SiGAN: Siamese Generative Adversarial Network for Identity-Preserving Face Hallucination

Chih-Chung Hsu, *Member, IEEE*, Chia-Wen Lin, *Fellow, IEEE*, Weng-Tai Su, *Student Member, IEEE*, and Gene Cheung, *Senior Member, IEEE*

***Abstract*— Though generative adversarial networks (GANs) can hallucinate high-quality high-resolution (HR) faces from low-resolution (LR) faces, they cannot ensure identity preservation during face hallucination, making the HR faces difficult to recognize. To address this problem, we propose a Siamese GAN (SiGAN) to reconstruct HR faces that visually resemble their corresponding identities. On top of a Siamese network, the proposed SiGAN consists of a pair of two identical generators and one discriminator. We incorporate reconstruction error and identity label information in the loss function of SiGAN in a pairwise manner. By iteratively optimizing the loss functions of the generator pair and the discriminator of SiGAN, we not only achieve visually-pleasing face reconstruction but also ensure that the reconstructed information is useful for identity recognition. Experimental results demonstrate that SiGAN significantly outperforms existing face hallucination GANs in objective face verification performance while achieving promising visual-quality reconstruction. Moreover, for input LR faces with unseen identities that are not part of the training dataset, SiGAN can still achieve reasonable performance.**

***Index Terms*— Face hallucination, convolutional neural networks, generative adversarial networks, super-resolution, generative model.**

I. Introduction

FACE hallucination that super-resolves a low-resolution (LR) face image to a high-resolution (HR) one has become an attractive technique in upscaling face photos due to its practicality in multiple application domains that require face images with fine details, such as video surveillance, face recognition, face tracking, facial expression estimation, etc. However, simple interpolation schemes cannot reconstruct fine details. Instead, example-based super-resolution (SR) schemes [1] can better reconstruct fine details from a LR image compared to interpolation-based schemes, provided that a comprehensive set of training HR/LR image pairs are used to learn the structures and patterns of face image pairs based on machine learning techniques.

However, the face hallucination problem is inherently different from the generic image SR problem, because faces have unified structures with which humans are very familiar. Even small structural errors on the face during reconstruction can cause visually disturbing artifacts. For example, geometric distortion in the mouth and eyes in the reconstructed face may only slightly reduce the objective quality of the image, but can significantly harm the perceived subjective quality. Hence, both the global face shape and textures and the local geometric structures (e.g., mouth, nose, and eyes) need to be treated carefully in face hallucination [2], [3].

To recognize the identity of a LR face captured by a surveillance camera is a challenging problem, since face images are often taken at a distance, making their spatial resolutions too low to provide sufficiently discriminative features. Recently, empirical studies [4] in face recognition showed that a minimum face resolution between 32×32 and 64×64 is required for effective face recognition, and an even lower resolution would degrade recognition performance significantly for existing recognition models. For captured images using low-cost/power surveillance system, the resolution of the facial part can be extremely low. This work aims to reconstruct facial details to aid humans in identity recognition for LR faces captured in unconstrained wide-field surveillance videos and images (e.g., group photos, face images captured by low-cost/power cameras, face images captured at a long distance). Further, recently an unconstrained face recognition challenge has been announced [5], in which the size of the face image can be very low. As demonstrated in [5], the resolution of a captured face image can be from 69×84 to 6×8. Hence an effective face SR technique to improve the face recognition rate is highly desirable [5].

Most of existing face hallucination methods [3], however, just focused on hallucinating visually pleasing HR details without considering whether the added details are helpful in recognizing the identity of a face. As illustrated in Fig. 1(b), such identity-unaware reconstructed faces, though with a higher resolution, usually cannot help boost face recognition/verification accuracy. Instead, identity-aware face hallucination—on that can hallucinate identity-preserving facial details as shown in Fig. 1(c)—can much better serve this purpose. Identity-preserving reconstruction

Manuscript received July 22, 2018; revised January 29, 2019 and May 4, 2019; accepted June 4, 2019. Date of publication June 28, 2019; date of current version September 4, 2019. This work was supported in part by the Ministry of Science and Technology, Taiwan, under Grant MOST 108-2634-F-007-009, Grant 107-2218-E-020-002-MY3, and Grant 107-2218-E-006-059. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Nikolaos Mitianoudis. *(Corresponding author: Chia-Wen Lin.)*

C.-C. Hsu is with the Department of Management Information Systems, National Pingtung University of Science and Technology, Neipu 91203, Taiwan (e-mail: cchsu@mail.npust.edu.tw).

C.-W. Lin is with the Department of Electrical Engineering, National Tsing Hua University, Hsinchu 30013, Taiwan, and also with the Institute of Communications Engineering, National Tsing Hua University, Hsinchu 30013, Taiwan (e-mail: cwlin@ee.nthu.edu.tw).

W.-T. Su is with the Department of Electrical Engineering, National Tsing Hua University, Hsinchu 30013, Taiwan.

G. Cheung is with the Department of Electrical Engineering and Computer Science, York University, Toronto, ON M4N 3M6, Canada (e-mail: genec@yorku.ca).

Digital Object Identifier 10.1109/TIP.2019.2924554

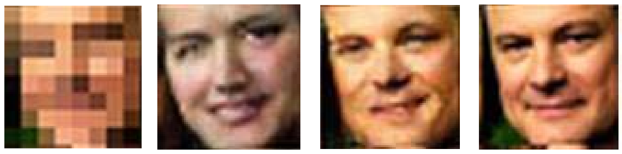


Fig. 1. Illustration of face hallucination: (a) input LR face (8×8), HR faces reconstructed by (b) identity-unaware face hallucination, (c) identity-aware face hallucination (our method), and (d) the ground-truth HR face.

is thus vital in face hallucination for many real-world applications [6], [7].

Hallucinating identify-preserving HR faces requires a labeled training set to learn identity-aware representations. There are many possible ways of embedding identity information in a face hallucination network. For example, one way is to embed full-class identity labels of the training faces (i.e., the true identity classes of individual training faces), which, however, usually consumes a huge labeling cost, especially for applications with a large-scale training set containing a large number of identities. Besides, giving true person identity labels often raises privacy concerns. Another alternative is to embed "weak" binary identity labels based on pairwise learning: a pair of two faces belonging to a same person, or belonging to different persons, thereby significantly reducing the labeling cost. Further, full-class label embedding may learn more discriminative representations if the number of training samples per class is large enough compared to weak binary-label embedding, but may find difficulty with a biased training set with some corner classes containing insufficient number of samples and has poorer scalability in training with newly added identity label classes.

To address the above problems, in this paper, we propose a novel Siamese generative adversarial network (SiGAN) to achieve visually-pleasing and identity-preserving HR face reconstruction. The training of our proposed SiGAN, thanks to its Siamese network structure, only relies on weak pairwise labels that signify whether a pair of two faces belongs to the same identity without the need to know the true identities of faces. Our contributions are summarized below:

- We propose a novel face hallucination GAN on top of a Siamese Network (namely SiGAN), upon which we can hallucinate HR faces to achieve high-quality and identity-preserving reconstruction.
- We embed weak binary pairwise label information in SiGAN without the need of full-class labels, which significantly reduces the labeling cost and increases the scalability of the method for super-revolving faces belonging to unseen identities.
- We conduct extensive experiments to evaluate the subjective and objective performances of the proposed SiGAN compared to several existing face hallucination GANs in terms of reconstruction quality and identity preservation ability.

The rest of this paper is organized as follows. Some most relevant works are surveyed in Sec. II. Sec. III presents the proposed SiGAN for identity-aware face hallucinations. In Sec. IV, experimental results are demonstrated. Finally, conclusions are drawn in Sec. V.

II. Related Work

Compared to traditional face hallucination schemes [3], deep learning-based approaches, particularly convolutional neural networks (CNNs), have proven to achieve state-of-the-art performance in face hallucination [7]–[13]. For example, a deep learning-based approach for joint face hallucination and recognition was proposed in [7], which consists of a SR network and a face recognition network. The two networks are optimized iteratively to achieve joint face hallucination and recognition. However, it adopts a relatively shallow CNN to hallucinate face images, resulting in possibly unsatisfactory visual quality of reconstructed faces. In contrast, [10] proposed a much deeper CNN to generate HR faces. To effectively upscale a LR face without introducing annoying artifacts, the method learns a dense correspondence field during training and upscale the LR face progressively by a cascading process. During a cascaded iteration, the dense correspondence field is first progressively refined with an increased face resolution, and then the face resolution is adaptively upsampled as guided by the refined dense correspondence field. To improve the fidelity of a hallucinated HR face, a two-stage method was proposed in [11], that reconstructs facial parts by using a deep CNN, followed by a fine-grained facial structure learner to further refine the reconstructed faces.

Recently, generative adversarial networks (GANs) have been successfully applied to various image processing applications such as image synthesis, image SR, and facial image generation [14]. A GAN is composed of a generator network and a discriminator network, in which the generator produces image contents based on a learned probability model, whereas the discriminator judges whether the generated contents are real or fake and decides to accept or reject the contents accordingly. By iterating the adversarial learning process between the generator and the discriminator, the generator will eventually be able to hallucinate high-quality image contents that can successfully confuse the discriminator.

For example, the SR GAN (SR-GAN) proposed in [15] was among the first to infer photo-realistic HR natural images for image SR. In SR-GAN, a perceptual loss function consisting of an adversarial loss term and a content loss term was proposed to push the solution to the natural image manifold using a discriminator network that is trained to differentiate between the super-resolved images and original photo-realistic images. However, this method is not suitable for face SR as explained in [9]. To overcome this problem, in [9] a pixel-wise L_2 regularization term is introduced in the generative model and the feedback of a discriminative network is exploited to make the upsampled face images more similar to real ones. Similarly, the method proposed in [8] utilizes deconvolutional layers to separately super-resolve the local and global parts and uses a discriminator to measure the visual quality of the hallucinated face image. The above-mentioned methods [8], [9], however, cannot guarantee faithful identity preservation of reconstructed faces since they do not provide any identity-aware guidance

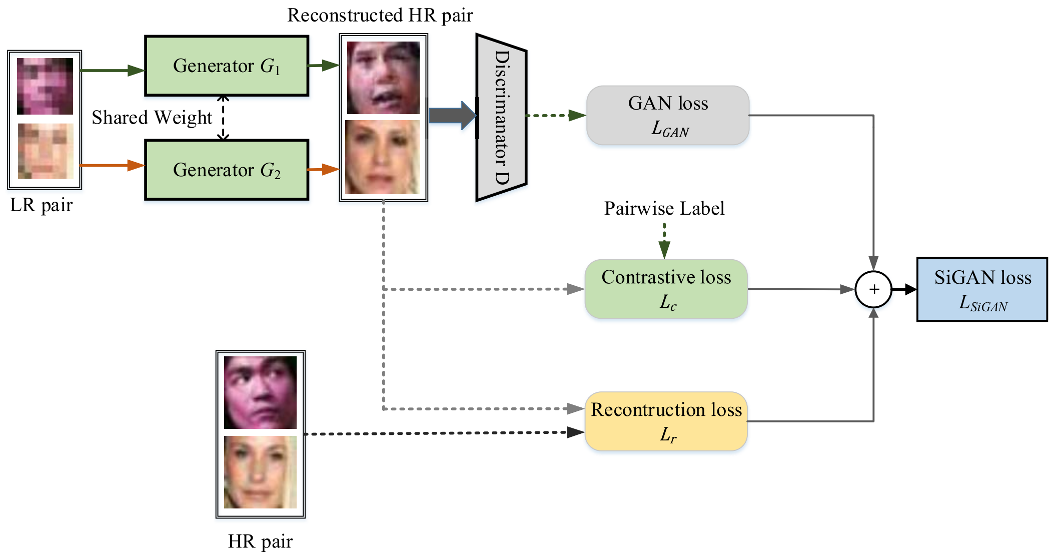


Fig. 2. Framework of the proposed Siamese GAN (SiGAN) with pairwise identity embedding for face hallucination.

to the learning of the discriminator/generator pair. Moreover, they often reconstruct unrealistic HR faces when the resolution of input LR face is extremely low, where much of facial structure information has been lost. In [16], a special network was proposed to learn the degradation models to make the face SR method suitable for real-world and arbitrary LR images. However, it also cannot guarantee identity-preserving reconstruction.

Similarly, [12] proposed an end-to-end GAN-based SR scheme which is combined with a face alignment network. The method utilizes heatmap loss to incorporate facial structural information by detecting facial landmarks so as to improve face hallucination results. The deep reinforcement learning method proposed in [13] hallucinates HR faces in an iterative reconstruction manner, that employs a recurrent policy network to reconstruct individual HR regions of a face based on previous reconstructions, followed by a local enhancement network to further refine facial details by considering the correlations between different facial parts. Reference [17] proposed a flexible neural network based on wavelet decomposition. The wavelet prediction network is then used to reconstruct smooth and detailed facial parts separately to improve the visual quality of a reconstructed face. In [18], the authors proposed a fine SR network with facial priors. It demonstrated that a proper facial prior can effectively improve the visual quality of a reconstructed face. Nevertheless, the above-mentioned methods [11]–[13] only focused on hallucinating visually pleasing HR details without considering whether the hallucinated HR details are helpful in recognizing the identity of a face.

III. Siamese GAN (SiGAN) for Identity-Aware Face Hallucination

A. Overview of Proposed SiGAN

To achieve visually-pleasing and identity-preserving reconstruction, as shown in Fig. 2, the proposed SiGAN adopts a pairwise learning scheme based on a Siamese network [19], which is composed of a pair of two identical spatial-upsampling generators G_1 and G_2 and a discriminator D. In the generator pair, a pair of LR faces are used as prior information to guide HR face generation. The generator pair are trained on same-identity and different-identity face pairs, each containing a pair of LR faces and their associated HR faces along with a binary pairwise identity indicator signifying whether the two LR faces belong to the same identity. To effectively learn identity-preserving representations, we introduce in the training process an identity-distinguishable contrastive energy function [20] which aims at decreasing the energy of same-identity pairs while increasing the energy of different-identity pairs. Combining the identity-distinguishable contrastive loss with the adversarial-loss and reconstruction-loss terms in SiGAN training can effectively boost the authenticity of reconstructed faces, while achieving good visual fidelity of hallucinated HR faces.

After training the SiGAN model using an iterative optimization process by minimizing the proposed loss function for both the discriminator and the generator pair, we can then use the learned generator to hallucinate HR faces from input LR faces, as elaborated below.

B. Network Models

SiGAN consists of a pair of twin generators, each comprising two/three residual blocks and upsampling blocks, followed by three convolutional layers and a sigmoid function, and a discriminator, which is a fully convolutional network. During training, the generator pair are used to hallucinate a pair of HR faces from a pair of input LR faces, and the discriminator is used to judge whether the two hallucinated HR faces are real or fake. The generator network and the discriminator network are described below.

1) Generator: The upper pipeline of Fig. 3 shows the generator of SiGAN based on ResNet in [21] is a SR CNN. In the

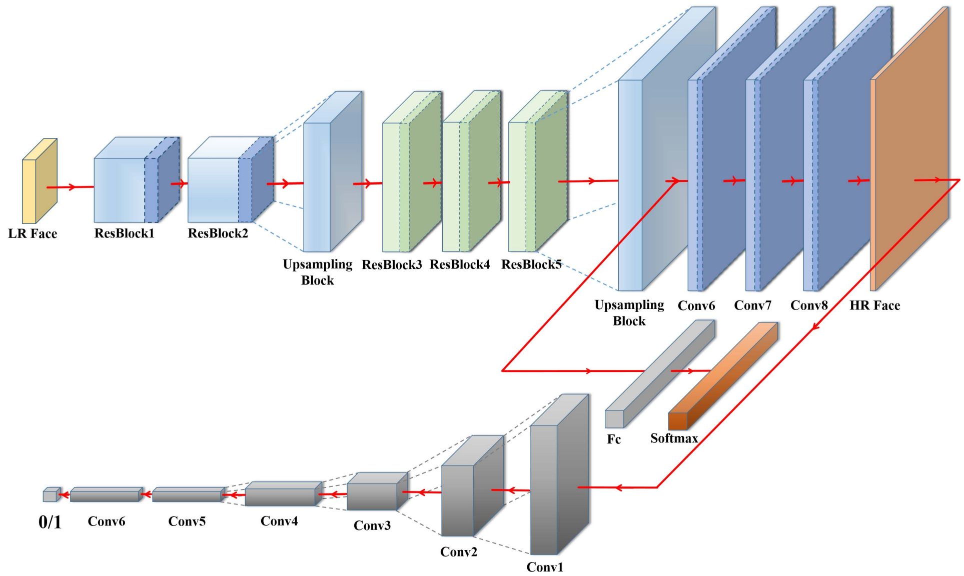


Fig. 3. Network models of the generator (the upper pipeline) and the discriminator (the lower pipeline) of SiGAN (ResNet).

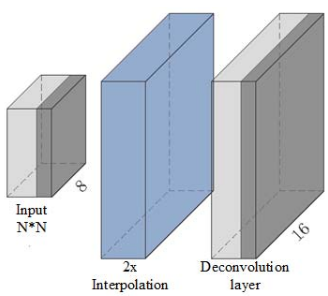


Fig. 4. Upsampler used in the generator of SiGAN.

generator, we insert two upsamplers to upscale the input faces by 4×. To effectively reconstruct HR faces, we replace the first two layers of the generator of DCGAN [22] with the residual blocks for faster convergence and better training performance. Then, an upsampler is inserted in between the second and the third layers to upscale the input feature maps. The third layer is then followed by three concatenated convolutional layers with a filter size of 3×3, and is finally concatenated with a convolutional layer with 1×1 kernels. Given an $N \times N$ face, the size of output face is $4N \times 4N$.

2) Upsampler: Since a CNN usually downscales the input image for extracting feature representations, for upscaling face images, as illustrated in Fig. 4, we adopt the upsampler proposed in [23] to gradually increase the spatial resolution layer by layer in the CNN. The image size is first linearly interpolated from $N \times N$ to $2N \times 2N$, followed by concatenating a batch normalization and an activation layer. Finally, a deconvolutional layer is used to learn deconvolution filters to produce a HR face with fine details.

3) Discriminator: Similar to the discriminator in DCGAN, as illustrated in the lower pipeline in Fig. 3, the discriminator is a fully convolutional network consisting of seven convolutional layers followed by an average polling layer. The output of the discriminator is a normalized value signifying whether the face generated by the generator is true or fake.

C. Training and Optimization

As illustrated in Fig. 2, the overall loss function $L_{SiGAN} = L_{GAN} + L_c + L_r$ contains three loss terms: the adversarial loss L_{GAN}, the reconstruction loss L_r, the contrastive loss L_c. The individual loss terms are elaborated below.

First, similar to [14], the adversarial loss L_{GAN} incorporated in a GAN is defined as

$$L_{GAN}(D, G) = E_D\left[\log D(\mathbf{x}_1^{HR})\right] + E_G\left[\log\left(1 - D(G(\mathbf{x}_1^{LR}))\right)\right], \quad (1)$$

where D and G represent the discriminator and generator, respectively, $G(\mathbf{x}^{LR})$ is the generative model used for hallucinating HR faces $\mathbf{x}^{SR}$, and $D(\mathbf{x})$ is the probability of data sample $\mathbf{x}$ being authenticated: $D(\mathbf{x}) = 1$ indicate that $\mathbf{x}$ is authenticated as a real sample; otherwise $D(\mathbf{x}) = 0$.

The reconstruction loss of the generator defined as the L_1 norm of the difference between a ground-truth HR face pair and its hallucinated version is used to maximize the fidelity of the reconstructed HR face pair:

$$L_r = \left\|\mathbf{x}_1^{HR} - \mathbf{x}_1^{SR}\right\|_1 + \left\|\mathbf{x}_2^{HR} - \mathbf{x}_2^{SR}\right\|_1, \quad (2)$$

where $\mathbf{x}_{y.i}^{HR}$ and $\mathbf{x}_i^{SR}$ respectively denote the HR ground-truth and its hallucinated version of the ith face of a training face pair.

To learn identity-aware features while training SiGAN, we introduce a contrastive loss term. Given ground-truth HR face pair $\mathbf{x}_1^{HR}$ and $\mathbf{x}_2^{HR}$ and the pairwise identity label y, where $y = 0$ indicates an impostor pair and $y = 1$ indicates a genuine pair, the contrastive loss L_c is defined as

$$L_c = (1 - y)L_I(E_w(\mathbf{x}_1^{SR}, \mathbf{x}_2^{SR})) + yL_G(E_w(\mathbf{x}_1^{SR}, \mathbf{x}_2^{SR})). \quad (3)$$

where $E_w = ||P(\mathbf{x}_1^{LR}) - P(\mathbf{x}_2^{LR})||_1^1$, $L_I = \frac{1}{2}[\max(0, m - E_w)]^2$, $L_G = \frac{1}{2}(E_W)^2$, and $m = 0.5$. We use the feature loss $E_w = ||P(\mathbf{x}_1^{LR}) - P(\mathbf{x}_2^{LR})||_1^1$ instead of the L_1 norm in the pixel domain (i.e., $E_w = ||\mathbf{x}_1^{SR} - \mathbf{x}_2^{SR}||_1^1$), because the feature distance is less sensitive to the variations in pose, lighting, and expression, thereby better capturing the semantic similarity between paired faces. The feature loss is obtained by concatenating a 128-neuron fully connected layer to the end of the second residual block to generate a 128-dimensional feature vector $P(\mathbf{x}^{LR})$ of input LR face $\mathbf{x}^{LR}$.

In (3), L_I is used to ensure that the distance between a different-identity paired faces will be larger than the predefined marginal value m. Moreover, L_G is used to keep the feature distance between a same-identity paired inputs as small as possible. In this manner, the contrastive loss term not only minimizes the marginal loss L_I between the reconstructed impostor pair $\mathbf{x}_1^{SR}$ and $\mathbf{x}_2^{SR}$, but also minimizes the loss L_G between the super-resolved genuine pair. If the reconstructed HR faces belong to different identities (i.e., $y = 0$), minimizing the contrastive loss E_C is equivalent to minimizing L_I. With the contrastive loss in (3), we can update the generator toward producing a better identity-preserving reconstruction.

By incorporating the three loss terms in (1), (2) and (3), the training of SiGAN is to solve the following min-max optimization problem.

$$\min_G \max_D L_{SiGAN}(D, G) = L_{GAN} + L_r + L_c, \tag{4}$$

We train SiGAN by iteratively optimizing the discriminator, generator, and contrastive loss functions using the stochastic gradient descent (SGD) algorithm proposed in [24]. In each iteration of optimization, we first update the discriminator by ascending its stochastic gradient calculated by

$$\nabla_{\theta_d} \frac{1}{b} \sum_{i=1}^{b} \left[\log D(\mathbf{x}_1^{HR})\right] + \left[\log\left(1 - D(G(\mathbf{x}_1^{LR}))\right)\right]. \tag{5}$$

Then, we update the generator pair by descending its gradient calculated by

$$\nabla_{\theta_g} \frac{1}{b} \sum_{i=1}^{b} \log\left(1 - D(G(\mathbf{x}_1^{LR}))\right). \tag{6}$$

Finally, we fix the updated results of the generator pair and discriminator, and update the generator pair based on the contrastive loss function by descending its gradient:

$$\nabla_{\theta_c} \frac{1}{b} \sum_{i=1}^{b} (1-y) L_I(E_w(P(\mathbf{x}_1^{LR}), P(\mathbf{x}_2^{LR}))) + y L_G(E_w(P(\mathbf{x}_1^{LR}), P(\mathbf{x}_2^{LR}))). \tag{7}$$

Taking several training epochs of the proposed SiGAN using SGD, we can learn the model of the generator pair that can hallucinate photo-realistic and identity-preserving HR faces.

IV. Experimental Results

For performance evaluation, we compare SiGAN with several existing methods including bicubic interpolation, ultra-resolution by deep facial component generation method (DFCG) [11], DCGAN [22], super-resolution GAN (SR-GAN) [15], Wavelet-SRNet [17], discriminative generative networks (UR-DGN) [9], and pixel recurrent super-resolution (PRSR) [25]. Since there is still no widely-accepted objective quality metric for face hallucination currently, besides the PSNR and SSIM (structural similarity) metrics for measuring reconstruction fidelity, we further perform face recognition and verification on reconstructed HR faces using the OpenFaces recognition engine [26], and use the face recognition/versification rate as an additional objective quality metric to evaluate whether the reconstructed HR details are useful for identity recognition. The compared methods are all trained and tested on a publicly available face dataset CASIA-WebFace [27] or simply CASIA. Besides the CASIA dataset, we also do performance evaluation against two faces-in-the-wild datasets: the Labeled Faces in the Wild (LFW) [28] and CelebA [29]. All face images are cropped to the size of 128×128 without any further preprocessing. The size of input LR face images is downscaled to 8×8 and 16×16 and then upscaled to 32×32 and 64×64, respectively, by various face hallucination schemes.

In order to demonstrate the effectiveness of the proposed Siamses architecture for face hallucination, we implement SiGAN on top of two CNN models, ResNet [21] and DenseNet [30], as the core of the generator pair, denoted "SiGAN (ResNet)" and "SiGAN (DenseNet)", respectively. In order to make a fair comparison between SiGAN (ResNet) and SiGAN (ResNet), we modify DenseNet by inserting a convolutional layer in between two successive blocks, so that the number of its last channels in a dense block is the same as that of the ResNet..

A. Subjective Visual Quality Evaluation

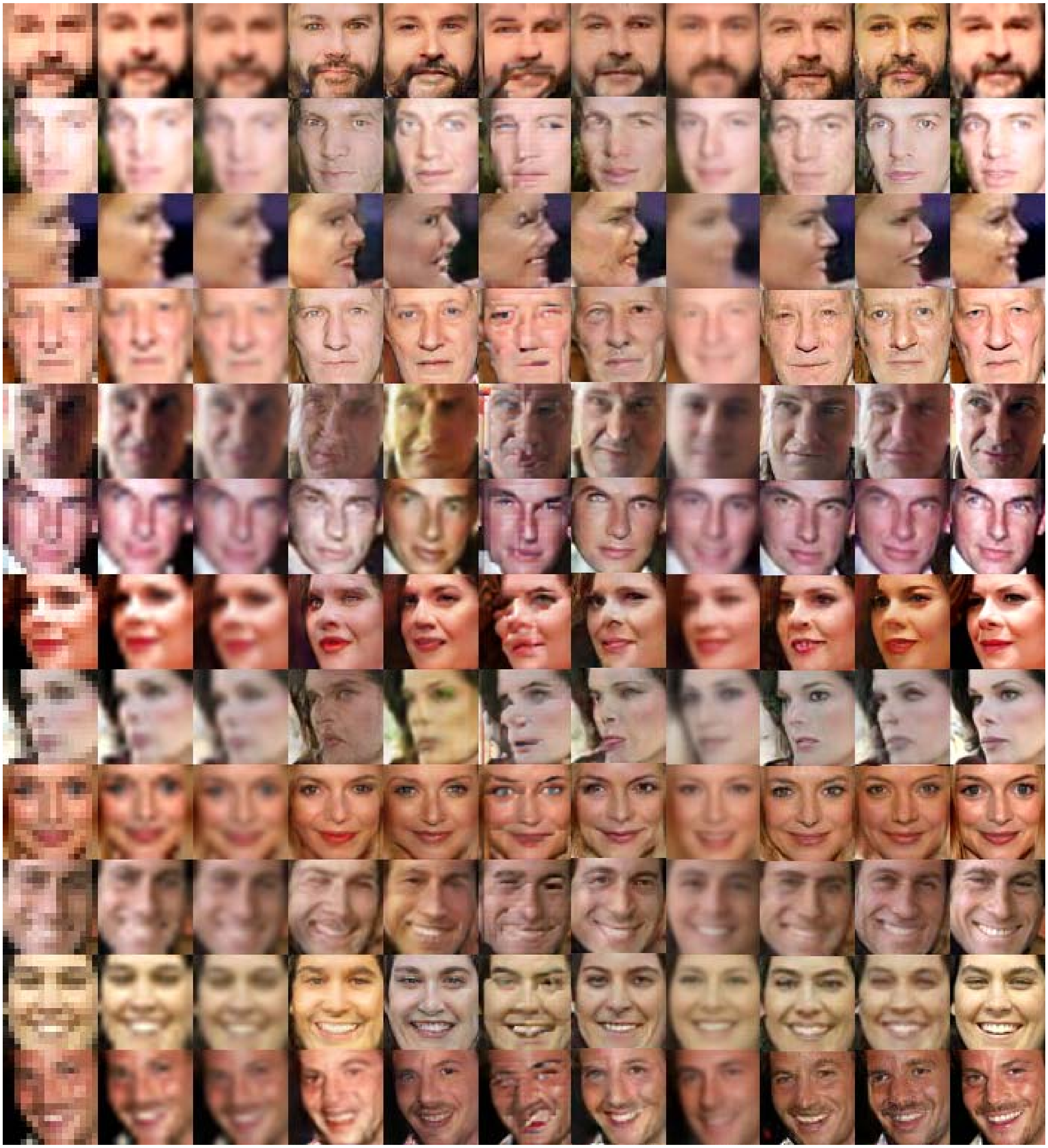
1) CASIA Dataset: The CASIA dataset [27] contains 494, 414 face images with various illuminations and poses captured from 10, 575 subjects. In each trial, we randomly select 491, 131 out of the 494, 414 face images for training and use the remaining 3, 283 images for testing. Fig. 5 illustrates the face hallucination results for 12 test faces upscaled from 8×8 to 32×32. In Fig. 5, since the resolution of LR faces is only 8×8, most of detailed facial information is missing. As a result, the HR faces reconstructed by DFCG [11] are blurry because the LR observations lack enough information for correctly estimating the initial facial parts, making the refiner in DFCG poorly hallucinate the HR details of facial parts. Wavelet-SRNet [17] also leads to blurry reconstructed HR faces. Although the DCGAN [22] can hallucinate fine details, the reconstructed HR faces are usually significantly dissimilar to their corresponding identities, as neither reconstruction loss nor identity information is considered in DCGAN. By contrast, UR-DGN [9] takes into account reconstruction loss in training to improve the fidelity of reconstructed HR faces, which, however, still leads to significantly dissimilar facial parts to their ground-truths due to the lack of identity information. Although PRSR [25] can produce fine and smooth details, it may generate severe artifacts if the initial HR face is not well inferred, which often causes serious error propagation in the succeeding step-by-step refinement. Besides, the lack of identity information in PRSR also makes the reconstructed HR faces poorly recognizable in identity. The visual quality of reconstructed HR faces using SR-GAN is poor when the size of input faces is 8×8, since the training of SR-GAN is based on a perceptual loss term, but such low input resolution makes

Fig. 5. Subjective visual quality comparison of various face hallucination methods for 12 identities selected from CASIA [27]. (a) The LR face images (8×8), (b)–(j) the reconstructed 32×32 HR faces using (b) bicubic interpolation, (c) DFCG [11], (d) DCGAN [22], (e) UR-DGN [9], (f) PRSR [25], (g) SR-GAN [15], (h) Wavelet-SRNet [17], (i) SiGAN (DenseNet), (j) SiGAN (ResNet), and (k) the ground-truths (32×32).

the learned features unrepresentative, thereby significantly reducing the effectiveness of perceptual loss.

Since SiGAN considers both the reconstruction loss and label information to overcome the above problems, besides successfully hallucinating fine details, the reconstructed HR facial parts more faithfully resemble their corresponding ground-truths. Fig. 6 illustrates the HR faces hallucinated from 16×16 to 64×64 for the same test faces in Fig. 5. Again, the results show that SiGAN outperforms the other schemes in both visual fidelity and authenticity of the reconstructed HR faces.

2) Faces in the Wild Datasets: Since in many applications the input LR faces often belong to unknown identities, we also evaluate the performances of hallucination methods on faces



(a) (b) (c) (d) (e) (f) (g) (h) (i) (j) (k)

Fig. 6. Subjective visual quality comparison of various face hallucination methods for 12 identities selected from CASIA [27]: (a) The LR face images (16×16), (b)–(j) the reconstructed 64×64 HR faces using (b) bicubic interpolation, (c) DFCG [11], (d) DCGAN [22], (e) UR-DGN [9], (f) PRSR [25], (g) SR-GAN [15], (h) wavelet-SRNet [17], (i) SiGAN (DenseNet), (j) SiGAN (ResNet), and (k) the ground-truths (64×64).

whose identities are not included in the training set to verify if these methods can be generalized to input faces with unseen identities. In the experiment, we randomly sample face images from two face-in-the-wild datasets, LFW [28] and CelebA [29], as test images to evaluate the generality of the compared methods which are all trained on the CASIA dataset. Fig. 7 illustrates the 8×8 to 32×32 face hallucination results of five difficult test faces (e.g., faces wearing glasses and non-frontal faces) selected from LFW [28] and CelebA [29]. We can observe that, for the tiny 8×8 faces with large poses and/or wearing sunglasses, all the compared methods produce visible artifacts on the HR faces, because the numbers of

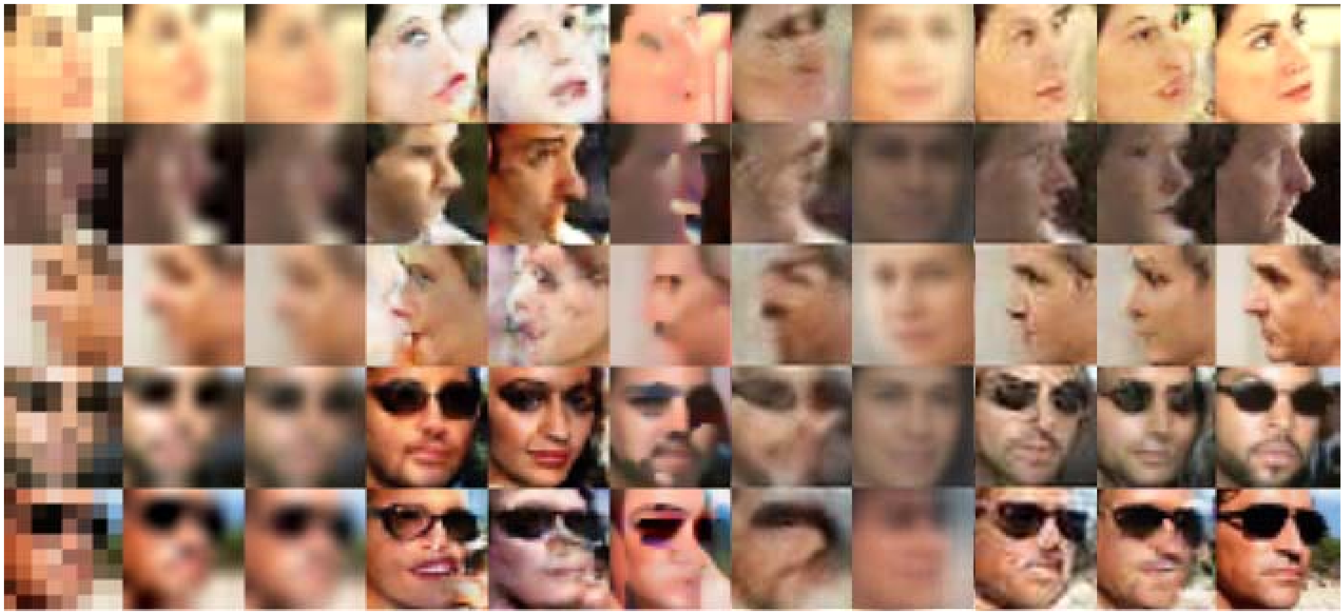


Fig. 7. Subjective visual quality comparison for five faces with unknown identities selected from LFW [28] and CelebA [29]: (a) the LR face images (8×8), (b)–(j) the reconstructed 32×32 HR faces using (b) bicubic interpolation, (c) DFCG [11], (d) DCGAN [22], (e) UR-DGN [9], (f) PRSR [25], (g) SR-GAN [15], (h) Wavelet-SRNet [17], (i) SiGAN (DenseNet), (j) SiGAN (ResNet), and (k) the ground-truths (32×32).

Fig. 8. Subjective visual quality comparison for five faces with unknown identities selected from LFW [28] and CelebA [29]: (a) the LR face images (16×16), (b)–(j) the reconstructed 64×64 HR faces using (b) bicubic interpolation, (c) DFCG [11], (d) DCGAN [22], (e) UR-DGN [9], (f) PRSR [25], (g) SR-GAN [15], (h) Wavelet-SRNet [17], (i) SiGAN (DenseNet), (j) SiGAN (ResNet), and (k) the ground-truths (64×64).

training samples in CASIA for such types of faces are very limited, making the generator difficult to well learn the face structures. For example, the fifth test face not only wears glasses but also involves some background information. In this case, all methods fail to correctly hallucinate HR facial parts. Nevertheless, compared to the other methods, SiGAN still achieves significantly better visual qualities. Fig. 8 shows the HR faces hallucinated from 16×16 to 64×64 for the same identities in Fig. 7. SiGAN achieves the best performance as well.

3) Comparison Between SiGAN and Its Baselines: We also compare SiGAN with two baseline methods to demonstrate the effectiveness of the proposed loss terms of SiGAN. We first replace the feature distance of SiGAN in the feature domain with the pixel distance measured in the pixel domain called "Baseline-I". We also remove the reconstruction loss

(a) (b) (c) (d) (e) (f) (g) (h) (i)

Fig. 9. Subjective visual quality comparison of the proposed method and the baselines: (a) the LR face images (8×8), (b)–(g) the reconstructed 32×32 HR faces using (b) bicubic interpolation, (c) baseline-I (ResNet), (d) baseline-II (ResNet), (e) baseline-I (DenseNet), (f) baseline-II (DenseNet), (g) SiGAN (DenseNet), (h) SiGAN (ResNet), and (i) the ground-truths (32×32).

from the loss function of SiGAN to obtain another baseline model called "Baseline-II." Besides, since the baseline model of SiGAN without the contrastive loss is equivalent to the DCGAN or SR-GAN with different network architectures, there is no need to conduct an additional experiment for this baseline method. Fig. 9 shows that Baseline-II severely distorts the reconstructed HR faces due to its lack of the reconstruction loss term in the training phase, indicating the high importance of reconstruction loss for face hallucination or super-resolution for both SiGAN (ResNet) and SiGAN (DenseNet). Although Baseline-I does a good job in identity preservation, it produces some artifacts on the reconstructed faces since the intra-person variation of the training images of an identity may be large, making the pixel-domain distance ineffective to capture the similarity between two paired faces. In contrast, the proposed feature distance used in the contrastive loss does a better job in measuring the similarity between two paired faces.

B. Objective Quality Evaluation

To evaluate the degree of authenticity of reconstructed HR faces compared to their ground-truth identity, we use a state-of-the-art CNN-based face recognition engine, OpenFaces [26], to evaluate the face recognition rate and verification rate for HR faces reconstructed by various face hallucination methods. We adopt two objective evaluation approaches. First, we employ OpenFaces [26] trained from training HR faces of CASIA to recognize the identities of the reconstructed HR faces and calculate the identity recognition rate. Second, following the standard face verification methodology described in [26], based on pair matching, we evaluate the accuracy of reconstructed HR faces being verified by OpenFaces as the same identity with their corresponding ground-truth face. We then further use the PSNR and SSIM metrics to evaluate the fidelity of a reconstructed HR faces compared with its ground-truth. All these metrics are used to evaluate the objective performances of various face hallucination methods against CASIA [27] and LFW [28]. Since CelebA does not provide identity labels, it is not used in the objective evaluation.

1) Face Recognition Performance Comparison: For the experiments on CASIA, we randomly sample $144,942$ images belonging to 671 identities to train OpenFaces. We then sample $2,000$ face images from the remaining images as the test dataset to evaluate the face recognition performance. Since the number of face images of some identities in CASIA is small, we only choose those identities with more than 120 face images in the dataset, as suggested in [31]. For the experiment on LFW, we first randomly sample $11,000$ face image belonging to 680 identities as the training set, and sample $2,000$ face images from the remaining as the test dataset. To train OpenFaces, all face images are resized to 96×96, as suggested in [26]. Similarly, in the testing stage, all hallucinated HR faces and LR faces are resized to 96×96.

We first evaluate the face recognition rates on hallucinated HR faces whose identities are included in the training set. Table I(a) compares the top-1, top-5, and top-10 face recognition rates for 32×32 HR faces upscaled from 8×8 LR faces by various methods. The result shows that, as evaluated by OpenFaces, the average recognition rate for the HR faces reconstructed by SiGAN is significantly higher than those achieved by the other methods. Among the remaining methods, compared to bicubic interpolation, UR-DGN [9], DFCG [11], DCGAN [22], and SR-GAN [15] all degrade face recognition performance, meaning that the HR details reconstructed by these methods are usually useless and even harmful for identity recognition. Table I(b) compares the average face recognition rates for 64×64 HR faces upscaled from 16×16 LR faces using various methods. Similarly, SiGAN achieves the best average face recognition performances.

Since in many applications an input LR face usually belongs to an unknown identity, Table II compares the performances of various hallucination methods on faces randomly sampled from LFW whose identities are not included in the training set of OpenFaces to verify the generality of these methods

TABLE I

COMPARISON OF FACE RECOGNITION RATES EVALUATED BY OPENFACES [26] FOR HR FACES RECONSTRUCTED BY VARIOUS FACE HALLUCINATION METHODS ON CASIA [27] BY UPSCALING: (a) FROM 8×8 TO 32×32; (b) FROM 16×16 TO 64×64

(a)

Method	Top-1	Top-5	Top-10
HR (32×32)	30.4%	51.2%	59.6%
LR (8×8)	10.7%	19.5%	33.1%
Bicubic	10.8%	20.1%	34.4%
DFCG [11]	9.3%	17.7%	21.4%
UR-DGN [9]	9.9%	18.6%	22.7%
DCGAN [22]	4.6%	10.9%	16.8%
PRSR [25]	10.8%	18.8%	24.4%
SR-GAN [15]	8.8%	11.1%	19.4%
Wavelet-SRNet [17]	12.8%	20.2%	30.3%
SiGAN (ResNet)	15.8%	27.5%	40.4%
SiGAN (DenseNet)	15.1%	26.8%	40.3%

(b)

Method	Top-1	Top-5	Top-10
HR (64×64)	36.8%	55.9%	63.8%
LR (16×16)	12.4%	27.4%	37.1%
Bicubic	11.6%	27.5%	37.6%
DFCG [11]	9.6%	23.7%	34.8%
UR-DGN [9]	12.2%	29.0%	38.7%
DCGAN [22]	9.3%	24.9%	33.9%
PRSR [25]	13.3%	29.7%	40.1%
SR-GAN [15]	11.6%	23.2%	36.3%
Wavelet-SRNet [17]	12.0%	25.5%	38.8%
SiGAN (ResNet)	17.9%	32.9%	48.1%
SiGAN (DenseNet)	18.3%	33.5%	50.0%

TABLE II

COMPARISON OF FACE RECOGNITION RATES EVALUATED BY OPENFACSES [26] FOR HR FACES RECONSTRUCTED BY VARIOUS FACE HALLUCINATION METHODS ON LFW [28] BY UPSCALING: (a) FROM 8×8 TO 32×32; (b) FROM 16×16 TO 64×64

(a)

Method	Top-1	Top-5	Top-10
HR (32×32)	32.2%	50.8%	56.7%
LR (8×8)	9.3%	17.4%	30.9%
Bicubic	9.6%	17.7%	30.4%
DFCG [11]	9.3%	16.9%	27.5%
UR-DGN [9]	7.9%	16.8%	20.1%
DCGAN [22]	4.7%	9.9%	14.6%
PRSR [25]	10.3%	19.8%	26.1%
SR-GAN [15]	9.1%	13.3%	22.6%
Wavelet-SRNet [17]	13.1%	22.7%	32.0%
SiGAN (ResNet)	14.5%	26.7%	39.2%
SiGAN (DenseNet)	15.3%	26.9%	40.7%

(b)

Method	Top-1	Top-5	Top-10
HR (64×64)	35.4%	51.4%	60.1%
LR (16×16)	14.8%	26.6%	35.3%
Bicubic	15.0%	26.4%	35.6%
DFCG [11]	13.2%	25.4%	34.7%
UR-DGN [9]	15.9%	30.2%	39.4%
DCGAN [22]	11.6%	24.3%	32.6%
PRSR [25]	18.3%	32.6%	45.5%
SR-GAN [15]	12.6%	26.5%	38.8%
Wavelet-SRNet [17]	15.1%	27.1%	40.2%
SiGAN (ResNet)	21.5%	40.5%	50.2%
SiGAN (DenseNet)	20.6%	38.8%	47.6%

to unseen faces. Again, SiGAN achieves the best average recognition rates, showing that even for unseen faces, it can still effectively reconstruct identity-preserving facial details.

TABLE III

COMPARISON OF FACE VERIFICATION AUC RATES EVALUATED BY OPENFACES [26] FOR VARIOUS FACE HALLUCINATION METHODS ON CASIA [27]

Methods	8×8 to 32×32	16×16 to 64×64
HR	83.3%	92.7%
LR	64.1%	64.3%
Bicubic	64.8%	63.7%
DFCG [11]	63.7%	64.0%
UR-DGN [9]	64.5%	67.7%
DCGAN [22]	60.9%	60.8%
PRSR [25]	70.0%	71.1%
SR-GAN [15]	66.0%	69.2%
Wavelet-SRNet [17]	70.2%	73.3%
SiGAN (ResNet)	81.2%	82.8%
SiGAN (DenseNet)	80.8%	82.6%

TABLE IV

COMPARISON OF FACE VERIFICATION AUC RATES EVALUATED BY OPENFACES [26] FOR VARIOUS FACE HALLUCINATION METHODS ON LFW [28]

Methods	8×8 to 32×32	16×16 to 64×64
HR	97.6%	98.8%
LR	70.7%	75.4%
Bicubic	70.8%	75.7%
DFCG [11]	68.6%	73.9%
UR-DGN [9]	67.7%	72.8%
DCGAN [22]	64.9%	74.8%
PRSR [25]	69.6%	76.9%
SR-GAN [15]	63.0%	66.8%
Wavelet-SRNet [17]	67.4%	70.7%
SiGAN (ResNet)	82.9%	83.4%
SiGAN (DenseNet)	81.8%	83.8%

2) Face Verification Performance Comparison: In this experiment, we first randomly sample $500,000$ and $200,000$ face pairs from CASIA and LFW, respectively, as the training sets to train the OpenFaces recognition engine with the settings specified in [26]. We then randomly sample $6,000$ faces from the remaining images of CASIA and LFW, respectively, as the test set to evaluate the face verification performance.

We first evaluate the area under curve (AUC) [28] of the trained face verification system for the hallucinated HR faces associated with identities that are included in the training set. Table III compares the AUC rates for 32×32 and 64×64 HR faces respectively reconstructed from 8×8 and 16×16 LR faces using various face hallucination methods. The result shows that, as evaluated by the OpenFaces recognition engine [26], the AUC for the HR faces reconstructed by SiGAN is significantly higher than those achieved by the other methods, meaning that SiGAN achieves a significantly higher degree of authenticity of reconstructed HR faces to their ground-truth identity. Table IV compares the AUC rates of various face hallucination methods on LFW. Again, SiGAN achieves the best AUC performance.

3) Reconstruction Fidelity Comparison: Table V(a) and (b) shows the PSNR and SSIM values of HR faces reconstructed by the compared face hallucination methods on CASIA and LFW, respectively.

C. Run-Time Complexity Analysis

Moreover, we compare the run-time complexity in the testing stage as shown in Table VI. The experiment is performed on a personal computer equipped with Intel Core i7-8700K

TABLE V

COMPARISON OF PSNR AND SSIM OF HR FACES RECONSTRUCTED BY VARIOUS FACE HALLUCINATION METHODS ON (a) CASIA [27] AND (b) LFW [28]

(a)

Method	8×8 to 32×32		16×16 to 64×64	
	PSNR	SSIM	PSNR	SSIM
Bicubic	23.7 dB	0.610	26.6 dB	0.701
DFCG [11]	22.2 dB	0.645	26.6 dB	0.664
UR-DGN [9]	21.5 dB	0.733	27.6 dB	0.778
DCGAN [22]	20.6 dB	0.693	23.1 dB	0.777
PRSR [25]	21.1 dB	0.779	28.7 dB	0.819
SR-GAN [15]	22.4 dB	0.684	23.3 dB	0.736
Wavelet-SRNet [17]	23.9 dB	0.687	24.3 dB	0.705
SiGAN (ResNet)	24.8 dB	0.807	25.3 dB	0.811
SiGAN (DenseNet)	24.1 dB	0.787	25.6 dB	0.816

(b)

Method	8×8 to 32×32		16×16 to 64×64	
	PSNR	SSIM	PSNR	SSIM
Bicubic	22.4 dB	0.629	23.3 dB	0.665
DFCG [11]	22.7 dB	0.633	23.3 dB	0.688
UR-DGN [9]	22.2 dB	0.687	23.0 dB	0.709
DCGAN [22]	18.7 dB	0.575	21.6 dB	0.622
PRSR [25]	22.1 dB	0.700	22.3 dB	0.712
SR-GAN [15]	21.4 dB	0.676	22.1 dB	0.689
Wavelet-SRNet [17]	22.1 dB	0.691	22.4 dB	0.702
SiGAN (ResNet)	23.8 dB	0.797	24.0 dB	0.805
SiGAN (DenseNet)	22.6 dB	0.768	24.8 dB	0.811

TABLE VI

RUN-TIME COMPLEXITY COMPARISON IN HALLUCINATING ONE HR FACES OF SiGAN AND THE COMPARED METHODS

Method	32×32	64×64
DFCG [11]	14.24 s	21.65 s
UR-DGN [9]	0.61 s	0.89 s
DCGAN [22]	0.55 s	0.96 s
PRSR [25]	227.12 s	1091.78 s
SR-GAN [15]	0.41 s	0.57 s
Wavelet-SRNet [17]	0.89 s	1.05 s
SiGAN (ResNet)	0.71 s	0.92 s
SiGAN (DenseNet)	0.35 s	0.48 s

CPU and a NVIDIA GeForce GTX 1080 Ti GPU. Among these methods, PRSR is based on a pixel-recurrent structure that needs to predict every pixel during hallucination, thereby consuming significantly longer time compared to the others. DFCG consumes the second highest computational complexity. In contrast, SiGAN (DenseNet) achieves the fastest processing speed and the remaining methods consume similar computational complexities.

V. CONCLUSION

We proposed an identity-preserving Siamese face hallucination GAN based on a novel pairwise learning scheme to effectively capture identity-aware facial representations for reconstructing photo-realistic and identity-preserving HR faces. We have also proposed a new loss function that integrates a reconstruction loss term, a pairwise identity loss term, and a GAN loss term to guide the training of the proposed SiGAN to significantly improve the realism of a hallucinated face and its authenticity to the identity. Experimental results demonstrate that our methods based on ResNet and DenseNet both significantly outperform state-of-the-art face hallucination networks in terms of objective face recognition/verification rate, while still achieving visually-pleasant reconstruction subjectively.

REFERENCES

[1] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 9, pp. 1167–1183, Sep. 2002.

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

[3] N. Wang, D. Tao, X. Gao, X. Li, and J. Li, "A comprehensive survey to face hallucination," *Int. J. Comput. Vis.*, vol. 106, no. 1, pp. 9–30, 2014.

[4] W. W. W. Zou and P. C. Yuen, "Very low resolution face recognition problem," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 327–340, Jan. 2012.

[5] Z. Cheng, X. Zhu, and S. Gong, "Surveillance face recognition challenge," Apr. 2018, *arXiv:1804.09691*. [Online]. Available: https://arxiv.org/abs/1804.09691

[6] P. H. Hennings-Yeomans, S. Baker, and B. V. K. V. Kumar, "Simultaneous super-resolution and feature extraction for recognition of low-resolution faces," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[7] J. Wu, S. Ding, W. Xu, and H. Chao, "Deep joint face hallucination and recognition," Nov. 2016, *arXiv:1611.08091*. [Online]. Available: https://arxiv.org/abs/1611.08091

[8] O. Tuzel, Y. Taguchi, and J. R. Hershey, "Global-local face upsampling network," Mar. 2016, *arXiv:1603.07235*. [Online]. Available: https://arxiv.org/abs/1603.07235

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

[10] S. Zhu, S. Liu, C. C. Loy, and X. Tang, "Deep cascaded bi-network for face hallucination," in *Proc. Eur. Conf. Comput. Vis.* Amsterdam, The Netherlands: Springer, 2016, pp. 614–630.

[11] Y. Song, J. Zhang, S. He, L. Bao, and Q. Yang, "Learning to hallucinate face images via component generation and enhancement," in *Proc. Int. Joint Conf. Artif. Intell.*, 2017, pp. 4537–4543.

[12] 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 Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 109–117.

[13] Q. Cao, L. Lin, Y. Shi, X. Liang, and G. Li, "Attention-aware face hallucination via deep reinforcement learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 690–698.

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

[15] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," Sep. 2016, *arXiv:1609.04802*. [Online]. Available: https://arxiv.org/abs/1609.04802

[16] 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.*, Sep. 2018, pp. 185–200.

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

[18] Y. Chen, Y. Tai, X. Liu, C. Shen, and J. Yang, "FSRNet: End-to-end learning face super-resolution with facial priors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2492–2501.

[19] J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, and R. Shah, "Signature verification using a 'Siamese' time delay neural network," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 737–744.

[20] S. Chopra, R. Hadsell, and Y. LeCun, "Learning a similarity metric discriminatively, with application to face verification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, pp. 539–546.

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

[22] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," Nov. 2015, *arXiv:1511.06434*. [Online]. Available: https://arxiv.org/abs/1511.06434

[23] A. Odena, V. Dumoulin, and C. Olah. (2016). Deconvolution and Checkerboard Artifacts. Distill. [Online]. Available: http://distill.pub/2016/deconv-checkerboard.

[24] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," Dec. 2014, *arXiv:1412.6980*. [Online]. Available: https://arxiv.org/abs/1412.6980

[25] R. Dahl, M. Norouzi, and J. Shlens, "Pixel recursive super resolution," Feb. 2017, *arXiv:1702.00783*. [Online]. Available: https://arxiv.org/abs/1702.00783

[26] B. Amos, B. Ludwiczuk, and M. Satyanarayanan, "Openface: A general-purpose face recognition library with mobile applications," School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA, Tech. Rep. CMU-CS-16-118, 2016.

[27] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Learning face representation from scratch," Nov. 2014, *arXiv:1411.7923*. [Online]. Available: https://arxiv.org/abs/1411.7923

[28] E. Learned-Miller, G. B. Huang, A. RoyChowdhury, H. Li, and G. Hua, "Labeled faces in the wild: A survey," in *Advances in Face Detection and Facial Image Analysis*. Amsterdam, The Netherlands: Springer, 2016, pp. 189–248.

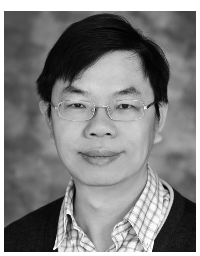
[29] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proc. Int. Conf. Comput. Vis.*, Dec. 2015, pp. 3730–3738.

[30] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 4700–4708.

[31] C. Ding and D. Tao, "Robust face recognition via multimodal deep face representation," *IEEE Trans. Multimedia*, vol. 17, no. 11, pp. 2049–2058, Nov. 2015.

Chih-Chung Hsu (S'11–M'14) received the B.S. degree in information management from the Ling-Tung University of Science and Technology, Taiwan, in 2004, the M.S. degree from the National Yunlin University of Science and Technology, in 2007, and the Ph.D. degree in electrical engineering from National Tsing Hua University (NTHU), Taiwan, in 2014.

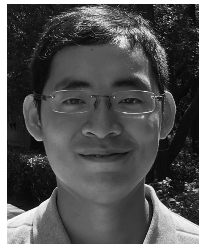
He was a Postdoctoral Researcher with the Institute of Communications Engineering, NTHU, from 2014 to 2017. He has been an Assistant Professor with the Department of Management Information Systems, National Pingtung University of Science and Technology, since 2018. His research interests mainly lie in computer vision and machine learning with applications to image and video processing. He received the Top 10% Paper Award from the IEEE MMSP 2013 and the First-Place Award from the ACM Multimedia Social Media Prediction Challenge, in 2017.



Chia-Wen Lin (S'94–M'00–SM'04–F'18) received the Ph.D. degree in electrical engineering from National Tsing Hua University (NTHU), Hsinchu, Taiwan, in 2000.

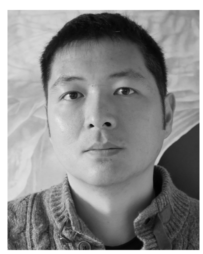
He is currently a Professor with the Department of Electrical Engineering, Institute of Communications Engineering, NTHU, where he is also the Deputy Director of the AI Research Center. He was with the Department of Computer Science and Information Engineering, National Chung Cheng University, Taiwan, from 2000 to 2007. Prior to joining academia, he was with the Information and Communications Research Laboratories, Industrial Technology Research Institute, Hsinchu, from 1992 to 2000. His research interests include image and video processing, computer vision, and video networking.

Dr. Lin has served as an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE MULTIMEDIA, and the *Journal of Visual Communication and Image Representation*. He was a Steering Committee Member of the IEEE TRANSACTIONS ON MULTIMEDIA from 2014 to 2015. He has been serving as a Distinguished Lecturer for the IEEE Circuits and Systems Society since 2018. He has been serving as the President for the Chinese Image Processing and Pattern Recognition Association, Taiwan, since 2019. He was the Chair of the Multimedia Systems and Applications Technical Committee, IEEE Circuits and Systems Society, from 2013 to 2015. He has served as the Technical Program Co-Chair for the IEEE ICME 2010. He will be the General Co-Chair of the IEEE VCIP 2018 and the Technical Program Co-Chair of the IEEE ICIP 2019. He has received the Young Investigator Award presented by the Ministry of Science and Technology, Taiwan, in 2006. His papers have received the Young Investigator Award from VCIP 2005, the Top 10% Paper Award from the IEEE MMSP 2013, and the Best Paper Award from the IEEE VCIP 2015.



Weng-Tai Su received the B.S. degree in electrical engineering from the National Yunlin University of Science and Technology, Yunlin, Taiwan, in 2012, Taiwan, the M.S. degree in electrical engineering from National Tsing Hua University (NTHU), Hsinchu, Taiwan, in 2014, where he is currently pursuing the Ph.D. degree with the Department of Electrical Engineering.

His research interests mainly lie in machine learning, image and video processing, and computer vision.



Gene Cheung (M'00–SM'07) received the B.S. degree in electrical engineering from Cornell University in 1995, and the M.S. and Ph.D. degrees in electrical engineering and computer science from the University of California at Berkeley, Berkeley, in 1998 and 2000, respectively.

He was a Senior Researcher with Hewlett-Packard Laboratories, Tokyo, Japan, from 2000 to 2009. He was a Faculty Member of the National Institute of Informatics, Tokyo, from 2009 to 2018. He is currently an Associate Professor with the Department of Electrical Engineering and Computer Science, York University, Toronto, Canada. He has been an Adjunct Associate Professor with The Hong Kong University of Science and Technology (HKUST) since 2015. His research interests include 3D imaging and graph signal processing.

Dr. Cheung has served as an Associate Editor for the IEEE TRANSACTIONS ON MULTIMEDIA from 2007 to 2011, DSP Applications Column of the *IEEE Signal Processing Magazine* from 2010 to 2014, the *Journal of Electronic Imaging* (SPIE) from 2014 to 2016, and the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY from 2016 to 2017. He has been serving as an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING since 2015 and the *APSIPA Journal on Signal and Information Processing* since 2011, and as an Area Editor for the *Signal Processing: Image Communication* (EURASIP) since 2011. He has served as a member for the Multimedia Signal Processing Technical Committee (MMSP-TC), IEEE Signal Processing Society, from 2012 to 2014, and the Image, Video, and Multidimensional Signal Processing Technical Committee (IVMSP-TC) from 2015 to 2017 and from 2018 to 2020. He is the coauthor of the Best Student Paper Award from IEEE Workshop on Streaming and Media Communications 2011 (in conjunction with ICME 2011), ICIP 2013, IVMSP 2016, and ICIP 2017, the Best Paper Runner-Up Award from ICME 2012, and the Best Paper Award from the IEEE Signal Processing Society (SPS) Japan 2016.