SlackedFace: Learning a Slacked Margin for Low-Resolution Face Recognition

Cheng Yaw Low[1]
chengyawlow@ibs.re.kr
Jacky Chen Long Chai[3]
jackyccl@yonsei.ac.kr
Jaewoo Park[3]
julypraise@yonsei.ac.kr
Kyeongjin An[2]
kyeongjin.an@kaist.ac.kr
Meeyoung Cha[1,2]
meeyoungcha@kaist.ac.kr

[1] Data Science Group,
Institute for Basic Science,
Daejeon, Republic of Korea.

[2] School of Computing,
Korea Advanced Institute of Science
and Technology,
Daejeon, Republic of Korea.

[3] School of Electrical and Electronic
Engineering,
Yonsei University,
Seoul, Republic of Korea.

Abstract

Low-resolution (LR) face recognition poses a significant challenge in embedding learning due to the severe loss of identity information. Recent softmax losses introduce a non-static margin that assigns greater importance or a larger margin to recognizable examples based on embedding norm as a measure of face recognizability. In this paper, we argue that face recognizability is more than just the embedding norm, as it does not capture the identity-level details that are important to embedding learning. We propose SlackedFace to induce a relaxed margin aligned with face recognizability and the model's confidence based on both embedding norm and embedding proximity for empowered embedding learning. We also put forward fast-hill climbing as an early calibration stage between pre-trained and randomly initialized modules. We show that SlackedFace outperforms the current best models on realistic LR face datasets when tested in practical open-set evaluation scenarios.

1 Introduction

Face recognition has made significant strides in recent years, primarily attributed to advancements in deep convolutional neural networks (CNNs) trained with powerful margin-based loss functions designed particularly for open-set evaluation tasks. However, given poor-quality face images contaminated with unrestricted noise and artifacts, e.g., low-resolution (LR) face images detected from surveillance footage, the generalizability of a deep face model trained on a million-scale face dataset, is not transferable to that of LR [18]. Compared to high-resolution (HR), the primary problem with LR face images is that they tend to be unrecognizable due to severe identity information vanishing. This impedes discriminative embedding learning, even with margin-based losses, e.g., SphereFace [17], NormFace [27] CosFace [28], ArcFace [8], to name a few.

© 2023. The copyright of this document resides with its authors.

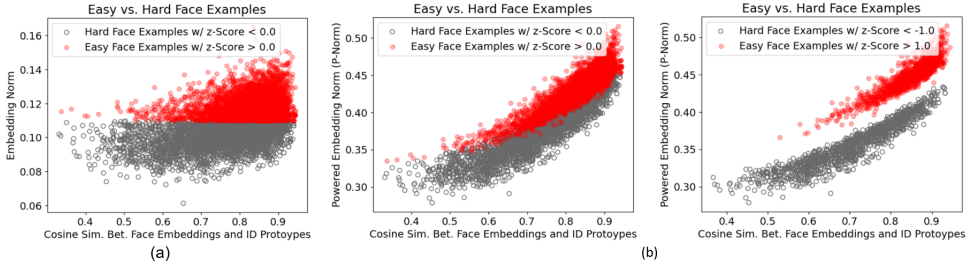

Figure 1: Scatter plots comparing face recognizability index to model predictions. The plot shows that embedding norm does not correlate with model predictions, especially for hard examples with low norm indexes (a), which is corrected by P-Norm (b).

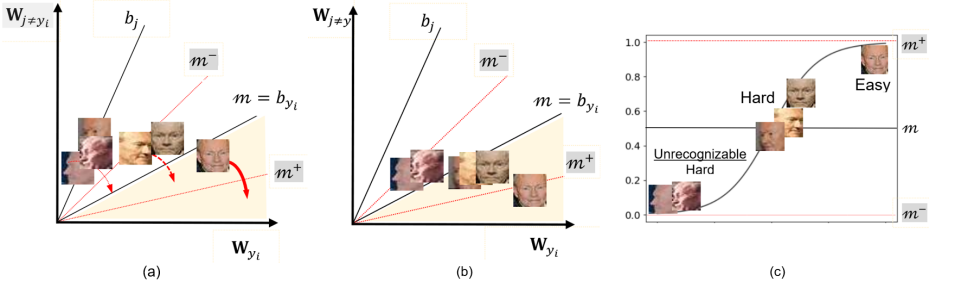

Figure 2: A conceptual diagram, where (a) is a geometrical interpretation of embedding space defined by SlackedFace with a predetermined static margin m, a positive margin m_+, and a negative margin m_-, given that $m^- < m < m^+$. The goal is to push (b) *unrecognizable* hard examples to m_-, *recognizable* hard examples close to m, and easy examples to m^+, based on a sigmoidal function in (c).

An alternative to the traditional static margin-based softmax loss is a non-static margin that attributes greater importance to easily recognizable examples based on the embedding norm as a proxy of face recognizability. Two leading works in this domain are, namely MagFace [21] and AdaFace [14]. Since these two margin-based softmax classifiers render a scale-invariant Cosine space, the embedding norm may not directly reflect face recognizability. Moreover, the embedding norm lacks correlation with the model prediction (see Figure 1(a)), which makes the margin elicited from the embedding norm susceptible to two drawbacks: (1) when an *unrecognizable* hard example is encoded with a large embedding norm (inducing a *large* margin), it may collapse towards its identity prototype; and (2) when a *recognizable* hard example is associated with a relatively poor embedding norm (resulting in a *small* margin), the model inclines to de-emphasize it, leading to limited learning from such examples. These scenarios are particularly detrimental to the ultimate generalization performance. We therefore hypothesize that the embedding norm alone may not capture all the essential information to distinguish between recognizable and unrecognizable faces, especially in the context of LR face images. To address this limitation, the margin definition should be aligned with the model's confidence in classifying a face image correctly for improved generalization (see Figure 1(b)).

This paper presents SlackedFace, a new non-static margin-based softmax loss with the following properties. We extend the concept of face recognizability by incorporating both

It may be distributed unchanged freely in print or electronic forms.

embedding norms and embedding proximity, known as the powered-embedding norm (P-Norm). We leverage P-Norm to enhance margin definition, resulting in the first relaxed margin strongly aligned with the model's confidence. We show that SlackedFace renders a dual-role face model that empowers embedding learning with face recognizability and serves face recognizability prediction for any unseen LR face images. We further introduce an early network calibration stage, dubbed fast hill-climbing (fast-HC), for compatibility learning between pre-trained modules (with prior knowledge) and other random initializations (without prior knowledge) to improve convergence.

We illustrate the high-level conceptual diagram in Figure 2. Our learning objective is to emphasize the importance of *recognizable* hard and easy examples by inducing a relatively large margin, close to or larger than the pre-determined static margin, using a sigmoidal function for empowered embedding learning. On the contrary, other *unrecognizable* hard examples are de-emphasized with a diminished margin to prevent learning noisy patterns from these examples.

2 Related Work

We revisit static and non-static margin-based softmax losses, followed by several representative face image quality assessment (FIQA) models for face recognition.

2.1 Static Margin-Based Softmax Loss

Different from the typical softmax loss, margin-based softmax losses [8, 17, 27, 28] introduce a margin penalty term to learn decision boundaries that maximize intra-class compactness and inter-class separation of face embeddings. This leads to enhanced model generalizability to tackle face recognition problems that involve a relatively large identity set in the underlying problem space.

Let $\mathbf{z}_i$ denote the face embedding for an arbitrary example $(\mathbf{z}_i, y_i)$, where y_i is the ground-truth identity label. The generic margin-based softmax function is defined in terms of cross-entropy loss as follows:

$$\mathcal{L}_i = -\log \frac{\exp(\mathcal{T}(\theta_{y_i}, s, m))}{\exp(\mathcal{T}(\theta_{y_i}, s, m)) + \sum_{j \neq y_i} s\cos\theta_j} \tag{1}$$

where $\mathcal{T}(.)$ is a margin function with three entities: the angle θ_{y_i} between the L2-normalized face embedding $\hat{\mathbf{z}}_i = \mathbf{z}_i / \|\mathbf{z}_i\|$ and its corresponding L2-normalized identity prototype $\hat{\mathbf{w}}_{y_i} = \mathbf{w}_{y_i} / \left\|\mathbf{w}_{y_i}\right\|$, a scaling term s, and a static margin term m. As a whole, the most representative instances are CosFace [28] and ArcFace [8] with the following margin functions:

$$\mathcal{T}(\theta_j, s, m)_{\text{CosFace}} = s(\cos\theta_{y_j} - m)\ ,\ \ \mathcal{T}(\theta_j, s, m)_{\text{ArcFace}} = s\cos(\theta_{y_i} + m) \tag{2}$$

2.2 Non-Static Margin-Based Softmax Loss

MagFace [21] and AdaFace [14] showed that the face models learned concerning a static margin are susceptible to overfitting when trained on noisy data, including LR face images. Hence, the notion of a non-static margin was proposed to relax the margin constraint based on face recognizability encoded by the embedding norm. However, MagFace and AdaFace are susceptible to the reliability of the embedding norm, a model-dependent quantity that various

factors can influence. Another two non-static margin-based losses are ElasticFace [2] and AdaptiveFace [16]. However, these variants are not intended to confront the challenges of LR face recognition.

2.3 Face Image Quality Assessment

There are two generic face image quality assessment (FIQA) models: (1) learning a typical regression problem concerning pre-annotated FIQA indexes, e.g., human ratings or pair-wise distance scores [1, 11, 22, 29], and (2) leveraging the intrinsic properties of embedding space, which have been shown to be correlated with face image quality, e.g., embedding variance, embedding norm, and embedding proximity [3, 5, 14, 21, 25, 26]. However, harnessing either of these model-dependent quantities as a sole indicator may not be reliable, especially when dealing with LR face images.

For clarity, we adopt "face recognizability" as the default term throughout this paper, as opposed to "face quality" which primarily refers to the visual appearance of face images. The important reason is that our focus is on face recognizability that pertains specifically to the ability to identify individuals based on discriminative face embeddings.

3 The SlackedFace Model

We disclose our proposed SlackedFace model in this section, which covers the mathematical formulation and principles of our slacked margin and how it differs from other relevant loss functions designated for face recognition.

3.1 Preliminaries

The key components of SlackedFace, depicted in Figure 3, comprise (1) a CNN backbone with a projection network and an embedding multilayer perceptron (MLP) for targeted embedding learning, denoted by $f(.)$; (2) a regression MLP attached to the projection network for learning face recognizability, collectively represented by $g(.)$; and (3) a softmax classifier learned with a set of c identity prototypes $\{\mathbf{w}_j\}_{j=1}^{c}$, where $\mathbf{w}_j \in \mathbb{R}^d$. Provided with a face example $\mathbf{x}_i$ with its identity label y_i, we encode $(\mathbf{x}_i)$ into a face embedding $\mathbf{z}_i = f(\mathbf{x}_i) \in \mathbb{R}^d$, alongside a learned face recognizability index $\sigma_i' = g(\mathbf{x}_i)$. Our goal is to induce a relaxed margin based on σ_i' for empowered embedding learning.

3.2 Powered Embedding Norm (P-Norm)

We reinterpret the notion of face recognizability based on two important properties: the embedding norm as the basis, and the embedding proximity learned in the latent space. For any given arbitrary face embedding $\mathbf{z}_i$, we compute the embedding proximity ρ_i as follows:

$$\rho_i = \Phi(\cos\theta_{y_i}(\cos\theta_{y_i} - \max_{j_{\max}\in\{1,\ldots,C\}\neq y_i} \cos\theta_{j_{\max}})) \tag{3}$$

where θ_{y_i} is the positive angle between $\mathbf{z}_i$ and its corresponding ground-truth identity (ID) prototype $\mathbf{w}_{y_i}$; $\theta_{j_{\max}}$ is the hard-negative angle between $\mathbf{z}_i$ and its closest negative ID prototype $\mathbf{w}_{j_{\max}}$. On the other hand, $\Phi(.;\Lambda)$ is a Sigmoid function with a steep slope parameter Λ. This is to enforce the constraint $\rho_i \in (0,1)$ for our slacked margin definition in (8) such that

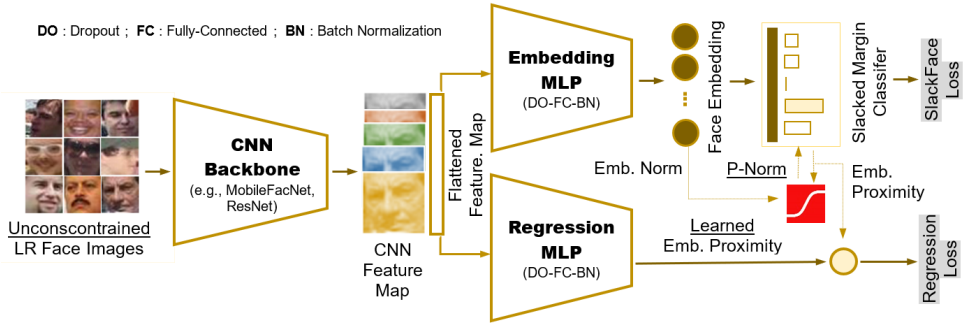

Figure 3: Generic network construction for SlackedFace includes a CNN backbone with (1) an embedding MLP for face representation learning based on a relaxed margin softmax classifier, and (2) a regression MLP for learning face recognizability for unseen examples.

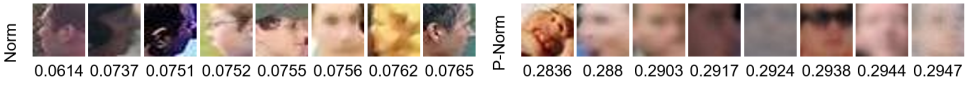

Figure 4: A collection of 8 hardest examples in TinyFace, indexed by Norm (left) and P-Norm (right), which demonstrate that P-Norm encodes meaningful indexes that reflect face recognizability.

$\rho_i \approx 1.0$ is induced for easy examples, $\rho_i \approx 0.5$ for recognizable hard examples, and $\rho_i \approx 0.0$ for unrecognizable hard examples (refer to Figure 2 for a schematic diagram).

As the embedding norm $\|\mathbf{z}_i\|$ is positive and unbounded such that $\|\mathbf{z}_i\| \in \mathbb{R}^+$, we constrain its range to the interval $(0,1)$ by imposing an upper bound τ for simple handling as follows:

$$\sigma_i = \max(0.0,\ \min(\ \|\mathbf{z}_i\| / \tau,\ 1.0\ )) \tag{4}$$

We empirically set $\tau = 1e^2$ to the default, which is generally applicable to a range of CNN backbones. Accordingly, we compute face recognizability $\mathcal{R}_i$ in terms of powered-embedding norm (P-Norm) by raising σ_i to the power of ρ_i as follows:

$$\mathcal{R}_i = \sigma_i^{(1.0\ -\ \rho_i)} \in (0,1) \tag{5}$$

Intuitions. P-Norm redefines embedding norm as follows: (1) P-Norm indexes remain small for *unrecognizable* hard examples characterized by low embedding norms; (2) P-Norm computes a relatively large index for the *recognizable* hard example with a low σ_i but a high ρ_i; and (3) P-Norm assigns easy examples, i.e., high for both σ_i and ρ_i, with large indexes, unless ρ_i and σ_i are not consistent with one another, e.g., mislabeled examples and under-sampled subject.

Properties of P-Norm. By our definition in (5), P-Norm outperforms Norm as a new proxy of face recognizability. We illustrate the hardest LR face images in the TinyFace dataset [7] ranked by embedding norm and P-Norm in Figure 4. It is shown that P-Norm singles out the unrecognizable examples with the lowest indexes and vice versa. Another useful property of the P-Norm is that it is strongly correlated to the model's predictions, resulting from inclusion (3) in our P-Norm definition, whereas the embedding norm has no such correspondence. Our theoretical analyses in our supplementary materials demonstrate that these properties are crucial for inducing a meaningful slacked margin that reflects face recognizability.

3.3 Slacked Margin Function

To increase the interpretability of $\mathcal{R}_i$ for our slacked margin definition (as both σ_i and ρ_i are model-dependent), we normalize it into a unit Gaussian, referred to as $\hat{\mathcal{R}}_i$ in the succeeding sections, using batch statistics as follows:

$$\hat{\mathcal{R}}_i = \begin{cases} 1 & \text{if } \alpha(\mathcal{R}_i) \geq 1 \\ \alpha(\mathcal{R}_i) & \text{if } -1 < \alpha(\mathcal{R}_i) < 1 \\ -1 & \text{otherwise} \end{cases} \;,\; \alpha(\mathcal{R}_i) = \frac{\mathcal{R}_i - \mu_{\mathcal{R}}}{\Sigma_{\mathcal{R}}} \tag{6}$$

where $\mu_{\mathcal{R}}$ and $\Sigma_{\mathcal{R}}$ indicate the exponential moving average (EMA) batch mean and standard deviation, respectively. To this end, $\mu_{\mathcal{R}}$ and $\Sigma_{\mathcal{R}}$ are computed over κ iterations (within each mini batch) as follows:

$$\mu_{\mathcal{R}} = \beta\ \mu_{\mathcal{R}}^{(\kappa)} + (1-\beta)\ \mu_{\mathcal{R}}^{(\kappa-1)} \;,\; \Sigma_{\mathcal{R}} = \beta\ \Sigma_{\mathcal{R}}^{(\kappa)} + (1-\beta)\ \Sigma_{\mathcal{R}}^{(\kappa-1)} \tag{7}$$

where β is a scalar momentum set to 0.99 in our experiments for all $\kappa > 0$. Note that we let $\beta = 1.0$ at the initial state, i.e., when $\kappa = 0$.

Given the normalized P-Norm indexes $\hat{\mathcal{R}}_i$ defined for each face embedding face embedding $\mathbf{z}_i$, the slacked margin function for ArcFace (2) is defined as follows:

$$\mathcal{T}(\theta_j, s, m)_{\text{SlackedFace}} = s\cos(\theta_{y_i} + \delta(m)) \;,\; \delta(m) = m + \eta\,\hat{\mathcal{R}}_i \tag{8}$$

Provided that $\hat{\mathcal{R}}_i \in [-1, 1]$ based on (6), we define each positive and negative slacked margin by $m_i^+ = m + \eta\,|\hat{\mathcal{R}}_i|$ and $m_i^- = m - \eta\,|\hat{\mathcal{R}}_i|$, respectively, where η indicates the degree of margin relaxation to be determined in our experiments. This induces $m_i^+ > m$ to only recognizable easy and hard examples to emphasize them more during embedding learning. On the contrary, the unrecognizable hard examples are de-emphasized by inducing $m_i^- < m$ to discourage learning from noisy patterns that may result in performance degradation. The rule of thumb is to set η to be sufficiently significant such that $\delta(m)$ does not estimate an overly large $m^+ \approx 2m$ that hinders convergence, or an insignificantly small $m^- \approx 0$ that collapses to the typical non-margin softmax configuration. We examine $\eta = \{0.025, 0.05, 0.10, 0.15, 0.20\}$ in our analyses provided in the supplementary materials. Aside from that of ArcFace, it is trivial to impose a slacked margin to the CosFace loss by substituting the static margin m in (2) with $\delta(m)$.

3.4 SlackedFace Losses

Softmax Loss. Accordingly, the SlackedFace loss is optimized in conjunction with $\mathcal{R}_i$ as a regularization term as follows:

$$\mathcal{L}_{\text{SlackedFace}} = \frac{1}{N}\sum_{i=1}^{N} -\log \frac{\exp(s\cos(\theta_{y_i} + \delta(m)))}{\exp(s\cos(\theta_{y_i} + \delta(m))) + \sum_{j \neq y_i} s\cos\theta_j} + \lambda\,\mathcal{L}_{\mathcal{R}} \tag{9}$$

where $\mathcal{L}_{\text{SlackedFace}} = \mathcal{L}_{\delta(m)} + \lambda\,\mathcal{L}_{\mathcal{R}}$, $\mathcal{L}_{\mathcal{R}} = \sum_i -\log(\mathcal{R}_i)$ is a regularization term, and λ is an empirical weighting factor. By optimizing $\mathcal{R}_i$ as regularization, our objective is to enhance the robustness of SlackedFace from two perspectives: (1) rectifying misclassified examples resulting from the overly relaxed margin by pushing them within their respective class (identity) boundary, and (2) standardizing embedding norm to ensure our learning stage is more robust to variations of the norm in both normalized and pre-normalized spaces [24]. These are supported by our theoretical analysis in the supplementary section.

Regression Loss. With the estimated embedding proximity ρ_i in (3), we define the learned embedding proximity $\hat{\rho}_i = \Phi(g(\mathbf{x}_i))$, where $g(\cdot)$ is the regression MLP responsible for learning the embedding proximity for unseen examples. We resolve this regression problem using Huber loss [13] as follows:

$$\mathcal{L}_{\text{Huber}} = \frac{1}{N}\sum_{i=1}^{N}\begin{cases}\frac{1}{2}(\rho_i - \hat{\rho}_i)^2, & \text{for } |\rho_i - \hat{\rho}_i| < \gamma \\ \gamma(|\rho_i - \hat{\rho}_i| - \frac{1}{2}\gamma), & \text{otherwise}\end{cases} \tag{10}$$

In principle, γ serves as a hyperparameter that governs the transitions from a quadratic function to a linear function. This is to address outliers and noises which are often prevalent in the LR face datasets, as in our particular case.

Overall Training Loss. In accordance with (9) and (10), we learn a SlackedFace model that empowers embedding learning based on the estimated face recognizability as follows:

$$\mathcal{L}_{Total} = \mathcal{L}_{\text{SlackedFace}} + \mathcal{L}_{\text{Huber}} \tag{11}$$

Compared to other static margin-based losses, the proposed SlackedFace requires tuning two additional hyperparameters, namely η in (8), and λ in (9). Other hyperparameters configured to the pre-determined default setting are $\Lambda = 6.0$ in (3), $\tau = 1e^2$ in (4) and $\gamma = 0.5$ in (10). We provide a comprehensive hyperparameter analysis in our supplementary materials.

3.5 Fast Hill-Climbing Optimization

Learning on a pre-trained HR face model alongside randomly initialized modules for a new downstream task often leads to a poor local optimum. To facilitate knowledge transfer, the pre-learned backbone is typically frozen for network calibration with other random modules. However, we observe that setting the backbone to the warm-up state (i.e., with a learning rate of $1e^{-06}$) converges to a better minimum.

Inspired by the linear probing proposed in [15] in conjunction with our empirical observations, we present a fast hill-climbing (fast-HC) search for network calibration prior to end-to-end learning. More specifically, during the calibration stage, we re-learn only pre-trained batch normalization parameters for several initial training epochs with the random regression MLP and softmax classifier. This is to adapt the prior knowledge marginally, while incrementally improving both regression MLP and softmax classifier from random initialization. We disclose in our experiments that a fast-HC calibration stage not only improves generalization performance by a remarkable margin but also offers a relatively stable calibration stage, compared to the one involving no calibration, and the two alternatives calibrated with a frozen and a warm-up backbone.

4 Experiments and Discussions

This section summarizes our experimental results, including ablation analyses, performance comparisons, and discussions. We compile in our supplementary materials [1] with additional details on benchmarking datasets, hyperparameter settings, and relevant analyses, along with our implementation materials for reproducibility.

[1] https://github.com/chengyawlow/SlackedFace

4.1 Benchmarking Datasets

We evaluate open-set face identification tasks with disjoint training and testing identities in the presence of unknown face images in the gallery set, referred to as distractors. This introduces additional challenges to the problem search space, simulating a more realistic deployment scenario. In a nutshell, our experiments involve three real-world LR face datasets, i.e., surveillance camera faces (SCFace) [10], TinyFace [2], and DroneFace [12].

SCFace, TinyFace, DroneFace. SCFace includes a gallery set with only a single high-resolution (HR) mugshot and three low-resolution (LR) probe sets, i.e., dubbed D1, D2, and D3, for an HR-LR identification task. On the contrary, **TinyFace** is a large-scale LR face repository with an LR gallery, and an LR probe set for an LR-LR identification task alongside 153,428 distractors. DroneFace, on the other hand, is only a test set for the HR-LR identification task. We, therefore, evaluate the generalization performance on DroneFace using the SCFace-learned models without additional fine-tuning.

Ad-Hoc Distractor Set. Since SCFace and DroneFace contain no distractors, we extend these datasets with an ad-hoc distractor set of 20,000 examples sampled from that of TinyFace in our experiments. We provide the random distractor set as supplementary materials.

4.2 Implementation

Experimental Setup. We employ MobileFaceNet [6] and ResNet50[20] as the backbone embedding encoders, which are pre-trained on a million-scale HR face dataset, known as VGGFace2 [4]. To instantiate a SlackedFace model, each pre-learned CNN backbone (with an embedding MLP) is stacked with a randomly initialized softmax classifier alongside a regression MLP. We perform a fast hill-climbing (fast-HC) search to re-learn only the batch normalization parameters in the CNN backbone for 8 initial training epochs, while incrementally improving both the random regression MLP and softmax classifier. This is followed by end-to-end fine-tuning for 32 epochs using the Adam optimizer.

Performance Metric. We evaluate the generalization performance in terms of rank-1 (R1) identification rate (%) for analyses and comparisons. To investigate the effectiveness of P-Norm as a metric of face recognizability, we provide the error rejection curves (ERC) [3, 11, 23] based on false non-matched rates (FNMR) and false matched rates (FMR), alongside the area under the ERC (AUERC) metric.

4.3 Performance Analysis

The complete configuration of SlackedFace is de-generalized into several baselines for our ablation analyses. We summarize our observations as follows:

Effect of Slacked Margin. We explore whether a slacked margin improves generalization performance in baseline I, II, and III, regardless of face recognizability metrics. Our findings disclose that a slacked margin, induced by Norm and P-Norm, elevates the average performance by de-emphasizing unrecognizable hard examples.

Effect of Norm vs. P-Norm. Based on our comparison between baseline II and III, we disclose that our proposed P-Norm is a better metric to encode face recognizability for slacked margin definition, particularly for hard examples. With P-Norm, the R1 identification rate is increased by 2% with MobileFaceNet and 1.25% with ResNet50 in the most challenging D1 probe set in terms of distortion severity level.

Baselines #	Backbone	Reg.Metric	Network Config.			Extended SCFace			
			fast-HC	$\mathcal{L}_{\delta(m)}$	$\mathcal{L}_{\mathcal{R}}$	D1	D2	D3	Mean
I	MobileFaceNet	-				50.00	89.75	89.00	76.25
II	MobileFaceNet	Norm		/		49.75	91.00	90.00	76.92
III	MobileFaceNet	P-Norm		/		51.75	91.50	89.75	77.67
IV	MobileFaceNet	P-Norm	/	/		59.75	**93.00**	90.25	81.00
SlackedFace	MobileFaceNet	P-Norm	/	/	/	**60.75**	92.50	**90.75**	**81.33**
I	ResNet50	-				71.25	93.75	97.25	87.42
II	ResNet50	Norm		/		72.00	94.75	97.50	88.08
III	ResNet50	P-Norm		/		73.25	94.50	98.25	88.67
IV	ResNet50	P-Norm	/	/		79.50	**95.25**	**99.00**	91.25
SlackedFace	ResNet50	P-Norm	/	/	/	**80.00**	95.00	**99.00**	**91.33**
SlackedFace*	ResNet50	P-Norm	/	/	/	77.75	94.75	98.25	90.25
SlackedFace**	ResNet50	P-Norm	/	/	/	79.25	95.00	98.25	90.83

Table 1: Performance analysis on extended SCFace for SlackedFace and other baseline models built on pre-trained MobileFaceNet and ResNet50 in terms of rank-1 identification rate (%). Notably, the SlackedFace variants marked with "*" and "**" are respectively trained with a frozen and a warm-up CNN backbone during network calibration in place of the proposed fast-HC.

Effect of Fast Hill-Climbing. In comparison to both baseline III and IV, we reveal that our proposed fast-HC improves the average R1 identification rate by at least 3%. This demonstrates that an early network calibration by the fast-HC enables convergence to a better local minimum when learning on pre-trained models along with randomly initialized modules.

Complete Configuration. We disclose that the SlackedFace model with its complete configuration, which includes (1) P-Norm as an objective metric of face recognizability, (2) the regularized slacked margin function, and (3) the fast-HC calibration, jointly empowers embedding learning by means of inducing a meaningful slacked margin to tackle the open-set LR face identification tasks. Moreover, we also demonstrate that the proposed fast-HC calibration enhances other common alternatives at reduced computation cost, particularly the warm-up strategy that requires fine-tuning on the entire backbone.

4.4 Performance Comparison and Discussions

For a fair comparison, we train the most relevant state-of-the-art (SoTA) models based on the respective non-static margin softmax functions derived from ArcFace [8], using the same pre-trained ResNet50 with the proposed fast-HC calibration stage.

LR Face Identification Tasks. Based on Table 2, the SlackedFace-trained models consistently perform well on all LR face datasets, demonstrating robustness to large-scale distractors. Our observations are as follows: (1) AdaptiveFace [16] induces the class margin based on class distribution rather than face recognizability, leading to poorer performance in our target tasks. (2) With a random Gaussian margin, ElasticFace [2] achieves an overall performance close to that of MagFace [21] and AdaFace [14] with a norm margin. However, the standard deviation parameter must be set to an insignificantly small value. (3) SlackedFace prevails over the recently proposed MagFace and AdaFace with a norm margin that may overly emphasize or de-emphasize the hard examples. Notably, SlackedFace is the first to combine embedding norm with embedding proximity (model's prediction) to encode face recognizability. Furthermore, SlackedFace achieves a new baseline performance on SCFace and TinyFace, suggesting that SlackedFace can serve as a new standard for future comparison. This potentially triggers further advances in the field, e.g., security surveillance and

Face Models	Ori. / Ext. SCFace				TinyFace	Ori. / Ext. DroneFace
	D1	D2	D3	Mean		
AdaptiveFace [CVPR'19] [14]	93.95 / 76.75	99.00 / 94.25	99.75 / 97.25	97.33 / 89.42	75.80 / 71.93	- / -
ElasticFace [CVPRW'22] [3]	95.25 / 78.50	**99.25 / 96.50**	99.50 / 97.50	98.00 / 90.83	76.56 / 72.92	- / -
MagFace [CVPR'21] [21]	94.75 / 77.75	99.00 / 94.50	99.50 / 98.00	97.75 / 90.08	76.15 / 72.13	95.45 / 65.15
AdaFace [CVPR'22] [12]	95.75 / 78.75	99.00 / 96.00	99.50 / 97.75	98.08 / 90.83	77.71 / 73.20	95.60 / 66.79
SlackedFace	**97.00 / 80.00**	**99.25** / 95.00	**100.00 / 99.00**	**98.75 / 91.33**	**78.43 / 73.87**	**96.41 / 68.48**

Table 2: Performance comparison with recent models by rank-1 identification rate (%).

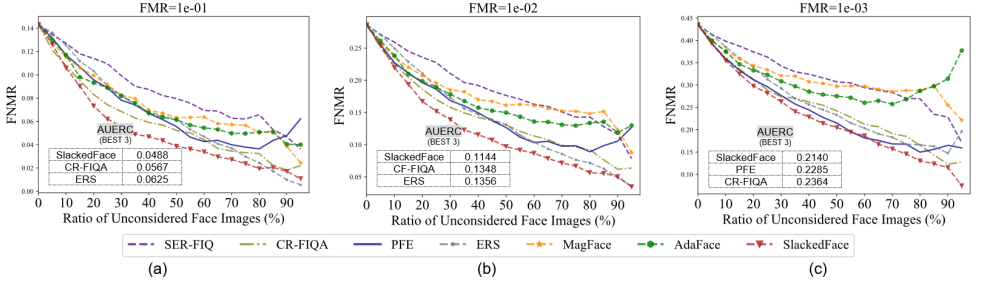

Figure 5: ERC in terms of FNMR at varying FMR rates discloses that SlackedFace encodes more reliable face recognizability indexes (P-Norm), compared to other existing metrics for face image quality assessments.

monitoring, forensic investigations, and mobile authentication.

Predicting Face Recognizability on Unseen Examples. To assess the reliability of P-Norm as a face recognizability index for unseen instances, we neglect the hard examples (by ratio) from the TinyFace test set with respect to the learned P-Norm to estimate FNMR and FMR from random positive and negative pairs. We observe from the ERC in Figure 5 that P-Norm reports the lowest FNMR at three pre-determined FMRs, indicating its superior performance in encoding face recognizability. This contributes to another observation that the proposed SlackedFace, with the learned P-Norm, achieves the lowest AUERC compared to other SoTAs, including SER-FIQ [26], CR-FIQA [3], PFE [25], ERS [3], MagFace, and AdaFace.

5 Conclusion

We proposed a new variant of the non-static margin-based softmax loss to tackle open-set low-resolution (LR) face identification tasks by empowered embedding learning. Our model stands out from other novel instances in two perspectives, marking several contributions. We are the first to introduce the notion of powered-embedding norm, which incorporates embedding norm and embedding proximity to reliably capture face recognizability. We are also the first to induce a slacked margin that is strongly correlated with modeling prediction. Our experiments demonstrated that the proposed model outperforms the existing best models in open-set LR face identification tasks and face recognizability prediction against unseen examples. We believe our findings are to trigger more advanced face applications that benefit humanity. In the future, we intend to integrate this model with supervised metric learning.

Acknowledgments. Meeyoung Cha is the corresponding author. This work was supported by the Institute for Basic Science (IBS-R029-C2).

References

[1] Lacey Best-Rowden and Anil K Jain. Learning face image quality from human assessments. *IEEE Transactions on Information Forensics and Security*, 13(12):3064–3077, 2018.

[2] Fadi Boutros, Naser Damer, Florian Kirchbuchner, and Arjan Kuijper. Elasticface: Elastic margin loss for deep face recognition. In *Proc. CVPR*, pages 1578–1587, 2022.

[3] Fadi Boutros, Meiling Fang, Marcel Klemt, Biying Fu, and Naser Damer. Cr-fiqa: Face image quality assessment by learning sample relative classifiability. *Proc. CVPR*, 2023.

[4] Qiong Cao, Li Shen, Weidi Xie, Omkar M Parkhi, and Andrew Zisserman. Vggface2: A dataset for recognising faces across pose and age. In *IEEE international conference on Automatic Face and Gesture Recognition*, pages 67–74, 2018.

[5] Jacky Chen Long Chai, Ng Tiong-Sik, Cheng-Yaw Low, and Andrew Beng Jin Teoh. Recognizability embedding enhancement for very low-resolution face recognition and quality estimation. In *Proc. CVPR*, pages 9957–9967, 2023.

[6] Sheng Chen, Yang Liu, Xiang Gao, and Zhen Han. Mobilefacenets: Efficient cnns for accurate real-time face verification on mobile devices. In *Proc. Chinese Conference on Biometric Recognition*, pages 428–438, 2018.

[7] Zhiyi Cheng, Xiatian Zhu, and Shaogang Gong. Low-resolution face recognition. In *Proc. ACCV*, pages 605–621, 2018.

[8] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proc. CVPR*, pages 4690–4699, 2019.

[9] Siqi Deng, Yuanjun Xiong, Meng Wang, Wei Xia, and Stefano Soatto. Harnessing unrecognizable faces for improving face recognition. In *Proc. CVPR*, pages 3424–3433, 2023.

[10] Mislav Grgic, Kresimir Delac, and Sonja Grgic. Scface–surveillance cameras face database. *Multimedia tools and applications*, 51(3):863–879, 2011.

[11] Javier Hernandez-Ortega, Javier Galbally, Julian Fierrez, Rudolf Haraksim, and Laurent Beslay. Faceqnet: Quality assessment for face recognition based on deep learning. In *International Conference on Biometrics (ICB)*, pages 1–8, 2019.

[12] Hwai-Jung Hsu and Kuan-Ta Chen. Droneface: an open dataset for drone research. In *Proc. ACMMSys*, pages 187–192, 2017.

[13] Peter J Huber. Robust estimation of a location parameter. *Breakthroughs in statistics: Methodology and distribution*, pages 492–518, 1992.

[14] Minchul Kim, Anil K Jain, and Xiaoming Liu. Adaface: Quality adaptive margin for face recognition. In *Proc. CVPR*, pages 18750–18759, 2022.

[15] Ananya Kumar, Aditi Raghunathan, Robbie Jones, Tengyu Ma, and Percy Liang. Fine-tuning can distort pretrained features and underperform out-of-distribution. In *Proc. ICLR*, 2022.

[16] Hao Liu, Xiangyu Zhu, Zhen Lei, and Stan Z Li. Adaptiveface: Adaptive margin and sampling for face recognition. In *Proc. CVPR*, pages 11947–11956, 2019.

[17] Weiyang Liu, Yandong Wen, Zhiding Yu, Ming Li, Bhiksha Raj, and Le Song. Sphereface: Deep hypersphere embedding for face recognition. In *Proc. CVPR*, pages 212–220, 2017.

[18] Cheng-Yaw Low and Andrew Beng-Jin Teoh. An implicit identity-extended data augmentation for low-resolution face representation learning. *IEEE Transactions on Information Forensics and Security*, 17:3062–3076, 2022.

[19] Cheng-Yaw Low and Andrew Beng-Jin Teoh. An implicit identity-extended data augmentation for low-resolution face representation learning. *IEEE Transactions on Information Forensics and Security*, 17:3062–3076, 2022.

[20] Ze Lu, Xudong Jiang, and Alex Kot. Deep coupled resnet for low-resolution face recognition. *IEEE Signal Processing Letters*, 25(4):526–530, 2018.

[21] Qiang Meng, Shichao Zhao, Zhida Huang, and Feng Zhou. Magface: A universal representation for face recognition and quality assessment. In *Proc. CVPR*, pages 14225–14234, 2021.

[22] Fu-Zhao Ou, Xingyu Chen, Ruixin Zhang, Yuge Huang, Shaoxin Li, Jilin Li, Yong Li, Liujuan Cao, and Yuan-Gen Wang. Sdd-fiqa: Unsupervised face image quality assessment with similarity distribution distance. In *Proc. CVPR*, pages 7670–7679, 2021.

[23] Fu-Zhao Ou, Xingyu Chen, Ruixin Zhang, Yuge Huang, Shaoxin Li, Jilin Li, Yong Li, Liujuan Cao, and Yuan-Gen Wang. Sdd-fiqa: unsupervised face image quality assessment with similarity distribution distance. In *Proc. CVPR*, pages 7670–7679, 2021.

[24] Jaewoo Park, Cheng Yaw Low, and Andrew Beng Jin Teoh. Divergent angular representation for open set image recognition. *IEEE Transactions on Image Processing*, 31: 176–189, 2021.

[25] Yichun Shi and Anil K Jain. Probabilistic face embeddings. In *Proc. CVPR*, pages 6902–6911, 2019.

[26] Philipp Terhorst, Jan Niklas Kolf, Naser Damer, Florian Kirchbuchner, and Arjan Kuijper. Ser-fiq: Unsupervised estimation of face image quality based on stochastic embedding robustness. In *Proc. CVPR*, pages 5651–5660, 2020.

[27] Feng Wang, Xiang Xiang, Jian Cheng, and Alan Loddon Yuille. Normface: L2 hypersphere embedding for face verification. In *Proc. ACM ICMM*, pages 1041–1049, 2017.

[28] 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 *Proc. CVPR*, pages 5265–5274, 2018.

[29] Weidi Xie, Jeffrey Byrne, and Andrew Zisserman. Inducing predictive uncertainty estimation for face verification. In *Proc. BMVC*, 2020.

Supplementary Materials

We offer additional resources in this section to enhance the understanding and reproducibility of this work. To summarize, our supplementary materials are presented as follows:

A. Theoretical Analysis

B. Benchmarking Datasets

C. Hyperparameter Analysis and Configuration

D. Stability Analysis

E. Model Extension

A. Theoretical Analysis

We provide the mathematical proofs for the conceptual principles that underlie SlackedFace, including the role of (1) slacked margin, and (2) regularization term. To simplify our analysis, we omit normalization procedures and subscript indexes.

A1. Slacked Margin

Assumption. A face example, either easy or hard, is ranked recognizable, if and only if two conditions are satisfied: (1) its embedding magnitude $\|\mathbf{z}\| > \tau_{\text{norm}}$, and the Cosine similarity between $\mathbf{z}$ and its true identity prototype $\cos\theta_y > \tau_{\text{cosine}}$.

Therefore, a high recognizability example has a large embedding magnitude and a large similarity score, while an unrecognizable example is deficient at both. We first show that the SlackedFace margin is proportional to these recognizability factors.

Proposition 1. *(Slacked margin) The slacked margin $m = \sigma^{1.0-\rho}$ is monotonically strictly increasing with respect to $\|\mathbf{z}\|$ and $\cos\theta_y$, if $\|\mathbf{z}\| < \tau$ and $\theta_y + m < \pi/2$.*

SlackedFace, hence, can accurately induce a margin that corresponds to recognizability, resulting in a model that produces a significantly large gradient update on recognizable examples and a small gradient update on unrecognizable examples during training.

Corollary. *(Gradient of SlackedFace Loss) The magnitude of SlackedFace loss gradient $\frac{\partial \mathcal{L}}{\partial \theta_y}$ on the target similarity angle θ_y is monotonically and strictly increasing with respect to $\|\mathbf{z}\|$ and $\cos\theta_y$, if $\|\mathbf{z}\| < \tau$ and $\theta_y + m < \pi/2$.*

Proof. Let $\mathcal{L} = -\log\exp(\cos(\theta_y+m))/[\sum_{k\neq y}\exp(\theta_k) + \exp(\theta_y+m)]$. Then,

$$\frac{\partial \mathcal{L}}{\partial \theta_y} = [1 - (\sum_{k\neq y}\exp(\cos\theta_k - \cos(\theta_y+m)))^{-1}]\sin(\theta_y+m). \tag{12}$$

Given $\theta_y + m < \pi/2$, $\sin(\theta_y+m)$ and $\sum_{k\neq y}\exp(\cos\theta_k - \cos(\theta_y+m))$ are strictly increasing with m. Similarly, the derivative of $\frac{\partial \mathcal{L}}{\partial \theta_y}$ with respect to m is also strictly increasing. Therefore, by the previous proposition, the gradient is also strictly increasing with embedding magnitude and cosine similarity. □

The above corollary demonstrates that *the SlackedFace margin is highly correlated with the recognizability of a face.* This means that SlackedFace induces a smaller margin for unrecognizable examples, and otherwise. As a result, SlackedFace can effectively promote empowered embedding learning based on face recognizability.

A2. Regularization Term

In accordance with (9), the SlackedFace loss incorporates a regularization term $\mathcal{L}_{\mathcal{R}}$ to maximize the embedding norm to an upper bound. We demonstrate that this improves learning stability and convergence by enforcing the model to focus on minimizing the relative angle between a face embedding and its true identity prototype.

Since SlackedFace is updated by a gradient-based optimizer, the gradient update for the embedding vector is performed by $\mathbf{z} \leftarrow \mathbf{z} - \alpha \frac{\partial \mathcal{L}_{class}}{\partial \mathbf{z}}$ at a certain learning rate. Therefore, learning of embedding vector is determined by the gradient $\frac{\partial \mathcal{L}_{class}}{\partial \mathbf{z}}$ in (13). We analyze this gradient as follows:

Proposition 2. *For the classification loss $\mathcal{L}_{class}$,*

$$\frac{\partial \mathcal{L}_{class}}{\partial \mathbf{z}} = \frac{1}{\|\mathbf{z}\|} \cdot \sum_k \frac{\partial \mathcal{L}_{class}}{\partial \cos\theta_k} (\widehat{\mathbf{w}}_k - \cos\theta_k \widehat{\mathbf{z}}) \tag{13}$$

where $\widehat{\mathbf{w}}_k$ and $\widehat{\mathbf{z}}$ are obtained by normalizing $\mathbf{w}_k$ and $\mathbf{z}$, respectively.

The proposition indicates that the gradient $\frac{\partial L_{class}}{\partial \mathbf{z}}$ for an embedding update is *disentangled* to two terms, specifically, $\|\mathbf{z}\|^{-1}$ and $\sum_k \frac{\partial L_{class}}{\partial \cos\theta_k} (\widehat{w}_k - \cos\theta_k \widehat{\mathbf{z}})$, where the former depends on the embedding norm and the latter does not. Since the regularization term reduces the reciprocal of embedding magnitude, the regularizer makes the gradient less variant to the magnitude:

Corollary. *Minimizing $\mathcal{L}_r$ reduces the magnitude of the embedding gradient, thereby making the embedding gradient invariant to embedding magnitude.*

Overall, the regularization term serves two important roles: (1) preventing the overly large gradient update, which helps to *stabilize the training stage*. (2) *marking the embedding learning more depend on the relative angle between face embeddings and true identity prototypes*. Increasing the angle-dependency of the embedding learning improves the generalization of the corresponding cosine similarity metric for open-set applications [24].

B. Benchmarking Datasets

We further elaborate on our benchmarking datasets, including SCFace, TinyFace, and DroneFace, for performance evaluation under the open-set deployment scenario.

SCFace. Real-world face recognition systems enroll individuals using high-resolution (HR) mugshots, leaving unseen (test) face images unrestricted. Hence, SCFace includes a gallery set with a high-resolution (HR) mugshot per identity (ID), and three low-resolution (LR) probe sets, namely D1, D2, and D3, to simulate a real-world HR (gallery)-LR (probe) identification task. As a whole, these probe sets are compiled with examples captured at standoff distances of 4.20m, 2.60m, and 1.00m, respectively. In compliance with [19], we allocate the first 50 subjects (from ID 01 to 050) for training, while the other 80 subjects (from ID 051 to 130) are reserved for testing.

TinyFace. Contrary to SCFace, TinyFace is a large-scale LR face dataset with both an LR gallery and an LR probe set for an LR-LR identification task. Overall, it is a composition of 7,804 / 8,171 face images annotated with 2,570 / 2,569 ID labels in each training and testing set, respectively. On average, the pixel resolution of these examples is limited to only 20×16 pixels. It is worth noting that its gallery search space is interfered with 153,428 distractors of unknown identities to simulate a more challenging real-world scenario.

DroneFace. DroneFace, on the other hand, is only a test set for an HR-to-LR identification task (similar to SCFace). As a whole, it consists of 11 subjects with 1,364 examples detected from drone footage (at 1.5m to 5m high, and 2m to 17m away from the subjects) in the probe set and 2 frontal mugshots per ID as the enrolled templates in the gallery set. We evaluate the generalization performance on DroneFace using the SCFace-learned models.

Ad-Hoc Distractor Set. As both SCFace and DroneFace contain no distractors, we extend these datasets with an ad-hoc distractor set of 20,000 unknown examples randomly sampled from that of TinyFace.

We summarize the data distribution for each dataset in Table 3. On the other hand, we portray 10 hardest and easiest examples indexed by Norm and P-Norm in Figure 6.

Datasets	Desc.	Train. Set	Test. Set			Eval. Protocol
			Gallery	Probe	Distract.	
SCFace	# IDs	50	80	80	-	HR-to-LR
	# Imgs.	800	80	1,200	20,000*	
TinyFace	# IDs	2,570	2,569	2,569	-	LR-to-LR
	# Imgs.	7,804	4,443	3,728	153,428	
DroneFace	# IDs	-	11	11	-	HR-to-LR
	# Imgs.	-	22	1,364	20,000*	

Table 3: Data distribution for our benchmarking datasets, including TinyFace, SCFace, and DroneFace. Note that "*" refers to a random distractor set of 20,000 LR face images sampled from that of TinyFace.

C. Hyperparameter Analysis and Configuration

Compared to other static margin-based softmax losses that involve two primary hyperparameters, i.e., the scaling factor s and the margin term m, training a SlackedFace model requires tuning two additional hyperparameters, specifically the degree of margin relaxation η in (8), and the regularization weighting factor λ in (9). Other default hyperparameters are the Sigmoid steep slope $\Lambda = 6.0$ in (3), the generic upper bound for the embedding norm $\tau = 10^2$ in (4), and the regression transition parameter $\gamma = 0.5$ in (10). Accordingly, we analyze in Table 4 the effect of η and λ, in addition to m. These are the key parameters to determine the generalization power of the SlackedFace models, achieving optimal performance with and without distractors in the most difficult SCFace D1 probe set. We summarize our hyperparameter configuration in Table 5.

D. Stability Analysis

To assess the model's stability against random initializations, we train the SlackedFace models and the comparing instances using 5 random seeds, i.e., 0, 1, 42, 1234, and 2023. Our experimental results in Table 6 reveal that the SlackedFace models exhibit the ideal robustness to multiple random initializations, underlining that SlackedFace is a reliable alternative to other non-static margin-based softmax losses.

Hyperparams.	Setting	Ori. / Ext. SCFace			
		D1	D2	D3	Mean
Effects of η ($m = 0.5$; $\lambda = 0.1$)	0.025	89.25 / 59.25	**98.50** / 92.50	**98.75** / 90.25	95.50 / 80.67
	0.05	89.50 / 60.25	98.25 / **93.00**	98.25 / 90.25	95.33 / 81.17
	0.10	**90.00** / 60.75	98.25 / 92.50	98.25 / **90.75**	**95.50** / **81.33**
	0.15	89.25 / **61.00**	98.00 / 92.25	98.25 / 89.50	95.17 / 80.92
	0.20	88.50 / 58.00	98.00 / 92.00	98.25 / 90.25	94.92 / 80.08
Effects of λ ($m = 0.5$; $\eta = 0.1$)	0.0	88.75 / 59.75	**98.50** / **93.00**	97.75 / 90.25	95.00 / 81.00
	0.05	89.75 / 61.00	98.25 / 92.25	98.50 / 90.00	**95.55** / 81.08
	0.10	**90.00** / 60.75	98.25 / 92.50	98.25 / **90.75**	95.50 / 81.33
	0.20	89.25 / **62.25**	98.25 / **93.00**	**98.75** / 89.75	95.42 / **81.67**
	0.50	88.25 / 61.50	**98.50** / 92.25	98.25 / 90.00	95.00 / 81.25
Effects of m ($\eta = 0.1$; $\lambda = 0.1$)	0.40	**90.00** / 59.75	98.25 / 90.75	**98.50** / 89.25	95.58 / 79.92
	0.45	**90.00** / 60.00	**98.50** / 92.00	**98.50** / 89.75	**95.67** / 80.58
	0.50	**90.00** / 60.75	98.25 / 92.50	98.25 / **90.75**	95.50 / **81.33**
	0.55	89.00 / **60.75**	98.25 / 92.50	**98.50** / 90.50	95.25 / 81.25
	0.60	88.75 / 61.00	**98.50** / **93.25**	**98.50** / 89.75	95.25 / **81.33**

Table 4: Hyperparameter analyses for SlackedFace (using pre-trained MobileFaceNet as the embedding encoder).

	Hyperparameters	SCFace		TinyFace
		MobileFaceNet	ResNet50	ResNet50
Basic	Mini Batch Size	32	32	32
	# Epochs (Fast-HC + End-to-End)	8 + 32	8 + 32	8 + 32
	Learning Rate	$1e^{-03}$	$1e^{-03}$	$1e^{-03}$
	Learning Rate Decay	0.1 / 8 epochs	0.1 / 4 epochs	0.1 / 6 epochs
	Weight Decay	$1e^{-04}$	$1e^{-04}$	$1e^{-04}$
	Dropout Rate	0.6	0.8	0.8
SlackedFace (Default)	Sigmoid Steep Slope Λ in (3)	6.0		
	Upper Bound for Norm τ in (4)	100		
	Regress. Transition Parameter. γ in (10)	0.5		
SlackedFace (Fine-Tuned)	Scaling s, Margin m in (8)	60, 0.50		
	Slacked Margin Degree η in (8)	0.10		
	Reg. Weighting Factor λ in (9)	0.10		

Table 5: Overall hyperparameter configuration in our experiments. We set the learning rate for the pre-trained backbone (inclusive of the embedding MLP) to 0.1x of the softmax classifier to prevent the prior knowledge from being distorted with noises in LR face examples.

Face Models	Ori. / Ext. SCFace			
	D1	D2	D3	Mean
ElasticFace	95.75±0.35 / 78.45±0.97	99.50±0.25 / 94.95±0.62	99.55±0.27 / 97.90±0.45	98.27±0.11 / 90.43±0.41
MagFace	95.80±0.62 / 77.65±1.24	99.50±0.18 / 94.65±0.58	99.60±0.22 / 98.10±0.38	98.30±0.24 / 90.13±0.59
AdaFace	95.90±0.55 / 77.95±0.65	**99.55±0.11** / 95.75±0.40	99.95±0.11 / 97.95±0.27	98.47±0.14 / 90.55±0.10
SlackedFace	**96.50±0.50** / **79.30±0.48**	99.50±0.18 / 96.00±0.59	**100.0±0.00** / **98.65±0.29**	**98.67±0.12** / **91.32±0.15**
SlackedCosFace	96.20±0.21 / 78.45±1.05	99.40±0.22 / **96.10±0.74**	**100.0±0.00** / 98.55±0.37	98.53±0.13 / 91.03±0.55

Table 6: Performance comparison for SlackedFace and SoTAs (using pre-trained ResNet50 as the embedding encoder) over 5 random initializations, in terms of averaged rank-1 identification rate (%) and standard deviation. Note that SlackedCosFace is an extended variant to be disclosed in Section E of this supplementary material.

E. Extension of SlackedFace

Aside from ArcFace (reported in our manuscript), we extend SlackedFace based on CosFace, termed SlackedCosFace in this section, for further exploration. Likewise, we substitute the static margin m in $\mathcal{T}(\theta_j, s, m)_{\text{CosFace}}$ (2) with a slacked margin term $\delta(m)$ as follows:

$$\mathcal{T}(\theta_j, s, m)_{\text{SlackedCosFace}} = s(\cos\theta_{y_i} - \delta(m)) \;\;,\;\; \delta(m) = m + \eta\,\hat{\sigma}_i \tag{14}$$

We also disclose in Table 6 that SlackedCosFace performs on a par with the ArcFace-learned

Figure 6: A collection of 10 hardest and easiest (unseen) test examples indexed by Norm and proposed P-Norm for each benchmarking dataset.

counterpart. More importantly, we demonstrate that the overall performance of the proposed SlackedFace models (both learned with ArcFace and CosFace) surpass other SoTAs by a significant margin in resolving open-set LR face identification tasks, especially in the most challenging D1 probe set with distractors.