QCFace: Image Quality Control for boosting Face Representation & Recognition

Duc-Phuong Doan-Ngo, Thanh-Dang Diep, Thanh Nguyen-Duc, Thanh-Sach LE, Nam Thoai
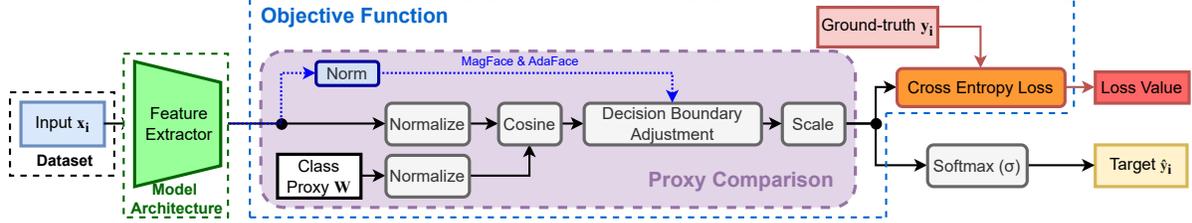
Supplementary Material

Figure 1. The training pipeline of face recognition models with class center loss.

A. The previous class center loss analysis

Fig. 1 clearly illustrates about four components of a training pipeline with *class-center-loss-based* methods, where we can target improvements to enhance face recognition performance, specifically, *model architecture*, *training dataset*, *training strategy* and *objective (loss) function* – the target of the following analysis. Based on the criteria mentioned in Sec. 1 in our main paper, the loss function shows its most potential approach. Recall that the general formula of *the class center loss* is shown as follows.

$$\mathcal{L}_i = \underbrace{-log \frac{e^{s.F(\mathbb{M},\theta_{\boldsymbol{w}_{y_i}},\boldsymbol{z}_i)}}{e^{s.F(\mathbb{M},\theta_{\boldsymbol{w}_{y_i}},\boldsymbol{z}_i)} + \sum_{j\neq y_i}^{C} e^{s.N(t,\theta_{\boldsymbol{w}_j},\boldsymbol{z}_i)}}}_{\mathcal{L}_{sm_i}} + \underbrace{\lambda_g . g(\boldsymbol{z}_i)}_{\mathcal{L}_{reg_i}}$$

(1)

where y_i is an identity label of an image i, and

$$F(\mathbb{M},\theta) = \cos(m_1\theta + m_2) - m_3 \text{ where } \mathbb{M} = \{m_1, m_2, m_3\} \quad (2)$$

where m_1, m_2 and m_3 express three margin types to adjust the decision boundary in $\mathcal{L}_{sm}$, Tab. 1 details margin components (*i.e.*, $\mathbb{M} = \{m_1, m_2, m_3\}$) in positive modulation function (*i.e.*, F), negative modulation function (*i.e.*, N) and regularization loss (*i.e.*, $\mathcal{L}_{reg}$) of previous studies (*e.g.*, [5, 10, 15, 20, 23, 24]).

B. Technical Terminology Definition

The following definitions are common in the face recognition community and are mentioned in our main paper.

- ***Knowledge accumulation (aka learning capacity)*** is a concept used to express the **model** property. It means how much knowledge (*aka* useful information) a deep learning model can extract (*aka* learn) from a dataset.
- ***Knowledge capacity*** is used to express the **data** quality. Specifically, it describes the degree of diversity and effect of noise (*e.g.,* mislabeled samples) of datasets.

- ***Image recognizability*** is a human-perceptual variable used to express the **image** property, which quantifies the ease of recognizing an image. Specifically, it is a measurement of how easy an image can be recognized.
- ***Recognizability capacity*** is used to express the usefulness of an embedded feature in representing a face image.
- ***Mislabeled sample*** refers to a sample with an incorrect label, which is a common problem in large-scale datasets.
- ***Hypersphere*** is an embedded feature space, which is technically defined by choosing the number of dimensions of embedded features. Higher values can represent more data diversity, but they are easily prone to overfitting and cause high computational cost. The suitable value is frequently chosen based on dataset scale and by conducting experiments. The value 512 is the common embedded dimension in most well-known studies (*e.g.*, [5, 10, 15, 20, 23, 24]).

C. Lemma & Property Proofs

Recall that the formulas of the gradients affecting the actual-class proxy, the non-actual (misclassified) -class proxy, feature direction and magnitude are respectively shown as follows.

$$\boldsymbol{g}_{ac}^i := s.(P_{y_i}^{(i)} - 1).\frac{\partial F(\mathbb{M},\theta_{\boldsymbol{w}_{y_i}},\boldsymbol{z}_i)}{\partial \cos(\theta_{\boldsymbol{w}_{y_i}},\boldsymbol{z}_i)}.\frac{\partial \cos(\theta_{\boldsymbol{w}_{y_i}},\boldsymbol{z}_i)}{\partial \boldsymbol{w}_{y_i}} \quad (3)$$

$$\boldsymbol{g}_{mc}^i := s.P_k^{(i)}.\frac{\partial N(t,\theta_{\boldsymbol{w}_k},\boldsymbol{z}_i)}{\partial \cos(\theta_{\boldsymbol{w}_k},\boldsymbol{z}_i)}.\frac{\partial \cos(\theta_{\boldsymbol{w}_k},\boldsymbol{z}_i)}{\partial \boldsymbol{w}_k} \text{ where } k \neq y_i \quad (4)$$

$$\boldsymbol{g}_{\theta}^i := s.\sum_{k=1}^{C}\left[P_k^{(i)} - \mathbf{1}_{k=y_i}\right].\frac{\partial Fnc_k^{(i)}}{\partial \cos(\theta_{\boldsymbol{w}_k},\boldsymbol{z}_i)}.\frac{\partial \cos(\theta_{\boldsymbol{w}_k},\boldsymbol{z}_i)}{\partial \boldsymbol{z}_i} \quad (5)$$

$$\boldsymbol{g}_{\|\boldsymbol{z}\|}^i := \underbrace{s.\sum_{k=1}^{C}\left[P_k^{(i)} - \mathbf{1}_{k=y_i}\right].\frac{\partial Fnc_k^{(i)}}{\partial \|\boldsymbol{z}_i\|}.\frac{\partial \|\boldsymbol{z}_i\|}{\partial \boldsymbol{z}_i}}_{\boldsymbol{g}_{\|\boldsymbol{z}\|-}^i} + \underbrace{\frac{\partial \mathcal{L}_{reg_i}}{\partial \|\boldsymbol{z}_i\|}.\frac{\partial \|\boldsymbol{z}_i\|}{\partial \boldsymbol{z}_i}}_{\boldsymbol{g}_{\|\boldsymbol{z}\|+}^i} \quad (6)$$

Table 1. The detail of each component in Eq. (1) of previous *class center loss* solutions. m expresses a constant coefficient, u_m, l_m, u_a and l_a are respectively upper and lower bounds of additive angular margin and feature magnitude $\|\boldsymbol{z}\|$, $\widehat{\|\boldsymbol{z}\|}$ is batch normalized value of $\|\boldsymbol{z}\|$, and t expresses mis-classified coefficient.

Method	Additive function components			N	g
	m_1	m_2	m_3		
SphereFace [15]	m	0	0	$cos(\theta_j)$	0
CosFace [23]	0	0	m	$cos(\theta_j)$	0
ArcFace [5]	0	m	0	$cos(\theta_j)$	0
MV-Arc-Softmax [24]	0	m	0	$t.cos(\theta_j) + t - 1$	0
CurricularFace [10]	0	m	0	$(t + cos(\theta_j)).cos(\theta_j)$	0
MagFace [17]	0	$\frac{u_m - l_m}{u_a - l_a}.(\|\boldsymbol{z}\| - l_a) + l_m$	0	$cos(\theta_j)$	$\frac{1}{\|\boldsymbol{z}\|} + \frac{\|\boldsymbol{z}\|}{u_a^2}$
AdaFace [13]	0	$-m.\widehat{\|\boldsymbol{z}\|}$	$m.(\widehat{\|\boldsymbol{z}\|} + 1)$	$cos(\theta_j)$	0
QCFace (Ours)	0	m	0	$cos(\theta_j)$	$g_{rc}(\|\boldsymbol{z}\|, p_{rc})$

where

$$Fnc_k^i = \begin{cases} F(\mathbb{M}, \theta_{\boldsymbol{w}_{y_i}, \boldsymbol{z}_i}) & k = y_i \\ N(t, \theta_{\boldsymbol{w}_k, \boldsymbol{z}_i}) & k \neq y_i \end{cases} \text{ and } \mathbf{1}_{k=y_i} = \begin{cases} 1 & k = y_i \\ 0 & k \neq y_i \end{cases} \tag{7}$$

The probability output at class k (*i.e.*, $P_k^{(i)}$) and derivatives of Fnc_k^i with respect to direction (*i.e.*, $cos(\theta_{\boldsymbol{w}_k, \boldsymbol{z}_i})$) and magnitude (*i.e.*, $\|\boldsymbol{z}_i\|$) are calculated as follows.

$$P_k^{(i)} = \frac{e^{Fnc_k^{(i)}}}{\sum_{j=1}^C e^{Fnc_j^{(i)}}} \tag{8}$$

$$\frac{\partial Fnc_k^{(i)}}{\partial cos(\theta_{\boldsymbol{w}_k, \boldsymbol{z}_i})} = \begin{cases} s.\frac{\partial cos(m_1.\theta_{\boldsymbol{w}_k, \boldsymbol{z}_i})}{\partial cos(\theta_{\boldsymbol{w}_k, \boldsymbol{z}_i})} \times f(\theta_{\boldsymbol{w}_k, \boldsymbol{z}_i}, \|\boldsymbol{z}_i\|) & k = y_i \\ s.\frac{\partial N(t, cos(\theta_{\boldsymbol{w}_k, \boldsymbol{z}_i}))}{\partial cos(\theta_{\boldsymbol{w}_k, \boldsymbol{z}_i})} & k \neq y_i \end{cases}$$

where
$$f(\theta_{\boldsymbol{w}_k, \boldsymbol{z}_i}, \|\boldsymbol{z}_i\|) = cos(m_2(\|\boldsymbol{z}_i\|)) + cot(m_1.\theta_{\boldsymbol{w}_k, \boldsymbol{z}_i}).sin(m_2(\|\boldsymbol{z}_i\|)) \tag{9}$$

$$\frac{\partial Fnc_k^{(i)}}{\partial \|\boldsymbol{z}_i\|} = \begin{cases} -s.sin(m_1.\theta_{\boldsymbol{w}_k, \boldsymbol{z}_i} + m_2(\|\boldsymbol{z}_i\|)).\frac{\partial m_2(\|\boldsymbol{z}_i\|)}{\partial \|\boldsymbol{z}_i\|} & k = y_i \\ 0 & k \neq y_i \end{cases} \tag{10}$$

Lemma 1. *Let $\boldsymbol{z}_1, \boldsymbol{z}_2 \in \mathbb{R}^d$ where d denotes the embedding dimension. If $\|\boldsymbol{z}_1\| > \|\boldsymbol{z}_2\|$, then $\|\boldsymbol{g}_\theta(\boldsymbol{z}_1)\| < \|\boldsymbol{g}_\theta(\boldsymbol{z}_2)\|$.*
Proof. Let $f(\|\boldsymbol{z}_i\|) = cos(\theta_{\boldsymbol{w}_k, \boldsymbol{z}_i})$. Subsequently, we have

$$f(\|\boldsymbol{z}_i\|) = \frac{\boldsymbol{z}_i^\top \boldsymbol{w}_k}{\|\boldsymbol{z}_i\|\|\boldsymbol{w}_k\|} = \frac{1}{\|\boldsymbol{w}_k\|}.\frac{\boldsymbol{z}_i^\top \boldsymbol{w}_k}{\|\boldsymbol{z}_i\|} = c.\frac{\boldsymbol{z}_i^\top \boldsymbol{w}_k}{\|\boldsymbol{z}_i\|}$$

where $c = \frac{1}{\|\boldsymbol{w}_k\|}$ is constant.

Then, we calculate the derivative of $cos(\theta_{\boldsymbol{w}_k, \boldsymbol{z}_i})$ with respect to $\|\boldsymbol{z}_i\|$ as follows.

$$\frac{\partial cos(\theta_{\boldsymbol{w}_k, \boldsymbol{z}_i})}{\partial \boldsymbol{z}_i} = \nabla_{\boldsymbol{z}_i} f(\boldsymbol{z}_i) = c.\nabla_{\boldsymbol{z}_i}\left(\frac{\boldsymbol{z}_i^\top \boldsymbol{w}_k}{\|\boldsymbol{z}_i\|}\right)$$

$$= c.\frac{\|\boldsymbol{z}_i\|\boldsymbol{w}_k - (\boldsymbol{z}_i^\top \boldsymbol{w}_k).\frac{\boldsymbol{z}_i}{\|\boldsymbol{z}_i\|}}{\|\boldsymbol{z}_i\|^2}$$

$$= c.\frac{\boldsymbol{w}_k - \frac{\boldsymbol{z}_i^\top \boldsymbol{w}_k}{\|\boldsymbol{z}_i\|^2}.\boldsymbol{z}_i}{\|\boldsymbol{z}_i\|}$$
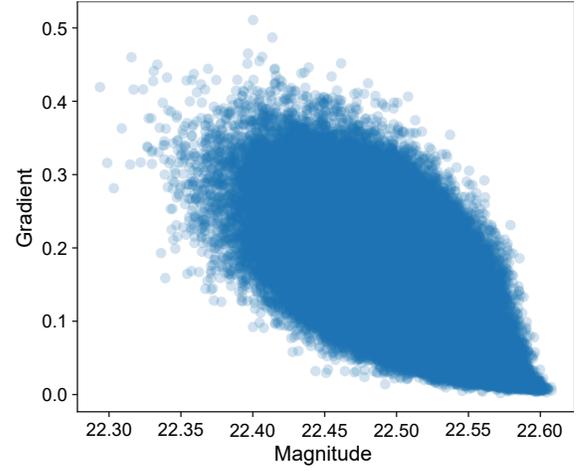

Figure 2. The plot of negative correlation between magnitude value of embedded features and mean values of gradient magnitude in an FR model trained with IResNet18 and CASIA-WebFace.

Now, we compute the norm of this gradient as follows to compare the magnitudes.

$$\|\nabla_{\boldsymbol{z}_i} f(\boldsymbol{z}_i)\| = c.\frac{\|\boldsymbol{w}_k - \frac{\boldsymbol{z}_i^\top \boldsymbol{w}_k}{\|\boldsymbol{z}_i\|^2}.\boldsymbol{z}_i\|}{\|\boldsymbol{z}_i\|}$$

Let us define the projection of $\boldsymbol{w}_k$ onto $\boldsymbol{z}_i$ as follows.

$$proj_{\boldsymbol{z}_i}(\boldsymbol{w}_k) = \frac{\boldsymbol{z}_i^\top \boldsymbol{w}_k}{\|\boldsymbol{z}_i\|^2}.\boldsymbol{z}_i$$

Then the numerator becomes $\boldsymbol{w}_k - proj_{\boldsymbol{z}_i}(\boldsymbol{w}_k)$, *i.e.* the component of $\boldsymbol{w}_k$ orthogonal to $\boldsymbol{z}_i$. Thus, we have

$$\|\nabla_{\boldsymbol{z}_i} f(\boldsymbol{z}_i)\| = c.\frac{\|\boldsymbol{w}_{k\perp}\|}{\|\boldsymbol{z}_i\|} = \frac{1}{\|\boldsymbol{w}_k\|}.\frac{\|\boldsymbol{w}_{k\perp}\|}{\|\boldsymbol{z}_i\|} \tag{11}$$

Based on (11), we derive that $\|\nabla_{\boldsymbol{z}_i} f(\boldsymbol{z}_i)\| \propto \frac{1}{\|\boldsymbol{z}_i\|}$. Besides, based on (5), we also have $\|\boldsymbol{g}_\theta\| \propto \|\nabla_{\boldsymbol{z}_i} f(\boldsymbol{z}_i)\|$,

thus, $\|\boldsymbol{g}_\theta\| \propto \frac{1}{\|\boldsymbol{z}_i\|}$. Therefore, we can derive that if $\|\boldsymbol{z}_1\| > \|\boldsymbol{z}_2\|$, then $\|\boldsymbol{g}_\theta(\boldsymbol{z}_1)\| < \|\boldsymbol{g}_\theta(\boldsymbol{z}_2)\|$, and Lemma 1 is completely demonstrated. The correctness of Lemma 1 is also validated by Fig. 2 experimentally.

Lemma 2. *Suppose that $a_1, a_2 \in \mathbb{R}$ are two independent variables, and a function $f(a_1, a_2) : \mathbb{R}^2 \to \mathbb{R}$. If f can be described as a form of $f(a_1, a_2) = f_1(a_1) + f_2(a_2)$ where $f_1, f_2 : \mathbb{R} \to \mathbb{R}$, then the mutual overlapping gradient calculated from the derivative of f does not exist.*

Proof. We first calculate the derivatives with respect to a_1 and a_2 as follows.

$$\begin{cases} \frac{\partial f}{\partial a_1} = \frac{\partial f_1}{\partial a_1} + \frac{\partial f_2}{\partial a_1} = \frac{\partial f_1}{\partial a_1} \quad := \boldsymbol{g}_{a_1} \\ \frac{\partial f}{\partial a_2} = \frac{\partial f_1}{\partial a_2} + \frac{\partial f_2}{\partial a_2} = \frac{\partial f_2}{\partial a_2} \quad := \boldsymbol{g}_{a_2} \end{cases}$$

Applying gradient descent for a_1 and a_2 to minimize f, the optimization process is shown as follows.

$$\begin{cases} a_1^{new} \leftarrow \mathcal{O}_1(a_1^{old}, \boldsymbol{g}_{a_1}) \\ a_2^{new} \leftarrow \mathcal{O}_2(a_2^{old}, \boldsymbol{g}_{a_2}) \end{cases} \quad (12)$$

where $\mathcal{O}_1$ and $\mathcal{O}_2$ are optimization (*aka* update) operations (*e.g.*, SGD, Adam, etc.) for two variables, a_1 and a_2. We can easily observe that

$$\begin{cases} \frac{\partial \boldsymbol{g}_{a_1}}{\partial a_2} = \frac{\partial^2 f_1}{\partial a_1 \partial a_2} = 0 \ \forall a_2 \in \mathbb{R} \quad \implies \frac{\partial \mathcal{O}_1}{\partial a_2} = 0 \\ \frac{\partial \boldsymbol{g}_{a_2}}{\partial a_1} = \frac{\partial^2 f_2}{\partial a_2 \partial a_1} = 0 \ \forall a_1 \in \mathbb{R} \quad \implies \frac{\partial \mathcal{O}_2}{\partial a_1} = 0 \end{cases} \quad (13)$$

Therefore, based on (13), the optimization processes of two variables (*i.e.*, a_1 and a_2) are not affected by each other's previous values, and Lemma 2 is completely demonstrated.

Property 1. *In Eq. (1), if m_2 is a strictly increasing convex function of $\|\boldsymbol{z}_i\|$ (i.e., $m_2(\|\boldsymbol{z}_i\|)$), then the norm $\|\boldsymbol{z}_i\|$ influences the gradient with respect to θ (i.e., $\boldsymbol{g}_\theta$), while θ in turn affects the gradient with respect to $\|\boldsymbol{z}_i\|$ (i.e., $\boldsymbol{g}_{\|\boldsymbol{z}\|}^i$). This manifests as a mutual overlapping gradient.*

Proof. Based on (5) and (6), the component in the loss function that may lead to a mutual overlapping gradient problem between $\boldsymbol{g}_\theta^i$ and $\boldsymbol{g}_{\|\boldsymbol{z}\|}^i$, is the margin function $F(\mathbb{M}, \theta)$ (see Eq. (2)). Specifically, if m_2 is a strictly increasing convex function of $\|\boldsymbol{z}_i\|$ (*i.e.*, $m_2(\|\boldsymbol{z}_i\|)$), based on (9) and (10), the gradients influencing the updates of two independent attributes of the embedded feature, *i.e.* direction (*i.e.*, $\cos(\theta_{\boldsymbol{w}_k, \boldsymbol{z}_i})$) and magnitude (*i.e.*, $\|\boldsymbol{z}_i\|$), depend on each other's previous values. In contrast to the findings of Lemma 2, we have

$$\begin{cases} \exists \|\boldsymbol{z}_i\| \in \mathbb{R}^+ \text{ such that } \frac{\partial \boldsymbol{g}_\theta^i}{\partial \|\boldsymbol{z}_i\|} \neq 0 \quad \implies \frac{\partial \mathcal{O}_\theta}{\partial \|\boldsymbol{z}_i\|} \neq 0 \\ \exists \theta_{\boldsymbol{w}_k, \boldsymbol{z}_i} \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \text{ such that } \frac{\partial \boldsymbol{g}_{\|\boldsymbol{z}\|}^i}{\partial \theta_{\boldsymbol{w}_k, \boldsymbol{z}_i}} \neq 0 \quad \implies \frac{\partial \mathcal{O}_{\|\boldsymbol{z}\|}}{\partial \theta_{\boldsymbol{w}_k, \boldsymbol{z}_i}} \neq 0 \end{cases}$$

where $\mathcal{O}_\theta$ and $\mathcal{O}_{\|\boldsymbol{z}\|}$ are update functions for direction and magnitude of embedded feature. Consequently, the mutual overlapping gradient exists in this situation.

Let us recall the formulas of regularization loss ($\mathcal{L}_{reg}$) and ***guidance value*** (p_d) of ***QCFace*** for the proofs of the following properties.

$$\mathcal{L}_{reg_i} = k.p_{d_i}.\left(\frac{1}{\|\boldsymbol{z}_i\|} + \frac{\|\boldsymbol{z}_i\|}{u_a^2}\right) + (1 - p_{d_i}).\left(\frac{1}{\|\boldsymbol{z}_i\|} + \frac{\|\boldsymbol{z}_i\|}{l_a^2}\right) - b \quad (14)$$

$$p_{d_i} = \frac{e^{s.cos(\theta_{\boldsymbol{w}_{y_i}, \boldsymbol{z}_i})}}{e^{s.cos(\theta_{\boldsymbol{w}_{y_i}, \boldsymbol{z}_i})} + \sum_{j \neq y_i}^n e^{s.cos(\theta_{\boldsymbol{w}_j, \boldsymbol{z}_i})}} \quad (15)$$

Property 2. *In Eq. (15), p_d is bounded in $[0, 1]$.*
Proof. We rewrite Eq. (15) as $p_{d_i} = \frac{e^{v_{y_i}}}{\sum_{j=1}^n e^{v_j}}$. Accordingly, we have

$$\begin{cases} e^{v_t} > 0 \ \forall v_t \in \mathbb{R} \quad \implies p_{d_i} > 0 \\ \sum_{j=1}^n e^{v_j} \geq e^{v_t} \ \forall v_t \in \mathbb{R} \quad \implies p_{d_i} \leq 1 \end{cases}$$

Therefore, p_{d_i} is a bounded value in $[0, 1]$.

Property 3. *Suppose that $\boldsymbol{z}_1, \boldsymbol{z}_2 \in \mathbb{R}^d$ where d denotes the embedding dimension, and let their corresponding guidance values (i.e., $p_{d1}, p_{d2} \in [0, 1]$) be computed by Eq. (15). If $p_{d1} > p_{d2}$, then $\|\boldsymbol{g}_\theta(\boldsymbol{z}_1)\| < \|\boldsymbol{g}_\theta(\boldsymbol{z}_2)\| \ \forall \boldsymbol{z}_1, \boldsymbol{z}_2$.*
Proof. We easily observe that both p_d and $P_k^{(i)}$ are bounded in $[0, 1]$ (see Property 1). Moreover, p_d and $\|\sum_{k=1}^n \left[P_k^{(i)} - \mathbb{1}(k = y_i)\right]\|$ exhibit opposite trends. Thus, we can derive as

$$p_{d1} > p_{d2} \implies$$
$$\left|\sum_{k=1}^n \left[P_k^{(i)} - \mathbb{1}(k = y_i)\right]\right|_{\boldsymbol{z} = \boldsymbol{z}_1} < \left|\sum_{k=1}^n \left[P_k^{(i)} - \mathbb{1}(k = y_i)\right]\right|_{\boldsymbol{z} = \boldsymbol{z}_2}$$

On the other hand, $\|\boldsymbol{g}_\theta\| \propto \left|\sum_{k=1}^n \left[P_k^{(i)} - \mathbb{1}(k = y_i)\right]\right|$ when $\theta_{\boldsymbol{w}_{y_i}, \boldsymbol{z}_i} \to 0$ and $\theta_{\boldsymbol{w}_k, \boldsymbol{z}_i} \to \pm\frac{\pi}{2} \ \forall k \neq y_i$. Consequently, if $p_{d1} > p_{d2}$, then $\|\boldsymbol{g}_\theta(\boldsymbol{z}_1)\| < \|\boldsymbol{g}_\theta(\boldsymbol{z}_2)\| \ \forall \boldsymbol{z}_1, \boldsymbol{z}_2$.

D. Constraints for QCFace

In this section, we establish and prove three constraints for the convergence of optimization and hyperparameter selection in ***QCFace***. Firstly, we demonstrate the convergence ability of ***QCFace*** (see Sec. D.1). Secondly, we establish a condition to ensure the linearity of magnitude encoding (see Sec. D.2). Lastly, we add a constraint for bias value to observe whether $\mathcal{L}_{reg}$ converges (see Sec. D.3).

D.1. Constraint for convergence

In Eq. (14), l_a and u_a are respectively set to 1 and 100, which are lower and upper bounds of the recognizability score of a face image (*i.e.*, $score \in [1, 100]$), according to the work by Chen *et al.* [2]. To certainly reach the expected convergence point in the optimization process, $\mathcal{L}_{reg}$ is required to have the following three properties for all images:
• $\mathcal{L}_{reg}$ is a strictly convex downward function.

- $\mathcal{L}_{reg}$ has an optimal unique value in range $[l_a, u_a]$ (i.e., $\|\mathbf{z}\|^* \in [l_a, u_a]$).
- The optimal magnitude value ($\|\mathbf{z}\|^*$) of $\mathcal{L}_{reg}$ for each value p_d is monotonically increasing as the recognizability expressed by **guidance value** (p_d) increases.

For the demonstrations, we first introduce and prove three lemmas.

Lemma 3. *If $k > 0$ and $p_d \in [0, 1]$, then $\mathcal{L}_{reg}$ is a strictly convex downward function.*

Proof. We first calculate the first and second order derivatives of the loss $\mathcal{L}_{reg}$ with respect to $\|\mathbf{z}\|$.

$$\frac{\partial \mathcal{L}_{reg}}{\partial \|\mathbf{z}\|} = k.p_d.\left(-\frac{1}{\|\mathbf{z}\|^2} + \frac{1}{u_a^2}\right) + (1 - p_d).\left(-\frac{1}{\|\mathbf{z}\|^2} + \frac{1}{l_a^2}\right)$$
$$\implies \frac{\partial^2 \mathcal{L}_{reg}}{(\partial \|\mathbf{z}\|)^2} = [1 + (k - 1).p_d] . \frac{2}{\|\mathbf{z}\|^3}$$

By analyzing each component in $\frac{\partial^2 \mathcal{L}_{reg}}{\partial \|\mathbf{z}\|^2}$, we have

$$\|\mathbf{z}\| \geq 0 \quad \forall \mathbf{z} \in \mathbb{R}^d \implies \frac{2}{\|\mathbf{z}\|^3} > 0$$

and

$$1 + (k - 1).p_d > 0 \quad \forall k > 0, \ \forall p_d \in [0, 1]$$

Subsequently, $\frac{\partial^2 \mathcal{L}_{reg}}{(\partial \|\mathbf{z}\|)^2} > 0$, thus, $\mathcal{L}_{reg}$ is a strictly convex downward function.

Lemma 4. *If $l_a < u_a$ and $k > 0$, then $\mathcal{L}_{reg}$ has an optimal unique magnitude value in range $[l_a, u_a]$ (i.e., $\|\mathbf{z}\|^* \in [l_a, u_a]$).*

Proof. According to Lemma 1, $\mathcal{L}_{reg}$ is a strictly convex downward function if $k > 0$, we have

$$\frac{\partial \mathcal{L}_{reg}}{\partial \|\mathbf{z}\|^i} > \frac{\partial \mathcal{L}_{reg}}{\partial \|\mathbf{z}\|^{ii}} \quad \forall \|\mathbf{z}\|^i, \|\mathbf{z}\|^{ii} \in [l_a, u_a]$$
$$if \ u_a \geq \|\mathbf{z}\|^i > \|\mathbf{z}\|^{ii} \geq l_a$$

Therefore, if there exists an optimal solution $\|\mathbf{z}\|^* \in [l_a, u_a]$, then it is an unique solution.
Let us consider the value of derivatives of $\mathcal{L}_{reg}$ of l_a, u_a

$$\begin{cases} \frac{\partial \mathcal{L}_{reg}}{\partial \|\mathbf{z}\|}(u_a) &= (1 - p_d).\left(-\frac{1}{u_a^2} + \frac{1}{l_a^2}\right) \\ \frac{\partial \mathcal{L}_{reg}}{\partial \|\mathbf{z}\|}(l_a) &= k.p_d.\left(-\frac{1}{l_a^2} + \frac{1}{u_a^2}\right) \end{cases}$$

We also have $u_a > l_a$ since they are upper and lower bounds of $\|\mathbf{z}\|$, respectively. Thus, we can derive

$$\begin{cases} \frac{\partial \mathcal{L}_{reg}}{\partial \|\mathbf{z}\|}(u_a) &> 0 \\ \frac{\partial \mathcal{L}_{reg}}{\partial \|\mathbf{z}\|}(l_a) &< 0 \end{cases}$$

As $\frac{\partial \mathcal{L}_{reg}}{\partial \|\mathbf{z}\|}$ is monotonically and strictly increasing, there must exist an optimal unique value $\|\mathbf{z}\|^* \in [l_a, u_a]$ which has a 0 derivative.

Lemma 5. *If $k > 0$ and $l_a < u_a$, then the optimal magnitude value (i.e., $\|\mathbf{z}\|^*$) of $\mathcal{L}_{reg}$ for each p_d value is monotonically increasing as p_d increases.*
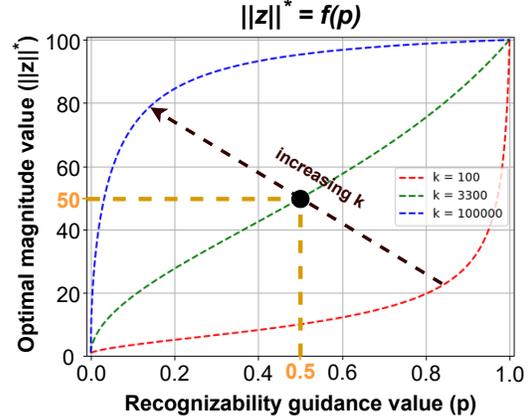

Figure 3. The visualization of function analysis of $f(p_d)$ $\forall p_d \in [0, 1]$.

Proof. Assume that k, l_a and u_a have already been chosen. To find the formula of the optimal magnitude value $\|\mathbf{z}\|^*$ depending on p_d, we resolve $\frac{\partial \mathcal{L}_{reg}}{\partial \|\mathbf{z}\|} = 0$

$$\implies \|\mathbf{z}\|^* = \sqrt{\frac{[1+(k-1).p_d].u_a^2.l_a^2}{u_a^2 + (k.l_a^2 - u_a^2).p_d}} \quad (16)$$
$$(\|\mathbf{z}\|^* > 0 \ \forall \mathbf{z} \in \mathbb{R}^d)$$
$$= f_{\mathbf{z}}(p_d)$$

To make $\|\mathbf{z}\|^*$ be monotonically increasing as p_d increases, we have to satisfy

$$\frac{\partial f_{\mathbf{z}}}{\partial p_d} > 0 \quad (17)$$
$$\iff k.(u_a^2 - l_a^2) > 0$$

We also have $u_a^2 - l_a^2 > 0$ since $u_a > l_a$, and $k > 0$ by the constraint of convergence (see Lemma 3), thus, the Proposition (17) is always true. Therefore, $\|\mathbf{z}\|^*$ is monotonically increasing as p_d increases.

Theorem 1. *The regularization loss function $\mathcal{L}_{reg}$ certainly converges at the expected point in training when $k > 0$, $l_a < u_a$ and $0 \leq p_d \leq 1$.*

Proof. Immediate from Lemmas 3, 4 and 5.

D.2. Constraint for balancing encoding value

To enhance the ability of recognizability representation, the $\|\mathbf{z}\|^*$ value must be guaranteed to be linearly increasing as p_d increases. This constraint makes the value of feature magnitude linearly distributed in the range value of p_d ($p_d \in [0, 1]$). Fig. 3 shows the effect of the value k on the linearity of $f_{\mathbf{z}}(p_d)$ (see Eq. (16)).

Therefore, to balance the range of encoded values, k must be chosen to satisfy

$$\begin{cases} f(p_d = 0) &= l_a \\ f(p_d = 1) &= u_a \\ f(p_d = 0.5) &= \frac{l_a + u_a}{2} \end{cases}$$
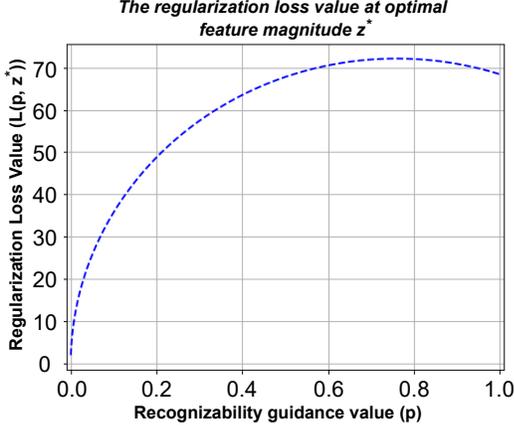
Figure 4. The visualization of loss value at optimal magnitude value $\|\boldsymbol{z}\|^*$, $\mathcal{L}_{reg}(\|\boldsymbol{z}\|^*)$.
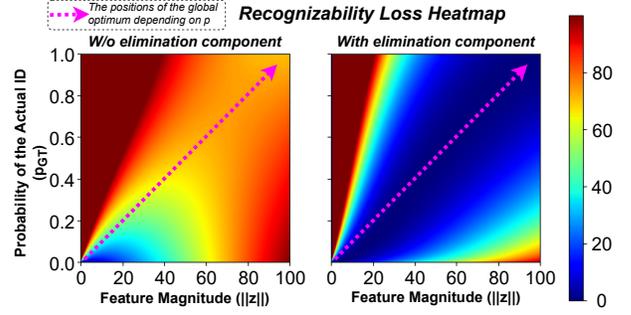

Figure 5. Illustration of two heatmaps of loss value without and with reduction offset. The pink dashed arrow expresses the optimum position (*i.e.*, $\langle p^*, \|\boldsymbol{z}\|^* \rangle$) where the loss function achieves minimum value. The color of heatmap expresses the value of regularization loss $\mathcal{L}_{reg}$.

Table 2. The details of evaluation datasets.

Dataset	No. Pairs		No. Images
	Positive	Negative	
AgeDB-30 [18]	3k	3k	12k
CFP-FP [21]	3.5k	3.5k	12k
LFW [9]	3k	3k	12k
CALFW [28]	3k	3k	12k
CPLFW [27]	3k	3k	14k
XQLFW [14]	3k	3k	13.233k
IJB-B [26]	10k	8M	228k
IJB-C [16]	19.5k	15M	469k

The first two constraints are always fulfilled when replacing $p_d = 0$ and $p_d = 1$ in $f(p_d)$. Thus, the third equation is a main condition for k selection. After resolving it, we have

$$k = \frac{u_a^2 \cdot \left[(u_a + l_a)^2 - 4 \cdot u_a^2 \cdot l_a^2 \right]}{l_a^2 \cdot \left[4 \cdot u_a^2 - (u_a + l_a)^2 \right]} \tag{18}$$

However, the choice of k must satisfy the constraint of convergence in training (see Lemma 3), specifically, $k > 0$. Let us consider the numerator and denominator of Eq. (18)

$$\begin{cases} u_a^2 \cdot \left[(u_a + l_a)^2 - 4l_a^2 \right] & > 0 \\ l_a^2 \cdot \left[4u_a^2 - (u_a + l_a)^2 \right] & > 0 \end{cases} \iff 2l_a < ua + l_a < 2u_a \tag{19}$$

Proposition (19) is always true since $u_a > l_a$. Therefore, the chosen k following Eq. (18) still satisfies the convergence constraint.

D.3. Constraint for tracking regularization loss

By replacing $\|\boldsymbol{z}\|$ in Eq. (14) by $\|\boldsymbol{z}\|^*$ in Eq. (16), setting value b to 0 and choosing l_a, u_a and k followed the two above constraints (*i.e.,* Secs. D.1 and D.2), the optimal values of $\mathcal{L}_{reg}$ are plotted by Fig. 4. However, this value makes them difficult to track the convergence point of $\mathcal{L}_{reg}$ since each value of p_d yields a distinct value of optimal loss value $\mathcal{L}_{reg}(p_d, \|\boldsymbol{z}\|^*)$, where $\|\boldsymbol{z}\|^*$ is an optimal magnitude value for each p_d. To get over this problem, b is used as a (reduction) offset value to make $\mathcal{L}_{reg}(p_d, \|\boldsymbol{z}\|^*) = 0$.

$$\implies b = k.p_d. \left(\frac{1}{\|\boldsymbol{z}\|^*} + \frac{\|\boldsymbol{z}\|^*}{u_a^2} \right) + (1 - p_d). \left(\frac{1}{\|\boldsymbol{z}\|^*} + \frac{\|\boldsymbol{z}\|^*}{l_a^2} \right)$$

Fig. 5 shows the heat maps of the loss value. With the reduction offset, the heat map turns out that $\mathcal{L}_{reg}$ achieves a value of zero at all optimal points. Besides, the gradients from the loss function of both situations (*i.e.,* with and without reduction offset) are similar to one another since b is a

constant value and does not contribute to the update gradient. Consequently, b is only used to monitor whether $\mathcal{L}_{reg}$ reaches a value of zero, without affecting the optimal result.

E. Datasets and augmentation

For evaluation datasets, AgeDB-30 [18], CFP-FP [21], LFW [9], CALFW [28], CPLFW [27], XQLFW [14], IJB-B [26], IJB-C [16] and TinyFace [3] are used as benchmark datasets, which are widely used in almost SoTA studies. They all play an important role in the comprehensive evaluation of model recognition ability and fair comparison with other methods. The details of each evaluation dataset are shown in Tab. 2. In addition, the details of augmentation, which also affect the learning improvement of a model, are shown in Tab. 3.

F. Warming-up in QCFace training

Based on the evaluation with IResNet18, the negative effect of proxy oscillation readily emerges, as demonstrated by the convergence evaluation results. Specifically, the FR model must be pre-trained with ArcFace before applying *QCFace* in order to obtain a meaningful *guidance value*,

Table 3. The details of the experimental augmentations.

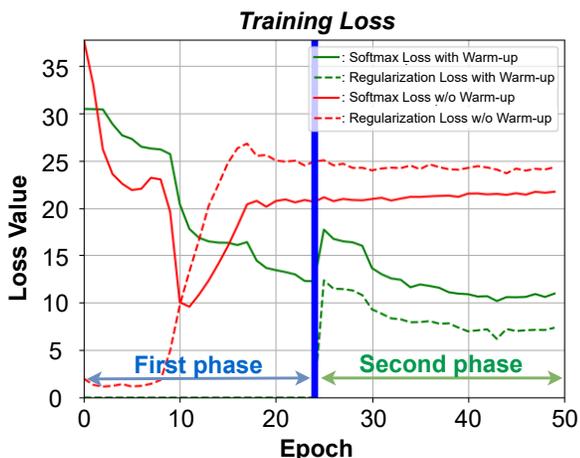| Augment Operator	Details
Horizontal & Vertical Flip	p=0.3
Rotation	p=0.3; angle=90°
Blur	p=0.3; type={Gaussian, Median}
CLAHE	p=0.3
Brightness & Contrast	p=0.3
Image Compression	p=1.0; quality (%) = [20, 100]

Figure 6. The training loss visualization of **QCFace**.[**]

Table 4. The verification performance of **QCFace-Arc** in different training configurations. Cfg-1, Cfg-2 and Cfg-3 express non-warm-up, non-augmentation and full setting, respectively.[**]

Config	AgeDB-30	CFP-FP	LFW	CALFW	CPLFW
Cfg-1	49.900	53.700	81.600	70.650	53.900
Cfg-2	91.217	88.457	98.783	91.683	82.200
Cfg-3	92.050	91.429	99.117	92.150	84.700

called *warm-up training phase*. Fig. 6 shows that without the *warm-up phase*, $\mathcal{L}_{sm}$ and $\mathcal{L}_{rc}$ do not decrease simultaneously in the first phase; even so, both loss values start increasing at epoch 11, and cannot be optimized later. In contrast, with the *warm-up*, the convergence process is smooth in both training phases. Tab. 4 further demonstrates in verification accuracy of the necessity of the *warm-up phase* for the **frozen proxy** guarantee, and provides a comparison between Cfg-1 and Cfg-2. However, as we mentioned in the main paper, this problem can be overcome by a large-size backbone and a large-scale training dataset.

G. Augmentation Effect

In Tab. 4, the use of augmentation in **QCFace** yields performance gains of 0.740%, 4.003%, 0.304%, 0.709%, and 3.211% on AgeDB-30, CFP-FP, LFW, CALFW, and CPLFW, respectively. This substantial effect appears when utilizing small-scale training datasets (*e.g.*, CASIA-WebFace [6]) and small-size backbones (*i.e.*, IResNet18).

H. Extended visualization & Experiments

In this section, we extend the visualization of the geometrical representation of the embedded feature (see Sec. H.1). Moreover, we show the comprehensive evaluation of verification and identification with two backbone scales of IResNet [8], *i.e.* IResNet18 and IResNet100 (see Sec. H.2).

H.1. Embedded feature distribution comparison

We add the embedded feature representations of ArcFace [5] and CurricularFace [10], whose embedded features are also uniformly distributed in the low range of magnitude value (< 15), similar to AdaFace (Figs. 7a, 7b and 7d). Meanwhile, **QCFace** shows its superiority with its widest range of feature magnitude for representing the diversity of the recognizability level of face images.

H.2. Verification and Identification evaluation

For the comprehensive evaluation, we also show the verification and identification benchmarks of both IResNet18 and IResNet100 based on high-quality (see Tabs. 5 and 6), mixed-quality and low-quality benchmark datasets (see Tabs. 7 and 8 for verification benchmarks, and Tabs. 9 and 10 for identification benchmarks as well as MegaFace Challenge [12]).

For high-quality benchmarks, the negative effect of proxy oscillation readily emerges, clearly shown by the evaluation results of **(Q)QCFace-Cur** and **(Q)QCFace-MVS** trained with IResNet18, while **(Q)QCFace-Arc** also achieves top-1 and top-2 verification accuracy.

The limitation in recognizing very low-quality images is clearly demonstrated by the evaluation results of **QCFace** using backbone IResNet18 as its feature extractor (see **Limitations** in main paper). Specifically, the recognition performance dramatically drops when benchmarking on the mixed-quality (*i.e.*, IJB-B and IJB-C) and low-quality (*i.e.*, TinyFace) datasets (see Tabs. 8 and 10). Consequently, the large-scale backbone can enhance generalization ability, mitigating the limited impact of low-recognizability images in the presence of mislabeled data in low-magnitude representative region.

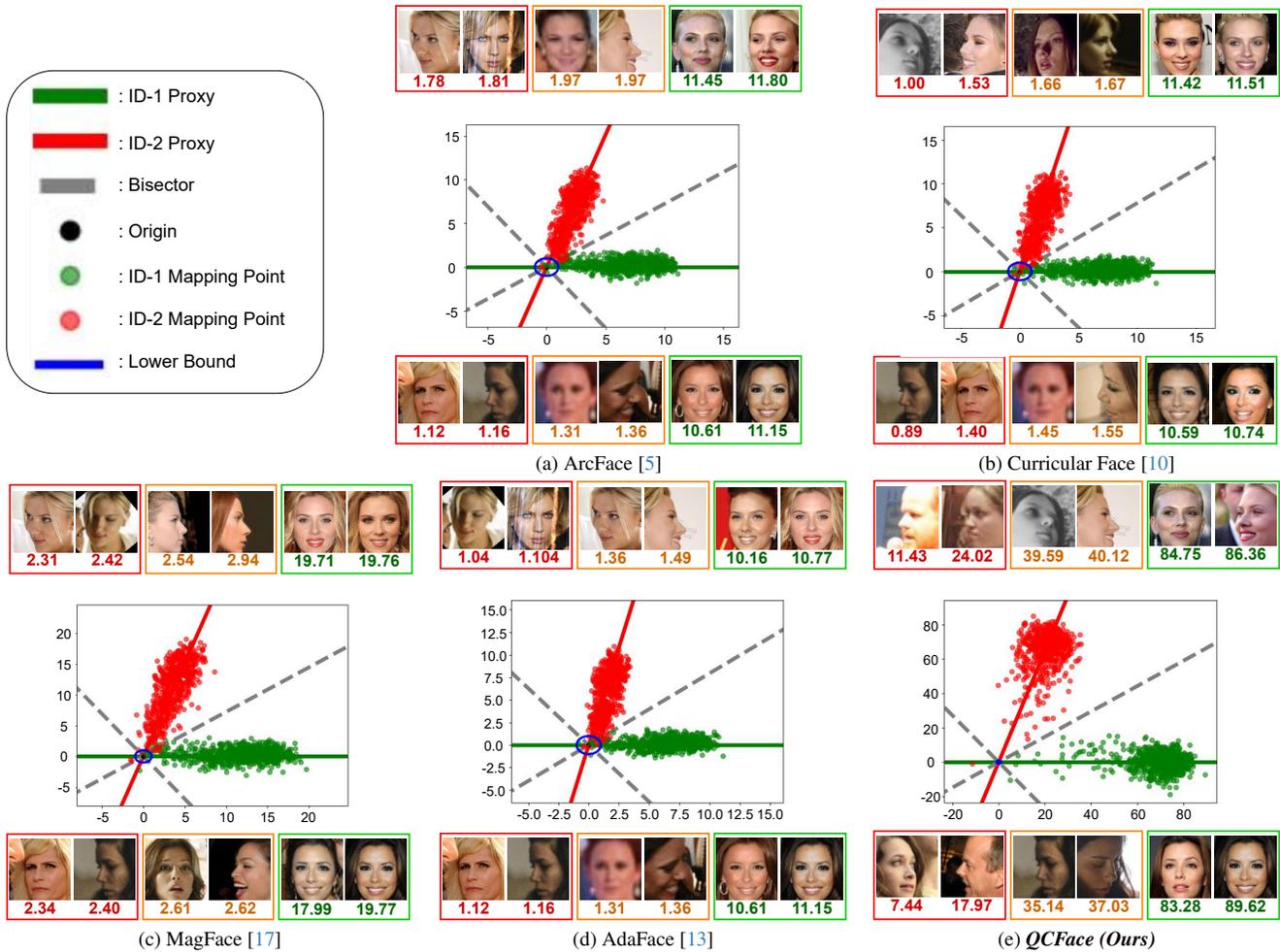[**]These experimental results are done with the IResNet18 backbone.

Figure 7. The non-normalized geometrical representation of the feature space optimized by **QCFace-Arc** and the cutting-edge methods. The feature vectors of each image extracted by face backbones are projected to the hyperplane formed by two proxies of two identities in the CASIA-WebFace dataset by the Gram-Schmidt algorithm. The **bold** red and green lines express the proxies of two identities. The actual proxy of each mapping feature point has the same color as that of feature. The gray dashed lines are the internal and external bisectors. The number below each image sample is the magnitude of the mapping feature vectors.

References

[1] Fadi Boutros, Naser Damer, Florian Kirchbuchner, and Arjan Kuijper. Elasticface: Elastic margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1578–1587, 2022. 8, 9, 10

[2] Jiansheng Chen, Yu Deng, Gaocheng Bai, and Guangda Su. Face image quality assessment based on learning to rank. *IEEE signal processing letters*, 22(1):90–94, 2014. 3

[3] Zhiyi Cheng, Xiatian Zhu, and Shaogang Gong. Low-resolution face recognition. In *Asian conference on computer vision*, pages 605–621. Springer, 2018. 5

[4] Jun Dan, Yang Liu, Jiankang Deng, Haoyu Xie, Siyuan Li, Baigui Sun, and Shan Luo. Topofr: A closer look at topology alignment on face recognition. In *Advances in Neural Information Processing Systems*, pages 37213–37240, 2024.

[5] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4690–4699, 2019. 1, 2, 6, 7, 8, 9, 10

[6] Jiankang Deng, Jia Guo, Debing Zhang, Yafeng Deng, Xiangju Lu, and Song Shi. Lightweight face recognition challenge. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pages 0–0, 2019. 6

[7] Jiankang Deng, Jia Guo, Jing Yang, Alexandros Lattas, and Stefanos Zafeiriou. Variational prototype learning for deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11906–11915, 2021. 8, 9, 10

[8] Ionut Cosmin Duta, Li Liu, Fan Zhu, and Ling Shao. Improved residual networks for image and video recognition.

Table 5. Comparison in verification accuracy and AUC-ROC on high-quality datasets with IResNet100. The **bold**, underlined and *italic* numbers correspondingly express **top-1**, top-2 and *top-3* accuracy. In notation *(Q)QCFace-X*, *(Q)* expresses the use of [22] as a similarity score calculation instead of cosine similarity in decision making, *X* expresses the name of softmax loss (*i.e.*, Arc - ArcFace, Cur - CurricularFace, MVS - MVSoftmax). Green and red backgrounds express the use of margin-based and misclassified softmax losses.

Year	Method	AgeDB-30		CFP-FP		LFW		CALFW		CPLFW		XQLFW	
		Ver	AUC	Ver	AUC	Ver	AUC	Ver	AUC	Ver	AUC	Ver	AUC
2015	FaceNet [20]	58.150	61.360	70.314	77.734	79.417	87.862	59.567	64.095	62.633	64.760	66.650	72.543
2017	SphereFace [15]	96.100	98.843	93.729	97.783	99.600	99.852	95.400	97.983	88.917	93.759	79.883	87.493
2018	CosFace [23]	97.033	98.866	95.743	98.497	99.750	99.862	95.467	97.866	90.317	94.635	81.783	90.128
2019	ArcFace [5]	97.133	98.917	95.100	98.521	99.633	99.878	95.517	97.921	90.217	94.707	81.833	90.087
2020	MV-Arc-Softmax [24]	97.100	98.972	95.757	98.541	99.700	99.826	95.617	97.918	90.233	94.414	83.150	90.814
2020	CurricularFace [10]	97.100	98.962	94.686	98.418	99.550	99.846	95.700	97.929	90.083	94.607	82.650	90.265
2021	MagFace [17]	96.867	98.797	95.357	98.593	99.600	99.889	95.517	97.797	89.583	94.105	81.233	89.090
2021	VPLFace [7]	97.183	98.984	94.257	98.163	99.700	99.865	95.683	97.858	89.617	94.483	83.117	90.677
2022	SphereFace2 [25]	96.000	98.777	94.029	98.210	99.583	99.878	95.283	97.949	90.100	94.677	83.933	91.551
2022	Elastic-Arc [1]	97.283	98.930	94.329	98.004	99.467	99.851	95.500	97.799	89.800	94.156	82.467	89.896
2022	Elastic-Cos [1]	97.417	98.979	95.357	98.432	99.700	99.890	95.733	97.837	89.400	94.395	82.350	89.729
2022	AdaFace [13]	96.817	98.924	95.614	98.723	99.717	99.866	95.733	97.790	90.917	94.957	82.017	89.839
2023	QAFace [19]	96.783	98.942	95.243	98.469	99.733	99.887	95.400	97.877	89.817	94.430	83.550	90.946
2023	QMagFace [22]	97.250	98.895	96.014	99.119	99.700	99.916	95.617	97.801	90.683	96.042	81.483	88.922
2023	UniFace [29]	97.083	98.903	95.686	98.749	99.617	99.829	95.667	97.856	90.683	95.029	82.700	90.755
2023	UniTSFace [11]	97.217	99.008	95.371	98.550	99.667	99.889	95.717	97.936	90.450	94.835	83.517	91.171
2024	TopoFR [4]	97.500	99.021	95.871	99.015	99.750	99.876	95.650	97.913	90.383	95.490	83.583	91.070
Now	**QCFace-Arc**	98.100	99.129	98.200	99.372	**99.800**	99.872	95.950	98.031	92.483	95.624	86.067	93.327
Now	**QQCFace-Arc**	**98.183**	**99.175**	**98.457**	99.524	**99.800**	99.879	95.950	**98.087**	92.833	96.403	**86.267**	**93.634**
Now	**QCFace-Cur**	97.700	98.979	98.200	99.515	99.783	99.895	**96.083**	98.031	92.383	96.404	83.283	91.060
Now	**QQCFace-Cur**	97.917	99.013	98.271	**99.528**	99.767	**99.900**	96.067	98.032	92.467	**96.718**	83.283	91.466
Now	**QCFace-MVS**	93.200	98.101	96.229	98.971	99.483	99.897	91.683	96.853	89.133	93.874	82.833	91.058
Now	**QQCFace-MVS**	93.517	98.066	96.186	99.001	99.467	99.892	91.833	96.757	88.817	93.692	84.300	92.166

Table 6. Comparison in verification accuracy and AUC-ROC on high-quality datasets with IResNet18. The **bold**, underlined and *italic* numbers correspondingly express **top-1**, top-2 and *top-3* accuracy. In notation *(Q)QCFace-X*, *(Q)* expresses the use of [22] as a similarity score calculation instead of cosine similarity in decision making, *X* expresses the name of softmax loss (*i.e.*, Arc - ArcFace, Cur - CurricularFace, MVS - MVSoftmax). Green and red backgrounds express the use of margin-based and misclassified softmax losses.

Year	Method	AgeDB-30		CFP-FP		LFW		CALFW		CPLFW		XQLFW	
		Ver	AUC	Ver	AUC	Ver	AUC	Ver	AUC	Ver	AUC	Ver	AUC
2015	FaceNet [20]	66.300	71.282	71.429	79.354	84.250	92.479	66.183	71.948	63.283	67.745	62.483	67.077
2017	SphereFace [15]	81.117	89.157	83.757	91.786	95.317	99.128	81.767	89.814	75.383	82.828	70.167	76.479
2018	CosFace [23]	89.633	95.895	82.114	91.183	98.433	99.833	90.683	96.340	79.867	85.792	64.950	67.418
2019	ArcFace [5]	90.017	96.285	78.243	88.674	98.500	99.818	91.217	96.381	79.150	83.315	64.150	67.732
2020	MV-Arc-Softmax [24]	86.383	93.664	80.629	89.069	97.600	99.677	87.767	94.898	76.033	82.560	64.867	67.384
2020	CurricularFace [10]	88.767	95.758	83.629	91.859	98.617	99.808	90.900	96.411	79.517	85.891	65.833	66.843
2021	MagFace [17]	89.483	96.100	83.286	92.320	98.667	99.804	91.767	96.567	79.550	86.057	65.933	70.257
2021	VPLFace [7]	91.150	96.989	87.671	94.571	98.550	99.759	92.050	96.774	83.567	89.949	71.983	79.611
2022	SphereFace2 [25]	91.683	96.875	88.214	94.793	98.717	99.774	92.000	96.599	82.850	89.916	75.317	82.526
2022	Elastic-Arc [1]	89.617	96.010	84.729	92.704	98.317	99.725	90.933	96.369	80.467	87.350	76.150	83.565
2022	Elastic-Cos [1]	88.167	95.249	84.357	91.783	97.933	99.692	90.533	96.146	80.833	87.970	76.233	84.555
2022	AdaFace [13]	89.000	95.262	83.014	88.513	98.433	99.873	91.117	96.502	80.133	85.569	66.433	66.590
2023	QAFace [19]	51.067	51.813	50.886	51.563	52.667	54.955	51.983	52.106	51.383	50.680	52.317	51.745
2023	QMagFace [22]	89.633	95.997	80.800	86.973	98.917	99.868	91.817	96.833	80.250	84.760	61.717	64.426
2023	UniFace [29]	89.167	95.638	87.600	94.662	98.617	99.808	91.400	96.366	82.533	89.709	70.050	76.759
2023	UniTSFace [11]	91.133	96.930	88.671	95.301	98.950	99.824	92.433	96.846	84.067	90.176	76.317	84.031
2024	TopoFR [4]	87.183	94.154	77.457	86.262	97.600	99.598	89.567	95.465	76.950	83.268	**79.667**	**87.483**
Now	**QCFace-Arc**	**92.067**	97.335	**91.429**	**97.143**	**99.083**	99.871	92.150	97.133	84.700	**91.298**	75.633	83.475
Now	**QQCFace-Arc**	91.833	**97.369**	90.900	96.888	99.000	**99.885**	**92.483**	**97.149**	**84.750**	90.755	74.883	82.362
Now	**QCFace-Cur**	79.783	88.209	81.314	89.590	96.150	99.317	82.917	90.405	75.217	83.205	72.117	79.008
Now	**QQCFace-Cur**	79.517	87.831	80.343	88.777	96.233	99.315	82.833	90.652	75.633	83.170	71.417	78.954
Now	**QCFace-MVS**	78.817	87.213	79.857	88.222	95.617	99.232	83.283	90.844	75.500	84.020	72.333	78.991
Now	**QQCFace-MVS**	78.167	86.612	78.829	87.097	95.733	99.178	83.300	90.906	76.283	83.773	72.017	78.753

In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 9415–9422. IEEE, 2021. 6

[9] Gary B Huang, Marwan Mattar, Tamara Berg, and Eric Learned-Miller. Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. In *Workshop on faces in'Real-Life'Images: detection, alignment, and recognition*, 2008. 5

[10] Yuge Huang, Yuhan Wang, Ying Tai, Xiaoming Liu,

Table 7. Comparison in verification performance on IJB dataset with IResNet100.

Year	Method	IJB-B (TAR@FAR)							IJB-C (TAR@FAR)						
		10^{-6}	10^{-5}	10^{-4}	10^{-3}	10^{-2}	10^{-1}	AUC	10^{-6}	10^{-5}	10^{-4}	10^{-3}	10^{-2}	10^{-1}	AUC
2015	FaceNet [20]	0.44	0.89	2.46	7.55	25.59	69.98	90.25	0.43	1.06	2.85	9.16	29.24	71.29	90.29
2017	SphereFace [15]	33.33	57.43	79.64	92.04	96.74	98.73	99.46	27.76	54.16	79.74	93.20	97.34	98.97	99.53
2018	CosFace [23]	37.81	84.36	91.82	95.27	97.45	98.81	99.44	82.44	90.02	94.07	96.45	98.01	99.06	99.57
2019	ArcFace [5]	36.71	85.15	91.61	95.09	97.23	98.74	99.47	82.85	89.40	93.66	96.22	97.88	99.04	99.59
2020	MV-Arc-Softmax [24]	39.02	80.41	90.87	94.73	97.23	98.70	99.44	79.61	87.82	93.26	96.05	97.97	99.09	99.59
2020	CurricularFace [10]	37.63	84.99	92.27	95.22	97.15	98.68	99.40	85.12	90.79	94.20	96.35	97.84	99.00	99.53
2021	MagFace [17]	38.10	83.59	91.47	95.16	97.38	98.64	99.32	81.64	88.81	93.38	96.17	97.93	99.00	99.51
2021	VPLFace [7]	36.90	85.55	91.64	95.19	97.36	98.70	99.42	86.52	90.26	93.79	96.29	97.94	99.05	99.55
2022	SphereFace2 [25]	36.79	80.86	90.10	94.50	97.26	98.93	99.50	78.70	86.67	92.34	95.73	97.87	99.20	99.63
2022	Elastic-Arc [1]	36.48	83.97	91.79	95.01	97.16	98.75	99.45	80.43	89.95	94.33	96.18	97.81	99.07	99.56
2022	Elastic-Cos [1]	34.37	83.72	91.67	95.03	97.30	98.79	99.50	82.48	89.65	93.80	96.19	97.95	99.05	99.56
2022	AdaFace [13]	35.01	85.07	92.44	95.31	97.27	98.52	99.38	80.43	89.95	94.33	96.49	98.05	98.98	99.56
2023	QAFace [19]	40.00	80.19	89.49	93.99	97.26	98.82	99.54	79.02	87.17	92.27	95.63	97.92	99.13	99.63
2023	QMagFace [22]	36.80	83.42	88.25	93.17	96.57	98.29	99.38	81.24	87.28	91.23	95.16	97.69	98.90	99.61
2023	UniFace [29]	37.16	85.46	92.16	95.30	97.45	98.74	99.41	82.62	90.35	94.21	96.47	98.08	99.15	99.57
2023	UniTSFace [11]	34.35	82.71	91.64	95.20	97.40	98.84	99.47	79.78	88.66	93.64	96.42	98.10	99.14	99.64
2024	TopoFR [4]	40.25	82.88	91.25	95.25	97.41	98.87	99.49	77.94	88.28	93.30	96.42	98.08	99.20	99.63
Now	QCFace-Arc	34.42	89.39	94.30	96.23	97.73	98.75	99.44	88.85	93.82	95.84	97.37	98.45	99.16	99.62
Now	QQCFace-Arc	37.89	88.80	93.93	95.90	97.28	98.51	99.48	80.57	91.70	95.48	97.01	98.12	98.93	99.66

Table 8. Comparison in verification performance on IJB dataset with IResNet18.

Year	Method	IJB-B (TAR@FAR)							IJB-C (TAR@FAR)						
		10^{-6}	10^{-5}	10^{-4}	10^{-3}	10^{-2}	10^{-1}	AUC	10^{-6}	10^{-5}	10^{-4}	10^{-3}	10^{-2}	10^{-1}	AUC
2015	FaceNet [20]	0.93	2.33	6.23	18.47	43.25	78.80	93.34	0.92	2.62	7.52	7.52	45.09	79.83	93.50
2017	SphereFace [15]	4.77	11.74	28.29	51.25	75.54	93.69	97.59	5.17	14.24	30.60	53.56	77.44	94.70	97.89
2018	CosFace [23]	0.10	0.22	1.22	15.84	70.88	95.23	97.72	0.05	0.18	0.91	15.73	69.41	95.37	97.70
2019	ArcFace [5]	0.03	0.15	0.33	9.60	71.94	94.70	97.55	0.04	0.10	0.23	9.22	70.35	94.97	97.54
2020	MV-Arc-Softmax [24]	0.26	1.03	7.92	37.19	77.31	95.23	97.98	0.24	1.54	8.84	37.55	76.41	95.21	98.00
2020	CurricularFace [10]	0.35	1.30	7.64	33.45	76.11	95.30	97.97	0.38	1.80	8.44	34.60	75.62	95.74	98.04
2021	MagFace [17]	0.11	0.37	2.22	28.19	79.66	95.71	98.15	0.10	0.28	2.17	28.70	79.12	95.97	98.19
2021	VPLFace [7]	31.95	60.81	73.35	84.00	92.15	97.49	99.00	56.51	68.83	78.13	86.72	93.77	97.96	99.21
2022	SphereFace2 [25]	28.22	59.82	74.55	84.99	92.91	97.70	99.10	54.87	68.08	78.73	87.57	94.23	98.18	99.23
2022	Elastic-Arc [1]	33.61	54.86	69.17	81.22	90.91	97.23	98.95	49.30	61.72	72.98	83.70	92.10	97.70	99.10
2022	Elastic-Cos [1]	27.56	52.57	68.95	81.29	90.93	97.14	98.91	47.67	60.88	73.62	84.21	92.67	97.71	99.10
2022	AdaFace [13]	0.13	0.25	2.41	36.39	82.28	96.04	98.23	0.06	0.27	2.85	37.02	82.14	96.42	98.31
2023	QAFace [19]	0.03	0.07	0.17	0.49	1.94	12.82	55.29	0.03	0.05	0.16	0.49	2.16	13.59	56.05
2023	QMagFace [22]	0.19	0.28	0.67	4.08	44.52	75.41	86.75	0.05	0.09	0.43	3.07	40.79	76.21	88.63
2023	UniFace [29]	6.97	26.88	57.83	78.58	90.64	97.40	98.98	13.47	35.51	60.35	79.63	91.34	97.75	99.09
2023	UniTSFace [11]	32.53	61.27	74.66	85.19	92.97	97.95	99.20	57.66	68.11	78.70	87.44	94.17	98.34	99.35
2024	TopoFR [4]	0.00	0.15	14.15	62.94	85.28	95.28	98.07	0.01	0.04	8.73	61.07	86.56	95.80	98.21
Now	QCFace-Arc	0.68	2.78	31.73	71.87	91.58	97.79	99.16	0.74	3.74	29.82	72.81	92.60	98.10	99.24
Now	QQCFace-Arc	0.69	2.59	29.10	67.52	89.52	96.25	98.39	0.70	3.28	25.86	68.36	91.11	97.10	98.77

Pengcheng Shen, Shaoxin Li, Jilin Li, and Feiyue Huang. Curricularface: adaptive curriculum learning loss for deep face recognition. In *proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5901–5910, 2020. 1, 2, 6, 7, 8, 9, 10

[11] Xi Jia, Jiancan Zhou, Linlin Shen, Jinming Duan, et al. Unitsface: Unified threshold integrated sample-to-sample loss for face recognition. *Advances in Neural Information Processing Systems*, 36:32732–32747, 2023. 8, 9, 10

[12] Ira Kemelmacher-Shlizerman, Steven M Seitz, Daniel Miller, and Evan Brossard. The megaface benchmark: 1 million faces for recognition at scale. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4873–4882, 2016. 6

[13] Minchul Kim, Anil K Jain, and Xiaoming Liu. Adaface: Quality adaptive margin for face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 18750–18759, 2022. 2, 7, 8, 9, 10

[14] Martin Knoche, Stefan Hormann, and Gerhard Rigoll. Cross-quality lfw: A database for analyzing cross-resolution image face recognition in unconstrained environments. In *2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021)*, pages 1–5. IEEE, 2021. 5

[15] Weiyang Liu, Yandong Wen, Zhiding Yu, Ming Li, Bhiksha Raj, and Le Song. Sphereface: Deep hypersphere embedding for face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 212–220, 2017. 1, 2, 8, 9, 10

[16] Brianna Maze, Jocelyn Adams, James A Duncan, Nathan Kalka, Tim Miller, Charles Otto, Anil K Jain, W Tyler Niggel, Janet Anderson, Jordan Cheney, et al. Iarpa janus benchmark-c: Face dataset and protocol. In *2018 international conference on biometrics (ICB)*, pages 158–165.

Table 9. Comparison in identification performance on IJB, TinyFace and MegaFace datasets with IResNet100. In the MegaFace benchmark, *Iden* expresses *Rank-1* in identification and *Veri* is TAR@FAR=10^{-6}. The similarity score calculation of [22] is not supported by MegaFace devkit, expressed by the notation "-".

Year	Method	IJB-B		IJB-C		TinyFace		MegaFace	
		Rank-1	*Rank-5*	*Rank-1*	*Rank-5*	*Rank-1*	*Rank-5*	*Iden*	*Veri*
2015	FaceNet [20]	8.997	19.628	7.788	16.240	11.534	17.167	0.724	2.209
2017	SphereFace [15]	89.075	93.204	90.088	93.411	51.261	57.350	89.597	91.387
2018	CosFace [23]	*93.554*	95.891	94.789	96.392	60.381	64.941	95.617	96.260
2019	ArcFace [5]	93.204	95.852	94.314	96.193	57.672	62.607	95.242	95.895
2020	MV-Arc-Softmax [24]	92.882	95.813	94.233	96.172	59.120	64.297	94.188	94.944
2020	CurricularFace [10]	93.476	95.794	94.794	96.366	60.649	64.941	95.637	96.316
2021	MagFace [17]	92.911	95.803	94.187	96.045	57.833	62.527	94.965	96.052
2021	VPLFace [7]	93.427	*96.076*	94.666	96.448	57.913	62.446	94.937	95.978
2022	SphereFace2 [25]	92.162	95.170	93.610	95.743	57.994	63.251	92.595	94.224
2022	Elastic-Arc [1]	92.911	95.813	94.391	96.203	57.779	62.607	95.674	96.429
2022	Elastic-Cos [1]	93.262	96.008	94.473	96.315	60.086	64.565	94.989	95.602
2022	AdaFace [13]	93.525	96.018	94.814	*96.524*	*60.837*	64.995	95.888	96.694
2023	QAFace [19]	92.687	95.764	94.018	96.320	57.967	62.983	93.510	94.738
2023	QMagFace [22]	91.870	95.560	93.707	96.213	57.833	62.527	-	-
2023	UniFace [29]	93.350	95.979	*94.886*	96.402	60.327	64.887	*95.961*	96.350
2023	UniTSFace [11]	93.155	95.891	94.442	96.366	60.408	*65.236*	94.409	95.737
2024	TopoFR [4]	93.019	95.813	94.243	96.289	59.871	64.029	*95.913*	*96.636*
Now	*QCFace-Arc*	94.761	96.884	96.157	97.468	62.634	**66.738**	**98.347**	**98.500**
Now	*QQCFace-Arc*	**94.898**	**96.923**	**96.172**	**97.484**	**64.941**	68.160	-	-

Table 10. Comparison in identification performance on IJB, TinyFace and MegaFace datasets with IResNet18. In the MegaFace benchmark, *Iden* expresses *Rank-1* in identification and *Veri* is TAR at FAR=10^{-6}. The similarity score calculation of [22] is not supported by MegaFace devkit, expressed by the notation "-".

Year	Method	IJB-B		IJB-C		TinyFace		MegaFace	
		Rank-1	*Rank-5*	*Rank-1*	*Rank-5*	*Rank-1*	*Rank-5*	*Iden*	*Veri*
2015	FaceNet [20]	20.643	36.943	18.057	32.813	17.543	25.161	*80.015*	*84.670*
2017	SphereFace [15]	52.561	69.961	50.094	67.187	15.397	23.095	18.430	20.975
2018	CosFace [23]	69.903	75.385	70.944	75.746	44.769	51.180	71.948	76.819
2019	ArcFace [5]	69.903	75.959	71.015	76.185	41.550	48.471	71.215	75.920
2020	MV-Arc-Softmax [24]	67.887	75.706	68.392	75.170	48.095	54.533	54.523	57.891
2020	CurricularFace [10]	70.156	75.871	71.464	76.446	45.681	52.173	72.807	74.928
2021	MagFace [17]	71.782	77.897	72.659	78.064	44.260	50.402	71.616	74.771
2021	VPLFace [7]	*83.846*	*90.224*	*85.638*	*90.726*	42.623	49.329	*79.485*	83.521
2022	SphereFace2 [25]	85.268	90.691	86.077	91.027	43.750	49.893	**80.574**	**84.818**
2022	Elastic-Arc [1]	82.210	88.802	83.050	88.950	41.229	47.720	74.005	79.325
2022	Elastic-Cos [1]	82.960	89.026	84.163	89.501	46.057	52.387	73.298	77.391
2022	AdaFace [13]	73.087	78.793	73.853	78.732	47.639	54.211	70.251	74.663
2023	QAFace [19]	0.341	0.818	0.311	0.658	3.782	5.392	0.069	0.022
2023	QMagFace [22]	66.095	71.694	65.467	70.745	44.260	50.402	-	-
2023	UniFace [29]	78.296	85.209	78.850	85.056	**49.249**	**56.330**	70.749	74.664
2023	UniTSFace [11]	**85.287**	**90.740**	**86.480**	**91.089**	48.686	55.660	79.008	82.475
2024	TopoFR [4]	74.839	81.831	75.047	81.514	47.452	53.943	65.542	0.048
Now	*QCFace-Arc*	80.253	85.910	81.305	86.021	*48.578*	*55.123*	80.017	84.681
Now	*QQCFace-Arc*	79.231	84.732	79.824	84.556	*48.578*	*55.123*	-	-

IEEE, 2018. 5

[17] Qiang Meng, Shichao Zhao, Zhida Huang, and Feng Zhou. Magface: A universal representation for face recognition and quality assessment. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14225–14234, 2021. 2, 7, 8, 9, 10

[18] Stylianos Moschoglou, Athanasios Papaioannou, Christos Sagonas, Jiankang Deng, Irene Kotsia, and Stefanos Zafeiriou. Agedb: the first manually collected, in-the-wild age database. In *proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages

51–59, 2017. 5

[19] Mohammad Saeed Ebrahimi Saadabadi, Sahar Rahimi Malakshan, Ali Zafari, Moktari Mostofa, and Nasser M Nasrabadi. A quality aware sample-to-sample comparison for face recognition. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 6129–6138, 2023. 8, 9, 10

[20] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823, 2015. 1, 8,

9, 10

[21] Soumyadip Sengupta, Jun-Cheng Chen, Carlos Castillo, Vishal M Patel, Rama Chellappa, and David W Jacobs. Frontal to profile face verification in the wild. In *2016 IEEE winter conference on applications of computer vision (WACV)*, pages 1–9. IEEE, 2016. 5

[22] Philipp Terhörst, Malte Ihlefeld, Marco Huber, Naser Damer, Florian Kirchbuchner, Kiran Raja, and Arjan Kuijper. Qmagface: Simple and accurate quality-aware face recognition. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3484–3494, 2023. 8, 9, 10

[23] Hao Wang, Yitong Wang, Zheng Zhou, Xing Ji, Dihong Gong, Jingchao Zhou, Zhifeng Li, and Wei Liu. Cosface: Large margin cosine loss for deep face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5265–5274, 2018. 1, 2, 8, 9, 10

[24] Xiaobo Wang, Shifeng Zhang, Shuo Wang, Tianyu Fu, Hailin Shi, and Tao Mei. Mis-classified vector guided softmax loss for face recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 12241–12248, 2020. 1, 2, 8, 9, 10

[25] Yandong Wen, Weiyang Liu, Adrian Weller, Bhiksha Raj, and Rita Singh. Sphereface2: Binary classification is all you need for deep face recognition. In *ICLR*, 2022. 8, 9, 10

[26] Cameron Whitelam, Emma Taborsky, Austin Blanton, Brianna Maze, Jocelyn Adams, Tim Miller, Nathan Kalka, Anil K Jain, James A Duncan, Kristen Allen, et al. Iarpa janus benchmark-b face dataset. In *proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 90–98, 2017. 5

[27] Tianyue Zheng and Weihong Deng. Cross-pose lfw: A database for studying cross-pose face recognition in unconstrained environments. *Beijing University of Posts and Telecommunications, Tech. Rep*, 5(7):5, 2018. 5

[28] Tianyue Zheng, Weihong Deng, and Jiani Hu. Cross-age lfw: A database for studying cross-age face recognition in unconstrained environments. *arXiv preprint arXiv:1708.08197*, 2017. 5

[29] Jiancan Zhou, Xi Jia, Qiufu Li, Linlin Shen, and Jinming Duan. Uniface: Unified cross-entropy loss for deep face recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 20730–20739, 2023. 8, 9, 10