Identity-Aware Deep Face Hallucination via Adversarial Face Verification

Hadi Kazemi
West Virginia University
hakazemi@mix.wvu.edu

Fariborz Taherkhani
West Virginia University
fariborztaherkhani@gmail.com

Nasser M. Nasrabadi
West Virginia University
nasser.nasrabadi@mail.wvu.edu

Abstract

In this paper, we address the problem of face hallucination by proposing a novel multi-scale generative adversarial network (GAN) architecture optimized for face verification. First, we propose a multi-scale generator architecture for face hallucination with a high up-scaling ratio factor, which has multiple intermediate outputs at different resolutions. The intermediate outputs have the growing goal of synthesizing small to large images. Second, we incorporate a face verifier with the original GAN discriminator and propose a novel discriminator which learns to discriminate different identities while distinguishing fake generated HR face images from their ground truth images. In particular, the learned generator cares for not only the visual quality of hallucinated face images but also preserving the discriminative features in the hallucination process. In addition, to capture perceptually relevant differences we employ a perceptual similarity loss, instead of similarity in pixel space. We perform a quantitative and qualitative evaluation of our framework on the LFW and CelebA datasets. The experimental results show the advantages of our proposed method against the state-of-the-art methods on the 8x downsampled testing dataset.

1. Introduction

Face hallucination methods deal with super-resolving a low-resolution (LR) face image and generating a high-resolution (HR) one. They have many applications such as face recognition, face tracking, security in surveillance video, and facial expression estimation. One of the most common issues in the practical face recognition systems is that they have low performance on low-resolution face images captured in the wild. Especially in the standard surveillance videos, detected faces might have a resolution of 20×20 pixels or smaller [49]. Such LR face images negatively affect the performance of the subsequent face recognition and analysis. Consequently, in the past few years generating HR face images from LR ones has attracted great research interests.

Traditional interpolation techniques, such as the nearest neighbor or bilinear up-scaling, are not able to reconstruct high-frequency details. On the contrary, frameworks which are based on example-based super-resolution (SR) schemes [4] have shown a good performance in fine detailed reconstruction from a LR image compared to the interpolation-based methods. This capacity is acquired by learning the patterns, textures, and geometrical characteristics of face images based on different machine learning techniques trained on a comprehensive pair of training HR/LR images.

Unlike the natural images, SR face hallucination images have similar structures. Employing only a reconstruction error may result in faces with visually undesirable artifacts. For example, small geometry distortion in face components which plays a critical role in person identification, such as the mouth and eyes, can degrade the subjective quality of the face hallucination. Therefore, the global face shape and local characteristics such as textures and local geometric structures (e.g., nose and eyes) need to be handled cautiously in face hallucination [3, 39].

Surveillance cameras, which usually provide low-resolution images, especially for small objects of interest such as faces taken at a distance, makes face identification a more challenging problem. This is due to the lack of sufficiently discriminative features in low-resolution face images. An empirical study [52] showed that for effective face identification, the minimum face resolution should be between 32 x 32 and 64 x 64 pixels. Thus, a lower resolution face will significantly degrade the recognition performance for the current recognition models. Consequently, an effective face hallucination framework is desirable.

Typically, for the high upscaling factors, the textural detail in the reconstructed SR images is absent. The restricted use of mean squared error (MSE) between the generated HR image and the ground truth as the only optimization target could be the reason for the missing detailed information in the reconstructed SR images. More specifically, the MSE lacks the ability to capture perceptually relevant dif-

978-1-7281-1522-1/19/$31.00

ferences, e.g., textural detail, as it is defined at image pixel level [40, 41].

Majority of existing face hallucination techniques [39] have been focused on hallucinating faces which are visually pleasant. In other words, they just generate HR details neglecting whether the added details are useful for face recognition. Such reconstructed faces, usually do not improve the face recognition/verification performance. On the contrary, incorporating the identity in face hallucination process enable the framework to preserve the facial details which play a crucial role in face recognition and serve the purpose much better. Therefore, for many real-world applications, preserving identity in face reconstruction is a vital step of hallucination process [14, 42].

In this work, first, we propose a multi-scale generative adversarial network architecture, for face hallucination with a high up-scaling ratio factor. The generator network has multiple outputs, that share most of their parameters, in a progressive structure. As shown in Fig. 1, the input to the network is a low-resolution face image, and multiple face images with different scaling factors are generated through different branches of the network. The deepest output of our generator has resolution equal to our high-resolution face image. The intermediate outputs, have a growing goal of synthesizing small to large images.

Second, we incorporate a face verifier with the original GAN discriminator and propose a novel discriminator which learns to discriminate betwen different identities while distinguishing fake generated HR face images from their corresponding ground truths. Correspondingly, the generator is trained to not only generate face images of high visual quality but also preserve the discriminative features in the hallucination process. Intuitively, improving the discriminator enhances the verification ability by infusing missing details to the LR image, and improving the verification performance boosts the discriminator (which trains the generator) to look for the quality of identity discriminative features in the generated images. In summary, our framework has three major contributions:

- We propose a novel identity-aware GAN for face super resolution which enable us to hallucinate photo-realistic HR faces while preserving the face identity.
- We combine the disciminator and face verifier by proposing a single network which performs both tasks simultaneously.
- Our discriminator jointly learns to distinguish face images at multiple scales. This unified multi-scale structure enables the discriminator to transfer information between generated face images of different scales.
- A series of qualitative and quantitative experiments proves the effectiveness of the proposed end-to-end framework.

1.1. Related Work

Prediction-based methods were among the early methods for single image super-resolution. However, these filtering approaches, oversimplify this problem and usually yield outputs with low details and blurry textures. Some other frameworks have been proposed in [1, 25] that focus particularly on edge-preservation. More effective approaches, which are usually data-driven, learn a complex mapping between low- to high-resolution images. Early approaches to the SR problem were developed based on compressed sensing [45, 47, 9]. Huang et al. [16] exploit self-similarity, where self dictionaries are extended for small transformations and shape variations. A method based on convolutional sparse coding is proposed in [12] to improve output consistency by processing the whole image at once instead of overlapping patches.

Deep learning-based approaches outperformed most of the traditional methods in computer vision [37, 33, 51, 35, 31, 36, 34, 32], more specifically face hallucination schemes [39, 42, 38, 46, 50]. In [42], a deep joint face hallucination and recognition scheme is proposed, which comprises two separate networks, namely SR and face recognition networks. They have jointly optimized the two networks iteratively, however, due to employing a relatively shallow CNN, it resulted in unsatisfactory visual quality in face reconstruction. A much deeper CNN is utilized in [50] to generate HR face image of higher visual quality. To this end, they trained a cascaded bi-network progressively to learn a dense correspondence during the training phase. Song et al. [30] proposed a two-stage face hallucination process that first reconstructs facial parts employing a deep CNN, and then refines the reconstructed faces using a fine-grained facial structure learner.

Recently, generative adversarial network (GAN) has been successfully adopted by many computer vision applications such as image synthesis, image SR, and in-painting [11, 21, 20]. The SR-GAN [24] is the pioneer in utilizing GAN in inferring photo-realistic high-resolution natural images from LR images. They incorporated the perceptual loss in addition to the adversarial loss to push the solution toward learning to preserve the content of images in the super-resolving process. However, Yu et al. [46] showed that this framework is not effective for super-resolving LR to face images. They introduced a pixel-wise L_2 regularization term to exploit the discriminator network feedback and produce faces with higher similarity to the real ones.

Similarly, in [38], deconvolutional layers are separately applied to super-resolve local and global parts. However, none of the mentioned methods guarantee identity preservation in the reconstruction process. Moreover, they often generate unrealistic low quality face images from the very

low resolution face images, as much of the facial structural information is missing.

An end-to-end GAN-based SR model combined with a face alignment network is proposed in [5] that employs a heatmap loss to integrate facial geometrical information in hallucination process by detecting facial landmarks. The application of deep reinforcement learning in HR face generation has also been investigated in [6]. They proposed to employ a recurrent policy network for individual HR face regions reconstruction based on previous regions reconstructions. Finally, they applied a local enhancement network to improve the facial details. Again, the importance of identity preservation has been neglected in these works.

2. Preliminaries

In this section, we provide some rudiments of GANs, necessary to understand the proposed preference-based image generation framework.

2.1. Generative Adversarial Networks (GANs)

GANs [11] are a type of generative models which learn the statistical distribution of the training data, allowing us to synthesize data samples by mapping a random noise z to an output image y: $G(z) : z \longrightarrow y$, where G is the generator network. GAN in its conditional setting (cGAN) is proposed in [18] which learns a mapping from an input x and a random noise z to the output image y: $G(x, z) : \{x, z\} \longrightarrow y$, using an autoencoder network. The generator model $G(x, z)$, is trained to generate images which are not distinguishable from the *real* samples by a discriminator network, D. Simultaneously, the discriminator is learning, adversarially, to discriminate between the *fake* generated images by the generator and the real samples from the training dataset. The objective function of GAN is given by:

$$l_{GAN}(G, D) = \mathbf{E}_{x,y\sim p_{data}}[\log D(x, y)] + \mathbf{E}_{x,z\sim p_z}[\log(1 - D(x, G(x, z)))], \quad (1)$$

where G attempts to minimize it and D tries to maximize it. Since the adversarial loss is not enough to guarantee that the trained network generates the desired output, one may add an extra Euclidean distance term to the objective function to generate images which are near the ground truth. Consequently, the final objective is defined as follows:

$$G^* = \arg\min_G \max_D l_{GAN}(G, D) + \lambda l_{L1}(G), \quad (2)$$

where $l_{L1}(G) = \| y - G(x, z) \|_1$ and λ is weighting factor.

3. Proposed Multi-Scale GAN Architecture

The goal of this work is to learn a generating function G to reconstruct a HR face images from a given LR input face image. The backbone of our deep generator network G, which is demonstrated in Figure 1, is a series of residual blocks with an identical layout. The residual blocks comprise of two 3×3 convolutional layers followed by batch-normalization layers [17] and ParametricReLU activation function [13]. We employ three pre-trained sub-pixel convolution layers [28] to gradually increase the resolution of the input image.

The preliminary results showed that the network is not able to generate high-quality images for upscaling factor of greater than 4x. Consequently, we propose to progressively learn a series of multi-scale images. As shown in Figure 1, our generator has multiple outputs at different resolutions. Each output of the generator learns the face image distribution at that scale. We also concatenate the images at different depths of the discriminator. This multi-scale structure improves the discriminator by jointly learning to distinguish face images at multiple scales. This enable us to transfer information between images of different scales.

The architecture of our discriminator is shown in Figure 1. Following the previous works in [27, 24], we use LeakyReLU activation ($\alpha = 0.2$) and avoid max-pooling in the architecture. However, our discriminator takes images of different resolutions as its inputs. To this end, we utilize strided convolutions ($s = 2$) in the main branch of the discriminator that reduce the spatial size of the feature maps by a factor of two. Simultaneously, images of lower resolutions are processed by convolutional layers ($s = 1$) which extract feature maps of the same size. We then concatenate the extracted feature maps of the lower resolution images with the feature maps of the same spatial size in the main branch.

4. Training Loss Function

To train the proposed network, we utilize multiple loss terms, including an adversarial face verification loss, perceptual loss, and color-consistency regularization.

4.1. Perceptual Loss

The Pixel-wise MSE loss is one of the most widely used loss terms for image super-resolution problems. However, despite the high PSNR, the learned solutions by MSE optimization often lack the high-frequency information, which results in unsatisfactory images of excessively smooth textures. Consequently, similar to [24], rather than relying on the pixel-level losses, we use MSE on the high level extracted features via a pre-trained VGG19 [29], denoted by Φ, which represents perceptual similarity between the generated HR image and its corresponding ground truth. Let $\Phi_j(x)$ denote the feature maps of the j^{th} layer of the loss network for the input image x. The perceptual loss, which has been introduced in [10], is defined as the Euclidean

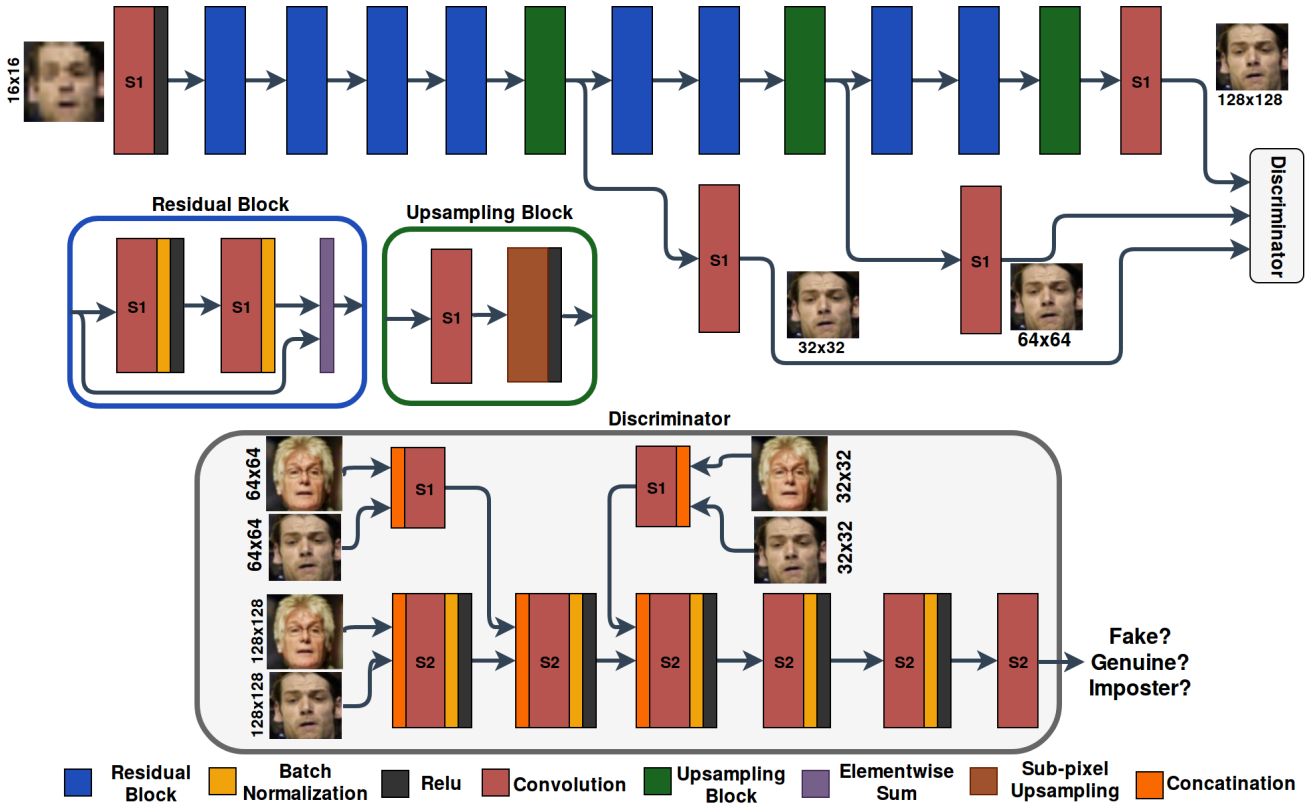

Figure 1: Architecture of the proposed network. Different branches of the generator produce images of different scales. The discriminator, then, learns to distinguish fake and real images, jointly in multiple scales.

distance between the feature representations of a super-resolved image $G(I_{LR})$ and the reference image I_{HR}:

$$l_p^j(G(I_{LR}), I_{HR}) = \frac{1}{N_j} \parallel \Phi_j(G(I_{LR})) - \Phi_j(I_{HR}) \parallel_2^2,$$

where N_j is the number of perceptrons in the jth layer.

4.2. Adversarial Face Verification Loss

In addition to the perceptual loss described above, one may want to add the adversarial loss of GAN, described by Equation 1, to train the generator network. This loss encourages the generator to favor, by seeking to fool the discriminator, production of images which reside on the manifold of natural face images.

However, in this work, we introduce a completely different loss function, namely adversarial face verification loss (AFVL). Our proposed AFVL follows two different goals simultaneously. First, it should help the generator to learn high-frequency information, which cannot be learned by the sole adoption of MSE or perceptual loss. In this way, the generator would be able to produce highly realistic HR images. However, forcing the generator to just care about visual realism may come with the cost of losing identity information. Moreover, incorporating identity information in the discriminator decisions results in training a generator which preserves the critical facial features, which matters in person identification, in super resolving process.

Our discriminator, instead of assigning a *fake* or *real* label to the patches of its HR input image, performs a three ways classification task. More specifically, it classifies the HR images into *fake*, *genuine*, and *imposter*. Particularly, the discriminator takes a pair of images instead of a single image. To train the discriminator, we define four different pairs as follows:

$$\begin{aligned} p_1 &= (I_{HR}^{i1}, G(I_{LR}^{i1})) \qquad p_2 = (I_{HR}^{i}, G(I_{LR}^{j})) \\ p_3 &= (I_{HR}^{i1}, I_{HR}^{i2}) \qquad p_4 = (I_{HR}^{i}, I_{HR}^{j}), \end{aligned} \tag{3}$$

where I_{HR}^{i} represents a high-resolution face image of the i^{th} identity in our dataset, and the number next to the i shows if the same image is used in both side of the pair or not. More specifically, the first pair, p_1, includes a HR face image of the i^{th} identity and the super-resolved image from the down-sampled version of the exact same image. On the contrary, the second pair p_2 includes a HR face image of the i^{th} identity and the super-resolved image from the down-sampled version of an image of another identity.

The discriminator is supposed to classify these two pairs as a *fake* samples. These pairs aim to help the discriminator to learn the visual realism of the generated images.

The third pair, p_3, comprises two different HR face images of the same identity. Note that none of the images are generated by the generator network. The discriminator is trained to classify this pair as *genuine*. On the contrary, the last pair p_4 does not share the identity between the HR images. Clearly, it is desired that the discriminator classify p_4 as *imposter*. Hence, the pairs p_3 and p_4 jointly enable the discriminator to capture critical features which are highly influential in face verification task.

To train the discriminator using these four pairs, we can define the adversarial face verification loss as:

$$l^d_{AFVL}(G,D) = \mathbf{E}_{(x,y)\sim p_1}[\log d_f(x,y)] + \mathbf{E}_{(x,y)\sim p_2}[\log d_f(x,y)] + \mathbf{E}_{(x,y)\sim p_3}[\log d_{gen}(x,y)] + \mathbf{E}_{(x,y)\sim p_4}[\log d_{imp}(x,y)], \tag{4}$$

where d_f, d_{gen}, and d_{imp} are the outputs of discriminator for *fake*, *genuine*, and *imposter* classes, respectively. Similar to the original GAN, the discriminator is trained to maximize this objective function. However, to train the generator, we only use the first two pairs p_1 and p_2. Similar to the training process of the discriminator, here, we train a generator which consider the identity-preservation in its face hallucination process. To this end, the generator is trained to maximize d_{gen} for p_1, and d_{imp} for p_2. In this way, the generator not only tries to fool the discriminator in terms of visual realism of the generated images, by minimizing d_f, but also takes into account the identity of the super-resolved image, through maximization of the verification objective function. In short, the generator maximize the following objective function:

$$l^g_{AFVL}(G,D) = \mathbf{E}_{(x,y)\sim p_1}[\log d_{gen}(x,y)] + \mathbf{E}_{(x,y)\sim p_2}[\log d_{imp}(x,y)]. \tag{5}$$

4.3. Color-consistency regularization

As we go deeper into our generator, the resolution of the generated image is also gradually increased. Since all the generated images belong to the same input but at different scales, they require to have similar structures and colors. To this end, we utilize color-consistency regularization term as an additional objective function to keep the generated samples of different scales from the same input to be more consistent in color. This can improve the quality of the generated images.

Let $\mu = \sum_k x_k/N$, and $\Sigma = \sum_k (x_k - \mu)(x_k - mu)^T/N$ represent the mean and covariance of pixels of the given image, respectively, where $x_k = (R,G,B)^T$ is a pixel in the generated image. Then, the color-consistency regularization term tries to minimize the differences of μ and Σ between the different generated images at various resolutions, which inspires the consistency:

$$l_{C_i} = \frac{1}{n}\sum_{j=1}^{n}\Big(\lambda_1 \parallel \mu_{s_i^j} - \mu_{s_{i-1}^j} \parallel_2^2 + \lambda_2 \parallel \Sigma_{s_i^j} - \Sigma_{s_{i-1}^j} \parallel_F^2 \Big), \tag{6}$$

where $\mu_{s_i^j}$ and $\Sigma_{s_i^j}$ represents the mean and covariance matrix of the j^{th} sample generated in i^{th} scale, and, n is the batch size. Note that images of different resolution are generated by different branches of the generator. In our work, since the generator produces images at three scales, we have two color-consistency regularization terms corresponding to $i = 1, 2$, where each i belongs to the image of size 2^{i+6} pixels.

4.4. Total Loss

While the discriminator is trained using only the AFVL loss l^d_{AFVL}, the total loss to train the generator is defined as:

$$l^g_t = \sum_j l^j_p + \lambda_c \sum_i l_{C_i} - \lambda_a l^g_{AFVL}(G,D). \tag{7}$$

Note that we minimize the perceptual loss in more than one layer to enforce fine and coarse perceptual similarity.

5. Experiments

Our experiments aim to show that our framework can generate high visual quality HR faces at different up-scaling factors while preserving the identity of the hallucinated faces. We compare our method against baselines both qualitatively and quantitatively.

5.1. Datasets

Our experiments are evaluated on the Labeled Faces in the Wild (LFW funneled) [23] and the BioID [19] datasets. The LFW dataset contains 13,233 face images which are collected from the web. Images in this dataset cover a vast variety of pose variations and facial expressions. This dataset comprises of four different parts, including the original set and three different aligned images. In this work, we only use the original ones to conduct our experiments. To generate the LR and HR pairs, we use the original aligned images of size 250×250 pixels, and extract the centric 128×128 image patches as the HR images. Then, we create the corresponding LR images by down-sampling the HR ones using a bilinear kernel with the down-sampling factor. Our training set includes 9,526 images which leaves us the remaining 3,707 images for testing.

The BioID dataset consists of 1,521 face images. We use 1,028 images for training and the remaining 493 images for

testing. This follows the same split provided by the LFW dataset. The images are aligned with SDM method [43] and then a patch with the size of 160×120 is cropped from the center of each image.

5.2. Visual Realism Evaluation

For visual realism evaluation, we evaluate our proposed SR network on two scaling factors of $4x$ and $8x$. Input low-resolution image is generated by resizing the original images with the scaling factors. Hence, to generate LR images, the HR images of BioID are resized to 40×30 and 20×15, and the HR images are resized to 32×32 and 16×16, respectively.

For the evaluation metrics, we adopt the widely used Peak Signal-to-Noise Ratio (PSNR), structural similarity (SSIM) as well as feature similarity (FSIM) [48]. We perform a comparison between our method and several state-of-the-art face hallucination and image super-resolution techniques. Particularly, we compare with the BCCNN [49], SFH [44], GLN [38], MZQ [26] face hallucination approaches and three general image super-resolution methods: A-FH [6], SRCNN [8], and VDSR [22].

Table 1 compares the performance of our method with other state-of-the-art techniques. Our proposed framework significantly outperforms all the other methods in terms of PSNR, SSIM and FSIM metrics on LFW and BioID datasets. Since the traditional face hallucination methods, i.e., SFH and MZQ, are highly dependent on the facial landmarks detection performance, and the landmarks detection is not quite reliable in very low-resolution images, their performances on 8x up-scaling factor are too low compared to the other methods. Among deep-learning based methods, our work outperforms the current state-of-the-art image super-resolution method (A-FH) on different experiments.

In addition, our method significantly outperforms state-of-the-art face hallucination methods, namely SiGAN and GLN. Figures 2 and 3 illustrates the qualitative comparisons of face hallucination results on the LFW dataset for 4x and 8x up-scaling factors, respectively. Our proposed framework generates face images that are more clear and sharper compared to the A-FH, GLN, and VDSR.

5.3. Identity Preserving Evaluation

To evaluate the performance of different methods in preserving the identity of the LR face in the hallucination process, we compare our method with several state-of-the-art face hallucination methods including DFCG [30], SiGAN [15], UR-DGN [46], GLN [38], A-FH [6], DCGAN [27], PRSR [7], and [24]. Note that we selected the methods whose performances are reported in the literature. In addition, we evaluated the performances of the top two methods with the highest visual realism scores in the evaluation section, namely A-FH and GLN.

5.3.1 Face Verification Performance

To compare the performance of the proposed method with the previous face hallucination techniques for face verification task, we employ a state-of-the-art CNN-based face recognition engine, the OpenFaces [2]. We report the face recognition rate and verification rate of the hallucinated faces by different methods. The accuracy of the hallucinated HR faces is evaluated, following the standard face verification methodology described in [2]. The accuracy is calculated which is based on if the OpenFaces verifies the hallucinated HR faces as the same identity as their corresponding ground-truth or not.

We setup our experiment by first randomly sampling 200,000 face pairs from LFW, similar to the training set of the OpenFaces recognition engine as described in [2]. Then, 6,000 faces are randomly sampled from the remaining face images of LFW for the face verification performance evaluation. Our evaluation metric is the area under curve (AUC) [23] of the trained face verification system based on the super-resolved HR faces.

Table 2 reports the AUCs for the generated HR faces of 128×128 and 64×64 pixels from 16×16 LR faces using different face hallucination techniques. The results show that the AUC for the generated HR faces by our method is significantly higher than the AUCs of the other methods. This proves the superiority of our method in preserving the identity of faces in the hallucination process

5.3.2 Face Recognition Performance

We also evaluate the performance of our framework for the face recognition task. We setup this experiment as suggested in [2]. For this experiment, the training set includes 11,000 face image of 680 different identities randomly sampled from the LFW dataset. The remaining 2,000 face images form our test set. The OpenFaces is trained on face images after resizing to 96×96. Likewise, all the hallucinated HR faces need to be resized to 96×96 at the test time to evaluate the performance.

Table 2 compares the top-1, top-5, and top-10 face recognition rates of different methods for 64×64, i.e., 4x upscaling factor, and 128×128, 8x up-scaling factor, HR faces upscaled from 16×16 LR faces. Note that for some methods the results for 8x upscaling factors are not reported in the original papers. The result prove the superiority of our method, in terms of the average recognition rates for the hallucinated HR faces, compared to the previous state-of-the-art methods. Note that, GLN and SR-GAN have worse performance on 8x compared to the 4x upscaling factor due to the generated artifacts by these methods at 8x super-resolved face images.

Compared to the bicubic interpolation, DFCG and DCGAN have lower face recognition rates, which shows the

Figure 2: Qualitative results on LFW-funneled with scaling factor of 4.

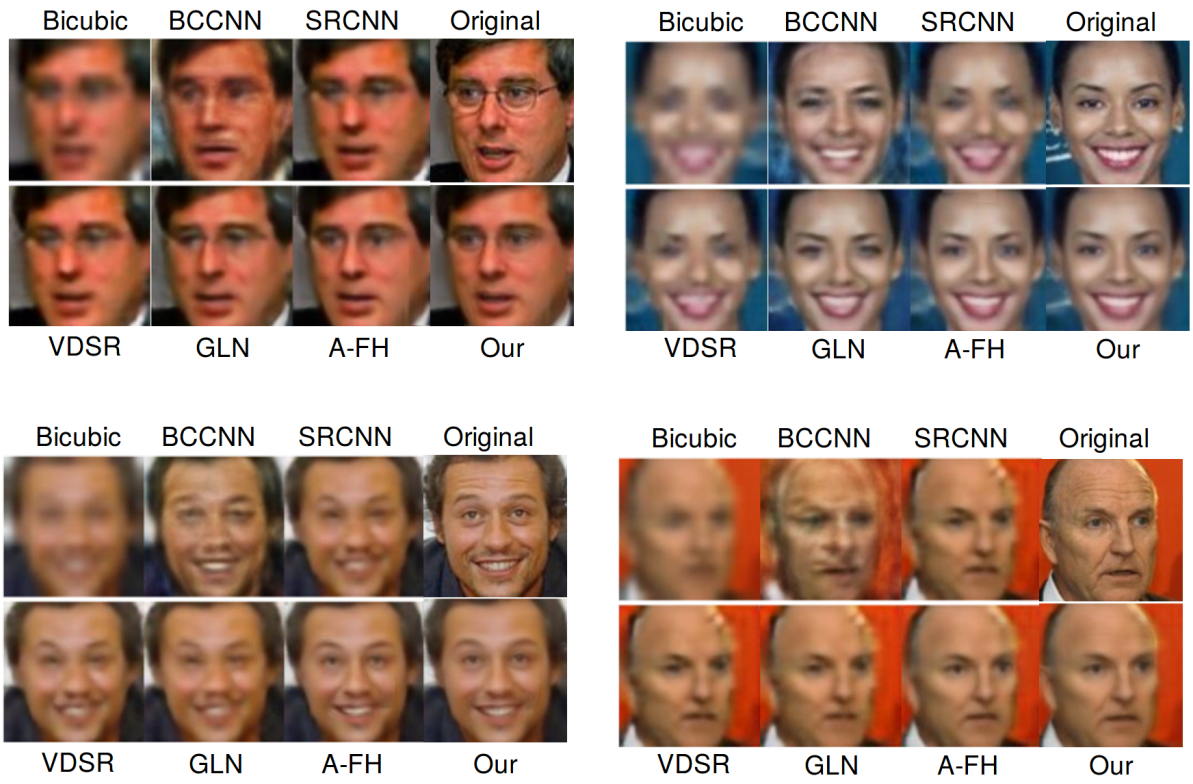

Figure 3: Qualitative results on LFW-funneled with scaling factor of 8.

Table 1: Comparison between our method and others in terms of PSNR, SSIM and FSIM evaluation metrics.

Method	LFW-funneled 4x			LFW-funneled 8x			BioID 4x			BioID 8x		
	PSNR	SSIM	FSIM	PSNR	SSIM	FSIM	PSNR	SSIM	FSIM	PSNR	SSIM	FSIM
Bicubic	26.79	0.8469	0.8947	21.92	0.6712	0.7824	25.18	0.8170	0.8608	20.68	0.6434	0.7539
SFH [44]	26.59	0.8332	0.8917	22.12	0.6732	0.7832	25.41	0.8034	0.8494	20.31	0.6234	0.7238
BCCNN [49]	26.60	0.8329	0.8982	22.62	0.6801	0.7903	24.77	0.8034	0.8421	21.40	0.6504	0.7621
MZQ [26]	25.93	0.8313	0.8865	22.12	0.6771	0.7802	24.66	0.8001	0.8573	21.11	0.6481	0.7594
SRCNN [8]	28.94	0.6363	0.9069	23.92	0.6927	0.8314	27.02	0.8517	0.8771	22.34	0.6980	0.8274
VDSR [22]	29.25	0.8711	0.9123	24.12	0.7031	0.8391	28.52	0.8627	0.8914	24.31	0.7321	0.8465
GLN [38]	30.34	0.8922	0.9151	24.51	0.7109	0.8405	29.13	0.8794	0.8966	24.76	0.7421	0.8525
A-FH [6]	32.93	0.9104	0.9427	26.17	0.7604	0.8630	31.56	0.9002	0.9343	26.56	0.7864	0.8747
Our	**33.59**	**0.9213**	**0.9601**	**26.94**	**0.7723**	**0.8772**	**32.49**	**0.9899**	**0.9481**	**27.83**	**0.7967**	**0.8914**

Table 2: Comparison of LFW face recognition rates for the hallucinated HR faces using different techniques

	AUC		Verification Acc. 4x			Verification Acc. 8x		
Method	4x	8x	Top-1	Top-5	Top-10	Top-1	Top-5	Top-10
HR	98.8%	99.1%	36.8%	55.9%	63.8%	37.5%	57.0%	66.2%
Bicubic	75.7%	76.0%	11.6%	27.5%	37.6%	11.7%	27.1%	36.4%
DFCG [30]	73.9%	-	9.6%	23.7%	34.8%	-	-	-
UR-DGN [46]	72.8%	-	12.2%	29.0%	38.7%	-	-	-
DCGAN [27]	74.8%	-	9.3%	24.9%	33.9%	-	-	-
PRSR [7]	76.9%	-	13.3%	29.7%	40.1%	-	-	-
SiGAN [15]	83.4%	-	17.9%	32.9%	48.1%	-	-	-
GLN [38]	80.6%	79.2%	17.5%	30.6%	45.7%	17.1%	29.9%	44.3%
SR-GAN [24]	81.8%	71.8%	18.3%	31.4%	47.6%	15.8%	26.7%	39.6%
A-FH [6]	85.2%	85.6%	18.8%	31.9%	48.4%	18.9%	32.4%	48.9%
Our	**86.0**%	**86.5**%	**19.5**%	**33.2**%	**49.5**%	**20.1**%	**33.5**%	**50.1**%

importance of reconstructing facial features that are critical in face re-identification. In other word, despite the higher level of HR details in these methods, compared to the bicubic, they are not useful for identity recognition.

6. Conclusion

In this paper, we have proposed a identity-preserving face hallucination GAN-based framework. We enabled our generator to up-scale LR face images by a factor of 8 and learn to jointly generate face images of progressive resolution. We have also proposed a new discriminator which can jointly learn to verify the identity of the generated images and check their visual quality. The new discriminator architecture enables the whole face hallucination process to be identity-preserving too. Experimental results on several LR version of face benchmarks have convincingly demonstrated the effectiveness of the proposed approach.

References

[1] J. Allebach and P. W. Wong. Edge-directed interpolation. In *Proceedings of 3rd IEEE International Conference on Image Processing*, volume 3, pages 707–710. IEEE, 1996.

[2] B. Amos, B. Ludwiczuk, M. Satyanarayanan, et al. Openface: A general-purpose face recognition library with mobile applications. *CMU School of Computer Science*, 6, 2016.

[3] S. Baker and T. Kanade. Hallucinating faces.

[4] S. Baker and T. Kanade. Limits on super-resolution and how to break them. In *Proceedings IEEE Conference on Com-*

puter Vision and Pattern Recognition. CVPR 2000 (Cat. No. PR00662), volume 2, pages 372–379. IEEE, 2000.

[5] 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 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 109–117, 2018.

[6] Q. Cao, L. Lin, Y. Shi, X. Liang, and G. Li. Attention-aware face hallucination via deep reinforcement learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 690–698, 2017.

[7] R. Dahl, M. Norouzi, and J. Shlens. Pixel recursive super resolution. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 5439–5448, 2017.

[8] C. Dong, C. C. Loy, K. He, and X. Tang. Learning a deep convolutional network for image super-resolution. In *European conference on computer vision*, pages 184–199. Springer, 2014.

[9] W. Dong, L. Zhang, G. Shi, and X. Wu. Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization. *IEEE Transactions on Image Processing*, 20(7):1838–1857, 2011.

[10] L. Gatys, A. S. Ecker, and M. Bethge. Texture synthesis using convolutional neural networks. In *Advances in neural information processing systems*, pages 262–270, 2015.

[11] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[12] S. Gu, W. Zuo, Q. Xie, D. Meng, X. Feng, and L. Zhang. Convolutional sparse coding for image super-resolution. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1823–1831, 2015.

[13] K. He, X. Zhang, S. Ren, and J. Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015.

[14] P. H. Hennings-Yeomans, S. Baker, and B. V. Kumar. Simultaneous super-resolution and feature extraction for recognition of low-resolution faces. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2008.

[15] C.-C. Hsu, C.-W. Lin, W.-T. Su, and G. Cheung. Sigan: Siamese generative adversarial network for identity-preserving face hallucination. *arXiv preprint arXiv:1807.08370*, 2018.

[16] J.-B. Huang, A. Singh, and N. Ahuja. Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5197–5206, 2015.

[17] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.

[18] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.

[19] O. Jesorsky, K. J. Kirchberg, and R. W. Frischholz. Robust face detection using the hausdorff distance. In *International conference on audio-and video-based biometric person authentication*, pages 90–95. Springer, 2001.

[20] H. Kazemi, S. M. Iranmanesh, and N. Nasrabadi. Style and content disentanglement in generative adversarial networks. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 848–856. IEEE, 2019.

[21] H. Kazemi, S. Soleymani, F. Taherkhani, S. Iranmanesh, and N. Nasrabadi. Unsupervised image-to-image translation using domain-specific variational information bound. In *Advances in Neural Information Processing Systems*, pages 10348–10358, 2018.

[22] J. Kim, J. Kwon Lee, and K. Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1646–1654, 2016.

[23] 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*, pages 189–248. Springer, 2016.

[24] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017.

[25] X. Li and M. T. Orchard. New edge-directed interpolation. *IEEE transactions on image processing*, 10(10):1521–1527, 2001.

[26] X. Ma, J. Zhang, and C. Qi. Hallucinating face by position-patch. *Pattern Recognition*, 43(6):2224–2236, 2010.

[27] A. Radford, L. Metz, and S. Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

[28] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang. Real-time single image and video super-resolution using an efficient subpixel convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1874–1883, 2016.

[29] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[30] Y. Song, J. Zhang, S. He, L. Bao, and Q. Yang. Learning to hallucinate face images via component generation and enhancement. *arXiv preprint arXiv:1708.00223*, 2017.

[31] F. Taherkhani and M. Jamzad. Restoring highly corrupted images by impulse noise using radial basis functions interpolation. *IET Image Processing*, 12(1):20–30, 2017.

[32] F. Taherkhani, H. Kazemi, and N. M. Nasrabadi. Matrix completion for graph-based deep semi-supervised learning. In *Thirty-Third AAAI Conference on Artificial Intelligence*, 2019.

[33] F. Taherkhani, N. M. Nasrabadi, and J. Dawson. A deep face identification network enhanced by facial attributes prediction. In *Proceedings of the IEEE Conference on Computer*

Vision and Pattern Recognition Workshops, pages 553–560, 2018.

[34] F. Taherkhani, V. Talreja, H. Kazemi, and N. Nasrabadi. Facial attribute guided deep cross-modal hashing for face image retrieval. In *2018 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–6. IEEE, 2018.

[35] V. Talreja, T. Ferrett, M. C. Valenti, and A. Ross. Biometrics-as-a-service: A framework to promote innovative biometric recognition in the cloud. In *2018 IEEE International Conference on Consumer Electronics (ICCE)*, pages 1–6. IEEE, 2018.

[36] V. Talreja, F. Taherkhani, M. C. Valenti, and N. M. Nasrabadi. Using deep cross modal hashing and error correcting codes for improving the efficiency of attribute guided facial image retrieval. In *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 564–568. IEEE, 2018.

[37] V. Talreja, M. C. Valenti, and N. M. Nasrabadi. Multibiometric secure system based on deep learning. In *2017 IEEE Global conference on signal and information processing (globalSIP)*, pages 298–302. IEEE, 2017.

[38] O. Tuzel, Y. Taguchi, and J. R. Hershey. Global-local face upsampling network. *arXiv preprint arXiv:1603.07235*, 2016.

[39] N. Wang, D. Tao, X. Gao, X. Li, and J. Li. A comprehensive survey to face hallucination. *International journal of computer vision*, 106(1):9–30, 2014.

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

[41] Z. Wang, E. P. Simoncelli, and A. C. Bovik. Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, volume 2, pages 1398–1402. Ieee, 2003.

[42] J. Wu, S. Ding, W. Xu, and H. Chao. Deep joint face hallucination and recognition. *arXiv preprint arXiv:1611.08091*, 2016.

[43] X. Xiong and F. De la Torre. Supervised descent method and its applications to face alignment. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 532–539, 2013.

[44] C.-Y. Yang, S. Liu, and M.-H. Yang. Structured face hallucination. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1099–1106, 2013.

[45] J. Yang, J. Wright, T. Huang, and Y. Ma. Image super-resolution as sparse representation of raw image patches. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. Citeseer, 2008.

[46] X. Yu and F. Porikli. Ultra-resolving face images by discriminative generative networks. In *European conference on computer vision*, pages 318–333. Springer, 2016.

[47] R. Zeyde, M. Elad, and M. Protter. On single image scale-up using sparse-representations. In *International conference on curves and surfaces*, pages 711–730. Springer, 2010.

[48] L. Zhang, L. Zhang, X. Mou, and D. Zhang. FSIM: A feature similarity index for image quality assessment. *IEEE transactions on Image Processing*, 20(8):2378–2386, 2011.

[49] E. Zhou, H. Fan, Z. Cao, Y. Jiang, and Q. Yin. Learning face hallucination in the wild. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.

[50] S. Zhu, S. Liu, C. C. Loy, and X. Tang. Deep cascaded bi-network for face hallucination. In *European conference on computer vision*, pages 614–630. Springer, 2016.

[51] F. Zohrizadeh, M. Kheirandishfard, and F. Kamangar. Class subset selection for partial domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019.

[52] W. W. Zou and P. C. Yuen. Very low resolution face recognition problem. *IEEE Transactions on image processing*, 21(1):327–340, 2012.