Towards Mask-robust Face Recognition

Tao Feng[1] Liangpeng Xu[1] Hangjie Yuan[2] Yongfei Zhao[1] Mingqian Tang[1] Mang Wang[1*]
[1]Alibaba Group [2]Zhejiang University
{shisi.ft,liangpeng.xlp,yongfei.zyf,mingqian.tmq,wangmang.wm}@alibaba-inc.com
hj.yuan@zju.edu.cn

Abstract

In this paper, we focus on the problem of mask-robust face recognition. Facial mask usually covers a major part of face, causing a significant reduction in extracting effective features. Due to such restriction, even the most advanced face recognition models are confronted with significant challenges. In light of this, this paper attempts to provide a reliable solution. Specifically, we introduce a mask-to-face image blending approach based on UV texture mapping, and a self-learning based cleaning pipeline for processing noisy training datasets. Then, considering the impacts of the long-tail distribution and hard faces samples, a loss function named Balanced Curricular Loss is introduced. Together with a bag of tricks is briefly presented. Experimental results show that the proposed solution separately achieved 84.528% @ Mask and 88.355% @ MR-ALL in InsightFace ms1m Track, which ranks 3rd when the paper submitted.

1. Introduction

Wearing mask during the COVID-19 pandemic is quite an effective measure for protection. However, the implementation of this measure has greatly degraded the performance of the most advanced face recognition model. It is thus imperative that existing face recognition models need to be upgraded to deal with the serious challenge. Firstly, the use of masks leads to nearly half of the useful facial information is reduced, which directly affects the quality of extracted face features by the model. Secondly, there lacks a large face dataset similar to Glint360k [1] and WebFace260M [19] for masked face, which is fatal to the model. To tackle the above difficulties, we propose a robust solution in the ICCV 2021-Masked Face Recognition (MFR) Challenge [2, 18]. This solution generates the masked face from the existing face data to solve the scarcity of raw data. Meanwhile, a cleaning pipeline is introduced to deal with the noise of the existing dataset. In addition, we propose Balanced Curricular Loss (BCL) to reduce the impacts of long-tailed distribution and hard example samples.

Paper Outline. We first adopt a mask-to-face image blending approach to generate masked faces in Section 2.1, and then present the details of cleaning by self-learning to purify the noise-infested datasets in Section 2.2. In Section 2.3 we set up a novel loss function called Balanced Curricular Loss. In Section 3 we briefly introduce a bag of tricks in our implementation. At last, the experimental results and ablation study is provided in Section 4.

2. Method

2.1. A Mask-to-Face Image Blending Approach

Intensive masked face dataset training is essential to obtain a reliable model on masked face recognition. To supply sufficient data for training, a feasible solution is to draw masks on human face images. Yet the position-paste approach popular today cannot provide real and flexible masked face data. Inspired by PRNet [7, 17], we adopt the mask-to-face image blending approach to get more natural masked faces.

To be specific, we adopt the mask-to-face texture mapping approach and then generate the masked face images with rendering. The principle of this technology is generating a UV texture map for the mask images and face images are re-constructed with the 3D approach. Then the masked face images are blended based on the texture spaces. The masks flexibly fit different areas on faces through the mask-to-face blending image approach we used.

Here, we use 8 types of masks, which yields 8 types of masked faces based on existing face databases, as shown in Figure 1. Ulteriorly, we use UV mapping to overcome uneven mask surfaces or curvy face edges of planar projection, which often leads to unsatisfactory fitting between human faces and masks. In addition, the rendering process is adopted to generate different rendering effects, such as adjusting mask angles and shadow effects.

[1]Corresponding author.

Figure 1. Some examples of the original faces and their masked versions generated with the mask-to-face image blending approach we used.

2.2. Self-learning based Cleaning

The purity of training data is an essential factor affecting the performance of state-of-the-art face recognition models [16, 12]. Most of face recognition models are built upon the assumption that the entire datasets are clean. While this assumption does not hold, and the poor purity of datasets will significantly lower model performance. The competition provides with noise-infested MS1M families [8, 4, 3] and WebFace260M [19] datasets, making it inevitable that highly noisy masked face datasets are generated. Therefore, we design a scalable and efficient self-learning [9] based cleaning approach to purify the noise-infested datasets [13].

The overall self-learning based cleaning framework is shown in Figure 2. The details are introduced as follows, (1) An initial model is first trained with the MS1M-V3 to clean the original dataset, which mainly consist of ID-Union, Inter-ID cleaning and Intra-ID cleaning. (2) Then the i-th model is trained on the cleaned datasets from (1). (3) We iterate this process by initializing the i-th model as the (i+1)-th model. Here, Inter-ID cleaning is first conducted and then Intra-ID.

ID-Union. In this paragraph, the ID-Union procedure is introduced to scan duplicate ID folders. Specifically,

1. DBSCAN [6] algorithm is performed to cluster faces in each ID folder.
2. The center feature of the maximum cluster obtained by DBSCAN is calculated.
3. ID information with duplicate center feature is removed.

Inter-ID Cleaning. In terms of the noise between different identities, we use the center feature between identities to purify the dataset, as shown in Figure 3. Specifically,

1. The similarity search is performed for the center feature after ID-union.
2. The folders are merged if the similarity is higher than 0.75.

After the Inter-ID cleaning procedure, the face samples with same id but different labels are merged. Thus the noise level in the inter-class is greatly filtered.

Intra-ID Cleaning. In terms of the data noise within the same identity, we use the center feature within the identity to obtain clean datasets, as shown in Figure 3. Specifically,

1. The center feature of the merged face cluster is updated.
2. A similarity search is performed for the max center feature and each face feature.
3. The faces are deleted if the similarity is lower than 0.25.

After Intra-ID Cleaning is performed, the face samples with incorrect labeling within the same identity are removed as outliers. This allows us to effectively suppress the data noise introduced by incorrect labeling.

Self-learning Cleaning. In this procedure, we perform self-learning by initializing the i-th model as the (i+1)-th model. Here, each Inter-ID and Intra-ID cleaning is conducted on the initial dataset by the learned model.

It is worth noting that the dataset cleaning pipeline we used relies on the center feature, which is different from the CAST [19]. The scheme has good robustness, making it easier to purify the dataset.

2.3. Balanced Curricular Loss based Model

Curricular Loss We use the Curricular Loss [11] with IR_SE-101 [3, 10] as our base model. More specifically, Curricular Loss is employed to emphasize adaptively adjust the relative importance of easy and hard samples during different training stages. To extract the model outputs from the CurricularFace framework, we start by generating the Curricular Loss in the same way as in [11],

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

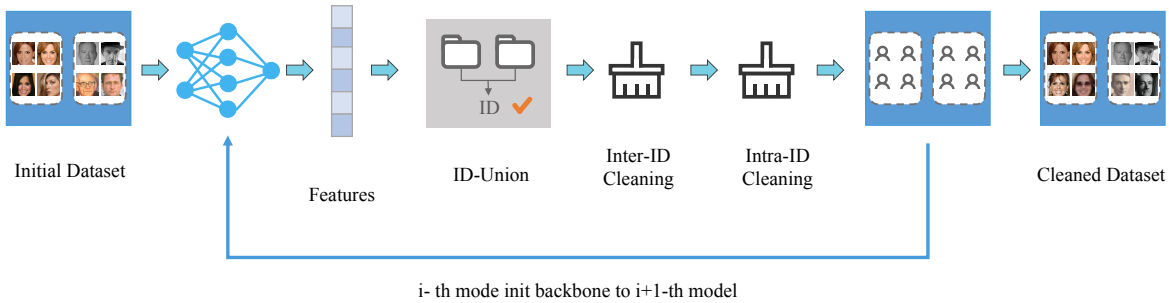

Figure 2. The proposed Self-learning Cleaning.

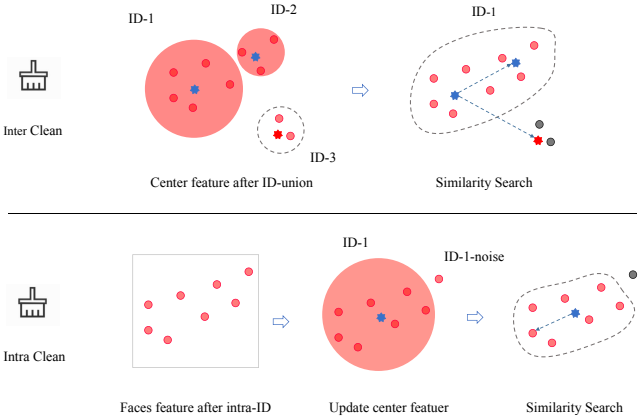

Figure 3. Diagram of Intra and Inter cleaning unit. (The star marks denote the center feature.)

where $\cos(\theta_{y_i} + m)$ and $N\left(t^{(k)}, \cos\theta_j\right)$ are the cosine similarity function of positive and negative [11]. However, it lacks consideration for the long-tail distribution of the provided dataset, which leads to a biased estimation of the model.

Balanced Softmax Loss. As far as we are concerned, long-tail distribution will pose formidable challenges for the masked face recognition task. In this section, a loss function named balanced softmax [15] is introduced. Unlike the mainstream multi-class loss supporting the algorithm to achieve satisfactory performance in class-balanced datasets, balanced softmax takes into account the problems concerning label distribution migration and gradient of the rare category during the training and testing process, the original formula is detailed as,

$$\mathcal{L}_{Balanced} = -\log\left(\hat{\phi}_y\right) = -\log\left(\frac{n_y e^{\eta_y}}{\sum_{i=1}^{k} n_i e^{\eta_i}}\right) \quad (2)$$

where $\hat{\phi}_y$ is the desired conditional probability of the imbalanced training set. And η is the output of the model, n is the number of face samples in current category.

Balanced Curricular Loss. Combining the CurricularLoss and Balanced Softmax, we present the detail of our proposed Balanced Curricular Loss. BCL fully exploited hard samples meanwhile emphasized the effects of long-tail distribution. We define the Balanced Curricular Loss as,

$$\mathcal{L} = -\log\frac{n_{y_i} e^{s\cos(\theta_{y_i}+m)}}{n_{y_i} e^{s\cos(\theta_{y_i}+m)} + \sum_{j=1, j\neq y_i}^{n} n_j e^{sN(t^{(k)}, \cos\theta_j)}} \quad (3)$$

Table 1. The results (%) on the proposed BCL.

Method	Mask	MR-ALL	Infer(ms)
CL	70.319	85.017	9.196
BCL	**73.693**	**86.154**	9.038

3. Bag of tricks

In this section, we present bag of tricks in our implementation.

3.1. Color jitter

Color jitter is adopted to randomly change the brightness, contrast, and saturation of the face images.

3.2. Horizontal flip

Horizontal flip is used during the training process for all settings.

3.3. Cutout strategy

Considering the reduction of facial features due to the mask, this is similar to bring the hard sample in the face recognition model. Therefore, we perform the cutout [5] strategy to simulate hard face samples to improve the performance of recognition models.

3.4. Label smoothing

We try the label smoothing [14] as a regularization strategy, and we find it brings a few improvements when this paper submitted.

4. Experimental results

In this section, we present the results of extensive experiments to demonstrate the performance of the solution we

Table 2. Perfomance (%) of the proposed approach on InsightFace-ms1m.

Method	Mask	Children	African	Caucasian	South Asian	East Asian	MR-ALL	Inter(ms)
w/o clean & w/o mask	69.522	66.299	81.540	89.183	64.771	64.771	85.381	9.001
clean & w/o mask	78.181	68.425	85.857	91.673	92.136	67.522	87.852	9.132
Ours	**84.528**	68.303	**86.820**	**92.251**	88.326	**68.595**	**88.355**	9.318

Table 3. The results (%) on InsightFace-ms1m under different proportions of masked faces in all datasets.

Mask/all	Mask	Children	African	Caucasian	South Asian	East Asian	MR-ALL	Infer(ms)
5%	78.856	68.006	82.957	90.365	89.974	65.591	**86.185**	8.834
10%	79.164	63.272	81.443	89.410	88.857	65.455	85.576	8.993
15%	82.065	67.512	82.098	89.978	88.635	65.716	85.764	9.128
20%	82.553	67.070	80.357	89.236	87.482	64.625	84.674	9.295
25%	**82.682**	60.132	77.510	87.562	84.217	59.873	81.700	9.105

Table 4. The results (%) on InsightFace-ms1m under different threshold of Intra-ID Cleaning.

Threshold	Mask	Children	African	Caucasian	South Asian	East Asian	MR-ALL	Infer(ms)
0.05	72.013	67.084	82.812	89.958	90.037	66.014	86.139	9.193
0.15	72.150	66.308	82.686	89.708	89.815	65.146	85.651	8.972
0.25	**73.693**	65.984	83.414	90.173	89.991	65.852	**86.154**	9.287
0.35	71.805	64.494	82.682	89.718	89.404	64.814	85.672	9.189
0.45	68.581	62.331	81.182	88.568	88.080	62.985	84.268	8.933
0.50	62.407	56.040	76.519	84.849	85.018	57.824	79.983	9.181

proposed. Then, we conduct a detailed ablation study to further investigate our scheme.

Training details. We train the model using 32 Tesla V100 GPUs for 24 epochs. The batch size is set to 160 on single GPU. The model is trained with SGD optimizer with momentum 0.9 and weight decay $5e$-4. The learning rate starts from 0.1 and is divided by 10 at 10, 18, 22 epochs. The scale s is set to 128 and margin m is set to 0.5.

4.1. Results on MFR challenge protocol

We evaluate the performance of the proposed approach on the MFR challenge protocol. Results on the InsightFace ms1m track are shown in Table 2. The test datasets mainly comes from InsightFace Recognition Test (IFRT). In Table 2, our approach achieves the 84.528% @ Mask and 88.355% @ MR-ALL performance, ranks 3rd on the challenge when this paper submitted. Compared with w/o clean and w/o mask scenarios, our approach shows better performance under both Mask and MR-ALL. The results in Table 2 demonstrate the superiority of approach we proposed. Meanwhile, the results on the proposed BCL are present in Table 1.

4.2. Different proportions of masked faces

Here, the impact of the mask added on the performance of the model is evaluated through different proportions of masked faces in all datasets. As shown in Table 3, the proportions are set to 5%, 10%, 15%, 20%, and 25%, respectively. It is clear that when the proportion is 5%, the MR-ALL reaches the maximum value of 86.185%. However, when the proportion is adjusted to 25, the MR-ALL drops to the lowest value of 81.7%, even though the MASK increased to 82.682%. This indicates that the percentage of masked faces is an important parameter for the trade-off of model efficiency when masked faces and normal faces are trained together.

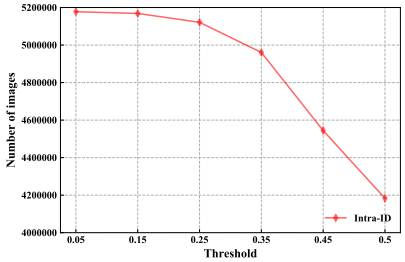

Figure 4. Effects of the different thresholds in Intra-ID cleaning.

4.3. The effect of various threshold settings

We validate the solution we proposed on different thresholds of Intra-ID clean, ranging from 0.05 to 0.5, reported in Table 4. Our approach achieves 73.693% @ mask and 86.153% @ MR-ALL with 0.25 threshold. As shown in Table 4, we observe that obvious wrong faces are cleaned reasonably well when a loose threshold is set. But a strict threshold leads to a negative effect, which is due to some hard samples being cleaned along with noise. These hard samples are significant to the model, while increase the dif-

ficulty of cleaning. Therefore, in terms of performance, we use the 0.25 threshold as an appropriate candidate. Figure 4 shows the changes in the number of images in the dataset for the corresponding threshold settings.

5. Conclusion

In this work, we present the details of our solution for working with masked face recognition. First, we use a novel mask-to-face mapping to generate good quality masked faces meanwhile adopt a cleaning method in a self-learning manner to purify the noise in datasets. Then we introduce a novel Balanced Curricular Loss to handling long-tail distribution and hard samples. Besides, we explore the optimal proportions of masked faces in all datasets when mixing them in training. Our solution significantly reduces the impact of noise on the model by an appropriate setting. The experimental results on the ICCV 2021-MFR challenge demonstrate the effectiveness of the proposed approach.

References

[1] Xiang An, Xuhan Zhu, Yang Xiao, Lan Wu, Ming Zhang, Yuan Gao, Bin Qin, Debing Zhang, and Ying Fu. Partial FC: training 10 million identities on a single machine. *CoRR*, abs/2010.05222, 2020.

[2] Jiankang Deng, Jia Guo, Xiang An, Zheng Zhu, and Stefanos Zafeiriou. Masked face recognition challenge: The insightface track report. In *Proceedings of the International Conference on Computer Vision Workshops*, 2021.

[3] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, 2019.

[4] Jiankang Deng, Yuxiang Zhou, and Stefanos Zafeiriou. Marginal loss for deep face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops*, 2017.

[5] Terrance Devries and Graham W. Taylor. Improved regularization of convolutional neural networks with cutout. *CoRR*, abs/1708.04552, 2017.

[6] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, KDD*, 1996.

[7] Yao Feng, Fan Wu, Xiaohu Shao, Yanfeng Wang, and Xi Zhou. Joint 3d face reconstruction and dense alignment with position map regression network. In *Computer Vision - ECCV - 15th European Conference, ECCV*, 2018.

[8] Yandong Guo, Lei Zhang, Yuxiao Hu, Xiaodong He, and Jianfeng Gao. Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, *Computer Vision - ECCV - 14th European Conference, ECCV*, 2016.

[9] Jiangfan Han, Ping Luo, and Xiaogang Wang. Deep self-learning from noisy labels. In *IEEE/CVF International Conference on Computer Vision, ICCV*, 2019.

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, 2016.

[11] Yuge Huang, Yuhan Wang, Ying Tai, Xiaoming Liu, Pengcheng Shen, Shaoxin Li, Jilin Li, and Feiyue Huang. Curricularface: Adaptive curriculum learning loss for deep face recognition. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR*, 2020.

[12] Youngdong Kim, Junho Yim, Juseung Yun, and Junmo Kim. NLNL: negative learning for noisy labels. In *IEEE/CVF International Conference on Computer Vision, ICCV*, 2019.

[13] Kuang-Huei Lee, Xiaodong He, Lei Zhang, and Linjun Yang. Cleannet: Transfer learning for scalable image classifier training with label noise. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, 2018.

[14] Rafael Müller, Simon Kornblith, and Geoffrey E. Hinton. When does label smoothing help? In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems, NeurIPS*, 2019.

[15] Jiawei Ren, Cunjun Yu, Shunan Sheng, Xiao Ma, Haiyu Zhao, Shuai Yi, and Hongsheng Li. Balanced meta-softmax for long-tailed visual recognition. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems, NeurIPS*, 2020.

[16] Fei Wang, Liren Chen, Cheng Li, Shiyao Huang, Yanjie Chen, Chen Qian, and Chen Change Loy. The devil of face recognition is in the noise. In *Computer Vision - ECCV - 15th European Conference, ECCV*, 2018.

[17] Jun Wang, Yinglu Liu, Yibo Hu, Hailin Shi, and Tao Mei. Facex-zoo: A pytorch toolbox for face recognition. *arxiv*, abs/2101.04407, 2021.

[18] Zheng Zhu, Guan Huang, Jiankang Deng, Yun Ye, Junjie Huang, Xinze Chen, Jiagang Zhu, Tian Yang, Jia Guo, Jiwen Lu, Dalong Du, and Jie Zhou. Masked face recognition challenge: The webface260m track report. In *Proceedings of the International Conference on Computer Vision Workshops*, 2021.

[19] Zheng Zhu, Guan Huang, Jiankang Deng, Yun Ye, Junjie Huang, Xinze Chen, Jiagang Zhu, Tian Yang, Jiwen Lu, Dalong Du, and Jie Zhou. Webface260m: A benchmark unveiling the power of million-scale deep face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, 2021.