

Complete Face Recovering: An Approach towards Recognizing a Person by a Single Partial Face Image without the Target Photo in Gallery

Yiu-ming Cheung ¹ and Mengke Li ¹

¹Affiliation not available

October 30, 2023

Abstract

Complete face recovering (CFR) is to recover the complete face image of a given partial face image of a target person whose photo may not be included in the gallery set. The CFR has several attractive potential applications but is challenging. As far as we know, the CFR problem has yet to be explored in the literature. This paper therefore proposes an identity-preserved CFR approach (IP-CFR) to addressing the CFR. First, a denoising auto-encoder based network is applied to acquire the discriminative feature. Then, we propose an identity-preserved loss function to keep the personal identity information. Furthermore, the acquired features are fed into a new variant of the generative adversarial network (GAN) to restore the complete face image. In addition, a two-pathway discriminator is leveraged to enhance the quality of the recovered image. Experimental results on the benchmark datasets show the promising result of the proposed approach.

Complete Face Recovering: An Approach towards Recognizing a Person by a Single Partial Face Image without the Target Photo in Gallery

Yiu-ming Cheung and Mengke Li

Department of Computer Science, Hong Kong Baptist University
Hong Kong SAR, China

ymc@comp.hkbu.edu.hk, csmkli@comp.hkbu.edu.hk

Abstract. Complete face recovering (CFR) is to recover the complete face image of a given partial face image of a target person whose photo may not be included in the gallery set. The CFR has several attractive potential applications but is challenging. As far as we know, the CFR problem has yet to be explored in the literature. This paper therefore proposes an identity-preserved CFR approach (IP-CFR) to addressing the CFR. First, a denoising auto-encoder based network is applied to acquire the discriminative feature. Then, we propose an identity-preserved loss function to keep the personal identity information. Furthermore, the acquired features are fed into a new variant of the generative adversarial network (GAN) to restore the complete face image. In addition, a two-pathway discriminator is leveraged to enhance the quality of the recovered image. Experimental results on the benchmark datasets show the promising result of the proposed approach.

Keywords: Complete face recovering, denoising auto-encoder, generative adversarial network

1 Introduction

In real-world human face identification applications, query face images are generally collected in a less constrained or unconstrained environment. Let us consider a scenario below: In the crowded area such as supermarket or railway station, a target person can easily hide in the crowd, and his/her photos captured by surveillance cameras are probably partial. These partial faces can be: 1) occluded by disguises of sunglasses, hats and scarfs, or faces of other individuals; 2) captured in large-scale poses without user awareness; 3) positioned partially outside camera's view, to name a few. Some examples of partial face images as illustrated in Figure 1. Under the circumstances, we can only acquire the partial face images of a target person as the query samples. More worse, the photos of the target person cannot be found in the gallery set. Generally speaking, such partial face images will certainly increase much difficulty for practical face recognition. In particular, there is only one single partial face image is available. Under the circumstance, the learning problem of how to recover a complete face

image for recognition becomes more challenging. We call such learning problem **Complete Face Recovering** (CFR). With this very limited input information, the fundamental challenges of CFR are not only to recover visually realistic and semantically plausible content for the missing part, but also to preserve the personal identity information.

In the literature, the most related task is image inpainting [5] which is, however, quite different from the CFR with two-fold: (1) Image inpainting methods usually fill in visually plausible content for the missing region of an image based on the overall scene. However, they fail to restore the complete photo from these partial query photos because they are unable to predict a large missing region in a face image without relevant information for reference. (2) Image inpainting focuses on generating visually realistic and semantically reasonable image without considering the preservation of the personal identity information during the face image restoration. It turns out that it is hard to identify the target person from the recovered face images. Recently, partial face recognition methods [39, 13, 17] have received considerable attention and achieved remarkable performance. However, these methods need the 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. We use a denoising auto-encoder [36] based network with an identity-preserved loss function as the objective function to seek the manifolds of the query images. To recover the complete face image, we propose a deep architecture exploiting a new variant of generative adversarial network (GAN) based encoder-decoder [4] structure with the feature representation of the acquired manifolds of the query image. Further, we enhance the global structure and local consistency of the recovered results by a two-pathway discriminator, i.e., a global discriminator and a local discriminator. Experiments have shown its promising result. The main contributions of this paper are summarized below:

- We attempt to address a new challenging and practical CFR problem;
- We propose a novel identity-preserved network to represent the features of a face image in manifold space so that the essential features of the face image can be obtained;
- A new GAN-based structure with a two-pathway discriminator is proposed for boosting the recovery result.

The rest of this paper is organized as following. Section 2 reviews the related work of GAN and image inpainting. Section 3 introduces the detail of the proposed IP-CFR model. Section 4 presents the evaluation and the experimental results and Section 5 gives the conclusion of our work.

2 Related Work

2.1 Generative Adversarial Network (GAN)

GAN [15] is a kind of generative model that has led to significant progress in image synthesis [26, 34, 9, 19], super resolution [25, 7, 10, 38, 41], image inpaint-



Fig. 1. Examples of partial face images collected in an unconstrained environment. These query photos suffer varying degrees of damage, including: 1) occluded by disguises of sunglasses, hats and scarfs, or faces of other individuals; 2) captured in large-scale poses without user awareness; 3) positioned partially outside camera' view.

ing [44, 20, 43, 27, 42], style transfer [30, 48, 21, 2, 22], to name a few. The min and max two-player game is an ingenious strategy to estimate the target distribution and then generate photorealistic images. A widely accepted cognitive approach is to treat GAN as a possibility based model (PBM) such as Deep Convolutional GAN (DCGAN) [33], Wasserstein GAN (WGAN) [1] and Improved Training of WGAN [16], etc. Namely, 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 three unsolved problems, i.e., the training difficulty, hard control of the generated image diversity, 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.* have first attempted to propose an energy-based GAN (EBGAN) [46] that gives a clearer physical explanation for GAN, Wasserstein distance, and gradient penalty. Following this work, a number of EBM variants, e.g. Margin Adaptation for GAN (MAGAN) [37], Loss-Sensitive GAN (LSGAN) [32] and Boundary Equilibrium GAN (BEGAN) [6], 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.

2.2 Image Inpainting

Image inpainting is a technique which aims to automatically fill-in a missing region of an image. According to [20], image inpainting methods can be roughly classified into two categories: classical methods and learning-based methods.

Classical methods diffuse the external information along the contour normal to the missing portion [35, 5, 8, 14], or fill the missing area by copying patches from the similar area within an image [40, 18, 18, 24, 3, 11]. However, these classical methods cannot well handle complex images such as non-stationary texture images because they simply manipulate low-level features and are unable to capture a high-level understanding of scenes.

More recently, deep learning-based methods, especially GAN-based methods, have achieved great success at the task of image inpainting. These schemes fill the missing pixels using the learned data distribution, and thus generating a new type of image inpainting methods called learning-based methods [31, 43, 28, 47]. This kind of methods utilizes global context information to generate local missing region, and make the restored image visually realistic. For example, Yu *et al.* [45] combined a coarse network with a contextual attention model-based fine work to reconstruct the global image as well as to fine-tune detail texture, which can fill-in the missing regions of arbitrary shapes. Zheng *et al.* [47] integrated a reconstruction network into a generation network, where the former network obtained the pixel distribution of the missing area while the generation network generated a real area image according to the distribution. Nevertheless, the GAN-based methods ignore preserving the personal identity information during image restoration, which is therefore unfavorable for the subsequent face recognition. Besides, it has been shown that these methods will fail to recover the missing region when this region possesses independent semantic (e.g., eyes, chin and forehead). In such cases, the generator in existing GAN-based inpainting methods may get out of control and generate the strange texture that mismatches with the surrounding region, as illustrated in Figure 4 and 5.

In summary, the traditional methods simply use low-level global information to restore the local missing region. Hence, they are incapable of dealing with complex non-stationary texture images and cannot recover unknown regions in images. For existing learning-based methods, they generate visually realistic and context plausible images, but without considering the personal identity information.

3 The Proposed Approach

The aim of CFR is to restore a query photo of a person (containing partial face), i.e., x_p into a desired complete face, i.e., \hat{x}_c , to make it close to the ground truth complete photo, i.e., x_c . Subsequently, it is expected that the restored query photo \hat{x}_c will be used to recognize x_c correctly. The architecture of the proposed approach, denoted as IP-CFR hereinafter, is illustrated in Figure 2.

3.1 Network Architecture

Identity-preserved Encoder: The proposed algorithm is inspired from the previous work of [36], where the partial face image x_p is treated as some kind of noising input. Based on this, we design a stochastic operator, i.e., $p(x_c|x_p) =$

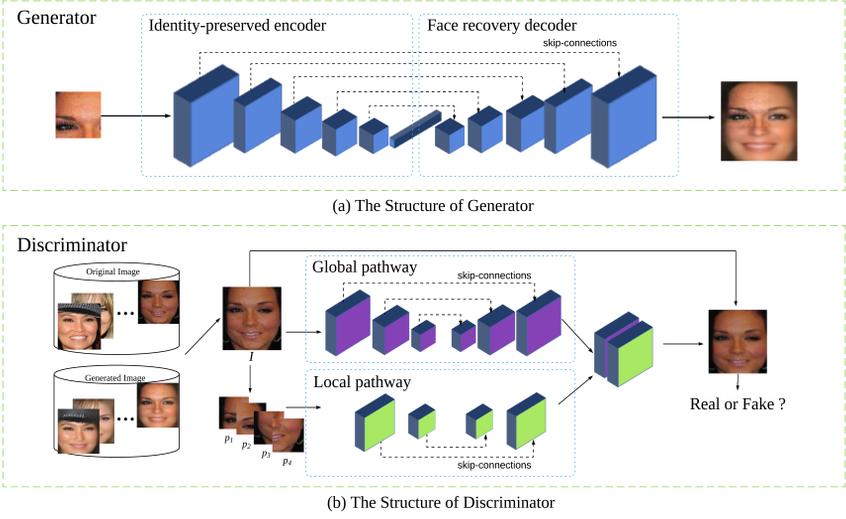


Fig. 2. The Architecture of IP-CFR

$B_{g_{\theta}(y)}(x_c)$, where $y = f_{\theta_f}(x_p)$, and treat it as an essential representation of x_p reside 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 [29]. Figure 3 gives the sketch of a map process. In the training stage, suppose the dataset contains M ($M > 1$) images for a person i ($i = 1, 2, \dots, N$). We select one image x_{g_i} as control image for the person. In a formal formula, x_{g_i} and $x_{p_i}^j$ ($j = 1, \dots, M$) should be mapped into the same manifold through the mapping function f_{θ_f} as follows:

$$f_{\theta_f}(x_{p_i}^j) = f_{\theta_f}(x_{g_i}), \quad (1)$$

where θ_f is the parameter of the encoder. Subsequently, we use the de-noising encoder to seek out the mapping function $f_{\theta_f}(\cdot)$ and use a decoder $g_{\theta_g}(\cdot)$ to restore the complete face image $\hat{x}_{c_i}^j$ as :

$$\hat{x}_{c_i}^j = g_{\theta_g}(f_{\theta_f}(x_{p_i}^j)), \quad (2)$$

where θ_g is the parameter of the decoder. Then, the proposed model can be trained in the training set that contains M pairs of images $x_{p_i}^j$ and $x_{c_i}^j$ for person i , and the objective function is as follows:

$$\hat{\theta}_G = \arg \min_{\theta_G} \frac{1}{M} \sum_{j=1}^M L(g_{\theta_g}(f_{\theta_f}(x_{p_i}^j)), x_{c_i}^j), \quad (3)$$

where $\theta_G = \{\theta_g, \theta_f\}$, and L is represented as:

$$L = \lambda_{ip} \frac{1}{N} \sum_{i=1}^N L_{ip}(x_{p_i}^j, x_{g_i}) + \lambda_{pw} \frac{1}{N} \sum_{i=1}^N L_{pw}(g_{\theta_g}(f_{\theta_f}(x_{p_i}^j)), x_{c_i}^j). \quad (4)$$

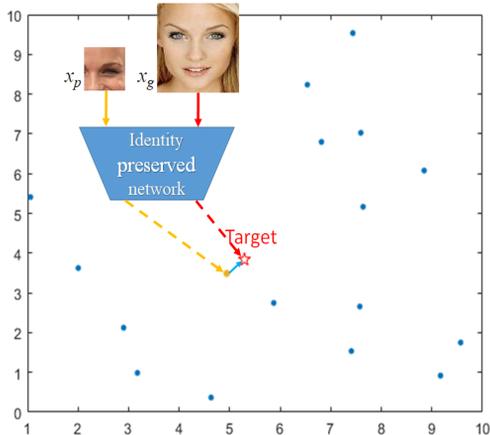


Fig. 3. The sketch of a manifold map, where each point represents the manifold of a person. Both complete and partial face images of one person should be mapped into the same low-dimensional manifold. The identity-preserved network is to seek the map function.

Note that L_{ip} is the identity-preserved constraint that guarantees the partial query photo and the control image are mapped into the same manifold, so as to preserve the target person’s identity. L_{pw} is the pixel-wise reconstruction residual. There are more detailed descriptions about the loss functions in Section 3.2. λ_{ip} and λ_{pw} are the hyper-parameters.

Face Recovery Decoder: To recover the face images from the manifold in the low-dimension space, this paper modifies the decoder by proposing a new recovery network architecture in virtue of EBM. Different from PBM, which estimate examples distribution, this paper introduces BEGAN to match the distributions of reconstruct errors. Besides, BEGAN exploits an encoder-decoder model as the classifier to match the reconstruct loss distribution based on Wasserstein distance. The discriminator tries to enlarge the distance between the reconstruct loss distributions of real samples and generated samples while the generator reduces this distance. This strategy makes the training process fast and stable. The objects of our discriminator and generator are formulized by:

$$\begin{cases} L_D = l_{rec, \theta_D}(x_c) - k_t \cdot l_{rec, \theta_D}(g_{\theta_g}(f_p)) & \text{for } \theta_D \\ L_g = l_{rec, \theta_D}(g_{\theta_g}(f_p)) & \text{for } \theta_g \\ k_{t+1} = k_t + \lambda_k (\gamma \cdot l_{rec, \theta_D}(x_c) - l_{rec, \theta_D}(g_{\theta_g}(f_p))) & \text{for each step } t \end{cases}, \quad (5)$$

where, θ_g and θ_D are the parameter of the face recovery decoder (g) and discriminator (D), respectively. $f_p = f_{\theta_f}(x_p)$ is the feature expression in the manifold

space, as described in Eq.(1). λ_k is the learning rate for k and is set to 0.001 in our experiment. The initial value of k is $k_0 = 0$.

Furthermore, $g : R^X \mapsto R^{W \times H \times C}$, decodes the features into original image space. The feature f_p is an X dimension vector and the recovered image has the size of $W \times H \times C$. The structure of g is shown in Figure 2 (a).

In addition, l_{rec, θ_D} denotes the reconstruction error which can be calculate by:

$$l_{rec, \theta_D}(x) = \|x - D_{\theta_D}(x)\|_{\eta}^2, \quad (6)$$

where $D_{\theta_D} : R^{W \times H \times C} \mapsto R^{W \times H \times C}$, is the encoder-decoder function parameterized by θ_D . $x \in R^{W \times H \times C}$. The structure of D will be described in detail in the following section. η represents the norm which can be 1 or 2. The hyper-parameter γ in Eq.(5) is defined as:

$$\gamma = \frac{E(l_{rec, \theta_D}(g_{\theta_g}(f_p)))}{E(l_{rec, \theta_D}(x_c))}. \quad (7)$$

Here, γ is used to control the balance between g and D , meanwhile balancing the diversity and quality of samples generated by g , which takes the value from $[0, 1]$. When γ is lower, g pays more attention to the reality of the decoded images, thus leading to the lower diversity of the generated images.

Dual Pipeline Discriminator: Besides the improved structure of recovery network, we also enhance the discriminator using a dual pipeline architecture [47] that contains a global and a local pathway. The global pathway network takes the whole image as the input to discern the overall consistency of the face image, while the local pathway observes on one quarter of the entire face image to determine the local consistency and detail information. The discriminator is also based on an encoder-decoder structure that reconstruct the input. Outputs of the two pathways then are synthesized together by a convolution layer that reconstruct the entire image. The skip-connections are introduced for relieving the network architecture and fusing multi-scale features. Finally, the distribution of reconstruct error can be acquired to distinguish the generated image from real image. An overview of the discriminator is shown in Figure 2 (b) and the layers of the discriminator architecture is described in Table 2. The layers of generator is also described and can be seen in Table 1. Here, the enc. and dec. are the abbreviations of encoder and decoder respectively. The conv. is short for convolutional layer and the deconv. is for deconvolution layer. k represents the kernel size, s denotes the stride and c is the number of channels. Every layer is followed by a Rectified Linear Unit (ReLU) activate layer except the synthesis layer. Every layer in the encoder is followed by a residual block before the ReLU layer, which is used for the skip connection. Compared with the generator, this structure of the discriminator is more simple so that does not introduce too much computation cost during the training stage.

The global pathway takes the entire image which has been resized to $128 \times 128 \times 3$ pixels as input. It consists of a down-sampling encoder and an up-

Table 1. Layers of the generator

	Type	k	s	c
enc.	conv.	4	2	64
	conv.	4	2	64
	conv.	4	2	128
	conv.	4	2	256
	conv.	4	2	512
	fc.	-	-	512
dec.	deconv.	4	2	256
	deconv.	4	2	128
	deconv.	4	2	64
	deconv.	4	2	32
	deconv.	4	2	16
	conv.	3	1	3

Table 2. layers of the discriminator

(a) Global Pathway				
	Type	k	s	c
enc.	conv.	3	1	32
	conv.	4	2	32
	conv.	4	2	64
dec.	deconv.	4	2	32
	deconv.	4	2	32
	conv.	3	1	3
(b) Local Pathway				
	Type	k	s	c
enc.	conv.	3	1	32
	conv.	4	2	64
dec.	deconv.	4	2	32
	conv.	3	1	3
(c) Synthesis Layer				
	Type	k	s	c
	conv.	3	1	3

sampling decoder. The local pathway uses 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 $\{p_1, p_2, p_3, p_4\}$ then input them into the local pathway. Each patch usually contains the recovered region. Finally, we fuse the reconstructed images from each pathway by simply feeding them into a successive convolution layer to obtain the final output image. The skip-connections are introduced for relieving the network architecture and fusing multi-scale features. The structure of the discriminator is shown in Fig. 2 (b).

3.2 Training Loss

The loss function is essential for training our network. We formulate this loss function as a weighted sum of terms in Eq.(4) and (5) to train the generator network and the following section will give a detailed description.

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

$$L_{ip}(x_p, x_g) = \|f_{\theta_f}(x_p) - f_{\theta_f}(x_g)\|_{\eta}^2, \quad (8)$$

η can be 1 or 2. Since the query image (which contains partial face) and the control image are the same person, thus should be mapped into the same manifold in low-dimensional space. $f_{\theta_f}(x_p)$ and $f_{\theta_f}(x_g)$ correspond to the outputs of

the hidden layer which can be seen as the feature expressions of the manifolds that are with the same identity. Therefore, they are desirable to be the same. This constraint helps to enforce the decoder robust to the variances and find the essential features of the query image.

Reconstruct Residual Loss: The pixel-wise reconstruct residual loss L_{pw} (Note that we use an abbreviated notation: $L_{pw} = L_{pw}(x_p, x_c)$) is expressed as:

$$L_{pw} = \frac{1}{W \times H} \sum_{w=1}^W \sum_{h=1}^H \left\| G_{\theta_G}(x_p)_{(w,h)} - x_{c,(w,h)} \right\|_{\eta}^2, \quad (9)$$

where W and H denote the with and height of the image respectively. G is the generator which is consisted of the identity-preserved encoder and the face recovery decoder. This reconstruct residual loss function measures the global similarity of the recovered image and the original complete image. Since the recovered face images should be similar to the input complete face images 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.

Adversarial Loss: Since there could be several possible mapping 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 besides the loss functions described so far, the generative part L_g mentioned above in Eq.(5) of the proposed GAN is also introduced to the loss function which serves as a supervision to motivate the network to generate natural and reliable images by trying to mislead the discriminator.

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

$$L_T = \lambda_{ip}L_{ip} + \lambda_{pw}L_{pw} + \lambda_gL_g. \quad (10)$$

3.3 Algorithm

Based on Section 3.1 and 3.2, the training and testing of the proposed approach are below:

Training: 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 .

Testing: When the network get trained, the discriminator can be discarded. We use the generator to recovery the complete face image with a partial face image as input.

Accordingly, the algorithm is given in Algorithm 1:

Algorithm 1 identity-preserved complete face recovering

Input: partial face image $x_{p_i}^j$,
 corresponding complete face image $x_{c_i}^j$,
 control image x_{g_i} , ($j = 1, \dots, M, i = 1, 2, \dots, N$),
 learning rate α .

Output: Recovered complete face image $\hat{x}_{c_i}^j$

- 1: $i \leftarrow 0$
 - 2: **while** not converged **do**
 - 3: **for** k steps **do**
 - 4: Compute $g_{\theta_{G_{i,k}}}$ -gradient based on Eq.(10): $g_{\theta_{G_{i,k}}} \leftarrow \nabla_{\theta_G} L_T$
 - 5: Perform Adam-updates for θ_G : $\theta_G \leftarrow \theta_G + \alpha \cdot g_{\theta_{G_{i,k}}}$
 - 6: **end for**
 - 7: Compute $g_{\theta_{D_i}}$ -gradient based on Eq.(5): $g_{\theta_{D_i}} \leftarrow \nabla_{\theta_G} L_D$
 - 8: Perform Adam-updates for θ_D : $\theta_D \leftarrow \theta_D + \alpha \cdot g_{\theta_{D_i}}$
 - 9: $i = i + 1$
 - 10: **end while**
-

4 Experiment

The extensive experiments are carry out to demonstrate the ability of the proposed model for recovering the missing contents on face images.

4.1 Datasets and Evaluation Criteria

The experiments use two benchmark face datasets: the CelebFaces Attributes (CelebA)¹ dataset and the Labeled Faces in the Wild (LFW)² dataset.

We use the recovered results as the qualitative index. As for the quantitative indicator, it is worth noting that our model aims at generating identity-preserved complete face images rather than the same pixels in the original images. Therefore, we use recognize accuracy to help evaluate the performance of our proposed model.

¹ <http://mmlab.ie.cuhk.edu.hk/projects/CelebA.html>

² <http://vis-www.cs.umass.edu/lfw/index.html>

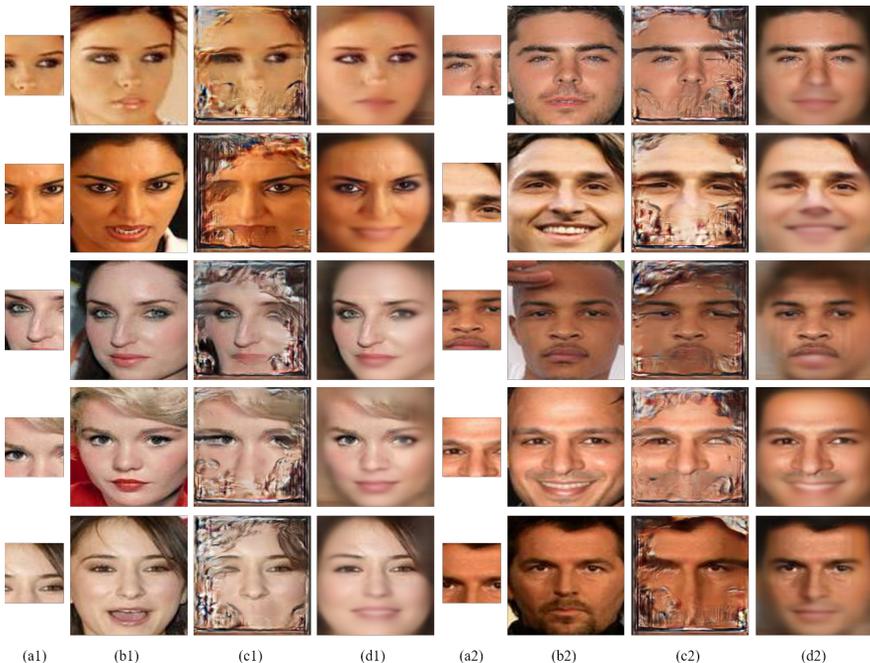


Fig. 4. Qualitative comparisons on CelebA dataset: (a1) and (a2) Input partial face images; (b1) and (b2) Original images; (c1) and (c2) Inpainting results by generative inpainting; (d1) and (d2) Recovered complete face images by IP-CFR.

4.2 Implementation Details

The algorithm evaluation of this paper is divided into two aspects:

1) The effectiveness of IP-CFR. We randomly select 90% of the persons to train our model and the rest of the objects is utilized for testing. 702 persons who contain more than 6 images are selected to calculate the recognize accuracy. Among these images, we stochastically select 6 images for each identity. For each person, we randomly select one image as control image (702 samples in total) then the rest 5 images (3510 samples in total) are used as query samples.

2) The transferability of IP-CFR. We use CelebA dataset for training and LFW for testing. To calculate the recognize accuracy, we choose 311 identities which have no less than 6 images for each. There are 5425 images in total. For each person, we randomly select one image as control image and the remaining 5114 images are set as query samples.

All the images are cropped over the face part and then re-scaled into the size of $128 \times 128 \times 3$ pixels. The partial face image x_p are randomly cropped from x_c with the size of $64 \times 64 \times 3$ to simulate the image occlusion.

The experiments are conducted by firstly extracting deep features with insightFace [12] and then use a cosine-distance metric to calculate the rank-1, rank-5 and rank-10 recognize accuracy.

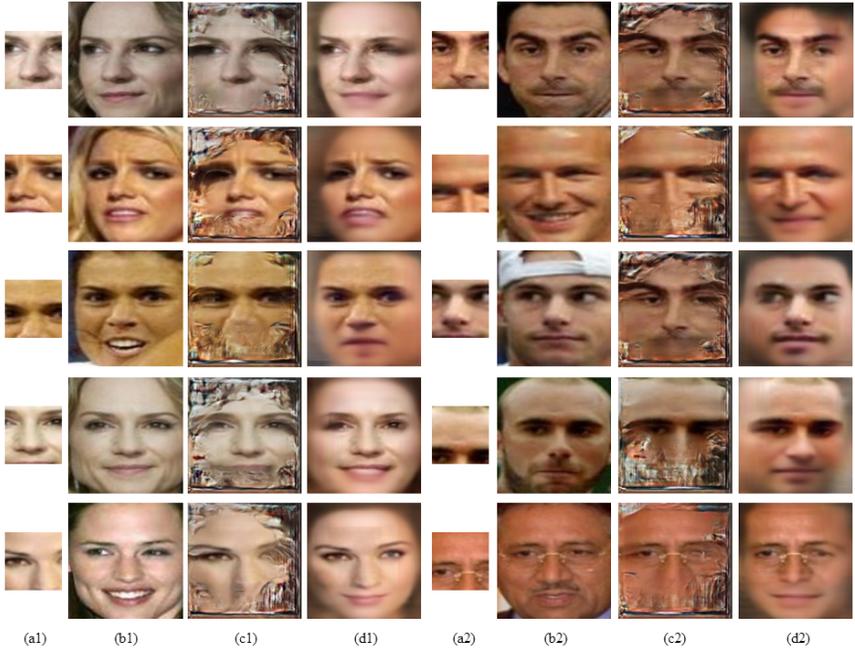


Fig. 5. Qualitative comparisons on LFW dataset: (a1) and (a2) Input partial face images; (b1) and (b2) Original images; (c1) and (c2) Inpainting results by generative inpainting; (d1) and (d2) Recovered complete face images by IP-CFR.

The hyper-parameters are set as: $\lambda_{ip} = 1$, $\lambda_{pw} = 1$ and $\lambda_g = 5 \times 10^{-3}$.

We compare our model with three different kinds of methods:

1. InsightFace is a easy and effective method for face recognition. It improves the speed and accuracy greatly. Moreover, this method has strong robustness under the large posture change (roll, yaw, pitch).

2. Dynamic Feature Matching (DFM) proposed by He *et al.* [17]. DFM is a partial face recognition method that can work for arbitrary input size without pre-alignment.

3. Generative inpainting proposed by Yu *et al.* [45] which is also trained on CelebA dataset. It is an GAN-based inpainting method which can generate realistic inpainting results for the missing pixels inside the input image.

For DFM and generative inpainting, we use the same setting recommended in the papers. For these three methods, we use the cropped images which are the same with the input of our proposed model as query samples.

4.3 Qualitative Comparison

We compare our model with generative inpainting. Figure 4 and 5 present examples of CFR results on CelebA dataset and LFW dataset. We can observe

Table 3. Recognition accuracy comparison with state-of-the-art on CelebA dataset of 702 classes

Method \ Acc.	Rank-1	Rank-5	Rank-10
insightFace	0.3106	0.4954	0.5755
DFM	0.1160	0.3020	0.4160
Generative inpainting	0.3385	0.4991	0.5684
IP-CFR	0.6632	0.7145	0.8589

Table 4. Recognition accuracy comparison with state-of-the-art on LFW dataset of 311 classes

Method \ Acc.	Rank-1	Rank-5	Rank-10
insightFace	0.3772	0.5501	0.6242
DFM	0.1390	0.3660	0.5086
Generative inpainting	0.4658	0.6144	0.6629
IP-CFR	0.6758	0.8641	0.9361

that generative inpainting has limited ability to predict external region of the input. It can generate only little more visual meaningful pixels. Because generative inpainting mainly utilizes the global image information to infer the inner missing pixels' distribution. Once missing region locates outside the input, the model has no global information as guidance, thus generates much meaningless texture. Therefore, the inpainting method cannot handle the CFR problem. Our method use an identity-preserved encoder to find the manifolds of the face images and then get the complete face images by a decoder. The global information is acquired by the encoder and the complete face image is recovered by the decoding process.

4.4 Quantitative Comparison

Table 4.4 and 4.4 show the recognize accuracy of the state-of-the-art methods and IP-CFR. The useful information for recognize of generative inpainting mainly comes from the input images. Therefore, the recognize accuracy increases a little by generative inpainting compared with directly using insightFace. The rank-10 recognize accuracy of DFM is acceptable but not good enough. We also implement the experiment which uses the original complete face images as query samples by insight face and get the Rank-10 results of 0.9063 on CelebA and 0.9969 on LFW. The recognize accuracy of our proposed method (0.8589 on CelebA and 0.9361 on LFW) is close to using the original complete face images as query samples.



Fig. 6. Examples of failed results: The images in the first row are the inputs; The second row shows the recovered results and the last row shows the original images.

4.5 Limitations and 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 difficulty of finding the feature expression of the low-dimensional manifold. The failed examples mainly contain the following situations: 1) contain less recognizable information; 2) are in low resolution. The sparse useful information is due to the randomly crop. Low resolution is caused by image cropping face part and resizing in the preprocessing step. The failed results give us an inspiration that the eye region contains more crucial information for deep network to recognize. Resolution is beyond our consideration. Mapping an image from low resolution to high resolution is a heterogeneous matching problem. We plan to investigate better identity-preserved network to seek more discriminative feature to address this issue in our future work. Figure 6 shows several examples of failed recovered results.

The other limitation is that the recovered results lose some subtle image detail. From Figure 4 and 5, we can see that the recovered parts to some extent become blurred, which can cause diminishment in visual quality. Limited available information causes this kind of visual blur. More prior information, for example the location of the facial features, can be considered to enhance the clarity of the image in our future work.

5 Conclusion

In this work, we have proposed the IP-CFR network to deal with the new but challenging CFR problem. In our model, the generator is composed of an identity-preserved encoder which is constrained by an identity-preserved loss and followed by a complete image recovery decoder. Then, a dual pipeline discriminator (global and local) is applied 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 partial face image. Experiments have shown the promising results of the proposed approach on the benchmark face datasets.

References

1. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein gan. In: 5th International Conference on Learning Representations (2017)
2. Azadi, S., Fisher, M., Kim, V.G., Wang, Z., Shechtman, E., Darrell, T.: Multi-content gan for few-shot font style transfer. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 7564–7573 (June 2018)
3. Barnes, C., Shechtman, E., Finkelstein, A., Goldman, D.B.: PatchMatch: A randomized correspondence algorithm for structural image editing. *ACM Transactions on Graphics* **28**(3), 1–12 (2009). <https://doi.org/10.1145/1531326.1531330>
4. Bengio, Y., Lamblin, P., Popovici, D., Larochelle, H.: Greedy layer-wise training of deep networks. *Advances in Neural Information Processing Systems* **19**, 153–160 (2007)
5. Bertalmio, M., Sapiro, G., Caselles, V., Ballester, C.: Image inpainting. In: Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques. pp. 417–424. SIGGRAPH '00, New York, NY, USA (2000). <https://doi.org/10.1145/344779.344972>
6. Berthelot, D., Schumm, T., Metz, L.: BEGAN: Boundary Equilibrium Generative Adversarial Networks (March 2017)
7. Bulat, A., Yang, J., Tzimiropoulos, G.: To learn image super-resolution, use a gan to learn how to do image degradation first. In: The European Conference on Computer Vision (ECCV). pp. 185–200 (September 2018)
8. Chan, T.F., Shen, J.: Nontexture inpainting by curvature-driven diffusions. *Journal of Visual Communication and Image Representation* **12**(4), 436–449 (2001). <https://doi.org/10.1006/jvci.2001.0487>
9. Chen, X., Duan, Y., Houthoofd, R., Schulman, J., Sutskever, I., Abbeel, P.: Info-gan: Interpretable representation learning by information maximizing generative adversarial nets. In: *Advances in Neural Information Processing Systems* 29, pp. 2172–2180. Curran Associates, Inc. (2016)
10. Dahl, R., Norouzi, M., Shlens, J.: Pixel recursive super resolution. In: The IEEE International Conference on Computer Vision (ICCV). pp. 5439–5448 (October 2017)
11. Darabi, S., Shechtman, E., Barnes, C., Dan B Goldman, Sen, P.: Image melding: Combining inconsistent images using patch-based synthesis. *ACM Transactions on Graphics* **31**(4) (2012). <https://doi.org/10.1145/2185520.2185578>
12. Deng, J., Guo, J., Xue, N., Zafeiriou, S.: Arcface: Additive angular margin loss for deep face recognition. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 4690–4699 (June 2019)
13. Duan, Y., Lu, J., Feng, J., Zhou, J.: Topology Preserving Structural Matching for Automatic Partial Face Recognition. *IEEE Transactions on Information Forensics and Security* **13**(7), 1823–1837 (2018). <https://doi.org/10.1109/TIFS.2018.2804919>
14. Esedoglu, S., Shen, J.: Digital inpainting based on the Mumford-Shah-Euler image model. *European Journal of Applied Mathematics* **13**(4), 353–370 (2002). <https://doi.org/10.1017/S0956792502004904>
15. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: *Advances in Neural Information Processing Systems* 27, pp. 2672–2680. Curran Associates, Inc. (2014)
16. Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A.C.: Improved training of wasserstein gans. In: *Advances in Neural Information Processing Systems* 30, pp. 5767–5777. Curran Associates, Inc. (2017)

17. He, L., Li, H., Zhang, Q., Sun, Z.: Dynamic feature matching for partial face recognition. *IEEE Transactions on Image Processing* **28**(2), 791–802 (2019)
18. Huang, J.B., Kang, S.B., Ahuja, N., Kopf, J.: Image completion using planar structure guidance. *ACM Transactions on Graphics* **33**(4), 1–10 (2014). <https://doi.org/10.1145/2601097.2601205>
19. Huang, R., Zhang, S., Li, T., He, R.: Beyond face rotation: Global and local perception gan for photorealistic and identity preserving frontal view synthesis. In: *The IEEE International Conference on Computer Vision (ICCV)*. pp. 2439–2448 (October 2017)
20. Iizuka, S., Simo-Serra, E., Ishikawa, H.: Globally and locally consistent image completion. *ACM Transactions on Graphics* **36**(4) (2017). <https://doi.org/10.1145/3072959.3073659>
21. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 1125–1134 (July 2017)
22. Jing, Y., Yang, Y., Feng, Z., Ye, J., Yu, Y., Song, M.: Neural Style Transfer: A Review. *IEEE Transactions on Visualization and Computer Graphics* pp. 1–1 (2019). <https://doi.org/10.1109/tvcg.2019.2921336>
23. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. *CoRR* **abs/1412.6980** (2014)
24. Kwatra, V., Essa, I., Bobick, A., Kwatra, N.: Texture optimization for example-based synthesis. *ACM Transactions on Graphics* **24**(3), 795–802 (2005). <https://doi.org/10.1145/1073204.1073263>
25. Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., Shi, W.: Photo-realistic single image super-resolution using a generative adversarial network. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 4681–4690 (July 2017)
26. Li, C., Wand, M.: Combining markov random fields and convolutional neural networks for image synthesis. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 2479–2486 (June 2016)
27. Li, Y., Liu, S., Yang, J., Yang, M.H.: Generative face completion. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 3911–3919 (July 2017)
28. Liu, G., Reda, F.A., Shih, K.J., Wang, T.C., Tao, A., Catanzaro, B.: Image inpainting for irregular holes using partial convolutions. In: *The European Conference on Computer Vision (ECCV)* (September 2018)
29. Lu, J., Tan, Y.p., Member, S., Wang, G.: Discriminative Multimanifold Analysis for Face Recognition from a Single Training Sample per Person. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **35**(1), 39–51 (2013). <https://doi.org/10.1109/TPAMI.2012.70>
30. Luan, F., Paris, S., Shechtman, E., Bala, K.: Deep photo style transfer. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 4990–4998 (July 2017)
31. Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., Efros, A.A.: Context Encoders: Feature Learning by Inpainting. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **2016-Decem**, 2536–2544 (2016). <https://doi.org/10.1109/CVPR.2016.278>
32. Qi, G.J.: Loss-Sensitive Generative Adversarial Networks on Lipschitz Densities (2017)
33. Radford, A., Metz, L., Chintala, S.: Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks (November 2016)

34. Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., Chen, X., Chen, X.: Improved techniques for training gans. In: *Advances in Neural Information Processing Systems 29*, pp. 2234–2242. Curran Associates, Inc. (2016)
35. Shen, J., Chan, T.F.: Mathematical Models for Local Nontexture Inpaintings. *SIAM Journal on Applied Mathematics* **62**(3), 1019–1043 (January 2002). <https://doi.org/10.1137/S0036139900368844>
36. Vincent, P., Larochelle, H., Bengio, Y., Manzagol, P.A.: Extracting and composing robust features with denoising autoencoders. In: *Proceedings of the 25th International Conference on Machine Learning*. pp. 1096–1103. ICML '08, ACM, New York, NY, USA (2008). <https://doi.org/10.1145/1390156.1390294>
37. Wang, R., Cully, A., Chang, H.J., Demiris, Y.: MAGAN: Margin Adaptation for Generative Adversarial Networks (April 2017)
38. Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., Change Loy, C.: Esrgan: Enhanced super-resolution generative adversarial networks. In: *The European Conference on Computer Vision (ECCV) Workshops* (September 2018)
39. Weng, R., Lu, J., Tan, Y.P.: Robust point set matching for partial face recognition. *IEEE Transactions on Image Processing* **25**(3), 1163–1176 (2016). <https://doi.org/10.1109/TIP.2016.2515987>
40. Wexler, Y., Shechtman, E., Irani, M.: Space-time completion of video. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **29**(3), 463–476 (2007). <https://doi.org/10.1109/TPAMI.2007.60>
41. Xie, Y., Franz, E., Chu, M., Thuerey, N.: tempoGAN: A Temporally Coherent, Volumetric GAN for Super-resolution Fluid Flow. *ACM Transactions on Graphics* **37**(4), 95:1–95:15 (January 2018). <https://doi.org/10.1145/3072959.3073643>
42. Yan, Z., Li, X., Li, M., Zuo, W., Shan, S.: Shift-net: Image inpainting via deep feature rearrangement. In: *The European Conference on Computer Vision (ECCV)*. pp. 1–17 (September 2018)
43. Yang, C., Lu, X., Lin, Z., Shechtman, E., Wang, O., Li, H.: High-resolution image inpainting using multi-scale neural patch synthesis. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. vol. 2017-Janua, pp. 4076–4084. IEEE (July 2017). <https://doi.org/10.1109/CVPR.2017.434>
44. Yeh, R., Chen, C., Lim, T., Hasegawa-Johnson, M., Do, M.: Semantic image inpainting with perceptual and contextual losses. *arxiv, 2016. arXiv preprint arXiv:1607.07539* (2016)
45. Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., Huang, T.S.: Generative image inpainting with contextual attention. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 5505–5514. IEEE (June 2018). <https://doi.org/10.1109/CVPR.2018.00577>
46. Zhao, J., Mathieu, M., Lecun, Y.: Energy-based generative adversarial networks (2017)
47. Zheng, C., Cham, T.J., Cai, J.: Pluralistic image completion. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 1438–1447 (June 2019)
48. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.E.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: *The IEEE International Conference on Computer Vision (ICCV)*. pp. 2223–2232 (October 2017)