# Semi-Cycled Generative Adversarial Networks for Real-World Face Super-Resolution

Hao Hou, Jun Xu, *Member, IEEE*, Yingkun Hou, *Senior Member, IEEE*, Xiaotao Hu, Benzheng Wei, and Dinggang Shen, *Fellow, IEEE*

***Abstract*— Real-world face super-resolution (SR) is a highly ill-posed image restoration task. The fully-cycled Cycle-GAN architecture is widely employed to achieve promising performance on face SR, but is prone to produce artifacts upon challenging cases in real-world scenarios, since joint participation in the same degradation branch will impact final performance due to huge domain gap between real-world and synthetic LR ones obtained by generators. To better exploit the powerful generative capability of GAN for real-world face SR, in this paper, we establish two independent degradation branches in the forward and backward cycle-consistent reconstruction processes, respectively, while the two processes share the same restoration branch. Our Semi-Cycled Generative Adversarial Networks (SCGAN) is able to alleviate the adverse effects of the domain gap between the real-world LR face images and the synthetic LR ones, and to achieve accurate and robust face SR performance by the shared restoration branch regularized by both the forward and backward cycle-consistent learning processes. Experiments on two synthetic and two real-world datasets demonstrate that, our SCGAN outperforms the state-of-the-art methods on recovering the face structures/details and quantitative metrics for real-world face SR. The code will be publicly released at https://github.com/HaoHou-98/SCGAN.**

***Index Terms*— Real-world face super-resolution, semi-cycled architecture, cycle-consistent generative adversarial networks.**

## I. Introduction

FACE is of central importance for human identity recognition. The low-resolution (LR) face images captured by camera sensors would largely degrade the corresponding identity information. Face super-resolution (SR) aims to estimate high-resolution (HR) face images from LR ones, to improve the image quality and performance of subsequent identity recognition tasks [1], [2], [3]. This task is very challenging upon complex real-world scenarios, where the degradation kernel is usually unknown. Traditional face SR methods can be roughly divided into local patch-based methods [4], [5], [6], global image-based methods [7], [8], [9], and hybrid methods taking advantage of global image consistency and local patch sparsity [10], [11], [12], [13]. However, these hand-crafted methods could hardly achieve satisfactory results upon diverse real-world scenarios [14].

Recently, the powerful learning capability of deep convolutional neural networks (CNNs) has been extensively exploited for face SR [16], [17], [18], [19], [20]. These discriminative CNNs mainly learn a direct enhancing mapping function between pairs of LR and HR face images. For objective evaluation, the LR face images are usually degraded by synthetic downsampling kernels from the HR ones. However, since it is difficult to obtain the corresponding HR face images for the real-world LR ones, the discriminative CNNs suffer from a huge performance gap between synthetic and practical degradation for real-world face SR. To this end, several methods [21], [22], [23] align LR face images with unpaired HR face images with the same identity. However, face alignment is often challenged by the insufficiently trained face SR models, due to short of HR face images in practical scenarios.

Compared to the discriminative competitors, generative CNNs like Generative Adversarial Networks (GANs) [24] are employed in [21], [25], [26], [27], [28], and [14] to perform blind face SR with complex degradations. To deal with unknown real-world degradation, several generative CNNs [15], [29], [30], [31] further implement unsupervised face SR by resorting to the insight of cycle-consistency developed for the unpaired image translation tasks [32]. LRGAN [15] is a representative work to utilize the cycle learning scheme [32] for real-world face SR, introducing a “learning-to-degrade” branch and a “learning-to-SR” branch to perform face image degradation and SR, respectively. However, since unpaired LR and HR face images suffer from a considerable gap on identity information, the two branches in LRGAN are consistent only for the HR face images and could hardly preserve well the face details and identity information of the LR face images.

Manuscript received 8 September 2022; revised 30 December 2022; accepted 20 January 2023. Date of publication 3 February 2023; date of current version 10 February 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 62002176, Grant 62176068, and Grant 61872225; and in part by the Natural Science Foundation of Shandong Province under Grant ZR2020MF038, Grant ZR2020KF013, and Grant ZR2020ZD44. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Diego Valsesia. *(Corresponding author: Jun Xu.)*

Hao Hou is with the College of Intelligence and Information Engineering and the Center for Medical Artificial Intelligence, Shandong University of Traditional Chinese Medicine, Jinan 250355, China.

Jun Xu is with the School of Statistics and Data Science, KLMDASR, LEBPS, and LPMC, Nankai University, Tianjin 300071, China, and also with the School of Biomedical Engineering, ShanghaiTech University, Shanghai 201210, China (e-mail: nankaimathxujun@gmail.com).

Yingkun Hou is with the School of Information Science and Technology, Taishan University, Taian 271000, China.

Xiaotao Hu is with the College of Computer Science, Nankai University, Tianjin 300071, China.

Benzheng Wei is with the Center for Medical Artificial Intelligence, Shandong University of Traditional Chinese Medicine, Jinan 250355, China.

Dinggang Shen is with the School of Biomedical Engineering, ShanghaiTech University, Shanghai 201210, China, also with Shanghai United Imaging Intelligence Co., Ltd., Shanghai 200230, China, and also with the Shanghai Clinical Research and Trial Center, Shanghai 201210, China.

This article has supplementary downloadable material available at https://doi.org/10.1109/TIP.2023.3240845, provided by the authors.

Digital Object Identifier 10.1109/TIP.2023.3240845

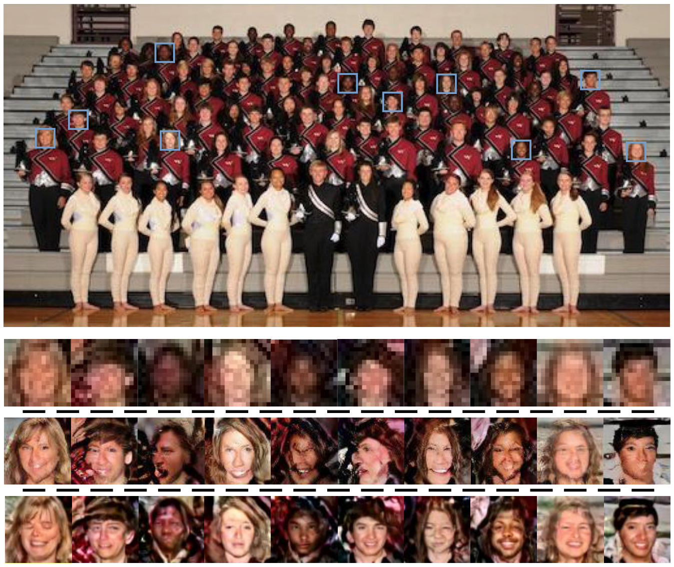


Fig. 1. Comparison of LRGAN [15] and our SCGAN on blind real-world face SR. We perform ×4 real-world face SR on $16 \times 16$ LR face images to obtain $64 \times 64$ HR ones. 1-st row: a real-world group photo crawled from the internet that suffers from complex and unknown degradation. 2-nd row: the LR face images from the photo. 3-rd row: the face SR results of LRGAN [15]. 4-th row: the SR results by our SCGAN.

Since the unpaired LR and HR face images suffer from uncertain relationship, employing a directional framework [15] or a fully-cycled bidirectional one [32] is not sufficient to simultaneously preserve the identity information of the LR and HR face images in real-world scenarios. To better alleviate the domain gap between unpaired LR and HR face images, in this paper, we introduce a Semi-Cycled Generative Adversarial Network (SCGAN) for real-world face SR, by extending the bidirectional cycle consistency scheme in [32] to a more flexible version. Specifically, we propose to learn three generative branches, instead of two in [32] and [15], for real-world HR and LR face image reconstructions: 1) a "learning-to-degrade" branch to obtain synthetic LR face images by degrading the HR ones, 2) a "learning-to-SR" branch to obtain the SR images by restoring the synthetic and real-world LR face images, and 3) another "learning-to-degrade" branch to degrade the SR images restored from the real-world LR images. Different from CycleGAN [32], our SCGAN is only coupled at the middle "learning-to-SR" branch, while learning the cycle consistency of LR and HR face image reconstructions by individual branches. For example, in Figure 1, we compare the real-world face SR performance between LRGAN [15] and our SCGAN. The real-world LR face images with severe degradation could hardly be restored by LRGAN to recover the identity structure and details. However, our SCGAN, benefited from the semi-cycle consistency insight, well preserves both aspects for face SR.

In summary, our contributions are mainly three-fold:

- **We develop a novel Semi-Cycled architecture to exploit GANs for real-world face super-resolution**. Our proposed Semi-Cycled GANs (SCGAN) well mitigate adverse effects of the degradation gap between real-world LR face images and synthetic ones, resulting in better preservation of identity and detailed information.
- **We study in-depth the roles of adversarial loss, pixel loss, and cycle-consistency loss** in our SCGAN for face image super-resolution. The adversarial loss reduces the domain gap between the HR images and those outputs by our SCGAN, and the pixel loss enriches the contextual details of the SR results, while the cycle-consistency loss helps to preserve the structural information.
- Experiments on five benchmark datasets show that **our SCGAN outperforms the state-of-the-art methods quantitatively and qualitatively** on real-world face SR. Application on downstream vision tasks of face detection, face verification and face landmark detection further validates the effectiveness of our SCGAN on face SR.

The rest of this paper is organized as follows. In §II, we summarize the related works. In §III, we introduce our SCGAN for real-world face SR. In §IV, experiments on benchmark datasets are conducted to evaluate the performance of different face SR methods. §V concludes this paper.

## II. Related Work

### A. Human Face Super-Resolution

Human face super-resolution (SR) aims to obtain visual-pleasing high-resolution (HR) face images from the low-resolution ones [33], [34]. Early face SR methods [4], [5], [6], [35], [36], [37] utilize hand-craft image priors and degradation models. For instance, Baker and Kanade [35] utilized Gaussian image pyramids for face SR, while Gunturk et al. [36] presented a Bayesian model for face SR from a global image-level perspective. To well recover local details, the methods in [4], [6], and [5] tackle the face SR by patch-wise modeling. Neighborhood embedding [6] is a representative work in this direction. Later, the methods of [10], [11], [12], [37], and [13] have been developed for face SR to simultaneously preserve local details and global structures. However, these methods do not perform well upon complex real-world cases.

Recent methods [16], [17], [18], [19], [20] employ deep convolutional neural networks (CNNs) for face SR. RBPNet [19] employs iterative back projection to directly learn the mapping from LR to HR face images. SPARNet [18] integrates the spatial attention mechanism into their framework to improve the representation ability of the network. WaSRNet [20] transforms the face image domain into the wavelet coefficient domain to preserve more details. Lu et al. [38] proposed a hybrid approach based on a global upsampling network and a local enhancement network to jointly enhance the facial contours and local details. The Residual Attribute Attention Network [39] employs a multi-block cascaded structure to extract pixel-level representation and semantic-level identification information from LR face images and restores high-resolution images via efficient feature fusion. The Facial Attribute Capsule Network [40] converts the extracted LR face image features into a set of facial attribute capsules by the proposed capsule generation block and utilizes the facial attribute information from both semantic space and probability space to generate the corresponding HR results. However, since they are trained on synthetic images, these discriminative learning based methods cannot be well generalized to real-world scenarios.

Generative models like Generative Adversarial Networks (GANs) [24] have achieved remarkable progress on face SR [15], [29], [30], [31], [43]. URDGN [44] is among the first work in this direction, but sensitive to the LR face images with large face rotations or poses. To alleviate this problem, Super-FAN [45] locates the key points of faces via heat map regression to deal with faces in different angles and poses, which needs large-scale annotations of face landmarks for model training. LRGAN [15] is an unsupervised face SR network by utilizing the architecture of cycle consistency [32]. But this method only exploits the consistency within the HR face images while ignoring the consistency within the LR ones. PULSE [27] often loses spatial information and identity consistency of face images, by randomly sampling the low-dimensional latent codes. The methods of GLEAN [46], GFPGAN [28], and GPEN [14] utilize a pre-trained StyleGAN [47] model for face SR, but show limited performance on LR face images with severe degradation. In this work, we propose to learn three forward or backward mappings, *i.e.*, two independent "learning-to-degrade" branches and one shared "learning-to-SR" branch, which are semi-cycled to maintain the consistency of both the HR and LR face image reconstructions.

### B. Generative Adversarial Networks

Generative Adversarial Networks (GANs) [24] have been widely utilized in unsupervised computer vision tasks with great success [15], [29], [30], [31], [32], [43], [48], [49], [50], [51], [52], [53]. InfoGAN [48] learns explainable feature representation by decomposing the input noise vector into incompressible noise and latent codes, to control semantic features of the generated images. Conditional GAN (cGAN) [49] adds to the original GAN an extra training supervision, achieving great success on image translation tasks [50], [51]. With the insight of cycle consistency, the methods of CycleGAN [32], DualGAN [52], and DiscoGAN [53] achieve promising performance on image translation tasks. This insight has also been resorted by many image restoration methods [15], [29], [30], [31], [43], [54]. Among them, LRGAN [15] introduces two cycle-consistent generators [32] for face SR: a "learning-to-degrade" branch for HR image degradation and a "learning-to-SR" branch for LR face image super-resolution. However, the two branches are only coupled for HR face image reconstruction, bringing a potential gap between unpaired LR and HR face images. In this work, we also exploit the powerful generative capability of CycleGAN [32] for unsupervised real-world face SR. Built upon LRGAN [15], our SCGAN introduces an additional "learning-to-degrade" branch to degrade the super-resolved face images, which are supervised by the real-world LR ones.

### C. Cycle-Consistent Learning

The framework of cycle-consistent learning has been developed originally for image-to-image translation [32] to jointly learn a paired of coupled branches under the process of backward domain transfer. From then on, researchers have exploited the cycle-consistent learning framework for many vision tasks such as image restoration [15], [29], [30], [31], [43]. For example, the methods of [29] and [43] simultaneously perform degradation on the LR images and also restoration on the degraded LR images with pseudo-supervision.

CinCGAN [55] first uses an inner CycleGAN to map a noisy LR image into a clean LR image, and then uses an outer CycleGAN to map the clean LR image into an HR one. MCinCGAN [56] obtains SR results with different upsampling factors by adjusting the number of recurrent GAN models. Lugmayr et al. [57] first utilizes the cycle consistency model to degenerate the HR image into a simulated LR image and then forms data pairs for supervised learning of the SR network. Similar ideas have also been studied in [29], [58], and [43]. Reference [59] replaces the adversarial loss in CycleGAN [32] with a dual back-projection loss to form an internal learning framework.

Guo et al. [30] introduced a U-Net like cycle-consistent network for image super-resolution, while Zhang et al. [31] employed the cycle-consistent learning for image deblurring. Yi et al. [54] proposed an asymmetric cycle-consistent architecture for face portrait line drawing, which is improved by a pre-trained Inception-V3 [60] under a knowledge distillation scheme [61]. The method of [62] employs the cycle-consistent architecture to firstly convert the input real-world LR image into a pseudo-clean LR image, and then produce the final SR result. Differently, our SCGAN simultaneously restores the real-world and synthetic LR face images to real-world HR face images.

Cycle-consistent learning is also effective for face SR. LRGAN [15] first learns to degrade the real-world HR face images to the synthetic LR ones by a "learning-to-degrade" sub-network, and then learns to restore the synthetic/real-world LR face images to the corresponding SR ones by a "learning-to-SR" sub-network. In this paper, we also employ the cycle-consistent learning framework. But different from fully-cycled CycleGAN, our semi-cycled SCGAN alleviates the adverse gap between real-world HR face images and SR ones by establishing independent degradation mappings.

## III. Proposed Method

In this section, we introduce the motivation of our Semi-Cycled Generative Adversarial Networks (SCGAN) for unsupervised face super-resolution (SR) in §III-A. Then we overview our SCGAN in §III-B. We present three degradation and restoration branches in §III-C, §III-D, and §III-E, respectively. Finally, the implementation details are provided in §III-F.

### A. Motivation

Our goal is to super-resolve real-world low-resolution (LR) face images into the identity preserving high-resolution (HR) face images, without the corresponding paired real-world HR face images. This task can be suitably tackled under the unsupervised cycle-consistent framework like CycleGAN [32].

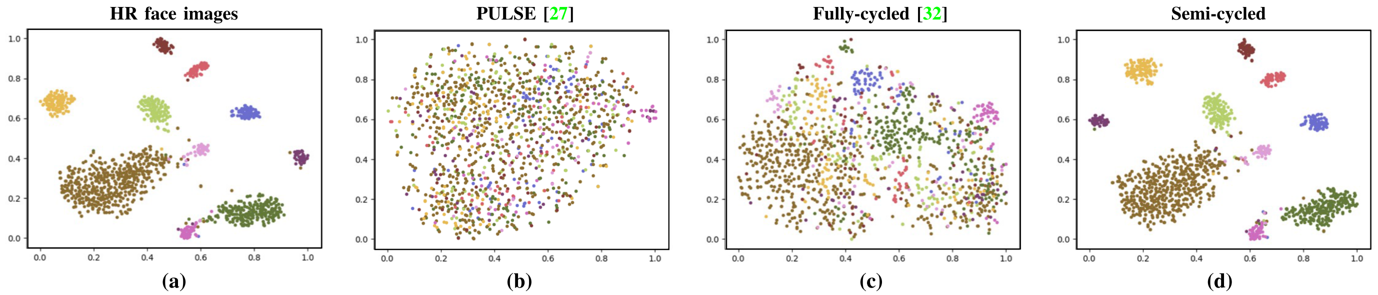


Fig. 2. Distributions of the feature maps extracted by ResNet-101 [41] from HR and SR face images using t-SNE [42]. (a) Visualization of the HR face images. (b) Visualization of the SR face images restored by state-of-the-art face SR method PULSE [27]. (c) Visualization of the SR face images restored by fully-cycled CycleGAN [32]. (d) Visualization of the SR face images restored by our semi-cycled SCGAN. Our semi-cycled architecture better retains the feature maps of the SR face images compared to the fully-cycled CycleGAN.

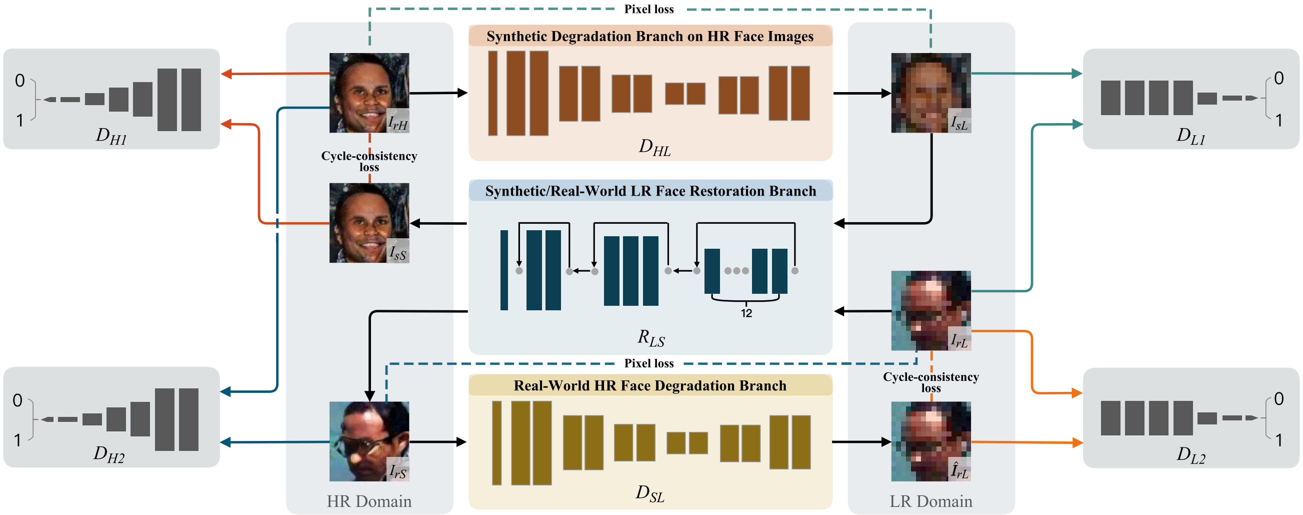


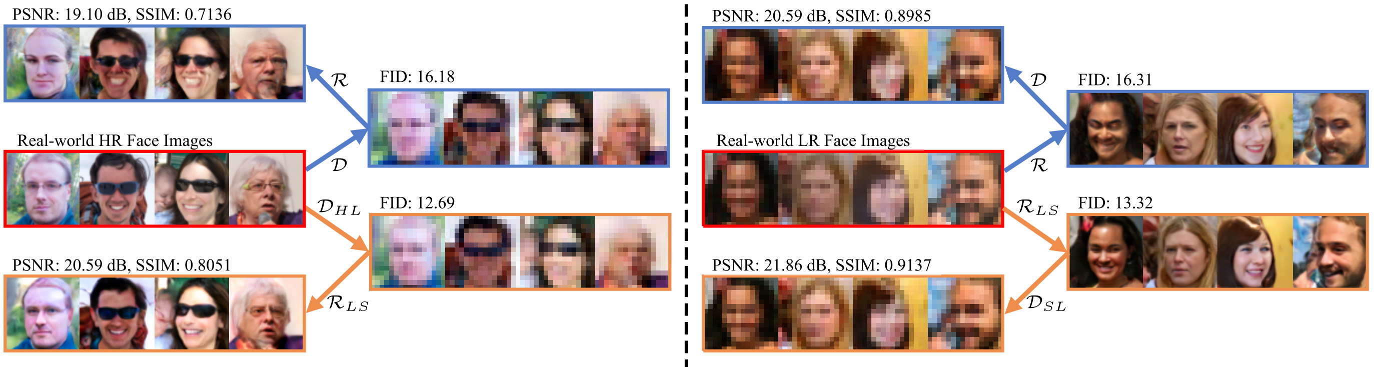
Fig. 3. Architecture of our Semi-Cycled Generative Adversarial Network (SCGAN) for unsupervised face super resolution. Given a real-world HR face image $\mathbf{I}_{rH}$, we first perform image degradation through the HR face degradation branch $\mathcal{D}_{HL}$ and compute the pixel loss between the downsampled $\mathbf{I}_{rH}$ and the obtained $\mathbf{I}_{sL}$ to preserve more details. Then $\mathbf{I}_{sL}$ is sent to the sub-network $R_{LS}$ to perform SR to obtain $\mathbf{I}_{sS}$. Here, we calculate the cycle consistency loss for $\mathbf{I}_{rH}$ and $\mathbf{I}_{sS}$ to maintain their identity consistency. The above process forms a forward cycle consistent GAN model. Among them, $\mathcal{D}_{L1}$ is responsible for distinguishing $\mathbf{I}_{sL}$ and unpaired input LR face image $\mathbf{I}_{rL}$, $\mathcal{D}_{H1}$ is responsible for distinguishing $\mathbf{I}_{rH}$ and paired SR result $\mathbf{I}_{sS}$. The backward cycle consistent GAN model is similar to the forward one. The most important difference with CycleGAN [32] is that, the backward model has an independent degradation sub-network $\mathcal{D}_{SL}$, while not in the fully-cycled situation.

With two fully-cycled generators, CycleGAN well preserves the consistency within the bidirectional translation between two different image domains. However, the fully-cycled CycleGAN is prone to get stuck upon real-world unsupervised face SR with unpaired LR and HR face images, since the complex degradation in real-world LR face images can hardly be well simulated by the generator simultaneously synthesizing the HR face degradation. Therefore, directly employing the fully-cycled architecture for real-world face SR inadvertently suffers from an inevitable problem on the degradation gap between synthetic LR images and real-world LR images. To address this problem, it is natural to model the synthetic and real-world degradations by different generators. To this end, our SCGAN is developed with two different degradation branches and one restoration branch to learn semi-cycled forward and backward cycle-consistent reconstruction processes. Our SCGAN is more flexible than the fully-cycled architecture with more accurate unsupervised real-world face SR performance. Besides, the two independent degradation branches in our SCGAN further facilitate our SCGAN to learn a stronger restoration branch for LR face image, making our SCGAN very robust on super-resolving real-world LR face images.

To globally compare our semi-cycled SCGAN with the fully-cycled CycleGAN, we perform degradation and restoration on HR face images of 10 identities from the LFW dataset [63], by the same strategy mentioned above. In Figure 2, we visualize the input HR face images, and the HR face images restored by PULSE [27], CycleGAN [32] and our SCGAN via t-SNE [42], using the one-dimensional vectors output by the last fully connected layer of a pre-trained Resnet-101 [41]. One can see that the distribution of HR face images restored by our SCGAN is more consistent than those restored by PULSE or CycleGAN, with the distribution of input HR face images. This validates the superiority of our SCGAN over the fully-cycled CycleGAN on the identity preservation of HR face image restoration.

### B. Network Overview

Our SCGAN contains two semi-cycled sub-networks consisting of two independent degradation branches coupled by a restoration branch. The overall network architecture is illustrated in Figure 3. The synthetic degradation branch $\mathcal{D}_{HL}$ and the restoration branch $\mathcal{R}_{LS}$ together perform forward cycle-consistent HR face image reconstruction, while the



(a) Degradation ⟶ Restoration

(b) Restoration ⟶ Degradation

Fig. 4. Comparison of the degradation and restoration results between the fully-cycled CycleGAN [32] and the proposed semi-cycled SCGAN. The input images are circled with red borders. (a) The "degradation-restoration" process of the real-world HR face image. CycleGAN (or Our SCGAN) degrades a real-world HR face image by a degradation branch $\mathcal{D}$ (or $\mathcal{D}_{HL}$) and restores the degraded image by a restoration branch $\mathcal{R}$ (or $\mathcal{R}_{LS}$). The FID (lower is better) is calculated between the degraded LR face image and the real-world LR face image, the PSNR and SSIM (higher is better) are calculated between the restored HR face image and the original HR face image. (b) The "restoration-degradation" process of the real-world LR face image. CycleGAN (or Our SCGAN) restores a real-world LR face image by a restoration branch $\mathcal{R}$ (or $\mathcal{R}_{LS}$) and degrades the restored image by a degradation branch $\mathcal{D}$ (or $\mathcal{D}_{SL}$). The FID is calculated between the restored HR face image and the real HR face image, and the PSNR and SSIM are calculated between the degraded LR face image and the original LR face image. Please zoom in for better view.

restoration branch $\mathcal{R}_{LS}$ and the real-world degradation branch $\mathcal{D}_{SL}$ together implement the backward cycle-consistent LR face image reconstruction. The two reconstruction sub-networks are semi-cycled to avoid the adverse effects of the domain gap between the synthetic and realistic LR face images, and to achieve robust yet accurate face SR performance.

*1) Synthetic HR Image Degradation Branch:* The HR face image degradation branch, denoted as $\mathcal{D}_{HL}$, degrades an HR face image $\mathbf{I}_{rH}$ to a synthetic LR face image. It is the degradation stage of the forward cycle-consistency learning process $\mathbf{I}_{rH} \rightarrow \mathcal{D}_{HL}(\mathbf{I}_{rH}) \rightarrow \mathbf{I}_{rH}$, in which the corresponding restoration stage is implemented by the LR face image restoration branch $\mathcal{R}_{LS}$ introduced as follows.

*2) LR Face Restoration Branch:* This branch is to enhance the quality of the synthetic LR face image $\mathbf{I}_{sL}$ generated by previous degradation branch $\mathcal{D}_{HL}$ and the real-world LR face image $\mathbf{I}_{rL}$ that is the input in the test stage. The restoration of synthetic LR face image comprises the forward cycle-consistent learning process "$\mathbf{I}_{rH} \rightarrow \mathcal{D}_{HL}(\mathbf{I}_{rH}) \rightarrow \mathcal{R}_{LS}(\mathcal{D}_{HL}(\mathbf{I}_{rH}))$", together with the previous synthetic degradation branch, and simultaneously comprises the backward cycle-consistency learning process "$\mathbf{I}_{rL} \rightarrow \mathcal{R}_{LS}(\mathbf{I}_{rL}) \rightarrow \mathcal{D}_{SL}(\mathcal{R}_{LS}(\mathbf{I}_{rL}))$", in which the corresponding degradation stage is implemented by the real-world HR face image degradation branch $\mathcal{D}_{SL}$ introduced as follows.

*3) Real-World HR Face Degradation Branch:* Since the real-world and synthetic LR face images suffer from an inevitable degradation gap, it is reasonable to separately degrade the real-world HR face image $\mathbf{I}_{rH}$ and the synthetic one $\mathbf{I}_{sH}$ generated from the restoration branch $\mathcal{R}_{LH}$ by respective branches. The real-world HR face degradation branch, with the restoration one, comprise the backward cycle-consistent learning process "$\mathbf{I}_{rL} \rightarrow \mathcal{R}_{LH}(\mathbf{I}_{rL}) \rightarrow \mathcal{D}_{SL}(\mathcal{R}_{LH}(\mathbf{I}_{rL}))$".

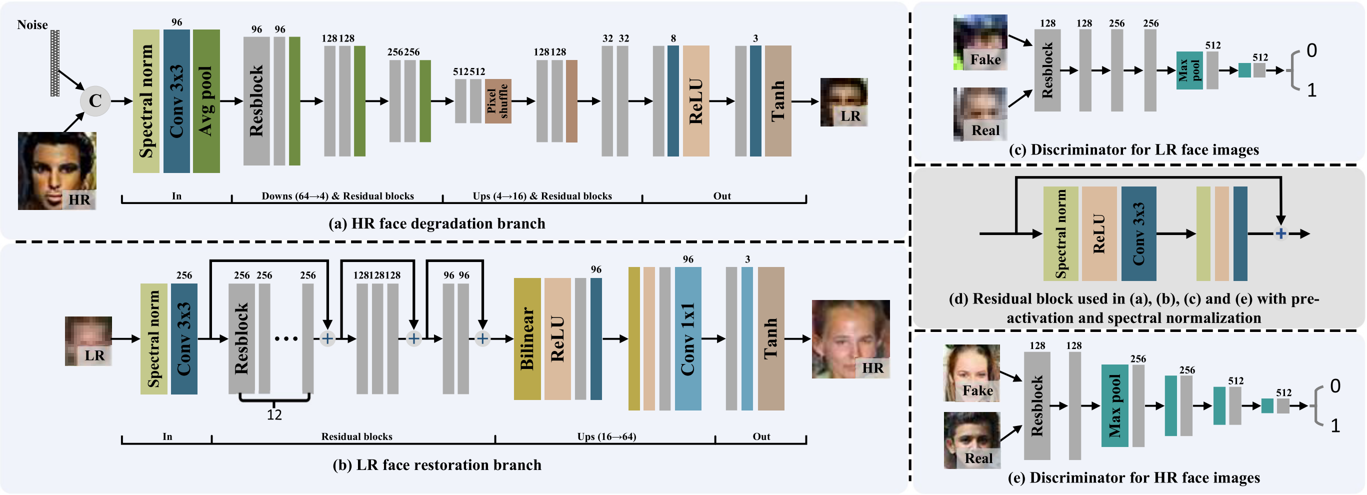
*4) Discussion:* We train the fully-cycled CycleGAN and our semi-cycled SCGAN with unpaired HR face images from the FFHQ dataset [47] and LR ones from the Widerface dataset [64]. To achieve better reconstruction quality, we train the fully-cycled CycleGAN and our semi-cycled SCGAN with an additional pixel-wise loss function, which will be introduced in §III-C. In Figure 4, we first compare the degraded LR images and the restored HR ones by CycleGAN and our SCGAN, respectively, on four typical real-world HR face images. The results of FID scores [65], PSNR, and SSIM [66] are also provided as references. One can see that the synthetic LR face images degraded from the HR ones in our SCGAN obtains lower FID score than those degraded in the fully-cycled CycleGAN, indicating that our degradation branch obtains more realistic LR face images than those generated by CycleGAN. To evaluate the reconstruction consistency, we also restore the two sets of synthetic LR face images by the corresponding restoration branches in CycleGAN and our SCGAN, respectively. Besides, we compare the restored SR images and the degraded LR ones by CycleGAN and our SCGAN, respectively, on four typical real-world LR face images. We observe that the face images restored by our SCGAN show clear improvement over those restored by CycleGAN on details recovery. All these results indicate the advantage of our semi-cycled SCGAN over the fully-cycled CycleGAN on unsupervised real-world face SR.

### *C. Synthetic Degradation Branch on HR Face Image*

This branch, denoted as $\mathcal{D}_{HL}$, aims to learn the degradation process from real-world HR face images to synthetic LR ones. Given a real-world HR face image $\mathbf{I}_{rH} \in \mathbb{R}^{H \times W \times 3}$, we randomly generate a noise vector $z \in \mathbb{R}^{HW}$, reshape it into the size of $H \times W$, and concatenate it with $\mathbf{I}_{rH}$ along the channel dimension. This is to simulate different degrees and types of noise contained in real-world LR face images, as suggested in [49]. The concatenated tensor $[\mathbf{I}_{rH}, z] \in \mathbb{R}^{H \times W \times 4}$ is then fed into the degradation branch $\mathcal{D}_{HL}$ to produce a synthetic LR face image $\mathbf{I}_{sL}$:

$$\mathbf{I}_{sL} = \mathcal{D}_{HL}([\mathbf{I}_{rH}, z], \mathbf{\Theta}_{HL}), \tag{1}$$

where $\mathbf{\Theta}_{HL}$ is the set of learnable parameters for $\mathcal{D}_{HL}$.



(a) HR face degradation branch

(b) LR face restoration branch

(c) Discriminator for LR face images

(d) Residual block used in (a), (b), (c) and (e) with pre-activation and spectral normalization

(e) Discriminator for HR face images

Fig. 5. Architectures of the synthetic and real-world HR face degradation branches $\mathcal{D}_{HL}$ and $\mathcal{D}_{SL}$ (a), and the LR face restoration branch $\mathcal{D}_{LS}$ (b). The discriminators are shown in **(c)** and **(e)**. The residual block used in them is shown in **(d)**. Please zoom in for the best view.

As shown in Figure 5 (a), the synthetic degradation branch $\mathcal{D}_{HL}$ is in an encoder-decoder architecture. The encoder begins with a Spectral Normalization (SN) [67], followed by a 3 × 3 convolutional layer (conv.) and a global average pooling (GAP). Then six residual blocks (Resblocks) are used to extract meaningful feature. As shown in Figure 5 (d), the ResBlock used in $\mathcal{D}_{HL}$ contains two successive sets of SN, ReLU, and 3 × 3 conv., with a skip connection for feature addition. Here, we use SN to mitigate unstable model training and gradient explosion with the 1-Lipschitz constraint [67]. GAP is used after every two ResBlocks to reduce feature resolution by a factor of 2. The decoder also has six Resblocks, with two Pixel-Shuffle operations used after the second and fourth ResBlocks to upsample feature resolution by a factor of 2. Finally, this branch has two groups of Resblock and 3 × 3 Conv., followed by a ReLU or a Tanh function for nonlinear activation, respectively, and outputs the degraded LR face image $\mathbf{I}_{sL}$.

To approximate the degradation in real-world LR face images, the synthetic degradation branch $\mathcal{D}_{HL}$ is learned with an adversarial loss function and a pixel loss function as:

$$l_{\mathcal{D}_{HL}} = \alpha l_{adv}^{\mathcal{D}_{L1}} + \beta l_{pix}^{\mathbf{I}_{sL}}, \tag{2}$$

where $\alpha$ and $\beta$ are the weights of the two loss functions.

*1) Adversarial Loss:* $l_{adv}^{\mathcal{D}_{L1}}$ uses a discriminator $\mathcal{D}_{L1}$ to predict the real-world LR face image $\mathbf{I}_{rL}$ as 1 and the synthetic LR one $\mathbf{I}_{sL}$ as 0, respectively. As shown in Figure 5 (c), the discriminator $\mathcal{D}_{L1}$ contains six Resblocks followed by a fully connected layer. The max-pooling is used before the last two Resblocks to reduce the resolution of the feature map. Similar to [67], we use the hinge loss as follows,

$$\begin{aligned} l_{adv}^{\mathcal{D}_{L1}} &= \mathbb{E}_{I_{rL} \sim \mathbf{P}_{rL}}[\min(0, \mathcal{D}_{L1}(I_{rL}) - 1)] \\ &\quad + \mathbb{E}_{I_{sL} \sim \mathbf{P}_{sL}}[\min(0, -1 - \mathcal{D}_{L1}(I_{sL}))], \end{aligned} \tag{3}$$

where $\mathbf{P}_{rL}$ and $\mathbf{P}_{sL}$ are the distributions of real-world LR face image $\mathbf{I}_{rL}$ and the synthetic one $\mathbf{I}_{sL}$ degraded by $D_{HL}$ from the real-world HR face image $\mathbf{I}_{rH}$, respectively.

*2) Pixel Loss:* $l_{pix}^{\mathbf{I}_{sL}}$ is calculated between the synthetic degradation image $\mathbf{I}_{sL}$ and the input HR face image $\mathbf{I}_{rH}$ downsampled to the same resolution with $\mathbf{I}_{sL}$ by average pooling. Here, we adopt the $\ell_1$ loss function that is widely used in image SR task [15], [68] to well recover image details.

### *D. Synthetic/Real-World LR Face Restoration Branch*

The LR face restoration branch $\mathcal{R}_{LS}$ is a hub shared by the forward and backward cycle-consistency learning processes. In the forward learning process, it restores the synthetic LR image $\mathbf{I}_{sL}$ degraded from the HR face image $\mathbf{I}_{rH}$ via $\mathcal{D}_{HL}$, while in the backward learning process, it restores the real-world LR face image $\mathbf{I}_{rL}$. Denote the SR image restored from $\mathbf{I}_{sL}$ as $\mathbf{I}_{sS}$ and the SR image restored from $\mathbf{I}_{rL}$ as $\mathbf{I}_{rS}$, the restoration process is as follows:

$$\mathbf{I}_{sS} = \mathcal{R}_{LS}(\mathbf{I}_{sL}, \mathbf{\Theta}_{LS}), \tag{4}$$

$$\mathbf{I}_{rS} = \mathcal{R}_{LS}(\mathbf{I}_{rL}, \mathbf{\Theta}_{LS}), \tag{5}$$

where $\mathbf{\Theta}_{LS}$ is the learnable parameters of the branch $\mathcal{R}_{LS}$.

As shown in Figure 5 (b), our restoration branch $\mathcal{R}_{LS}$ also begins with a Spectral Normalization [67], followed by a 3 × 3 convolutional layers. Then three groups of 12, 3, and 2 Resblocks are used to extract meaningful features, and in each group the input and output of each group have a skip connection for feature addition to preserve high-frequency details. To enhance its resolution, the feature map is upsampled by a factor of 4 by two bilinear interpolations, followed by a group of "ReLU-Resblock-3 × 3 Conv." and a group of "ReLU-Resblock-1 × 1 Conv.", respectively. Finally, this branch outputs the restored HR face image through a Resblock, a 1 × 1 Conv., and a Tanh activation function.

The restoration branch $\mathcal{R}_{LS}$ aims to generate high-quality face images, shared by the forward and backward learning processes. We use the combination of adversarial loss $l_{adv}^{\mathcal{D}_{H1}}$ and cycle-consistency loss $l_{cyc}^{\mathbf{I}_{sS}}$ in the forward learning process, and use the combination of adversarial loss $l_{adv}^{\mathcal{D}_{H2}}$ and pixel loss $l_{pix}^{\mathbf{I}_{rS}}$ in the backward learning process. The overall loss function for this branch is

$$l_{\mathcal{R}_{LS}} = \theta l_{\mathcal{R}_{LS}}^{\mathbf{I}_{sS}} + \gamma l_{\mathcal{R}_{LS}}^{\mathbf{I}_{rS}}, \tag{6}$$

where $\theta$ and $\gamma$ are the corresponding weights, and

$$l_{\mathcal{R}_{LS}}^{\mathbf{I}_{sS}} = \alpha l_{adv}^{\mathcal{D}_{H1}} + \beta l_{cyc}^{\mathbf{I}_{sS}}, \tag{7}$$

$$l_{\mathcal{R}_{LS}}^{\mathbf{I}_{rS}} = \alpha l_{adv}^{\mathcal{D}_{H2}} + \beta l_{pix}^{\mathbf{I}_{rS}}. \tag{8}$$

*1) Adversarial Loss:* We use a discriminator $\mathcal{D}_{H1}$ to predict the real-world HR face image $I_{rH}$ as 1 and the synthetic SR image $I_{sS}$ as 0, respectively. Similarly, we use a discriminator $\mathcal{D}_{H2}$ to predict the real-world HR face image $I_{rH}$ as 1 and the real-world SR image $I_{rS}$ as 0, respectively. The adversarial losses $l_{adv}^{\mathcal{D}_{H1}}$ and $l_{adv}^{\mathcal{D}_{H2}}$ are computed as follows,

$$\begin{aligned} l_{adv}^{\mathcal{D}_{H1}} &= \mathbb{E}_{I_{rH} \sim \mathbf{P}_{rH}}[\min(0, \mathcal{D}_{H1}(I_{rH}) - 1)] \\ &\quad + \mathbb{E}_{I_{sS} \sim \mathbf{P}_{sS}}[\min(0, -1 - \mathcal{D}_{H1}(I_{sS}))], \end{aligned} \tag{9}$$

$$\begin{aligned} l_{adv}^{\mathcal{D}_{H2}} &= \mathbb{E}_{I_{rH} \sim \mathbf{P}_{rH}}[\min(0, \mathcal{D}_{H2}(I_{rH}) - 1)] \\ &\quad + \mathbb{E}_{I_{rS} \sim \mathbf{P}_{rS}}[\min(0, -1 - \mathcal{D}_{H2}(I_{rS}))]. \end{aligned} \tag{10}$$

Here, $\mathbf{P}_{rH}$, $\mathbf{P}_{sS}$, and $\mathbf{P}_{rS}$ are the distributions of real-world HR face image $\mathbf{I}_{rH}$, synthetic SR image $\mathbf{I}_{sS}$ restored by $\mathcal{R}_{LS}$ from the synthetic LR face image $\mathbf{I}_{sL}$, and real-world SR image $\mathbf{I}_{rS}$ restored by $\mathcal{R}_{LS}$ from the real-world LR face image $\mathbf{I}_{rL}$, respectively. The discriminators $\mathcal{D}_{H1}$ and $\mathcal{D}_{H2}$ are in the same structure, which contains six Resblocks followed by a fully connected layer and uses max-pooling before the last four Resblocks, as shown in Figure 5 (e).

*2) Cycle-Consistency Loss:* $l_{cyc}^{\mathbf{I}_{sS}}$ is an $\ell_1$ loss function used here to make our restoration branch $\mathcal{R}_{LS}$ well preserve the identity information and well recover the face details.

*3) Pixel Loss:* $l_{pix}^{\mathbf{I}_{rS}}$ is an $\ell_1$ loss function to penalize the difference between real-world HR face image $\mathbf{I}_{rH}$ and SR one $\mathbf{I}_{rL}$ (upsampled to the same size of $\mathbf{I}_{rH}$ by bicubic interpolation).

### *E. Real-World HR Face Degradation Branch*

This branch, denoted as $\mathcal{D}_{SL}$, learns to degrade the real-world SR face image $\mathbf{I}_{rS}$ restored from the real-world LR image $\mathbf{I}_{rL}$ via $\mathcal{R}_{SL}$ as follows,

$$\hat{\mathbf{I}}_{rL} = \mathcal{D}_{SL}(\mathbf{I}_{rS}, \mathbf{\Theta}_{SL}), \tag{11}$$

where $\mathbf{\Theta}_{SL}$ is the learnable parameters. As shown in Figure 5 (b), the architecture of $\mathcal{D}_{SL}$ is the same as that of the synthetic HR face degradation branch $\mathcal{D}_{HL}$ introduced in §III-C.

To make the branch $\mathcal{D}_{SL}$ generate degradation results that are close to real-world LR face images, here, we employ the adversarial loss $l_{adv}^{\mathcal{D}_{L2}}$ and the cycle-consistency loss $l_{cyc}^{\hat{I}_{rL}}$ between the output LR image $\hat{\mathbf{I}}_{rL}$ and the real-world one $\mathbf{I}_{rL}$, which are computed as follows,

$$l_{\mathcal{D}_{SL}} = \alpha l_{adv}^{\mathcal{D}_{L2}} + \beta l_{cyc}^{\hat{I}_{rL}}. \tag{12}$$

*1) Adversarial Loss:* $l_{adv}^{\mathcal{D}_{L2}}$ uses a discriminator $D_{L_2}$ to predict the real-world LR face image $\mathbf{I}_{rL}$ as 1 and the output LR one $\hat{\mathbf{I}}_{rL}$ as 0, respectively. The architecture of $D_{L2}$ is the same as that of $D_{L1}$ introduced in §III-C. Similar to Eq. (3), the adversarial loss $l_{adv}^{\mathcal{D}_{L2}}$ is computed as follows,

$$\begin{aligned} l_{adv}^{\mathcal{D}_{L2}} &= \mathbb{E}_{I_{rL} \sim \mathbf{P}_{rL}}[\min(0, \mathcal{D}_{L2}(I_{rH}) - 1)] \\ &\quad + \mathbb{E}_{\hat{I}_{rL} \sim \mathbf{P}_{\hat{rL}}}[\min(0, -1 - \mathcal{D}_{L2}(\hat{I}_{rL}))], \end{aligned} \tag{13}$$

where $\mathbf{P}_{rL}$ and $\mathbf{P}_{r\hat{L}}$ are the distributions of real-world LR face image $\mathbf{I}_{rL}$ and synthetic one $\hat{I}_{rL}$ degraded by $D_{SL}$ from the real-world SR face image $\mathbf{I}_{rS}$, respectively.

*2) Cycle-Consistency Loss:* $l_{cyc}^{\hat{I}_{rL}}$ is an $\ell_1$ loss function to penalize the difference between the LR image $\hat{\mathbf{I}}_{rL}$ degraded by this branch and the corresponding real-world LR face image $\mathbf{I}_{rL}$.

### *F. Implementation Details*

The parameters of all three branches in our SCGAN are initialized by Kaiming initialization [69], and optimized by Adam [70] with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. We set $\alpha = 1$, $\beta = 0.05$ in Eqs. (2)-(8) and $\theta = 1, \gamma = 0.05$ in Eq. (6). Our SCGAN is trained for 200 epochs. The learning rate is initialized as $1 \times 10^{-4}$ and decayed to $1 \times 10^{-5}$ with the cosine annealing scheme at every 10 epochs. The batch size is set as 64. We implement our SCGAN in PyTorch [71] and train it on a Tesla V100 GPU, which takes about 42 hours.

## IV. Experiments

In this section, we first introduce the experimental setup, including the dataset and evaluation metrics in §IV-A. We then conduct a comprehensive ablation study in §IV-B to validate the role of each component of our SCGAN on face SR. Comparison with the state-of-the-art methods on real-world face SR are presented in §IV-C. Finally, we apply our SCGAN into three other vision tasks, *e.g.*, face detection, face verification, and face landmark detection in §IV-D.

### *A. Dataset and Evaluation Metric*

*1) Training Set:* We train our SCGAN, its variants (to be introduced in §IV-B), and all the comparison methods (to be introduced in §IV-C) with the 20,000 high-quality, high-resolution (HR) face images from the real-world FFHQ dataset [47] and the 4,000 low-quality, low-resolution (LR) face images from the real-world Widerface dataset [64].

*2) Test Set:* We evaluate the comparison methods on four popular face SR datasets, including two synthetic datasets, *i.e.*, *LS3D-W balanced* [72] and *FFHQ* [47], and two real-world datasets, *i.e.*, *Widerface* [64] and our newly collected *Webface*:

- *LS3D-W balanced* [72] contains 7,200 HR face images taken in different scenes and poses. We randomly select 1000 face images and perform simple bilinear downsampling to produce synthetic LR face images.
- *FFHQ* [47] contains 70,000 HR face images, 20,000 of which are used as the training set. We randomly select 2,500 images from the remaining images to perform

random degradation $I_{sL} = ((I_{rL} \otimes k) \downarrow + n_{\delta})_{JPEG_q}$ to produce the synthetic LR face images, as suggested in [15]. Here, $k$ is a Gaussian blur kernel, $\downarrow$ is a downsampling operation randomly chosen from bilinear or bicubic at a scaling factor of 4, $n_{\delta}$ is additive white Gaussian noise, and $JPEG_q$ is the JPEG compression with quality factor $q$. For each degradation, we randomly sample $k \in [0.5, 8]$, $\delta \in [1, 25]$, and $q \in [30, 95]$, respectively.

- *Widerface* [64] contains 32,203 real-world LR face images from 62 versatile scenes, and we randomly select 2,000 images with unknown yet complex degradation process.
- *WebFace*. We crawled 1,028 real-world LR face images, with different genders, ages, races, expression, postures, and unknown degradation process, from the internet.
- *DroneSURF* [73] contains more than 720,000 images with faces from drone-captured videos in the wild, we randomly selected 1,000 images, and directly cropped the patches of size 16 × 16 that contain human faces.

*3) Evaluation Metrics:* We employ feature-level and image-level metrics to objectively and comprehensively evaluate results of different methods. On all test sets, we use the Frechet Inception Distance (FID) [65] and Kernel Inception Distance (KID) [74] to evaluate the distribution distance between the output SR images and real-world HR face images on diversity and visual quality, respectively. On two synthetic test sets, we also use the Learned Perceptual Image Patch Similarity (LPIPS) [75] to measure the distance of human perception between the SR images and the corresponding ground-truth ones. On two real-world test sets, we also use the widely used Natural Image Quality Evaluator (NIQE) [76] to evaluate the naturalness of restored face images. Besides, we compute the detection accuracy of the method based on RetinaFace [77] on the SR face images from each dataset, which indirectly measures the capability of face SR methods on identity preservation.

### B. Ablation Study

To study the role of each component in our SCGAN to its effectiveness on real-world face SR, here, we conduct detailed examinations of our SCGAN on different LR face image test sets. Specifically, we access a) the benefits of our semi-cycle architecture; b) whether to share parameters of two degradation branches or not in our SCGAN; c) how different loss functions (adversarial loss, pixel loss, and cycle consistency loss) contribute to our SCGAN; d) how different combinations of adversarial losses influence our SCGAN; e) how different structures of the degradation branch influence our SCGAN; f) how about using two independent restoration branches with a shared degradation branch; g) how to determine the weights of different loss functions.

TABLE I

QUANTITATIVE RESULTS ON TWO SYNTHETIC AND TWO REAL-WORLD DATASETS BY OUR SCGAN AND ITS VARIANTS WITH DIFFERENT ARCHITECTURES. THE BEST, SECOND BEST, AND THIRD BEST RESULTS ARE HIGHLIGHTED IN RED, BLUE AND **BOLD**, RESPECTIVELY

| Dataset | Variant | FID ↓ | KID ↓ | LPIPS ↓ |
|---|---|---|---|---|
| LS3D-W balanced [72] | SCGAN-w/o-$\mathcal{D}_{SL}$ | 43.24 | 3.64±0.10 | **0.081** |
| | SCGAN-w/o-$\mathcal{D}_{HL}$ | **27.49** | **3.51±0.10** | 0.129 |
| | SCGAN-fc | **25.84** | **1.99±0.07** | **0.088** |
| | SCGAN | **22.55** | **1.26±0.06** | **0.068** |
| FFHQ [47] | SCGAN-w/o-$\mathcal{D}_{SL}$ | 27.57 | **3.18±0.09** | **0.242** |
| | SCGAN-w/o-$\mathcal{D}_{HL}$ | **19.25** | 3.44±0.12 | 0.310 |
| | SCGAN-fc | **15.15** | **2.00±0.08** | **0.247** |
| | SCGAN | **9.06** | **0.94±0.05** | **0.197** |
| Dataset | Variant | FID ↓ | KID ↓ | NIQE ↓ |
| Widerface [64] | SCGAN-w/o-$\mathcal{D}_{SL}$ | 33.96 | **3.14±0.11** | **6.5738** |
| | SCGAN-w/o-$\mathcal{D}_{HL}$ | **20.53** | 3.36±0.12 | 6.7653 |
| | SCGAN-fc | **16.02** | **1.65±0.06** | **6.7538** |
| | SCGAN | **13.32** | **1.08±0.05** | **6.6192** |
| WebFace | SCGAN-w/o-$\mathcal{D}_{SL}$ | 40.57 | 3.75±0.10 | **6.5697** |
| | SCGAN-w/o-$\mathcal{D}_{HL}$ | **25.29** | **3.32±0.10** | **6.7358** |
| | SCGAN-fc | **23.31** | **2.10±0.07** | 6.7464 |
| | SCGAN | **21.06** | **1.39±0.06** | **6.5835** |

**a) How the semi-cycled architecture benefits our SCGAN on real-world face SR?** To answer this question, we develop three variants of our SCGAN. 1) We remove the real-world HR face degradation branch $\mathcal{D}_{SL}$ introduced in §III-E, and only train the forward cycle-consistency reconstruction process "$\mathcal{D}_{HL} \rightarrow \mathcal{R}_{LS}$". This variant is denoted as "SCGAN-w/o-$\mathcal{D}_{SL}$". 2) We remove the synthetic degradation branch $\mathcal{D}_{HL}$ introduced in §III-C, and only train the backward cycle-consistency reconstruction process "$\mathcal{R}_{LS} \rightarrow \mathcal{D}_{SL}$". This variant is denoted as "SCGAN-w/o-$\mathcal{D}_{HL}$". 3) We share the parameters of synthetic HR face degradation branch $\mathcal{D}_{HL}$ and real-world one $\mathcal{D}_{SL}$, and jointly train the forward cycle-consistency reconstruction process $\mathcal{D}_{HL} \rightarrow \mathcal{R}_{LS}$ as well as the backward one $\mathcal{R}_{LS} \rightarrow \mathcal{D}_{SL}$. This variant is denoted as "SCGAN-fc". The quantitative results are listed in Table I. We observe that our SCGAN achieves better results in term of FID, KID, and LPIPS, with comparable NIQE results, than the other variants. This demonstrates that our semi-cycled architecture really benefits the real-world face SR task.

**b) Whether to share parameters or not in the two degradation branches in our SCGAN?** Here, we study whether to share parameters or not in the two degradation branches $\mathcal{D}_{HL}$ and $\mathcal{D}_{SL}$ of our semi-cycled SCGAN. To this end, we design four other variants of our SCGAN with degradation branches $\mathcal{D}_{HL}$ sharing partial parameters (except for the input head and output head). 1) We share the parameters of all layers except the input and output heads (denoted as "In" and "Out" in Figure 5) in $\mathcal{D}_{HL}$ and $\mathcal{D}_{SL}$, and denote this variant as "SCGAN-SA". 2) We share the parameters of the encoder layers in $\mathcal{D}_{HL}$ and $\mathcal{D}_{SL}$, and denote this variant as "SCGAN-SE". 3) We share the parameters of the decoder layers in $\mathcal{D}_{HL}$ and $\mathcal{D}_{SL}$, and denote this variant as "SCGAN-SD". 4) We share the parameters of the middle part, *i.e.*, the last two groups of Resblocks in the encoder and the first two groups in the decoder, in $\mathcal{D}_{HL}$ and $\mathcal{D}_{SL}$. We denote this variant as "SCGAN-SM". The objective results are listed in Table II. One can see that our SCGAN with two independent

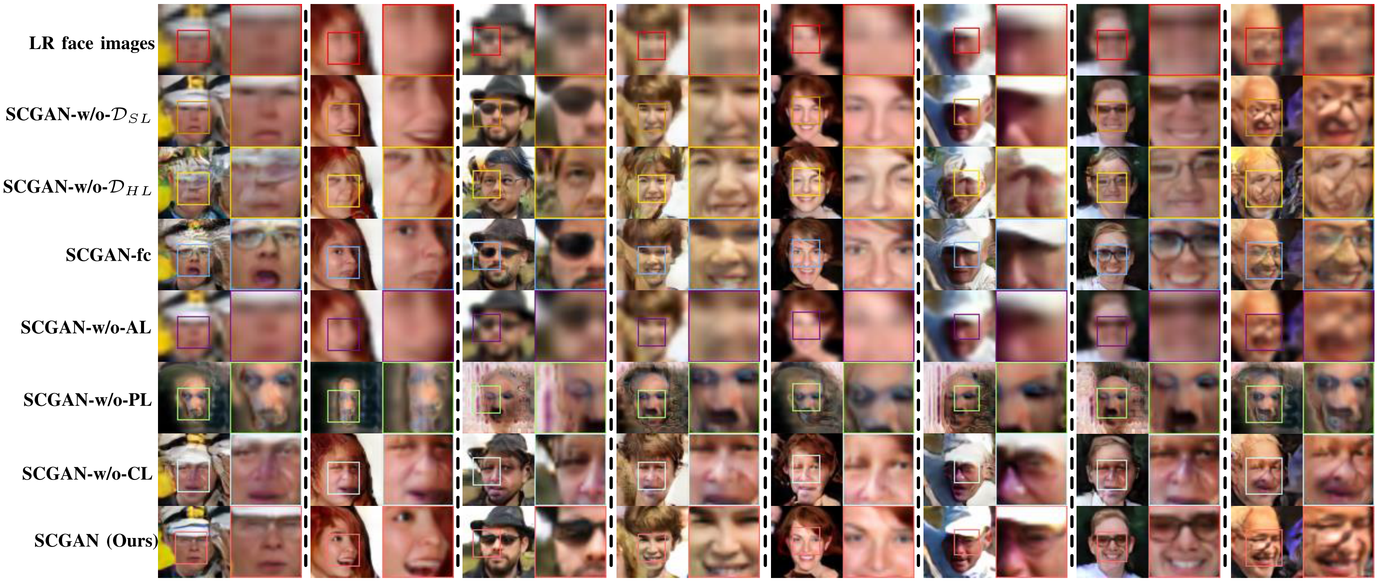


Fig. 6. Comparison results by different variants of our SCGAN on representative LR face images from the Widerface [64] dataset.

TABLE II

QUANTITATIVE RESULTS ON TWO SYNTHETIC AND TWO REAL-WORLD DATASETS BY OUR SCGAN AND ITS VARIANTS WITH DIFFERENT PARAMETER-SHARING SCHEMES IN TWO DEGRADATION BRANCHES. THE BEST, SECOND BEST, AND THIRD BEST RESULTS ARE HIGHLIGHTED IN **RED**, **BLUE** AND **BOLD**, RESPECTIVELY

| Dataset | Variant | FID ↓ | KID ↓ | LPIPS ↓ |
|---|---|---|---|---|
| LS3D-W balanced [72] | SCGAN-SA | 25.54 | 2.10±0.10 | **0.073** |
| | SCGAN-SE | 23.47 | 2.19±0.08 | **0.073** |
| | SCGAN-SD | **23.33** | **1.39±0.08** | **0.068** |
| | SCGAN-SM | **23.12** | **1.28±0.06** | 0.077 |
| | SCGAN | **22.55** | **1.26±0.06** | **0.068** |
| FFHQ [47] | SCGAN-SA | **10.65** | **1.05±0.05** | **0.200** |
| | SCGAN-SE | **10.59** | 1.64±0.07 | 0.209 |
| | SCGAN-SD | 10.96 | **1.11±0.06** | **0.198** |
| | SCGAN-SM | 11.57 | 1.31±0.07 | 0.219 |
| | SCGAN | **9.06** | **0.94±0.05** | **0.197** |
| Dataset | Variant | FID ↓ | KID ↓ | NIQE ↓ |
| Widerface [64] | SCGAN-SA | 15.36 | 1.45±0.08 | **6.6553** |
| | SCGAN-SE | **14.30** | 1.64±0.09 | **6.6534** |
| | SCGAN-SD | **13.61** | **1.11±0.05** | 6.6741 |
| | SCGAN-SM | 14.58 | **1.23±0.07** | 6.7038 |
| | SCGAN | **13.32** | **1.08±0.05** | **6.6192** |
| WebFace | SCGAN-SA | 21.77 | 1.86±0.08 | **6.6563** |
| | SCGAN-SE | 22.04 | 2.14±0.08 | **6.5968** |
| | SCGAN-SD | **21.49** | **1.43±0.07** | 6.6875 |
| | SCGAN-SM | **21.69** | **1.70±0.07** | 6.7028 |
| | SCGAN | **21.06** | **1.39±0.06** | **6.5835** |

TABLE III

QUANTITATIVE RESULTS ON TWO SYNTHETIC AND TWO REAL-WORLD DATASETS BY OUR SCGAN AND ITS VARIANTS WITH DIFFERENT LOSS FUNCTIONS. THE BEST, SECOND BEST, AND THIRD BEST RESULTS ARE HIGHLIGHTED IN **RED**, **BLUE** AND **BOLD**, RESPECTIVELY

| Dataset | Variant | FID ↓ | KID ↓ | LPIPS ↓ |
|---|---|---|---|---|
| LS3D-W balanced [72] | SCGAN-w/o-AL | 235.60 | **10.47±0.10** | **0.343** |
| | SCGAN-w/o-PL | **185.45** | 24.48±0.24 | 0.360 |
| | SCGAN-w/o-CL | **33.62** | **4.11±0.09** | **0.117** |
| | SCGAN | **22.55** | **1.26±0.06** | **0.068** |
| FFHQ [47] | SCGAN-w/o-AL | 235.76 | **11.81±0.18** | 0.916 |
| | SCGAN-w/o-PL | **181.00** | 24.89±0.35 | **0.877** |
| | SCGAN-w/o-CL | **24.29** | **4.21±0.12** | **0.289** |
| | SCGAN | **9.06** | **0.94±0.05** | **0.197** |
| Dataset | Variant | FID ↓ | KID ↓ | NIQE ↓ |
| Widerface [64] | SCGAN-w/o-AL | 233.94 | **11.00±0.16** | 7.7086 |
| | SCGAN-w/o-PL | **186.03** | 24.88±0.30 | **6.8927** |
| | SCGAN-w/o-CL | **26.49** | **4.21±0.12** | **6.7678** |
| | SCGAN | **13.32** | **1.08±0.05** | **6.6192** |
| WebFace | SCGAN-w/o-AL | 239.39 | **12.49±0.12** | 7.7209 |
| | SCGAN-w/o-PL | **189.44** | 24.55±0.26 | **6.9038** |
| | SCGAN-w/o-CL | **31.68** | **4.48±0.10** | **6.7251** |
| | SCGAN | **21.06** | **1.39±0.06** | **6.5835** |

degradation branch achieves the best results among all these variants in terms of all four objective metrics.

**c) How different loss functions (*i.e.*, adversarial loss, pixel loss and cycle consistency loss) contribute to our SCGAN on face SR?** To understand the role of different loss functions, we design three variants of our SCGAN: 1) we remove all adversarial losses in our SCGAN, and denote this variant as “SCGAN-w/o-AL”; 2) we remove all pixel losses in our SCGAN, and denote this variant as “SCGAN-w/o-PL”; 3) we remove all cycle-consistency losses in our SCGAN, and denote this variant as “SCGAN-w/o-CL”. The results of FID, KID and NIQE listed in Table III show that our SCGAN without either loss function achieves inferior performance to the original SCGAN. The visual comparison results on Widerface [64] are shown in Figure 6. We observe that the variant “SCGAN-w/o-AL” fails to recover well the face details, and the variant “SCGAN-w/o-PL” could not guarantee the contextual consistency with the input real-world LR face images, while the variant “SCGAN-w/o-CL” hard to preserve structural consistency on identity. On the contrary, by integrating all three loss functions, our SCGAN recovers well the contextual

TABLE IV

QUANTITATIVE RESULTS ON TWO REAL-WORLD LR FACE IMAGE DATASETS BY OUR SCGAN AND ITS 15 MORE VARIANTS WITH DIFFERENT COMBINATIONS OF ADVERSARIAL LOSSES. THE BEST, SECOND BEST, AND THIRD BEST RESULTS ARE HIGHLIGHTED IN **RED**, **BLUE** AND **BOLD**, RESPECTIVELY

| Removal number | Variant | $l_{adv}^{\mathcal{D}_{L1}}$ | $l_{adv}^{\mathcal{D}_{L2}}$ | $l_{adv}^{\mathcal{D}_{H1}}$ | $l_{adv}^{\mathcal{D}_{H2}}$ | Widerface [64] | | | WebFace | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | FID ↓ | KID ↓ | NIQE ↓ | FID ↓ | KID ↓ | NIQE ↓ |
| 4 | $l_{adv}$-4-1 | ✗ | ✗ | ✗ | ✗ | 233.94 | 11.00±0.16 | 7.7086 | 239.39 | 12.49±0.12 | 7.7209 |
| 3 | $l_{adv}$-3-1 | ✗ | ✗ | ✗ | ✓ | 31.68 | 3.31±0.11 | **6.7656** | 41.60 | 4.07±0.11 | 6.7961 |
| | $l_{adv}$-3-2 | ✗ | ✗ | ✓ | ✗ | 222.02 | 9.55±0.15 | 7.0507 | 228.72 | 10.94±0.12 | 7.0907 |
| | $l_{adv}$-3-3 | ✗ | ✓ | ✗ | ✗ | 230.46 | 10.40±0.15 | 7.1812 | 236.07 | 11.86±0.12 | 7.2225 |
| | $l_{adv}$-3-4 | ✓ | ✗ | ✗ | ✗ | 98.33 | 4.98±0.12 | 6.8866 | 108.11 | 5.65±0.11 | 6.8991 |
| 2 | $l_{adv}$-2-1 | ✗ | ✗ | ✓ | ✓ | **14.00** | **1.09±0.05** | **6.6933** | **21.51** | **1.50±0.06** | **6.6727** |
| | $l_{adv}$-2-2 | ✗ | ✓ | ✗ | ✓ | 201.45 | 8.60±0.14 | 6.9884 | 207.31 | 9.81±0.11 | 7.0166 |
| | $l_{adv}$-2-3 | ✗ | ✓ | ✓ | ✗ | 213.81 | 9.15±0.14 | 7.0584 | 220.66 | 10.23±0.11 | 7.0915 |
| | $l_{adv}$-2-4 | ✓ | ✗ | ✗ | ✓ | 34.27 | 4.12±0.12 | 6.8002 | 43.71 | 4.75±0.12 | 6.8045 |
| | $l_{adv}$-2-5 | ✓ | ✗ | ✓ | ✗ | 215.76 | 9.65±0.14 | 7.0541 | 222.01 | 10.29±0.11 | 7.1055 |
| | $l_{adv}$-2-6 | ✓ | ✓ | ✗ | ✗ | 233.59 | 10.94±0.16 | 7.2024 | 238.47 | 12.29±0.12 | 7.2139 |
| 1 | $l_{adv}$-1-1 | ✗ | ✓ | ✓ | ✓ | **17.44** | **1.52±0.07** | 6.7663 | **24.46** | **1.85±0.08** | **6.7707** |
| | $l_{adv}$-1-2 | ✓ | ✗ | ✓ | ✓ | 231.55 | 10.97±0.16 | 7.0690 | 236.46 | 12.42±0.13 | 7.0960 |
| | $l_{adv}$-1-3 | ✓ | ✓ | ✗ | ✓ | 36.59 | 4.53±0.12 | 6.8268 | 46.41 | 5.09±0.12 | 6.8242 |
| | $l_{adv}$-1-4 | ✓ | ✓ | ✓ | ✗ | 208.74 | 8.77±0.13 | 7.0242 | 217.33 | 9.66±0.11 | 7.0768 |
| 0 | SCGAN | ✓ | ✓ | ✓ | ✓ | **13.32** | **1.08±0.05** | **6.6192** | **21.06** | **1.39±0.06** | **6.5835** |

TABLE V

QUANTITATIVE RESULTS ON TWO REAL-WORLD DATASETS BY OUR SCGAN AND ITS VARIANTS WITH DIFFERENT STRUCTURES IN TWO DEGRADATION BRANCHES. THE BEST, SECOND BEST, AND THIRD BEST RESULTS ARE HIGHLIGHTED IN **RED**, **BLUE** AND **BOLD**, RESPECTIVELY

| Dataset | Variant | FID ↓ | KID ↓ | NIQE ↓ |
|---|---|---|---|---|
| Widerface [64] | SCGAN-$D_{HL}$-6 | **21.47** | **1.89±0.07** | **6.8726** |
| | SCGAN-$D_{SL}$-6 | **15.00** | **1.34±0.07** | **6.7011** |
| | SCGAN | **13.32** | **1.08±0.05** | **6.6192** |
| WebFace | SCGAN-$D_{HL}$-6 | **27.53** | **2.11±0.06** | **6.8160** |
| | SCGAN-$D_{SL}$-6 | **21.99** | **1.46±0.07** | **6.6485** |
| | SCGAN | **21.06** | **1.39±0.06** | **6.5835** |

and detailed information to preserve the face identity. These demonstrate that the adversarial loss is mainly used to recover details, and the pixel loss is mainly used to preserve the contextual information, while the cycle-consistency loss is mainly used to keep the structural consistency.

**d) How different combinations of adversarial losses influence our SCGAN?** Our SCGAN has 4 adversarial losses. To study this problem, we design 15 more variants of our SCGAN in 4 categories, according to the number of removed adversarial losses. The variants are denoted as "$l_{adv}$-a-b", where "a" represents the number of removed adversarial losses, and "b" represents the possible combination of the remaining 4-a adversarial losses. The details of the variants and the objective results are summarized in Table IV. We have four main observations: 1) With all the 4 adversarial losses, our SCGAN achieves better results than the other 15 variants. This demonstrates the essential role of every adversarial loss in our SCGAN for promising face SR performance. 2) By removing one adversarial loss, our SCGAN without $\mathcal{D}_{L2}$ or $\mathcal{D}_{H2}$ degrades greatly in terms of all four evaluation metrics. This reveals the dominate role of the adversarial losses in the backward cycle-consistency learning process for effective real-world LR face restoration. 3) By removing two adversarial losses, our SCGAN without $\mathcal{D}_{L1}$ and $\mathcal{D}_{L2}$ ("$l_{adv}$-2-1") achieves slightly inferior results than our original SCGAN. This shows the essential role of high-quality HR face images on the guidance of learning an effective LR face restoration branch. 4) By removing three adversarial losses, the variant "$l_{adv}$-3-1" performs better than the other three variants of "$l_{adv}$-3-2", "$l_{adv}$-3-3", or "$l_{adv}$-3-4". This shows that the adversarial loss in $\mathcal{D}_{H2}$ plays a dominant role in optimizing the real-world LR face restoration branch.

**e) How different structures of the degradation branch influence our SCGAN?** To this end, we conducted experiments to explore the impact of degradation branch architectures on the performance of our SCGAN. Specifically, we designed two variants of our SCGAN with different degradation structures: 1) "SCGAN-$D_{HL}$-6": the number of residual blocks in the encoder-decoder architecture of the synthetic degradation branch $D_{HL}$ of our SCGAN is reduced from 12 to 6; 2) "SCGAN-$D_{SL}$-6": the number of residual blocks in the encoder-decoder architecture of the real-world HR face degradation branch $D_{SL}$ of our SCGAN is reduced from 12 to 6. In Table VI, we list the quantitative results on two real-world datasets Widerface [64] and WebFace by the two variants "SCGAN-$D_{HL}$-6" and "SCGAN-$D_{SL}$-6", as well as our SCGAN. One can see that the variants of "SCGAN-$D_{HL}$-6" and "SCGAN-$D_{SL}$-6" achieve lower FID, KID, and NIQE scores than the proposed SCGAN. This indicates that our SCGAN performs better when the two degradation branches both have 12 residual blocks than those with only 6 blocks in the synthetic (or real-world) HR face degradation branch.

**f) How about using two independent restoration branches with a shared degradation branch?** To answer this question, we performed experiments by designing our SCGAN with a shared degradation branch and two independent restoration branches. We denote this varaint as "SCGAN-2SR", it contains two independent restoration branches,

TABLE VI

QUANTITATIVE RESULTS ON TWO REAL-WORLD DATASETS BY OUR SCGAN AND ITS VARIANT WITH TWO INDEPENDENT RESTORATION BRANCHES AND A SHARED DEGRADATION BRANCH. THE BEST, SECOND BEST, AND THIRD BEST RESULTS ARE HIGHLIGHTED IN **RED**, **BLUE** AND **BOLD**, RESPECTIVELY

| Dataset | Variant | FID ↓ | KID ↓ | NIQE ↓ |
|---|---|---|---|---|
| Widerface [64] | SCGAN-2SR ($\mathcal{R}_{LS}$) | **397.81** | **51.88±0.27** | **7.8605** |
| | SCGAN-2SR ($\mathcal{R}_{RS}$) | **91.86** | **9.55±0.18** | **6.4488** |
| | SCGAN | **13.32** | **1.08±0.05** | **6.6192** |
| WebFace | SCGAN-2SR ($\mathcal{R}_{LS}$) | **375.92** | **51.63±0.25** | **7.7761** |
| | SCGAN-2SR ($\mathcal{R}_{RS}$) | **87.29** | **8.59±0.14** | **6.3390** |
| | SCGAN | **21.06** | **1.39±0.06** | **6.5835** |

TABLE VII

QUANTITATIVE RESULTS ON TWO SYNTHETIC AND TWO REAL-WORLD DATASETS BY OUR SCGAN AND ITS VARIANTS WITH DIFFERENT WEIGHTS OF DIFFERENT LOSS FUNCTIONS. THE BEST, SECOND BEST, AND THIRD BEST RESULTS ARE HIGHLIGHTED IN **RED**, **BLUE** AND **BOLD**, RESPECTIVELY

| Dataset | $\beta$ | $\gamma$ | FID ↓ | KID ↓ | NIQE ↓ |
|---|---|---|---|---|---|
| Widerface [64] | 1 | 0.05 | 387.24 | 48.47±0.23 | 7.3197 |
| | 0.5 | 0.05 | 56.55 | 5.01±0.13 | **6.6583** |
| | 0.1 | 0.05 | **15.90** | **1.34±0.07** | **6.6016** |
| | 0.01 | 0.05 | 18.78 | 1.86±0.08 | 6.6731 |
| | 0.05 | 1 | 27.56 | 2.76±0.11 | 6.7708 |
| | 0.05 | 0.5 | 23.82 | 2.20±0.09 | 6.7727 |
| | 0.05 | 0.1 | **15.51** | **1.28±0.06** | 6.7463 |
| | 0.05 | 0.01 | 17.21 | 1.67±0.08 | 6.7041 |
| | 0.05 | 0.05 | **13.32** | **1.08±0.05** | **6.6192** |
| WebFace | 1 | 0.05 | 366.09 | 48.24±0.21 | 7.4073 |
| | 0.5 | 0.05 | 30.62 | 5.08±0.11 | 6.6584 |
| | 0.1 | 0.05 | **23.41** | **1.73±0.07** | **6.5631** |
| | 0.01 | 0.05 | 30.00 | 2.60±0.08 | 6.6624 |
| | 0.05 | 1 | 33.39 | 3.35±0.10 | 6.7788 |
| | 0.05 | 0.5 | 31.99 | 2.71±0.08 | 6.7178 |
| | 0.05 | 0.1 | **25.31** | **1.68±0.06** | 6.7091 |
| | 0.05 | 0.01 | 25.94 | 2.07±0.08 | **6.6479** |
| | 0.05 | 0.05 | **21.06** | **1.39±0.06** | **6.5835** |

denoted as $\mathcal{R}_{LS}$ and $\mathcal{R}_{RS}$. Please note that, since this variant "SCGAN-2SR" contains two restoration branches $\mathcal{R}_{LS}$ and $\mathcal{R}_{RS}$, here we performed independent face image super-resolution test by $\mathcal{R}_{LS}$ and $\mathcal{R}_{RS}$, respectively. In Table VI, we list the quantitative results on two real-world datasets Widerface [64] and WebFace by the restoration branch $\mathcal{R}_{LS}$ or $\mathcal{R}_{RS}$ in the variant "SCGAN-2SR", as well as our SCGAN. It can be seen that, our SCGAN achieves better results than the restoration branches $\mathcal{R}_{LS}$ and $\mathcal{R}_{RS}$ of the variant "SCGAN-2SR". This demonstrates that using two independent SR branches with one shared degradation branch largely degrades the performance of our SCGAN on real-world face image super-resolution task.

**g) How to determine the weights of different loss functions?** The proposed SCGAN has four weights ($\alpha$, $\beta$, $\theta$, and $\gamma$) for different loss functions. To determine these parameters, we have conducted more ablation studies with different weights $\beta = 0.01, 0.1, 0.5, 1$ and $\gamma = 0.01, 0.1, 0.5, 1$ by fixing one parameter as 0.05. The quantitative results presented in Table VII show that, although our SCGAN achieves the best NIQE score when $\beta = 0.1$ and $\gamma = 0.05$, our SCGAN obtains the best FID and KID scores on both datasets when $\beta = 0.05$ and $\gamma = 0.05$. Overall, we set $\beta = 0.05$ and $\gamma = 0.05$.

### C. Comparisons With State-of-the-Art Methods

Here, we compare our SCGAN with the state-of-the-art methods on both synthetic and real-world ×4 face SR tasks (16 × 16 LR face images to 64 × 64 HR ones). To comprehensively evaluate the performance of different methods on face SR, we perform face SR with three different degradation settings: 1) *simple* degradation with randomly bilinear or bicubic downsampling; 2) *complex* degradation with blur kernel, downsampling, synthetic noise, and JPEG compression; and 3) *real-world* unknown degradation.

*1) Comparison Methods:* We compare our SCGAN with Bicubic Interpolation and other state-of-the-art methods, such as DFDNet [23], HifaceGAN [78], Real-ESRGAN [79], GFPGAN [28], LRGAN [15], PULSE [27], GCFSR [80], and RestoreFormer [81]. We also compare our SCGAN with the fully-cycled CycleGAN [32] to verify the effectiveness of our semi-cycled SCGAN on face super-resolution. Here, DFDNet [23] firstly detects and crops out faces using a face detector and then performs face super-resolution on the cropped images, and would be limited by the accuracy of employed face detector. The PULSE [27] utilized a pre-trained StyleGAN [47], and thus can only generate HR images of size 1024 × 1024. For a fair comparison, we resize its results to the same size (*e.g.*, 64 × 64) as those obtained by the comparison methods.

*2) Face SR on Simple Degradation:* Here, the *simple* degradation is performed by random bilinear or bicubic downsampling on the 1,000 LR face images in LS3D-W balanced dataset [72] as the test set, as described in §IV-A, and we evaluate the performance of different methods on it. For a fair comparison, all the comparison methods are re-trained carefully to achieve their best results. The quantitative results of FID, KID, and LPIPS are summarized in Table VIII (2-nd row). It can be seen that our SCGAN obtains higher indices than the other competiting methods. The qualitative results of visual quality are presented in Figure 7 (left part). One can see that DFDNet [23] and HifaceGAN [78] produce blurry results similar to those produced by Bicubic Interpolation. Real-ESRGAN, CycleGAN, and LRGAN fail to preseve well the facial structure. We also observe that, although PULSE and GFPGAN have amazing generalization ability in face super-resolution, their performance still has room for improvement for very low-resolution face images (16 × 16) that suffer from severe degradation. On the contrary, our SCGAN generates realistic results in ensuring the structure and detail consistency to that of the ground-truth HR face images.

*3) Face SR on Complex Degradation:* Before generalizing our SCGAN to real-world face SR, we perform experiments on blind face SR with random *complex* degradation. Here, the *complex* degradation with random blur kernel, downsampling, synthetic noise, and JPEG compression is performed on the 2,500 face images in FFHQ as the test set [47], as described

TABLE VIII

QUANTITATIVE RESULTS OF OUR SCGAN AND OTHER STATE-OF-THE-ART METHODS ON TWO SYNTHETIC AND TWO REAL-WORLD DATASETS. THE BEST, SECOND BEST, AND THIRD BEST RESULTS ARE HIGHLIGHTED IN **RED**, **BLUE** AND **BOLD**, RESPECTIVELY

| Dataset | Metric | Bicubic Interpolation | DFDNet [23] | HifaceGAN [78] | Real-ESRGAN [79] | CycleGAN [32] | GFPGAN [28] | LRGAN [15] | PULSE [27] | GCFSR [80] | RestoreFormer [81] | SCGAN (Ours) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *LS3D-W balanced* [72] | FID↓ | 274.78 | 277.38 | 238.75 | 57.20 | 51.04 | 51.02 | **33.67** | **33.65** | 70.58 | 72.89 | **22.55** |
| | KID↓ | 15.31±0.11 | 16.60±0.11 | 13.56±0.12 | **3.08±0.07** | 5.81±0.11 | 4.45±0.11 | 3.95±0.10 | **3.49±0.16** | 5.37±0.10 | 5.45±0.11 | **1.26±0.06** |
| | LPIPS↓ | 0.343 | 0.387 | 0.268 | 0.114 | **0.086** | 0.094 | 0.118 | **0.106** | 0.125 | 0.142 | **0.068** |
| *FFHQ* [47] | FID↓ | 277.19 | 285.93 | 241.69 | 43.75 | 36.22 | 43.86 | **18.39** | **27.81** | 42.57 | 46.72 | **9.06** |
| | KID↓ | 16.32±0.22 | 23.84±0.28 | 14.52±0.19 | **3.13±0.10** | 6.14±0.15 | 3.88±0.12 | 3.22±0.13 | 3.48±0.19 | 4.00±1.13 | 3.84±1.18 | **0.94±0.05** |
| | LPIPS↓ | 0.927 | 0.959 | 0.719 | 0.336 | **0.234** | **0.299** | 0.302 | 0.329 | 0.317 | 0.336 | **0.197** |
| *Widerface* [64] | FID↓ | 270.69 | 271.74 | 157.34 | 41.73 | 39.55 | 59.34 | **19.35** | **28.27** | 54.69 | 59.08 | **13.32** |
| | KID↓ | 16.23±0.21 | 17.34±0.22 | 17.54±0.24 | **2.77±0.10** | 6.26±0.12 | 3.79±0.12 | **3.20±0.14** | 3.36±0.17 | 4.80±0.12 | 4.83±0.13 | **1.08±0.05** |
| | NIQE↓ | 18.8702 | 7.9981 | **5.2294** | 6.7595 | 6.8953 | 6.9437 | 6.6446 | 6.7473 | **5.9455** | 6.7386 | **6.6192** |
| *WebFace* | FID↓ | 273.96 | 274.49 | 238.10 | 51.57 | 44.94 | 88.71 | **26.49** | **36.02** | 70.57 | 76.51 | **21.06** |
| | KID↓ | 18.02±0.15 | 19.08±0.15 | 14.46±0.15 | **3.62±0.09** | 6.54±0.11 | 6.45±0.11 | 4.45±0.11 | **3.60±0.16** | 5.89±0.11 | 5.70±0.10 | **1.39±0.06** |
| | NIQE↓ | 7.4657 | 7.7320 | **6.0178** | 6.6773 | 6.8132 | 6.8841 | **6.4868** | 6.7039 | **5.9873** | 6.8277 | 6.5835 |
| *DroneSURF* [73] | FID↓ | 173.06 | 247.77 | 153.01 | **58.47** | 87.81 | 114.05 | 59.25 | **27.91** | 116.05 | 147.58 | **48.08** |
| | KID↓ | 21.27±1.50 | 35.54±1.58 | 20.33±1.59 | **4.96±1.09** | 8.94±1.52 | 15.03±1.25 | 7.13±1.25 | **4.04±1.36** | 14.61±1.26 | 15.85±1.35 | **4.62±1.14** |
| | NIQE↓ | 6.4464 | 7.9855 | 5.8811 | **4.9336** | 7.1707 | **5.7217** | 6.6369 | 6.8865 | 6.5101 | 6.5475 | **5.1488** |

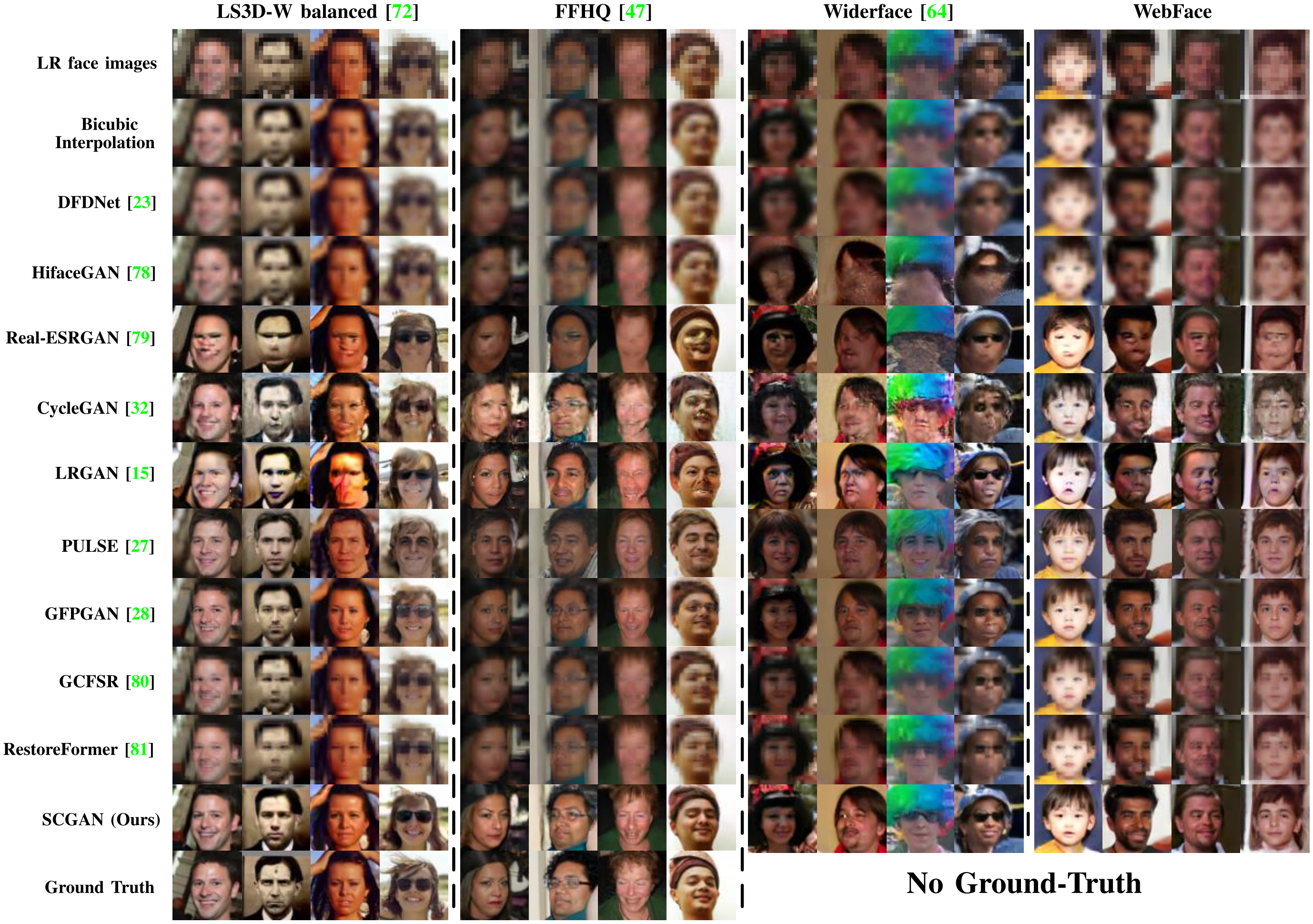


Fig. 7. Comparison of visual quality by our SCGAN and other face SR methods on LS3D-W balanced [72], FFHQ [47], Widerface [64], and WebFace datasets, respectively (from left to right). Please zoom in for better view.

in §IV-A. The quantitative results are listed in Table VIII (3-rd row). It can be seen that our SCGAN obtains clearly lower scores of FID, KID and LPIPS than the other methods. In Figure 7 (right part), we compare the face SR results of different methods on representative face samples in FFHQ [47]. We observe that Bicubic Interpolation, DFDNet [23] and HifaceGAN [78] still produce blurry results. Besides, Real-ESRGAN [79] and CycleGAN [32] fail to recover the face contexts. LRGAN [15] tends to produce incomplete face structure, and PULSE [27] is prone to lose the identity information or important face components like eyeglasses, while GFPGAN [28] fails to recover important face details. On the contrary, our SCGAN generates high-quality and realistic HR face images with accurate face structure and fine-grained face details. All these results validate that our SCGAN is more robust to the complex random degradation, and can produce high-quality HR face images more realistically to the real-world HR face images, than all the comparison methods.

*4) Face SR on Real-World Degradation:* Now we compare different methods on the Widerface [64], our Webface and DroneSURF [73] dataset for real-world face SR with complex and unknown degradation, where the experimental settings are described in §IV-A. The quantitative results are presented in Table VIII.

One can see that our SCGAN achieves higher FID and KID results than the other methods on Widerface and our WebFace datasets. At the same time, it is only inferior to PULSE on DroneSURF [73] dataset. As shown in Figure 7, though achieving the best NIQE results in the Widerface dataset, HifaceGAN is prone to produce blurry face images, similar to Bicubic Interpolation, DFDNet, and Real-ESRGAN. Though PULSE achieves the best FID and KID scores on DroneSURF dataset, its results produce noticeable structural changes from the input LR face images. Although Real-ESRGAN achieves the best NIQE score on DroneSURF dataset, it is difficult to well recover some key facial features such as eyes, nose, etc. The methods of LRGAN, PULSE, and GFPGAN produce either artifacts or color bias. After all, our SCGAN not only restores the facial structure and details, but also preserves human identity of real-world LR face images, when compared to the other comparison methods.

## D. Application on Downstream Vision Tasks

In this section, we apply our SCGAN and state-of-the-art methods to downstream vision tasks. We conduct experiment on face detection, face verification and face landmark detection in §IV-D.1, §IV-D.2 and §IV-D.3, respectively. On all tasks, we compare our SCGAN with the methods of Bicubic Interpolation, DFDNet [23], HifaceGAN [78], CycleGAN [32], LRGAN [15], and GFPGAN [28] on face SR.

*1) Application on Face Detection:* Face detection is to predict the bounding boxs around the faces in the images. To validate the effectiveness of these methods on face SR, we perform face detection with the state-of-the-art face detection method of RetinaFace [77], on the SR face images by different methods. Here, we define face detection accuracy as the ratio of the number of face images successfully predicted by bounding boxes to the total number of the input face images, and each input image has exactly one face. In Table IX, we list the detection accuracies on the SR face images by different methods from the datasets of LS3D-W balanced [72], FFHQ [47], Widerface [64] and WebFace, as described in §IV-A. One can see that, the model of RetinaFace [77] consistently achieves the highest detection accuracy on the SR face images by our SCGAN. This validates the effectiveness of our SCGAN on the face structure preservation for face detection.

*2) Application on Face Verification:* Face verification is a binary classification task to determine whether the pair of output and reference face images have the same identity or not. Here, we first restore the synthetic DroneSURF test set [73], as introduced in §IV-A, by different face SR methods. Then we perform face verification on the restored face images by the

TABLE IX

Accuracy (%) of Face Detection on the SR Results Restored by Our SCGAN and Other Methods on Two Synthetic and Two Real-World Datasets. The Best, Second Best, and Third Best Results Are Highlighted in **Red**, **Blue** and **Bold**, Respectively

| Method | LS3D-W balanced [72] | FFHQ [47] | Widerface [64] | WebFace |
|---|---|---|---|---|
| Bicubic Interpolation | 42.80 | 42.04 | 53.90 | 46.49 |
| DFDNet [23] | 32.90 | 31.64 | 41.90 | 35.99 |
| HifaceGAN [78] | 43.10 | 41.36 | 42.05 | 45.53 |
| CycleGAN [32] | **84.90** | **86.92** | **87.90** | **86.18** |
| LRGAN [15] | 75.60 | 67.56 | **83.90** | **84.63** |
| GFPGAN [28] | **92.10** | **91.76** | 83.55 | 74.90 |
| SCGAN (Ours) | **96.60** | **95.40** | **97.75** | **96.89** |

TABLE X

Accuracy of Face Verification by FaceNet on the Super-Resolved Face Images in DroneSURF Test Set [73] Restored by Different Face Super-Resolution Methods. The Best, Second Best, and Third Best Results Are Highlighted in **Red**, **Blue** and **Bold**, Respectively

| Method | Accuracy (%) |
|---|---|
| Bicubic Interpolation | 37.67 |
| DFDNet [23] | 27.83 |
| HifaceGAN [78] | **40.33** |
| CycleGAN [32] | 18.17 |
| LRGAN [15] | **47.83** |
| GFPGAN [28] | 46.17 |
| SCGAN(Ours) | **56.83** |

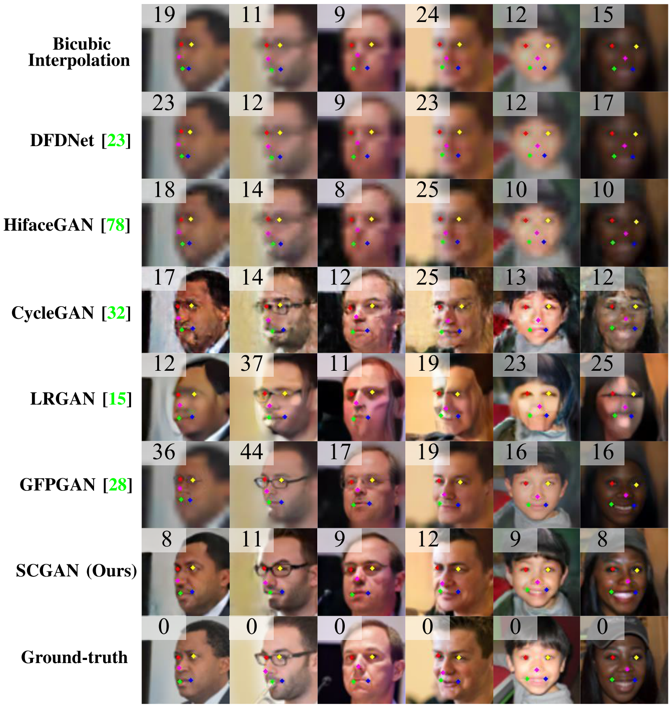


Fig. 8. Comparison results of landmark detection and the corresponding $\ell_1$ norm errors on the face images in synthetic FFHQ [47] dataset restored by different methods. Please zoom in for better view.

state-of-the-art face verification method of FaceNet [82]. The accuracies of FaceNet on the SR images by different methods are listed in Table X. It can be seen that, the accuracy on the SR images of our SCGAN is clearly higher than those of the other face SR methods. This demonstrates the superiority of our SCGAN over the other competitors on preserving the consistency of face identity information for the face SR task.

*3) Application on Face Landmark Detection:* Face landmark detection aims to locate the key facial components of face images. We restore the LR images into HR ones with more facial details and then use the state-of-the-art face landmark detection method of RetinaFace [77] for face landmarks detection. We compare the face SR methods on six representative LR face images degraded from FFHQ test set, as described in §IV-A. The landmark detection results of six representative face images restored by different methods and the $\ell_1$ norm errors (lower is better) between them and the corresponding ground-truth landmarks are shown in Figure 8. We observe that, compared with the SR results of other methods, the landmarks detected by the face images restored by our SCGAN are closer to those detected on the original HR face images, and the corresponding $\ell_1$ errors are also the lowest among all comparison methods. This shows that our SCGAN recovers the key facial components more detectable than the other methods for face SR.

## V. Conclusion

In this paper, we proposed a novel Semi-Cycled Generative Adversarial Network (SCGAN) to alleviate the domain gap between unpaired LR and HR face images for real-world face super-resolution (SR). Our SCGAN contains three independent branches to learn the forward and backward cycle-consistent reconstruction processes. Specifically, a synthetic degradation branch learns to generate synthetic LR face images by degrading the real-world HR ones, a restoration branch recovers SR face images from the synthetic/real-world LR face images, and a real-world degradation branch degrades the SR face images restored from the real-world LR ones. The restoration branch is coupled and regularized by the two independent degradation branches, making our SCGAN robust to super-resolve synthetic and real-world LR face images. Experiments on two synthetic and two real-world datasets demonstrated that our semi-cycled SCGAN outperforms the state-of-the-art methods on synthetic and real-world face SR tasks, in terms of structure preservation, detail recovery, and standard objective metrics. Three downstream vision tasks on face detection, face verification, and face landmark detection reveal that the effectiveness of our SCGAN better recovers the face structure, identity, and details than the other face SR methods.

## References

[1] B. Fasel and J. Luettin, "Automatic facial expression analysis: A survey," *Pattern Recognit.*, vol. 36, no. 1, pp. 259–275, Jan. 2008.

[2] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Comput. Surv.*, vol. 35, no. 4, pp. 399–458, 2003.

[3] X. Xiong and F. De la Torre, "Supervised descent method and its applications to face alignment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 532–539.

[4] W. Liu, D. Lin, and X. Tang, "Neighbor combination and transformation for hallucinating faces," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2005, pp. 1–4.

[5] S. W. Park and M. Savvides, "Robust super-resolution of face images by iterative compensating neighborhood relationships," in *Proc. Biometrics Symp.*, Sep. 2007, pp. 1–5.

[6] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2004, p. 1.

[7] X. Wang and X. Tang, "Hallucinating face by eigentransformation," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 35, no. 3, pp. 425–434, Aug. 2005.

[8] A. Chakrabarti, A. N. Rajagopalan, and R. Chellappa, "Super-resolution of face images using kernel PCA-based prior," *IEEE Trans. Multimedia*, vol. 9, no. 4, pp. 888–892, Jun. 2007.

[9] J.-S. Park and S.-W. Lee, "An example-based face hallucination method for single-frame, low-resolution facial images," *IEEE Trans. Image Process.*, vol. 17, no. 10, pp. 1806–1816, Oct. 2008.

[10] Y. Hu, K. M. Lam, G. Qiu, T. Shen, and H. Tian, "Learning local pixel structure for face hallucination," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2010, pp. 2797–2800.

[11] Y. Hu, K.-M. Lam, G. Qiu, and T. Shen, "From local pixel structure to global image super-resolution: A new face hallucination framework," *IEEE Trans. Image Process.*, vol. 20, no. 2, pp. 433–445, Feb. 2011.

[12] Y. Li, C. Cai, G. Qiu, and K.-M. Lam, "Face hallucination based on sparse local-pixel structure," *Pattern Recognit.*, vol. 47, no. 3, pp. 1261–1270, Mar. 2014.

[13] J. Shi, X. Liu, and C. Qi, "Global consistency, local sparsity and pixel correlation: A unified framework for face hallucination," *Pattern Recognit.*, vol. 47, no. 11, pp. 3520–3534, 2014.

[14] T. Yang, P. Ren, X. Xie, and L. Zhang, "GAN prior embedded network for blind face restoration in the wild," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 672–681.

[15] A. Bulat, J. Yang, and G. Tzimiropoulos, "To learn image super-resolution, use a GAN to learn how to do image degradation first," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 185–200.

[16] D. Huang and H. Liu, "Face hallucination using convolutional neural network with iterative back projection," in *Proc. Chin. Conf. Biometric Recognit.*, 2016, pp. 167–175.

[17] Y. Liu, Z. Dong, K. Pang Lim, and N. Ling, "A densely connected face super-resolution network based on attention mechanism," in *Proc. 15th IEEE Conf. Ind. Electron. Appl. (ICIEA)*, Nov. 2020, pp. 148–152.

[18] C. Chen, D. Gong, H. Wang, Z. Li, and K.-Y.-K. Wong, "Learning spatial attention for face super-resolution," *IEEE Trans. Image Process.*, vol. 30, pp. 1219–1231, 2021.

[19] X. Chen, X. Wang, Y. Lu, W. Li, Z. Wang, and Z. Huang, "RBPNET: An asymptotic residual back-projection network for super-resolution of very low-resolution face image," *Neurocomputing*, vol. 376, pp. 119–127, Feb. 2020.

[20] H. Huang, R. He, Z. Sun, and T. Tan, "Wavelet-SRNet: A wavelet-based CNN for multi-scale face super resolution," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 1689–1697.

[21] X. Li, M. Liu, Y. Ye, W. Zuo, L. Lin, and R. Yang, "Learning warped guidance for blind face restoration," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 272–289.

[22] B. Dogan, S. Gu, and R. Timofte, "Exemplar guided face image super-resolution without facial landmarks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 1–10.

[23] X. Li, C. Chen, S. Zhou, X. Lin, W. Zuo, and L. Zhang, "Blind face restoration via deep multi-scale component dictionaries," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 399–415.

[24] I. Goodfellow et al., "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, pp. 1–15.

[25] X. Li, W. Li, D. Ren, H. Zhang, M. Wang, and W. Zuo, "Enhanced blind face restoration with multi-exemplar images and adaptive spatial feature fusion," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2706–2715.

[26] C. Chen, X. Li, L. Yang, X. Lin, L. Zhang, and K.-Y.-K. Wong, "Progressive semantic-aware style transformation for blind face restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 11896–11905.

[27] S. Menon, A. Damian, S. Hu, N. Ravi, and C. Rudin, "PULSE: Self-supervised photo upsampling via latent space exploration of generative models," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2437–2445.

[28] X. Wang, Y. Li, H. Zhang, and Y. Shan, "Towards real-world blind face restoration with generative facial prior," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 9168–9178.

[29] S. Maeda, "Unpaired image super-resolution using pseudo-supervision," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 291–300.

[30] Y. Guo et al., "Closed-loop matters: Dual regression networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 5407–5416.

[31] K. Zhang et al., "Deblurring by realistic blurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2020, pp. 2737–2746.

[32] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2223–2232.

[33] J. Jiang, C. Wang, X. Liu, and J. Ma, "Deep learning-based face super-resolution: A survey," *ACM Comput. Surv.*, vol. 55, no. 1, pp. 1–36, Jan. 2023.

[34] X. Hu, J. Xu, S. Gu, M.-M. Cheng, and L. Liu, "Restore globally, refine locally: A mask-guided scheme to accelerate super-resolution networks," in *Proc. Eur. Conf. Comput. Vis.*, 2022, pp. 74–91.

[35] S. Baker and T. Kanade, "Hallucinating faces," in *Proc. 4th IEEE Int. Conf. Autom. Face Gesture Recognit.*, Jul. 2000, pp. 83–88.

[36] B. K. Gunturk, A. U. Batur, Y. Altunbasak, M. H. Hayes, and R. M. Mersereau, "Eigenface-domain super-resolution for face recognition," *IEEE Trans. Image Process.*, vol. 12, no. 5, pp. 597–606, May 2003.

[37] Y. Zhuang, J. Zhang, and F. Wu, "Hallucinating faces: LPH super-resolution and neighbor reconstruction for residue compensation," *Pattern Recognit.*, vol. 40, no. 11, pp. 3178–3194, 2007.

[38] T. Lu, J. Wang, J. Jiang, and Y. Zhang, "Global-local fusion network for face super-resolution," *Neurocomputing*, vol. 387, pp. 309–320, Apr. 2020.

[39] J. Xin, N. Wang, X. Gao, and J. Li, "Residual attribute attention network for face image super-resolution," in *Association for the Advancement of Artificial Intelligence*, vol. 33, no. 1. Cambridge, MA, USA: MIT Press, 2019, pp. 9054–9061.

[40] J. Xin, N. Wang, X. Jiang, J. Li, X. Gao, and Z. Li, "Facial attribute capsules for noise face super resolution," in *Association for the Advancement of Artificial Intelligence*, vol. 34, no. 7. Cambridge, MA, USA: MIT Press, 2020, pp. 12476–12483.

[41] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[42] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.

[43] S. Chen et al., "Unsupervised image super-resolution with an indirect supervised path," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR) Workshops*, Jun. 2020, pp. 468–469.

[44] X. Yu and F. Porikli, "Ultra-resolving face images by discriminative generative networks," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 318–333.

[45] A. Bulat and G. Tzimiropoulos, "Super-FAN: Integrated facial landmark localization and super-resolution of real-world low resolution faces in arbitrary poses with GANs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 109–117.

[46] K. C. K. Chan, X. Wang, X. Xu, J. Gu, and C. C. Loy, "GLEAN: Generative latent bank for large-factor image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 14245–14254.

[47] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4401–4410.

[48] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel, "InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 1–9.

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

[50] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1125–1134.

[51] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional GANs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8798–8807.

[52] Z. Yi, H. Zhang, P. Tan, and M. Gong, "DualGAN: Unsupervised dual learning for image-to-image translation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2849–2857.

[53] T. Kim, M. Cha, H. Kim, J. K. Lee, and J. Kim, "Learning to discover cross-domain relations with generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 1857–1865.

[54] R. Yi, Y.-J. Liu, Y.-K. Lai, and P. L. Rosin, "Quality metric guided portrait line drawing generation from unpaired training data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 1, pp. 905–918, Jan. 2023.

[55] Y. Yuan, S. Liu, J. Zhang, Y. Zhang, C. Dong, and L. Lin, "Unsupervised image super-resolution using Cycle-in-Cycle generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 701–710.

[56] Y. Zhang, S. Liu, C. Dong, X. Zhang, and Y. Yuan, "Multiple cycle-in-cycle generative adversarial networks for unsupervised image super-resolution," *IEEE Trans. Image Process.*, vol. 29, pp. 1101–1112, 2020.

[57] A. Lugmayr, M. Danelljan, and R. Timofte, "Unsupervised learning for real-world super-resolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 3408–3416.

[58] W. Sun, D. Gong, Q. Shi, A. van den Hengel, and Y. Zhang, "Learning to zoom-in via learning to zoom-out: Real-world super-resolution by generating and adapting degradation," *IEEE Trans. Image Process.*, vol. 30, pp. 2947–2962, 2021.

[59] J. Kim, C. Jung, and C. Kim, "Dual back-projection-based internal learning for blind super-resolution," *IEEE Signal Process. Lett.*, vol. 27, pp. 1190–1194, 2020.

[60] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.

[61] G. Hinton et al., "Distilling the knowledge in a neural network," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, pp. 1–14.

[62] G. Kim et al., "Unsupervised real-world super resolution with cycle generative adversarial network and domain discriminator," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, Jul. 2020, pp. 456–457.

[63] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database forstudying face recognition in unconstrained environments," in *Proc. Workshop Faces 'Real-Life' Images, Detection, Alignment, Recognition*, 2008, pp. 1–15.

[64] S. Yang, P. Luo, C. C. Loy, and X. Tang, "WIDER FACE: A face detection benchmark," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 5525–5533.

[65] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local Nash equilibrium," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–5.

[66] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[67] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, "Spectral normalization for generative adversarial networks," in *Proc. Int. Conf. Learn. Represent.*, 2018, pp. 1–26.

[68] G. Xu, J. Xu, Z. Li, L. Wang, X. Sun, and M.-M. Cheng, "Temporal modulation network for controllable space-time video super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 6388–6397.

[69] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. Int. Conf. Comput. Vis.*, 2015, pp. 1026–1034.

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

[71] A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 1–12.

[72] A. Bulat and G. Tzimiropoulos, "How far are we from solving the 2D & 3D face alignment problem? (And a dataset of 230,000 3D facial landmarks)," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 1021–1030.

[73] I. Kalra, M. Singh, S. Nagpal, R. Singh, M. Vatsa, and P. B. Sujit, "DroneSURF: Benchmark dataset for drone-based face recognition," in *Proc. 14th IEEE Int. Conf. Autom. Face Gesture Recognit.*, May 2019, pp. 1–7.

[74] M. Binkowski, D. J. Sutherland, M. Arbel, and A. Gretton, "Demystifying MMD GANs," in *Proc. Int. Conf. Learn. Represent.*, 2018, pp. 1–15.

[75] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 586–595.

[76] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Oct. 2012.

[77] J. Deng, J. Guo, E. Ververas, I. Kotsia, and S. Zafeiriou, "RetinaFace: Single-shot multi-level face localisation in the wild," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 5203–5212.

[78] L. Yang et al., "HiFaceGAN: Face renovation via collaborative suppression and replenishment," in *Proc. ACM Int. Conf. Multimedia*, 2020, pp. 1551–1560.

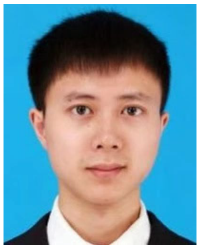
[79] X. Wang, L. Xie, C. Dong, and Y. Shan, "Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2021, pp. 1905–1914.

[80] J. He, W. Shi, K. Chen, L. Fu, and C. Dong, "GCFSR: A generative and controllable face super resolution method without facial and GAN priors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 1889–1898.

[81] Z. Wang, J. Zhang, R. Chen, W. Wang, and P. Luo, "RestoreFormer: High-quality blind face restoration from undegraded key-value pairs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 17512–17521.

[82] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 815–823.

**Hao Hou** received the B.Sc. degree in information and computing science from the School of Mathematical Science, University of Jinan, in 2019. He is currently pursuing the M.Sc. degree with the College of Intelligence and Information Engineering, Shandong University of Traditional Chinese Medicine, Jinan, China. His current research interests are in the areas of image processing and medical image analysis.



**Jun Xu** (Member, IEEE) received the B.Sc. and M.Sc. degrees from the School of Mathematics Science, Nankai University, Tianjin, China, and the Ph.D. degree from the Department of Computing, The Hong Kong Polytechnic University, in 2018. He worked as a Research Scientist at IIAI, Abu Dhabi, United Arab Emirates. He is an Associate Professor with the School of Statistics and Data Science, Nankai University.

**Yingkun Hou** (Senior Member, IEEE) received the Ph.D. degree from the School of Computer Science and Technology, Nanjing University of Science and Technology, in 2012. He is currently a Professor with the School of Information Science and Technology, Taishan University, Taian, China. His current research interests are in the areas of image processing, pattern recognition, and artificial intelligence.

**Xiaotao Hu** received the B.Sc. degree from the School of Software Engineering, Dalian University of Technology, in 2021. He is currently pursuing the M.Sc. degree with the College of Computer Science, Nankai University, Tianjin, China. His current research interests are in the areas of image processing and medical image processing and analysis.

**Benzheng Wei** received the B.S. degree in computer science from the School of Computer Science, Shandong Institute of Light Industry, Jinan, China, in 2000, the M.S. degree in computer science from the School of Computer Science and Technology, Shandong University, Jinan, in 2007, and the Ph.D. degree in precision instrument and machinery from the College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2013. He is a Professor with the Shandong University of Traditional Chinese Medicine. He is also acting as a Director of the Center for Medical Artificial Intelligence and the Computational Medicine Laboratory, Shandong University of Traditional Chinese Medicine. He has published over 80 papers in refereed international leading journals/conferences, such as *Medical Image Analysis*, IEEE TRANSACTIONS ON MEDICAL IMAGING, *Neurocomputing*, IPMI, and MICCAI. His current research interests are in artificial intelligence, medical information engineering, and computational medicine.

**Dinggang Shen** (Fellow, IEEE) received the Ph.D. degree in electrical engineering from Shanghai Jiao Tong University, Shanghai, China, in 1995. He is currently the Founding Dean of the School of Biomedical Engineering, ShanghaiTech University, Shanghai. He has published more than 1100 peer-reviewed papers in the international journals and conference proceedings, with H-index 127. His research interests include medical image analysis, computer vision, and pattern recognition. He served on the Board of Directors at the Medical Image Computing and Computer Assisted Intervention (MICCAI) Society, from 2012 to 2015. He is a fellow of the American Institute for Medical and Biological Engineering and the International Association for Pattern Recognition. He was the General Chair of MICCAI 2019. He serves as an editorial board member for eight international journals.