MFGAN: A Novel CycleGAN-Based Network for Masked Face Generation

Weiming Xiong, Mingyang Zhong, Cong Guo, Huamin Wang, Libo Zhang*
College of Artificial Intelligence
Southwest University
Chongqing, China
lbzhang@swu.edu.cn

***Abstract*—In the post-epidemic era, Masked Face Recognition (MFR) is of great significance to our daily life, but it confronts a severe challenge of lacking real-world large-scale masked face datasets with identity labels. Moreover, mask enhances the diversity of face images and further improves the requirements for datasets. To address the above problem, we propose a novel CycleGAN-based masked face generation method MaskedFace-GAN (MFGAN), which is able to generate correct, authentic-looking and type-diverse masked face while ensuring the invariance of facial features. We design a three-stage training pipeline for MFGAN, which corresponds to three modules, respectively. Specifically, a facial feature detector is adopted to guide the model to generate the correct mask in the correct position. Then, by utilizing a mask binary segmentation module, the authenticity of generated images can be guaranteed. Lastly, with mask style encoder, the model can be optimized towards generating type-diverse masked faces. Finally, comparing with advanced masked face synthesis and generation methods comprehensively, our MFGAN achieves the best results. Then we apply the generated masked face datasets to MFR model training, which further proves the feasibility of training MFR models on generated datasets and the effectiveness and advancement of MFGAN compared with other state-of-the-art methods.**

***Index Terms*—Masked Face Generation, Image-to-Image Translation, Mask Style Encoder**

I. Introduction

In recent years, benefiting from the advancement of Convolutional Neural Networks (CNNs), face recognition has developed rapidly [1], [2]. Nowadays, people wear mask in reaction to global pandemics such as COVID-19, but it poses a great challenge to face recognition [3]. National Institute of Standards and Technology (NIST) found that, as most of the facial region are occluded by mask, the discriminative features that can be extracted by face recognition models are reduced, which leads to the degradation of the recognition performance of masked face [4]. However, taking off the mask for face recognition will increase the risk of infection, especially in crowded places such as airport [5]. Therefore, masked face recognition (MFR) is an urgent topic to research [3]. Furthermore, two possible solutions, occlusion robust face recognition (OFR) [6] and partial face recognition (PFR) [7] are not applicable, for they address different problems, as shown in Fig. 1. Mask occlusion is a kind of fixed position, large area, continuous and diverse occlusion, by contrast, random in OFR [6]. Moreover, masked face preserves the facial contour well, which cannot be guaranteed in PFR [7].

Fig. 1. Samples of occluded faces, partial faces and masked faces.

Despite its importance, MFR is still a challenging task due to the absence of large-scale real-world masked face datasets. In recent years, Generative Adversarial Networks (GANs) [8] has greatly promoted the development of image generation methods, and the use of them to generate dataset has gradually been widely adopted [9]. Therefore, GANs methods are naturally adopted to generate masked face, but they cannot be applied smoothly. Thereinto, masked face synthesis methods [10] directly overlay mask on face, which is prone to produce unnatural masked face. And masked face generation (MFG) methods confront two inevitable problems. Firstly, it is difficult to simultaneously generate authentic-looking mask and preserve the invariance of facial features [9], [11], [12]. Secondly, the generated mask types are not abundant and the methods cannot be applied to all kinds of datasets [13]–[15]. Recently, CycleGAN-based methods IAMGAN [15] and SimGAN [12] are specially proposed to generate masked face. However, they still suffer from the problems of incorrect wearing, sharpness distortion, etc, as shown in Fig. 2. Note that MFG is different from the other face generation tasks for the reason that the mask cover half of the face, and have various types, and are greatly influenced by the face posture, illumination and angle.

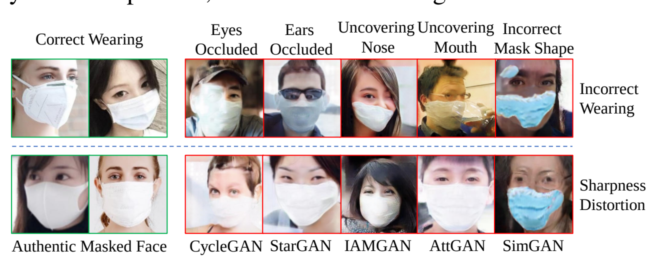

Fig. 2. Example results of some advanced masked face synthesis and generation methods reproduced in our experiments.

To address the aforementioned problems, we propose a novel CycleGAN-based masked face generation method

*Corresponding author: Libo Zhang (lbzhang@swu.edu.cn)
DOI reference number: 10.18293/SEKE2022-016

MaskedFaceGAN (MFGAN), which aims to generate correct, authentic-looking and type-diverse masked face images. In addition, in order to make the generated images meet the standard of training set, we consider more generation details, such as mask bandage, occlusion position, sharp edge, opaque, facial fidelity, fold feeling, multi-style, etc. Note that face recognition is a task involving major ethical issues, and this paper is devoted to the research of masked face generation. This paper only uses the official face datasets collected by legal means to conduct experiments, and only applies the generated datasets for MFR model training. Then we guarantee that the proposed method will not be applied to private face images, so as to protect personal privacy as much as possible. The generated samples on *FFHQ* [16] are shown in Fig. 3. Obviously, MFGAN can generate natural-looking mask on different ages, genders and complexion face.

Female Children Male

Fig. 3. Example results of generated masked face by proposed MFGAN.

In order to optimize MFGAN in three directions: correctness, authentic-looking and type-diversity, we design a three-stage training pipeline that introduces specific modules in different stages, detailed in Chapter III. (1) Firstly, in order to generate correct mask, we propose a facial feature detector, which can detect whether the facial features are occluded or not on the generated images according to the standard for correct mask wearing, so as to guide the model to generate the correct mask in the correct position. (2) Secondly, in order to generate authentic-looking mask, we propose a mask binary segmentation module to measure the fidelity of facial features in non-mask areas, so as to guide the model to generate mask without losing facial features as much as possible. (3) Thirdly, in order to generate type-diverse mask, we propose a mask style encoder, which can extract mask style code from the real-world referenced masked face, so as to guide the model to generate mask of corresponding style on the input face.

Finally, through qualitative and quantitative comparative experiments of the generated images, the proposed MFGAN achieves the best results. In addition, we further apply the generated images for MFR model training, the results show that the images generated by MFGAN is more suitable as training dataset for MFR model. The main contributions of this paper are listed as follows:

1) MFGAN is proposed to generate correct, authentic-looking and type-diverse masked face as a remedy of training dataset absence. We adopt it to generate two masked face datasets (*Masked-FFHQ*, *Masked-CelebA*) and publish them to facilitate future research [1].
2) We design a three-stage training pipeline. The facial detector ensures the correctness of masks, the mask binary segmentation module preserves the non-mask facial area, and the mask style encoder guides model to generate diverse styles of masks.

[1] https://github.com/MySky37/MySky.github.io

The outline of this paper is as follows. In Section II, we survey the related works. Then we introduce the proposed methods MFGAN in detail in Section III. Next, in Section IV, we conduct the experiments for generated images. Finally, we conclude the paper in Section V.

II. Related Work

A. Masked Face Dataset

Wang et al. [17] proposed a Real-World Masked Face Dataset (*RMFRD*), including 5,000 images of 525 people with mask and 90,000 images of the same people without mask. Anwar et al. [10] proposed *MFR*2 dataset that includes 269 images of 53 politicians and celebrities from the Internet. However, the datasets proposed above are small-scale and not enough for MFR model training, but they are barely suitable as test sets. Therefore, we follow the previous work and adopt *RMFRD* as our test benchmark.

B. Masked Face Synthesis and Generation Method

Anwar et al. [10] proposed an open-source tool MaskTheFace, which is used to synthesize masked face. MaskedFaceNet [18] is a masked face synthesis method, which can synthesize masked face with different wearing postures, but its mask type is too single and not authentic-looking enough. Geng et al. [15] proposed a masked face generation method named Identity Aware Mask GAN (IAMGAN) with segmentation guided multi-level identity preserve module, which gained certain performance improvement compared to traditional CycleGAN [11]. However, the above methods have various disadvantages, as shown in Fig. 2. Following the previous work, we propose a novel CycleGAN-based masked face generation method MaskedFaceGAN (MFGAN).

III. Method

CycleGAN [11] is a classic unpaired image-to-image translation method, which is suitable for masked face generation, so we adopt it as the backbone of our MFGAN. Correctness, authenticity and type-diversity are three basic standards of masked face generation, so we specially design a three-stage training pipeline for MFGAN, as shown in Section III-A. In Section III-B, we propose a facial feature detector to guide the model to generate the correct mask in the correct position. In Section III-C, we propose a mask binary segmentation module to guide the model to generate authentic-looking mask without losing facial features as much as possible. In Section III-D, we propose a mask style encoder to guide the the model to generate diverse styles of masks.

A. Three-stage training pipeline

We design a three-stage training pipeline for MFGAN, as shown in Fig. 4, that is, we carry out gradual training for it, so that it can "learn" correctness first, then authenticity, and finally type-diversity. The computational complexity is also increasing gradually, detailed in Chapter IV.

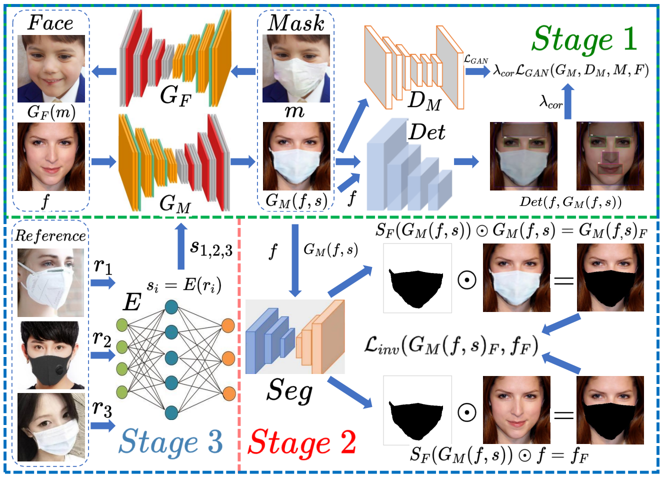

Fig. 4. A three-stage training pipeline of MFGAN. Each stage is represented by different colors and boxes. There are three image sets namely $Face$, $Mask$, and $Reference$. G_F and G_M are generators, D_M is a discriminator, Det is a facial feature detector, Seg is a mask binary segment module, and E is a mask style encoder. s_i is the mask style code of referenced masked face r_i. $G_M(f,s)_F$ and f_F represent the face with mask region removal.

B. Stage1: Generate Correct Mask

For masked face generation, the first requirement is to generate correct mask in the correct position. However, as shown in Fig. 2, some existing methods cannot guarantee the above premise well, and expose the problems of incorrect wearing, incorrect mask shape, etc. Therefore, we experimentalized and found that without additional supervision information, the generator is often prone to "make mistakes", as shown in Fig. 2. Therefore, we consulted the literature and learnt that the medical standard for wearing mask is to cover the mouth and nose, not eyes, and bandages should be hung on the ears. We also learnt that Yolov3 is a popular face detection method, and it can achieve good results in facial feature detection task. Therefore, based on a pretrained Yolov3 model, we construct a facial feature detector to check whether the mask in the generated image is worn correctly or not, as shown in Fig. 4 (Stage 1).

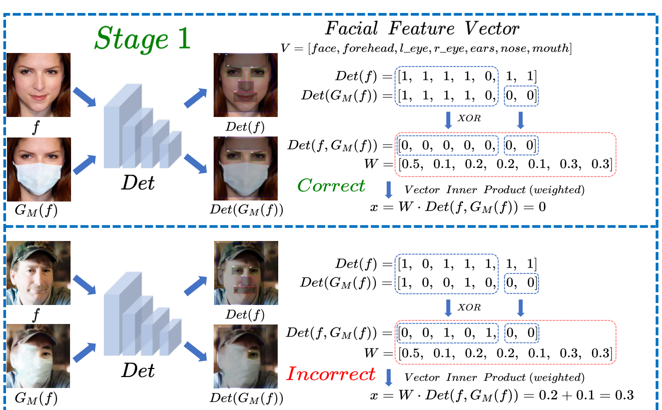

Fig. 5. Examples of detection result calculation. XOR represents exclusive OR operation, the same is 0 and the difference is 1. W means the penalty weight vector and x means the final penalty value.

The detection process is shown in Fig. 5 and described in detail as follows: Firstly, the proposed detector Det is used to detect facial features of face f and corresponding generated masked face $G_M(f)$, including face, eyes, nose, mouth, ears, forehead, etc. Secondly, the output item $D_i, i \in face, eye, mouth, ...$ with consistent feature detection results is 0, otherwise 1. For example, if eyes can be detected on face while not on masked face, which means the detection results are inconsistent and generated mask is incorrect, so the output is $D_{eye} = 1$. Then, all the outputs are formed into a vector $Det(f, G_M(f)) = [D_{face}, D_{eye}, D_{mouth}, ...]$, and further multiplied by the penalty weight vector W to obtain the final penalty value $x = W * Det(f, G_M(f))$. Thirdly, we input x into a revised adaptive correction function, which is an extended version of adaptive loss function [19], to get the adaptive correction factor λ_{cor}, the formula is as follows:

$$\lambda_{cor} = 1 + \frac{|\alpha - 2|}{\alpha}\left(\left(\frac{x^2}{c^2|\alpha - 2|} + 1\right)^{\alpha/2} - 1\right) \quad (1)$$

where $\lambda_{cor} = 1$ means correct mask wearing and $\lambda_{cor} > 1$ means incorrect. $\alpha \in R$ is a shape parameter that controls the specific form of the function. $c > 0$ is a scale parameter. Finally, we multiply the adaptive correction factor λ_{cor} with the adversarial loss of the discriminator $\mathcal{L}_{\text{GAN}}$ to obtain the corrected adversarial loss $\lambda_{cor}\mathcal{L}_{GAN}$. It provides additional supervisory information for discriminator to "learn" to distinguish true or false of the generated image according to the mask wearing condition. Then, based on the antagonistic game mechanism, by enhancing the distinctive ability of discriminator, the generator is forced to generate correct mask in correct position. The full objective of stage 1 is as follows:

$$\begin{aligned}\mathcal{L}_{GAN}(G_M, G_F, D_M, D_F) = {} & \lambda_{cor}\mathcal{L}_{GAN}(G_M, D_M, M, F) \\ & + \mathcal{L}_{GAN}(G_F, D_F, F, M) \\ & + \lambda\mathcal{L}_{cyc}(G_M, G_F)\end{aligned} \quad (2)$$

where $\mathcal{L}_{GAN}$ is the standard adversarial loss [8], $\mathcal{L}_{cyc}$ is the cycle consistency loss [11], and λ controls the relative importance of the two objectives.

C. Stage2: Generate Authentic-Looking Mask

Authentic-looking is the key element of masked face generation. Synthesis methods directly overlay the mask on face, which is prone to produce unnatural masked faces. Generation methods will inevitably lose facial features in the generation process, as shown in Fig. 6.

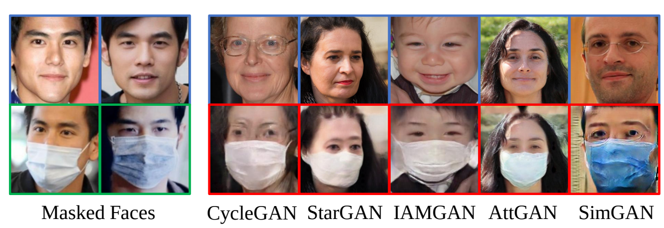

Fig. 6. Examples of facial features loss results of some generation models (CycleGAN, StarGAN, IAMGAN, AttGAN) from our experiments.

Geng et al. [15] adopted U-Net to guide masked face generation and achieved certain effect. Therefore, based on a pre-trained U-Net, we construct a mask binary segmentation module to segment the mask area on the input face and the generated masked face, as shown in Fig. 4 (Stage 2). Given a masked face image $G_M(f)$, Seg predicts a binary segmentation map $S_F(G_M(f))$, where pixel value 0 and 1 represent the mask and non-mask region, respectively. Then we use the element-wise multiplication between $G_M(f)$ and $S_F(G_M(f))$, as well as f and $S_F(G_M(f))$, to obtain the image of mask area removal. Further, by calculating the similarity difference of the non-mask region before and after the generation, the local invariance loss function can be constructed as follows:

$$\mathcal{L}_{inv}(G_M(f)_F, f_F) = \mathbb{E}_{f\sim p_{\text{data}}(f)}\left[\|G_M(f)_F - f_F\|_2^2\right] \quad (3)$$

where $G_M(f)_F$ represents $G_M(f) \odot S_F(G_M(f))$ that means generated masked face of mask region removal, while f_F represents $f \odot S_F(G_M(f))$ that means input face of mask region removal. Then, we add $\mathcal{L}_{inv}$ into full objective of stage 2 and derive the following formula:

$$\begin{aligned}\mathcal{L}(G_M, G_F, D_M, D_F) = {} & \lambda_{cor}\mathcal{L}_{\text{GAN}}(G_M, D_M, M, F) \\ & + \mathcal{L}_{\text{GAN}}(G_F, D_F, F, M) \\ & + \lambda\mathcal{L}_{cyc}(G_M, G_F) \\ & + \mu\mathcal{L}_{inv}(G_M(f)_F, f_F)\end{aligned} \quad (4)$$

where μ controls the relative importance of $\mathcal{L}_{inv}$.

D. Stage3: Generate Type-Diverse Mask

Type-diversity is an indispensable element in masked face datasets construction, because people wear masks in different types, colors and postures in the real world. However, most existing methods do not take it into consideration. Therefore, inspired by StarGANv2 [20] on diversified image translation between multiple domains, we propose a mask style encoder specially designed for masked face generation, which is used to instruct generator to generate mask in the direction of multi-style, as shown in Fig. 4 (Stage 3).

Generator: We extend the form of the input and output for generator G_M, which translates a face image f into a masked face image $G_M(f, s)$ according to domain-specific style code s provided by mask style encoder.

Mask Style Encoder: Given a referenced masked face r, our mask style encoder E extracts mask style code $s = E(r)$. E can produce diverse mask style codes using different referenced masked faces. This allows G_M to generate multi-style masked face reflecting the mask style s of the referenced masked face r. Our mask style encoder consists of a CNN with K output branches, where K is the number of mask style, and one of which is selected when training the corresponding mask domain. To make encoder more suitable for our task, we extend the pre-activation residual blocks to two ResStage blocks [2], and each block includes a Start ResBlock [2], a Middle ResBlock [2] and an End ResBlock [2], and output shape is changed correspondingly. Two ResStage block are also shared among all domains, followed by one specific fully connected layer for each domain. For the loss function, we adopt style reconstruction loss and style diversification loss, which are proposed by StarGANv2 [20].

Style reconstruction loss: It is designed to force the generator G_M to utilize the style code s when generating the masked face $G_M(f, s)$. The formula is as follows:

$$\mathcal{L}_{sty}(G_M, F, s) = \mathbb{E}_{f\sim p_{\text{data}}(f)}\left[\|s - E(G_M(f, s))\|_1\right] \quad (5)$$

Style diversification loss: It is designed to enable the generator G_M to produce diverse styles of masked face images according to the mask style codes. The formula is as follows:

$$\mathcal{L}_{ds}(G_M, F, s_1, s_2) = \mathbb{E}_{f\sim p_{\text{data}}(f)}\left[\|G_M(f, s_1) - G_M(f, s_2)\|_1\right] \quad (6)$$

where the target style codes s_1 and s_2 are produced by E, $s_i = E(r_i)$ for $i = 1, 2$. The goal of maximizing the loss is to force G_M to explore the image space and discover meaningful style features from the input masked face dataset for generating diverse masked face images. Then, we add $\mathcal{L}_{sty}$ and $\mathcal{L}_{ds}$ into full objective of stage 3 and derive the following formula:

$$\begin{aligned}\mathcal{L}(G_M, G_F, D_M, D_F) = {} & \lambda_{cor}\mathcal{L}_{\text{GAN}}(G_M, D_M, M, F) \\ & + \mathcal{L}_{\text{GAN}}(G_F, D_F, F, M) \\ & + \lambda\mathcal{L}_{cyc}(G_M, G_F) \\ & + \mu\mathcal{L}_{inv}(G_M(f)_F, f_F) \\ & + \lambda_{sty}\mathcal{L}_{sty}(G_M, F, s) \\ & - \lambda_{ds}\mathcal{L}_{ds}(G_M, F, s_1, s_2)\end{aligned} \quad (7)$$

where λ_{sty} and λ_{ds} are the weights of the corresponding items.

IV. Experiments

To evaluate the masked face generation performance of the proposed MFGAN and baselines, we made qualitative and quantitative comparative analysis and diversity display in Section IV-C. In Section IV-D, we compared the performance of MFR models trained on different generated datasets. Lastly, we conducted ablation study on the effectiveness of the proposed modules in MFGAN, detailed in Section IV-E.

A. Datasets and Implementation Details

Based on CycleGAN, MFGAN applies adaptive instance normalization (AdaIN) for up-sampling blocks in generator, and adds a convolution layer in discriminator. MFGAN is trained on *FFHQ* (as $Face$) and *RMFRD* (as $Mask$) with adam optimizer for 450K steps totally with batch size 1, and the training steps for three stages are 100K, 200K and 150K, respectively. Additionally, a facial feature detector Yolov3 is pretrained for 62.5K steps with adam optimizer on the detection version of *RMFRD*, and a mask binary segmentation module U-Net is pretrained for 20K steps with SGD optimizer on the segmentation version of *RMFRD*.

For the training of MFR model, we select a public large-scale face dataset *CelebA* and adopt three masked face generation and synthesis methods, MaskTheFace(MTF) [10], IAMGAN [15] and MFGAN, to construct three versions of *Masked-CelebA*[1], respectively.

B. Evaluation Metrics and Baselines

We evaluate the generated images quantitatively and qualitatively, and measure the masked face recognition (MFR) performance of the model trained on generated datasets. For quantitative evaluation, SSIM is used to measure the structural similarity between input face and generated masked face in non-mask area, PSNR is used for image quality evaluation, and FID is adopted to measure the data distribution distance between real-world masked face images and generated images. For qualitative evaluation, it mainly includes feature fidelity, mask transparency, mask type diversity, etc. For MFR performance evaluation, we choose *RMFRD* as test benchmark, and adopt verification accuracy, TAR@FAR=1e-3 and Rank-5 accuracy as evaluation metrics. For baselines, we compare MFGAN with two domain translation methods (CycleGAN [11] and StarGAN [9]), two facial attribute editing methods (AttGAN [13] and SaGAN [14]), and two CycleGAN-based methods (IAMGAN [15] and SimGAN [12]).

C. Comparison of Masked Face Image Generation Effect

In this section, we compare the masked face generated by our MFGAN and other generation and synthesis methods on *FFHQ*, quantitatively and qualitatively.

TABLE I
QUANTITATIVE COMPARISON RESULTS OF SOME ADVANCED MASKED FACE GENERATION METHODS.

Methods	SSIM	PSNR	FID
CycleGAN [11]	0.723	18.42dB	64.29
SimGAN [12]	0.701	16.42dB	83.19
SaGAN [14]	0.742	21.77dB	48.16
AttGAN [13]	0.781	24.81dB	35.77
StarGAN [9]	0.732	21.19dB	51.10
IAMGAN [15]	0.801	26.33dB	27.38
MFGAN	**0.838**	**29.52dB**	**21.73**

1) Quantitative Comparison: As shown in Table I, our MFGAN achieves the best results and outperforms the second best model IAMGAN by a large margin. The largest SSIM and PSNR indicate that, in masked face images generated by MFGAN, the feature information of the non-mask area is the best preserved, and the visual quality is better than others. Then combined with the smallest FID and the visual effect of the generated images, MFGAN can generate the most authentic-looking masked face images.

Fig. 7. Comparison with some state-of-the-art methods on masked face generation. MFGAN is able to generate authentic-looking masked face and preserve the facial features well.

2) Qualitative Comparison: For fair comparison, we randomly select four images from *FFHQ* and feed them into the compared models to generate the corresponding masked face images. From Fig. 7, we observe that the masked face generated by MFGAN are the most natural-looking, and have the best retention effect on the feature information of the non-mask area, while the baselines all have various shortcomings.

3) Diversity Display: Furthermore, most baselines cannot control the style of generated masked face, but our MFGAN can generate masked face with reference to real-world masked face, as shown in Fig. 8.

Fig. 8. Diversity display of masked face images generated by MFGAN.

Obviously, MFGAN can refer to many types of real-world masked faces and generate type-diverse masked faces. It not only proves the effectiveness of proposed mask style encoder, but also further proves the superiority of our MFGAN.

D. Comparison of MFR training effect on generated dataset

We adopt three methods to construct masked face datasets, and then apply them to the masked face recognition (MFR) training of four face recognition (FR) models, respectively.

TABLE II
TRAINING EFFECT OF FR MODELS ON *Mask-CelebA*, WHICH IS SYNTHESIZED OR GENERATED BY THE LEFTMOST METHODS.

Datasets	Methods	Acc	TAR@FAR=1e-3	Rank-5
MTF [10]	Softmax	77.2	65.1	61.2
	Triplet [21]	77.8	66.2	64.6
	CosFace [22]	78.4	67.9	66.5
	ArcFace [23]	78.5	67.9	66.3
IAMGAN [15]	Softmax	82.1	69.2	75.1
	Triplet	83.2	70.4	77.1*
	CosFace	83.6	71.5	72.9
	ArcFace	83.7*	71.6*	73.1
MFGAN	Softmax	89.2	75.9	80.3
	Triplet	90.1	76.7	**82.8(+5.7)**
	CosFace	90.4	78.1	78.1
	ArcFace	**90.9(+7.2)**	**78.4(+6.8)**	78.9

As shown in Table IV-D, we conclude the following results: (1) For verification task, face recognition (FR) models can extract distinguishing features from the full face to accurately judge whether the identities of two face are the same or not, but when confronting the masked face, the models can only extract a few features from non-mask areas, which easily leads to lower verification accuracy. (2) For recognition task, FR models can easily find the best matching identity from the face database, but when confronting the masked face, the models are more likely to be misled by similar faces to make wrong judgments, so the Rank-5 accuracy is relatively low. (3) However, with the same FR models, the training effect on the masked face dataset generated by MFGAN has achieved remarkable performance improvement, which fully proves the feasibility of training MFR models on the generated datasets and the effectiveness and advancement of our MFGAN.

E. Ablation Study

Finally, to analyze the function of different modules in MFGAN, we train three variants of it by removing $\mathcal{L}_{inv}$, λ_{cor}, and $\mathcal{L}_{sty}$, which controls the correctness, authenticity and type-diversity of the generated masked face. Additionally, we use CycleGAN as baseline, which lacks the above three modules simultaneously. The results are shown in Table III.

TABLE III
COMPARISON OF TRAINING EFFECT AND VISUAL QUALITY BETWEEN DIFFERENT VARIANTS OF MFGAN.

Methods	Performance			Visual Quality		
	Acc	TAR@FAR=1e-3	Rank-5	SSIM	PSNR	FID
CycleGAN	76.3	64.8	65.7	0.723	18.42dB	64.29
w/o $\mathcal{L}_{inv}$	80.3	68.5	70.8	0.747	20.44dB	58.46
w/o λ_{cor}	85.4	73.9	76.3	0.764	23.31dB	47.66
w/o $\mathcal{L}_{sty}$	87.2	75.3	80.1	0.813	27.31dB	25.98
All	**90.9**	**78.4**	**82.8**	**0.838**	**29.52dB**	**21.73**

Obviously, without $\mathcal{L}_{inv}$, MFGAN occurs serious performance degradation. Without λ_{cor}, MFGAN loses the ability to accurately control the generated position of mask. Without $\mathcal{L}_{sty}$, MFGAN cannot optimize in the direction of generating type-diverse masks. In general, lacking any modules will directly affect the model performance, which validates the effectiveness of the proposed modules and MFGAN.

V. CONCLUSION

In this paper, to alleviate the challenge of lacking large-scale real-world masked face datasets, we propose a novel CycleGAN-based masked face generation method Masked-FaceGAN (MFGAN), which enables to generate correct, authentic-looking and type-diverse masked face images. A three-stage training pipeline combined with facial feature detector, mask binary segmentation module and mask style encoder is designed to gradually optimize MFGAN. In addition, the masked face version of *FFHQ* and *CelebA* generated by MFGAN are publicly available to facilitate future research. Extensive experiments from quantitative, qualitative and diversity aspects have proved the practical significance and performance advantages of MFGAN and its corresponding modules. However, due to the lack of large area facial features, the masked face recognition (MFR) task is inherently difficult, and the performance of existing methods is still unsatisfactory. But the performance improvement of MFR models training on the datasets generated by MFGAN fully proves the feasibility of training MFR models on generated datasets. In the future, we will further enhance the robustness of MFGAN, and conduct in-depth research on MFR model.

ACKNOWLEDGMENT

This work is supported by the Fundamental Research Funds for the Central Universities (Nos. SWU021001 and SWU021002), the National Nature Science Foundation of China (No. 62106205), the Project of Science and Technology Research Program of Chongqing Education Commission of China (No. KJZD-K202100203 and KJQN202100207), and the Natural Science Foundation of Chongqing (Nos. cstc2021jcyj-msxmX0824 and cstc2021jcyj-msxmX0565).

REFERENCES

[1] Q. Zhang, Q. Cai, and J. Zheng, "Inspect defect of power equipment via deep learning method."

[2] I. C. Duta, L. Liu, F. Zhu, and L. Shao, "Improved residual networks for image and video recognition," in *2020 25th International Conference on Pattern Recognition*, 2021, pp. 9415–9422.

[3] A. Alzu'bi, F. Albalas, T. Al-Hadhrami, L. B. Younis, and A. Bashayreh, "Masked face recognition using deep learning: A review," *Electronics*, vol. 10, no. 21, p. 2666, 2021.

[4] M. L. Ngan, P. J. Grother, K. K. Hanaoka *et al.*, "Ongoing face recognition vendor test (frvt) part 6b: Face recognition accuracy with face masks using post-covid-19 algorithms," 2020.

[5] N. Damer, J. H. Grebe, C. Chen, F. Boutros, F. Kirchbuchner, and A. 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)*. IEEE, 2020, pp. 1–6.

[6] D. Zeng, R. Veldhuis, and L. Spreeuwers, "A survey of face recognition techniques under occlusion," *arXiv preprint arXiv:2006.11366*, 2020.

[7] S. A. Abushanap, A. M. Abdalla, A. A. Tamimi, and S. Alzu'bi, "A survey of human face recognition for partial face view," in *2021 International Conference on Information Technology (ICIT)*. IEEE, 2021, pp. 571–576.

[8] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in neural information processing systems*, vol. 27, 2014.

[9] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, "Stargan: Unified generative adversarial networks for multi-domain image-to-image translation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8789–8797.

[10] A. Anwar and A. Raychowdhury, "Masked face recognition for secure authentication," *arXiv preprint arXiv:2008.11104*, 2020.

[11] J.-Y. Zhu, T. Park, P. Isola, and 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.

[12] S. Mumford, "Generation of realistic facemasked faces with gans," http://cs230.stanford.edu/projects_winter_2021/reports/70681837.pdf, 2021.

[13] Z. He, W. Zuo, M. Kan, S. Shan, and X. Chen, "Attgan: Facial attribute editing by only changing what you want," *IEEE transactions on image processing*, pp. 5464–5478, 2019.

[14] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," in *International conference on machine learning*, 2019, pp. 7354–7363.

[15] M. Geng, P. Peng, Y. Huang, and Y. Tian, "Masked face recognition with generative data augmentation and domain constrained ranking," in *Proceedings of the 28th ACM International Conference on Multimedia*, 2020, pp. 2246–2254.

[16] T. Karras, S. Laine, and 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.

[17] Z. Wang, G. Wang, B. Huang, Z. Xiong, Q. Hong, H. Wu, P. Yi, K. Jiang, N. Wang, Y. Pei *et al.*, "Masked face recognition dataset and application," *arXiv preprint arXiv:2003.09093*, 2020.

[18] A. Cabani, K. Hammoudi, H. Benhabiles, and M. Melkemi, "Maskedface-net – a dataset of correctly/incorrectly masked face images in the context of covid-19," *Smart Health*, 2021.

[19] J. T. Barron, "A general and adaptive robust loss function," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2019, pp. 4331–4339.

[20] Y. Choi, Y. Uh, J. Yoo, and J.-W. Ha, "Stargan v2: Diverse image synthesis for multiple domains," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2020, pp. 8188–8197.

[21] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 815–823.

[22] H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, Z. Li, and 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.

[23] J. Deng, J. Guo, N. Xue, and 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.