Experiment | Experimental algorithms |
|---|---|
| 1 | DeepFace + Softmax |
| 2 | FaceNet + Triplet Loss |
| 3 | Resnet34 + BN + Dropout + Softmax |
| 4 | IR-Resnet34 + BN + Dropout + Softmax |
| 5 | IR-Resnet34 + BN + Dropout + Arcface |
| 6 | IR-Resnet34 + BN + Dropout + SE+Arcface |

### 5.3 Experimental Analysis

Set each iteration one hundred times as an epoch. Analysis of Fig. 5(b), the loss drop plot of Fig. 5(b) echoes the CASIA-WebFace training set accuracy increase of (a). The large initial value of the loss plot in Fig. 5(b) is due to the fact that the loss value at this point is composed of two parts, Arcface and Softmax loss functions. The reason for the small fluctuations in the loss map in Fig. 5(b) is that it is being at some local optimum at the time of training. As the training time goes on, the loss gradually decreases as the model continues to train out of its influence. At this point, there is a flat change in the loss value and training accuracy that is not rising. In experiment 6 (IR-ResNet-SE algorithm) the training accuracy is 99.90%.




(a) Training accuracy (b) Loss curve

**Fig. 5.** Experimental training accuracy and loss

We use 2 evaluation metrics - confusion matrix and accuracy rate. Each column of the confusion matrix represents the prediction category, and each column sums to 1. Each row represents the true attribution category of the data. To extract more information about the model performance, the confusion matrix can be used. The confusion matrix helps us to identify whether the model is "confused" when distinguishing between two categories. The labels in the two rows and two columns are the true label and the predicted label, respectively.

The four elements of the confusion matrix represent the four indicators that calculate the number of correct and incorrect predictions of the model. The four indicators in the confusion matrix are: top-left (TP); top-right (FN); bottom-left (FP); and bottom-right (TN). The accuracy rate can be calculated by these four indicators.

$$\text{Accuracy:} \qquad ACC = \frac{TP + TN}{TP + TN + FN + FP} \,. \tag{9}$$

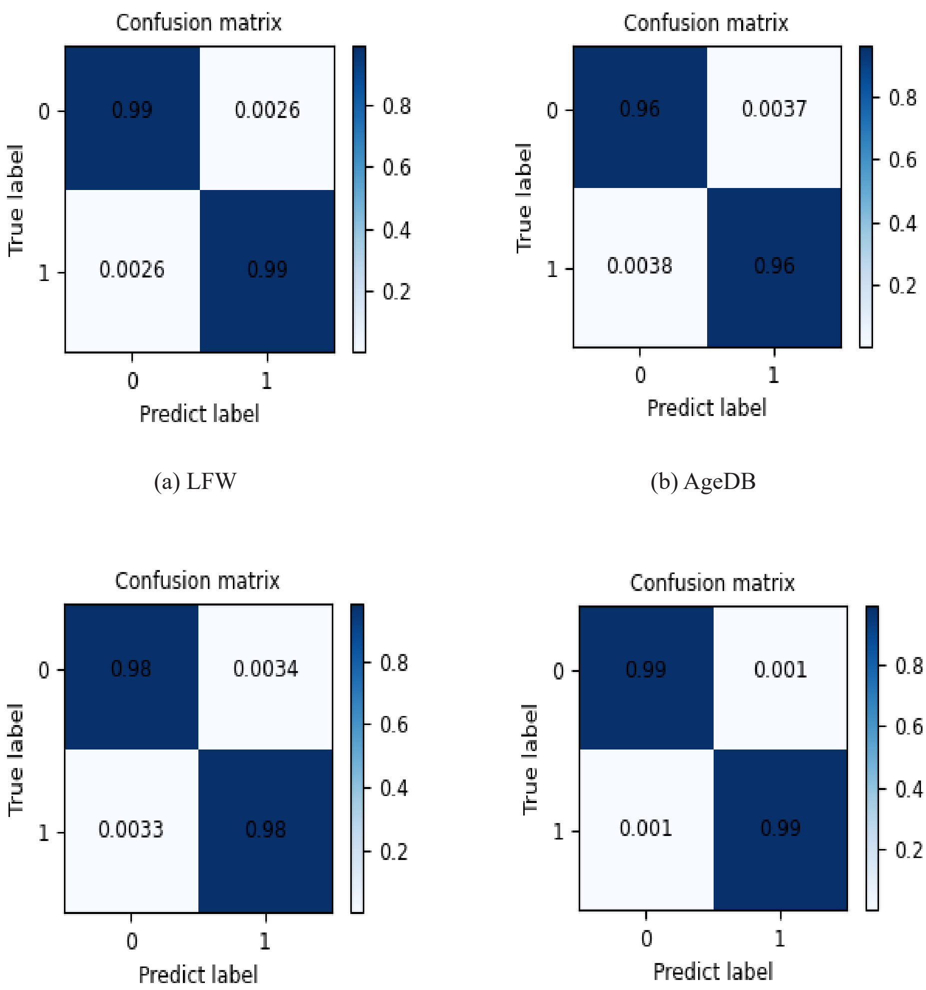
The number of equation (9) represents the following meaning.

TP: the number of positive classes predicted as positive classes.

FN: the number of positive classes predicted to negative classes.

FP: the number of negative classes predicted to positive classes.

TN: the number of negative classes will be predicted as negative classes.




(a) LFW (b) AgeDB

(c) AR (occlusion) (d) AR (illumination, expression)

**Fig. 6.** Confusion matrix

The values inside the confusion matrix are normalized to facilitate the calculation of accuracy. The confusion matrix can display the results of a classifier. The accuracy of the classification result can be represented in the confusion matrix. The four confusion matrices in Fig. 6 are the results obtained by the algorithm IR-ResNet-SE on three unrestricted datasets, LFW, AgeDB, and AR. The smaller value of the subdiagonal of the confusion matrix represents the higher accuracy.

**Table 3. Results of the comparison experiments**

| Experiment | LFW | AgeDB | AR (occlusion) | AR (illumination, expression) |
|---|---|---|---|---|
| 1 | 97.35% | 89.965 | 91.43% | 98.24% |
| 2 | 98.70% | 91.33% | 93.56% | 99.18% |
| 3 | 99.12% | 94.79% | 96.89% | 99.29% |
| 4 | 99.14% | 94.75% | 96.78% | 99.34% |
| 5 | 99.66% | 96.20% | 98.26% | 99.76% |
| 6 | 99.74% | 96.63% | 98.67% | 99.90% |

The experimental results in Table 3 are all derived under the same conditions and parameters. This experiment uses the face dataset under unrestricted conditions. The LFW dataset features faces with different poses, which can be used well for face recognition under pose conditions. The AgeDB dataset features faces with age changes, which can be well used for face recognition under age conditions. The AR has light and expression changes, which can be well used for face recognition under expression and light conditions. Experiment 1 and Experiment 2 are both deep learning based face recognition algorithms. The accuracy in three unrestricted datasets, much lower than the results of Experiment 3, and also lower than the accuracy of this paper's algorithm IR-ResNet-SE. Experiment 3 uses batch normalization and Dropout inside the 34-layer residual network, which can effectively reduce over-fitting, and has an accuracy of 99.12% in the LFW dataset. The latest accuracy of the 34-layer residual network in the Circle Loss paper is 99.18% in the LFW dataset. Replacing the Softmax accuracy with the Arcface loss function improves the accuracy by 0.52%. This is because the Softmax loss function lacks constraints on the intra-class, while the Arcface loss function can improve the classification ability of the network. Adding the IR structure can improve the accuracy to some extent in data sets with little intra-class variation, but data sets with large intra-class variation, such as AgeDB, will reduce the accuracy instead. The comparison between experiment 5 and experiment 6 shows that after adding the SE attention mechanism, there is an improvement of more than 0.08% on LFW, AgeDB, and AR datasets. The above experimental results show that the IR-ResNet-SE algorithm designed in this paper can achieve higher accuracy on the unrestricted face dataset without adding the SE attention mechanism compared with the residual network of Arcface function.

## 6 Conclusion

The algorithm in this paper increase the accuracy of face recognition under unrestricted conditions. Inserting the SE attention mechanism in IR-Resnet can increase the effective feature weights and decrease the invalid feature weights. Meanwhile, the Arcface loss function replaces the Softmax loss function in the fully connected layer. Arcface can reduce the intra-class and expand the inter-class distances of the same class and improve the classification ability of the face recognition model. The combination of the two achieves high accuracy in face recognition under unrestricted conditions. Of course, there are some minor problems, and the reasonable setting of hyperparameters requires a lot of experiments. For example, how to choose hyperparameters needs to be studied in depth. In addition, the recognition effect is not satisfactory in the presence of occlusions and large changes in age, for example, the accuracy rate of the AgeDB dataset is lower than LFW. The algorithm in this paper is not ideal for age and occlusion recognition, and the algorithm needs further improvement. Our future research will use GAN networks to solve the problem of different pose and age variations to improve the robustness of face recognition algorithms in real-world scenarios.

## 7 Acknowledgement

Sichuan Science and Technology Program of China (2020YFSY0027); The Scientific Research Foundation of Sichuan University of Science and Engineering under Grant 2019RC11 and Grant 2019RC12; The artificial intelligence key laboratory of Sichuan province Foundation (2020RZY02); The Opening Project of Key Laboratory of Higher Education of Sichuan Province for Enterprise Informationalization and Internet of Things (2021WYY01).

## References

[1] L. Li, X. Mu, S. Li, H. Peng, A review of face recognition technology, IEEE Access 8(2020) 139110-139120.
[2] X. Han, H.-Y. Zhang, Y.-Y. Zhang, Facial expression recognition based on high efficient channel attention network, Transducer and Microsystem Technologies 40(1)2021 118-121.
[3] T. Ojala, M. Pietikainen, T. Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Transactions on pattern analysis and machine intelligence 24(7)(2002) 971-987.
[4] J. Platt, Sequential minimal optimization: A fast algorithm for training support vector machines, Technical Report MSR-TR-98-14, 1998.
[5] A. Alzu'bi, F. Albalas, T. AL-Hadhrami, L.B. Younis, A. Bashayreh, Masked Face Recognition Using Deep Learning: A Review, Electronics 10(21)(2021) 2666.
[6] C.-C. Yu, H.-B. Li, Deep Learning Based 2D Face Recognition: a Survey, Chinese Journal of Engineering Mathematics 38(4)(2021) 451-469.
[7] H.-R. Jing, H.-J. Zhang, T.-T. Chen, L. Chen, Q. Hu, Research on Face Recognition Method Based on Deep Learning, International Core Journal of Engineering 7(7)(2021) 165-170.
[8] M. Wang, W. Deng, Deep face recognition: A survey, Neurocomputing 429(2021) 215-244.
[9] Y. Yan, H. Choi, H. Suk, Exploring the Facial Color Representative Regions Using the Humanae Images, Journal of Imaging Science and Technology 64(6)(2020) 060406.1-060406.7.
[10] S.J. Oh, R. Benenson, M. Fritz, B. Schiele, Person Recognition in Personal Photo Collections, IEEE Transactions on Pattern Analysis and Machine Intelligence 42(1)(2020) 203-220.
[11] Y. Sun, Y. Chen, X. Wang, X. Tang, Deep learning face representation by joint identification-verification, in: Proc. Advances in neural information processing systems, 2014.
[12] Y. Sun, X. Wang, X. Tang, Deeply learned face representations are sparse, selective, and robust, in: Proc. the IEEE conference on computer vision and pattern recognition, 2015.
[13] S. Saypadith, S. Aramvith, Real-time multiple face recognition using deep learning on embedded GPU system, in: Proc. 2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2018.
[14] F.-W. Gu, J. Lu, G.-H. Xia, Face recognition method based on facenet Pearson discrimination network, CAAI Transactions on Intelligent Systems 17(1)(2022) 107-115.
[15] L.-Z. Wang, H.-C. Lai, Face Recognition Based on Multi-Task Cascade Convolution CNN and Center Loss Function, Computer Simulation 37(8)(2020) 398-403.
[16] X. Zhou, Z. Chen, Q. Guo, C. Liu, W. He, Lightened SphereFace for face recognition, Journal of electronic imaging 29(6)(2020) 063010.1-063010.14.
[17] F.-Z. Zeng, L. Zou, Y. Zhou, T.-D. Qiu, J.-W. Chen, Application of Lightweight GAN super-resolution image reconstruction algorithm in real-time face recognition, Journal of Chinese Computer Systems 41(9)(2020) 1993-1998.
[18] Y. Sun, C. Cheng, Y. Zhang, C. Zhang, L. Zheng, Z. Wang, Y. Wei, Circle loss: A unified perspective of pair similarity optimization, in: Proc. the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020.
[19] Q. Meng, S. Zhao, Z. Huang, F. Zhou, Magface: A universal representation for face recognition and quality assessment, in: Proc. the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021.
[20] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proc. the IEEE conference on computer vision and pattern recognition, 2016.
[21] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, in: Proc. International conference on machine learning. PMLR, 2015.
[22] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proc. the IEEE conference on computer vision and pattern recognition, 2018.
[23] S. Woo, J. Park, J.-Y. Lee, I.S. Kweon, Cbam: Convolutional block attention module, in: Proc. the European conference on computer vision (ECCV), 2018.
[24] Y. Wen, K. Zhang, Z. Li, Y. Qiao, A discriminative feature learning approach for deep face recognition, in: Proc. European conference on computer vision, 2016.
[25] J. Deng, J. Guo, N. Xue, S. Zafeiriou, Arcface: Additive angular margin loss for deep face recognition, in: Proc. the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019.