Identity-Preserved Complete Face Recovering Network for Partial Face Image

Mengke Li, *Graduate Student Member, IEEE*, and Yiu-ming Cheung, *Fellow, IEEE*

***Abstract*—Complete face recovering (CFR) is to recover the face image of a given partial face image of a target person whose photo may not be included in the gallery set. CFR has several attractive potential applications in surveillance, personal identification in forensics, to name a few, but it is challenging because of little information revealed from a single partial face image. Furthermore, the facial identity may get lost when recovering the complete face image. As far as we know, CFR problem has yet to be explored in the literature. This paper therefore proposes an identity-preserved CFR approach (IP-CFR) to tackle this problem. Accordingly, a denoising auto-encoder based network is applied. We propose an identity-preserved loss function to constrain the features in latent space of decoder, whereby maintaining the personal identity information. Then, to better restore the complete face image, the acquired features are further fed into a decoder with an adversarial structure that takes a new variant of discriminator. That is, we borrow the idea from energy based GAN that utilize an auto-encoder structure discriminator. It can produce very different gradient directions within the minibatch and therefore can make the model be trained stably. Further, we propose a novel dual-pipeline structure in the discriminator, which is leveraged to enhance the quality of the recovered image. Experimental results on the benchmark datasets show the superiority of IP-CFR.**

***Index Terms*—Complete face recovering, Face recognition, Identity preservation, Generative adversarial network.**

I. Introduction

In real-world human face identification applications, query face images are generally collected in a less constrained or unconstrained environment. Let us consider the following scenario: In a crowded area such as supermarket and railway station, a target person can easily hide in the crowd, and his/her photos captured by surveillance cameras are probably partial. Further, the photos of a target person do not exist in the gallery set. Under the circumstances, towards recognizing a person based on a single partial face image, a key but challenging problem we have to tackle is how to perform an identity-preserved face recovering of this partial face image. We call this problem **C**omplete **F**ace **R**ecovering (CFR), which not only recovers visually realistic and semantically plausible content for the missing part, but also preserves the personal identity information. To the best of our knowledge, the CFR has yet to be explored thus far.

In the literature, image inpainting (also called image completion) [1], [2] and super resolution (SR) [3] are two kinds of technologies that are related to CFR. They both aim at enhancing the visual effect of the input image, which are, however, quite different from the CFR. Image inpainting/completion methods usually fill in the missing parts inside the image. SR methods reconstruct a higher-resolution image from the observed low-resolution images. These two technologies synthesize the missing details based on the complete structure information of the input corrupted image. Nevertheless, they fail to infer the global image structure of the partial query photo whose overall image structure is destroyed. Under the circumstances, they cannot restore the complete photo well. Recently, partial face recognition methods [4] have received considerable attention and achieved remarkable performance. However, these methods need photos of a target person to be included in the gallery set. Otherwise, they cannot acquire the complete face image by retrieving one from the gallery set, and thus resulting in an incorrect solution.

In this paper, we therefore focus on studying the CFR problem without the target photo included in the gallery. We use a denoising auto-encoder [5] based network for recovering the complete face image. Different from the previous works, we additionally minimize the identity difference between the partial face and its corresponding complete face to obtain the robust identity information. To this end, we define an identity-preserved loss in latent space to seek the manifold of the partial face image. To recover the complete face image, we propose a new variant of generative adversarial network (GAN). Specifically, in order to make the model trained stably, inspired by [6], we use the auto-encoder structure discriminator which can produce different gradient directions within a minibatch. Further, we propose a novel dual-pipeline structure in the discriminator to enhance the global structure and local consistency of the recovered results. The main contributions of this paper are summarized below:

- This is the first attempt to address a new challenging and practical CFR problem;
- We introduce an identity-preserved loss function to constrain the features in manifold space, which is a simple but effective strategy to obtain the essential discriminative features of the face image;
- We improve the discriminator by using the auto-encoder structure and introducing a dual-pipeline to boost the recovery result.

II. Related Work

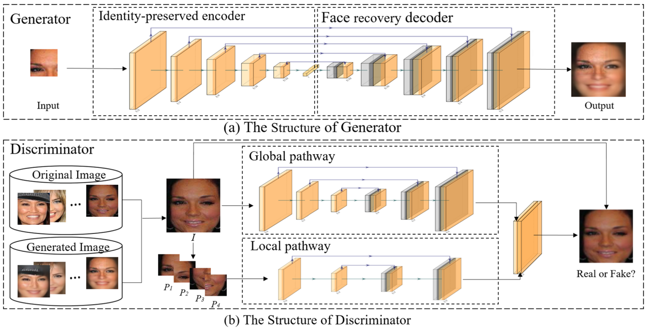
A. Generative Adversarial Network (GAN)

GAN [7] is a kind of generative model that has led to significant progress in image/video inpainting [8], machine translation [9] and haze removal [10], to name a few. A widely accepted cognitive approach is to treat GAN as a probability-based model (PBM) such as Deep Convolutional GAN (DCGAN) [11] and Wasserstein GAN (WGAN) [12]. That is, the discriminator calculates the conditional probability $p(y|x)$ that the sample x belongs to the category y and the generator estimates the joint probability $p(x, y)$. PBM has at least two unsolved problems, i.e. the training difficulty and model collapse. Subsequently, researchers have tried to study GAN along another cognitive perspective, i.e. energy-based model (EBM), which utilizes an energy function to replace probability calculation. Zhao *et al.* [6] had first attempted to propose an energy-based GAN (EBGAN), which gives a clear physical explanation for GAN, Wasserstein distance, and gradient penalty. Following this work, a number of EBM variants, e.g. Margin

Manuscript received 15 March 2021; revised 27 June 2021; accepted 20 July 2021. Date of publication 17 August 2021; date of current version 27 March 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 61672444, in part by Hong Kong Baptist University (HKBU) under Grants RC-FNRA-IG/18-19/SCI/03 and RC-IRCMs/18-19/SCI/01, and in part by the Innovation and Technology Fund of Innovation and Technology Commission of the Government of the Hong Kong SAR under Grant ITS/339/18. *(Corresponding author: Yiu-ming Cheung.)*

The authors are with the Department of Computer Science, Hong Kong Baptist University, Kowloon Tong, Hong Kong SAR, China (e-mail: csmkli@comp.hkbu.edu.hk; ymc@comp.hkbu.edu.hk).

Digital Object Identifier 10.1109/TETCI.2021.3100646



(a) The Structure of Generator

(b) The Structure of Discriminator

Fig. 1. The architecture of the proposed IP-CFR approach.

Adaptation for GAN (MAGAN) [13] and Boundary Equilibrium GAN (BEGAN) [14], have been proposed. Compared with the PBMs, the merits of these EBMs are three-fold: stable convergence, easy for training, and robust against hyper-parameters. In this paper, we will therefore develop our CFR approach based on EBM.

B. Image Inpainting and Super Resolution

Image inpainting aims to automatically fill in a missing region inside an image, and super resolution (SR) aims at recovering a high-resolution image from a given low-resolution one. The goal of these two technologies is to increase the visual effect of the give incorrupt images. Thanks to the revolution of deep generative networks, a number of image inpainting [2], [15]–[18] and SR works [19]–[21] have been proposed with success in their application domains. For example, Yu *et al.* [15] proposed contextual attention which combines a coarse network with a contextual attention model-based fine work to reconstruct the global image as well as fine-tune the detailed texture. This method can fill in the missing regions of arbitrary shapes. Zhang *et al.* [21] proposed Re-CPGAN stacking multiple external copy and paste units, which can produce photo-realistic image as given a very low resolution input. The aforementioned methods generate local image details based on the overall sematic information. However, they do not perform well when dealing with the partial input that lacks global context structure.

III. The Proposed Approach

Given a query photo x_{p_i} of a target person i, where x_{p_i} is a partial face image, the aim of CFR is to recover a complete face, i.e. $\hat{x}_{c_i}$, so that it is close to the complete ground-truth photo, i.e. x_{c_i}, as much as possible with the preservation of personal identity. Subsequently, it is expected that this restored query photo $\hat{x}_{c_i}$ can be used to recognize the target person i correctly.

A. Basic Notations

Suppose we have N subjects. Subject i $(i = 1, 2, \ldots, N)$ has M_i $(M_i \geq 1)$ images for training. Let $X_g = \{x_{g_1}, x_{g_2}, \ldots, x_{g_N}\}$ be the control image set that the images are clean and frontal. X_g is used to control the identity. The partial face image set are denoted as $X_p = \{X_{p_1}, X_{p_2}, \ldots, X_{p_N}\}$, where $X_{p_i} = \{x^1_{p_i}, x^2_{p_i}, \ldots, x^{M_i}_{p_i}\}$. The corresponding complete face image set is denoted as $X_c = \{X_{c_1}, X_{c_2}, \ldots, X_{c_N}\}$, where $X_{c_i} = \{x^1_{c_i}, x^2_{c_i}, \ldots, x^{M_i}_{c_i}\}$.

B. Network Architecture

As illustrated in Fig. 1, the proposed IP-CFR approach first utilizes an identity-preserved encoder to seek the features of the query images. Then, an adversarial structure with the obtained features as condition is utilized. The discriminator exploits dual pipelines. Their details are as follows.

1) Identity-Preserved Encoder: Inspired by the previous work of [5], where the partial face image is treated as some kind of noising input, we design a mapping function, i.e. $Y_p = f_{\theta_f}(X_p)$, and treat it as an essential representation of X_p residing in the manifold. Moreover, it has been shown that both complete and partial face images of one person can be mapped into the same low-dimensional manifold [22]. That is, a partial face image sample x_{p_i} and the corresponding control image x_{g_i} should be mapped into the same manifold through the mapping function f_{θ_f}, where θ_f is the parameter of this mapping function. For each partial face image $x_{p_i} \in X_p$, the low-dimensional representation y_{p_i} is obtained by $y_{p_i} = f_{\theta_f}(x_{p_i})$. The parameter θ_f can be obtained by:

$$\theta_f \ = \arg\min_{\theta_f} L_{ip}(y_{p_i}, y_{g_i}), \tag{1}$$

where L_{ip} is the identity-preserved loss. The details of L_{ip} will be described in following section. The de-noising encoder is utilized to seek out the mapping function $f_{\theta_f}(\cdot)$.

2) Adversarial Structure: The obtained y_{p_i} is utilized as the condition to decode the complete face image $\widehat{x}_{c_i}$ by:

$$\widehat{x}_{c_i} = g_{\theta_g}\left(z|y_{p_i}\right), \tag{2}$$

where $z \sim \mathbf{N}(0, 1)$, θ_g is the parameter of the decoder. Subsequently, the proposed model can be trained using the training set that contains M_i pairs of images $x_{p_i} \in X_p$ and $x_{c_i} \in X_c$ for person i. The objective function is given by:

$$\theta_g \ = \arg\min_{\theta_g} \frac{1}{N}\sum_{i=1}^{N}\frac{1}{M_i}\sum_{j=1}^{M} L_{pw}\left(\widehat{x}^j_{c_i}, x^j_{c_i}\right), \tag{3}$$

where L_{pw} is the pixel-wise reconstruction residual loss that will be described in the following section.

The CFR is an under-determined problem. We therefore modify the decoder by proposing a new recovery network architecture in virtue of EBM to recover the face images from the manifold in the low-dimensional space. The generator is to map a known distribution (e.g., Uniform distribution or Gaussian distribution) to the data distribution. Finding a proper distance measurement between distributions is difficult. Therefore, a parameterized network called discriminator is utilized to directly learn a measurement of the similarity between distributions. This distance should be constrained to avoid divergence. The regular discriminator in GAN [11] uses sigmoid function to limit the distance to the interval [0,1]. However, the gradients provided by the logistic loss in regular discriminator of different samples in a minibatch are not orthogonal, which leads to inefficient training. In order to solve this drawback, we improve the discriminator by using the auto-encoder structure based discriminator D_{θ_D} in [6]. It uses a reconstruction error L_{rec} as the measurement of the distribution similarity:

$$L_{rec}(x) = \|x - D_{\theta_D}(x)\|^2_2 + \beta_\theta\|\theta_D\|^2_2, \tag{4}$$

where $D_{\theta_D} : R^{W\times H\times C} \mapsto R^{W\times H\times C}$, is the discriminator with auto-encoder structure and is parameterized by θ_D, $x \in R^{W\times H\times C}$ is a sample from the image space. β_θ is the hyper-parameter. This L_{rec} can produce different gradient directions within a minibatch.

Traditionally, auto-encoder has been used to represent energy-based model. Therefore, L_{rec} is an energy-based measurement naturally. For real data, we assign low-level energy through D_{θ_D}. In contrast, fake data is allocated high-level energy. The generator tries to recover low-energy images, which are close to the real data. Therefore, the loss function of the discriminator and the generator are formulated by:

$$\begin{cases} L_D &= L_{rec}\left(x_c\right) - k_t \cdot L_{rec}\left(G_{\theta_G}\left(z, x_p\right)\right) & \text{for } \theta_D \\ L_G &= L_{rec}\left(G_{\theta_G}\left(z, x_p\right)\right) & \text{for } \theta_G \end{cases} \tag{5}$$

where k_t is a parameter that varies with the training step t, and used to balance diversity and reality of recovered images. We introduce the diversity ratio γ proposed in [14] to calculate k_t:

$$k_{t+1} = k_t + \lambda_k \left(\gamma \cdot L_{rec}\left(x_c\right) - L_{rec}\left(G_{\theta_G}\left(z, x_p\right)\right)\right) \text{for step } t \tag{6}$$

where λ_k is the learning rate for k_t and is set at 0.001 in our experiment. The initial value of k_t is $k_0 = 0$. γ is defined as:

$$\gamma = \frac{\mathbb{E}_{x_p \sim X_p}\left(L_{rec}\left(G_{\theta_G}\left(z, x_p\right)\right)\right)}{\mathbb{E}_{x_c \sim X_c}\left(L_{rec}\left(x_c\right)\right)}, \tag{7}$$

where $G_{\theta_G}(z, x) = g_{\theta_g}(z|f_{\theta_g}(x))$, is the aforementioned encoder-decoder based identity-preserved generator, $\theta_G = \{\theta_g, \theta_f\}$. γ controls the balance between diversity and quality of generated samples through controlling the balance between G_{θ_G} and D_{θ_D}. It takes the value within the range of [0,1]. When γ is lower, G_{θ_G} pays more attention to the reality of the decoded images, thus leading to the lower diversity of the generated images.

3) Dual-Pipeline Discriminator: Besides using the energy based measurement of the discriminator, we also improve the discriminator structure by using a dual-pipeline that contains a global and a local pathway. The global pathway takes the entire image as input to discern the overall consistency of the face image, while the local pathway observes on one-quarter of the complete face image to determine the local consistency and details. The discriminator also exploits an encoder-decoder structure that reconstructs the input. The global pathway takes the entire image which has been resized to $128 \times 128 \times 3$ pixels as input. It comprises a down-sampling encoder and an up-sampling decoder. The local pathway uses a similar structure with the global pathway except that the input is $64 \times 64 \times 3$-pixel patch. As the recovered region is large, we divide the entire image I into four equally sized patches $\{p_1, p_2, p_3, p_4\}$ and then input them into the local pathway. Each patch usually contains the recovered region. Outputs of the two pathways are subsequently synthesized together by a successive convolution layer to obtain the final reconstructed image. The skip-connections are introduced to relieve the network architecture and fuse multi-scale features. Finally, the distribution of the reconstruction error can be acquired to distinguish the recovered image from the real image. A sketch of the discriminator is shown in Fig. 1 (b). Compared with the generator, the structure of the discriminator is more simple so that it does not introduce too much computation cost during the training stage.

C. Selection of Loss Function

The loss function is essential for training the proposed network. We formulate the loss function as a weighted sum of three individual loss functions below to train the CFR network.

1) Identity-Preserved Loss: The identity-preserve loss function L_{ip} is formulated as:

$$L_{ip}\left(x_{p_i}, x_{g_i}\right) = \mathbb{E}_{x_{p_i} \sim X_{p_i}} \left\| f_{\theta_f}\left(x_{p_i}\right) - f_{\theta_f}\left(x_{g_i}\right) \right\|_2^2, \tag{8}$$

Since the query image, which is a partial face, and the control image represent the same identity, they should be mapped into the same manifold in low-dimensional space. f_{θ_f} is the de-noising encoder in Section III-B. $f_{\theta_f}(x_{p_i})$ and $f_{\theta_f}(x_{g_i})$ correspond to the outputs of the hidden layer, which can be regarded as the features of the same person with the different fields of view. Therefore, it is desirable to make them close as much as possible. This constraint helps to enforce the encoder robust against the variances, and find the essential features of the query image.

2) Reconstruction Residual Loss: The pixel-wise reconstruction residual loss L_{pw} (We use an abbreviated notation: $L_{pw} = L_{pw}(x_{p_i}, x_{c_i})$) is expressed as:

$$L_{pw} = \frac{1}{W \times H} \sum_{w=1}^{W} \sum_{h=1}^{H} \left\| G_{\theta_G}\left(z, x_{p_i}\right)_{(w,h)} - x_{c_i,(w,h)} \right\|_2^2, \tag{9}$$

where W and H denote the width and height of the image, respectively. This reconstruction residual loss function measures the global similarity between the recovered image and the original complete image. Since the recovered face image should be similar to the original complete face image in terms of pixel intensities, we employ the L_2 norm to enforce this similarity. It means that the missing areas of the query image should be repaired after passing through the generator.

3) Adversarial Loss: Since there could be several possible mappings between partial and complete face images, the CFR inherently belongs to an under-determined problem. Conducting the identity features and the pixel intensities similarities may not guarantee that the decoder can output realistic and reliable recovered face images. Thus, the generative part L_G mentioned in Eq.(5) is also introduced to the loss function, which serves as a supervision to motivate the network to generate natural and reliable images by minimizing the distribution distance learned by discriminator.

4) Overall Loss Function: The overall loss function is a weighted sum of the foregoing losses and is expressed as:

$$L_T = \beta_{ip} L_{ip} + \beta_{pw} L_{pw} + \beta_G L_G. \tag{10}$$

D. Algorithm

The proposed approach consists of the training and testing phases, respectively.

1) Training Phase: The training set contains M_i different images for a person i. The input $x_{p_i}^j$ $(j = 1, 2, \ldots, M_i)$ is a random patch cropped from the complete image $x_{c_i}^j$. The control image x_{g_i} is a clean and frontal image selected to control the identity.

The standard forward-backward optimization paradigm is utilized to train the network. In the forward direction, the IP-CFR network takes x_p as input and outputs $\hat{x}_c$. In the backward direction, the networks parameters are updated based on Eq.(5) and (10) over the recovered image. Specifically, the Adam optimization algorithm [23] is utilized to update the parameters θ_G and θ_D. The details of the training phase are summarized in Algorithm 1.

2) Testing Phase: There is no intersection of the objects between the training set and the testing set. When the network has been trained completely, the discriminator can be discarded. We use the generator to recover the complete face image with a partial face image as input.

IV. Experiments

The experiments are conducted to demonstrate the performance of the proposed approach recovering the missing contents on face images, in comparison with the existing state-of-the-art methods. More results can be found in the Web link.[1]

A. Datasets and Experimental Setting

Two benchmark datasets, i.e. CelebFaces Attributes (CelebA) dataset [24] and Labeled Faces in the Wild (LFW) dataset [25], are

[1] https://github.com/Conexpres/experimental_results

Algorithm 1: Identity-Preserved Complete Face Recovering.

```
Input: Partial face image set $X_p$,
corresponding complete face image set $X_c$,
control image set $X_g$,
learning rate $\alpha$.
Output: Recovered complete face images.
$i \leftarrow 0$
while not converged do
  for $k$ steps do
    Compute $g_{\theta_{G_{i,k}}}$-gradient based on Eq. (10):
    $g_{\theta_{G_{i,k}}} \leftarrow \nabla_{\theta_G} L_T$
    Perform Adam-updates for $\theta_G$: $\theta_G \leftarrow \theta_G + \alpha \cdot g_{\theta_{G_{i,k}}}$
  end for
  Compute $g_{\theta_{D_i}}$-gradient based on Eq.(5): $g_{\theta_{D_i}} \leftarrow \nabla_{\theta_G} L_D$
  Perform Adam-updates for $\theta_D$: $\theta_D \leftarrow \theta_D + \alpha \cdot g_{\theta_{D_i}}$
  $i = i + 1$
end while
```

used in the experiments. All the images are cropped out the face part and then re-scaled into the size of $128 \times 128 \times 3$ pixels. The partial face image x_{p_i} is randomly cropped from x_{c_i} with the size of $64 \times 64 \times 3$ to simulate the image occlusion.

The hyper-parameters are empirically set to: $\beta_\theta = 3 \times 10^{-4}$, $\beta_{ip} = 1$, $\beta_{pw} = 1$ and $\beta_G = 5 \times 10^{-2}$.

B. Comparison Methods

In experiments, we compare our IP-CFR with two groups of methods:

Partial face recognition methods. To prove the identity-preserved effect of our IP-CFR, we empirically compared our approach with two state-of-the-art partial face recognition methods: a low-rank regression algorithm GD-HASLR [26] and a learning based method DFM [4]. Moreover, InsightFace [27], as a representative of CNN-based algorithms, is compared as well.

Image restoration methods. We also compare with the most related works, including an image inpainting method, namely Contextual Attention (CA) [15], and a super resolution method, namely Recursive Copy and Paste Generative Adversarial Network (Re-CPGAN) [21], to demonstrate the effect of complete face recovery. The models are both retained with the partial face images as input.

C. Performance Evaluation

The recovered results are utilized as the qualitative index. The peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) are used as quantitative indicator. Recognition accuracy is also used to help evaluate the identity-preserved performance. The recognition accuracy of the recovered images is calculated by firstly extracting deep features with insightFace and then using a cosine-distance metric to calculate the rank-1, rank-5 and rank-10 recognition accuracy.

The qualitative and quantitative performance of the proposed algorithm is evaluated in the following two aspects:

1) The effectiveness: We randomly select 90% of the persons to train our model and the rest of the objects are utilized for testing. 702 subjects who contain no less than 6 images are selected to calculate the recognition accuracy. Among these images, we stochastically select 6 images for each identity. For each person, we randomly select one image as the gallery sample and the remaining 5 images are used as query samples.

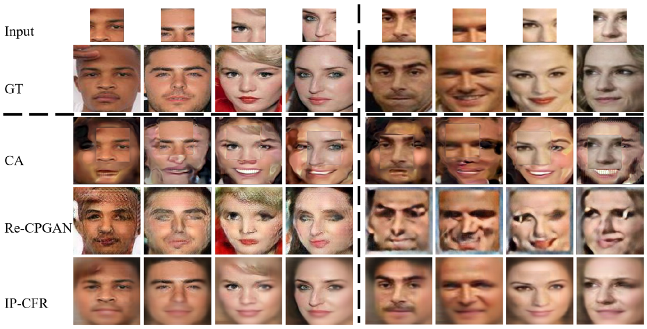


Fig. 2. Qualitative comparisons: The images in left part are from CelebA and in right part are from LFW.

TABLE I
QUANTITATIVE EVALUATION IN TERMS OF PSNR AND SSIM

Metric / Method	PSNR	SSIM
CelebA dataset		
CA	12.0585	0.3546
Re-CPGAN	13.0696	0.5744
IP-CFR	**17.3739**	**0.6649**
LFW dataset		
CA	13.6612	0.4836
Re-CPGAN	14.6169	0.4696
IP-CFR	**16.7954**	**0.6493**

2) The transferability: We use CelebA dataset for training and LFW for testing. To calculate the recognition accuracy, we choose 311 identities that have no less than 6 images of each. Similarly, we randomly select one image as the gallery sample of each person and the rest are set as query samples.

1) Qualitative Evaluation: Fig. 2 presents the examples of CFR results on CelebA dataset and LFW dataset. We can observe that CA cannot preserve the identity information well, even though it can generate visually meaningful pixels. In general, CA mainly utilizes the partial image information to infer the facial components distribution without considering personal identity preserving. Therefore, the inpainting method is inferior to our proposed method. Re-CPGAN needs the guidance of an external CPnet, which affects the prediction of missing parts. There is unpleasant texture recovered by Re-CPGAN. By contrast, our method uses an identity-preserved encoder to find the manifolds of the face images and then get the complete face images by an adversarial structure. The global information is acquired by the encoder and the complete face image is recovered by the generator.

2) Quantitative Evaluation: We compare the PSNR and SSIM of contextual attention and the proposed IP-CFR. These two metrics are used to quantify the similarity between the recovered complete images and the ground truth images. The results are summarized in Table I. We can find that our IP-CFR obtain higher PSNR and SSIM values, which verifies the better performance of IP-CFR.

Noted that PSNR and SSIM favor the images which are exactly the same as the ground truth. Our model aims at generating identity-preserved complete face images rather than the same pixels in the original images. Therefore, we also use the recognition accuracy as an evaluation indicator.

Table III shows the recognition accuracy of the state-of-the-art methods and IP-CFR. The rank-10 recognition accuracy of GD-HASLR and DFM are not competent due to the large occlusion rate of the face images. We also implemented the experiment which adopts the original complete face images as query samples by insightFace and get the Rank-10 results of 90.63% on CelebA and 99.69% on LFW, respectively. By contrast, the recognition accuracy of our proposed

TABLE II
QUANTITATIVE EVALUATION IN TERMS OF PSNR AND SSIM FOR DIFFERENT COMPONENTS OF IP-CFR

Metric / Method	PSNR	SSIM
w/o L_{ip} and L_G	15.0015	0.5306
w/o L_G	15.9854	0.5615
w/o L_{ip}	16.1274	0.6157
regular D	16.0054	0.6051
single pipeline D	16.4216	0.6142
IP-CFR	**16.7954**	**0.6493**

TABLE III
QUANTITATIVE EVALUATION IN TERMS OF RECOGNITION ACCURACY

Acc. / Method	Rank-1	Rank-5	Rank-10
CelebA dataset			
GD-HASLR	11.25 %	22.50%	37.90 %
DFM	11.60 %	30.20%	41.60 %
insightFace	31.06 %	49.54%	57.55 %
CA	33.85 %	49.91%	56.84 %
Re-CPGAN	43.64%	54.55%	63.64%
IP-CFR	**62.52**%	**71.29**%	**80.19**%
LFW dataset			
GD-HASLR	14.79 %	26.05%	50.79%
DFM	13.90 %	36.60%	50.86%
insightFace	37.72 %	55.01%	62.42%
CA	46.58 %	64.44%	66.29%
Re-CPGAN	56.83 %	69.09%	71.82%
IP-CFR	**65.05**%	**84.41**%	**90.48**%

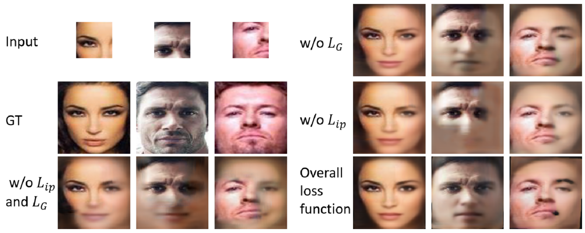


Fig. 3. The effectiveness of different components in the loss function.

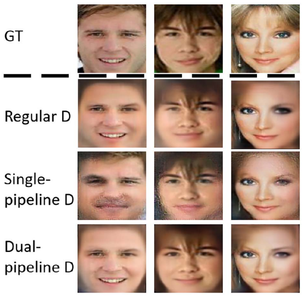


Fig. 4. Comparison of different discriminator structures.

approach (80.19% on CelebA and 90.48% on LFW) is close to using the original complete face images as query samples.

D. Ablation Experiment

To verify the effectiveness of different components of IP-CFR, we conduct ablation experiment using CelebA dataset. Table II, Fig. 3 and 4 show the results.

1) Loss Function: To illustrate the effectiveness of the proposed loss function, we use different loss functions to train the model. Fig. 3 shows the results obtained by training with different losses. We can observe that the recovered image is an unclear average face without L_{ip} and adversarial loss L_G. The face edge and details are blurry without L_G. The detail is more clear but some images cannot be recovered without L_{ip}, and some recovered parts are average face.

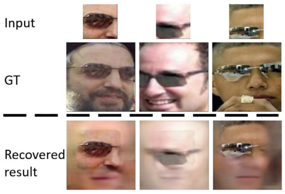


Fig. 5. Examples of failed results.

2) Discriminator Structure: Fig. 4 shows the results obtained by different discriminators. Using the improved discriminator with single pipeline discriminator (single-pipeline D) can benefit learning sharper edges and more detailed textures compared with the regular discriminator. However, there is unpleasant texture in the recovered part.

E. Discussion

Even though our approach can handle various partial face images, it has some limitations. The input with too sparse useful information cannot be repaired due to the difficulty of finding the feature expression of the low-dimensional manifold. The failed examples would mainly be caused by either containing less recognizable information or the images in low resolution. We plan to investigate a better identity-preserved network to seek more discriminative features to address this issue in our future work. Fig. 5 shows several examples of failed recovered results.

The other limitation is that the recovered results lose some subtle image details. From Fig. 2, we can observe that the recovered parts are somewhat blurred, which may diminish visual quality. Limited available information and high uncertainty cause this kind of visual blur. In our future work, more prior information, e.g., the location of the facial features, will be considered to enhance image clarity.

V. CONCLUSION

In this work, we have proposed the IP-CFR network to deal with the new and challenging CFR problem. In our model, an identity-preserved encoder is exploited to seek the features of the query images. Then, an adversarial structure is utilized to recover the complete face image. The discriminator of the adversarial structure adopts an auto-encoder structure combined with a dual-pipeline, which can discriminate features globally and locally to enhance the recovered results. The proposed framework is capable of recovering the complete face image as well as preserving the identity of the query sample that is a partial face image. Experiments have shown the promising results of the proposed approach on the benchmark face datasets.

REFERENCES

[1] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in *Proc. 27th Annu. Conf. Comput. Graph. Interactive Techn.*, 2000, pp. 417–424.

[2] Y. Li, S. Liu, J. Yang, and M.-H. Yang, "Generative face completion," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2017, pp. 3911–3919.

[3] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2017, pp. 4681–4690.

[4] L. He, H. Li, Q. Zhang, and Z. Sun, "Dynamic feature matching for partial face recognition," *IEEE Trans. Image Process.*, vol. 28, no. 2, pp. 791–802, Feb. 2018.

[5] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proc. 25th Int. Conf. Mach. Learn.*, New York, NY, USA: ACM, 2008, pp. 1096–1103.

[6] J. Zhao, M. Mathieu, and Y. LeCun, "Energy-based generative adversarial network," in *Proc. Int. Conf. Learning Representations*, Apr. 2017, pp. 1–17.

[7] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, Curran Associates, Inc., 2014, pp. 2672–2680.

[8] Y. Yeh, Y. Liu, W. Chiu, and Y. F. Wang, "Static2Dynamic: Video inference from a deep glimpse," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 4, no. 4, pp. 440–449, Aug. 2020.

[9] C. Mi, L. Xie, and Y. Zhang, "Improving adversarial neural machine translation for morphologically rich language," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 4, no. 4, pp. 417–426, Aug. 2020.

[10] A. Dudhane, P. W. Patil, and S. Murala, "An end-to-end network for image de-hazing and beyond," *IEEE Trans. Emerg. Topics Comput. Intell.*, early access, 2020, doi: 10.1109/TETCI.2020.3035407.

[11] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," in *Proc. Int. Conf. Lear. Representations*, May 2016, pp. 1–16.

[12] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein gan," in *Proc. Int. Conf. Mach. Learning*, vol. 70, Aug. 2017, pp. 214–223.

[13] R. Wang, A. Cully, H. J. Chang, and Y. Demiris, "Magan: Margin adaptation for generative adversarial networks," 2017. [Online]. Available: http://arxiv.org/abs/1704.03817

[14] D. Berthelot, T. Schumm, and L. Metz, "Began: Boundary equilibrium generative adversarial networks," 2017. [Online]. Available: http://arxiv.org/abs/1703.10717

[15] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, "Generative image inpainting with contextual attention," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2018, pp. 5505–5514.

[16] Y. Ma, X. Liu, S. Bai, L. Wang, D. He, and A. Liu, "Coarse-to-fine image inpainting via region-wise convolutions and non-local correlation," in *Proc. Int. Joint Conf. Artif. Intell. Int. Joint Conf. Artif. Intell. Org.*, 7 2019, pp. 3123–3129.

[17] P. Saxena, R. Gupta, A. Maheshwari, and S. Maheshwari, "Semantic image completion and enhancement using gans," in *High Performance Vis. Intell.*, Springer, 2020, pp. 151–170.

[18] O. Elharrouss, N. Almaadeed, S. Al-Maadeed, and Y. Akbari, "Image inpainting: A review," *Neural Process. Lett.*, vol. 51, no. 2, pp. 2007–2028, 2020.

[19] X. Yu, B. Fernando, R. Hartley, and F. Porikli, "Semantic face hallucination: Super-resolving very low-resolution face images with supplementary attributes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 11, pp. 2926–2943, Nov. 2020.

[20] X. Yu, F. Porikli, B. Fernando, and R. Hartley, "Hallucinating unaligned face images by multiscale transformative discriminative networks," *Int. J. Comput. Vision*, vol. 128, no. 2, pp. 500–526, 2020.

[21] Y. Zhang, I. Tsang, Y. Luo, C. Hu, X. Lu, and X. Yu, "Recursive copy and paste GAN: Face hallucination from shaded thumbnails," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, 2021, doi: 10.1109/TPAMI.2021.3061312.

[22] J. Lu, Y.-p. Tan, S. Member, and G. Wang, "Discriminative multimanifold analysis for face recognition from a single training sample per person," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 39–51, Jan. 2013.

[23] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Representations*, May 2015, pp. 1–15.

[24] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proc. IEEE Int. Conf. Comput. Vision*, 2015, pp. 3730–3738.

[25] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," University of Massachusetts, Amherst, Tech. Rep. 0 7–49, Oct. 2007.

[26] C. Y. Wu and J. J. Ding, "Occluded face recognition using low-rank regression with generalized gradient direction," *Pattern Recognit.*, vol. 80, pp. 256–268, 2018.

[27] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive angular margin loss for deep face recognition," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2019, pp. 4690–4699.