

Received June 24, 2020, accepted July 10, 2020, date of publication July 24, 2020, date of current version August 7, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3011699

# ID Preserving Face Super-Resolution Generative Adversarial Networks

JINNING LI<sup>1</sup>, YICHEN ZHOU<sup>2</sup>, JIE DING<sup>3</sup>, (Member, IEEE),  
CEN CHEN<sup>4</sup>, AND XULEI YANG<sup>4</sup>, (Senior Member, IEEE)

<sup>1</sup>Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

<sup>2</sup>Department of Computer Science, National University of Singapore, Singapore 119077

<sup>3</sup>Department of Electronic Engineering, School of Information Science and Engineering, Fudan University, Shanghai 200433, China

<sup>4</sup>Institute for Infocomm Research (I2R), Agency for Science, Technology and Research (A\*STAR), Singapore 138632

Corresponding authors: Jie Ding (jding002@e.ntu.edu.sg) and Xulei Yang (yangxulei@pmail.ntu.edu.sg)

**ABSTRACT** We propose an ID Preserving Face Super-Resolution Generative Adversarial Networks (IP-FSRGAN) to reconstruct realistic super-resolution face images from low-resolution ones. Inspired by the success of generative adversarial networks (GAN), we introduce a novel ID preserving module to help the generator learn to infer the facial details and synthesize more realistic super-resolution faces. Our method produces satisfactory visual results and also quantitatively outperforms state-of-the-art super-resolution methods on the face datasets including CASIA-Webface, CelebA, and LFW datasets under the metrics of PSNR, SSIM, and cosine similarity. In addition, we propose a framework to apply IP-FSRGAN model to address the face verification task on low-resolution face images. The synthesized  $4\times$  super-resolution faces achieve a verification accuracy of 97.6%, improved from 92.8% of low resolution faces. We also prove by experiments that the proposed IP-FSRGAN model demonstrates excellent robustness under different downsample scaling factors and extensibility to various face verification models.

**INDEX TERMS** ID preserving, face super-resolution, generative adversarial networks, face verification.


## I. INTRODUCTION

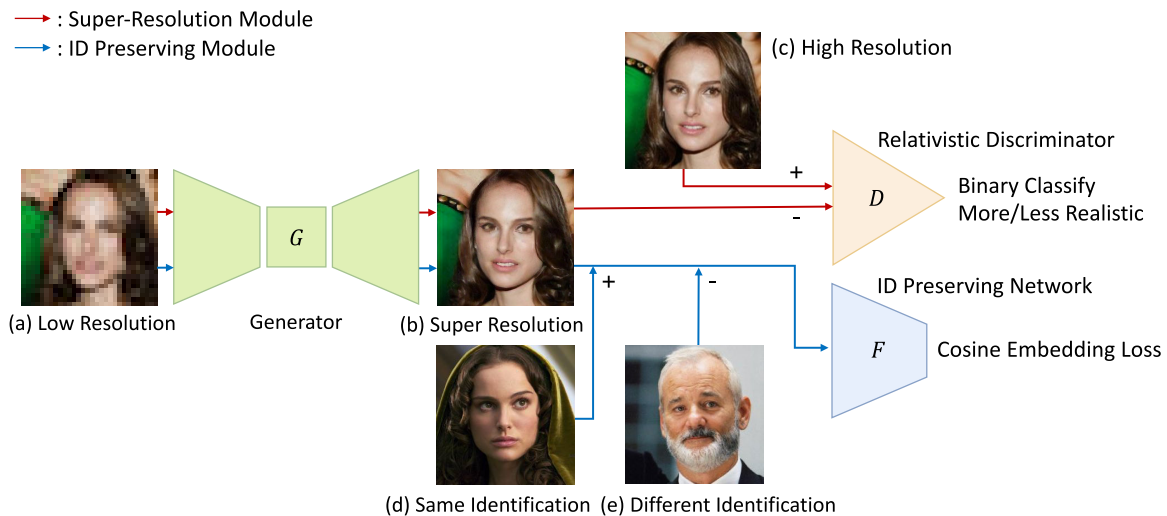
Recent advancements in deep neural networks and generative adversarial networks (GANs) have brought tremendous successes in the super-resolution task, which reconstructs high-resolution images from low-resolution images [1]. The super-resolution task is one of the most challenging tasks due to the information missing of the input low-resolution images. In addition to the general image super-resolution task, face super-resolution is much more challenging because the tolerance of distortion for face images is much lower.

The existing techniques, especially those based on GANs [2], [3], could already produce both visually and quantitatively plausible results. These super-resolution GANs are proved to be effective in the reconstruction of low-resolution images about general objects such as natural scenery and city landscape. However, their performance is still *limited on low-resolution human faces*. These super-resolution GANs methods often produce distortion and unrealistic patterns. Admittedly, slight distortion is *acceptable for landscapes synthesis*, such as trees, buildings, etc. This kind of distortion could even sometimes gives the image a sense of art and

thus brings additional benefits and applications, such as the style transfer task [4]. However, even slight distortion on the *human faces* will make them unrecognizable and sometimes horrible. One of the major reasons is that the existing super-resolution GANs don't learn to preserve the identity of the person. Instead, they only learn to build higher resolution without properly inferring and synthesize the facial details of the person, especially eyes and mouths.

In this paper, we propose an ID Preserving Face Super-Resolution Generative Adversarial Networks (IP-FSRGAN) to address the distortion problem and reconstruct realistic high-resolution faces. We embed a pre-trained face recognition network to the framework of conditional GANs [5] and design an ID preserving loss to help the generator learn to preserve the identity of the face during the reconstruction, as shown in Fig. 1. While training the GANs, the face recognition network extracts the representations of one pair of face images consisting of a synthesized super-resolution image and a real high-resolution one. The ID preserving loss is designed to minimize the distance between the synthesized image and high-resolution image if they belong to the same identity and otherwise maximize the distance. We conduct experiments to evaluate the performance of IP-FSRGAN. When compared with several

The associate editor coordinating the review of this manuscript and approving it for publication was Peter Peer .



**FIGURE 1.** Overview of the ID Preserving Face Super-Resolution GANs (IP-FSRGAN). (a): low-resolution faces as inputs; (b): super-resolution faces generated by IP-FSRGAN; (c): high-resolution faces as ground truth; (d) and (e): faces with the same or different identities. IP-FSRGAN contains a super-resolution module and an ID preserving module. The super-resolution module contains a framework of cGANs reconstructing (b) from (a). The ID preserving module contains an ID preserving network trained to minimize the distance between positive pairs (b, d) while maximizing that of negative pairs (b, e).

state-of-the-art models, the proposed IP-FSRGAN could not only produce visually satisfactory super-resolution faces but also obtain the highest PSNR, SSIM, and cosine similarity on CASIA-Webface [6], CelebA [7], and LFW [8] datasets.

We prove by experiments that the proposed IP-FSRGAN could improve face verification on low-resolution face image. The face verification model receives two face images and judges whether they belong to the same identity. Existing face verification methods could already verify the identities accurately and be widely used in industrial scenarios. However, these face verification methods often fail to verify low-resolution face images. In this paper, we propose a framework to first increase the resolution of the low-resolution face and then use the reconstructed super-resolution face images for face verification. We apply the proposed framework to the SphereFace [9], a face verification model that achieved relatively high accuracy on the LFW dataset. We improve its verification accuracy on low-resolution images from 92.8% to 97.6%. IP-FSRGAN also shows good robustness for different scaling factors and adaptation for various face verification models.

The main contributions of this paper are three folds: Firstly, we propose IP-FSRGAN, the first to integrate id preserving loss to the setting of GANs, for face super-resolution and design the corresponding training strategy. Secondly, we design a framework to successfully apply the trained IP-FSRGAN to improve the performance of face verification task on a public dataset. Lastly, we publicly share<sup>1</sup> the source codes of the implementation of the proposed IP-FSRGAN, interested readers may re-use the source codes for their own face super-resolution tasks. The rest of this paper is organized as follows: Section II reviews related works of

super-resolution, identity preserving, and face verification. Our proposed IP-FSRGAN for face super-resolution and its application for face verification are presented in Section III. Extensive experiments are conducted in Section IV and V to verify the effectiveness of the proposed approaches. Lastly, the conclusion is given in Section VI.

## II. RELATED WORK

### 1) SUPER-RESOLUTION

SRCNN [1] is the pioneer to apply an end-to-end deep convolutional neural network (DCNN) for super-resolution task by minimizing mean square error (MSE) or maximizing PSNR between synthesized super-resolution (SR) image and real high-resolution (HR) image. However, the generated SR images are still blurred and unsatisfactory enough. The subsequent works have tried to improve the DCNN architecture by applying residual blocks [2], Laplacian pyramid structure [10], residual dense network [11], recursive learning [12], [13], and deep back projection [14]. There is another research direction to deal with super-resolution with higher magnification (e.g. 8X) by using cascade strategy [15], [16] or assisted by facial component heatmap and attribute information [17], [18]. In [19], the author proposed a perceptual loss to minimize the perceptual similarity [20] between SR and HR images for image super-resolution, which improve the visual effects of synthesized SR images. SRGAN [2] trains the GANs architecture with both the adversarial loss and perceptual loss, which proves that the framework of GANs could improve the photo-realism of synthesized SR images. In ESRGAN [3], the author improves SRGAN by introducing a novel Residual-in-Residual Dense Block (RRDB) network architecture, removing batch normalization layers [21], and applying the relativistic GAN [22]. ESRGAN has produced realistic SR images on landscape

<sup>1</sup> <https://github.com/jinningli/IP-FSRGAN>

images, however, its application on face super-resolution is still limited.

### A. IDENTITY PRESERVING

The idea of identity preserving has been used for various tasks. There are mainly two kinds of identity preserving methods. The first kind applies a pre-trained neural network to extract the features from two faces and minimize their distance, such as the IP-CGAN [23] for face aging task, TP-GAN [24] for frontal view synthesis, and SICNN for face hallucination [25]. The second kind uses a classifier to classify the synthesized face to the correct category, like the FaceID-GAN [26]. In this paper, we propose that the face super-resolution task could also benefit from the idea of identity preserving. Instead of using L1 or L2 loss between the extracted features, we apply the cosine embedding loss to minimize the distance for faces with the same identity and simultaneously maximize the distance for faces with the different identities.

### B. FACE VERIFICATION

Face verification and recognition algorithms are already widely used in the industry. Current face verification and recognition techniques have achieved a high accuracy close to human-level performance, such as DeepFace [27], FaceNet [28], SphereFace [9] and LightCNN [29]. However, the performance of these models on very low-resolution faces is still limited. In this work, we prove by experiments that the proposed IP-FSRGAN could not only generate realistic HR faces but also improve the performance of face verification and recognition.

### III. THE PROPOSED IP-FSRGAN

GANs based techniques are already proved to be effective on the general super-resolution task [2], [3]. However, the GANs based models usually lead to blurring and distortion due to information missing of the low-resolution faces. This limits its application in some advanced facial tasks such as face recognition. In this work, we propose IP-FSRGAN to address the blurring and distortion problems. The framework of IP-FSRGAN is shown in Fig. 1, which contains a super-resolution module and an ID preserving module. The super-resolution module is a conditional generative adversarial network, which reconstructs super-resolution faces with low-resolution faces while the ID preserving module helps to supervise the generator to produce identity invariant super-resolution faces.

Mathematically, we use  $x$  to denote one low-resolution (LR) face image which is drawn from an underlying space  $\mathcal{X}$ . The ground truth - high-resolution (HR) face image is denoted by  $y$ , which is drawn from an underlying space  $\mathcal{Y}$ . The training data  $d = \{(x_1, y_1), \dots, (x_m, y_m)\}$  is drawn from the joint space  $\mathcal{X} \times \mathcal{Y}$ . The objective of face super-resolution is to find a mapping  $h \in \mathcal{H}$  from  $\mathcal{X}$  to  $\mathcal{Y}$ , which agrees the mappings in training dataset. Let  $\mathcal{L}$  be a loss function between  $h(x)$  and the target  $y$ . The mapping  $h(x)$  could be found by minimizing the expected loss:  $\min_h \mathbb{E}_{(x,y) \sim P_{xy}(x,y)} [\mathcal{L}(h(x), y)]$ .

In the setting of IP-FSRGAN, the loss function contains four elements including adversarial loss  $\mathcal{L}_{adv}$ , ID preserving loss  $\mathcal{L}_{id}$ , perceptual loss  $\mathcal{L}_p$ , and pixel loss  $\mathcal{L}_1$ . The objective of IP-FSRGAN is to find the optimal parameters of generator  $G$  and discriminator  $D$  which satisfy:

$$\begin{aligned} \mathcal{G}_{loss} &= \lambda \mathcal{L}_{adv,G} + \gamma \mathcal{L}_{id} + \eta \mathcal{L}_p + \xi \mathcal{L}_1; \\ \mathcal{D}_{loss} &= \mathcal{L}_{adv,D}, \end{aligned} \quad (1)$$

where  $\lambda$ ,  $\gamma$ ,  $\eta$  and  $\xi$  control to what extent each loss contributes to the final one. In experiments, we empirically determine the optimal  $\lambda$ ,  $\gamma$ ,  $\eta$  and  $\xi$ .

### A. SUPER-RESOLUTION MODULE

The super-resolution module is a generative adversarial network. It contains a generator  $G : \mathcal{X} \rightarrow \mathcal{Y}$  which transforms a LR face  $x_i$  to a fake super-resolution (SR) face  $G(x_i)$  and a discriminator  $D : \mathcal{Y} \rightarrow \{0, 1\}$  which tries to differentiate the synthesized SR faces from HR faces. We adopt the design of relativistic discriminator [22]. Instead of predicting whether a face image is real or fake, the relativistic discriminator tries to figure out whether a face image is relatively more realistic or not. Assume  $\tilde{D}(\cdot)$  represents the output of discriminator network before applying the sigmoid function  $\sigma$ . The relativistic discriminator could be formulated by:

$$D(x_1, x_2) = \sigma(\tilde{D}(x_1) - \mathbb{E}[\tilde{D}(x_2)]). \quad (2)$$

Note that  $D(x_1, x_2) \neq D(x_2, x_1)$ . The goal of the generator  $G$  is to reconstruct fake SR face  $G(x_i)$  so that it is hard for the discriminator to tell apart from the HR face  $y_i$ . The loss function for the generator can be formulated by:

$$\begin{aligned} \mathcal{L}_{adv,G} &= -\mathbb{E}_{(x,y) \sim P_{x,y}(x,y)} [1 - \log(D(y, G(x)))] \\ &\quad - \mathbb{E}_{(x,y) \sim P_{x,y}(x,y)} [\log(D(G(x), y))]. \end{aligned} \quad (3)$$

The discriminator, on the other hand, tries to separate the HR faces from the fake ones. The loss function for the discriminator can be formulated by:

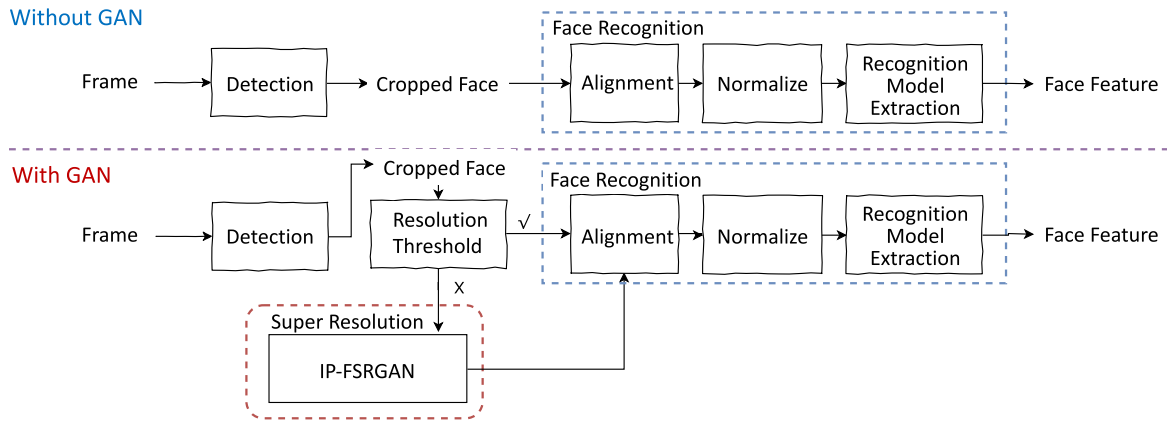
$$\begin{aligned} \mathcal{L}_{adv,D} &= -\mathbb{E}_{(x,y) \sim P_{x,y}(x,y)} [\log(D(y, G(x)))] \\ &\quad - \mathbb{E}_{(x,y) \sim P_{x,y}(x,y)} [1 - \log(D(G(x), y))]. \end{aligned} \quad (4)$$

When the competitive training between the generator and discriminator achieves an equilibrium, the generator will learn to reconstruct relatively realistic SR faces.

### B. IDENTITY PRESERVING MODULE

The idea of identity preserving module is based on an assumption that the identity of the face should remain invariant before and after the super-resolution transformation. This assumption contains two phases. First, the distance between positive samples (faces of the same ID) should be as close as possible. Second, the distance between negative samples (faces of different IDs) should be as far as possible.

To measure the distance, we introduce a pre-trained ID preserving network  $F$  to extract the ID feature of faces and then calculate the cosine similarity to further measure the distance. Assume  $G(x)$  is the fake SR face reconstructed by



**FIGURE 2.** Improve the performance of face verification with the proposed IP-FSRGAN. The general framework of face verification and recognition including face detection, face alignment, face normalization, and feature extraction by recognition model. To improve face verification, we could set a threshold of resolution for the detected faces and apply IP-FSRGAN to reconstruct the faces with low resolution.

the generator and  $y$  is the reference face image. The cosine similarity between their ID feature is  $\cos(F(G(x)), F(y))$ . We are going to maximize the similarity (minimize the distance) for the positive samples and minimize the similarity otherwise.

In our experiments, we adapt LightCNN [29], a pre-trained face recognition network, as the ID preserving network. We use  $I_{i,j}$  to represent the ID function.  $I_{i,j} = 1$  if the faces  $y_i$  and  $y_j$  drawn from  $\mathcal{Y}$  belong to the same identity, otherwise  $I_{i,j} = 0$ . The ID preserving loss can be formulated as:

$$\mathcal{L}_{id} = \mathbb{E}_{i,j}[I_{i,j} \cdot (1 - \cos(F(G(x_i)), F(y_j)))] + \mathbb{E}_{i,j}[(1 - I_{i,j}) \cdot \max(0, \cos(F(G(x_i)), F(y_j)) - m)] \quad (5)$$

where  $m$  is the margin value. By minimize the ID preserving loss,  $G$  will be optimized to keep the identity during transformation. This technique forces  $G$  to infer the details which are strongly related to identity such as the shape of eyes and mouth from the LR face images. In this way, the generator could synthesize much more realistic SR faces.

In addition to ID preserving loss, we also introduce the perceptual loss [19], [30], [31] to stabilize the training of GANs and improve the quality of reconstructed SR images. The perceptual loss is calculated with a pre-trained convolutional neural networks, usually the VGG network. We use  $\phi_j$  to denote the feature map obtained after the  $j$ -th convolutional layer. The perceptual loss between the fake SR face  $G(x)$  and HR face  $y$ :

$$\mathcal{L}_p = \sum_j \mathbb{E}_{(x,y) \sim P_{x,y}(x,y)} \left[ \frac{1}{C_j H_j W_j} \|\phi_j(G(x)) - \phi_j(y)\|_2^2 \right], \quad (6)$$

where  $C_j$ ,  $H_j$ , and  $W_j$  denote the shape of the respective feature maps after the  $j$ -th convolutional layer of VGG network. The reconstruction loss is also used to force the generated SR image to be close to the real HR image, which is defined by

$$\mathcal{L}_1 = \mathbb{E}_{(x,y) \sim P_{x,y}(x,y)} [\|y - G(x)\|_1]. \quad (7)$$

The L1-norm distance rather than L2-norm distance used here encourages less blurring.

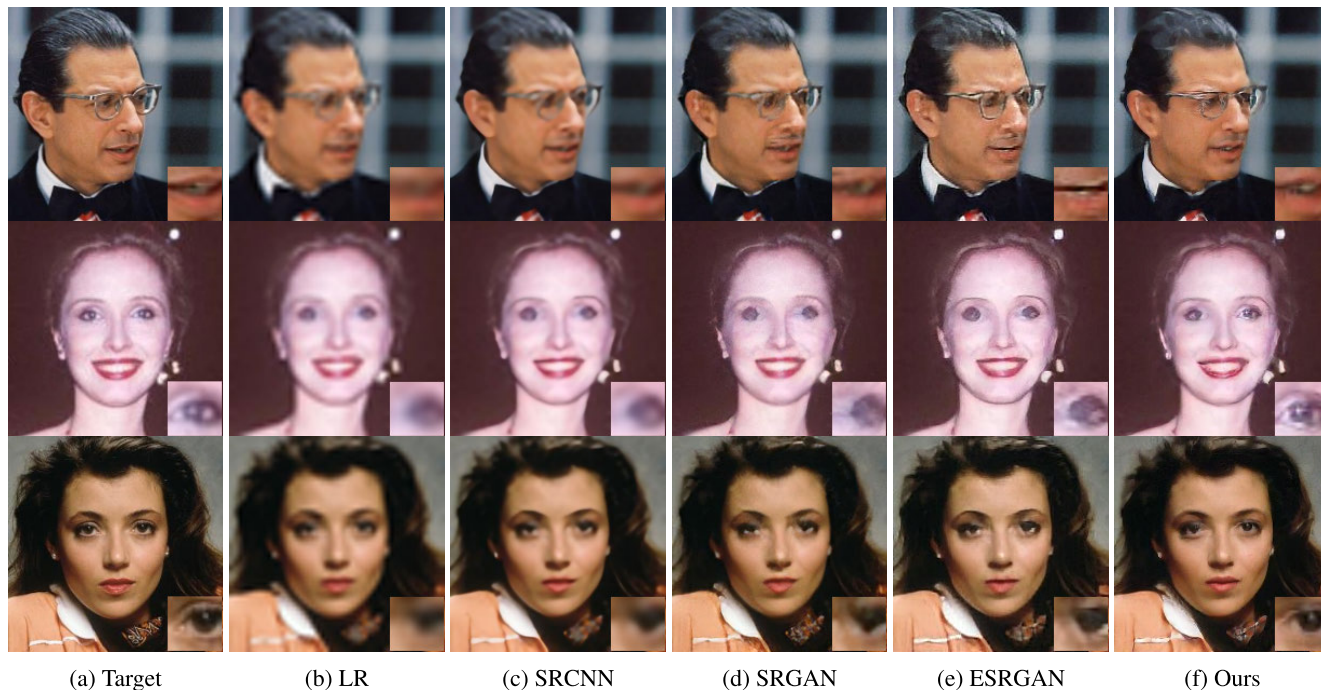
### C. IMPROVING FACE VERIFICATION WITH IP-FSRGAN

Face super-resolution is an important task that could not only reconstruct visually clearer facial images but also be applied to improve the performance of many other computer vision tasks such as face detection [32]. However, it is much more challenging to improve face verification with super-resolution because verification requires high-quality faces and precise details. Few previous works have successfully applied super-resolution to face verification. Recently, Ataer-Cansizoglu *et al.* [33] has tried to embed a deep SR network before a recognition network and train directly on the recognition criterion. However, this model cannot produce reasonable SR face. This means their SR module could only be used together with their proposed recognition network. On the contrary, IP-FSRGAN is a universal face super-resolution model which could not only reconstruct realistic SR faces but also be applied to improve various existing face recognition models.

As is shown in Fig 2, the original face verification receives two face images. After applying a series of operations including detection, alignment, normalization, and recognition network, two facial features are extracted from the given face images. Cosine similarity between the extracted face features could be viewed as the probability that they are of the same identity.

We propose a framework to improve face verification (See Fig. 2). After face detection on the input image, we filter the detected face images with a threshold of resolution. For those faces of low resolution, we first reconstruct SR faces with a super-resolution technique and then apply a face recognition network to extract features. For those faces whose resolution is higher than the threshold, we apply the face recognition network directly. Although most super-resolution models also work on HR images, they could not add more information to high-resolution images.





**FIGURE 3.** Synthesized super-resolution faces on CASIA-Webface dataset. (a) high-resolution image as the objective target. (b) low-resolution image downsampled from (a). The following columns are the synthesized faces by different models. The result of (c) SRCNN is relatively blur compared to the GAN based models. (d) SRGAN and (e) ESRGAN produce some unreasonable distortions such as the mouth and the eyes. By training with the id-preserving module, the proposed (f) IP-FSRGAN model could infer and synthesize much more realistic facial details. Please zoom in for better comparison.

What’s worse, super-resolution models may produce some noise to high-resolution images, which may obstruct the face verification. By setting the threshold, the noise problem could be avoided and the computational complexity is also reduced.

#### IV. EXPERIMENTS

##### A. DATASETS

For training, we mainly use the Large-scale CelebFaces Attributes (CelebA) Dataset [7]. We randomly separate CelebA into 162019 face images for training and 40580 face images for testing. We test these models on CelebA, CASIA-Webface [6], and LFW [8]. Our model is trained in RGB channels.

##### B. IMPLEMENTATION DETAILS

We have tried to train the network with local patches of size  $128 \times 128$  randomly cropped from the whole image, which is adopted in ESRGAN [3]. However, we find its performance is quite limited on the face dataset since the patches do not contain the whole face. In our experiments, we first detect the faces with a square bounding box and then resize these faces into  $128 \times 128$  images. We do not apply to face alignment. We use bicubic interpolation to downsample the faces with a scaling factor of 4 and receive LR faces image of size  $32 \times 32$ . As for the network architecture, we use the RRDB network proposed in [3] and a VGG discriminator. During training, we use a pre-trained LightCNN [29] as our ID-preserving network. We do not pre-train the generator and discriminator. The learning rate is initialized as  $2 \times 10^{-4}$  and decayed

as 70% every  $2 \times 10^5$  iterations. We set  $\lambda, \gamma, \eta$  and  $\xi$  in Eqn. 1 as 0.005, 10, 0.01, and 1. We use Adam optimizer with  $\beta_1 = 0.9, \beta_2 = 0.999$ .

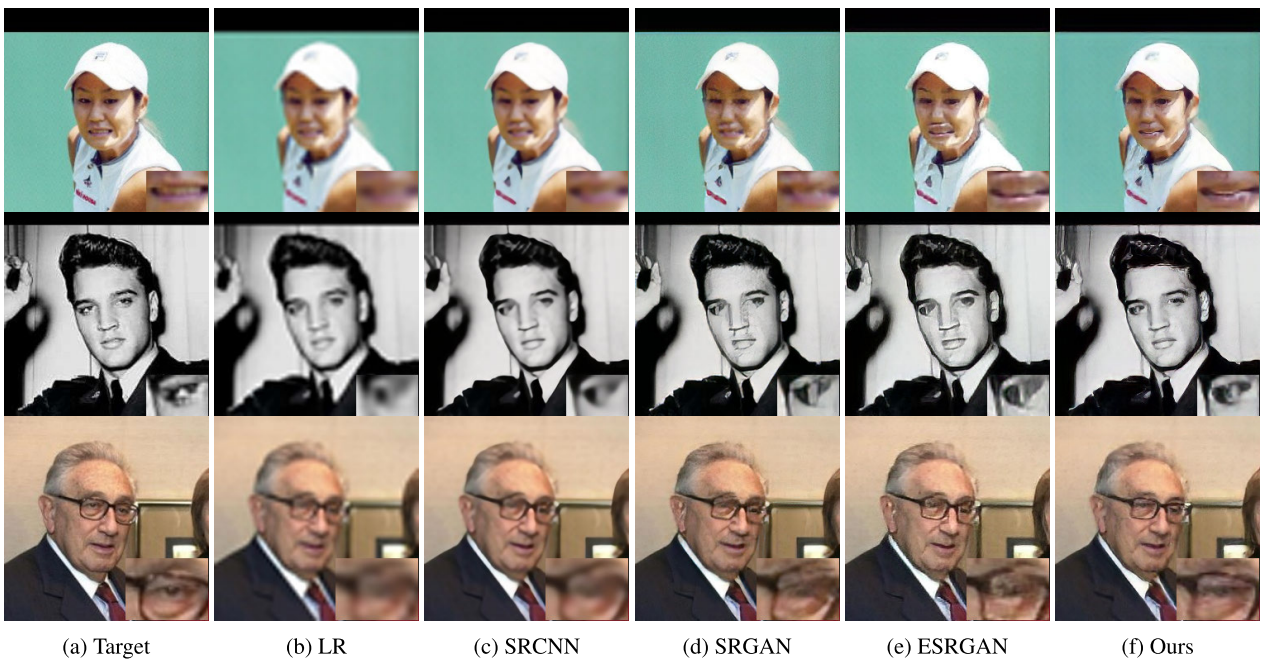
##### C. FACE SUPER-RESOLUTION

We compare the proposed IP-FSRGAN with several state-of-the-art super-resolution models including SRCNN [1], SRGAN [2], and ESRGAN [3]. All these models are trained on the CelebA dataset for 500000 mini-batches. The testing results on CASIA-Webface, CelebA, and LFW are shown in Fig. 3, Fig. 4, and Fig. 5. We prove by experiments that the superiority of IP-FSRGAN is reflected in its strong ability to infer from the low-resolution faces and synthesize realistic facial details including eyes, noses, mouths, and glasses. During this process, the identity before and after the super-resolution synthesis remains invariant. We will further discuss later that maintaining the facial details is critical for applying face super-resolution to the face verification task.

We report PSNR and SSIM metrics in Table 1 of each method on CASIA-Webface, CelebA, and LFW. The proposed IP-FSRGAN achieves the highest PSNR and SSIM, which should be attributed to the correctly synthesized facial details. We also notice that SRCNN obtains better PSNR and SSIM values than SRGAN and ESRGAN. This is mainly due to the fact that SRCNN is trained directly to optimize the PSNR metric. To further compare the performance, we report the cosine similarity (See Table 2) between the synthesized SR face and HR ground truth. We use a pre-trained



**FIGURE 4.** Synthesized super-resolution faces on CelebA dataset. In the first row, the noise synthesized by SRGAN and ESRGAN looks a bit unrealistic. In the second row, the shape of the eyes is changed by SRGAN and ESRGAN model. In the last row, the shape of the beard in the LR image confuses SRCNN, SRGAN, and ESRGAN models, which results in a horizon line under the nose. On the contrary, the IP-FSRGAN model could recognize and synthesize the beard correctly. Please zoom in for better comparison.



**FIGURE 5.** Synthesized super-resolution face images on LFW dataset. Line 1: ESRGAN model produces unrealistic mouth while the images synthesized by SRCNN and SRGAN are relatively blurred. Line 2: SRGAN and ESRGAN produce sharp edges on the face and fail to synthesize realistic eyes. Line 3: IP-FSRGAN is the only model to successfully infer and complement the glass. Please zoom in for better comparison.

LightCNN [29] network to extract the features. A larger cosine similarity represents a higher confidence that the SR face and HR face have the same identity. The proposed

IP-FSRGAN achieves the highest cosine similarity values as compared to others. The performance of SRCNN is much worse than other three methods in terms of the cosine

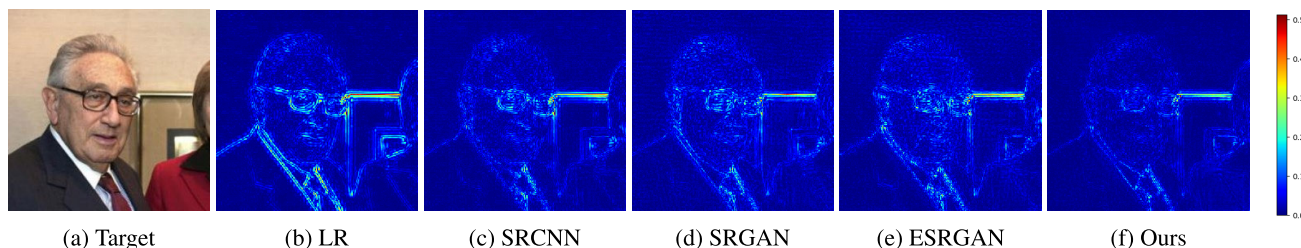


**TABLE 1.** The PSNR (dB) and SSIM of super-resolution models on CASIA-Webface, CelebA, and LFW datasets, the higher the better.

Dataset	LR	SRCNN	SRGAN	ESRGAN	IP-FSRGAN (Ours)
CASIA	29.87/0.793	33.17/0.904	32.92/0.892	31.14/0.877	<b>34.42/0.917</b>
CelebA	25.32/0.778	27.49/0.81	27.77/0.812	26.82/0.799	<b>28.91/0.845</b>
LFW	26.13/0.804	31.12/0.893	30.66/0.875	29.05/0.855	<b>32.58/0.908</b>

**TABLE 2.** Cosine similarity between SR and HR images on CASIA-Webface, CelebA, and LFW datasets. The larger the better.

Dataset	LR	SRCNN	SRGAN	ESRGAN	IP-FSRGAN (Ours)
CASIA	0.708	0.851	0.925	0.945	<b>0.968</b>
CelebA	0.647	0.814	0.906	0.921	<b>0.944</b>
LFW	0.682	0.828	0.915	0.929	<b>0.953</b>



**FIGURE 6.** Heatmap showing the pixel-wise difference compared with the target HR face in Figure ?? . Refer to Table 4 for numerical comparisons.

similarity metric. This is because the cosine similarity is more related to the visual effects, on which GANs based methods have been proven to work better than CNNs based methods in previous works [2], [3].

We further plot the heatmap in Figure 6 to show the difference between the synthesized images and the initial HR images in Figure 6a. The cool color represents a lower difference while the warm color represents a higher difference. The difference value  $\epsilon_{i,j} = \max_c \{|y_{i,j,c} - x_{i,j,c}|\}$  is calculated as the maximum pixel-wise difference under the three RGB channels  $c$ . The difference between the result of IP-FSRGAN in Figure 6f and the initial high resolution (Target) image is the lowest, which means the result of IP-FSRGAN is the closest to HR image, especially on the area of faces.

We report the MSE, PSNR, and SSIM values for each of the generated images with respect to the initial HR image in Table 4. From the table, we could find the result of LR image leads to the largest difference with respect to the HR face image, mainly due to the information loss during Bicubic downsampling and interpolation process. In contrast, SRCNN, SRGAN, and ESRGAN models produce much better results compared with Bicubic LR images, this is because the neural network of these models successfully infer partial information from the LR face images as inputs. However, their performance on the face area is still limited, especially in the area of eyeglasses. On the contrary, IP-FSRGAN could successfully infer the details, especially the facial details, from the input LR images and synthesize much more reliable super-resolution results.

### V. FACE VERIFICATION IMPROVEMENT

SphereFace [9] is one of the state-of-the-art face verification model. We use a pre-trained SphereFace network to extract the features of SR faces. We follow the 10-fold evaluation procedure of LFW dataset to calculate the accuracy.

**TABLE 3.** Average accuracy of face verification after applying different downsampling factor to LR. Factor=1 represents not applying further downsample for the LR images.

Model	Avg. Acc.	Factor=1
LR	0.585	0.928
SRCNN	0.601	0.948
SRGAN	0.620	0.962
ESRGAN	0.631	0.969
IP-FSRGAN	<b>0.656</b>	<b>0.976</b>

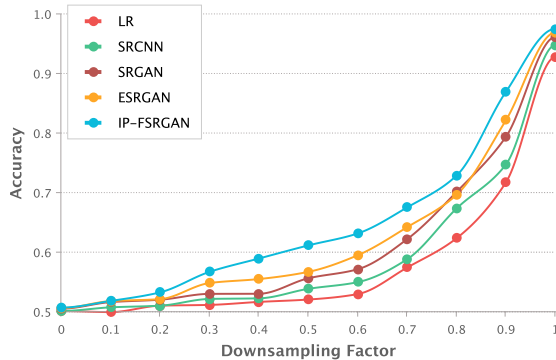
In addition, we find the performance of the face super-resolution model varies significantly with the resolution of input LR images. In order to evaluate the performance of face SR models under different resolutions of LR image, we first further downsample all the LR images in the test set by a downsample scaling factor  $\alpha$  between 0 and 1. For each  $\alpha$ , we applied the super-resolution algorithm and evaluate the synthesized SR image with a pre-trained face verification model. For every  $\alpha$ , we could calculate a corresponding accuracy. We could draw the accuracy- $\alpha$  curve (See Fig. 7), and observe how accuracy varies with  $\alpha$ . We also report the average accuracy upon different downsampling factors (See Table 3).

From Fig. 7, we could observe that the curve of IP-FSRGAN is above all the other models, which means IP-FSRGAN achieves the highest verification accuracy under different levels of blurring. Among the super-resolution models, the highest accuracy 97.6% is achieved by IP-FSRGAN, which is quite close to the performance of HR images (99.3%). Note that although IP-FSRGAN is trained using LightCNN as the ID preserving network, it still generalizes well to the SphereFace model, which proves that the ID preserving capability is transferable.

For the real-world application, we use our industrial partner’s face verification pipeline to detect and crop human faces, the cropped face image is then used in super-resolution, which works very well for most scenarios.

**TABLE 4.** The MSE, PSNR (dB) and SSIM of super-resolution models on the face image shown in Figure 6. MSE: the lower the better. PSNR and SSIM: the higher the better.

Metric	LR	SRCNN	SRGAN	ESRGAN	IP-FSRGAN (Ours)
MSE	108.09	99.14	82.62	102.44	72.15
PSNR	27.79	28.17	28.96	28.03	29.73
SSIM	0.860	0.868	0.867	0.857	0.898



**FIGURE 7.** The accuracy- $\alpha$  curve. The y axis is the accuracy of verification and the x axis is downsampling factor.

The only requirement has to be assured is the training and testing data have to use the same cropping method. In addition, evaluating the effective resolution of one given image may further help improve the result of face verification, we leave it for our future research work.

## VI. CONCLUSION

In this work, we address the challenge of information missing for super-resolution face generation by introducing a novel ID preserving module to help the generator learn to infer the facial details. The induced IP-FSRGAN model produces more realistic face images as compared to state-of-the-art super-resolution methods and improves the accuracy and robustness for the face verification task on low-resolution face images. Experimental results have demonstrated that the proposed IP-FSRGAN has excellent robustness for different downsample scaling factors and extensibility to various face verification models.

## REFERENCES

- [1] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2014, pp. 184–199.
- [2] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4681–4690.
- [3] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. C. Loy, "ESRGAN: Enhanced super-resolution generative adversarial networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 63–79.
- [4] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1125–1134.
- [5] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, *arXiv:1411.1784*. [Online]. Available: <http://arxiv.org/abs/1411.1784>
- [6] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Learning face representation from scratch," 2014, *arXiv:1411.7923*. [Online]. Available: <http://arxiv.org/abs/1411.7923>
- [7] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 3730–3738.
- [8] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Univ. Massachusetts, Amherst, MA, USA, Tech. Rep. 07-49, Oct. 2007.
- [9] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, "SphereFace: Deep hypersphere embedding for face recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 212–220.
- [10] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep Laplacian pyramid networks for fast and accurate super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 624–632.
- [11] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2472–2481.
- [12] J. Kim, J. K. Lee, and K. M. Lee, "Deeply-recursive convolutional network for image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1637–1645.
- [13] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3147–3155.
- [14] M. Haris, G. Shakhnarovich, and N. Ukita, "Deep back-projection networks for super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1664–1673.
- [15] N. Ahn, B. Kang, and K.-A. Sohn, "Image super-resolution via progressive cascading residual network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 904–9048.
- [16] K. Grm, W. J. Scheirer, and V. Štruc, "Face hallucination using cascaded super-resolution and identity priors," *IEEE Trans. Image Process.*, vol. 29, pp. 2150–2165, 2020.
- [17] X. Yu, B. Fernando, B. Ghanem, F. Porikli, and R. Hartley, "Face super-resolution guided by facial component heatmaps," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 217–233.
- [18] X. Yu, B. Fernando, R. Hartley, and F. Porikli, "Super-resolving very low-resolution face images with supplementary attributes," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 908–917.
- [19] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 694–711.
- [20] L. Gatys, A. S. Ecker, and M. Bethge, "Texture synthesis using convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 262–270.
- [21] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 136–144.
- [22] A. Jolicœur-Martineau, "The relativistic discriminator: A key element missing from standard GAN," 2018, *arXiv:1807.00734*. [Online]. Available: <http://arxiv.org/abs/1807.00734>
- [23] X. Tang, Z. Wang, W. Luo, and S. Gao, "Face aging with identity-preserved conditional generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7939–7947.
- [24] R. Huang, S. Zhang, T. Li, and R. He, "Beyond face rotation: Global and local perception GAN for photorealistic and identity preserving frontal view synthesis," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2439–2448.
- [25] K. Zhang, Z. Zhang, C.-W. Cheng, W. H. Hsu, Y. Qiao, W. Liu, and T. Zhang, "Super-identity convolutional neural network for face hallucination," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 183–198.
- [26] Y. Shen, P. Luo, P. Luo, J. Yan, X. Wang, and X. Tang, "FaceID-GAN: Learning a symmetry three-player GAN for identity-preserving face synthesis," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 821–830.



[27] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1701–1708.

[28] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," 2015, *arXiv:1503.03832*. [Online]. Available: <http://arxiv.org/abs/1503.03832>

[29] X. Wu, R. He, Z. Sun, and T. Tan, "A light CNN for deep face representation with noisy labels," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 11, pp. 2884–2896, Nov. 2018.

[30] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2414–2423.

[31] J. Bruna, P. Sprechmann, and Y. LeCun, "Super-resolution with deep convolutional sufficient statistics," 2015, *arXiv:1511.05666*. [Online]. Available: <http://arxiv.org/abs/1511.05666>

[32] Y. Bai, Y. Zhang, M. Ding, and B. Ghanem, "Finding tiny faces in the wild with generative adversarial network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 21–30.

[33] E. Ataer-Cansizoglu, M. Jones, Z. Zhang, and A. Sullivan, "Verification of very low-resolution faces using an identity-preserving deep face super-resolution network," 2019, *arXiv:1903.10974*. [Online]. Available: <http://arxiv.org/abs/1903.10974>



**JINNING LI** is currently pursuing the bachelor's degree with Shanghai Jiao Tong University, Shanghai, China. His research interests include computer vision and data mining. He has worked on the research on these fields for two years and has published three scientific articles.



**YICHEN ZHOU** is currently pursuing the Ph.D. degree in computer science with the National University of Singapore, Singapore. His research interests include computer vision and deep learning. His current research interest includes applications of deep learning in video analysis.



**JIE DING** (Member, IEEE) received the B.Eng. degree from Harbin Engineering University, China, in 2012, and the Ph.D. degree from Nanyang Technological University, Singapore, in 2018. She was a Scientist with the Institute for Infocomm Research (I2R), Agency for Science, Technology and Research (A\*STAR), Singapore, until August 2019. Since September 2019, she has been with the Department of Electronic Engineering, Fudan University, China, where she is currently a pre-tenure Associate Professor. Her research interests include machine learning, pattern recognition, control and optimization, and complex networks.



**CEN CHEN** received the Ph.D. degree in computer science from Hunan University, China. He currently works as a Scientist II with the Institute for Infocomm Research (I2R), Agency for Science, Technology and Research (A\*STAR), Singapore. He has published several research papers in international conference and journals of machine learning algorithms and parallel computing, such as the IEEE TRANSACTIONS ON COMPUTERS (IEEE-TC), the IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS (IEEE TPDS), AAAI, ICDM, ICPP, and many more. His research interests include parallel and distributed computing, machine learning, and deep learning.



**XULEI YANG** (Senior Member, IEEE) is currently a Senior Research Scientist and the Programme Head with the Institute for Infocomm Research (I2R), A\*STAR, with more than 13 years of research and development experiences in image/signal analysis and deep/machine learning. He is also the Kaggle Competition Master. He is the Principal Investigator for various projects involved in providing deep learning solutions for healthcare, Fintech, and computer vision. He has published more than 60 scientific articles and international patents. His current research interests include deep learning for image/signal analysis, recommendation, and prediction.

...