This ICCV workshop paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

Boosting Fairness for Masked Face Recognition

Jun Yu
University of Science and Technology of China
Hefei, China
harryjun@ustc.edu.cn

Xinlong Hao*
University of Science and Technology of China
Hefei, China
haoxl@mail.ustc.edu.cn

Zeyu Cui
University of Science and Technology of China
Hefei, China
mg980806@mail.ustc.edu.cn

Peng He
University of Science and Technology of China
Hefei, China
hp0618@mail.ustc.edu.cn

Tongliang Liu
Trustworthy Machine Learning Lab, The University of Sydney
Sydney, Australia
tongliang.liu@sydney.edu.au

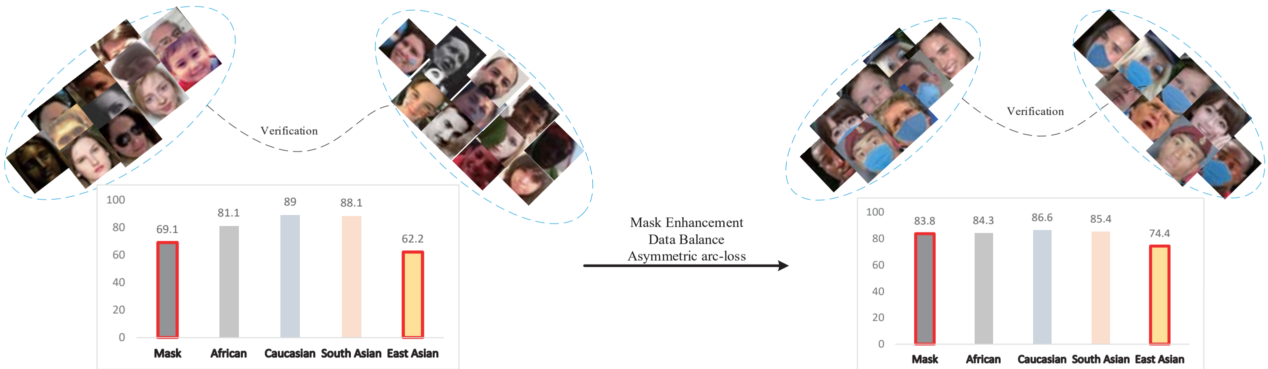

Figure 1: An overview of the proposed method. When using the original dataset to train a masked face recognition model, there would be unfairness between different ethnic groups. Using mask enhancement, data balance and asymmetric arc-loss, our method obtains a fairer and better masked face recognition results.

Abstract

Face recognition achieved excellent performance in recent years. However, its potential for unfairness is raising alarm. For example, the recognition rate for the special group of East Asian is quite low. Many efforts have spent to improve the fairness of face recognition. During the COVID-19 pandemic, masked face recognition is becoming a hot topic but brings new challenging for fair face recognition. For example, the mouth and nose are important to recognizing faces of Asian groups. Masks would further reduce the recognition rate of Asian faces. To this end, this paper proposes a fair masked face recognition system. First, an appropriate masking method is used to generate masked faces. Then, a data re-sampling approach is employed to balance the data distribution and reduce the bias based on the analysis of training data. Moreover, we propose an asymmetric-arc-loss which is a combination of arcface loss and circle-loss, it is useful for increasing recognition rate and reducing bias. Integrating these techniques, this paper obtained fairer and better face recognition results on masked faces.

1. Introduction

Face recognition has achieved excellent performance in recent years [7, 9, 15, 34], along with the development of deep neural networks [18, 28, 30, 33, 39]. However, as its wider and wider applications, the potential of unfairness for face recognition is raising alarm [1, 4, 27, 29]. For instance, according to [11], a year-long research investigation across 100 police departments indicated that the face recognition system has racial bias. Obviously, it is particularly important to obtain a fair face recognition system.

During the COVID-19 coronavirus epidemic,wearing facial masks is often required in public areas, which poses a huge challenge to face recognition. Traditional face recognition systems may not effectively recognize the masked faces, but removing the mask for authentication will increase the risk of virus infection. The COVID-19 pandemic has driven a need to understand how face recognition technology deals with occluded faces, often with just the periocular area and above visible. Recently, some commercial providers have announced the availability of face recognition algorithms capable of handling face masks, and an increasing number of research publications have surfaced on the topic of face recognition on people wearing masks. However, the problem of fairness has not been well studied. Moreover, according to our experimental analysis, masked faces may cause greater racial bias, for example, the mouth and nose are important to recognizing faces of Asian groups. Masks would further reduce the recognition rate of Asian faces. Therefore, a fair masked face recognition system is very necessary.

In this paper, we present a face recognition method to achieve a fair masked face recognition system. First, publicly masked face dataset is few, so we use mask augmentation tool to generate some masked faces. Then, after obtaining the face dataset, as diversity between different ethnic groups is very large, a series of face preprocessing methods are used to reduce bias and improve accuracy at the same time. For instance, we use a data re-sampling method to balance the data distribution by under-sampling the majority class. Training data enhancement and test time augmentation are used for obtaining improved accuracy. Then training data would be used to train face recognition models, trained models are used to extract features of test data. Next, by calculating the cosine similarity between two feature vectors, confidence scores of test data would be generated, which indicate the degree two faces belong to the same person.

The main contributions of this paper are summarized as follows:

(1) An appropriate mask enhancement method is used to generate masked faces based on MS1M dataset, it is helpful for us to build a masked face recognition system.

(2) To acquire low bias and high accuracy, not only multiple training data enhancement methods are used, but also a data re-sampling approach is adopted to balance the data distribution.

(3) An asymmetric-arc-loss is used to obtain fairer and better results, which is a combination of arc-face loss [9] and circle-loss [32].

2. Related Work

Face recognition. State-of-the-art face recognition algorithms are based on deep learning models. These models learn to extract the important features from a face image and embed them into an n-dimensional vector with small intra-class and large inter-class distance. These models are trained mainly following two approaches. The first one consists on training a multi-class classifier considering one class for each identity in the training dataset, normally using a softmax function [9,25]. In the second one, the embedding is learnt directly, comparing the results of different inputs to minimize the intra-class distance and to maximize the inter-class distance, for example using the triplet loss [31]. Both softmax-loss-based and triplet-loss-based models suffer from face-mask occlusions in terms of accuracy, as reported by [8] and [22]. However, as stated in [9], triplet-loss-based models require a data preparation step prior to the training phase, in order to select the triplets correctly. For this reason, we decided to address the problem using a softmax-loss approach. More specifically, we selected ArcFace [9] as our baseline, since it has been proven to be the approach that reports the best results for the face recognition task.

Fair face recognition. As point out by some works [1, 2, 36], the main cause of the model bias between well-represented groups and under-represented groups is the distribution of training dataset. As shown in [36], we can easily observe that the commonly used face recognition datasets [6, 14, 19, 21, 40] are dominated by Caucasian identities, since the dataset is mainly formed by Caucasian subjects. Face recognition models performance on Caucasian outperforms that on other groups of people, such as African, Asian, and Indian. Similarly, gender is another aspect of face recognition datasets imbalance, i.e., dataset mainly consists of male faces.

To solve these problems, many efforts on face recognition aim to tackle the class imbalance problem on training data. For example, in prior-DNN era, Zhang et al. [43] proposed a cost-sensitive learning framework to reduce misclassification rate of face identification. To correct the skew of separating hyperplanes of SVM on imbalanced data, Liu et al. [26] proposed Margin-Based Adaptive Fuzzy SVM that obtains a lower generalization error bound. In the DNN era, face recognition models are trained on large-scale face datasets with highly-imbalanced class distribution. Range Loss [42] learns a robust face representation that makes the most use of every training sample. To mitigate the impact of insufficient class samples, center-based feature transfer learning [41] and large margin feature augmentation [37] are proposed to augment features of minority identities and equalize class distribution. Besides, the FRVT 2019 [13] shows the demographic bias of over 100 face recognition algorithms. To uncover deep learning bias, Alexander et al. [2] developed an algorithm to mitigate the hidden biases within training data. Wang et al. [36] proposed a domain adaptation network to reduce racial bias in face recognition. They recently extended their work using reinforcement learning to find optimal margins of additive angular margin based loss functions for different races [35].

Masked face recognition. Since the rise of COVID-19, several works have been presented in order to solve masked face recognition task. The proposed methods tackle the problem following different approaches that can be categorized in three groups. The first group uses generative adversarial networks (GAN) to unmask faces prior to feeding them to the face recognition model [10, 23]. Using this approach it is not necessary to retrain the recognition model. However, the reconstructed faces are synthetic and their reliability depends on the quality of the data, the network and the training process. In addition, the process of removing the mask noticeably increases the computation time.

The approach adopted by the second group consists of extracting features only from the upper part of the face [17]. As the processed region of the face is smaller, the trained network performs faster. Nevertheless, this causes an important drop of information when dealing with unmasked faces, so it is not suitable for applications mixing both use cases.

Finally, the last group tackles the problem training the face recognition network with a combination of masked and unmasked faces [3, 12]. In [3] they combined the VGG2 dataset [6] with augmented masked faces and train the model following the original pipeline described in FaceNet [31]. This way, the model learns to distinguish when a face is wearing a mask and to trust more in the features of the upper half of the face, but still extracts information from the whole face. On the other hand, Geng et al. [12] defined two centers for each identity which correspond to the full face images and the masked face images respectively. They used Domain Constrained Ranking for forcing the feature of masked faces getting closer to its corresponding full face center and vice-versa.

Masked face datasets. For the methods previously described, there is a need of masked face datasets. Some recent works have contributed to this task. For instance, Geng et al. [12] presented a dataset where each identity has masked and full face images with various orientations. However, the dataset contains only 11,615 images and 1,004 identities, which is not enough data for training a complex network such as ResNet-50 [16, 18] . In [5], the authors presented a dataset composed of 137,016 masked faces divided in two groups: correctly and incorrectly masked. Nevertheless, the dataset does not contain information about the identity of any of the subjects, so it cannot be used for the face recognition task. In [38], two additional datasets are presented: Real-world Masked Face Recognition Dataset (RMFRD), with 95,000 images and 525 identites, and Simulated Masked Face Recognition Dataset (SMFRD), with 500,000 and 10,000 subjects. Although the latter dataset contains a great number of samples, it is not yet sufficient to train a complex network, for example if we compare it with MS1MV2 dataset used in ArcFace [9], which contains 5.8 million images and 85,000 identities.

On the other hand, Anwar and Raychowdhury [3] present a tool for masking faces in images. It uses a face landmarks detector to identify the face tilt and six key features of the face necessary for adjusting and applying a mask template. This tool supports different types and colors of masks. In this work, we imitate this tool to generate a masked version of the face recognition datasets used for training and evaluation.

3. Dataset Description

For training data, Microsoft Celeb (MS-Celeb-1M) is a dataset of 10 million face images harvested from the Internet for the purpose of developing face recognition technologies. According to Microsoft Research, who created and published the dataset in 2016, MS Celeb is the largest publicly available face recognition dataset in the world, containing over 10 million images of nearly 100,000 individuals. Microsoft's goal in building this dataset was to distribute an initial training dataset of 100,000 individuals' biometric data to accelerate research. But actually, there are many noises in the original MS1M dataset, so the original MS1M dataset can not be used directly.

MS1M-V3 (MS1M-RetinaFace) [14] is created by the organization of insightface, they clean the original MS1M dataset to obtain a better dataset. The detailed introduction is shown in Table 1, besides, we show some samples from MS1M dataset as Fig. 2.

For test data, test dataset mainly comes from IFRT, contains 6,964 identities, 6,964 masked images and 13,928 non-masked images. There are totally 13,928 positive pairs and 96,983,824 negative pairs. The globalised multi-racial testset contains 242,143 identities and 1,624,305 images. The detailed information is shown in Table 2.

4. Proposed Method

The proposed approach consists of three parts, they are mask augmentation, face preprocess, training module and

Figure 2: Some samples of MS1M dataset. Face images without a mask (left) and with a mask (right).

Table 1: Comparison of different face datasets.

Dataset	Identity	Image	Racial distribution (%)			
			Caucasian	Asian	Indian	African
LFW [19]	5,749	13,233	69.9	13.2	2.9	14.0
IJB-A [20]	500	5,396	66.0	9.8	7.2	17.0
VGGFace2 [6]	8,631	3,086,894	74.2	6.0	4.0	15.8
CASIA-Webface [40]	10,575	494,414	84.5	2.6	1.6	11.3
MS1M [14]	**85,742**	**5,822,653**	**76.3**	**6.6**	**2.6**	**14.5**

Table 2: Information of testset.

Race-set	Identities	Images	Positive Pairs	Negative Pairs
African	43,874	298,010	870,091	88,808,791,999
Caucasian	103,293	697,245	2,024,609	486,147,868,171
Indian	35,086	237,080	688,259	56,206,001,061
Asian	59,890	391,9702	1,106,078	153,638,982,852
ALL	242,143	1,624,305	4,689,037	2,638,360,419,683

inference module, as shown in Fig. 3. In this work, we use a mask augmentation method to generate masked face. Besides, a data re-sampling approach is employed to balance the data distribution and reduce the bias based on the analysis of training data. IR_100 is used as backbone for feature extraction. In the training module, arcface is used as model head. Then in the inference module, test data feature vectors are generated based on trained backbone. Next, the final prediction is obtained by calculating the cosine similarity between two feature vectors. The prediction is defined as follows:

$$prediction = \frac{A \cdot B}{\|A\| \times \|B\|} \tag{1}$$

where A represents a feature vector, B represents another feature vector.

4.1. Mask Augmentation

Insightface presents a tool for masking faces in images. This tool supports different types and colors of masks, such as mask_white, mask_blue, mask_black and mask_green. And one choice of the mask_probs is 0.4, 0.4, 0.1, 0.1. In this work, we use this tool to generate a masked version of the face recognition dataset used for training and evaluation. We decide to generate a masked twin dataset from the original one and to combine them during the training process. Both datasets are shuffled separately using the same seed,

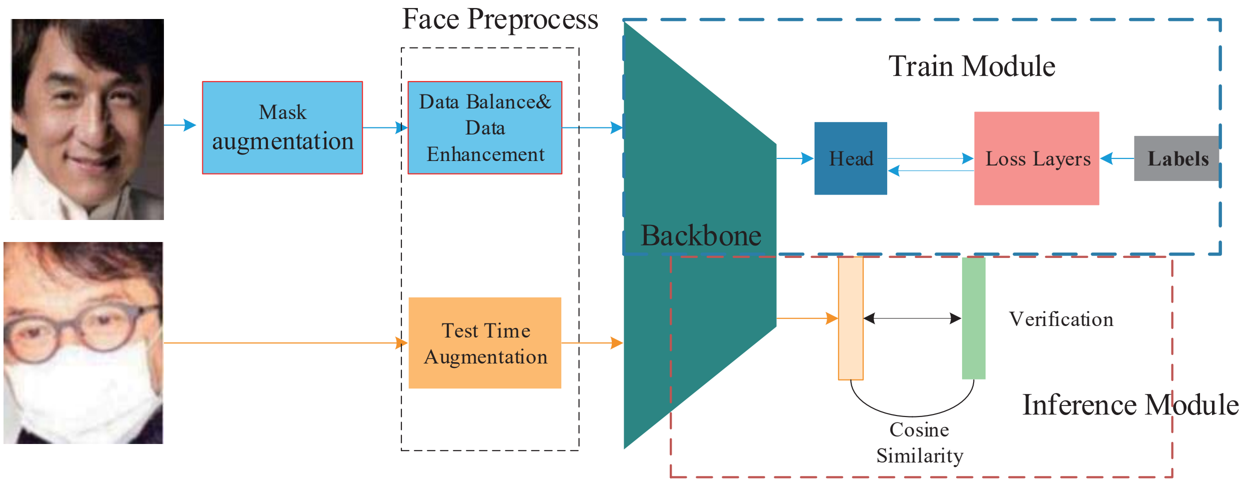

Figure 3: Pipeline of our method.

and for every new face image selected for the input batch, we decide whether the image is taken from masked dataset with a probability of 20%.

4.2. Face Preprocess

4.2.1 Data Balance

As shown in Table 1, the racial distribution is imbalanced, for instance, the number of Caucasian is ten times bigger than the number of Asian. The deep learning model is data-driven, so the gap of the face recognition result between Caucasian and Asian is also big. According to [1], if the face data is balanced in different ethnic groups, the accuracy of the face recognition result would be similar. So the intuitive idea is using a re-sampling method to get a balanced data and obtain a fair result finally.

For masked data, there is something different, the mouth and nose are more important to recognition a face for Asian than other ethnic groups [1], so the mask would reduce the recognition rate of Asian more. According to that, we mainly increase the weight of Asian base on data balance. One choice of our weight rate is Caucasian: Asian: Indian: African=0.05:1:1:0.2, the re-sample method could help us obtain a fair score.

4.2.2 Data Enhancement

In addition to data balance, multiple attempts have been made in the data enhancement phase. For training data enhancement, multiple enhanced methods, such as RandomHorizontalFlip (RHF), ColorJitter (CJ) and RandomBlur (RB), are used to improve data diversity.

In the inference stage, test time augmentation is used, we put the original face and the face which is flipped horizontally into backbone respectively, then both of the output feature vectors would be added together to produce the final feature vector.

4.3. Train Module and Inference Module

In train module, ir_100 is used as backbone. We proposed asymmetric-arc-loss for training, it contain arcface loss and circle loss.

The arcface loss is as follows:

$$L_{arc} = -\frac{1}{m}\Sigma_{i=1}^{m} log \frac{e^{s(cos(\theta_{y_i}+m))}}{e^{s(cos(\theta_{y_i}+m))} + \Sigma_{j=1,j\neq y_i}^{n} e^{scos\theta_j}}, \tag{2}$$

subject to

$$W_j = \frac{W_j}{||W_j||}, x_i = \frac{x_i}{||x_i||}, cos\theta_j = W_j^T x_i. \tag{3}$$

We assume θ_{y_i} as θ_p and others as θ_n. It's easy to analyze that the loss is monotonically increasing with respect to the θ_p while $\theta_p + m < \pi$ and monotonically decreasing with respect to θ_n, its convergence target is to maximize θ_n and to minimize θ_p. Then we take a look at Circle loss, which is:

$$L_{cir} = log[1 + \Sigma_{j=1}^{L} e^{(\gamma\alpha_n^j(s_n^j - \Delta_n))} \Sigma_{i=1}^{K} e^{(-\gamma\alpha_p^i(s_p^i - \Delta_p))}], \tag{4}$$

where s_n means negative similarity and s_p means positive. And in the class-level style, there is only one s_p so the loss can be shown as:

$$L_{cir} = -\frac{1}{m}\Sigma_{i=1}^{m} log \frac{e^{\gamma\alpha_p^{y_i}(s_p^{y_i} - \Delta_p)}}{e^{\gamma\alpha_p^{y_i}(s_p^{y_i} - \Delta_p)} + \Sigma_{j=1,j\neq y_i}^{n} e^{\gamma\alpha_n^j(s_n^j - \Delta_n)}}, \tag{5}$$

subject to

$$\alpha_p^i = |O_p - s_p^i|_+, \tag{6}$$

$$\alpha_n^j = |s_n^j - O_n|_+, \tag{7}$$

$$O_p = 1 + m, O_n = -m, \Delta_p = 1 - m, \Delta_n = m, \tag{8}$$

$$W_j = \frac{W_j}{||W_j||}, x_i = \frac{x_i}{||x_i||}, s^j = W_j^T x_i. \tag{9}$$

Based on the previous analysis, we can get two insights on improving the loss fuction.

Combination of advantages. Circle-loss provides the weight for s_n and s_p. We can also analyze that the loss is monotonically increasing with respect to the s_n and monotonically decreasing with respect to s_p, while both s_p and s_n are in $(0, 1)$. From the angel view, its convergence target is to maximize θ_n to $\pi/2$ and to minimize θ_p to 0.

Convergence target shift. From the previous anylasis, the convergence target of circle-loss is to maximize θ_n to $\pi/2$ and the convergence target of arc-loss is even maximize θ_n to π. But in fact, we do not alwayas need to maximize θ_n to $\pi/2$ or π. Since in face rocognition situation, we can not make sure that people in different sub ids are not similar at all, it is usual that two different people have some simliarity, like 0.3 or 0.2, and try to minimise this simliarity may make model pay useless attention on easy negtive samples. To solve this problem, we give a shift on the convergence target for negtive and make easy negtive samples contribute less to the final grad.

Since arc-loss provides an additive angular margin and circle-loss provides weight in training, we can make a combination for these two loss to use both of their advantages.

The asymmetric-arc-loss can be shown like this:

$$L = -\frac{1}{m}\Sigma_{i=1}^{m} log \frac{e^{\gamma \alpha_p^{y_i} cos(\theta_p^{y_i} + \Delta_p)}}{e^{\gamma \alpha_p^{y_i} cos(\theta_p^{y_i} + \Delta_p)} + \Sigma_{j=1, j \neq y_i}^{n} e^{\gamma \alpha_n^j cos(\theta_n^j + \Delta_n)}}, \tag{10}$$

subject to

$$\alpha_p^i = |O_p + \theta_p^i|_+, \tag{11}$$

$$\alpha_n^j = |O_n - \theta_n^j|_+, \tag{12}$$

$$O_p = \pi - tm, O_n = tm, \Delta_p = tm, \Delta_n = \pi - tm, \tag{13}$$

$$W_j = \frac{W_j}{||W_j||}, x_i = \frac{x_i}{||x_i||}, cos\theta^j = W_j^T x_i, \tag{14}$$

where γ and tm are hyperparameters and $^{y_i}_p + \Delta p$, $\theta_n^j + \Delta_n$ are clip to $(0, \pi)$.

Then we could make a analysis on this loss. First, just like circle-loss, θ_n and θ_p obtain weight based on their own value via α. Since O_n and O_p are fixed, the higher value of θ_p, which is more difficult obtain higher weights and lower value of θ_n, is also difficult to obtain more weights. And turn to the easy samples, for positive, the weights are still kept, and for negative samples, if $\theta_n^j > O_n$, their weights will become 0. Then we can see that this loss give a margin on θ instead of similarity, just like the arc-loss, to obtain an additive cosine margin. The decision boundary is achieved at:

$$\gamma(\alpha_p cos(\theta_p + \Delta_p) - \alpha_n cos(\theta_n + \Delta_n)) = 0. \tag{15}$$

What's more, seen from the grad, we take a look at item about θ_n, we assume that $v_n = \alpha_n cos(\theta_n + \Delta_n) = (tm - \theta_n)cos(\theta_n + \pi - tm)$ and $\frac{\partial v_n}{\partial \theta_n} = cos(\theta_n - tm) - (\theta_n - tm)sin(\theta_n - tm) = 0$ so the loss get min value for $cos(\theta_n - tm) - (\theta_n - tm)sin(\theta_n - tm) = 0$, in our hyperparameter setting where $tm = 0.65\pi$, the θ_n is at about 0.38π, and this target can shift base on the value of tm so this loss can focus less on easy negative samples since their grad are smaller.

Besides, the focal loss [24] is used to alleviate the significant imbalance of the proportion of positive and negative samples, which can be formulated as:

$$\mathcal{L}_{focal} = -a_t(1 - p_t)^{\gamma} \log(p_t), \tag{16}$$

$$a_t = \begin{cases} a, if\ y = 1 \\ 1 - a, otherwise \end{cases}, \tag{17}$$

$$p_t = \begin{cases} p, if\ y = 1 \\ 1 - p, otherwise \end{cases}. \tag{18}$$

In inference module, after obtaining features, Eq. 1 is used to calculate cosine similarity scores.

5. Experiments

A workstation with 4*NVIDIA A100 GPUs is used for experiments. We evaluate our method on the MS1M dataset and the given test dataset of MFR21 (ICCV21 masked face recognition challenge). The essential ablation studies are elaborately designed, as well as quantitative evaluations with other contestants.

For implementation details, all face images are resized to 112x112 pixels as the inputs. The number of image pairs or number of image for the training set, validation set, and test set are shown in Table 2. For optimization, we choose the SGD optimizer, where the initial learning rate is set to 0.01 and weight decay is set to 5e-4, momentum parameter is set to 0.9. The model is trained based on pytorch. In addition, we also use Partial-FC to obtain better speed and accuracy. In this challenge, for mask set, TAR (True Accept Rate) is measured on mask-to-nonmask 1:1 protocal, with FAR (False Accept Rate) is less than 0.0001(e-4). For other sets, TAR is measured on all-to-all 1:1 protocal, with FAR less than 0.000001(e-6).

5.1. The Effect of Hyper-parameters

In the training stage, image is taken from masked dataset with a probability of λ. When the parameter λ increases, the TAR@Mask (TAR of mask face) would increase, but the

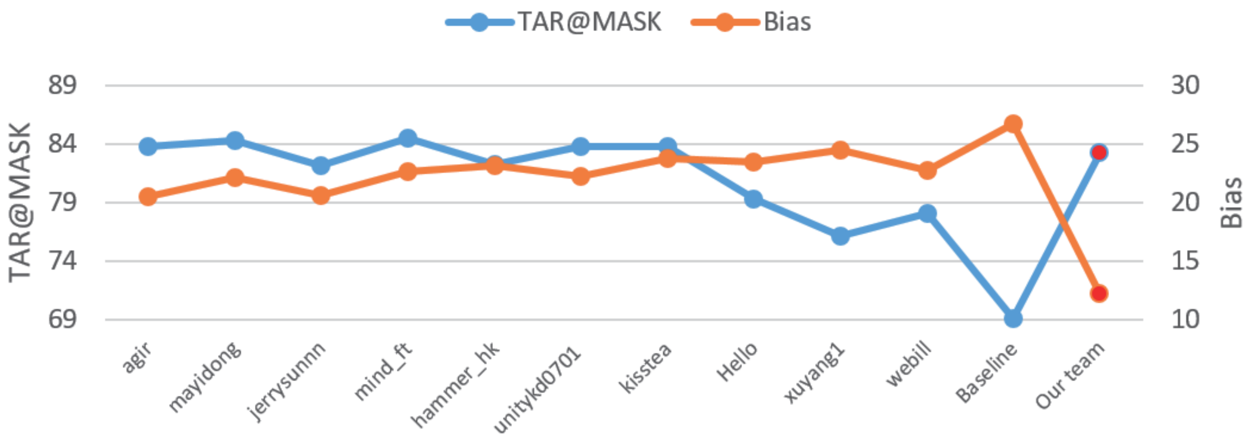

Figure 4: Verification score line chart, including top 10 contestants, baseline and our team in the leaderboard of MS1M track of ICCV21 masked face recognition challenge.

TAR@MR-All (TAR of all face) would decrease. We test the effect of λ in the Table 3, and can find when λ=20%, the final score is best. So we choose λ=20% as shown in Table 3.

Table 3: The effect of hyper-parameter λ.

λ	TAR@MASK	TAR@MR-Alls
0	69.1	84.3
10	73.2	83
15	76.3	82.5
20	**80.0**	**82.1**
25	80.5	80.5
30	81.1	79.3

Besides, the weight rate of re-sampling also has important influence, we compare the result of original sampling method and our sampling method Caucasian: Asian: Indian: African=0.05:1:1:0.2, the detailed information is given in Table 4. It shows that the TAR of East Asian would increase sharply, the bias would decrease using our re-sampling method. Besides, the TAR of MR-ALL would increase at the same time. It shows that our re-sampling method is effective.

Table 4: The effect of our re-sampling method.

	Original sampling	**Our sampling**
African	81.1	**82.0**
Caucasian	89.0	**86.1**
South Asian	88.1	**85.2**
East Asian	62.2	**74.0**
MR-All	84.3	**86.5**

5.2. Ablation Study

In this work, we mainly use three technologies, namely mask augmentation method, the data re-sampling approach and the symmetric-arc-loss. They all has different functions to the final model. We do ablation study as Table 5 to validate our ideas and find the best system. Bias is the deviation between Caucasian and East Asian.

5.3. Quantitative Evaluations

We evaluate our model based on two evaluation index: TAR@MASK and Bias. TAR@MASK is the tar score of masked test faces. Bias is the TAR bias between Caucasian and East Asian. Table 5 shows detailed information of different module combination. We can find Mask augmentation would be useful for TAR@MASK, but it would slightly increase the bias at the same time. The re-sampling method could decrease the bias. Asymmetric-arc-loss could increase TAR@MASK and Bias at the same time. Integrating these techniques, we get TAR@MASK=83.3, and bias=12.2%.

Moreover, the leaderboard of MS1M track of ICCV21 masked face recognition challenge is illustrated in Table 6, where we show top 10 contestants and the baseline. For our method, we get a great TAR@MASK score and a lower bias compared with other contestants. The line chart of verification score is drawn in Fig. 4, which obviously shows that the TAR@MASK scores of top contestants are all great, but the bias between the TAR score of Caucasian and the TAR score of East Asian is almost bigger than 20%. Our team get a lower bias of 12.2%, indicating our method is helpful to get a fairer mask face recognition result.

To further validate our approach, the method is also tested in glint360k track of ICCV21 masked face recognition challenge. Based on baseline, our method uses the same technologies in MS1M track of ICCV21 masked face

Table 5: Ablation study of our method.

Mask augmentation	Re-sample	Asymmetric-arc-loss	TAR@MASK	Bias
			69.1	27.2
✓			80.0	28.5
	✓		70.2	14.0
		✓	73.1	25.2
✓	✓		79.1	15.4
✓	✓	✓	83.3	12.2

Table 6: Leaderboard of MS1M track of ICCV21 masked face recognition challenge in the test stage.

Team	Mask	African	Caucasian	South Asian	East Asian	Bias
agir	83.8	87.8	93.0	93.2	72.5	20.5
mayidong	84.3	86.3	92.2	91.2	70.0	22.2
jerrysunnn	82.2	85.4	92.1	91.3	71.5	20.6
mind_ft	84.5	86.8	92.3	88.3	69.6	22.7
hammer_hk	82.3	85.0	91.7	90.5	68.5	23.2
unitykd0701	83.8	84.1	91.1	89.2	68.8	22.3
kisstea	83.8	83.8	90.9	90.0	67.1	23.8
Hello	79.3	86.0	92.2	92.6	68.7	23.5
xuyang1	76.2	88.0	93.3	92.6	68.8	24.5
webill	78.1	86.0	92.1	91.2	69.3	22.8
Baseline	69.1	81.1	89.0	88.1	62.2	26.8
Our team	83.8	84.3	86.6	85.4	74.4	12.2

Table 7: Leaderboard of glint360k track of ICCV21 masked face recognition challenge in the test stage.

Team	Mask	African	Caucasian	South Asian	East Asian	Bias
jerrysunnn	86.4	92.2	95.8	94.8	76.1	20.5
mayidong	86.0	92.6	96.2	94.6	75.6	22.2
derron	84.3	92.0	95.5	95.1	77.2	20.6
DongWang	83.5	92.2	95.7	94.8	76.4	22.7
didujustfart	89.1	90.4	95.2	93.6	72.6	23.2
yossi_avram	81.5	93.0	96.2	95.7	76.6	22.3
deepcam	84.4	90.9	94.6	94.2	75.8	23.8
helloface	87.3	88.9	93.8	92.0	73.2	23.5
suanying	77.6	92.8	96.3	95.5	77.6	24.5
sgglink	84.5	90.2	94.9	93.4	72.7	22.8
Baseline	75.6	89.5	94.3	93.4	72.5	26.8
Our team	84.5	91.2	92.3	91.7	78.0	14.3

recognition challenge. The leaderboard is illustrated in Table 7, where we show top 10 contestants and the baseline. We obtained great TAR@MASK score and a lower bias compared with other top contestants. The result proves the effectiveness of our method again.

6. Conclusion

This paper proposes a face recognition system with fairness and robustness. The highlights are the combination of the mask augmentation method, the data re-sampling approach and the symmetric-arc-loss. Integrating these techniques, this paper obtained fairer face recognition results and better results on masked faces.

7. Acknowledgement

This work was supported by the Anhui Province Key Research and Development Program (202104a05020007) and USTC Research Funds of the Double First-Class Initiative (YD2350002001)

References

[1] Mohsan Alvi, Andrew Zisserman, and Christoffer Nellåker. Turning a blind eye: Explicit removal of biases and variation from deep neural network embeddings. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 0–0, 2018.

[2] Alexander Amini, Ava P Soleimany, Wilko Schwarting, Sangeeta N Bhatia, and Daniela Rus. Uncovering and mitigating algorithmic bias through learned latent structure. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, pages 289–295, 2019.

[3] Aqeel Anwar and Arijit Raychowdhury. Masked face recognition for secure authentication. *arXiv preprint arXiv:2008.11104*, 2020.

[4] Joy Buolamwini and Timnit Gebru. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency*, pages 77–91, 2018.

[5] Adnane Cabani, Karim Hammoudi, Halim Benhabiles, and Mahmoud Melkemi. Maskedface-net–a dataset of correctly/incorrectly masked face images in the context of covid-19. *Smart Health*, 19:100144, 2021.

[6] Qiong Cao, Li Shen, Weidi Xie, Omkar M Parkhi, and Andrew Zisserman. Vggface2: A dataset for recognising faces across pose and age. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, pages 67–74. IEEE, 2018.

[7] Ken Chen, Yichao Wu, Haoyu Qin, Ding Liang, Xuebo Liu, and Junjie Yan. R3 adversarial network for cross model face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9868–9876, 2019.

[8] Naser Damer, Jonas Henry Grebe, Cong Chen, Fadi Boutros, Florian Kirchbuchner, and Arjan Kuijper. The effect of wearing a mask on face recognition performance: an exploratory study. In *2020 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–6. IEEE, 2020.

[9] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4690–4699, 2019.

[10] Nizam Ud Din, Kamran Javed, Seho Bae, and Juneho Yi. A novel gan-based network for unmasking of masked face. *IEEE Access*, 8:44276–44287, 2020.

[11] Clare Garvie. *The perpetual line-up: Unregulated police face recognition in America*. Georgetown Law, Center on Privacy & Technology, 2016.

[12] Mengyue Geng, Peixi Peng, Yangru Huang, and Yonghong Tian. Masked face recognition with generative data augmentation and domain constrained ranking. In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 2246–2254, 2020.

[13] Patrick J Grother, Mei L Ngan, and Kayee K Hanaoka. Face recognition vendor test part 3: Demographic effects. 2019.

[14] Yandong Guo, Lei Zhang, Yuxiao Hu, Xiaodong He, and Jianfeng Gao. Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In *European conference on computer vision*, pages 87–102. Springer, 2016.

[15] Chunrui Han, Shiguang Shan, Meina Kan, Shuzhe Wu, and Xilin Chen. Face recognition with contrastive convolution. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 118–134, 2018.

[16] Dongyoon Han, Jiwhan Kim, and Junmo Kim. Deep pyramidal residual networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5927–5935, 2017.

[17] Walid Hariri. Efficient masked face recognition method during the covid-19 pandemic. *arXiv preprint arXiv:2105.03026*, 2021.

[18] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[19] Gary B Huang, Marwan Mattar, Tamara Berg, and Eric Learned-Miller. Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. 2008.

[20] Nathan D Kalka, Brianna Maze, James A Duncan, Kevin OConnor, Stephen Elliott, Kaleb Hebert, Julia Bryan, and Anil K Jain. Ijb–s: Iarpa janus surveillance video benchmark. In *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–9. IEEE, 2018.

[21] Brendan F Klare, Ben Klein, Emma Taborsky, Austin Blanton, Jordan Cheney, Kristen Allen, Patrick Grother, Alan Mah, and Anil K Jain. Pushing the frontiers of unconstrained face detection and recognition: Iarpa janus benchmark a. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1931–1939, 2015.

[22] Langford Lane. Nist finds flaws in facial checks on people with covid masks. *Biometric Technology Today*, 2020.

[23] Chenyu Li, Shiming Ge, Daichi Zhang, and Jia Li. Look through masks: towards masked face recognition with de-occlusion distillation. In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 3016–3024, 2020.

[24] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017.

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

[26] Yi-Hung Liu and Yen-Ting Chen. Face recognition using total margin-based adaptive fuzzy support vector machines. *IEEE Transactions on Neural Networks*, 18(1):178–192, 2007.

[27] Steve Lohr. Facial recognition is accurate, if youre a white guy. *New York Times*, 9, 2018.

[28] Ningning Ma, Xiangyu Zhang, Hai-Tao Zheng, and Jian Sun. Shufflenet v2: Practical guidelines for efficient cnn architec-

ture design. In *Proceedings of the European conference on computer vision (ECCV)*, pages 116–131, 2018.

[29] Mike Orcutt. Are face recognition systems accurate? depends on your race. *MIT Technology Review 2016*, 2016.

[30] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4510–4520, 2018.

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

[32] Yifan Sun, Changmao Cheng, Yuhan Zhang, Chi Zhang, Liang Zheng, Zhongdao Wang, and Yichen Wei. Circle loss: A unified perspective of pair similarity optimization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6398–6407, 2020.

[33] Mingxing Tan and Quoc V Le. Efficientnet: Rethinking model scaling for convolutional neural networks. *arXiv preprint arXiv:1905.11946*, 2019.

[34] Hao Wang, Yitong Wang, Zheng Zhou, Xing Ji, Dihong Gong, Jingchao Zhou, Zhifeng Li, and Wei Liu. Cosface: Large margin cosine loss for deep face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5265–5274, 2018.

[35] Mei Wang and Weihong Deng. Mitigating bias in face recognition using skewness-aware reinforcement learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9322–9331, 2020.

[36] Mei Wang, Weihong Deng, Jiani Hu, Xunqiang Tao, and Yaohai Huang. Racial faces in the wild: Reducing racial bias by information maximization adaptation network. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 692–702, 2019.

[37] Pingyu Wang, Fei Su, Zhicheng Zhao, Yandong Guo, Yanyun Zhao, and Bojin Zhuang. Deep class-skewed learning for face recognition. *Neurocomputing*, 363:35–45, 2019.

[38] Zhongyuan Wang, Guangcheng Wang, Baojin Huang, Zhangyang Xiong, Qi Hong, Hao Wu, Peng Yi, Kui Jiang, Nanxi Wang, Yingjiao Pei, et al. Masked face recognition dataset and application. *arXiv preprint arXiv:2003.09093*, 2020.

[39] Zhaohui Yang, Yunhe Wang, Chuanjian Liu, Hanting Chen, Chunjing Xu, Boxin Shi, Chao Xu, and Chang Xu. Legonet: Efficient convolutional neural networks with lego filters. In *International Conference on Machine Learning*, pages 7005–7014, 2019.

[40] Dong Yi, Zhen Lei, Shengcai Liao, and Stan Z Li. Learning face representation from scratch. *arXiv preprint arXiv:1411.7923*, 2014.

[41] Xi Yin, Xiang Yu, Kihyuk Sohn, Xiaoming Liu, and Manmohan Chandraker. Feature transfer learning for face recognition with under-represented data. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5704–5713, 2019.

[42] Xiao Zhang, Zhiyuan Fang, Yandong Wen, Zhifeng Li, and Yu Qiao. Range loss for deep face recognition with long-tailed training data. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 5409–5418, 2017.

[43] Yin Zhang and Zhi-Hua Zhou. Cost-sensitive face recognition. *IEEE transactions on pattern analysis and machine intelligence*, 32(10):1758–1769, 2009.