

# *ReCurricularFace: Revisiting CurricularFace for Hard Sample Mining*

Meng Sang<sup>1,2,a</sup>, Yang Yang<sup>1,2,b,\*</sup>

<sup>1</sup>*School of Information Science and Technology, Yunnan Normal University, Kunming, China*

<sup>2</sup>*Laboratory of Pattern Recognition and Artificial Intelligence, Yunnan Normal University, Kunming, China*

<sup>a</sup>*sangmeng.one@gmail.com*, <sup>b</sup>*yyang\_ynu@163.com*

*\*Corresponding author*

**Keywords:** Face recognition, deep learning, loss function

**Abstract:** Mining of hard samples has always been a challenge in the field of face recognition. Mining-based methods have achieved promising results on the challenge of hard samples. However, current methods all suffer from the problem of not thinking, about when the hard samples should be close to the target class center and when they should be close to the non-target class center. Therefore, this work is based on CurricularFace by analyzing the logit and gradient, to carry out the boundary of judging the hard samples to be close to the center of the target class and non-target class to be close to the center, and based on the boundary to revisit the CurricularFace, to obtain a revised CurricularFace (ReCurricularFace), which is named as ReCurricularFace. We find through comparison experiments that ReCurricularFace obtains a huge improvement in the face benchmark.

## 1. Introduction

Face recognition is an important direction of biometric identification, in the financial, hotel, retail, security, and other fields, face recognition has a huge application value. Although face recognition[1] has great application value, in complex application scenarios, especially in uncontrolled scenarios, a large number of hard samples that are difficult to recognize will be generated, which poses a great challenge to face recognition algorithms. Therefore, step by step, reducing the accuracy of hard samples being recognized incorrectly, thus improving the accuracy of face recognition algorithms, will further expand the application of face recognition in various fields of application scenarios.

Some of the classic works based on hard sample mining, such as MV-Softmax[2] and CurricularFace[3], have further improved the accuracy of face recognition. MV-Softmax[2] addresses the challenge of effectively learning hard samples by adding different penalties between the target class center and the non-target class center. However, this approach can lead to training instability. In the subsequent work, CurricularFace introduces curriculum learning to address the instability issue of MV-Softmax. Although previous works like MV-Softmax[2] and CurricularFace[3] have further enhanced the performance of face recognition, CurricularFace, during the learning process, narrows down the decision boundary range by reducing the size of the

target class center. This limitation prevents the network from reaching its optimal state.

In the CurricularFace framework, the CurricularFace loss[3] is defined as follows:

$$\mathcal{L} = -\log \frac{e^{sT(\cos \theta_{y_i})}}{e^{sT(\cos \theta_{y_i}) + \sum_{j=1, j \neq y_i}^n e^{sN(t, \cos \theta_j)}}} \quad (1)$$

where  $T(\cos(\theta_{y_i})) = \cos(\theta + m)$  and  $N(t, \cos(\theta_j)) = \cos(\theta) * I(t, \cos(\theta))$  are two functions representing the positive and negative cosine similarities, respectively.  $I(t, \cos(\theta))$  denotes the coefficient for negative cosine similarity. Considering the distance between the target class center and the non-target class center as  $\frac{\pi}{2}$ , with an added margin function of  $m = 0.5$ , the samples lie on a plane formed by the target and the non-target class center. We can obtain  $\theta_{y_i} = \frac{\pi}{2} - \theta_j$ . Consequently, the relationship between  $N(t, \cos(\theta_j))$  and  $T(\cos(\theta_{y_i}))$  is illustrated in Figure 1. The intersection points of  $N$  and  $T$  represent the distance from the sample to the target center where both classifiers yield equal probabilities, corresponding to the actual decision boundary of the target class center. The yellow line indicates the ideal distance from the non-target class center to the decision boundary of the target class center. When  $N$  is above, the sample is classified as belonging to the non-target class center, and when  $T$  is above, it is classified as belonging to the target class center. Through Figure 1, it can be observed that CurricularFace reduces the actual decision boundary of the target class center, leading to the neural network failing to learn the optimal state.

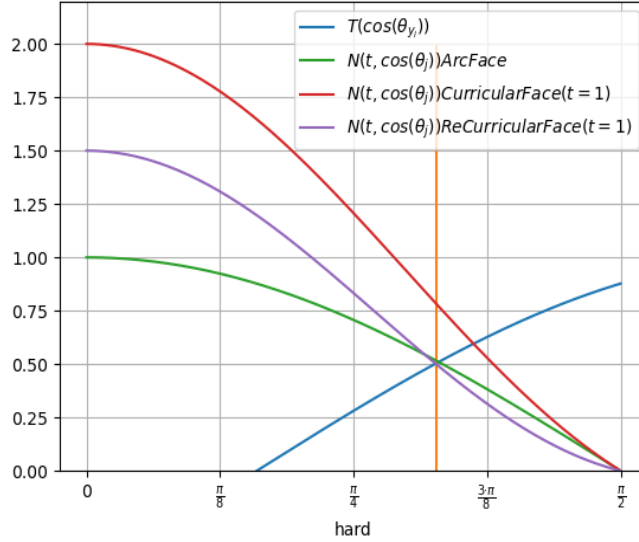


Figure 1: The relationship between  $N(t, \cos(\theta_j))$  and  $T(\cos(\theta_{y_i}))$ .

To address the issue mentioned above, we have modified CurricularFace to ensure correct classification. Firstly, considering the decision boundary perspective, when the sample is too close to the non-target center, we allow to take any value in the domain, and . Next, we employ the Curriculum Learning (CL) method. Initially, we provide a larger decision boundary for the target class center, enabling the network to learn between the class center and the samples. Additionally, during the convergence of the network, we gradually shrink the decision boundary of the target class center towards the ideal state.

## 2. Related Work

The design of the loss function is very important, and the general loss function has good generalization for different tasks. Research on loss functions for face recognition can be classified into loss functions based on margin and loss functions based on mining.

Loss function based on margin: The loss function based on margin evolved from classification loss and has become the mainstream direction of face loss research[1] in the field of face recognition. With the step-by-step exploration of face loss based on margin, the accuracy of face recognition has been gradually improved, which also verifies the importance of the idea based on margin. Previous face recognition mainly explored the design of the margin from the perspective of the target class center, while our work explored the design of the margin and the correctness of the margin design from the perspective of the non-target class center.

Loss function based on mining: Compared to the loss design based on margin, the face loss based on mining is relatively less. Among them, the more classic works are MV-Softmax[2] and CurricularFace[3]. MV-Softmax is the first to integrate the two ideas of margin and mining. CurricularFace brings the idea of curricular learning into face recognition. We redesign CurricularFace from the perspective of decision boundaries and the non-target class center.

## 3. Method

Although the relationship between samples, the target class center, the and non-target class center is defined in the same plane in the introduction section, in actual training on networks, a more complex situation is encountered. Based on previous work, we can infer that the class centers are assumed to be distributed on a hypersphere. Under this assumption, the distance between adjacent class centers is defined as follows:

$$\mathbb{E}[\theta(W_j)] \rightarrow n^{-\frac{2}{d-1}} \Gamma\left(1 + \frac{1}{d-1}\right) \left(\frac{\Gamma(\frac{d}{2})}{2\sqrt{\pi}(d-1)\Gamma(\frac{d-1}{2})}\right)^{-\frac{1}{d-1}} \quad (2)$$

Where  $d$  represents the dimensionality,  $n$  represents the number of class centers, and  $\theta(W_j) = \min_{1 \leq i, j \leq n, i \neq j} \arccos(W_i, W_j)$  for each  $i, j$ . Additionally, the distance from the target class to the decision boundary is  $\frac{\mathbb{E}[\theta(W_j)] - m}{2}$ . Therefore, when  $\cos(\theta_j) > \cos(\frac{\mathbb{E}[\theta(W_j)] - m}{2})$ , samples will be classified within the decision boundary of the non-target class, and we should ensure that  $I(t, \cos(\theta_j)) > 1$  to rapidly increase the distance between the sample and the non-target center. Thus, we scale  $\cos(\theta_j)$  by a coefficient. Therefore, we can obtain:

$$I\left(t, \cos(\theta_j)\right) = \frac{1}{\cos\left(\frac{\mathbb{E}[\theta(W_j)] - m}{2}\right)} \cos(\theta_j) \quad (3)$$

Though the above  $I(t, \cos(\theta_j))$  can rapidly increase the distance between the sample and the non-target center, under this condition, according to  $N(t, \cos(\theta_j)) = T(\cos(\theta_{y_i}))$ , the actual decision boundary generated by class centers will expand, making it easier for more hard samples to be attended to by the target class centers. However, this would cause the network to lose some discriminative power. To shrink the decision boundary to an ideal state, we introduce Curricular Learning, gradually restoring the decision boundary of class centers to the ideal state. For this purpose, we introduce a  $\alpha t$  bias term to shift the decision boundary to the ideal state. Therefore, we can obtain an equation:

$$\frac{1}{\cos\left(\frac{\mathbb{E}[\theta(W_j)]-m}{2}\right)} * \cos(\theta_j) + \alpha t = \cos\left(\frac{\mathbb{E}[\theta(W_j)]+m}{2}\right) \quad (4)$$

Where  $t \in [0,1]$ . As the network converges,  $t \rightarrow 1$ , we should ensure that the decision boundary tends toward the ideal state. By assuming  $\theta_j = \frac{\mathbb{E}[\theta(W_j)]+m}{2}$ , we can obtain the value of  $\alpha$ , represented as follows:

$$\alpha \approx \left(1 - \frac{\mathbb{E}[\theta(W_j)]-m}{2}\right) \frac{\mathbb{E}[\theta(W_j)]+m}{2} \quad (5)$$

Therefore, summarizing the above, we define  $I(t, \cos(\theta_j))$  as follows:

$$I(t, \cos(\theta_j)) = \frac{1}{\cos\left(\frac{\mathbb{E}[\theta(W_j)]-m}{2}\right)} * \cos(\theta_j) + t \left(1 - \frac{\mathbb{E}[\theta(W_j)]-m}{2}\right) \frac{\mathbb{E}[\theta(W_j)]+m}{2} \quad (6)$$

## 4. Experiments

### 4.1. Datasets

**Training Datasets:** The training set is trained using CASIA-WebFace[4].

**Testing Datasets:** Based on previous AdaFace, we divided the dataset into high-quality, mixed-quality, and low-quality datasets. The high-quality dataset includes LFW[5], AgeDB-30[6], and CFP-FP[7]. The mixed-quality dataset comprises IJB-B[8] and IJB-C[9], while the low-quality dataset consists of TinyFace[10].

### 4.2. Implement details

According to ArcFace[1], it can be inferred that  $\mathbb{E}[\theta(W_j)] \approx 80$ . Therefore,  $I(t, \cos(\theta_j)) = 1.1\cos(\theta_j) + 0.4t$ . Other relevant training details are as follows: The backbone network is trained using ResNet18. The batch size is set to 512, training for 34 epochs with a learning rate of 0.1, and the learning rate is reduced by a factor of 0.1 at epochs 20, 28, and 32.

### 4.3. Evaluation Protocol

In the testing set, we partition the dataset into high-quality, mixed-quality, and low-quality datasets. For the high-quality and mixed-quality datasets, we employ a face verification protocol (Accuracy) to assess the model's performance. For the low-quality dataset, we utilize a face recognition protocol (Rank-n) for evaluation.

### 4.4. Results on LFW, AgeDB-30, and CFP-FP

Table 1: Results on LFW, AgeDB-30, and CFP-FP. Accuracy is used to evaluate the performance of face recognition, with the best metric indicated in bold red font.

| Method           | LFW(%) | AgeDB-30(%) | CFP-FP (%) | AVG(%) |
|------------------|--------|-------------|------------|--------|
| ArcFace          | 99.25  | 93.67       | 93.12      | 95.35  |
| CurricularFace   | 99.22  | 92.53       | 93.33      | 95.03  |
| ReCurricularFace | 99.25  | 94.03       | 93.35      | 95.54  |

From Table. 1, it can be observed that for the same environment, ReCurricularFace demonstrates a significant improvement over ArcFace and CurricularFace on the high-quality dataset. This

indicates the effectiveness of ReCurricularFace.

Table 2: Results on IJB-B and IJB-C. Tar@Far is used to evaluate the performance of face recognition, with the best metric indicated in bold red font.

| Datasets\Methods |                  | Tar@Far(%) |       |       |       |       |       |
|------------------|------------------|------------|-------|-------|-------|-------|-------|
|                  |                  | 1e-06      | 1e-05 | 1e-04 | 1e-03 | 1e-02 | 1e-01 |
| IJB-B            | ArcFace          | 0.05       | 0.21  | 1.41  | 40.54 | 89.20 | 96.58 |
|                  | CurricularFace   | 0.68       | 6.99  | 39.24 | 80.78 | 93.45 | 97.81 |
|                  | ReCurricularFace | 19.76      | 60.53 | 77.96 | 87.41 | 94.05 | 98.27 |
| IJB-C            | ArcFace          | 0.07       | 0.13  | 1.57  | 38.44 | 89.11 | 96.79 |
|                  | CurricularFace   | 1.72       | 10.54 | 42.94 | 82.52 | 94.34 | 98.08 |
|                  | ReCurricularFace | 54.23      | 71.83 | 82.40 | 90.09 | 95.22 | 98.47 |

#### 4.5. Results on IJB-B and IJB-C

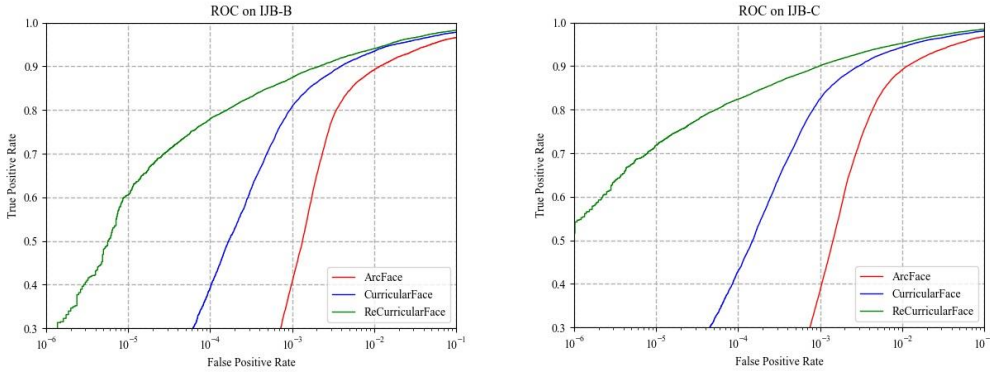


Figure 2: ROC of IJB-B and IJB-C.

From Table. 2, it can be observed that on the mixed dataset, ReCurricularFace exhibits a widening performance gap relative to ArcFace and CurricularFace as the constraint conditions increase. Therefore, this suggests that ReCurricularFace enables the network to better learn hard samples. Figure 2 illustrates the same results.

#### 4.6. Results on TinyFace

From Table. 3, it is evident that both CurricularFace and ReCurricularFace show improvements relative to ArcFace. Particularly, CurricularFace and ReCurricularFace exhibit similar performance on the low-quality dataset.

Table 3: Results on TinyFace. Rank-n is used to evaluate the performance of face recognition, with the best metric indicated in bold red font.

| Method           | Rank-1(%) | Rank-5(%) | Rank-20(%) | AVG(%) |
|------------------|-----------|-----------|------------|--------|
| ArcFace          | 0.42      | 0.48      | 0.54       | 0.48   |
| CurricularFace   | 0.46      | 0.51      | 0.55       | 0.51   |
| ReCurricularFace | 0.44      | 0.51      | 0.56       | 0.50   |

### 5. Conclusion

In this study, we introduced modifications to CurricularFace from the perspective of non-class

centers and decision boundaries, resulting in ReCurricularFace. Experimental results demonstrate significant improvements in ReCurricularFace over CurricularFace on both high-quality and mixed-quality datasets. Therefore, we can conclude that when designing face loss functions, it is essential to ensure that the actual decision boundaries of target and non-target class centers align with the ideal decision boundaries. This conclusion can provide further guidance for the design of loss functions in subsequent research. Additionally, for future work, we consider incorporating quality attributes into the loss function to further guide the network's learning process.

## References

- [1] 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.
- [2] Wang X, Zhang S, Wang S, et al. Mis-classified vector guided softmax loss for face recognition[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2020, 34(07): 12241-12248.
- [3] Huang Y, Wang Y, Tai Y, et al. Curricularface: adaptive curriculum learning loss for deep face recognition[C]//proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 5901-5910.
- [4] Yi D, Lei Z, Liao S, et al. Learning face representation from scratch[J]. arXiv preprint arXiv:1411.7923, 2014.
- [5] Huang G B, Mattar M, Berg T, et al. Labeled faces in the wild: A database for studying face recognition in unconstrained environments[C]//Workshop on faces in 'Real-Life' Images: detection, alignment, and recognition. 2008.
- [6] Moschoglou S, Papaioannou A, Sagonas C, et al. Agedb: the first manually collected, in-the-wild age database[C]//Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2017: 51-59.
- [7] Sengupta S, Chen J C, Castillo C, et al. Frontal to profile face verification in the wild[C]//2016 IEEE winter conference on applications of computer vision (WACV). IEEE, 2016: 1-9.
- [8] Whitelam C, Taborsky E, Blanton A, et al. Iarpa janus benchmark-b face dataset[C]//proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2017: 90-98.
- [9] Maze B, Adams J, Duncan J A, et al. Iarpa janus benchmark-c: Face dataset and protocol[C]//2018 international conference on biometrics (ICB). IEEE, 2018: 158-165.
- [10] Cheng Z, Zhu X, Gong S. Low-resolution face recognition[C]//Computer Vision—ACCV 2018: 14th Asian Conference on Computer Vision, Perth, Australia, December 2–6, 2018, Revised Selected Papers, Part III 14. Springer International Publishing, 2019: 605-621.