

# A Face Recognition Algorithm Based on Improved Resnet

Hongrong Jing<sup>1</sup>, Guojun Lin<sup>1,2,\*</sup>, Hongjie Zhang<sup>1</sup>, Tiantian Chen<sup>1</sup>

<sup>1</sup> Department of Automation and Information Engineering, Sichuan University of Science & Engineering, Yibin 633000, Sichuan, China

<sup>2</sup> Artificial Intelligence Key Laboratory of Sichuan Province, Sichuan University of Science & Engineering, Yibin 633000, Sichuan, China

\* Corresponding author: Guojun Lin (Email: 386988463@qq.com)

**Abstract:** Regarding the problem that the increasing number of layers of CNN (convolutional neural network) leads to the decline of accuracy, an improved loss function algorithm based on the Resnet-50 model is proposed. The Softmax loss function lacks constraints on the distance within the same class and between different classes. Replacing the Softmax layer with improved Arcface loss enables the neural network to learn more distinguishing features. Experiments on LFW and AgeDB data sets show that the algorithm can not only learn deep-face characteristics but also efficiently improve the accuracy of face recognition compared with ordinary CNN. In the meantime, the improved Resnet also obtains a higher discerning rate under the conditions of occlusions, illumination, expression, Age.

**Keywords:** Deep learning; Residual; Face recognition; Softmax; Improved loss function.

## 1. Introduction

In recent decades, traditional face recognition methods are used to extract features and classify, such as LBP [1] and SVM [2]. In the case of a small number of samples, this kind of method has a good result, but with the increase of face data, the traditional methods are far from meeting the requirements. With the vigorous development of computer vision and deep learning algorithms in recent years, especially the development of CNN, a powerful classification method commonly used to recognize and verify images [3], face recognition technology [4] has also developed rapidly and gradually stepped out to the life. Compared with the traditional face recognition methods, the face recognition algorithm based on deep learning has better recognition accuracy. Thus, it can be seen that deep learning plays a great role in face recognition [5]. In view of the convenience, uniqueness, and non-repeatability of human faces, face recognition technology is widely used in many aspects of society, such as security, finance, scientific research, and so on. Face recognition has become the future development direction with many potential application prospects [6]. Over the past five years, it has made a qualitative leap [7]. Since the Resnet was put forward in 2015, more and more excellent algorithms based on Resnet have been proposed and good achievements have been made in the field of face recognition. Under the situation that it is difficult to further upgrade and optimize the network structure, researchers gradually turn their attention to the field of loss function and attention network [8]. Starting from the improved loss function, this paper uses a residual network different from the original algorithm for face recognition. The experimental results show that the improved loss function increases the accuracy of face recognition.

## 2. Principle

### 2.1. Residual Principle

Resnet is a network model raised by The Kaiming [9] in 2015. Directly increasing the network depth to improve the

accuracy will lead to two problems, vanishing gradient and exploding gradient problem and accuracy decline. The latter is not due to over-fitting. It is caused by saturation or even a decline in accuracy. For problem 1, Batch Norm [10] is adopted. The residual learning mechanism can be used to solve performance degradation problems caused by the increase of the alternating convolutional layer like problem 2. If  $H(x)$  is regarded as the desired actual mapping, that is, the stacked multi-layer nonlinear network is used to represent the fitting of the mapping relationship, then the multi-layer network can gradually approach a complex function. It is assumed to be equivalent to the approximation residual function  $F(x)$ , where  $x$  refers to the input of the first layer and  $F(x)$  represents the residual function, then the actual mapping relationship can be expressed as:

$$H(x) = F(x) + x. \quad (1)$$

The advantage of using residual block is that there are no redundant parameters and the computational complexity will not be increased. The following figure is the most basic residual learning unit whose idea is to assume that there is identity mapping between models, namely, to solve the identity mapping function. Because it is difficult to obtain  $H(x)$  directly, the residual unit is used to pass through the lines of short-cut connections. Formula (1) can be realized by adding a feed-forward neural network of short-cut connections. In other words, the input of each layer is the superposition of mapping and input instead of the input mapping of a traditional neural network.

### 2.2. Batch Normalization

Batch Normalization, also known as the BN layer, is a data pre-processing method. After the input data is normalized, the output range is between 0-1. BN layer eliminates the differences, reduces the interference of useless data, and speeds up network convergence.

### 2.3. Attention mechanism

The SE module proposed by CVPR in 2017 only solved the shortcomings of traditional convolution from the perspective

of channel, not from the perspective of space. In 2018, the author of CVPR proposed the sSE attention mechanism based on the SE attention mechanism to solve the shortcomings of convolution from the perspective of space.

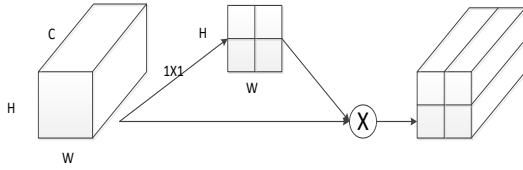


Figure 1. sSE attention mechanism

The difference between the two attention mechanisms is their different operations on dimensionality reduction. The SE attention mechanism uses global average pooling for dimensionality reduction, while the sSE attention mechanism uses conv 1\*1 for dimensionality reduction. However, their similarity is that they all use the Sigmoid function activation to compress the weight between 0 and 1, which is convenient for multiplying and stacking with the original feature.

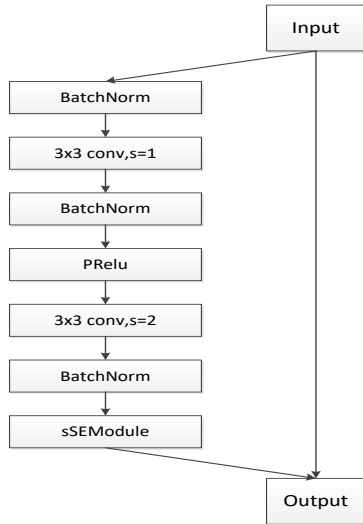


Figure 2. Network Structure

### 3. Improved Loss Function

#### 3.1. Softmax function

Sigmoid [13] function, also called Logistic function, is a binary classification problem. The logistic function fails to meet the requirement in the case of multi-classification of face images. Softmax[14] function is usually viewed as the last classifier in CNN for multi-classification tasks. Softmax loss function can ensure a good separability between classes. However, the within-class distance of features is scattered in a large range, in-class features are not compact enough, and the distance between some in-class features is even longer than that between classes. The function is expressed as follows:

$$LOSS_{\text{softmax}} = -\frac{1}{N} \sum_{i=1}^N \log \left( \frac{e^{w_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^n e^{w_{y_j}^T x_i + b_{y_j}}} \right) \quad (2)$$

N: batch size

n: number of categories

In Formula (2),  $w_{y_j}^T x_i + b_{y_j}$  represents the output of the fully connected layer. The value of  $w_{y_j}^T x_i + b_{y_j}$  has to be

increased to reduce Loss and all faces belonging to this kind of sample are included in this kind of decision boundary. Softmax mainly takes correct classification into consideration while lacking constraints on in-class and inter-class distances. Generally speaking, it is poor in classifying similar faces. Wen Yandong[15], the author of this thesis, finds that there is still a large in-class distance in traditional Softmax through experiments. That is to say, by improving loss function, it can add constraints on in-class distance and improve network performance as well.

#### 3.2. Improved Softmax Function

Because Softmax mainly considers whether the samples can be correctly classified, but lacks the distance limitation within and between classes, this paper raises an ArcFace[16]. Arcface improves the recognition capability of the training model by reducing the in-class distance and increasing the inter-class distance. By observing the relationship between weight and class center, Sphereface[17], an epoch-making paper is put forward. Moreover, the important concept of angular margin is introduced. However, some approximate calculations needed will lead to training instability. The Arc-cosine function is used to calculate the angle between the current feature and the weight and an extra angular margin  $m$  is added to the target angle. To reduce the computational complexity, the offset  $b_j = 0$ , the inner product of weight and input features is as follows:

$$W_j^T x_i = \|W_j\| \|x_i\| \cos \theta_j. \quad (3)$$

Regularizing the weight and feature L2, namely,  $\|W_j\| = 1$  and  $\|x_i\| = 1$ , and  $\theta_j$  is the angle between  $W_j$  and  $x_i$ , therefore, the learned embedding features are distributed on the hyper-sphere with radius  $s$ . Since the embedding features are distributed around each feature center on the hyper-sphere, an extra  $m$  is added between the weight and feature. Arcface loss function is shown as:

$$LOSS_{\text{Arcface}} = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{s(\cos(\theta_{y_i} + m))}}{e^{s(\cos(\theta_{y_i} + m))} + \sum_{j=1, j \neq y_i}^n e^{s \cos \theta_j}} \quad (4)$$

Formula (4) enhances the in-class compactness and expands the inner-class differences almost simultaneously. Improved Softmax loss function pays attention to maximizing the classification boundary directly in the angular space. Combined with Cosface[18] and the usage of angular margin  $m$  in Arcface, this paper points out a modified loss function for Arcface loss.

The details are as follows:

$$Loss = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{s(\cos(\theta_{y_i} + m) + c)}}{e^{s(\cos(\theta_{y_i} + m) + c)} + \sum_{j=1}^n e^{s \cos \theta_j}} \quad (5)$$

s: radius of hypersphere

m: angle margin

c: cosine distance.

$$s((\cos(\theta_1 + m) + c) - \cos \theta_2) = 0. \quad (6)$$

Formula (5) is the final loss function and Formula (6) is the

corresponding classification boundary.  $c$  is a cosine distance with a value of 0.3. Adding  $c$  can further compress the intra-class distance, and compressing the intra-class distance is equivalent to expanding the inter-class distance. If  $\cos(\theta_1 + m) + c > \cos\theta_2$ , then the face input belongs to category 1, otherwise, it belongs to category 2.

## 4. Experimental Results and Analysis

### 4.1. Introduction to Data Set

LFW data set including a total of 13000 face images of about 5000 people. CASIA-WebFace[19] data set is a large-scale face data set published in 2014. Face images collected from the Internet, including 494414 images of 10575 people, are used as the training set of Resnet in this experiment. There are 4000 face images of 100 people in the AR face database. It not only contains four basic expressions like nature, joy, anger, and surprise but also includes face images under various illumination conditions. Besides, it also contains partially occluded face images with sunglasses or scarves. Normal images and partially occluded images with sunglasses or scarves of 100 people are selected from the AR database. AgeDB [20] is a dataset of different ages, containing about 16,000 images of human faces ranging in age from 1 to 101, with an average age of 30 for each face.

### 4.2. Experimental Results

Based on the pytorch framework of deep learning, the experiment is carried out in the windows environment with RTX2060 GPU and 6GB RAM. The Batch size = 64, the learning rate = 1/e, and the optimizer adopts SGD. Accuracy:

$$ACC = \frac{TP+TN}{TP+TN+FN+FP} \quad (7)$$

The meanings represented by the symbols of formulas (7) are as follows:

TP: True-Negative  
 FN: True-Negative  
 FP: False-Positive  
 TN: True-Negative

**Table 1.** Accuracy Rate of various algorithms in LFW

Face recognition algorithm	Accuracy rate (%)
Deepface [21]	97.35%
Facenet [22]	98.87%
Softmax+center Loss [23]	98.78%
Arcface[16]	99.53%

**Table 2.** Accuracy Rate of the algorithm (s=30, m=0.3, c=0.3) in various data sets

Network+ Loss function	Data set	Accuracy rate (%)
Network structure + formula (4)	LFW	99.57%
Network structure + formula (5)	LFW	99.60%
Network structure + formula (5)	AgeDB	94.93%
Network structure + formula (5)	AR(illumination and expression)	99.99%
Network structure + formula (5)	AR(occlusions)	98.8%

Table 2 shows the comparative experiments of this paper and the accuracy of the improved loss function algorithm on four data sets. By comparing the accuracy, it can be seen that the improved loss function slightly improves the accuracy compared to the Arcface loss function set, which proves the effectiveness of the improved loss function.

## 5. Conclusion

Since the lack of distance constraints within the same class and between different classes, Softmax loss fails to improve accuracy no matter how good the model is used in training. The algorithm proposed in this paper uses the improved Arcface loss function to solve the non-ideal classification effect of the Softmax layer based on Resnet and has achieved great results. A large number of experiments shall be conducted to set super-parameter  $m$  reasonably. As the angular margin  $m$  increases, the model is hard to be trained. Further study is needed on how to select  $m$ . Last but not least, improving the low accuracy of the AgeDB data set also needs to be solved.

## References

- [1] Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns[J]. IEEE Transactions on pattern analysis and machine intelligence, 2002, 24(7): 971-987.
- [2] Platt J. Sequential minimal optimization: A fast algorithm for training support vector machines[J]. 1998.
- [3] Alzu'bi A, Albalas F, AL-Hadhrani T, et al. Masked Face Recognition Using Deep Learning: A Review[J]. Electronics, 2021, 10(21): 2666.
- [4] Yu Cuicancui & Li Huibin. Review of Face Recognition Methods Based on Deep Learning [J]. Chinese Journal of Engineering Mathematics, 2021,38(4): 451-469.
- [5] Tiantian Chen, Hongrong Jing , Hongjie Zhang. Research on Face Recognition Method Based on Deep Learning [J]. Scientific Journal of Intelligent Systems Research,2021,3(7):
- [6] Li L, Mu X, Li S, et al. A review of face recognition technology [J]. IEEE Access, 2020, 8: 139110-139120.
- [7] Wang M, Deng W. Deep face recognition: A survey[J]. Neurocomputing, 2021, 429: 215-244.
- [8] Han X., Zhang H.Y., Zhang Y.Y. Facial Expression Recognition Based on Efficient Channel Attention Network[J]. Transducer and Microsystem Technologies, 2021(1).
- [9] K. He, X. Zhang, S. Ren, et al. Deep residual learning for image recognition [C]. Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, USA, 2016, 770-778.
- [10] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift [C]. International conference on machine learning. PMLR, 2015: 448-456.
- [11] Hu J, Shen L, Sun G. Squeeze-and-excitation networks [C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.
- [12] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module [C]. Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.
- [13] Yin X, Goudriaan J A N, Lantinga E A, et al. A flexible sigmoid function of determinate growth [J]. Annals of botany, 2003, 91(3): 361-371.

- [14] Jang E, Gu S, Poole B. Categorical reparameterization with gumbel-softmax [J]. arXiv preprint arXiv:1611.01144, 2016.
- [15] Wen Y, Zhang K, Li Z, et al. A discriminative feature learning approach for deep face recognition[C]. European conference on computer vision. Springer, Cham, 2016: 499-515.
- [16] Deng J, Guo J, Xue N, et al. Arcface: Additive angular margin loss for deep face recognition[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 4690-4699.
- [17] Liu W, Wen Y, Yu Z, et al. Sphereface: Deep hypersphere embedding for face recognition[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 212-220.
- [18] Wang H, Wang Y, Zhou Z, et al. Cosface: Large margin cosine loss for deep face recognition[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 5265-5274.
- [19] Yi D, Lei Z, Liao S, et al. Learning face representation from scratch[J]. arXiv preprint arXiv:1411.7923, 2014.
- [20] Li J.Y.(24), Li Zh.H., Xie L.C., etc. Research Progress of Cross-age Face Recognition Based on Aging Model[J]. Computer Engineering and Applications, 2021(24).
- [21] Taigman Y, Yang M, Ranzato M, et al. DeepFace: Closing the gap to human-level performance in face verification [C]. Computer Vision and Pattern Recognition. IEEE, 2014: 1701-1708.
- [22] Schroff F, Kalenichenko D, Philbin J. FaceNet: A unified embedding for face recognition and clustering [C]. IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2015: 815-823.
- [23] SUN Yi, WANG Xiaogang, TANG Xiaoou. Deep Learning Face Representation from Predicting 10,000 Classes [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2014: 1891-1898.