

RepFace: Refining Closed-Set Noise with Progressive Label Correction for Face Recognition

Jie Zhang^{1, 2, 3}, Xun Gong^{1, 2, 3*}, Zhonglin Sun^{1, 4}

¹ School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu 611756, China

² Engineering Research Center of Sustainable Urban Intelligent Transportation, Ministry of Education, Chengdu 611756, China

³ Manufacturing Industry Chains Collaboration and Information Support Technology Key Laboratory of Sichuan Province, Southwest Jiaotong University, Chengdu 611756, China

⁴Queen Mary university of London, UK

zhangjie-18@my.swjtu.edu.cn, xgong@swjtu.edu.cn, zhonglin.sun@qmul.ac.uk

Abstract

Face recognition has made remarkable strides, driven by the expanding scale of datasets, advancements in various backbone and discriminative losses. However, face recognition performance is heavily affected by the label noise, especially closed-set noise. While numerous studies have focused on handling label noise, addressing closed-set noise still poses challenges. This paper identifies this challenge as training isn't robust to noise at the early-stage training, and necessitating an appropriate learning strategy for samples with low confidence, which are often misclassified as closed-set noise in later training phases. To address these issues, we propose a new framework to stabilize the training at early stages and split the samples into clean, ambiguous and noisy groups which are devised with separate training strategies. Initially, we employ generated auxiliary closed-set noisy samples to enable the model to identify noisy data at the early stages of training. Subsequently, we introduce how samples are split into clean, ambiguous and noisy groups by their similarity to the positive and nearest negative centers. Then we perform label fusion for ambiguous samples by incorporating accumulated model predictions. Finally, we apply label smoothing within the closed set, adjusting the label to a point between the nearest negative class and the initially assigned label. Extensive experiments validate the effectiveness of our method on mainstream face datasets, achieving state-of-the-art results.

Code — <https://gitee.com/swjtugx/classmate/tree/master/OurGroup/RepFace>

Extended version — <http://arxiv.org/abs/2412.12031>

Introduction

Face recognition is a prominent research area in pattern recognition, drawing significant attention and advancements in recent years. The application of FR consists of security and surveillance, and mobile unlocking *et al.* The remarkable progress in face recognition technology can be largely attributed to three key factors: the emergence of vast, large-scale face datasets (Yi et al. 2014; Guo et al. 2016;

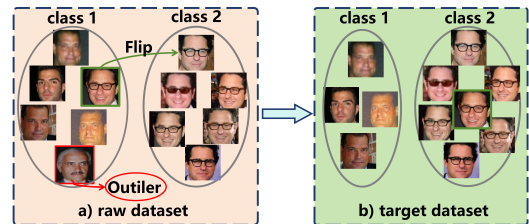


Figure 1: Noise category: a) Label noise in the face recognition training set, including both closed-set (Flip) and open-set (Outlier) noise. b) Target dataset obtained through noise filtering and label correction.

Kemelmacher-Shlizerman et al. 2016; Cao et al. 2018; Zhu et al. 2022; Liu et al. 2015), the pivotal role played by backbones (He et al. 2016; Howard et al. 2019; Tan and Le 2019; Dosovitskiy et al. 2020; Sun and Tzimiropoulos 2022; Sun et al. 2024), and the advancement in loss functions (Liu et al. 2017; Deng et al. 2019; Wang et al. 2018b; Kim, Jain, and Liu 2022; Wang et al. 2020; Huang et al. 2020; Wen et al. 2021; Saadabadi et al. 2023; Deng et al. 2021; Meng et al. 2021; An et al. 2021).

With the increasing size of face recognition datasets, the amount of label noise is also rising (Wang et al. 2018a), dramatically preventing the FR from producing more effective performance. Typically, these datasets are compiled by retrieving labeled images from web search engines and cleaned using automated procedures, e.g. MS-Celeb-1M (Guo et al. 2016). The MS1M (Guo et al. 2016) dataset is available in several versions: e.g. MS1Mv2, and MS1Mv3. MS1Mv2 refines the dataset by removing samples that are distant from the feature center using the model proposed by (Deng, Zhou, and Zafeiriou 2017). MS1Mv3 further enhances the dataset by employing RetinaFace (Deng et al. 2020b) for face detection and a face recognition (FR) model to eliminate noise and duplicate samples. Another way to refrain from the impact of noise is manually labelling, however such label cleaning process requires significant human effort, which is time-consuming and resource-intensive. Therefore, it is vital to explore effective training methods for the automatic and efficient handling of label noise in large-

*Corresponding author

scale datasets.

Label noise falls into two main categories, as detailed in Fig 1 (a): closed-set noise and open-set noise. Closed-set noise refers to the misassignment of a person’s identity to another existing identity within the dataset. In contrast, open-set noise occurs when the true/actual label of an image is absent from the dataset but is erroneously treated as an existing identity of the dataset. Studies for tackling label noise have employed two primary strategies: **noise cleaning** and **label correction**.

Recent research on **noise cleaning** predominantly adopts the seminal approaches of identifying noisy samples by adopting noise filtering techniques, subsequently these identified noisy samples are discarded or re-weighted to minimize their impact during training (Hu et al. 2019; Zhong et al. 2021; Deng et al. 2020a; Wang et al. 2022, 2019). The above-mentioned solutions are reasonable for open-set noise scenarios, however, the majority proportion of label noise distributed in the dataset is close-set noise (Wei et al. 2021; Wang et al. 2018a), which brings greater impact than open-set noise, thereby reducing the effectiveness of these methods when applied to real-world applications.

To address the problem of closed-set noise, **label correction** are adopted to assign the closed-set noise sample with their estimated label, allowing these samples to contribute to the facial recognition training without being discarded. BoundaryFace (Wu and Gong 2022) pioneering focuses on detecting and correcting closed-set noise from the perspective of the decision boundary. However, this method requires FR model to obtain a certain discriminative ability in the early-stage training, but may result in inaccurate noise assessment during later-stage of training if early-stage training is not converged enough. Furthermore, this approach is dedicated to the identification and correction of clean and noisy samples, which is only effective in addressing significantly noisy samples that can be clearly identified as noise. It does not accommodate ambiguous ones that lack the requisite confidence level for definitive classification as noise.

To address the abovementioned problems, we first propose a module Auxiliary Sample Cleaning (ASC) aimed at reducing the influence of noise in early-stage training. This enables the model to learn effective performance, thereby providing a discriminative ability to identify noise in the subsequent stages. Subsequently, this paper partitions the dataset into three categories: clean samples, ambiguous samples, and closed-set noise samples according to the cosine distance between the nearest negative and positive class centers. Each category is subjected to different supervision strategies. Specifically, based on the split categories, for ambiguous samples, we introduce Label Robust Fusion (LRF) method merges the prediction results of the model with the ground-truth label, aiming to maximize the use of these samples. Additionally, Smoothing Label Correction (SLC) is proposed to rectify closed-set noise samples. The main contributions of our work are as follows:

- We propose the ASC module to effectively identify and remove noisy samples during the early-stage of training, allowing the model to concentrate on effectively and ef-

ficiently learning generalized representation from clean samples. Consequently, this approach enhances model generalization and improves the model’s performance in subsequent noise detection tasks.

- To maximize the utilization of training samples, we introduce two techniques: LRF uses a memory bank to accumulate model’s prediction and fuse it with ground-truth labels, enabling more effective utilization of ambiguous samples. SLC applies an exponential smoothing strategy among the ground-truth label and corrected label to minimize the impact of incorrect corrections
- We conduct extensive experimentation on datasets with varying noise ratios, which achieves state-of-the-art results in well-established benchmarks, validating the effectiveness of the proposed method in handling closed-set label noise.

Related Work

Face Recognition with Label Noise

Researches on label noise processing methods primarily focus on cleaning label noise and correcting closed-set noisy labels. For label noise cleaning, Noise-Tolerant (Hu et al. 2019) find that the θ -distribution of training samples implicitly reflects their clean probability. Based on this finding, they propose a new paradigm for dynamically adjusting the training weights of samples. Co-Mining (Wang et al. 2019) trains two networks simultaneously, and categorizes samples into three parts: noisy, high-confidence clean, and clean samples. Based on the difference in results between the two networks, different loss functions and training strategies are used for each category of samples. SubCenter-ArcFace (Deng et al. 2020a) trains multiple sub-centers, while discarding non-dominant sub-centers and high-confidence noisy samples to enhance internal compactness and reduce the influence of noise. The noise is dynamically determined by DTDD (Zhong et al. 2021) through the calculation of the cumulative average difference between the maximum cosine similarity and the label-specified cosine similarity, which is then discarded during training. RV-Face (Wang et al. 2022) uses the OTSU algorithm to compute the noise threshold based on the θ distribution. During the training process, samples are categorized into simple, semi-hard, and ambiguous samples, with higher weights assigned to semi-hard samples.

For the correction of close-set noise labels, BoundaryFace (Wu and Gong 2022) analyzes the relationship between the nearest negative and ground-truth classes from a decision boundary perspective to identify and correct closed-set noise. Part of our proposed method is further optimized based on BoundaryFace by using the smoothing technique while scaling the criteria for judging noise. The goal of our optimization is to enhance the model’s capability to detect label noise and minimize the impact of misjudged noise.

Methodology

Firstly, the first sub-section describes the Auxiliary Sample Cleaning (ASC) module for filtering closed-set noise during the early-stage of training. Then, we introduce how we

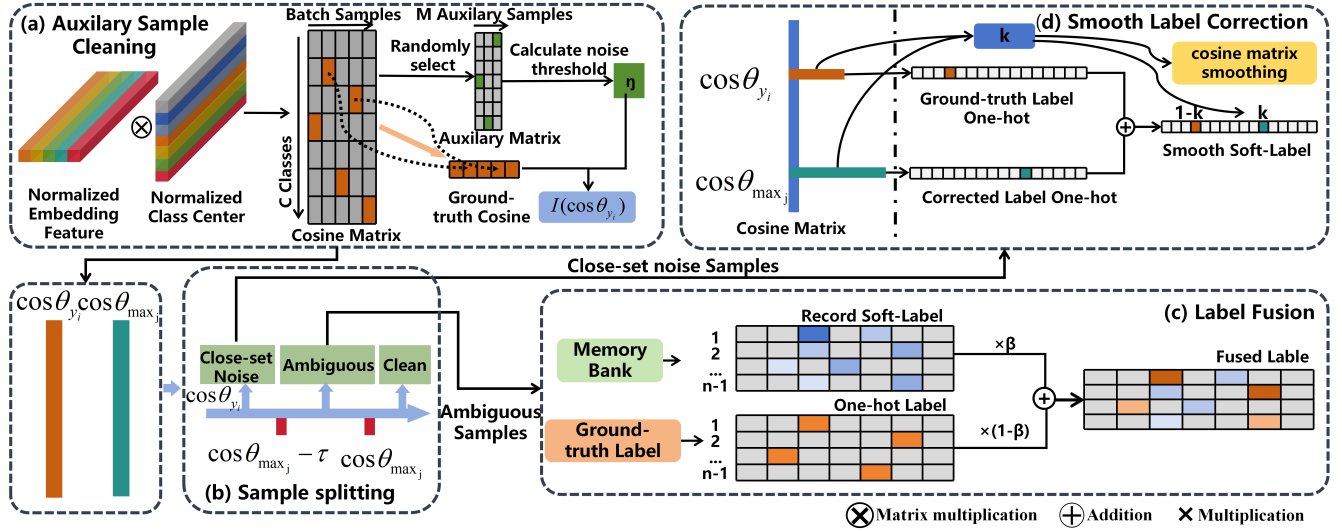


Figure 2: The overall framework: (a) Auxiliary Sample Cleaning: Randomly labeled samples are compared with class centers using cosine similarity and evaluated by a dynamic threshold. (b) Sample Splitting: Samples are categorized as clean, ambiguous, or noisy based on similarity to positive and nearest negative centers. (c) Label Fusion: A memory bank stabilizes training on ambiguous samples with accumulated predictions. (d) Smooth Label Correction: Labels are smoothed between positive and nearest negative classes, adjusting cosine similarity logits.

divide samples into three categories (clean, ambiguous and close-set noise) in the second sub-section, and thus propose separate learning strategies for each category. Specifically, For ambiguous samples, we use a Label Fusion method namely Label Robust Fusion(LRF) to integrate stored predicted cosine logits with the labels as illustrated in the second sub-section. Finally, in the third sub-section for closed-set noise samples, we apply the smooth technique to correct the labels. The pseudocode of this paper is available at the *SUPPLEMENTARY MATERIAL* Algorithm 1.

Auxiliary Sample Cleaning

Recent face recognition methods for label correction, such as BoundaryFace (Wu and Gong 2022), identify and correct label noise by analyzing the cosine similarity between sample labels and their nearest negative class, after the model has attained a certain level of recognition capability. However, during the early stages of training, when the model is still developing its learning capabilities, it remains susceptible to label noise. This early-stage noise can subsequently impair the model’s ability to accurately identify label noise in the later stage of training. To address this issue, we propose an Auxiliary Sample Cleaning (ASC) strategy as illustrated in the left part of Fig 2. This approach aims to filter out noisy samples and focus on clean samples during the early-stage training, facilitating the learning of effective facial embeddings that can be utilized for noise detection in later stage. Firstly, random M samples are selected in each mini-batch and assigned with random labels y_r , denoted as Auxiliary Samples in this paper. Subsequently, to identify the injected Auxiliary Samples during training, we create a matrix to record the cosine similarity between the Auxiliary Samples and their randomly assigned label, which are em-

ployed to filter out the noisy samples. Specifically, the cosine similarities are calculated between each facial embedding produced by the under-training FR backbone and the weight of the linear layer representing the estimated class center.

The average cosine similarity between samples and assigned label center in the matrix is used as a threshold, which is formulated as follows:

$$\eta = \frac{1}{M} \sum_{i=1}^M \cos \theta_{y_i} + \alpha \quad (1)$$

where M denotes the number of auxiliary samples generated in the mini-batch, $\cos \theta_{y_i}$ denotes the cosine similarity between the samples and their assigned labels, and α is a parameter that stabilizes the threshold for random noise. Finally, the sample indicator can be obtained by comparing the cosine similarity between the sample and ground-truth label with the threshold calculated by the ASC:

$$I(\cos \theta_{y_i}) = \begin{cases} 0, & \cos \theta_{y_i} < \eta \\ 1, & \cos \theta_{y_i} \geq \eta \end{cases} \quad (2)$$

where $\cos \theta_{y_i}$ denotes the cosine similarity between features and learned class centers corresponding to the ground-truth (actual) labels, $I(\cos \theta_{y_i}) = 1$ indicates the sample is clean without label perturbation, $I(\cos \theta_{y_i}) = 0$ suggest the sample is regarded as noisy sample. We multiply $I(\cos \theta_{y_i}) = 0$ by the final loss function to filter out samples identified as noisy. It is worth noting that our auxiliary sample strategy remains effective throughout the entire training process.

Sample Splitting

In BoundaryFace (Wu and Gong 2022), samples are only categorized as either clean or noisy. However, this method

may treat noisy samples with insufficient confidence (a subset of ambiguous samples) as clean, leading to instability in the later stages of training. To address this issue, we categorize the samples into three groups: clean samples (including easy and hard sample), ambiguous samples, and closed-set noise samples, by examining the cosine similarity between the positive class and the nearest negative class, as shown in the left bottom (b) part of Fig 2.

$$d_i = \cos \theta_{max_j} - \cos \theta_{y_i} \quad (3)$$

where $\cos \theta_{max_j}$ denotes the cosine similarity between the sample's feature to the estimated nearest negative class center. Specifically, samples are regarded as clean if d_i is less than 0; And samples are treated as ambiguous when $0 < d_i < \tau$, where τ is a learned threshold for determination (discussed in Ablation Study). Finally, samples with a distance larger than τ are classified as noise samples.

Ambiguous Sample Label Fusion

The ambiguous samples are distributed close to the decision boundary, thus the actual label of these samples cannot be explicitly applied for training. To appropriately tackle ambiguous samples, inspired by SELC (Lu and He 2022), which update the original noisy labels through ensemble predictions, we propose the Label Robust Fusion (LRF) module, illustrated in the right bottom(c) part of Fig 2. It effectively fuses the model's cosine prediction logits and labels of ambiguous samples, aiming at adopting the accumulated prediction at training for stabilizing the learning of ambiguous samples.

Specifically, we use a memory bank to store the model's prediction results during training, by recording the maximum cosine logits from the model's prediction results, representing the distance between the sample features and the learned class center. If a cosine logits for a class has already been recorded, it is updated according to the formula 4. Otherwise, the new logits $\cos(\theta_{max_j})$ is stored directly.

$$\cos \theta_p = (1 - \beta) * \cos(\theta_p) + \beta * \cos(\theta_{max_j}) \quad (4)$$

where $\cos \theta_p$ represents the value of the recorded cosine logits, $\cos \theta_{max_j}$ represents the value of the maximum cosine similarity between the sample features and the learned weights at the current prediction, and β is a balancing parameter.

During the training process, we continuously update the model's max cosine prediction logits in the memory bank. It is worth noting that a sample may have multiple different logits, we normalize these logits in order to generate soft labels. Subsequently, ground-truth one-hot encoding is adopted for fusion with the generated soft labels, as depicted in Fig 2 (Label Fusion(c)) and the fusion formula is as follows.

$$\mathbf{q}_r = \beta \mathbf{p} + (1 - \beta) \mathbf{q}_{y_i} \quad (5)$$

where \mathbf{p} and \mathbf{q}_{y_i} represent soft labels generated by the stored cosine logits and the one-hot ground-truth label, respectively, and β is the same as Eq. 4.

Ambiguous samples could be either clean or noisy. Therefore, if the samples are actually clean, the preserved cosine

logits' related category often aligns with the ground-truth label. In such cases, the fused label aligns with the class existing in the dataset. Conversely, if they are closed-set noise samples, the stored cosine logits' related category often represents the real label. By fusing labels together, we enhance the robustness of our model.

Smoothing Label Correction

In our method, we utilize the disparity between the cosine similarity of the nearest negative class center and positive class center for each sample (using Eq. 3, denote as d_i) as a metric for assessing noise.

Although BoundaryFace (Wu and Gong 2022) has demonstrated a commendable ability to detect and correct label noise, an analysis of its label detection formula reveals certain limitations: *if* $\max \{\cos(\theta_k + m) \text{ for all } k \neq y_i\} - \cos \theta_{y_i} > 0 : y_i = k$, where the m is the margin in the loss function. According to the formula, the smaller θ_{y_i} is, the more stringent the criteria for determining noise will be. This indicates that the method imposes stricter criteria on samples that cannot be accurately predicted, resulting in label correction only for samples with high confidence. Consequently, this can leave some label noise uncorrected during training. To address this issue and maximize label noise correction while maintaining model stability, we propose the Smoothing Label Correction (SLC) module, shown in the top right part(d) of Fig 2. A fixed, small threshold τ is used as the criterion for identifying closed-set noise. While lowering this threshold can relax the noise detection criteria, it also increases the risk of misjudgment. To address this, we propose a smoothing method that relaxes the criteria while reducing the impact of misjudgment errors. For hard sample mininig, MV-Softmax (Wang et al. 2020) is used as our baseline, which is formulated as follows.

$$N_1(t, \cos \theta_j) = \begin{cases} \cos \theta_j & T(\cos \theta_{y_i}) \geq \cos \theta_j \\ \cos \theta_j + t & T(\cos \theta_{y_i}) < \cos \theta_j \end{cases} \quad (6)$$

where $T(\cos \theta_{y_i}) = \cos(\theta_{y_i} + m)$ follow MV-softmax to use the m in ArcFace (Deng et al. 2019).

To balance the ratio between the two parts of the label and the nearest negative class, we compute a dynamic balancing parameter $k = \text{Sigmoid}(10 \cdot d_i)$, where the Sigmoid function is used to keep the ratio between 0 and 1, and multiplying the distance by 10 is used to expand the influence factor of the cosine-valued distance. Due to the label correction, we need to perform a smoothing weight adjustment for noising sample mining. The cosine matrix smoothing process is:

$$N_2(t, \cos \theta_j) = (1 - k)N_1(t, \cos \theta_1) + kN_1(t, \cos \theta_2) \quad (7)$$

where $N_1(t, \cos \theta_1)$ indicates the hard samples that are computed based on the given labels, and $N_1(t, \cos \theta_2)$ indicates the hard samples computed based on the corrected label which is refined by Eq. 8. After applying a smoothing operation to the noise sample mining, and equally importantly, we employ a similar approach to smooth the labels by transforming hard labels into soft labels. The formula for label correction by using the smoothing technique is as follows:

$$\mathbf{q}_s = (1 - k)\mathbf{q}_{y_i} + k\mathbf{q}_{y_j} \quad (8)$$

where \mathbf{q}_{y_i} and \mathbf{q}_{y_j} represent the one-hot encoding of the true label and the nearest negative label for closed-set noise, respectively.

By dividing the samples into clean, closed-set noise and ambiguous samples, for which we use different strategies to enhance the resilience of the network model against closed-set noise, the overall framework is referred to as RepFace, which can be summarised as follows:

$$\mathcal{L}_1 = \begin{cases} -I(\cos \theta_{y_i}) \cdot \mathbf{q}_s \cdot \log P_{y_j} & d_i > \tau \\ -I(\cos \theta_{y_i}) \cdot \mathbf{q}_r \cdot \log P_{y_i} & \tau \geq d_i \geq 0 \\ -I(\cos \theta_{y_i}) \cdot \mathbf{q}_{y_i} \cdot \log P_{y_i} & \text{others} \end{cases} \quad (9)$$

$$P_{y_j} = \frac{e^{s \cdot T(\cos \theta_{y_i})}}{e^{s \cdot T(\cos \theta_{y_i})} + \sum_{j=1, j \neq i}^C e^{s \cdot N_2(t, \cos \theta_j)}}$$

$$P_{y_i} = \frac{e^{s \cdot T(\cos \theta_{y_i})}}{e^{s \cdot T(\cos \theta_{y_i})} + \sum_{j=1, j \neq i}^C e^{s \cdot N_1(t, \cos \theta_j)}}$$

Experiments

Experiment Settings

Datasets. Our training set is the widely-used MS1MV2 (Guo et al. 2016; Deng et al. 2019) and CASIA-WebFace (Yi et al. 2014), which contains approximately 5.8M images of 85K individuals and approximately 0.5M images of 10K individuals, respectively. To simulate label noise within the dataset, we manually synthesize closed-set noise for constructing a noisy dataset. Specifically, we randomly select a certain percentage of samples from each person in the dataset and assign them with random labels. According to (Wang et al. 2018a), CASIA-WebFace contains 9.3%-13% noise, therefore, we artificially synthesize noisy datasets with 10% and 20% proportions to reproduce similar noise volume to CASIA-WebFace. Finally, we perform experiments using synthetic datasets and clean datasets. Our test datasets are several popular benchmarks in face recognition including LFW (Huang et al. 2008), SLLFW (Deng et al. 2017), CALFW (Zheng, Deng, and Hu 2017), CPLFW (Zheng and Deng 2018), AgeDB (Moschoglou et al. 2017), CFP-FP (Sengupta et al. 2016), and RFW (Wang et al. 2018c), IJB-B (Whitelam et al. 2017), and IJB-C (Maze et al. 2018). The metrics for evaluating each dataset are, TAR@FAR=1e-5 and 1e-4 for evaluation on IJB-B, and IJB-C; 1:1 verification accuracy are adopted for LFW, SLLFW, CALFW, AgeDB, CFP-FP, CPLFW and RFW.

Training details. We follow the ArcFace (Deng et al. 2019) to crop and resize faces in a size of 112×112 . IR-ResNet50 (Duta et al. 2020) is the backbone with a feature output dimension of 512. The models are optimized by the typical stochastic gradient descent (SGD) algorithm with a momentum of 0.9 and weight decay of $5e-4$. The learning rate starts at 0.1 and is divided by 10 at the 9th, 18th, and 26-th epochs, with a total of 30 training epochs on CASIA-WebFace, while the learning rate is divided by 10 at the 10th, 18th, and 22-nd epochs, with totally 24 training epochs on

MS1MV2 (Deng et al. 2019). We set the scale hyperparameters to $s = 64$, margin is $m = 0.5$.

Ablation Studies

A number of studies are carried out to demonstrate the effectiveness of each part of our proposed RepFace. The noise ratio employed in this study is 20% unless specified. The default backbone is IR-50 (Duta et al. 2020). Due to page limitation, we put the discussion of M , the effectiveness of our method on another advanced loss and the impact of open-set noise in the SUPPLEMENTARY MATERIAL.

Hyperparameter Analysis

Impact of τ In our comprehensive study, we initially analyze the critical threshold value τ , which is used to differentiate and assess closed-set noise by thoroughly comparing the threshold utilized by the Boundaryface algorithm. In our innovative approach to enhance closed-set noise detection, we meticulously refine and narrow judgment thresholds. Intuitively, we set τ to four distinct values: 0.16, 0.18, 0.20, and 0.22 as given in Fig 3. Experiments on datasets with 10% closed-set noise enabled rigorous evaluation. Among the four τ values tested, we can find that, if τ set too large, it will reduce the number of detected label noise instances, while a smaller value increases the risk of incorrect detections. Based on these findings, we decide to set $\tau = 0.2$ for all subsequent experiments.

Impact of α In the first sub-section of Methodology, we initially screen the samples by generating auxiliary noise samples, and use Equation 1 to calculate the threshold, where α to stabilize the threshold calculation. We use 0.03, 0.05 and 0.07 as alternatives to conduct experiments. From the experimental results Table 1, we can find that in each verification dataset, when α is 0.05, our method achieves the best effect. In addition, the performance of our method will degrade when α is greater than or less than 0.05, so we adopt the value of α to be 0.05 in subsequent experiments.

Impact of β Based on our analysis above, in the critical matter of selecting the balance parameter β , we have chosen to use values of 0.85, 0.9, and 0.95, for our experimentation. A meticulous examination of the results, as detailed in Table 2, reveals a clear and definitive trend: our RepFace demonstrates its superiority in performance when the value of β is

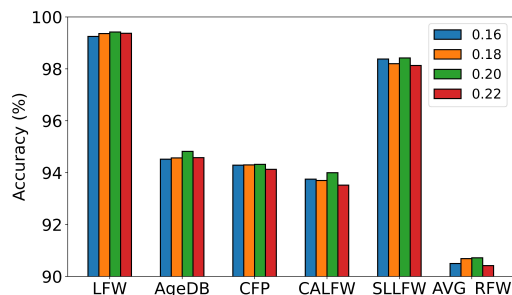


Figure 3: Test results on LFW, AgeDB, CFP-FP, CALFW, and SLLFW with different hyperparametric τ .

α	<i>LFW</i>	<i>AgeDB</i>	<i>CFP</i>	<i>CPLFW</i>	<i>RFW(Avg.)</i>
0.03	99.23	94.23	93.31	88.20	89.59
0.05	99.40	94.43	93.44	88.57	90.10
0.07	99.22	94.32	93.89	88.25	89.72

Table 1: Verification performance(%) with different α .

set to 0.9. Consequently, for the remainder of this study, the value of β has been firmly opted for at 0.9, providing a solid foundation for further exploration and analysis.

Methods	<i>LFW</i>	<i>AgeDB</i>	<i>CFP</i>
$\beta = 0.95$	99.33	94.32	93.30
$\beta = 0.9$	99.40	94.43	93.44
$\beta = 0.85$	99.33	94.37	93.40

Table 2: Verification performance(%) of different β .

Impact of Each Module

Effect of each module To rigorously ascertain the efficacy of each individually proposed module, a set of experiments are executed. This involves incorporating each module sequentially into the baseline MV-Softmax. Initially, MV-Softmax is trained solely on the noisy dataset. Subsequently, we augment this baseline with the ASC module, followed by the SLC and LRF modules, each added sequentially. Table 3 serves to elucidate the distinct contributions made by each module.

Methods	<i>LFW</i>	<i>AgeDB</i>	<i>CALFW</i>	<i>RFW(Avg.)</i>
MV-Softmax	99.12	92.58	92.45	88.00
ASC	99.13	92.87	92.93	88.45
ASC+SLC	99.10	93.83	93.12	88.78
ASC+SLC+LRF	99.40	94.43	93.50	90.10

Table 3: Validation performance(%) of progressively adding ASC,SLC and LRF to baseline MV-softmax.

It’s worth noting that the performance of our models does not show a substantial improvement with just the ASC and SLC modules included. However, there is a marked and significant enhancement in the performance of our trained models once all three modules are assimilated into the proposed framework, the RepFace.

Effect of sample and label smoothing in SLC Herein, we ablate the choice of smoothing option described from the fourth sub-section in Methodology. Specifically, for identified samples are provided three options:(a) directly assign the nearest negative class’s label to the sample; (b) correct the label with smoothing technology; (c) incorporate the cosine matrix smoothing approach while assigning the label as (a); And finally, combine the (b) and (c) as the complete SLC module provided by Section . The results are available at Tabel 4, revealing that (b) and (c) attribute the training performance under noisy data, and (d) indeed validate the effectiveness of the SLC.

Methods	<i>LFW</i>	<i>AgeDB</i>	<i>CALFW</i>	<i>RFW(Avg.)</i>
(a) directly assign	99.32	93.00	93.03	88.30
(b) label smoothing	99.25	93.60	93.05	92.89
(c) cosine matrix smoothing	99.33	93.90	93.05	93.11
(d) SLC	99.40	94.43	93.50	93.50

Table 4: Validation performance(%) of effective of smoothing technology in SLC.

Comparisons with SOTA Methods

In this section, we evaluate our innovatively designed RepFace framework on two synthetic, closed-set noise datasets and the raw CASIA-WebFace dataset. These datasets are capable of modelling the label noise present in realistic applications where the noise ratio varies according to different collection environments. Open-set noise is not considered as the prominent disturbance for effective FR training is closed-set noise. Table 5 illustrates the comparison between our methodology with several state-of-the-art (SOTA) alternatives, including ArcFace, MV-Softmax, CurricularFace, RVFace, AdaFace, and BoundaryFace. Our proposed RepFace achieves 96.26 on *avg*, 88.42 on Asian, 94.87 on Caucasian, 91.72 on Indian and 89.58 on African. In terms of injecting noise, experiment, our method improves the previous BoundaryFace by 0.19 on the *avg*, and the gap is also large when using the heavy noisy ratio 20%. Another interesting conclusion can be drawn by comparing the gap between the real and 20% noisy dataset, for instance, AdaFace decreases by 1.58, however RepFace drops by 0.5, demonstrating the robustness to noise of our method.

Observing suggests that our methodology demonstrates superior performance compared to the state-of-the-art (SOTA) method across both synthetic closed-set noise sets and the primitive CASIA-WebFace dataset, validating feasibility and effectiveness on synthetic noisy datasets.

Generalizability Evaluation

To validate the generalizability of our method, we conduct further experiments of training on the MS1Mv2 dataset. Specifically, our experiments are divided into two parts: training on the original MS1Mv2 dataset and training on a synthetic dataset with 20% closed-set noise. We make evaluation experiments on the mainstream verification datasets, IJB-B, and IJB-C datasets.

For the original dataset, we compare our method with mainstream techniques for handling hard samples and label noise. For the noisy dataset, we compare our method with leading label noise processing approaches. The results of model trained on the original dataset are illustrated in Table 6, we achieved results comparable to state-of-the-art (SOTA) methods on *LFW*, *AgeDB-30*, *CFP-FP* and *CALFW*, with our average performance being the best, surpassing the baseline MV-Softmax by 0.13. Similarly, our method demonstrates the highest performance across all metrics in the IJB-B and IJB-C datasets. These findings indicate that our method consistently outperforms other approaches when trained on the original dataset.

Although our results are comparable to state-of-the-art

Ratio	Method	LFW	AgeDB	CFP	CALFW	SLLFW	Avg.	RFW			
								Asian	Caucasian	Indian	African
0%	ArcFace	99.25	94.20	94.70	93.37	98.00	95.90	86.83	93.22	90.10	87.68
	MV-Softmax	99.37	93.80	94.76	93.07	98.17	95.83	87.23	93.67	89.95	87.88
	Curricular	99.28	94.48	94.37	93.25	98.22	95.92	87.03	93.57	90.67	88.53
	RVFace	99.42	94.40	95.01	93.40	98.25	96.10	87.75	93.75	91.03	88.27
	BoundaryFace	99.33	94.05	95.26	93.43	97.98	96.01	87.50	94.00	90.55	87.72
	AdaFace	99.50	94.63	94.91	93.58	98.32	96.19	87.20	94.45	90.52	87.83
	RepFace(ours)	99.47	94.77	94.71	93.82	98.53	96.26	88.42	94.87	91.72	89.58
10%	ArcFace	99.22	93.40	93.29	92.90	97.45	95.25	85.87	92.42	89.82	86.23
	MV-Softmax	99.32	93.97	93.81	93.25	98.00	95.67	86.72	93.22	90.13	87.38
	Curricular	99.30	93.40	93.70	93.03	97.85	95.46	85.93	92.78	89.93	87.05
	RVFace	99.27	94.17	94.47	93.43	97.73	95.81	87.30	93.42	90.60	87.63
	BoundaryFace	99.25	94.33	94.73	93.50	97.98	95.96	87.47	93.18	90.15	87.53
	AdaFace	99.25	93.82	94.07	93.02	97.85	95.60	86.22	93.27	90.20	87.32
	RepFace(ours)	99.33	94.78	94.30	94.05	98.30	96.15	87.58	94.50	91.48	88.95
20%	ArcFace	98.73	92.30	91.79	92.48	97.10	94.48	85.18	91.53	89.20	84.45
	MV-Softmax	99.12	92.58	92.34	92.45	97.15	94.73	85.68	91.83	89.50	84.97
	Curricular	99.02	91.35	91.56	92.25	96.52	94.14	84.48	90.82	88.52	84.38
	RVFace	99.10	93.60	93.86	93.10	97.62	95.45	86.15	92.80	89.92	87.08
	BoundaryFace	99.38	94.22	93.89	93.40	97.90	95.76	86.23	93.22	90.00	87.27
	AdaFace	98.97	92.10	92.26	92.57	97.15	94.61	84.98	91.17	88.73	84.57
	RepFace(ours)	99.40	94.43	93.44	93.50	98.03	95.76	87.25	93.97	90.83	88.35

Table 5: Verification performance (%) of different methods trained on CASIA-WebFace.

Ratio	Method	LFW	AgeDB-30	CFP-FP	CALFW	SLLFW	Avg.	IJB-B		IJB-C	
								1.E-05	0.0001	1.E-05	0.0001
0%	ArcFace	99.75	97.80	95.21	95.90	99.53	97.64	88.94	94.31	93.58	95.80
	MV-Softmax	99.76	97.82	95.34	95.92	99.37	97.64	87.60	94.11	93.28	95.68
	Curricular	99.71	97.80	95.37	95.78	99.45	97.60	87.93	94.25	93.21	95.75
	BoundaryFace	99.63	97.82	95.46	95.75	99.32	97.60	88.69	94.20	93.36	95.73
	RVFace	99.78	97.77	95.74	95.88	99.53	97.74	86.21	93.62	91.68	94.99
	AdaFace	99.75	97.81	95.84	95.82	99.60	97.76	87.22	94.20	92.82	95.66
	RepFace(ours)	99.79	97.82	95.86	95.92	99.48	97.77	88.98	94.67	94.01	96.14
20%	ArcFace	99.65	97.65	95.05	95.97	99.48	97.56	80.13	92.71	88.41	94.39
	MV-Softmax	99.64	97.37	95.20	95.70	99.34	97.45	80.30	92.79	88.67	94.65
	Curricular	99.65	97.55	95.22	95.68	99.30	97.48	81.40	92.92	89.00	94.89
	BoundaryFace	99.68	97.69	95.35	95.97	99.45	97.63	84.49	93.11	90.79	94.70
	RVFace	99.68	97.66	95.26	95.75	99.50	97.57	84.09	92.93	90.60	94.43
	AdaFace	99.65	97.27	93.91	95.73	99.37	97.19	75.40	90.31	83.19	92.27
	RepFace(ours)	99.70	97.68	95.35	96.10	99.50	97.67	84.49	93.12	90.91	94.77

Table 6: Evaluation outcomes for MS1Mv2 Raw Dataset and 20% Closed-Set Noise-Infused Synthetic Dataset.

(SOTA) methods on the original dataset, the advantages of our method become more pronounced when the noise level is increased to 20%. In the experiments with the synthetic 20% closed-set noise dataset, our method exhibits better performance compared to previous label noise processing methods. Our method achieves the best results in both average validation sets and IJB-B/IJB-C test sets, especially in the IJB-C dataset where TAR@FAR=1e-4, surpassing BoundaryFace and RVFace by 0.07% and 0.34%, respectively. These results confirm that our method remains effective in large-scale datasets, thereby validating its robustness and generalizability.

Conclusion

This paper proposes a novel closed-set noise training framework, namely RepFace. Firstly, we use generated auxiliary

closed-set noise samples to simulate closed-set noise sample, and propose an ASC model to identify and discard noisy samples, especially during the early-stage training of the model. Subsequently, we partition the datasets into three parts (clean, ambiguous, and noise) based on the cosine similarity difference between the nearest negative and positive classes. We introduce a label robust fusion module by injecting a memory bank solution to accumulate model prediction for effectively handling these ambiguous samples. Moreover, we apply smooth label correction using the smooth parameter k to correct closed-set noise samples. This framework minimizes the impact of label noise on the model during early training and maximizes utilization of ambiguous and noise samples during post-training. Extensive experiments validate our method’s efficacy and achieve state-of-the-art results.

Acknowledgments

This work is partially supported by National Natural Science Foundation of China (62376231), Sichuan Science and Technology Program (24NSFSC1070), Tangshan Basic Research Science and Technology Program (23130230E), Tangshan Lunan District Science and Technology Program. Fundamental Research Funds for the Central Universities (2682023ZDPY001).

References

- An, X.; Zhu, X.; Gao, Y.; Xiao, Y.; Zhao, Y.; Feng, Z.; Wu, L.; Qin, B.; Zhang, M.; Zhang, D.; and Fu, Y. 2021. Partial FC: Training 10 Million Identities on a Single Machine. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*.
- Cao, Q.; Shen, L.; Xie, W.; Parkhi, O. M.; and Zisserman, A. 2018. Vggface2: A dataset for recognising faces across pose and age. In *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*, 67–74. IEEE.
- Deng, J.; Guo, J.; Liu, T.; Gong, M.; and Zafeiriou, S. 2020a. Sub-center arcface: Boosting face recognition by large-scale noisy web faces. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XI 16*, 741–757. Springer.
- Deng, J.; Guo, J.; Ververas, E.; Kotsia, I.; and Zafeiriou, S. 2020b. Retinaface: Single-shot multi-level face localisation in the wild. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5203–5212.
- Deng, J.; Guo, J.; Xue, N.; and Zafeiriou, S. 2019. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 4690–4699.
- Deng, J.; Guo, J.; Yang, J.; Lattas, A.; and Zafeiriou, S. 2021. Variational prototype learning for deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11906–11915.
- Deng, J.; Zhou, Y.; and Zafeiriou, S. 2017. Marginal Loss for Deep Face Recognition. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*.
- Deng, W.; Hu, J.; Zhang, N.; Chen, B.; and Guo, J. 2017. Fine-grained face verification: FGLFW database, baselines, and human-DCMN partnership. *Pattern Recognition*, 66: 63–73.
- Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.
- Duta, I. C.; Liu, L.; Zhu, F.; and Shao, L. 2020. Improved Residual Networks for Image and Video Recognition. *arXiv preprint arXiv:2004.04989*.
- Guo, Y.; Zhang, L.; Hu, Y.; He, X.; and Gao, J. 2016. Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part III 14*, 87–102. Springer.
- He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.
- Howard, A.; Sandler, M.; Chu, G.; Chen, L.-C.; Chen, B.; Tan, M.; Wang, W.; Zhu, Y.; Pang, R.; Vasudevan, V.; et al. 2019. Searching for mobilenetv3. In *Proceedings of the IEEE/CVF international conference on computer vision*, 1314–1324.
- Hu, W.; Huang, Y.; Zhang, F.; and Li, R. 2019. Noise-tolerant paradigm for training face recognition CNNs. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 11887–11896.
- Huang, G. B.; Mattar, M.; Berg, T.; and Learned-Miller, E. 2008. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. In *Workshop on faces in 'Real-Life' Images: detection, alignment, and recognition*.
- Huang, Y.; Wang, Y.; Tai, Y.; Liu, X.; Shen, P.; Li, S.; Li, J.; and Huang, F. 2020. Curricularface: adaptive curriculum learning loss for deep face recognition. In *proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5901–5910.
- Kemelmacher-Shlizerman, I.; Seitz, S. M.; Miller, D.; and Brossard, E. 2016. The megaface benchmark: 1 million faces for recognition at scale. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4873–4882.
- Kim, M.; Jain, A. K.; and Liu, X. 2022. AdaFace: Quality Adaptive Margin for Face Recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
- Liu, W.; Wen, Y.; Yu, Z.; Li, M.; Raj, B.; and Song, L. 2017. SpheroFace: Deep hypersphere embedding for face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 212–220.
- Liu, Z.; Luo, P.; Wang, X.; and Tang, X. 2015. Deep Learning Face Attributes in the Wild. In *Proceedings of International Conference on Computer Vision (ICCV)*.
- Lu, Y.; and He, W. 2022. SELC: Self-Ensemble Label Correction Improves Learning with Noisy Labels. *IJCAI*.
- Maze, B.; Adams, J.; Duncan, J. A.; Kalka, N.; Miller, T.; Otto, C.; Jain, A. K.; Niggel, W. T.; Anderson, J.; Cheney, J.; et al. 2018. Iarpa janus benchmark-c: Face dataset and protocol. In *2018 international conference on biometrics (ICB)*, 158–165. IEEE.
- Meng, Q.; Zhao, S.; Huang, Z.; and Zhou, F. 2021. Mag-Face: A universal representation for face recognition and quality assessment.
- Moschoglou, S.; Papaioannou, A.; Sagonas, C.; Deng, J.; Kotsia, I.; and Zafeiriou, S. 2017. Agedb: the first manually collected, in-the-wild age database. In *proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 51–59.
- Saadabadi, M. S. E.; Malakshan, S. R.; Zafari, A.; Mostofa, M.; and Nasrabadi, N. M. 2023. A quality aware sample-to-sample comparison for face recognition. In *Proceedings of*

- the *IEEE/CVF Winter Conference on Applications of Computer Vision*, 6129–6138.
- Sengupta, S.; Chen, J.-C.; Castillo, C.; Patel, V. M.; Chellappa, R.; and Jacobs, D. W. 2016. Frontal to profile face verification in the wild. In *2016 IEEE winter conference on applications of computer vision (WACV)*, 1–9. IEEE.
- Sun, Z.; Feng, C.; Patras, I.; and Tzimiropoulos, G. 2024. LAFS: Landmark-based Facial Self-supervised Learning for Face Recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1639–1649.
- Sun, Z.; and Tzimiropoulos, G. 2022. Part-based Face Recognition with Vision Transformers. In *33rd British Machine Vision Conference 2022, BMVC 2022, London, UK, November 21-24, 2022*, 611. BMVA Press.
- Tan, M.; and Le, Q. 2019. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, 6105–6114. PMLR.
- Wang, F.; Chen, L.; Li, C.; Huang, S.; Chen, Y.; Qian, C.; and Loy, C. C. 2018a. The devil of face recognition is in the noise. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 765–780.
- Wang, H.; Wang, Y.; Zhou, Z.; Ji, X.; Gong, D.; Zhou, J.; Li, Z.; and Liu, W. 2018b. Cosface: Large margin cosine loss for deep face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 5265–5274.
- Wang, M.; Deng, W.; Hu, J.; Peng, J.; Tao, X.; and Huang, Y. 2018c. Racial faces in-the-wild: Reducing racial bias by deep unsupervised domain adaptation. *arXiv preprint arXiv:1812.00194*, 5.
- Wang, X.; Wang, S.; Liang, Y.; Gu, L.; and Lei, Z. 2022. Rvface: Reliable vector guided softmax loss for face recognition. *IEEE Transactions on Image Processing*, 31: 2337–2351.
- Wang, X.; Wang, S.; Wang, J.; Shi, H.; and Mei, T. 2019. Co-mining: Deep face recognition with noisy labels. In *Proceedings of the IEEE/CVF international conference on computer vision*, 9358–9367.
- Wang, X.; Zhang, S.; Wang, S.; Fu, T.; Shi, H.; and Mei, T. 2020. Mis-classified vector guided softmax loss for face recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 12241–12248.
- Wei, H.; Tao, L.; Xie, R.; and An, B. 2021. Open-set label noise can improve robustness against inherent label noise. *Advances in Neural Information Processing Systems*, 34: 7978–7992.
- Wen, Y.; Liu, W.; Weller, A.; Raj, B.; and Singh, R. 2021. Sphereface2: Binary classification is all you need for deep face recognition. *arXiv preprint arXiv:2108.01513*.
- Whitelam, C.; Taborsky, E.; Blanton, A.; Maze, B.; Adams, J.; Miller, T.; Kalka, N.; Jain, A. K.; Duncan, J. A.; Allen, K.; et al. 2017. Iarpa janus benchmark-b face dataset. In *proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 90–98.
- Wu, S.; and Gong, X. 2022. BoundaryFace: A mining framework with noise label self-correction for Face Recognition. In *European Conference on Computer Vision*, 91–106. Springer.
- Yi, D.; Lei, Z.; Liao, S.; and Li, S. Z. 2014. Learning face representation from scratch. *arXiv preprint arXiv:1411.7923*.
- Zheng, T.; and Deng, W. 2018. Cross-pose lfw: A database for studying cross-pose face recognition in unconstrained environments. *Beijing University of Posts and Telecommunications, Tech. Rep*, 5(7).
- Zheng, T.; Deng, W.; and Hu, J. 2017. Cross-age lfw: A database for studying cross-age face recognition in unconstrained environments. *arXiv preprint arXiv:1708.08197*.
- Zhong, Y.; Deng, W.; Fang, H.; Hu, J.; Zhao, D.; Li, X.; and Wen, D. 2021. Dynamic training data dropout for robust deep face recognition. *IEEE Transactions on Multimedia*, 24: 1186–1197.
- Zhu, Z.; Huang, G.; Deng, J.; Ye, Y.; Huang, J.; Chen, X.; Zhu, J.; Yang, T.; Du, D.; Lu, J.; et al. 2022. Webface260M: A benchmark for million-scale deep face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(2): 2627–2644.