Protecting Your Faces: MeshFaces Generation and Removal via High-order Relation-preserving CycleGAN

Zhihang Li[1,2,3], Yibo Hu[1,2,3], Man Zhang[1,2,3], Min Xu[4], Ran He[1,2,3]
[1]National Laboratory of Pattern Recognition, CASIA
[2]Center for Research on Intelligent Perception and Computing, CASIA
[3]University of Chinese Academy of Sciences, Beijing, 100049, China
[4]College of Information Engineering, Capital Normal University, Beijing 100048, China
{zhihang.li, zhangman, rhe}@nlpr.ia.ac.cn, yibo.hu@cripac.ia.ac.cn, xumin@cnu.edu.cn

***Abstract*—Protecting person's face photos from being misused has been an important issue as the rapid development of ubiquitous face sensors. MeshFaces provide a simple yet inexpensive way to protect facial photos and have been widely used in China. This paper treats MeshFace generation and removal as a dual learning problem and proposes a high-order relation-preserving CycleGAN framework to solve this problem. First, dual transformations between the distributions of MeshFaces and clean faces in pixel space are learned under the CycleGAN framework, which can efficiently utilize unpaired data. Then, a novel High-order Relation-preserving (HR) loss is imposed on CycleGAN to recover the finer texture details and generate much sharper images. Different from the $L1$ and $L2$ losses that result in image smoothness and blurry, the HR loss can better capture the appearance variation of MeshFaces and hence facilitates removal. Moreover, Identity Preserving loss is proposed to preserve both global and local identity information. Experimental results on three databases demonstrate that our approach is highly effective for MeshFace generation and removal.**

***Keywords*-MeshFaces; High-order; CycleGAN; de-mesh; de-noise;**

I. Introduction

Blind image inpainting, as a common image restoration problem, aims to restore the original image from a signal corrupted by noise, which is essential for many visual tasks including face alignment, face verification, etc. In real-world scenarios, due to faulty imaging sensors, environment and other human purposes, images are usually corrupted by noise. Learning directly from corrupted images will severely deteriorate the performance of a model [25]. However, most existing algorithms assume that training and testing data are noise-free, where there is still a gap in practice. Therefore, blind image inpainting catches amount of attention in computer vision community.

In real world application, ID photos provided by Chinese business organizations are corrupted by random wavy lines to protect private information from abusing or being illegally distributed. [26] denotes this type of corrupted photos as MeshFaces, as shown in Figure 1. Face verification using MeshFaces directly results in poor performance, because the identity of face has changed greatly in feature space.

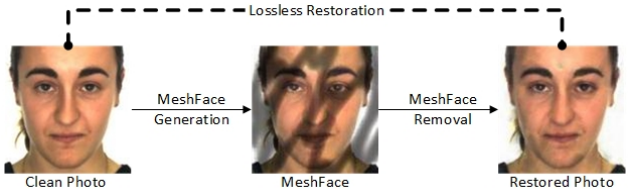


Figure 1. MeshFaces (Middle) refer to the images corrupted by randomly generated mesh-like lines or watermarks. Restored Photo (Right) is generated from MeshFace by our method.

Furthermore, experiments [3], [16] have demonstrated that slight perturbations lead to significant deterioration in performance for several machine learning models. Thus, how to recover the clean face photos from the corrupted ones so as to improve the recognition performance remains challenging.

Some efforts [25], [26], [19], [13] have been devoted to addressing this problem. [25] proposes a multi-task architecture to restore the clean face photos where detecting and reconstructing corrupted pixels are conducted simultaneously. They take blind face inpainting as an image reconstruction problem. In addition to removing mesh, Zhang et al. [26] constrain the pixel level similarity and the feature level similarity jointly to improve the face verification performance. However, we argue that different from the common image restoration problems, MeshFaces contain some important identity information. Therefore, the blind face inpainting cannot be treated as a single low-level vision problem. Mesh removing and identity preserving play an equally important role.

On the other hand, most of the blind inpainting methods [25], [26] need paired training data. However, for many tasks, collecting massive paired data is labor-intensive or even unavailable. Moreover, although some research efforts have been devoted to blind face inpainting which restores the clean face from MeshFaces, few attention is payed to the inverse problem, i.e. adding random patterns to images, which is useful for information security in many cases, e.g. ID photo.

0-7695-6403-8/18/31.00 ©2018 IEEE
DOI 10.1109/ICB2018.2018.00020

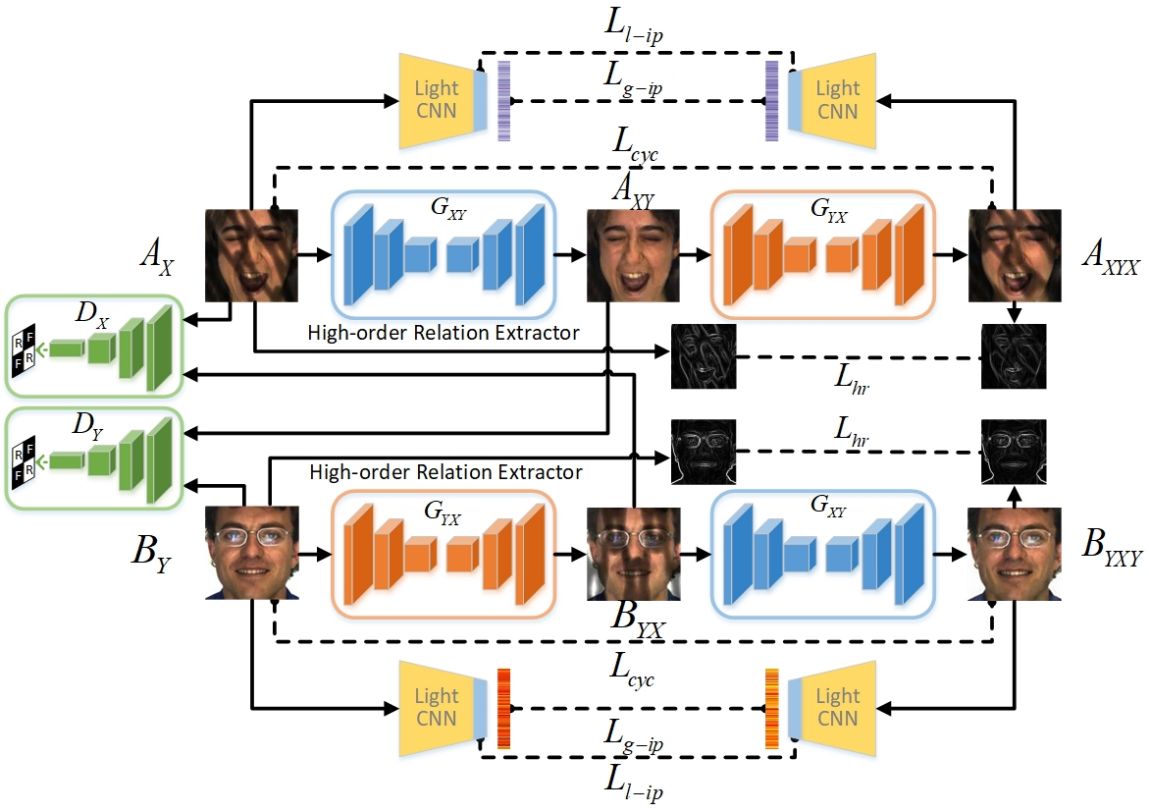


Figure 2. The framework of our proposed HRCycleGAN. Generating and removing MashFace simultaneously using unpaired data, HRCycleGAN relies on CycleGAN to learning the translation between the distributions of MeshFaces and clean faces. A high-order relation loss is further introduced to prevent image smoothness and blurry. Moreover, an identity preserving loss is leveraged to preserve global and local identity information in feature space. Best viewed in color.

To address the aforementioned problems, we propose a High-order Relation-preserving CycleGAN (HRCycleGAN) for generating and removing MeshFaces simultaneously (as shown in Figure 2). Specifically, we treat the MeshFaces generation and removal as two opposite tasks and unify them into a framework by dual learning [5]. First, due to lack of paired data, we rely on CycleGAN [27] to learning a translation between the distributions of MeshFaces and clean faces in pixel space. Second, as the commonly used $L1$ and $L2$ loss usually lead to smoothness and blurry [8], a High-order Relation-preserving loss is developed to enhance image and recover the finer texture details. Finally, Identity Preserving loss is proposed to preserve both global and local identity information. Extensive experimental results demonstrate the effectiveness of the proposed HRCycleGAN.

The contributions of this paper are summarized as follows:

1) We address the blind face inpainting problem with unpaired data from a view of generative model and develop a novel approach for simultaneously MeshFaces generation and removal.

2) We propose a High-order Relation-preserving loss to handle the problem of image smoothness and blurry. It pays more attention to the saliency information, e.g. edge, which derives finer texture details and generates much sharper images.

3) To preserve high level identity information of face images, an Identity Preserving loss is developed. Experimental results show that our HRCycleGAN can remove MeshFaces with identity maintained.

II. Related Work

A. Blind Inpainting

Blind inpainting refers to an image reconstruction process where some missing or corrupted areas on images need to be restored and their locations are not known in advance. Compared with conventional non-blind inpainting where exact positions of missing or damaged pixels are provided, blind inpainting is more difficult because local adjancency information is not available with no position prior. There are some works to handle this problem in recent years. Xie et al. [22] proposed a sparse auto-encoder model to move the imposed texts on images. Authors in [18] introduced

translation variant interpolation into convolutional neural networks (CNNs) for blind inpainting. Yang et al. [23] and Fu et al. [1] focus on image de-rain, which is a special blind inpainting problem. Zhang et al. [25] presented a multi-task CNN arthitecture for blind face inpainting with paired MeshFaces and clean faces. In this paper, we concentrate on the same issue with [25], but our approach is a generative model with adversarial property and addresses unpaired blind face inpainting problem with identity preserving where the MeshFaces are not paired with clean faces.

B. Generative Adversarial Network

Generative Adversarial Network (GAN) [2] was first proposed by Goodfellow et al. in 2014. It originally aims at generating realistic images through a zero-sum game between a generator and a discriminator. The generator produces realistic image samples to fool the discriminator, while the adversarial discriminator tries its best to distingush the generative samples from those real image samples. Mirza et al. [15] introduced the conditional version of GAN, which can generate images under a certain condition such as an image or a class label. Isola et al. [9] used conditional GAN for image-to-image translation in which source images (such as Aerial) are transformed to target images (such as Map). Their model requires paired images to learn such transformtion mapping. However, collecting paired samples is time and labor consuming in pratical application. To address it, Zhu et al. [27] designed a new GAN framework with cycle constraint, named CycleGAN, for unpaired image-to-image translation. Our approach takes CycleGAN as the base model, and transforms between MeshFaces and clean faces in an identity preserving manner, which goes beyond the originally CycleGAN.

III. The Proposed Method

In this section, we describe our proposed method. Specifically, the Cycle-Consistent Generative Adversarial Network (CycleGAN) [27] is introduced firstly. Then we detail the formulation of our proposed method. The architecture of our method is shown in Figure 2.

A. CycleGAN

Generative Adversarial Network (GAN) [2] consists of two networks: a generator G and a discriminator D, which are alternatively trained in a two-player min-max game manner. Generator G is optimized to generate samples as real as possible to fool discriminator D while D is trained to distinguish between the real sample x and the generated one $\hat{x}$. Overall, the objective function of GAN can be formulated as follow:

$$\begin{aligned} L_{GAN}(G,D) = \min_G \max_D E_{y\sim P_{data}(y)}[\log D(y)] \\ +E_{x\sim P_{data}(x)}[\log(1-D(G(x))] \end{aligned} \tag{1}$$

CycleGAN [27] is an extension of GAN [2] for unpaired data, which contains two generators G_{XY}, G_{YX} and two discriminators D_X, D_Y for two domains X, Y respectively. G_{XY} and G_{YX} are $X \to Y$ and $Y \to X$ mappings. The task of CycleGAN is to learn two mappings from domain X to domain Y and from domain Y to domain X simultaneously. Therefore, Cycle Consistency loss is proposed to enforce forward-backward consistency. Mathematically, the objective function of CycleGAN is written as:

$$\begin{aligned} L_{CycleGAN} = L_{GAN}(G_{XY}, D_Y) + L_{GAN}(G_{YX}, D_X) \\ +\lambda L_{cyc}(G_{XY}, G_{YX}) \end{aligned} \tag{2}$$

Where L_{cyc} is the Cycle Consistency loss, defined as below

$$\begin{aligned} L_{cyc} = E_{x\sim P_{data}(x)}[||G_{YX}(G_{XY}(x)) - x||_1] \\ +E_{y\sim P_{data}(y)}[||G_{XY}(G_{YX}(y)) - y||_1] \end{aligned} \tag{3}$$

B. High-order Relation-preserving CycleGAN

We regard the MeshFaces domain and clean face domain as domain X and domain Y respectively. It is noted that data in domain X and domain Y are unpaired. Given a MeshFace image x in domain X, the goal of our model is to generate a clean face image and preserve the identity, and vice versa.

1) High-order Relation-preserving Loss: For each sample x from domain X, CycleGAN restrains that x should be brought back to original domain X through transformation cycle with $L1$ loss to measure the similarity of these two images. However, it is well known that $L1$ and $L2$ loss produce blurry results on image generation problems [9], which are also inconsistent with human visual system [17], [11]. We argue that there exist two main problems in $L1$ and $L2$ loss. First, each pixel intensity in the image is processed separately, but few attention is paid to the relations between pixel intensities. Second, in fact, natural image contains abundant structures (e.g. texture, color, brightness, etc), but such high-order structural relationship between two images is not utilized. In addition, according to the research in [17], [11], human visual system is robust to the subtle variation of pixel and noise. When image changes slightly or is corrupted by noise, human visual neurons are firstly activated by some salient parts (e.g. edge) of objects. Then, based on structural relationship of the image, the changing or corrupted parts are filled through a neuronal filling-in mechanism [17], [11].

For MeshFaces generation and removal, the random patterns in MeshFaces are salient parts, where the corrupted pixels can be restored by comparing them to other pixels in the image that have similar neighborhoods. Therefore, we propose a novel high-order relation-preserving loss, which models the structural relationship between two images rather than each pixel intensity within a single image. It is formulated as follows:

$$L_{hr} = \frac{1}{n}\sum_{i,j}\left|HR(Z_{i,j}) - HR(\hat{Z}_{i,j})\right| \tag{4}$$

Where Z is the ground truth image, $\hat{Z}$ is the generated image by generator, n is the number of pixels in images and $Z_{i,j}$

is the pixel value of the image in position (i,j), which is easy to extend to patches. $HR(\cdot)$ is a high-order relation extractor, which can be defined based on different tasks. In our experiment, the high-order relation extractor is defined as follow:

$$HR(Z_{i,j}) = \frac{1}{2^n}\left[\left|\sum_{k=0}^{n}\binom{n}{k}(-1)^k Z_{i+n-k-1,j}\right| + \left|\sum_{k=0}^{n}\binom{n}{k}(-1)^k Z_{i,j+n-k-1}\right|\right] \tag{5}$$

Where n is the order of the relation. Our high-order loss considers the neighboring spatial dependencies, which is effective to infer the corrupted pixels.

2) Identity Preserving Loss: Although the CycleGAN and the high-order relation loss can generate very realistic and clean face images from MeshFaces in visual perceptual, they cannot ensure that the synthesized image is close to the real data in high-level semantic space. Therefore, we develop an Identity Preserving loss which considers the global and local identity information. We formulate the identity perserving loss as the sum of a global and a local identity perserving loss components as:

$$L_{ip} = L_{g-ip} + L_{l-ip} \tag{6}$$

In the following, the global identity perserving loss L_{g-ip} and the local identity perserving loss L_{l-ip} are described in detail.

A face recognition model maps face images to an embedding space where the $L2$ distance is used to compare the similarity of these images. Our model not only maintains the simialrity in pixel space, but also preserves identity. While achieving particularly high PSNR, CycleGAN based model cannot preserve the identity information. Thus, we propose the global identity preserving loss in high-level semantic space to improve the performance of face recognition and verification, which is described as follows:

$$L_{g-ip}(\hat{Z}, Z) = ||F(\hat{Z}) - F(Z)||_1 \tag{7}$$

Where F denotes the feature extractor network and projects a image to the embedding space. We hope the restored image $\hat{Z}$ is closer to the original input Z in the feature space of F.

As the global identity preserving loss mainly focuses on abstract semantic information, we extend the upper-level feature map constraint proposed in [12] to the finer identity preserving. Introducing an additional supervised constraint is beneficial to improving the convergence speed and the final performance. Therefore, we define the local identity preserving loss as:

$$L_{l-ip}(\hat{Z}, Z) = ||\phi(\hat{Z}) - \phi(Z)||_1 \tag{8}$$

Where ϕ maps images to the upper-level feature maps. This loss can guarantee the finer identity information is close between restored images and the original ones.

3) Overall Objective Function: To keep the balance and stability of training on two domains, a weighted sum of all the losses defined above is added to both domains X, Y. The final objective function is:

$$\begin{aligned} L_{HRCycleGAN} &= L_{GAN}(G_{XY}, D_Y) + L_{GAN}(G_{YX}, D_X) \\ &+\lambda L_{cyc}(G_{XY}, G_{YX}) + \lambda_1 L_{hr} + \lambda_2 L_{g-ip} + \lambda_3 L_{l-ip} \end{aligned} \tag{9}$$

where λ, λ_1, λ_2 and λ_3 are the trade-off parameters. $L_{GAN}(G_{XY}, D_Y)$ and $L_{GAN}(G_{YX}, D_X)$ are used to learn $X \to Y$ and $Y \to X$ mappings. The aim of $L_{cyc}(G_{XY}, G_{YX})$ is to enforce forward-backward consistency. Instead of measuring the pixel-to-pixel distance of two images, we develop the high-level relation-preserving loss L_{hr} by considering the structural relationship between two images. Except for maintaining the relationship of samples in pixel space, the identity preserving loss L_{ip} including global L_{g-ip} and local L_{l-ip} is introduced to preserve identity in the embedding space.

IV. Experiments

In this section, we evaluate the proposed method on three datasets. These datasets and test protocols are briefly introduced firstly. Then, we present the baseline methods and implementation details. At last, a comprehensive experimental analysis is conducted on qualitative synthesis results and quantitative face verification results.

A. Datasets and Protocols

The AR face database [14]. This dataset consists of over 4,000 color images of 126 people. Each people includes frontal view faces of different facial expressions, lighting conditions and occlusions. Moreover, AR provides 130 landmarks for frontal faces with different expressions. In our experiments, we only use the face images with landmark points to obtain normalized face images via landmarks as in [21]. Finally, 112 people with 895 normalized face images are obtained including 56 people with 448 face images for training and the remaining 56 people with 447 face images for testing. 30 random corrupted images are synthesized for each face image on training and test dataset as in [25], resulting in 13,440 MeshFaces for training and 13,410 MeshFaces for testing.

The CMU MultiPIE face database [4]. It contains 750,000 images of 337 people. Each people has multiple images under 15 view points, 19 illumination conditions and 6 different facial expressions. The corresponding 68 landmark points for each image are provided. In our experiments, we select images with frontal view and balanced illumination from MultiPIE. Following the same preprocessing as the AR dataset, frontal face images with landmark annotation are chosen and normalized as in [21], resulting in 337 subjects with 2,403 images. We split it into a training set of 237 subjects with 1,541 images and a testing set of 100 subjects

Method	TPR@FPR=1%		TPR@FPR=0.1%		TPR@FPR=0.01%		PSNR		SSIM	
	AR	MultiPIE	AR	MultiPIE	AR	MultiPIE	AR	MultiPIE	AR	MultiPIE
Corrupted	82.51	60.87	63.48	34.30	41.89	18.87	14.04	15.14	0.7331	0.6973
Clean	**97.44**	**90.94**	**90.54**	**83.60**	**85.68**	**72.05**	-	-	-	-
CycleGAN	87.37	80.58	74.27	61.10	49.77	38.69	21.95	22.88	0.6763	0.8286
RPCycleGAN	90.90	81.92	80.46	63.25	61.89	42.28	24.83	23.47	0.9221	0.8857
IPCycleGAN	93.55	83.31	82.92	66.48	72.48	42.34	25.25	22.59	0.9246	0.8484
HRCycleGAN	**95.35**	**87.95**	**87.01**	**75.62**	**74.32**	**59.62**	**26.89**	**24.44**	**0.9453**	**0.8771**

Table I
VERIFICATION PERFORMANCE ON AR AND MULTIPIE DATASETS AND INPAINTING RESULTS ON THE TESTING SET

with 862 images. Similar to [25], each clear face image is used to synthesize 30 totally different and random mesh face images. As a result, 46,230 mesh face images are used as training and 25,860 mesh face images for testing.

The LFW face database [7]. Besides face images under constrained conditions, we also conduct an experiment in the wild. LFW dataset is a standard test set for verification in unconstrained conditions, which contains 13,233 images of 5,749 people. Firstly, 5 facial points are extracted by MTCNN [24], then faces are normalized to two eyes points being horizontal as in [21]. Following the verification protocol [7], 6,000 face pairs with 3,000 positive pairs and 3,000 negative pairs are provided to evaluate our model. It is worth noting that our model is not trained on the LFW dataset. That is to say, LFW dataset is only used as the test set.

In the testing phase, the frontal view and neutral expression face image of each individual is chosen as the gallery set and the remaining face images are randomly corrupted as the probe set on AR and MultiPIE dataset. For each pair of images on LFW, one image of them is randomly selected and corrupted as a MeshFace. The face verification performance between the gallery set and recovered clean faces is compared through qualitative and quantitative analysis. The TPR@FPR=1% (true positive rate when false positive rate is 1%), TPR@FPR=0.1% and TPR@FPR=0.01% are used as evaluation criteria of face verification. The PSNR [dB] and SSIM [20] are reported to evaluate generated image quality.

B. Baselines and Implementation Details

To the best of our knowledge, this work makes the first attempt to remove mesh and generate mesh simultaneously using unpaired data. Although [25], [26] propose their methods to recover a clear face image from a corrupted one, their methods need to be trained on paired data. Therefore, our method cannot compare with them. In our work, we will evaluate each part of the proposed model through designing various configurations including CycleGAN (CycleGAN), CycleGAN with the high-order relation-preserving loss (RP-CycleGAN), CycleGAN with the identity preserving loss (IPCycleGAN). All three variable configurations are based on CycleGAN for blind face inpainting task.

The feature extraction network is pre-trained on the MS-Celeb-1M dataset followed by the instructions in [21]. All the face images are rotated to two eye points to be horizontal to overcome the pose variations [21], and then resized to 148x148. Finally, they are randomly cropped into 128x128 and mirrored with 0.5 probability for data augmentation. For all experiments, all networks are trained using ADAM solver [10] with batch size 16 and an initial learning rate of 0.0002 for the first 50 epochs. The learning rate is linearly decayed over the next 50 epochs. The trade-off parameters λ, λ_1, λ_2 and λ_3 are assigned to 1.0. The order of the relation in $HR(\cdot)$ is set to 1. Following the way of generating random patterns in [25], we add a random mask to the clean photo to generate a MeshFace.

Our generative networks take the architecture of ResNet [6] with 3 residual blocks and discriminator contains 4 convolutional layers. In the identity preserving module, we employ the light CNN-29 [21] as feature extractor with weights fixed during training, which includes 29 convolution layers, 4 max-pooling and one fully-connected layer. The output of the fully-connected layer is chosen as the global identity information with 256 dimensions. We select the ouput of the last pool layer as the local identity. All the experiments are conducted with PyTorch framework on a single GTX Titan X GPU.

C. Evaluation of Verification Results

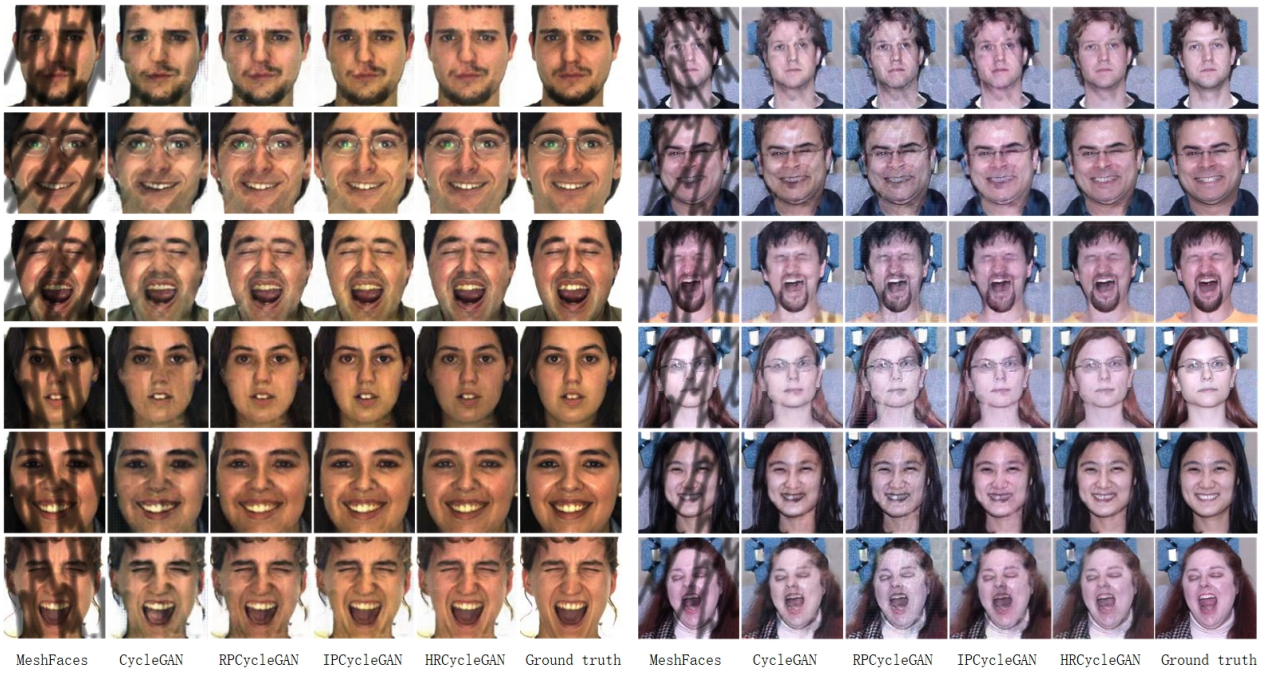
We evaluate our approach on three datasets, i.e., AR dataset, MultiPIE dataset and LFW dataset. The recovered face photos are employed on face verification task according to the aforementioned test protocol. Moreover, the verification performances of clean faces and MeshFaces are presented for fair comparison.

From Table I, we can observe that using MeshFaces for verification directly declines sharply in performance on all datasets owing to random mesh. Obviously, when MeshFaces are processed by blind face inpainting model, the face verification performance on recovered faces is significantly improved. This is because HRCycleGAN pulls the distribution of recovered face images into clear face images by adversarial learning procedure.

Experiments on AR dataset and MultiPIE dataset are under constrained condition. From Table I, it is obvious

Method	TPR@FPR=1%	TPR@FPR=0.1%	TPR@FPR=0.01%	PSNR	SSIM
Corrupted	59.93	28.40	7.47	14.02	0.4131
Clean	**93.73**	**92.53**	**89.67**	-	-
CycleGAN	73.57	58.63	29.20	15.82	0.4470
RPCycleGAN	77.53	63.47	35.87	16.54	0.4690
IPCycleGAN	80.87	61.50	39.57	17.04	0.5152
HRCycleGAN	**81.90**	**64.73**	**48.03**	**17.14**	**0.6831**

Table II
VERIFICATION PERFORMANCE ON LFW DATASETS AND INPAINTING RESULTS ON THE TESTING SET. NOTE THAT OUR MODEL IS TRAINED ON MULTIPIE DATASET, BUT DIRECTLY TESTED ON LFW DATASET.



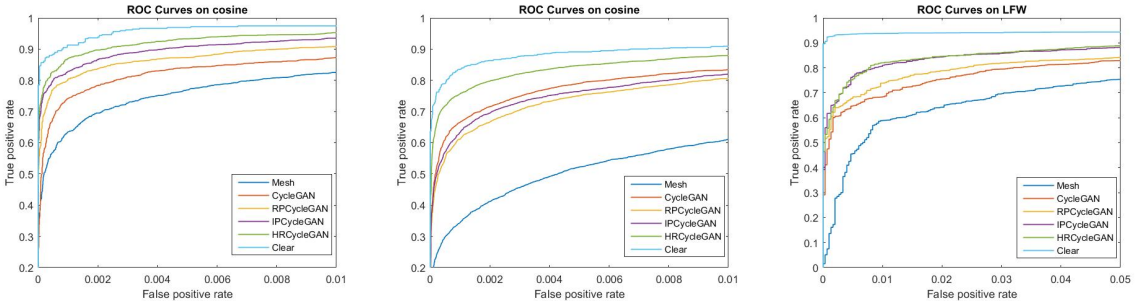
(a) AR dataset (b) MultiPIE dataset

Figure 3. A visual comparison of the inpainting results on the test set of AR and MultiPIE dataset.

that CycleGAN based mothods earn big advantages over the corrupted one on two datasets. Especially on MultiPIE, CycleGAN has improvement about 20% at TPR@FPR=1%, 30% at TPR@FPR=0.1% and 20% at TPR@FPR=0.01%, but is inferior to RPCycleGAN and IPCycleGAN. The reason is that RPCycleGAN considers the contextual similarity instead of pixel-to-pixel similarity. For random wavy lines in MeshFaces, contextual information is useful to recover clean faces. While IPCycleGAN can grasp more discriminative identity information than CycleGAN, which makes recovered face images by IPCycleGAN closer to the clean faces in a low-dimensional feature space. Our proposed HRCycleGAN achieves state-of-the-art result on both datasets, almost doubling the TPR@FPR=0.1% and even three times at TPR@FPR=0.01% on MultiPIE. Compared with CycleGAN, HRCycleGAN surpasses about 8% at TPR@FPR=1%, 13% at TPR@FPR=0.1% and 25% at TPR@FPR=0.01%. That means our proposed HRCycleGAN not only restores clean face images, but also preserves identity.

To evaluate the generalization capability of our approach, we directly perform blind face inpainting task on LFW dataset under unconstrained condition using model trained on MultiPIE dataset. From Table II, we can observe that the performance on LFW dataset is a little lower than on MultiPIE dataset, about 22% improvement at TPR@FPR=1%, doubling improvement at TPR@FPR=0.1% and seven times improvement at TPR@FPR=0.01%. The underlying reasons may be the cross-database test. Additionally, low-resolution and large-pose variances further make it harder to recovery clean face images.

In addition, the corresponding ROC curves for three datasets are also plotted in Figure 4. It is worth noting that the performance on clean face images is an upper limit



(a) AR dataset (b) MultiPIE dataset (c) LFW dataset

Figure 4. ROC curves for AR, MultiPIE and LFW dataset.

for face verification. From Table I, although we train our model on unpaired data, the gap between our model and clear face image at TPR@FPR=1% is very small (about 3% for MultiPIE dataset, 2% for AR dataset), which validates the effectiveness of the proposed method.

D. Evaluation of MeshFace Generation and Removal

In this section, we conduct extensive quantitative and qualitative evaluations on AR [14], MultiPIE [4] and LFW [7] dataset. For MeshFaces removal, Figure 3 presents some visual results of the compared models. It is observed that CycleGAN based models can generally recover most of the corrupted areas. Because CycleGAN based models rely on adversarial learning to push the distribution of recovered face images to the real distribution. In comparison, RPCygleGAN generates much cleaner and sharper images than simple CycleGAN. It suggests that just using CycleGAN cannot recover all corrupted areas completely, while RPCycleGAN with the high-order preserving loss can handle all the corrupted areas well because it measures the similarity between neighboring pixels instead of single pixel. However, as IPCycleGAN with identity preserving loss in feature space focuses on high level semantic information, it is so robust to pixel level changes that contain more artifacts than HRCycleGAN. As a whole, our model can recover more clean and photorealistic images than other compared methods due to considering identity and high-order relation information jointly.

Furthermore, we use the metrics PSNR and SSIM [20] to quantitatively evaluate the recovered face image quality, where higher values of PSNR and SSIM indicate better results. As is showed in Table I, the quantitative results are consistent to our visual perception.

Generating random meshes is equally important in particular field (e.g. information proctection), which is also a method to measure whether model has learnt the random patterns. Figure 5 shows some visual results of the proposed method. It can be observed that our model can generate more random patterns without undermining the subject information of image, which has never occurred in training set. It demonstrates that our model learns the distribution of the random patterns instead of just remembering those patterns in training set.

Figure 5. Visual inspection of MeshFace generation on the test set of AR dataset.

Figure 6. Some failure examples.

Failure Case. Although our method has achieved com-

pelling results, there exist several failure cases shown in Figure 6. In the top row, the corrupted areas on clothes and beard are not recovered. One possible reason is that because of the color of mesh is similar to the beard, our method hardly detects the corrupted regions. The bottom row shows the color deviation, which may result from inconsistency of color distribution among MultiPIE and LFW dataset.

V. Conclusion

In this paper, we have proposed a novel method for generating and removing MeshFaces simultaneously using unpaired data. CycleGAN is employed to learn a transformation between the distributions of MeshFaces and clean faces. To generate finer texture details and much sharper images, a high-order relation-preserving loss is introduced to prevent image smoothness and blurry. Moreover, an identity preserving loss is developed to preserve global and local identity information. Experimental results on three datasets have demonstrated the effectiveness of the proposed method for unpaired data.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (Grant No. 61622310, 61473289) and State Key Development Program (Grant No. 2016YFB1001001).

References

[1] X. Fu, J. Huang, D. Zeng, Y. Huang, X. Ding, and J. Paisley. Removing rain from single images via a deep detail network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[2] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Proceedings of the Annual Conference on Neural Information Processing Systems*, pages 2672–2680, 2014.

[3] I. J. Goodfellow, J. Shlens, and C. Szegedy. Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572*, 2014.

[4] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-pie. *Image and Vision Computing*, 28(5):807–813, 2010.

[5] D. He, Y. Xia, T. Qin, L. Wang, N. Yu, T. Liu, and W.-Y. Ma. Dual learning for machine translation. In *Proceedings of the Annual Conference on Neural Information Processing Systems*, pages 820–828, 2016.

[6] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[7] 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. Technical report, University of Massachusetts, Amherst, 2007.

[8] H. Huang, R. He, Z. Sun, and T. Tan. Wavelet-srnet: A wavelet-based cnn for multi-scale face super resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1689–1697, 2017.

[9] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017.

[10] D. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[11] H. Komatsu. The neural mechanisms of perceptual filling-in. *Nature Reviews. Neuroscience*, 7(3):220, 2006.

[12] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017.

[13] Z. Li, Y. Hu, and R. He. Learning disentangling and fusing networks for face completion under structured occlusions. *arXiv preprint arXiv:1712.04646*, 2017.

[14] A. Martinez and R. Benavente. The ar face database. *CVC technical report*, 1998.

[15] M. Mirza and S. Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.

[16] A. Nguyen, J. Yosinski, and J. Clune. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 427–436, 2015.

[17] M. A. Paradiso and K. Nakayama. Brightness perception and filling-in. *Vision research*, 31(7):1221–1236, 1991.

[18] J. S. Ren, L. Xu, Q. Yan, and W. Sun. Shepard convolutional neural networks. In *Proceedings of the Annual Conference on Neural Information Processing Systems*, pages 901–909, 2015.

[19] K. Wang, R. He, L. Wang, W. Wang, and T. Tan. Joint feature selection and subspace learning for cross-modal retrieval. *IEEE transactions on pattern analysis and machine intelligence*, pages 2010–2023, 2016.

[20] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

[21] X. Wu, R. He, Z. Sun, and T. Tan. A light cnn for deep face representation with noisy labels. *arXiv preprint arXiv:1511.02683*, 2015.

[22] J. Xie, L. Xu, and E. Chen. Image denoising and inpainting with deep neural networks. In *Proceedings of the Annual Conference on Neural Information Processing Systems*, pages 341–349, 2012.

[23] W. Yang, R. T. Tan, J. Feng, J. Liu, G. Zongming, and S. Yan. Joint rain detection and removal from a single image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[24] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10):1499–1503, 2016.

[25] S. Zhang, R. He, Z. Sun, and T. Tan. Multi-task convnet for blind face inpainting with application to face verification. In *Proceedings of the IEEE International Conference on Biometrics*, pages 1–8, 2016.

[26] S. Zhang, R. He, and T. Tan. Demeshnet: Blind face inpainting for deep meshface verification. *IEEE Transactions on Information Forensics and Security*, 2017.

[27] 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.