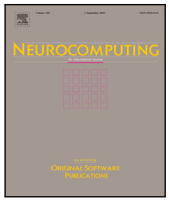
Neurocomputing 600 (2024) 128129

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom



Ethics-aware face recognition aided by synthetic face images

Xiaobiao Du [a], Xin Yu [b], Jinhui Liu [c], Beifen Dai [d], Feng Xu [c,*]

[a] Faculty of Engineering and Information Technology, University of Technology Sydney, 15 Broadway, Sydney, 2007, Australia
[b] School of Information Technology and Electrical Engineering, University of Queensland, 308 Queen, Brisbane, 4072, Australia
[c] School of Software and BNRist, Tsinghua University, 30 Shuangqing Rd, Beijing, 100084, China
[d] Institute for Advanced Studies in Humanities and Social Sciences, Beihang University, 37 Xueyuan Rd, Beijing, 100083, China

ARTICLE INFO

Communicated by H. Yu

Keywords:
Ethics-aware face recognition
Face synthesis
Privacy preservation

ABSTRACT

Current face recognition models trained on large-scale face datasets have achieved promising performance. However, using face images to train a face recognition model without consent would lead to severe privacy and ethical issues. Moreover, existing face recognition models also exhibit uneven performance on different races, thus perplexing vulnerable populations. To address the aforementioned two issues, this work investigates an ethics-aware face recognition method and examines whether we can leverage synthesized faces to achieve a high-accuracy racial balanced recognition model. In a nutshell, we introduce a race-controllable and identity-innumerable face synthesis approach to generate synthetic face images, and then employ the synthesized images to improve face recognition accuracy and mitigate recognition imbalance among different races despite the scarcity of consenting images (less than 100 individuals). More importantly, the synthetic data enable us to analyze the potential impacts of races on face recognition models quantitatively and facilitate the eradication of racial imbalance in face recognition. Extensive experiments demonstrate that employing our synthetic face data improves face recognition accuracy by a large margin while mitigating the recognition imbalance across different race groups.

1. Introduction

Face recognition has attracted great attention in recent years due to its wide applications in various domains, such as security, forensics and community safety, as well as its close relationship with biological representations and cognitive characteristics [1–3]. Benefiting from deep convolutional neural networks, face recognition techniques have achieved remarkable accuracy in recent years. Recent deep face recognition models often require large-scale face image datasets for training [4–9], and the training face images are generally source-crowded from the Internet without the acquisition of individuals' permission. The collected data not only include personal biometric information but also have been distributed without the owners' approval, thus leading to severe privacy leakage and ethical concerns. Due to the increased ethical awareness of the research community, many large-scale face image datasets have been taken down, such as MS-Celeb-1M [10] and MegaFace [11]. As a result, researchers can only leverage a small number of face images, which are relatively easy to be approved for usage, to develop face recognition models. Because of the scarcity of training data, face recognition performance has been dramatically limited.

In addition, it is notorious that face recognition models perform unequally across different races, resulting in racial discrimination issues [12]. As reported in [13], the error rates of recognizing non-Caucasians tend to be higher than identifying Caucasians for 106 face recognition models. Previous studies have investigated eliminating recognition gaps by incorporating more under-performance race data and designing race-aware recognition algorithms [14,15]. However, collecting sufficient real face images of a specific race might breach privacy policies. Although several works [16–19] have been proposed to alleviate recognition bias by modifying the network architectures or introducing new loss functions, they still rely on large-scale face datasets. Furthermore, when the amount of training data is scarce, current face recognition methods will suffer not only drastic performance degradation but also severer recognition bias. As a consequence, race-aware face recognition development faces the dilemma of insufficient data and ethical issues of data collection.

Recently, DigiFace-1M [20] has been proposed to address the privacy and ethical concerns of facial images via computer graphical rendering techniques. They train their model on the unrealistic facial features that would generate a domain gap. Thus their model cannot

* Corresponding author.
E-mail addresses: xiaobiao.du@student.uts.edu.au (X. Du), xin.yu@uq.edu.au (X. Yu), ljh_cs_polaris@163.com (J. Liu), daibeifen@buaa.edu.cn (B. Dai), feng-xu@tsinghua.edu.cn (F. Xu).

https://doi.org/10.1016/j.neucom.2024.128129
Received 26 May 2023; Received in revised form 15 April 2024; Accepted 29 June 2024
Available online 6 July 2024

generalize well to real-world images. In addition, some works [21–25] exploit facial attribute transferring techniques, such as Cycle-GAN [26] or pre-trained attribute networks, to enrich the number of under-performance racial images. Although under-represented racial images have been largely synthesized, the variety of identities is still limited and thus the performance improvements seem marginal. Furthermore, tackling racial imbalanced recognition performance in the data-scarce scenario is even more challenging since there are not sufficient balanced racial images and identities to achieve satisfactory face recognition performance. In this work, we aim to showcase how synthetic face images can greatly improve face recognition performance and mitigate the unbalance performance of the face recognition models on different races when only a few consented face images are available.

To this end, we introduce a racial de-biased face recognition solution in face identity scarce scenarios. Inspired by the face synthesis works [27–35] and facial attribute analysis works [36–38], we intend to synthesize race-specified face images with various identities and then employ them to train racial balanced recognition models. Before learning such a race-controllable and identity-diversified synthesis network, we will disentangle the race from the identity representation of a face image. Since race and identity features always couple together in real images, this entanglement would restrict the diversity of generated face images. The generated images with limited identities might prevent the attainment of balanced high face recognition performance.

In order to separate race and identity information from images, we design a facial race disentanglement network (FRD). Our FRD is composed of an encoder and a decoder. Here, we adopt the 3D morphable model (3DMM) [39] to represent face images and its coefficient is employed as the identity representation. Then, our FRD encoder is designed to explicitly decompose an identity representation into a race indicator and a race-irrelevant representation. Afterwards, FRD decoder reconstructs the original face representation (*i.e.*, 3DMM coefficients) from the race-irrelevant latent codes and the ground-truth race. Note that the ground-truth races can be either obtained through a pre-trained attribute classifier [38] automatically or labeled manually, and thus the attainment process of race information does not need the identity information of face images.

Once we successfully decouple the race from the identity representation, we develop a race-controllable face synthesis network (RCFS) to generate realistic face images with a specified race in various conditions, such as various poses and expressions. Specifically, we firstly leverage the decoder of FRD to produce a face representation from a specified race and a randomly sampled race-irrelevant representation that encodes identity information. Then, our RCFS takes as input the generated face representation and additional controllable facial attributes, such as head poses and expressions, to generate diverse face images. Our proposed RCFS can be utilized to synthesize images with desired race ratios and then use them along with a limited number of real images to train face recognition models. In this fashion, we improve the overall recognition accuracy and significantly alleviate racial unbalanced recognition performance on different race groups especially in face identity data scarce scenarios.

2. Related work

2.1. Deep face recognition

Since the introduction of deep convolution neural networks, face recognition techniques achieve remarkable accuracy in recent years. According to training losses, the current face recognition works can be fallen broadly into two branches.

One branch utilizes metric-learning based losses to train face recognition models, including contrastive loss [40], and triplet loss [41]. However, these methods are usually inefficient for large-scale datasets. The other branch is the usage of the margin-based loss. In contrast to the metric-learning based losses, the margin-based softmax loss achieves effective training, which enhances the intra-class compactness and the inter-class discrepancy to learn more discriminative face representations [42–47]. [42] propose a center loss to enhance intra-class compactness. [43] propose SphereFace that adds an angular margin to the origin softmax loss to achieve intra-class compactness. Furthermore, several works exploit the angular margin-based loss for face recognition, such as CosFace [44], ArcFace [45] and CurricularFace [48]. [49] design a face sketch recognition method that encodes the context of facial images with discriminative information. [50] propose a new coupled attribute learning to address the different modalities of faces for heterogeneous face recognition. These methods achieve state-of-the-art performance on various benchmarks. Unfortunately, these methods all require large-scale face datasets for training. Once these methods meet with limited face training data, the performance will suffer a significant drop. There are some works [51–54] that address the synthetic face dataset, but they do not focus on the unbalance of racial distribution.

2.2. Race-aware face recognition

Racial unbalance performance of face recognition mainly comes from two aspects, *i.e.*, training data, and algorithms. Thus previous studies on this topic can be categorized into two branches [17,55–58]. The first branch solves the problem by collecting ethnicity-aware datasets. [56] introduce two ethnicity-aware training datasets, called BUPT-Globalface and BUPT-Balancedface. The BUPT-Globalface dataset is built upon the population ratio of the demographic distribution in the world, and the BUPT-Balancedface strictly balances the number of samples among different races. Wang et al. further construct a Racial Faces in-the-Wild (RFW) testing database to validate the racial recognition deviations of face recognition algorithms [55]. The second branch mitigates unbalanced recognition performance by modifying the network architectures and loss functions. [57] adopt an adversarial learning paradigm and propose a de-biasing adversarial network to learn disentangled feature representations for unbiased face recognition. Recently, [17] propose a group adaptive classifier to mitigate recognition bias by employing adaptive convolution kernels and attention mechanisms. [58] develop a novel penalty term, a false positive rate penalty loss, into the softmax loss function to alleviate bias and improve the fairness performance in face recognition. However, these methods mitigate recognition bias by leveraging large-scale datasets. Tackling racial unbalance recognition in the data-scarce scenario still remains a problem.

2.3. Face synthesis

Due to the great success of generative adversarial networks (GAN) [59–68], a range of face synthesis techniques have been developed and achieved remarkable progress. Recently, attribute-preserving face synthesis becomes quite popular [69–73]. Specifically, [73] first disentangle the identity attribute from the face image for identity-preserving face synthesis. FaceID-GAN [71] generates identity-preserving faces by introducing an identity classifier, which distinguishes identities of real and synthesized faces. Recent works take a step further and manage to control more attributes of face images, such as poses, expressions and illuminations [69,74,75]. [69] propose to generate synthetic faces in a disentangled manner with multiple disentangled latent spaces that characterize different perspectives of a face image. They introduce an imitative-contrastive learning paradigm with 3D priors for disentangled face generation, which enables precise control of identity, pose, expression, and illumination. [29] propose StyleGAN2, which integrates a new generator normalization method and confines the generator to map latent codes to images in good conditioning for unconditional image modeling. Different from the previous methods using GAN and 3D models to generate synthetic datasets, [76] propose DCface based on the diffusion-based model. [77] propose a heterogeneous representation method to learn plain and interpretable representation for face recognition and face synthesis tasks. However, to the best of our knowledge, race-controllable face synthesis is not solved by previous works.

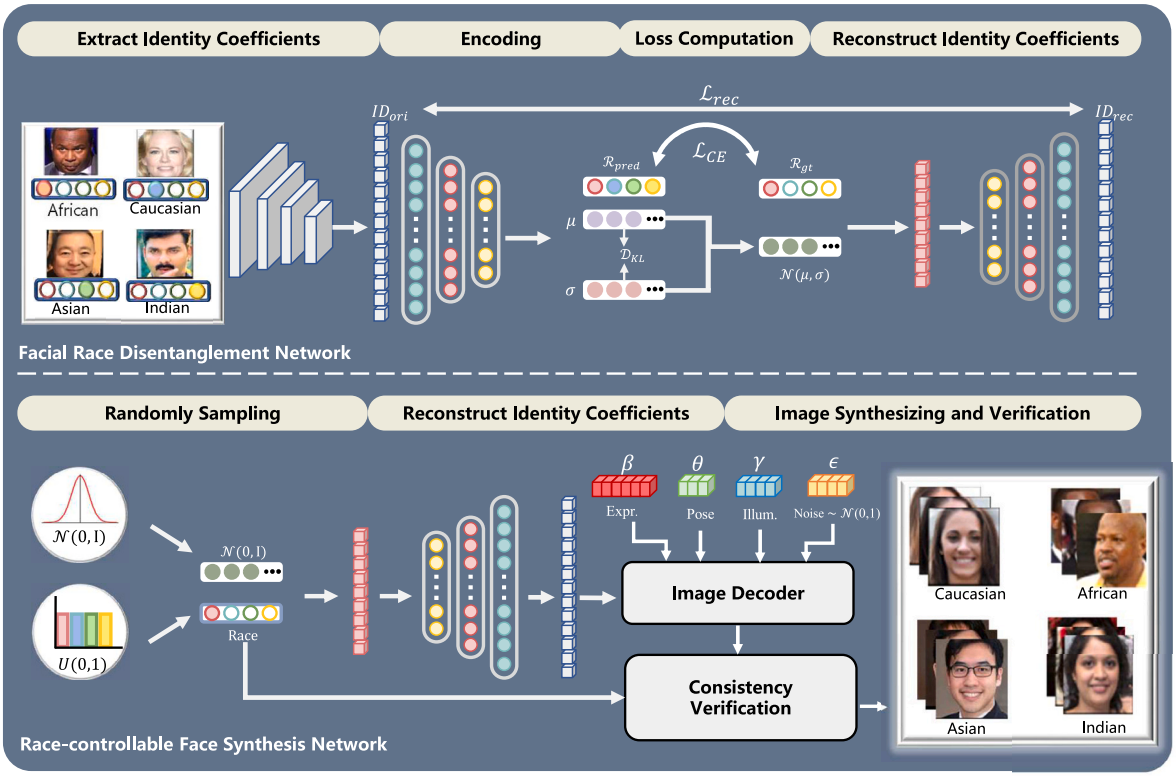


Fig. 1. An overview of our proposed method. It contains a facial race disentanglement network (FRD) and a race-controllable face synthesis network (RCFS). We enforce FRD to explicitly disentangle identity coefficients into a 4-dimensional race vector $\mathcal{R}_{pred}$ and a race-irrelevant vector following a normal distribution parameterized by the mean $\boldsymbol{\mu}$ and the variance $\boldsymbol{\sigma}$. Once race can be explicitly decoupled, we unitize the decoder of FRD to generate race-specified identities by randomly sampling, and feed them into RCFS to produce race-specified synthesized face images. To further guarantee the correctness of the race-specified images, we employ a race discriminator as a race filter that removes incorrect face images.

3. Identity-diversified and race-controllable face synthesis method

In this section, we introduce a facial race disentanglement network (FRD) to decouple the race and race-relevant information, which allows us to generate synthetic face images with free control of both races and identities at a later stage. Next, a race-controllable face synthesis network (RCFS) is developed to generate authentic face images with a specified race. To increase the diversity of synthesized face images, we further sample new identities and facial variations, such as poses, expressions and illuminations, and then feed them to RCFS. In this manner, our RCFS is able to synthesize diverse racial faces with various identities and facial variations. At last, the synthesized face images will be employed to train a racial de-biased face recognition network.

3.1. Facial race disentanglement network

Similar to previous identity-preserving face generators, we also parameterize a face image with a 3DMM model [39]. 3DMM exploits a 160-dimensional coefficient vector to represent an identity, known as identity coefficients. To achieve race and identity disentanglement, we design a facial race disentanglement network (FRD) that can disentangle race information from the identity representations of face images. As illustrated in Fig. 1, FRD is composed of an encoder and a decoder. The encoder learns to decompose identity coefficients into a race vector and a race-irrelevant representation, while the decoder reconstructs the original identity representation from the decoupled race-irrelevant representation and the ground-truth race. Race annotations can be obtained from the race-categorized dataset. In other words, the decoder enforces the encoder to disentangle the race and race-irrelevant information from face identity representations. Here, the ground-truth races of face images can be either obtained by a pre-trained race classifier [38] or manually labeled. In practice, we only assign race labels with high-confidence to face images, and they are deemed to be high-quality labels. Although manually labeling race groups is preferred, the identity information of face images will not be released to annotators. Thus, the process of race information attainment is ethics-aware.

To enable our method to synthesize diverse identity representations, we construct FRD based on the theory of variational autoencoder [78]. Specifically, in the training phase, FRD encodes the identity coefficients into two parts: a predicted race vector $R_{pred} \in \mathbb{R}^4$, and a latent race-irrelevant vector, which follows a Gaussian distribution represented by a mean vector $\boldsymbol{\mu} \in \mathbb{R}^{160}$ and a variance vector $\boldsymbol{\sigma} \in \mathbb{R}^{160}$. Then, the ground-truth race vector R_{gt} and a latent vector sampled from the Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\sigma})$ are fed into the FRD decoder to reconstruct the original identity coefficients. In this way, FRD explicitly learns to disentangle race from identity coefficients. The overall training loss function for FRD is formulated as follows:

$$\mathcal{L}_{FRD} = \lambda_c \mathcal{L}_{CE} + \lambda_k \mathcal{D}_{KL} + \lambda_r \mathcal{L}_{rec}, \tag{1}$$

where the cross entropy loss $\mathcal{L}_{CE}$ is employed to minimize the difference between R_{pred} and R_{gt}. The mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\sigma}$ are enforced to follow a normal distribution by a Kullback–Leibler (KL) divergence loss $\mathcal{D}_{KL}$. The reconstruction loss $\mathcal{L}_{rec}$ is formulated as the Euclidean distance between the reconstructed identity coefficients and its original counterparts. These three loss functions are expressed by:

$$\mathcal{L}_{CE} = -\sum_{i=1}^{C} R_{gt}^{i} \log \frac{\exp\left(R_{pred}^{i}\right)}{\sum_{j=1}^{C} \exp\left(R_{pred}^{j}\right)}, \tag{2}$$

$$\begin{aligned} \mathcal{D}_{KL} &= KL(\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}) \mid \mathcal{N}(\mathbf{0}, \mathbf{I})) \\ &= \frac{1}{2} \sum_{i=1}^{160} (\sigma_i^2 + \mu_i^2 - 1 - \ln \sigma_i^2), \end{aligned} \tag{3}$$

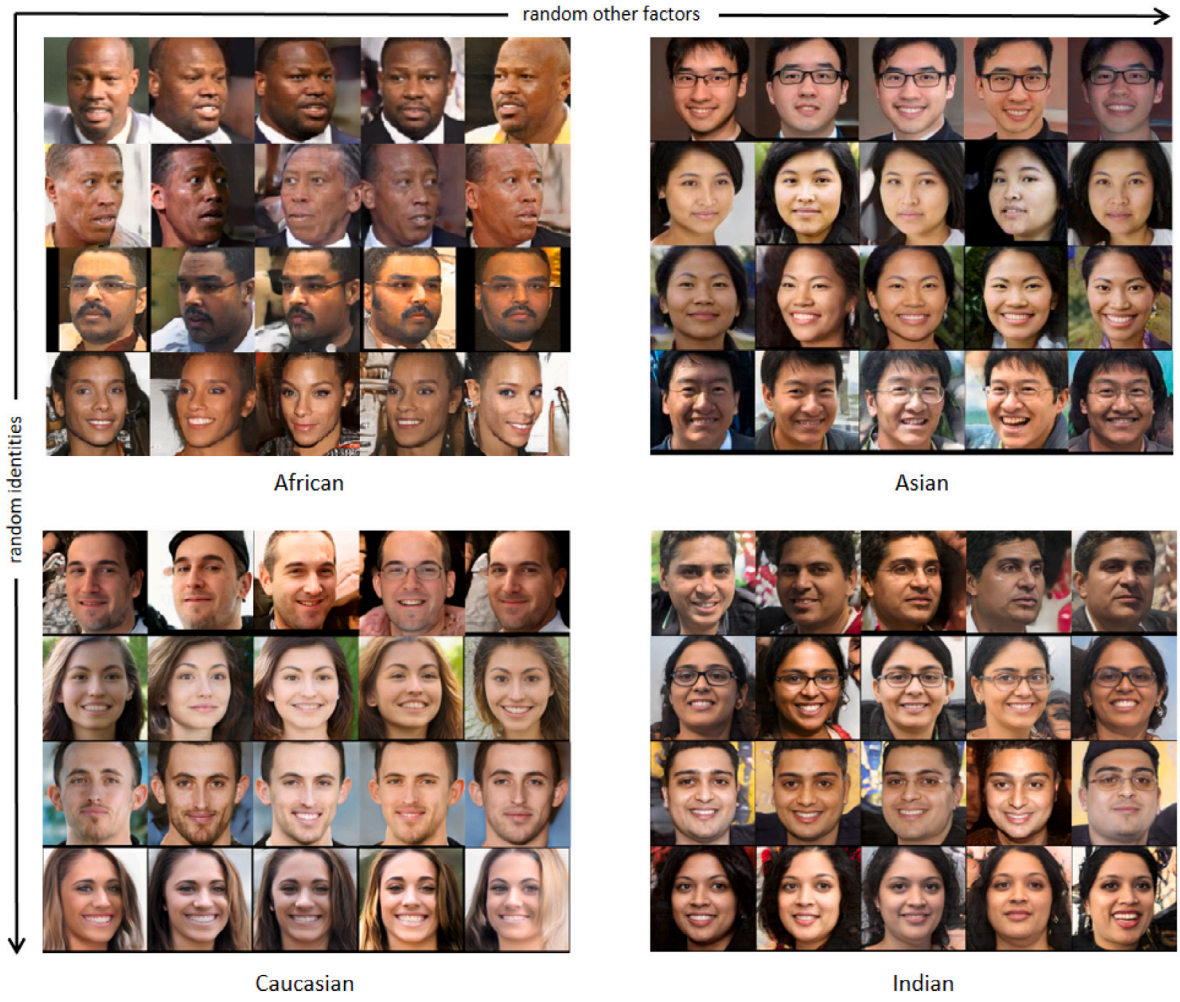


Fig. 2. Face images generated by our race-controllable face synthesis network (RCFS). We can control not only the identity and race of the generated images but also their expressions and poses.

$$\mathcal{L}_{rec} = \|ID_{rec} - ID_{ori}\|_2, \tag{4}$$

where C represents the number of race groups, R_{pred} is a C-dimensional probability vector and R_{gt} is represented by one-hot category vector. ID_{ori} and ID_{rec} are the original and reconstructed identity coefficients. Here, the original or ground-truth identity coefficients are obtained from a 3D face reconstruction network [79]. For each race group, we collect 100K face images regardless of identity information and employ the 3D face reconstruction network to obtain the identity coefficients of all the training images. Note that, in this process we do not need to know the real-world identity information of the persons, but estimate a kind of identity codes in the digital world.

Once the FRD has been trained, the encoder of FRD is discarded. We only employ the FRD decoder to produce identity representations by taking as input a specified race vector and a race-irrelevant vector sampled from a normal distribution. For instance, our FRD can be applied to the pre-trained face generator DiscoFaceGAN [27] to control the identity and race of a generated face image without the need of modifying its generator.

3.2. Race-controllable face synthesis network

To synthesize face images from the race-specified identity coefficients acquired from the FRD as well as other facial variations, we design a race-controllable face synthesis network (RCFS). The architecture of our RCFS is illustrated in Fig. 1.

We generate face images from five different coefficients: identity $\boldsymbol{\alpha} \in \mathbb{R}^{160}$, expression $\boldsymbol{\beta} \in \mathbb{R}^{64}$, pose $\boldsymbol{\theta} \in \mathbb{R}^{3}$, illumination $\boldsymbol{\gamma} \in \mathbb{R}^{27}$ and noise $\boldsymbol{\epsilon} \in \mathbb{R}^{32}$. The last four coefficients are together referred to as the variation coefficients, *i.e.*, $\upsilon \doteq [\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{\epsilon}]$. The identity and variation coefficients are concatenated together as a face representation vector $\boldsymbol{\lambda} \doteq [\boldsymbol{\alpha}, \upsilon]$ and then it is fed into our RCFS to synthesize face images. Here, the facial variations, including expression, pose and illumination, as well as noise are used to improve the diversity of the synthesized faces. In this way, the synthesized faces can resemble faces captured in real scenarios.

Though our RCFS can directly adopt the pre-trained model of the face generator DiscoFaceGAN, the training dataset of DiscoFaceGAN, *i.e.*, the FFHQ [28] dataset, exhibits a large domain gap with the in-the-wild face recognition datasets. Therefore, to minimize the domain gap, we can further fine-tune our RCFS on faces captured in the wild following the training protocols of DiscoFaceGAN. Similar to the training process of FRD, we do not need to know the real-world identities of the faces in training RCFS. Finally, we can generate a large number of synthetic face images while controlling the race and identity information. The samples of our generated face images are shown in Fig. 2. In Fig. 2, we can synthesize numerous identities in different races, and the synthesized faces exhibit different expressions and poses and undergo various lighting conditions, such as shading or specular.

To further guarantee the quality of our synthesized face images in terms of race correctness, we employ the race discriminator (in Section 5) as a race filter. It will remove face images with different

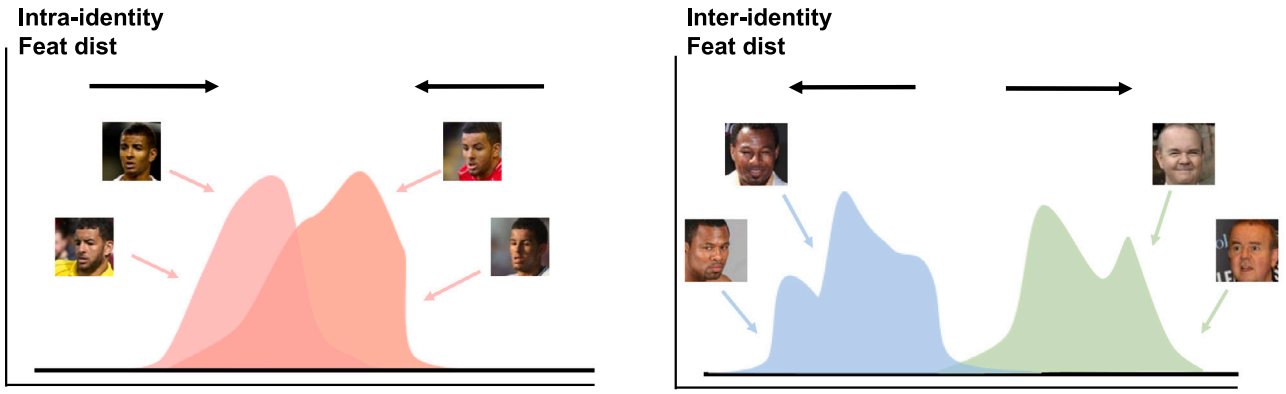


Fig. 3. Illustration of identity consistency verification via K-L divergence and MMD. The colored areas illustrate the distributions of sample faces. If the same identity is used to generate face images, their feature distributions should be similar. Otherwise, their feature distributions should be dissimilar.

race labels but with the same identity coefficients. After this, we can obtain synthetic face images from the same identity representation and with a specified race.

3.3. Identity consistency verification

For a synthetic identity representation, we will synthesize multiple face images of the same identity by changing the facial expression, pose and lighting coefficients and adding noise. Furthermore, we also check the identity consistency of the synthesized multiple images, and ensure that they are suitable for training face recognition.

To this end, we design two verification processes. First, we calculate the intra- and inter-identity distribution divergence of our synthesized face images. The identity is considered consistent if the divergence of the intra-identity distribution is less than that of the inter-identity distribution. This means the generated images from the same identity are close enough while images from different identities lie apart. Though this operation may remove some challenging cases, at this stage we mainly focus on achieving synthesized faces with consistent identities. We measure the distribution divergence by two metrics: Kullback–Leibler (KL) divergence and Maximum Mean Discrepancy (MMD). The definition of these two metrics are expressed as:

$$\mathrm{MMD}[\mathcal{F}, p, q] := \sup_{f \in \mathcal{F}} \left(\mathbf{E}_p[f(x)] - \mathbf{E}_q[f(y)] \right), \tag{5}$$

$$D_{KL}(p \parallel q) = \sum_{i=1}^{n} \left[p\left(x_i\right) \log p\left(x_i\right) - p\left(x_i\right) \log q\left(y_i\right) \right], \tag{6}$$

where $\left(x_1, x_2, \ldots, x_n\right) \sim p(x)$ and $\left(y_1, y_2, \ldots, y_n\right) \sim q(y)$ indicate face images sampled from an identity representation x and y, p and q represent the distribution of x and y respectively, n is the number of sampled images, $\mathcal{F}$ represents the Hilbert Space of a mapping function $f(\cdot)$ and $\mathbf{E}$ indicates Expectation.

We randomly select 1K identities from our synthesized images. For each identity, 20 face images are chosen to verify identity consistency. We utilize an off-the-shelf model ArcFace34 [4] to extract features from these face images and calculate K-L divergence and MMD. For intra-identity divergence, we measure the K-L divergence and MMD of image features from the same identities, while for inter-identity divergence, we calculate the distribution divergence of features across different identities. After the computation, we found the average distribution divergence of inter-identity is obviously larger than that of intra-identity, indicating our synthetic images preserve identity information. As illustrated in Fig. 3, for the intra-identity case, the discrepancy between the sampled feature distributions should be smaller, indicating two faces are more likely from the same identity. On the contrary, the discrepancy of feature distributions of different identities should be larger.

Second, we build a face verification dataset with our synthetic images similar to the LFW [80] dataset and follow the evaluation protocols of LFW. We then test the recognition accuracy of a pre-trained model on our synthetic data. If a pre-trained model achieves accurate face recognition performance, we consider the identities are consistent. Thus, we produce 3K identities and then construct 3K positive pairs and 3K negative pairs. The face verification accuracy is over 97%, demonstrating that our synthesized images successfully preserve identity information.

3.4. Face recognition in data-scarce scenario

We notice that face recognition models exhibit low recognition accuracy and severer racial imbalance performance in the data-scarce scenarios. Therefore, we combine our synthetic images with the limited real images to verify the effectiveness of our synthetic images on improving recognition performance and mitigating racial imbalance.

As aforementioned, we construct a race-aware synthetic dataset via our RCFS network for face recognition in real-world data-scarce scenarios. In training a face recognition model, we select $p \in \{10, 50, 100, 200\}$ persons/identities from each race group to simulate various degrees of data scarcity. The widely used margin-based face recognition loss function is used to train face recognition models. The unified formulation for margin-based objective functions is written as follows:

$$L_{margin} = -\frac{1}{N} \sum_{i=1}^{N} \log \frac{e^{s\left(\cos\left(m_1 \theta_{y_i} + m_2\right) - m_3\right)}}{e^{s\left(\cos\left(m_1 \theta_{y_i} + m_2\right) - m_3\right)} + \sum_{j=1, j \neq y_i}^{n} e^{s \cos \theta_j}}, \tag{7}$$

where y_i indicates the ground-truth label, m_1, m_2 and m_3 respectively indicate the margins of the SphereFace [9], CosFace [5], and ArcFace [4], N is the number of training samples, n indicates the number of identities, θ represents the angle between the feature and normalized weight, and s is a scale factor. When a face recognition method is employed, its corresponding margin is adopted while the other margins are set to 0 in Eq. (7).

In order to train our face recognition models, all the face images are cropped and aligned to 112 × 112 pixels by the state-of-the-art face alignment method img2pose [81]. We use the ResNet-34 [82] as our backbone. The margin m and scale s hyperparameters for different recognition models are set following their original settings. The Stochastic Gradient Descent (SGD) algorithm is used to optimize the neural network with momentum 0.9 and weight decay 5e-4. The batch size is set to 128. The learning rate is initialized to 0.01 and divided by 10 at the 20th, 32th, and 36th epochs. The training process terminates at the 40th epoch.

4. Experiments

In this section, we first introduce the settings of our experiments, including the datasets used for training and testing, and the hyperparameters and the implementation details. Next, we demonstrate that our synthetic data can successfully improve the face recognition accuracy in data-scarce scenarios. Besides, as we can arbitrarily control the race ratio of the synthetic training data, we significantly reduce the bias in recognition accuracy among different races in various scenarios. Finally, to ensure our solution is applicable to different face recognition models, we test the aforementioned two improvements on three recognition models. Notice that all accuracy values reported in this section are the average results of three random experiments since we aim to obtain solid and reliable conclusions from the experiments.

4.1. Experimental setting

4.1.1. Implementation details

All face images are cropped and aligned to 112 × 112 pixels by img2pose [83]. We use the ResNet-34 [84] as our backbone. The margin m and scale s hyperparameters for ArcFace are set as m=0.35 and s=32. The Stochastic Gradient Descent (SGD) algorithm is used to optimize the training procedure with momentum 0.9 and weight decay 5e-4. The batch size is set to 128. The learning rate starts from 0.01 and is divided by 10 at the 20th, 32th and 36th epochs. The training process is finished at the 40th epochs. We conduct all the experiments with the Pytorch framework [85].

4.1.2. Dataset

To mimic data-scarce scenarios, we employ some relatively small number of real face images from the BUPT-Balancedface dataset, which consists of four race groups: African, Asian, Caucasian and Indian, and each race group contains 7000 identities with around 1.2M face images in total. In order to build our race-aware synthetic face dataset, we utilize our RCFS network to synthesize 2000 identities for each race group and generate 50 various face images for each identity.

To evaluate the face recognition models trained on our synthetic data, we adopt two public test datasets, *i.e.*, LFW [80] and RFW [15], and a cleaned High-Quality (HQ) test dataset. The LFW is a race-unaware dataset and is only used for evaluating our method. The RFW dataset consists of four race groups similar to the BUPT-Balancedface dataset. Each race group contains about 10K images of 3K individuals. Since the RFW dataset contains many low-quality face images, they are very challenging for face recognition. Thus, we further clean its testing data and provide a High-Quality (HQ) test dataset for evaluation, where 5K high-quality face images are selected from 1K identities of each race group. The number of positive and negative pairs of HQ dataset is as the same as the RFW dataset.

4.1.3. Evaluation metric

Following the methods [14,15,80], we adopt the face verification metric to evaluate the performance of recognition models. There are various formulations for face verification metrics. One face verification protocol [86,87] is to verify whether a new input image has a corresponding identity within a pre-defined gallery. This metric might require multiple images of an individuals in the gallery to improve the robustness of evaluation, and the gallery set should contain the same identity as the query image. Pair matching [80] is an alternative evaluation metric of face verification. This evaluation protocol focuses on distinguishing whether two images are from the same identity or not. These two images could be never seen before. In this work, we opt to choose pair matching [80] as the evaluation metric since this metric is not restricted by the number of identities.

Specifically, there is a probe set $\mathcal{P}$ and a galley set $\mathcal{G}$. Note that, the identities in $\mathcal{P}$ and $\mathcal{G}$ are different from the training set. We randomly choose one image from the probe set $\mathcal{P}$ and gallery set $\mathcal{G}$, respectively to construct a pair, and our test set consists of 54K pairs in total (*i.e.*, 50% matched pairs and 50% unmatched pairs). Then, we measure the matching score of each pair from $\mathcal{P}$ and $\mathcal{G}$. The verification accuracy $Acc(\tau)$ is computed as follows:

$$Acc(\tau) = \frac{|\{p_i : s_{ij} \geq \tau, ID(g_j) = ID(p_i)\}|}{|\mathcal{P}|} + \frac{|\{p_i : s_{ij} < \tau, ID(g_j) \neq ID(p_i)\}|}{|\mathcal{P}|}, \tag{8}$$

where $p_i \in \mathcal{P}$, $g_j \in \mathcal{G}$, τ denotes a threshold that is used to determine the similarity between p_i and g_j, and ID indicates the identity of a sample. The first and second items represent the True Positive Rate (TPR) and True Negative Rate (TNR), respectively. TPR denotes the rate that a model predicts the matched pairs correctly. TNR represents the rate that a model predicts the unmatched pairs correctly.

Table 1
The ablation study of different margin-based loss functions The training dataset is set as every 100 identities across different races. We report the accuracy of the RFW dataset.

Method	m_1	m_2	m_3	Accuracy
$\cos(\theta)$				60.24
$\cos(m_1\theta)$	✓			61.07
$\cos(\theta + m_2)$		✓		62.14
$\cos(\theta) - m_3$			✓	62.95
$\cos(m_1\theta + m_2) - m_3$	✓	✓	✓	63.97

4.2. Ablation study

In this section, we conduct an ablation study of our margin-based objective function. As shown in Table 1, we train the model on every 100 identities across different races and report the results on the RFW dataset. This ablation study is conducted by modifying different margin-based losses. Here we only display the inner part of the cosine function of Eq. (7) for simplicity. $\cos(\theta)$ does not use any margin-based modification. In this case, it is the standard Cross-Entropy loss. When the loss function is added m_1, it is the SphereFace [9]. When the loss function is added m_2, that is the ArcFace [4]. When we involve $m3$, CosFace [5] would be our loss function. When we put m_1, m_2, and m_3 together, that is our margin-based loss function. Obviously, we can find that our method shows better results than others, which demonstrates that our method can facilitate the model to learn a more discriminative representation across different races.

4.3. Analysis of recognition accuracy improvement

In this section, we quantitatively analyze the effect of our race-aware synthetic technique on improving face recognition performance. We construct five different real face datasets from the BUPT-Balancedface dataset and seven synthetic datasets based on our race-aware synthetic technique. Specifically, to mimic the different data-scarce scenarios, we randomly select 0, 10, 50, 100 and 200 identities for each race group from the BUPT-Balancedface dataset, forming five real face *training* datasets. Besides, we synthesize 0, 50, 100, 200, 500, 1000 and 2000 identities for each race group, forming the seven synthetic training datasets with different data scale. They are designed to analyze the effectiveness and impacts of our synthesized data. Under each data-scarce scenario, we add the seven synthetic datasets to train the ArcFace-34 recognition models.

We evaluate the models on all three test datasets, as shown in Fig. 4. The horizontal axis indicates the number of synthesized identities increases from 0 to 2000. We show the impacts of synthesized identities on recognition accuracy along with different numbers of real identities. For the HQ and RFW datasets, we calculate the average accuracy across the four race groups as results. As the number of synthesized identities increases, we observe that the face recognition performance improves consistently. This demonstrates the effectiveness of our synthesized face images on improving recognition accuracy.

Table 2
Racial imbalance mitigation results on our HQ and RFW datasets.

	training real IDs				training syn IDs							
	Af.	As.	Ca.	In.	Af.	As.	Af.	As.	Ca.	In.	Mean	Std
HQ	100	100	100	100	0	0	72.22	79.34	80.75	81.69	78.50	4.30
	100	100	100	100	500	0	84.55	87.44	89.13	89.42	87.63	2.25 (↓ 47.67%)
	100	100	100	100	1000	0	87.09	88.96	90.39	90.76	89.30	**1.70** (↓ 60.47%)
	0	0	400	0	0	0	70.16	78.61	82.21	81.18	78.04	5.47
	0	0	400	0	500	0	82.99	86.43	90.40	88.94	87.19	3.25 (↓ 40.59%)
	0	0	400	0	1000	0	85.00	87.57	91.55	90.07	88.55	**2.89** (↓ **47.17**%)
	0	0	200	200	0	0	71.70	79.12	80.45	83.67	78.74	5.06
	0	0	200	200	1500	500	85.78	88.73	89.38	90.38	88.57	**1.98** (↓ **60.87**%)
	0	100	100	100	0	0	69.88	77.46	79.33	80.71	76.85	4.83
	0	100	100	100	500	0	81.99	86.57	88.14	88.28	86.25	2.95 (↓ 38.92%)
	0	100	100	100	1000	0	84.45	88.04	89.61	89.79	87.97	**2.48** (↓ **48.65**%)
RFW	100	100	100	100	0	0	58.53	64.95	66.90	65.48	63.97	3.72
	100	100	100	100	500	0	64.68	68.75	72.74	71.15	69.33	3.53 (↓ 5.11%)
	100	100	100	100	1000	0	66.28	70.11	73.81	71.89	70.52	**3.23** (↓ **13.17**%)
	0	0	400	0	0	0	56.58	64.52	67.33	64.38	63.20	4.62
	0	0	400	0	500	0	63.75	68.04	73.77	70.63	69.05	4.26 (↓ 7.79%)
	0	0	400	0	1000	0	65.06	69.15	74.66	71.73	70.15	**4.09** (↓ **11.47**%)
	0	0	200	200	0	0	57.43	63.37	65.53	64.17	62.63	3.58
	0	0	200	200	1500	500	64.37	67.33	72.00	69.32	68.26	**3.22** (↓ **10.06**%)
	0	100	100	100	0	0	57.50	64.01	66.05	64.73	63.07	3.85
	0	100	100	100	500	0	63.24	68.11	71.86	70.01	68.30	3.73 (↓ 3.12%)
	0	100	100	100	1000	0	64.66	69.69	73.10	71.27	69.68	**3.64** (↓ **5.45**%)

Af., As., Ca. and In. respectively indicate African, Asian, Caucasian and Indian.

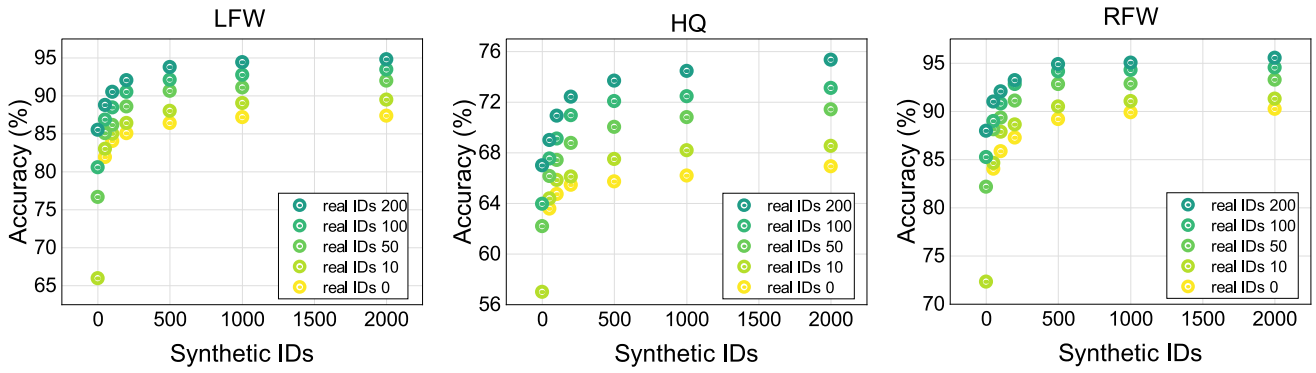


Fig. 4. Face recognition accuracy on LFW, our HQ, and RFW datasets. The Synthetic IDs indicate the number of synthetic identities, which range from 0 to 2000. The real IDs represent the number of real identities.

4.4. Racial imbalance mitigation

In this subsection, we discuss the effectiveness of our synthetic data on mitigating the racial imbalance under the data-scarce scenarios. We use real images to construct four different scenarios: (a) 100 identities from each race group; (b) 400 identities only from Caucasian; (c) 200 identities from Caucasian and 200 identities from Indian; (d) 100 identities per race from Caucasian, Asian and Indian. Overall, we constrain the total number of real people to be 400. The model trained by the real images of each scenario exhibits a noticeable performance gap among different races. Then, we add synthetic images to the race group that yields the lowest recognition accuracy to mitigate the recognition imbalance.

As shown in Table 2, we evaluate the racial imbalance mitigation performance on HQ and RFW datasets because these two datasets contain ground-truth race labels. Following prior works [18], we adopt the standard deviation (std) across the four race groups to indicate the racial imbalance. We can see that in all the four scenarios, the accuracy of African is always lower than the other three groups. Thus we add African synthetic face images to the real datasets in order to mitigate the racial imbalance. In the fourth scenario, we also add 500 identities of Asian synthetic data because we do not have real data from Asian group either. The results show that our synthetic face images can reduce the racial imbalance by over 40% on the HQ dataset and around 10% on the RFW dataset. The results demonstrate that our race-aware synthetic data effectively mitigates the racial imbalance while improving face recognition accuracy.

On the other hand, we can obtain different synthesizing strategies in Table 2 to improve the face recognition performance of different races. As we can see in scenario (b) comparing with others, if we want to enhance the face recognition accuracy of a certain race, we should add more real and synthetic data about that race. If we hope to improve the overall performance across different races, we should add more real data across different races, which is implied by scenario (a). Comparing scenarios (b) and (c) with others, we can infer that though we do not train the model on the specific race, the model still can achieve good results on the unseen race, which attribute to the strong facial representation of face recognition models.

4.5. Performance on other face recognition models

To further demonstrate the effectiveness of our technique on various face recognition models, we adopt two other face recognition models, *i.e.*, CosFace [5] and CurricularFace [6]. We conduct similar

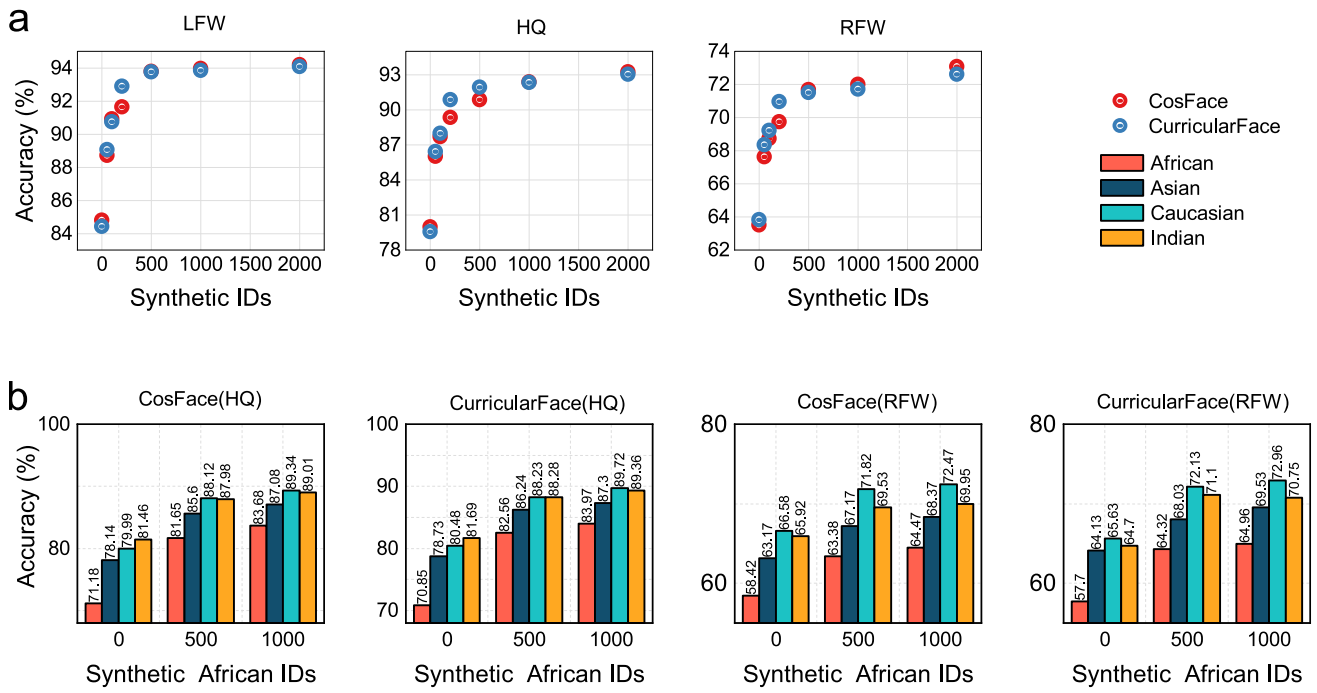


Fig. 5. The recognition accuracy results of other face recognition models, *i.e.*, CosFace and CurricularFace. a. Three datasets were utilized to evaluate the accuracy of these two models across all racial groups. **b.** The racial imbalance mitigation results of these two models on HQ and RFW datasets. The first two columns indicate the results on the HQ dataset. The last two columns illustrate the results on the RFW dataset.

experiments in Section 4.3 and Section 4.4 to evaluate our synthetic data on recognition accuracy improvement and racial imbalance mitigation. We also sample 100 identities from each race group as the real dataset for the accuracy improvement experiment, and then different scale synthetic datasets are added to judge the accuracy improvement. The fourth scenario introduced in Section 4.4 are considered for the racial imbalance mitigation experiment: 100 identities per race from Caucasian, Asian and Indian groups.

We also show the recognition accuracy over 3 test datasets, *i.e.*, LRW, HQ, and RFW. The recognition accuracy improvements are shown in Fig. 5**a**. For both of the face recognition models, *i.e.*, CosFace and the CurricularFace models, the performance improves consistently on each dataset when increasing the number of synthesized identities. In addition, Fig. 5**b** shows the experimental results of the racial imbalance mitigation for CosFace and the CurricularFace models. Similar to the results of ArcFace, the racial imbalance has been significantly mitigated in the HQ and RFW datasets for these two models. According to the results, our method exhibits promising performance in improving recognition accuracy and alleviating racial imbalance for different face recognition models.

5. Further analysis

To verify our hypothesis that race information is entangled in the identity representation extracted by 3DMM, we randomly sample 10K identity coefficients from the standard normal distribution and then generate synthetic face images from these identity coefficients by a pre-trained face generator DiscoFaceGAN [27]. Meanwhile, we employ a pre-trained race discriminator to identify the race group of each identity. We found that among the 10K identities, 50% are Caucasians, 26% are Asians, 14% are Indians and 10% are Africans. We observe that the most frequently generated race is Caucasian and the least generated ones are African and Indian. This is mainly because DiscoFaceGAN is trained on FFHQ [28] dataset which has the largest proportion of Caucasian face images and only a few African and Indian face images. This indicates there is a strong bias in existing face generators. As a result, simply applying a face generator may not help to achieve racial balanced recognition performance. Although we can apply sample faces multiple times and then balance the race among generated faces, the underrepresented racial images may still suffer the limited diversity, such as identities. This phenomenon motivates us to disentangle race from identity representations. Thus, we need control the identity and race independently rather than sampling from the coupled race-identity distribution.

To verify the race-specified face generation performance of our RCFS, we randomly sample 10K identities from each race group and evaluate the race correctness of generated face images of these identities. We adopt the same race discriminator to distinguish the race of each identity. The result shows that our RCFS achieves 100% accuracy on generating Caucasian and Asian, 95.8% and 93.4% on synthesizing African and Indian, respectively. The lower accuracy on African and Indian validates the challenges of generating images of those racial groups. Fortunately, we can obtain enough images with new identities and those identities cover the race groups more evenly. This attributes to our proposed two components: (i) our facial race disentanglement network (FRD) decouples race information from identity representations and thus enables us to choose the race of synthesized faces in generating identity representations; (ii) Our race-controllable face synthesis network (RCFS) effectively synthesizes face images with the control of race and identity information while increasing the diversity of face images by introducing facial variations, such as illumination, expressions and poses.

To further demonstrate the effectiveness of our proposed method, we conduct a quantitative experiment between different image generation models. Here, we introduce 4 state-of-the-art image generation models, Pix2Pix [88], Cycle-GAN [26], StyleGAN2 [29] and DCFace [76]. Note that our proposed model synthesizes the face images of different races by disentangling the race coefficient, while Pix2Pix, Cycle-GAN, StyleGAN2 transfer different racial images via complexion conversion. Fig. 6 illustrates face recognition results on the RFW dataset by ArcFace-34 recognition model. In scenario A, we employ 100 realistic identities from each race (400 identities in total) and then show the face recognition results as the number of synthetic African identities increases. In scenario B, we use 200 realistic identities

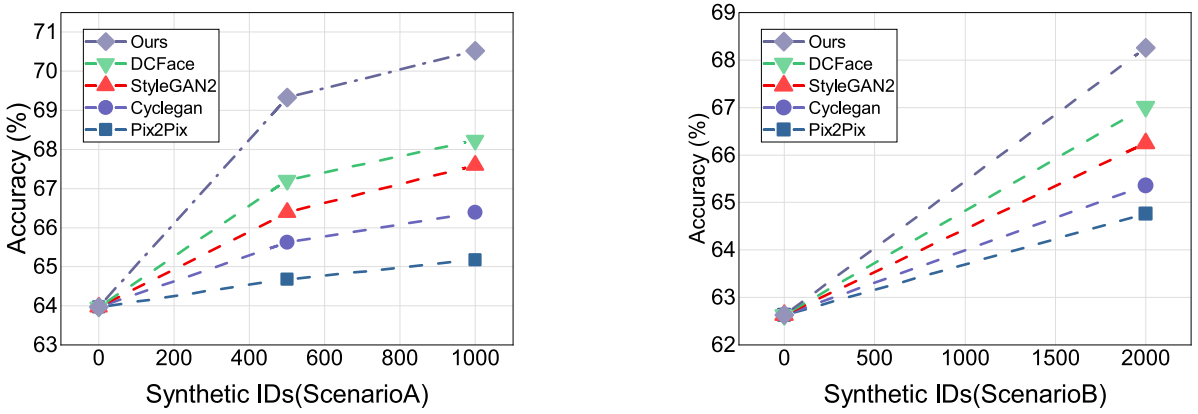


Fig. 6. Face recognition results on the RFW dataset with training face images generated by Pix2Pix, Cycle-GAN, StyleGAN2, DCFace and our method, respectively. Scenario A denotes the results based on 100 real identities of each race and the synthetic African identities from 0 to 1000. Scenario B denotes the results based on 400 real identities (200 Caucasians and 200 Indians) and 0/2000 synthetic identities (1500 Africans and 500 Asians).

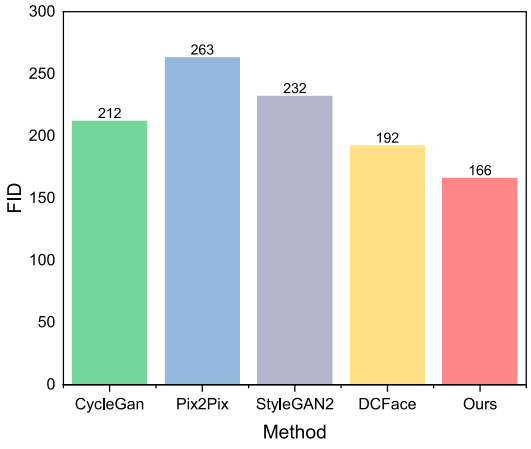


Fig. 7. The FID metric comparison of Cycle-GAN, Pix2Pix, StyleGAN2, DCFace and our method, respectively. The lower FID means better results.

from Caucasian and Indian respectively (also 400 identities in total) and demonstrate the face recognition results with additional 1500 synthetic African identities and 500 synthetic Asian identities. It is obvious that when synthetic identities are employed for training, our method achieves better recognition accuracy than that of Pix2Pix [88], Cycle-GAN [26], StyleGAN2 [29] and DCFace [76]. Thanks to our proposed facial race disentanglement, we can generate authentic face images across different races and the generated faces also achieve richer identity diversity.

In Fig. 7, we show the Fréchet inception distance (FID) [89] results on our method compared with Cycle-GAN, Pix2Pix, StyleGAN2, DCFace. The FID [89] metric describes two distributions how different in feature space. A large FID score means the synthetic image is more fake. As shown in Fig. 7, We use the synthetic images of each model to compute the FID with realistic images. Our method (red column) obtains the lowest FID result, which means our method generates more realistic images than other methods.

In Fig. 8, we compare our synthesized face images with those of Pix2Pix [88], Cycle-GAN [26], StyleGAN2 [29] and DCFace [76] in the feature space via t-SNE [90]. It can be seen that our method synthesizes more realistic face images across different races. In addition, the visualization of t-SNE projection indicates that our method successfully synthesizes different racial images and these images lie closely to the distributions of real images of corresponding races. Cycle-GAN transforms the Caucasian complexion into an African-like complexion but still retains obvious Caucasian facial features. As a result, the distribution of its synthesized images is far away from the latent Caucasian distribution. Pix2Pix struggles to learn a mapping from Caucasian into African as the paired images are not available. Thus, it produces distorted and low-quality face images, which might not be able to alleviate racial imbalance recognition performance. StyleGAN2 hardly transforms Caucasian features into African. ADFace seemingly synthesize realistic African faces but the t-SNE results demonstrate its generated African distribution has a large margin with the realistic one. Therefore, this experiment implies that our method is more effective than the complexion conversion based methods.

6. Conclusion

With the race-aware synthetic data generated by the proposed RCFS, existing face recognition models can be trained with only a few real face images and achieve promising recognition performance. Moreover, we improve face recognition accuracy noticeably and mitigate racial imbalance significantly in data-scarce scenarios with the help of our synthesized face images. The additional synthetic data alleviate ethical issues, such as imbalanced recognition performance against certain race groups and usage of personal images without consent.

CRediT authorship contribution statement

Xiaobiao Du: Writing – review & editing, Writing – original draft, Visualization, Software. **Xin Yu:** Data curation, Conceptualization. **Jinhui Liu:** Resources, Investigation. **Beifen Dai:** Validation, Supervision. **Feng Xu:** Validation, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Feng Xu reports financial support was provided by Tsinghua University. Feng Xu reports a relationship with Tsinghua University that includes: board membership and employment.

Data availability

No data was used for the research described in the article.

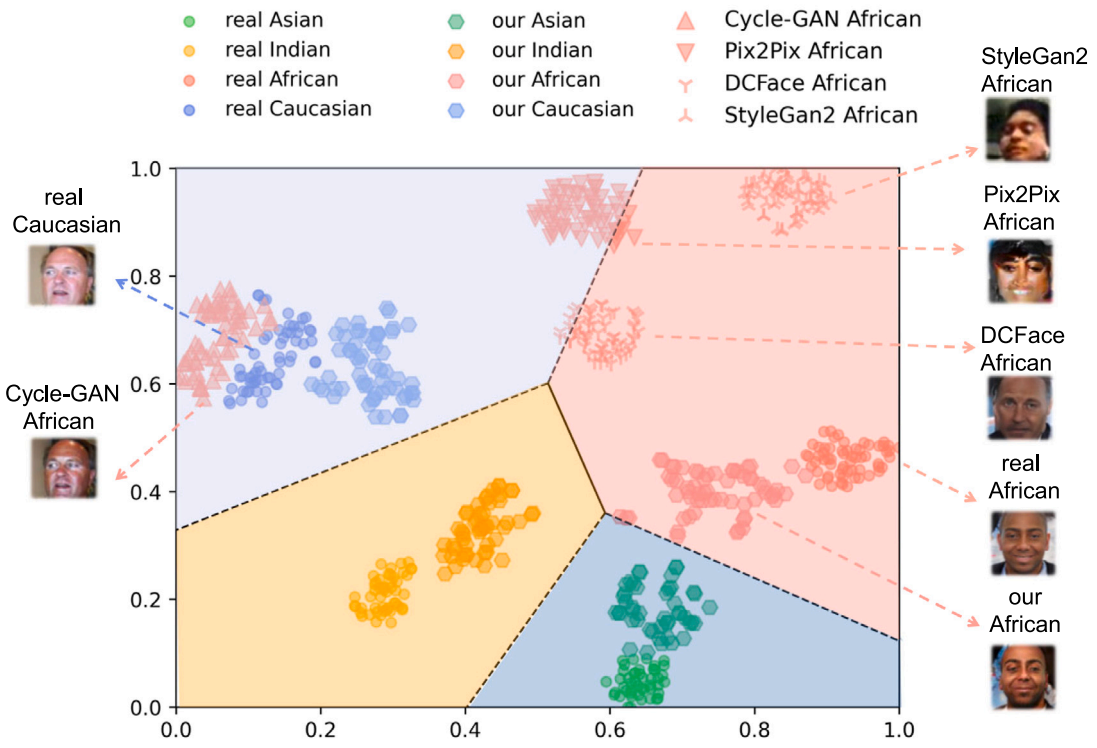


Fig. 8. t-SNE visualization of real face images and synthetic face images generated by Cycle-GAN, Pix2Pix, StyleGAN2, DCFace and our method, respectively.

Acknowledgments

This work is funded by National Key R&D Program of China (2018YFA0704000 and 2023YFC3305600), Australian Research Council (ARC) Discovery Project (DP220100800) and DECRA (DE230100477). It is also supported by the Beijing Natural Science Foundation (M22024), the NSFC (62021002), the the Zhejiang Provincial Natural Science Foundation (LDT23F02024F02) and the Key Research and Development Project of Tibet Autonomous Region (XZ202101ZY0019G). This work was also supported by THUIBCS, Tsinghua University, and BLBCI, Beijing Municipal Education Commission.

References

[1] D. Sero, A. Zaidi, J. Li, J.D. White, T.B.G. Zarzar, M.L. Marazita, S.M. Weinberg, P. Suetens, D. Vandermeulen, J.K. Wagner, et al., Facial recognition from DNA using face-to-DNA classifiers, Nat. Commun. 10 (1) (2019) 2557.

[2] S. Poltoratski, K. Kay, D. Finzi, K. Grill-Spector, Holistic face recognition is an emergent phenomenon of spatial processing in face-selective regions, Nat. Commun. 12 (1) (2021) 4745.

[3] J. Ding, K. Chen, H. Liu, L. Huang, Y. Chen, Y. Lv, Q. Yang, Q. Guo, Z. Han, M.A. Lambon Ralph, A unified neurocognitive model of semantics language social behaviour and face recognition in semantic dementia, Nature Commun. 11 (1) (2020) 2595.

[4] J. Deng, J. Guo, N. Xue, S. Zafeiriou, Arcface: Additive angular margin loss for deep face recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 4690–4699.

[5] H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, Z. Li, W. Liu, Cosface: Large margin cosine loss for deep face recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 5265–5274.

[6] Y. Huang, Y. Wang, Y. Tai, X. Liu, P. Shen, S. Li, J. Li, F. Huang, Curricularface: adaptive curriculum learning loss for deep face recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2020, pp. 5901–5910.

[7] F. Schroff, D. Kalenichenko, J. Philbin, Facenet: A unified embedding for face recognition and clustering, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 815–823.

[8] Q. Meng, S. Zhao, Z. Huang, F. Zhou, Magface: A universal representation for face recognition and quality assessment, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2021, pp. 14225–14234.

[9] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, L. Song, Sphereface: Deep hypersphere embedding for face recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 212–220.

[10] Y. Guo, L. Zhang, Y. Hu, X. He, J. Gao, Ms-celeb-1m: A dataset and benchmark for large-scale face recognition, in: European Conference on Computer Vision, 2016, pp. 87–102.

[11] I. Kemelmacher-Shlizerman, S.M. Seitz, D. Miller, E. Brossard, The megaface benchmark: 1 million faces for recognition at scale, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 4873–4882.

[12] A. Najibi, Racial discrimination in face recognition technology, Harvard Online: Science Policy and Social Justice 24 (2020).

[13] M. Ngan, P. Grother, Face recognition vendor test (FRVT) - performance of automated gender classification algorithms, 2015, http://dx.doi.org/10.6028/NIST.IR.8052.

[14] M. Wang, W. Deng, Mitigate bias in face recognition using skewness-aware reinforcement learning, 2019, arXiv preprint arXiv:1911.10692.

[15] M. Wang, W. Deng, J. Hu, X. Tao, Y. Huang, Racial faces in the wild: Reducing racial bias by information maximization adaptation network, in: Proceedings of the IEEE International Conference on Computer Vision, 2019, pp. 692–702.

[16] S. Gong, X. Liu, A.K. Jain, Jointly de-biasing face recognition and demographic attribute estimation, in: European Conference on Computer Vision, 2020, pp. 330–347.

[17] S. Gong, X. Liu, A.K. Jain, Mitigating face recognition bias via group adaptive classifier, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2021, pp. 3414–3424.

[18] X. Xu, Y. Huang, P. Shen, S. Li, J. Li, F. Huang, Y. Li, Z. Cui, Consistent instance false positive improves fairness in face recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2021, pp. 578–586.

[19] X. Yu, F. Shiri, B. Ghanem, F. Porikli, Can we see more? Joint frontalization and hallucination of unaligned tiny faces, IEEE Trans. Pattern Anal. Mach. Intell. 42 (9) (2019) 2148–2164.

[20] G. Bae, M. de La Gorce, T. Baltrušaitis, C. Hewitt, D. Chen, J. Valentin, R. Cipolla, J. Shen, DigiFace-1M: 1 million digital face images for face recognition, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2023, pp. 3526–3535.

[21] S. Yucer, S. Akçay, N. Al-Moubayed, T.P. Breckon, Exploring racial bias within face recognition via per-subject adversarially-enabled data augmentation, in:

Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 18–19.
[22] J. Joo, K. Kärkkäinen, Gender slopes: Counterfactual fairness for computer vision models by attribute manipulation, in: Proceedings of the International Workshop on Fairness, Accountability, Transparency and Ethics in Multimedia, 2020, pp. 1–5.
[23] M. Georgopoulos, J. Oldfield, M.A. Nicolaou, Y. Panagakis, M. Pantic, Mitigating demographic bias in facial datasets with style-based multi-attribute transfer, Int. J. Comput. Vis. 129 (7) (2021) 2288–2307.
[24] X. Yu, B. Fernando, R. Hartley, F. Porikli, Super-resolving very low-resolution face images with supplementary attributes, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 908–917.
[25] F. Shiri, X. Yu, F. Porikli, R. Hartley, P. Koniusz, Recovering faces from portraits with auxiliary facial attributes, in: 2019 IEEE Winter Conference on Applications of Computer Vision, WACV, IEEE, 2019, pp. 406–415.
[26] J.-Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2223–2232.
[27] Y. Deng, J. Yang, D. Chen, F. Wen, X. Tong, Disentangled and controllable face image generation via 3D imitative-contrastive learning, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2020, pp. 5154–5163.
[28] T. Karras, S. Laine, T. Aila, A style-based generator architecture for generative adversarial networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 4401–4410.
[29] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, T. Aila, Analyzing and improving the image quality of StyleGAN, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2020, pp. 8110–8119.
[30] T. Karras, M. Aittala, S. Laine, E. Härkönen, J. Hellsten, J. Lehtinen, T. Aila, Alias-free generative adversarial networks, Adv. Neural Inf. Process. Syst. 34 (2021) 852–863.
[31] F. Shiri, X. Yu, P. Koniusz, F. Porikli, Face destylization, in: 2017 International Conference on Digital Image Computing: Techniques and Applications, DICTA, IEEE, 2017, pp. 1–8.
[32] F. Shiri, X. Yu, F. Porikli, R. Hartley, P. Koniusz, Identity-preserving face recovery from portraits, in: 2018 IEEE Winter Conference on Applications of Computer Vision, WACV, IEEE, 2018, pp. 102–111.
[33] F. Shiri, X. Yu, F. Porikli, R. Hartley, P. Koniusz, Identity-preserving face recovery from stylized portraits, Int. J. Comput. Vis. 127 (2019) 863–883.
[34] X. Yu, F. Porikli, Imagining the unimaginable faces by deconvolutional networks, IEEE Trans. Image Process. 27 (6) (2018) 2747–2761.
[35] X. Yu, F. Porikli, B. Fernando, R. Hartley, Hallucinating unaligned face images by multiscale transformative discriminative networks, Int. J. Comput. Vis. 128 (2) (2020) 500–526.
[36] A. Toisoul, J. Kossaifi, A. Bulat, G. Tzimiropoulos, M. Pantic, Estimation of continuous valence and arousal levels from faces in naturalistic conditions, Nat. Mach. Intell. 3 (1) (2021) 42–50.
[37] M.Q. Hill, C.J. Parde, C.D. Castillo, Y.I. Colon, R. Ranjan, J.-C. Chen, V. Blanz, A.J. O'Toole, Deep convolutional neural networks in the face of caricature, Nat. Mach. Intell. 1 (11) (2019) 522–529.
[38] X. Yu, B. Fernando, R. Hartley, F. Porikli, Semantic face hallucination: Super-resolving very low-resolution face images with supplementary attributes, IEEE Trans. Pattern Anal. Mach. Intell. 42 (11) (2019) 2926–2943.
[39] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, T. Vetter, A 3D face model for pose and illumination invariant face recognition, in: IEEE International Conference on Advanced Video and Signal Based Surveillance, 2009, pp. 296–301.
[40] S. Chopra, R. Hadsell, Y. LeCun, Learning a similarity metric discriminatively, with application to face verification, in: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR'05, Vol. 1, 2005, pp. 539–546, http://dx.doi.org/10.1109/CVPR.2005.202.
[41] F. Schroff, D. Kalenichenko, J. Philbin, FaceNet: A unified embedding for face recognition and clustering, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2015, http://dx.doi.org/10.1109/cvpr.2015.7298682.
[42] Y. Wen, K. Zhang, Y. Qiao, A discriminative feature learning approach for deep face recognition, in: LNCS, Vol. 9911, 2016, pp. 499–515, http://dx.doi.org/10.1007/978-3-319-46478-7-31.
[43] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, L. Song, SphereFace: Deep hypersphere embedding for face recognition, 2017, http://dx.doi.org/10.48550/ARXIV.1704.08063, URL https://arxiv.org/abs/1704.08063.
[44] H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, Z. Li, W. Liu, CosFace: Large margin cosine loss for deep face recognition, 2018, http://dx.doi.org/10.48550/ARXIV.1801.09414, URL https://arxiv.org/abs/1801.09414.
[45] J. Deng, J. Guo, N. Xue, S. Zafeiriou, ArcFace: Additive angular margin loss for deep face recognition, 2018, http://dx.doi.org/10.48550/ARXIV.1801.07698, URL https://arxiv.org/abs/1801.07698.
[46] Q. Meng, S. Zhao, Z. Huang, F. Zhou, MagFace: A universal representation for face recognition and quality assessment, 2021, http://dx.doi.org/10.48550/ARXIV.2103.06627, URL https://arxiv.org/abs/2103.06627.
[47] X. Yu, F. Porikli, Ultra-resolving face images by discriminative generative networks, in: European Conference on Computer Vision, Springer International Publishing Cham, 2016, pp. 318–333.
[48] Y. Huang, Y. Wang, Y. Tai, X. Liu, P. Shen, S. Li, J. Li, F. Huang, CurricularFace: Adaptive curriculum learning loss for deep face recognition, 2020, http://dx.doi.org/10.48550/ARXIV.2004.00288, URL https://arxiv.org/abs/2004.00288.
[49] D. Liu, X. Gao, N. Wang, C. Peng, J. Li, Iterative local re-ranking with attribute guided synthesis for face sketch recognition, Pattern Recognit. 109 (2021) 107579.
[50] D. Liu, X. Gao, N. Wang, J. Li, C. Peng, Coupled attribute learning for heterogeneous face recognition, IEEE Trans. Neural Netw. Learn. Syst. 31 (11) (2020) 4699–4712.
[51] F. Boutros, V. Struc, J. Fierrez, N. Damer, Synthetic data for face recognition: Current state and future prospects, Image Vis. Comput. (2023) 104688.
[52] P. Melzi, C. Rathgeb, R. Tolosana, R. Vera-Rodriguez, A. Morales, D. Lawatsch, F. Domin, M. Schaubert, Synthetic data for the mitigation of demographic biases in face recognition, in: 2023 IEEE International Joint Conference on Biometrics, IJCB, IEEE, 2023, pp. 1–9.
[53] H. Qiu, B. Yu, D. Gong, Z. Li, W. Liu, D. Tao, Synface: Face recognition with synthetic data, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 10880–10890.
[54] F. Boutros, J.H. Grebe, A. Kuijper, N. Damer, Idiff-face: Synthetic-based face recognition through fizzy identity-conditioned diffusion model, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 19650–19661.
[55] M. Wang, W. Deng, J. Hu, X. Tao, Y. Huang, Racial faces in-the-wild: Reducing racial bias by information maximization adaptation network, 2019, arXiv:1812.00194.
[56] M. Wang, W. Deng, Mitigate bias in face recognition using skewness-aware reinforcement learning, 2019, arXiv:1911.10692.
[57] S. Gong, X. Liu, A.K. Jain, Jointly de-biasing face recognition and demographic attribute estimation, 2020, arXiv:1911.08080.
[58] X. Xu, Y. Huang, P. Shen, S. Li, J. Li, F. Huang, Y. Li, Z. Cui, Consistent instance false positive improves fairness in face recognition, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2021, pp. 578–586.
[59] M. Arjovsky, S. Chintala, L. Bottou, Wasserstein GAN, 2017, arXiv:1701.07875.
[60] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks, 2014, arXiv:1406.2661.
[61] Z. Chen, C. Wang, B. Yuan, D. Tao, Puppeteergan: Arbitrary portrait animation with semantic-aware appearance transformation, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, IEEE Computer Society, Los Alamitos, CA, USA, 2020, pp. 13515–13524, http://dx.doi.org/10.1109/CVPR42600.2020.01353, URL https://doi.ieeecomputersociety.org/10.1109/CVPR42600.2020.01353.
[62] M. Mirza, S. Osindero, Conditional generative adversarial nets, 2014, arXiv:1411.1784.
[63] J. Bao, D. Chen, F. Wen, H. Li, G. Hua, CVAE-GAN: Fine-grained image generation through asymmetric training, 2017, arXiv:1703.10155.
[64] T. Karras, S. Laine, T. Aila, A style-based generator architecture for generative adversarial networks, 2018, http://dx.doi.org/10.48550/ARXIV.1812.04948, URL https://arxiv.org/abs/1812.04948.
[65] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, T. Aila, Analyzing and improving the image quality of StyleGAN, in: Proc. CVPR, 2020.
[66] T. Karras, M. Aittala, S. Laine, E. Härkönen, J. Hellsten, J. Lehtinen, T. Aila, Alias-free generative adversarial networks, in: Proc. NeurIPS, 2021.
[67] X. Yu, F. Porikli, Face hallucination with tiny unaligned images by transformative discriminative neural networks, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 31, (1) 2017.
[68] P. Li, X. Yu, Y. Yang, Super-resolving cross-domain face miniatures by peeking at one-shot exemplar, in: Proceedings of IEEE International Conference on Computer Vision, 2021.
[69] Y. Deng, J. Yang, D. Chen, F. Wen, X. Tong, Disentangled and controllable face image generation via 3D imitative-contrastive learning, in: IEEE Computer Vision and Pattern Recognition, 2020.
[70] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, P. Abbeel, InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets, 2016, arXiv:1606.03657.
[71] Y. Shen, P. Luo, P. Luo, J. Yan, X. Wang, X. Tang, FaceID-GAN: Learning a symmetry three-player GAN for identity-preserving face synthesis, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 821–830, http://dx.doi.org/10.1109/CVPR.2018.00092.
[72] X. Yin, X. Yu, K. Sohn, X. Liu, M. Chandraker, Towards large-pose face frontalization in the wild, 2017, arXiv:1704.06244.
[73] J. Bao, D. Chen, F. Wen, H. Li, G. Hua, Towards open-set identity preserving face synthesis, 2018, arXiv:1803.11182.
[74] L. Li, S. Wang, Z. Zhang, Y. Ding, Y. Zheng, X. Yu, C. Fan, Write-a-speaker: Text-based emotional and rhythmic talking-head generation, in: AAAI 2021, 2021.

[75] Z. Wang, X. Yu, M. Lu, Q. Wang, C. Qian, F. Xu, Single image portrait relighting via explicit multiple reflectance channel modeling, ACM Trans. Graph. 39 (6) (2020) 1–13.
[76] M. Kim, F. Liu, A. Jain, X. Liu, Dcface: Synthetic face generation with dual condition diffusion model, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 12715–12725.
[77] D. Liu, X. Gao, C. Peng, N. Wang, J. Li, Heterogeneous face interpretable disentangled representation for joint face recognition and synthesis, IEEE Trans. Neural Netw. Learn. Syst. 33 (10) (2021) 5611–5625.
[78] D.P. Kingma, M. Welling, Auto-encoding variational Bayes, stat 1050 (2014) 1.
[79] Y. Deng, J. Yang, S. Xu, D. Chen, Y. Jia, X. Tong, Accurate 3d face reconstruction with weakly-supervised learning: From single image to image set, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2019, pp. 1–11.
[80] G.B. Huang, M. Mattar, T. Berg, E. Learned-Miller, Labeled faces in the wild: A database forstudying face recognition in unconstrained environments, in: Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition, 2008.
[81] V. Albiero, X. Chen, X. Yin, G. Pang, T. Hassner, Img2pose: Face alignment and detection via 6dof, face pose estimation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2021, pp. 7617–7627.
[82] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[83] V. Albiero, X. Chen, X. Yin, G. Pang, T. Hassner, img2pose: Face alignment and detection via 6DoF, face pose estimation, 2021, arXiv:2012.07791.
[84] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, 2015, arXiv preprint arXiv:1512.03385.
[85] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, S. Chintala, Pytorch: An imperative style, high-performance deep learning library, in: H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché Buc, E. Fox, R. Garnett (Eds.), Advances in Neural Information Processing Systems 32, Curran Associates, Inc., 2019, pp. 8024–8035, URL http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.
[86] P.J. Phillips, H. Moon, S.A. Rizvi, P.J. Rauss, The FERET evaluation methodology for face-recognition algorithms, IEEE Trans. Pattern Anal. Mach. Intell. 22 (10) (2000) 1090–1104.
[87] M.A. Turk, A.P. Pentland, Face recognition using eigenfaces, in: Proceedings. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE Computer Society, 1991, pp. 586–587.
[88] P. Isola, J.-Y. Zhu, T. Zhou, A.A. Efros, Image-to-image translation with conditional adversarial networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1125–1134.
[89] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, S. Hochreiter, Gans trained by a two time-scale update rule converge to a local nash equilibrium, Adv. Neural Inf. Process. Syst. 30 (2017).
[90] L. van der Maaten, G. Hinton, Visualizing data using t-SNE, J. Mach. Learn. Res. 9 (86) (2008) 2579–2605, URL http://jmlr.org/papers/v9/vandermaaten08a.html.

Xiaobiao Du is a phd student at University of Technology Sydney. His interests are face recognition, face synthesis, and 3D reconstruction.

Xin Yu is a Senior Lecturer at the University of Queensland. His research interests are face recognition and face synthesis.

Jinhui Liu obtain his master degree at Tsinghua University. His research interest is face recognition.

Beifen Dai is a assistant professor at Beihang University. Her research interest is ethics.

Feng Xu is a associate professor at Tsinghua University. His research interests are 3D Vision and Graphics Digital Health.