Journal of Artificial Intelligence Research 74 (2022) 1715-1737 Submitted 03/2022; published 08/2022

C-Face: Using Compare Face on Face Hallucination for Low-Resolution Face Recognition

Feng Han FENGHAN@SMAIL.NJU.EDU.CN
Xudong Wang WANGXD@SMAIL.NJU.EDU.CN
State Key Laboratory for Novel Software Technology,
Nanjing University, Nanjing 210046, China
Department of Computer Science and Technology,
Nanjing University, Nanjing 210046, China

Furao Shen *(corresponding author)* FRSHEN@NJU.EDU.CN
State Key Laboratory for Novel Software Technology,
Nanjing University, Nanjing 210046, China
School of Artificial Intelligence,
Nanjing University, Nanjing 210046, China

Jian Zhao *(corresponding author)* JIANZHAO@NJU.EDU.CN
School of Electronic Science and Engineering,
Nanjing University, Nanjing 210046, China

Abstract

Face hallucination is a task of generating high-resolution (HR) face images from low-resolution (LR) inputs, which is a subfield of the general image super-resolution. However, most of the previous methods only consider the visual effect, ignoring how to maintain the identity of the face. In this work, we propose a novel face hallucination model, called C-Face network, which can generate HR images with high visual quality while preserving the identity information. A face recognition network is used to extract the identity features in the training process. In order to make the reconstructed face images keep the identity information to a great extent, a novel metric, i.e., C-Face loss, is proposed. We also propose a new training algorithm to deal with the convergence problem. Moreover, since our work mainly focuses on the recognition accuracy of the output, we integrate face recognition into the face hallucination process which ensures that the model can be used in real scenarios. Extensive experiments on two large scale face datasets demonstrate that our C-Face network has the best performance compared with other state-of-the-art methods.

1. Introduction

Face hallucination is a domain-specific super-resolution (SR) task, which aims to super-resolve a low-resolution (LR) face image into a high-resolution (HR) one. Face hallucination model has many applications in practice. Specifically, it can help to recognize faces with a very low resolution. Although face recognition methods such as SphereFace (Liu et al., 2017) or ArcFace (Deng et al., 2019) have already achieved an impressive achievement and surpassed human-level performance, most of them have a very poor performance when the resolution of the input image is low. However, there are many cases where we only have LR images to identify (e.g., video surveillance). The most effective and direct way to solve this

Figure 1: Comparison of face hallucination: (a) LR face (16×16), SR faces reconstructed by (b) general face hallucination (e.g., (Zhang et al., 2018)), (c) identity-preserving face hallucination (e.g., (Zhang et al., 2018a)), (d) identity-preserving face hallucination with compare face (our method), and (e) the ground-truth HR face.

problem is to reconstruct the HR images using face hallucination methods, so that they can be recognized by face recognition models with high accuracy.

With the development of deep learning, many recent works in the field of super-resolution resort to artificial neural network to learn the feature relationship between the HR images and their LR counterparts. However, the super-resolution of face images, i.e., face hallucination, is different from the super-resolution of ordinary images, because the distribution of geometric features is different when faces occupy most of the images. As shown in Fig. 1(b), although using the general super-resolution model to super-resolve the face images can produce visually pleasing results, it cannot help improve the face recognition accuracy. Thus, face hallucination methods should have more considerations about maintaining facial features compared with general super-resolution methods.

There are also many works, such as the work of Huang et al. (2017) and Bulat and Tzimiropoulos (2018), focusing on super-resolving face images by considering some characteristics of the human faces. Through the capabilities of neural networks, most of them have achieved better results than traditional methods. However, most of the previous face hallucination methods only pay attention to visual quality and ignore how to maintain identity information during reconstruction.

Whether the reconstructed images can facilitate the performance of face recognition depends on whether the identity information can be preserved in the face hallucination process. Therefore, the identity-preserving face hallucination methods have attracted increasing research attention (Zhang et al., 2018a; Hsu et al., 2019) in recent years. However, most of the existing identity-preserving face hallucination methods only minimize the distance of the SR face image and its corresponding ground-truth HR face image in the identity feature space. The identity loss objective function of these methods is likely to be trapped into a local optimum. These methods cannot be applied to recognize other face images of the same identity, which may have distinct appearance, as shown in Fig. 1(c). Unlike those methods, our work compares the output face image with other face images of the same identity, i.e., images from the same person, during the training phase. In this way, our proposed method has a better ability to maintain the identity information compared with previous methods, as shown in Fig. 1(d).

According to the previous idea, we propose a face hallucination model with a novel loss function, as well as a new training method. The loss function we propose is called the C-Face loss, where C-Face stands for "compare face". In the face dataset, some images that belong to the same person can be used as an identity reference for each other during the training phase. For each image in the training dataset, we call its reference image as the "compare face". In C-Face loss, we calculate the difference between the output SR image and its target, i.e., the ground-truth HR face image, as well as the difference from the compare face. By using this loss function to train our network, the resulting model have a better ability to maintain the identity information from the LR input.

Directly combining the C-Face loss with other common loss function to train the network can improve the performance of the model. However, the loss function will not decrease to a sufficiently small value for the reason that the output images and the original HR images are in different manifolds in the high-dimensional space of the identity feature. To solve this problem, we propose a new training algorithm with four stages. Experiments show that the model has a better performance after our new training process is applied.

We use a generative adversarial network mentioned in Ledig et al. (2017) as the base architecture of our model, and we name our model as C-Face network. The model is trained on the CASIA-WebFace dataset (Yi et al., 2014), and we test it on the LFW dataset (Huang et al., 2008) including the standard protocol (with only face verification) as well as the LFW-BLUFR protocol (Liao et al., 2014). Moreover, we also verify the performance of our proposed model on the Celebrity Face Attributes (CelebA) dataset (Liu et al., 2015) which is a more challenging in the wild face images dataset. We compare the results of our model with some previous works to show that our method has the best ability to preserve the identity information in face hallucination. The main contributions of this paper can be summarized as follows:

- We propose a model named C-Face network with a novel loss function that can maintain the identity information when super-resolving the LR face images. Our loss function not only penalizes the distance of identity features between the SR face image and its corresponding ground-truth HR face image, but also decreases the distance with other examples of the same identity.

- A four-stage training algorithm is proposed to deal with the convergence problem for our model. By using this multi-stage training strategy, the hallucinated SR face images and the ground-truth HR face images will gradually move closer to each other in the same identity feature space.

- We conduct extensive experiments on three large-scale datasets to evaluate the effectiveness of our proposed model. Compared with several existing state-of-the-art face hallucination models, the C-Face network has the best performance on identity preservation ability.

The rest of this paper is organized as follows. Section 2 introduces some work that is related to this paper. Our method is presented in Section 3 and the experimental results are given in Section 4 and 5 which prove the validity of our method. Finally, Section 6 concludes our work.

2. Related Work

In this section, we review the literature on the most related work of general image super-resolution, identity-unaware face hallucination, and identity-aware face hallucination and put this work in an appropriate context.

2.1 General Image Super-Resolution

The image super-resolution field has achieved impressive success since deep learning based methods. The method in Dong et al. (2014) first uses a CNN model to map the LR images to HR ones. Motivated by the idea of residual learning, the very deep super resolution network (VDSR) (Kim, Lee, & Lee, 2016a) and the deeply-recursive convolutional network (DRCN) (Kim, Lee, & Lee, 2016b) tried to explore deeper architectures and improved accuracy. To achieve visually-pleasing results, Ledig et al. (2017) employ the generative adversarial network (GAN) (Goodfellow et al., 2014) with the adversarial loss to construct the super-resolution model (SRGAN). The SRGAN model also uses the perceptual loss (Johnson, Alahi, & Fei-Fei, 2016), which can generate photo-realistic HR images but cannot achieve a high grade in PSNR or SSIM indicators. In order to generate more perceptually satisfying results, Wang et al. (2018) proposed an enhanced SRGAN (ESRGAN) to improve SRGAN performance by a relativistic adversarial loss. In contrast to adversarial loss, Mechrez et al. (2018) proposed a contextual loss to maintain natural image statistics by measuring the distribution of features. Lim et al. (2017) enhanced the previous framework by removing batch normalization layers in conventional residual networks and expanding the model size.

2.2 Identity-Unaware Face Hallucination

Aiming at reconstructing the LR face images, face hallucination is a common research topic in both traditional image processing and deep learning. Based on the idea of sparse representation, Farrugia and Guillemot (2017) used a sparse coding scheme to form the HR patches by globally optimal LR patches. Jiang et al. (2017b) used two kinds of smoothing constraints to alleviate noise effects during the reconstruction. However, these traditional methods show limited advantages compared with deep learning models.

The performance of face hallucination has been improved significantly by using deep neural network. Yu and Porikli (2016) utilized a discriminative generative network to improve the scaling factors from $2 \sim 4\times$ to $8\times$ so as to ultra-resolve a very low resolution face image. To use facial features in the deep neural network, Huang et al. (2017) combined the wavelet transform and the CNN model to construct a new architecture, which can generate the HR images with the predicted wavelet coefficient. Chen et al. (2018) and Bulat and Tzimiropoulos (2018) used some geometry prior, such as facial landmarks or feature heatmaps, to calculate parts of the loss function during the training process. Yu et al. (2018a) proposed a method that used multi-task convolutional neural networks to explicitly incorporate the structural information of faces into the face super-resolution process. Focusing on one-to-many ambiguity in face hallucination tasks, Yu et al. (2018b) found that using some supplementary attributes during the face super-resolution process can significantly reduce that ambiguity. Yu et al. (2020) constructed an attribute-embedded

upsampling network, using supplementary residual images with additional facial attribute information to reduce the ambiguity in face hallucination. Conventional face hallucination methods require accurate alignment of low-resolution faces before upsampling them. In order to directly super-resolve unaligned face images, Yu and Porikli (2017) proposed an end-to-end transformative discriminative neural network to allow local receptive fields to line-up with similar spatial supports by embedding spatial transformation layers into the upsampling network. Motivated by this idea, the follow-up work (Yu et al., 2020) proposed a multiscale transformative discriminative neural network, which can be used for super-resolving small face images of different resolutions. In order to solve the problem that the performance of the existing face hallucination methods based on convolutional neural networks is significantly degraded under low and non-uniform illumination conditions, Zhang et al. (2020) and Zhang et al. (2021) proposed a copy and paste generative adversarial network to offset illumination and enhance facial details.

2.3 Identity-Aware Face Hallucination

The above-mentioned face hallucination methods have considered the face structure prior and attribute information. However, most of them did not consider the identity information of the face images or did not fully consider two similar tasks, face hallucination and face recognition, together. The identity information contains identity-aware details which is essential for boosting down-stream face recognition accuracy after reconstruction (Jiang et al., 2021).

Face recognition technology has made rapid progress in the past few years. The work of Taigman et al. (2014) achieved a breakthrough performance on recognition accuracy by using the CNN model and softmax loss function. However, softmax loss may be insufficient in distinguishing thousands or more classes. Schroff, Kalenichenko, and Philbin (2015) proposed the triplet loss to minimize the distance between the features from the same identity and maximize the features from different identities. Inspired by triplet loss, we try to improve the identity loss in Zhang et al. (2018a) and finally make the network model have a better ability for maintaining identity information in this paper. In Liu et al. (2017) and Deng et al. (2019), the angular-margin-based loss functions were used to make the inter-class distance larger than the intra-class distance. Zangeneh et al. (2020) proposed a coupled architecture for low resolution face recognition using two branches of deep convolutional neural networks, which can project the HR and LR faces into a common feature space. In our method, we use the recognition model in Liu et al. (2017) to extract identity information and use it in the new loss function we propose. Thus, our method has an excellent ability to maintain identity information during reconstruction and improve the performance of down-stream face recognition.

As mentioned before, identity-preserving face hallucination has received increasing attention recently because of its benefits in maintaining the identity information. Jiang et al. (2017a) used a method based on smooth regression to learn the relationship between facial images with different resolution and proved the validity of their method on some public face recognition datasets. Zhang et al. (2018a) and Huang et al. (2019) used the identity loss to add a recognition constraint during the training process of the face hallucination model. Instead of directly obtaining identity features from the original SR face image

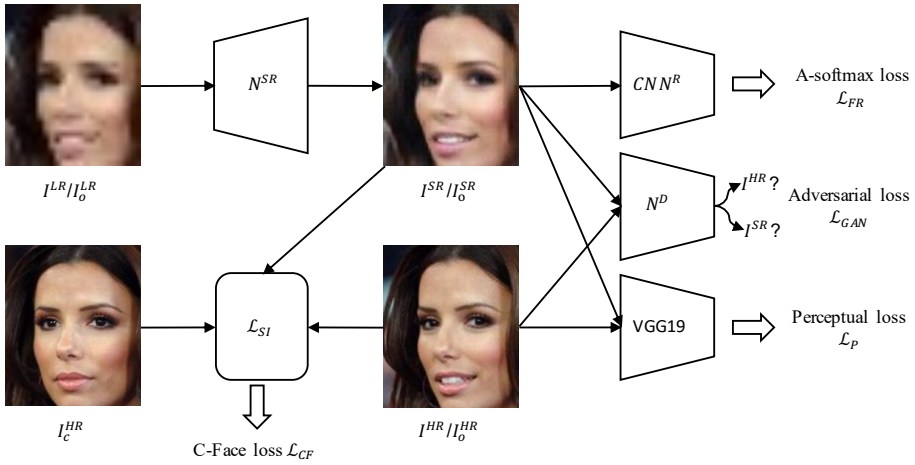


Figure 2: Overview of our proposed approach. It consists of a face hallucination GAN and a face recognition network and a pretrained perceptual feature extractor network. During the training phase, we employ 1) angular softmax (A-Softmax) loss to learn angularly discriminative features, 2) adversarial loss and perceptual loss to differentiate between the SR images and original HR images, and 3) our C-Face loss to retain face identity.

and the HR face image, Grm et al. (2020) extracts them from constructed residual images between SR (or HR) and LR. The training of face recognition-based methods utilizing identity loss needs well-labeled datasets which are costly. Instead, Hsu et al. (2019) proposed a siamese generative adversarial networks (SiGAN) to reconstruct high-resolution faces while preserving identity information with weak binary pairwise label. Specifically, SiGAN designs an identity-distinguishable contrastive loss that not only decreases the difference of same-identity pairs but also increases the difference of different-identity pairs. Different from existing identity-preserving face hallucination methods that just calculate the identity loss between the SR face image and its corresponding HR face image, our proposed method also minimizes the difference with the compare face image (another HR face image of the same person) at the same time. In this manner, our proposed method can produce a better identity-preserving reconstruction while improving the generalization ability of the hallucination model.

3. C-Face Network

In this section, we provide the details of our proposed C-Face network, as shown in Fig. 2. We first describe the identity-preserving SRGAN which is the backbone network of our C-Face network. Then we introduce the proposed C-Face loss to achieve further effectively identity-preserving reconstruction. After that, we show the selection method of compare face images in detail and analyze the effects of different selection strategies. Finally, we introduce the proposed training procedure to overcome the convergence problem.

3.1 Identity-Preserving SRGAN

The proposed C-Face network adopts the SRGAN network (Ledig et al., 2017) as the backbone network, which is a generative adversarial network for image super-resolution with perceptual loss. The perceptual loss is formulated as the weighted sum of an adversarial loss and a content loss term. In order to achieve identity-preserving reconstruction, we employ the SphereFace model (Liu et al., 2017) to extract the identity features from face images during training. This subnetwork is trained by A-Softmax loss function which can further increase the angular margin of learned face features.

3.1.1 Adversarial Loss

Adversarial loss was first introduced in generative adversarial networks (GANs) (Goodfellow et al., 2014) to improve the data generator N^{SR} until it generates something that resembles the real data. It fools the discriminator network, N^D, that classifies an image as real or fake. The adversarial loss function in our model can be expressed as follows:

$$\mathcal{L}_{GAN}(I^{LR}, I^{HR}) = \mathbb{E}[\log(N^D(I^{HR}))] + \mathbb{E}[\log(1 - N^D(N^{SR}(I^{LR})))], \tag{1}$$

where I^{LR} is the LR input of the N^{SR} and I^{HR} is the counterpart HR images, which can be seen as the target. The output image of the N^{SR} is denoted as I^{SR} in the following.

Through the continuous confrontation during training, N^{SR} will eventually generate the high quality output so that N^D cannot discriminate whether it is the original HR image or a super-resolved one.

3.1.2 Content Loss

We use the perceptual loss as the content loss, which was first proposed in Johnson et al. (2016), to make sure that the output of our model has a good visual quality. The perceptual loss focuses on the similarity of the perceptual or the texture representations by comparing the feature maps extracted by the pretrained 19 layers VGG network. We use $\varphi_{i,j}(I)$ to denote the feature map of the input I after the activation of the j-th convolution layer and before the i-th max pooling layer within the VGG19 network. The perceptual loss calculates the Euclidean distance between two feature maps:

$$\mathcal{L}_P(I^{SR}, I^{HR}) = \frac{1}{W_{i,j}H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\varphi_{i,j}(I^{HR})_{x,y} - \varphi_{i,j}(I^{SR})_{x,y})^2. \tag{2}$$

$W_{i,j}$ and $H_{i,j}$ correspond to the length and width of $\varphi_{i,j}(.)$ respectively. In our experiments, we set $i = 3$ and $j = 5$ for the perceptual loss when training the C-Face network.

3.1.3 A-Softmax Loss

Similar to Zangeneh et al. (2020) and Zhang et al. (2018a), we use a face recognition model, which is denoted as CNN^R, to assist the training. In theory, CNN^R could be any face recognition network that can output the identity features. In our work, we use the SphereFace (Liu et al., 2017) as CNN^R, which is trained by the A-Softmax loss function.

For the input I_i that belongs to the y_i-th identity, the A-Softmax loss is represented as:

$$\mathcal{L}_{FR} = \frac{1}{N}\sum_{i} -\log\Big(\frac{e^{\|\phi(I_i)\|\psi(m\theta_{y_i})}}{e^{\|\phi(I_i)\|\psi(m\theta_{y_i})} + \sum\limits_{j\neq y_i} e^{\|\phi(I_i)\|\cos(\theta_{y_j})}}\Big). \tag{3}$$

We set $m = 4$ and $\psi(.)$ is a monotonically decreasing angle function. $\phi(I_i)$ denotes the identity feature of the input image I_i, which is the output of the fully connected layer before the classification in CNN^R. The $\mathcal{L}_{FR}$ is not used to train N^{SR} or N^D directly. Instead, we use it to fine-tune the CNN^R in Section 3.4.

Directly using $\mathcal{L}_{GAN}$ and $\mathcal{L}_P$ to train the GAN model can produce the output with good visual quality, but this has not taken the identity information into consideration. A direct idea is to use the $\mathcal{L}_{FR}$ to train the GAN model to make the output I^{SR} to have the same classification result as the input human face. However, face hallucination and face recognition are two different tasks with different objectives, so $\mathcal{L}_{FR}$ has a limited effect when it is used to train the face hallucination model directly.

3.2 C-Face

To further effectively maintain the identity information, we introduce in the training process an C-Face loss which aims at penalizing the distance of SR/HR image pairs while decreasing the distance between the SR and other examples with the same identity.

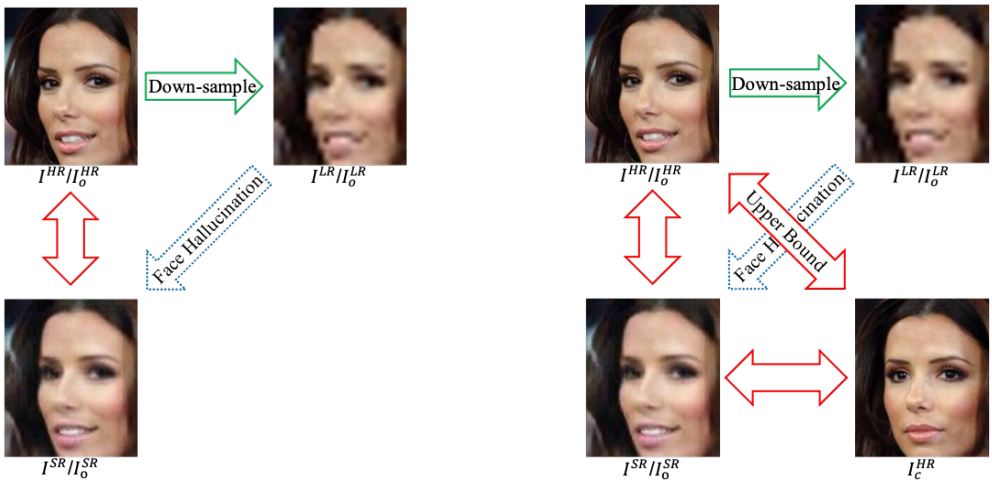
Zhang et al. (2018a) proposed the super-identity loss (abbreviated as $\mathcal{L}_{SI}$) that can help to keep the identity information. The core idea of $\mathcal{L}_{SI}$ is to penalize the normalized Euclidean distance between the features of I^{SR} and I^{HR}. The $\mathcal{L}_{SI}$ can be written as:

$$\mathcal{L}_{SI}(I^{HR}, I^{SR}) = \Big\| \frac{\phi(I^{SR})}{\| \phi(I^{SR}) \|_2} - \frac{\phi(I^{HR})}{\| \phi(I^{HR}) \|_2} \Big\|_2^2. \tag{4}$$

As mentioned before, $\phi(I)$ represents the identity feature of the input image I extracted by CNN^R.

The $\mathcal{L}_{SI}$ only tries to reduce the distance of identity information between the original image and the output result. It does not take full advantage of the face recognition model as well as the characteristic of the human face datasets. In face recognition datasets, different images in the same class are the photos of the same person, identity features of these images can be compared with each other with a relatively higher accuracy. In order to utilize this characteristic in our face hallucination model, we want $\phi(I^{SR})$ to be not only close to the $\phi(I^{HR})$, but also close the feature vector of the images that have the same identity label with I^{HR}. Inspired by this, for each I^{HR} in training data, we pick an image I_C^{HR} from the images that have the same identity label as I^{HR} in the training set. In other words, the I_C^{HR} is an image that belongs to the same person and is sufficiently similar to I^{HR}. During training, we want $\phi(I^{SR})$ to be similar to both $\phi(I^{HR})$ and $\phi(I_C^{HR})$. We will give the selection method of I_C^{HR} and explain it later in Section 3.3.

To avoid confusions, we use I_o^{LR} to represent the LR input in the following which was denoted as I^{LR} in (1) and (4). Similarly, I_o^{HR} and I_o^{SR} are used to replace I^{HR} and I^{SR}, respectively. By picking an I_c^{HR} for every input I_o^{LR}, we penalize the distance between the identity features of I_o^{HR} and I_c^{HR}. In this way, we have a novel loss function named



(a) Traditional Method

(b) C-Face Loss

Figure 3: Sketch map for the core idea of C-Face loss. All the images above belong to the same person. The red arrow means the difference between two images. The C-Face loss concerns about the difference between two different images that belong to the same person.

C-Face loss, where "C-Face" is short for "compare face". Our study found that using the normalized Euclidean distance to measure the distance can get a better result than other measures such as the L2-loss or the cosine distance. As defined by (4), we use $\mathcal{L}_{SI}(I_1, I_2)$ to denote the normalized Euclidean distance between the identity features.

During training, we penalize the distance between I_o^{SR} and I_o^{HR}, as well as the distance between I_o^{SR} and I_c^{HR}. In this way, we expect that the identity information of the output I^{SR} is similar to those images that have the same identity label with I^{HR}. Now, we have the first version of the C-Face loss function $\mathcal{L}_{CF}$, which can be represented as:

$$\mathcal{L}_{CF} = \gamma_1 \mathcal{L}_{SI}(I_o^{SR}, I_o^{HR}) + \gamma_2 \mathcal{L}_{SI}(I_o^{SR}, I_c^{HR}) \tag{5}$$

where the hyper-parameters γ_1 and γ_2 are constants during training.

In super-resolution tasks, the high-resolution image I_o^{HR} is usually regarded as the upper bound of the reconstructed super-resolution image I_o^{SR} in terms of visual quality and the identity information it contains. Thus, we require that the distance between I_o^{HR} and I_c^{HR} should be smaller than the distance between I_o^{SR} and I_c^{HR}. This is illustrated in Fig. 3. Therefore, we modified the first version of the C-Face loss and use $\mathcal{L}_{SI}(I_o^{HR}, I_c^{HR})$ as the lower bound of the second term in (5). The final version of the C-Face loss is represented as:

$$\mathcal{L}_{CF} = \gamma_1 \mathcal{L}_{SI}(I_o^{SR}, I_o^{HR}) + \gamma_2 \max\Big(\mathcal{L}_{SI}(I_o^{SR}, I_c^{HR}) - \mathcal{L}_{SI}(I_o^{HR}, I_c^{HR}), 0\Big). \tag{6}$$

As shown in Fig. 4, other methods only make I_o^{SR} to be as close to I_o^{HR} as possible. After some training, I_o^{SR} will stay close to I_o^{HR}, but they may be sparse with each other. However, by using the C-Face loss function, the output images of our method will not only stay close to their I_o^{HR}, but also be close to images in the same class (I_c^{HR}).

(a) Other Methods (b) Our Method

Figure 4: The design idea of our C-Face loss function. Assume that the feature space of the identity information is a two-dimensional plane. The blue star icons represent the position of the feature vector for some facial images that belong to the same person. The green hollow triangle symbols represent the feature vectors of the output I^{SR} images and the solid green triangle symbols represent the reconstruction results after some training. Fig. 4a represents other methods such as SICNN. Fig. 4b represents our method. Our method can make the reconstruction results stay closer to each other when they belong to the same identity label.

To sum up, the loss function that is used to train our GAN-based face hallucination model has three different parts, which can be formulated as:

$$\mathcal{L}_{total} = \alpha \mathcal{L}_{GAN} + \beta \mathcal{L}_P + \mathcal{L}_{CF} \tag{7}$$

where the hyperparameters α and β are constants. In order to optimize the model, we use stochastic gradient descent. The loss function $\mathcal{L}_{total}$ is differentiable with regard to all the parameters. In the end, N^{SR} is able to super-resolve face images while maintaining the identity information. We name the currently obtained N^{SR} as C-Face-v1.

3.3 Selection of the Compare Images

Now we discuss how to choose I_c. In the face dataset, the images of the same person may be quite different due to factors such as angle or expression. So a randomly chosen I_c^{HR} may be quite different from I_o^{HR}, and $\mathcal{L}_{CF}$ will not drop during training. For similar reasons, we cannot directly use the $\mathcal{L}_{FR}$ to train the face hallucination model. Before training, we use CNN^R to get the identity feature vector for every image I_o^{HR} in the training set. Then we calculate the normalized Euclidean distance between the feature vectors of any two images that belong to the same person and record the k most similar images. For every input I_o^{LR} during training, we randomly choose an I_c^{HR} from those k images that are most similar to I_o^{HR}. Through experimental exploration, we found that $k = 3$ has the best performance.

The Fig. 5 and Fig. 6 give an explanation of the result on k. As we have discussed before, the core idea of the C-Face loss is that we picked a similar image that can be used to compare with the output images. If the compare image I_c has a large difference with the original image I_o, the face hallucination model, which is used to reconstruct the image, may not have the ability to catch the similarity between I_o and I_c. As shown in Fig. 5, when the k value is large or we just randomly choose the I_c, there is a certain probability of choosing an image that has a large difference with I_o. When we use $k < 3$, we will pick the image that is very close to the I_o. However, as can be seen in Fig. 6, the I_c will be too similar to the I_o so that the C-Face loss function will not work with this choosing method.

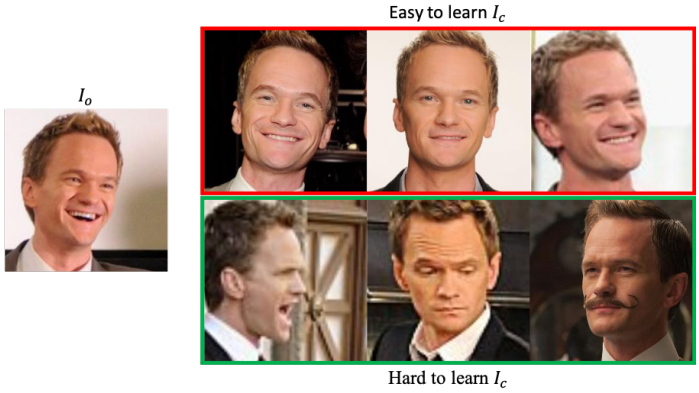


Figure 5: For the I_o, we randomly choose I_c from the three images in the first row that are most similar to I_o. However, if we randomly pick the I_c, images in the second row may be picked and it will be so different with I_o that the face hallucination model cannot capture the similarity.

Figure 6: The second row is the most similar face image in the dataset of the first row. If we choose the I_c from the second row, the C-Face loss will not work.

3.4 Training Procedure

Although C-Face-v1 can achieve a better result compared with most of the previous works, the C-Face loss in (7) cannot decrease low enough by using this direct training method. The CNN^R model is only trained by the original HR images, so the feature vectors of I^{SR} and I^{HR} extracted by CNN^R have different feature distributions. I^{SR} and I^{HR} stay in different domains of the feature space, even though they have the same identity.

Following this idea, we propose a new training algorithm with several stages as shown in Algorithm 1. In stage 3 of Algorithm 1, we fine-tune the CNN^R to make it map the original images and the super-resolved images into the same feature space. In this way, the CNN^R will better recognize I^{SR}. For similar reasons, we also fine-tune the CNN^R in each training step in stage 4. Using the fine-tuned CNN^R to calculate the feature vector for the C-Face loss, the I^{SR} output by N^{SR} will be more discriminative. Trained by this method, we get the N^{SR} as our final C-Face network. In order to prevent confusion with C-Face-v1 as well as to emphasize the effectiveness of the training algorithm, we call it C-Face-v2 in the following.

Algorithm 1 Training Procedure of C-Face Network

Input: CNN^R pretrained by HR face images, the GAN model including N^{SR} and N^D, a set of three images including I_o^{LR}, I_o^{HR} and I_c^{HR}.

1: Train N^{SR} and N^D with the loss function in (7);
2: Use the N^{SR} that we obtained after stage 1 to process all the LR images. Now, we have two datasets: the original HR dataset $\mathbb{D}^{HR}$ and the super-resolved dataset $\mathbb{D}^{SR}$;
3: Mix $\mathbb{D}^{HR}$ and $\mathbb{D}^{SR}$ together, and use the combined dataset to fine-tune CNN^R;
4: Fine-tune N^{SR} and N^D by using the CNN^R after stage 3 to extract features. In each step, we update N^{SR} and N^D by descending the loss function (7), we also update the CNN^R with $\mathcal{L}_{FR}$ in (3);

Output: The final C-Face network.

4. Experiments and Analysis

We compared the C-Face network on several public benchmark datasets with state-of-the-art general image super-resolution methods including Bicubic interpolation, VDSR (Kim et al., 2016a), EDSR (Lim et al., 2017), SRGAN (Ledig et al., 2017), DBPN (Haris, Shakhnarovich, & Ukita, 2018), RDN (Zhang et al., 2018), ESRGAN (Wang et al., 2018), and SRFBN (Li et al., 2019), and also with face super-resolution approaches including Wavelet-SRNet (Huang et al., 2017), and SICNN (Zhang et al., 2018a).

4.1 Datasets and Evaluation Metrics

In our experiments, we use the CASIA-WebFace dataset (Yi et al., 2014) with $494,414$ images as our training set and use the LFW dataset (Huang et al., 2008) with $13,233$ images as our testing set. Moreover, we also verify the performance on CelebA (Liu et al., 2015) which is a more challenging in the wild face images dataset. Before training, we first crop all the images from these datasets into 128×128 pixels. The image set we get after cropping is named as the HR-set. The size of input LR face images is down-sampled to 16×16, 32×32, and 64×64 and then upscaled to 128×128, respectively, by various scaling factor.

For the evaluation, we perform face recognition and verification on reconstructed SR face images, and use the recognition/verification accuracy as the indicator to evaluate whether the reconstructed SR images are suitable for identity recognition. We also adopt the widely-used pixel-wise metrics, Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM), for measuring reconstruction fidelity.

4.2 Experimental Settings

We implement our C-Face networks with the PyTorch framework and train them using four NVIDIA 1080Ti GPUs. We use Adam optimizer with a decayed learning rate of 0.99 and the batch size is 64 in every experiment. The learning rates of the GAN in stage 1 and stage 4 are 0.0002 and 1×10^{-5} respectively. We train the GAN model for 10 epochs in stage 1 and 5 epochs in stage 4. The initial learning rate of stage 3 is 0.1 but it will be multiplied by 0.1 after each epoch and the CNN^R will be fine-tuned for 10 epochs. We

Table 1: Comparison of face verification rates evaluated by SphereFace for various face hallucination models on standard LFW protocol.

Models	Scale=2	Scale=4	Scale=8
HR (128 × 128)	0.9508	0.9508	0.9508
Bicubic	0.9472	0.8775	0.6460
VDSR	0.9487	0.9152	0.7442
Wavelet-SRNet	0.9483	0.9340	0.8688
EDSR	0.9468	0.9330	0.8460
SRGAN	0.9495	0.9315	0.8437
DBPN	0.9487	0.9323	0.8633
RDN	0.9482	0.9365	0.8750
SICNN	0.9457	0.9103	0.7243
ESRGAN	**0.9505**	0.9338	0.8747
SRFBN	0.9463	0.9338	0.8603
C-Face (Ours)	0.9503	**0.9403**	**0.8897**

set $\alpha = 0.1$ and $\beta = 1.0$ for (7) in each stage. For (6), we set $\gamma_1 = 0.05$, $\gamma_2 = 0.1$ in stage 1 and $\gamma_1 = \gamma_2 = 0.05$ in stage 4. For the sake of fairness, all the models in the comparison experiments are trained by the CASIA-WebFace dataset from scratch after the data pre-process as mentioned above.

4.3 Evaluation on LFW

4.3.1 Standard Protocol

To make sure that our model has state-of-the-art performance in maintaining identity information, we use each model to reconstruct the downscaled LFW dataset and take the resulting images for face recognition. The SphereFace model (Liu et al., 2017) is used as the recognition model. Although the CNN^R, which is also the SphereFace model in our experiments, has been fine-tuned in the training algorithm as we proposed. We use the original SphereFace model trained by the CASIA-WebFace dataset in the testing experiments for fair comparisons.

We first use the standard LFW protocol (Liao et al., 2014) to test the results of each model and the testing results are shown in Table 1. The accuracy of the C-Face network is higher than all the comparing models on upscaling factor of 4× and 8×. All models perform very closely on upscaling factor of 2×, our model is only 0.0002% behind the best model.

We also give the testing results of the average PSNR and SSIM on LFW in Table 2. Our method is not superior in these two indicators, but it has maintained a relatively acceptable level. As has been mentioned in Ledig et al. (2017) and Zhang et al. (2018b), the level of

Table 2: Comparison of PSNR and SSIM of SR faces reconstructed by various face hallucination models on LFW.

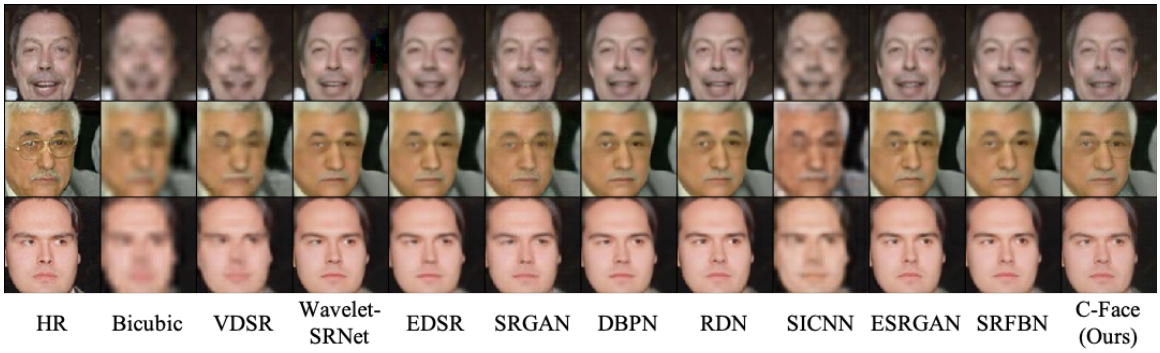
Models	Scale=2		Scale=4		Scale=8	
	PSNR (dB)	SSIM	PSNR (dB)	SSIM	PSNR (dB)	SSIM
Bicubic	36.5394	0.9596	29.9898	0.8659	24.4935	0.6563
VDSR	37.3361	0.9648	31.6599	0.8974	25.7980	0.7069
Wavelet-SRNet	37.8071	0.9669	33.2989	0.9232	27.9486	0.7989
EDSR	38.0262	0.9675	33.2594	0.9198	27.5588	0.7820
SRGAN	37.7331	0.9653	32.6060	0.9094	27.0180	0.7547
DBPN	38.1096	0.9679	33.5679	0.9236	28.0302	0.7976
RDN	**38.1444**	**0.9680**	**33.6732**	**0.9245**	**28.2989**	**0.8088**
SICNN	36.0144	0.9544	30.8252	0.8822	25.1120	0.6978
ESRGAN	37.5245	0.9645	33.4906	0.9233	28.0794	0.8001
SRFBN	38.0842	0.9679	33.2063	0.9199	27.6855	0.7888
C-Face (Ours)	37.8452	0.9665	32.9545	0.9132	27.5794	0.7792

PSNR and SSIM is inconsistent with human observation. These two metrics cannot measure the visual quality and the recognition accuracy of the reconstructed SR face images. By comparing the results of several methods in Table 1 and Table 2, we can see that whether the reconstruction results can help face recognition tasks is also independent of these two indicators.

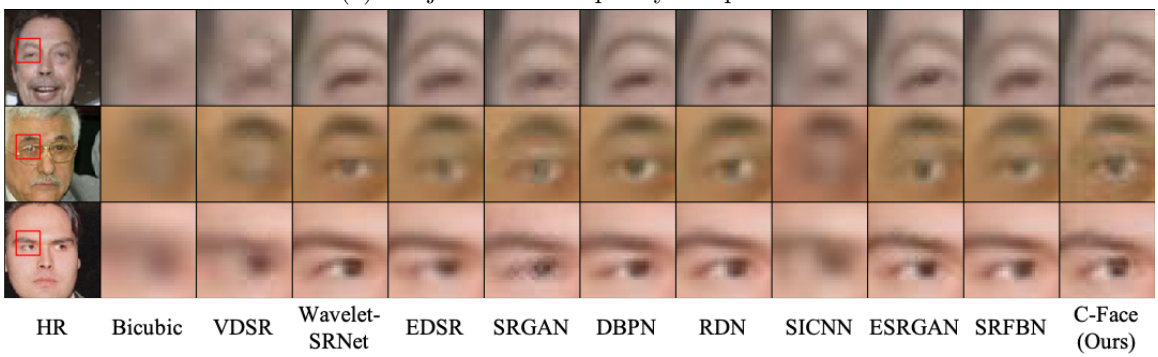
The qualitative comparisons of some generated images are shown in Fig. 7. Compared with deep learning methods, the effect of the traditional interpolation method is not ideal. The results of Wavelet-SRNet and SRGAN are relatively noisy compared with the other deep learning methods. The VDSR model can make the reconstructed results smoother but also lack more details. In contrast, our C-Face network can produce faithful and realistic facial identity details, such as eyes, mouth, and eyebrows. As shown in the second row of Fig. 7b, C-Face network also performs better on face images with occlusion (e.g., eyeglass) while other methods fail to recover the eyes with glasses.

4.3.2 BLUFR Protocol

The standard LFW protocol is not a difficult test for the accuracy of face hallucination. Even the traditional interpolation method can achieve a good accuracy and the difference between the results of each deep learning model is not obvious. With 3000 genuine and 3000 impostor matches for a model to recognize, the accuracy of the standard LFW protocol may be too optimistic for the reason that the protocol does not consider a low false accept rate (FAR) which is important in actual use. Therefore, we introduce a benchmark named LFW-BLUFR (Liao et al., 2014) which includes two test indicators: the verification rate (VR)



(a) Subjective visual quality comparison.



(b) Details of zoomed in regions.

Figure 7: Qualitative comparison on the LFW for 8× face super-resolution. Our C-Face network produces realistic details in identity facial components, such as eyes, mouth, and eyebrows.

at FAR = 0.1% for the face verification and the detection and identification rate (DIR) at Rank-1 corresponding to a FAR = 1% for open-set identification. The LFW-BLUFR is a tougher recognition test compared with the standard LFW protocol.

Table 3 shows the comparison results between the C-Face network and other methods on the LFW-BLUFR protocol. Our C-Face model significantly outperforms other models in both two indicators of the LFW-BLUFR. This result shows that our model has the best ability to maintain identity information in face hallucination. The C-Face network is able to recognize the low-resolution faces by reconstructing high-resolution outputs. Our C-Face network has a more noticeable improvement for the open-set identity test compared with the verification test on LFW-BLUFR.

The different effects under three different scaling factors (2×, 4×, and 8×) show that our proposed loss function as well as the training algorithm can be used as a general mechanism in face-hallucination. When the LR images contain enough identity information, the C-Face loss function will make the output result more recognizable.

Table 3: Comparison of verification rate (VR) at FAR = 0.1% for the face verification and the detection and identification rate (DIR) at Rank-1 corresponding to an FAR = 1% for open-set identification for various face hallucination models on LFW-BLUFR.

Models	Scale=2		Scale=4		Scale=8	
	VR	DIR	VR	DIR	VR	DIR
HR (128 × 128)	72.1141	36.2393	72.1141	36.2393	72.1141	36.2393
Bicubic	68.1243	32.6806	27.7407	3.8953	0.8822	0.3114
VDSR	70.0296	33.9750	46.2373	12.5788	1.0314	0.2968
Wavelet-SRNet	70.7067	35.4183	61.4134	25.3992	27.8256	9.3985
EDSR	70.9778	35.6552	58.5329	22.2884	19.6761	5.8050
SRGAN	71.1067	34.9254	57.7483	20.6052	18.9507	6.3029
DBPN	71.2264	36.2539	60.8500	23.8256	25.2759	9.5752
RDN	70.5374	35.0639	61.1352	24.2950	29.0899	9.1727
SICNN	69.8254	33.5369	42.4736	8.6124	1.0663	0.4661
ESRGAN	71.0270	36.1958	61.4914	25.0623	26.0105	9.3612
SRFBN	71.1204	35.8408	59.4563	22.1702	21.8458	8.3347
C-Face (Ours)	**71.2884**	**36.3369**	**63.2833**	**26.0716**	**30.8874**	**10.3348**

4.4 Evaluation on CelebA

In order to test the generalization ability, we also evaluate the hallucination models on CelebA faces dataset whose identities are not included in the training set. Following the standard protocol of LFW, we construct the 10-fold cross-validation scheme on the CelebA dataset. We define 3000 pairs of genuine comparisons and 3000 pairs of impostor comparisons, and then divide them into 10 disjoint subsets for cross validation, with each subset containing 300 genuine pairs and 300 impostor pairs. The quantitative results are shown in Table. 4. Our C-Face network achieves superior performance on all the three scaling factor (2×, 4×, and 8×), showing its significant generalization capability. C-Face network achieves the highest verification rate, indicating that our results retain better identity.

Not surprisingly, our C-Face network achieves lower PSNR and SSIM than general image super-resolution methods as shown in Table. 5. As mentioned before, this is mainly because PSNR and SSIM may be not good assessment metrics for the task of identity-preserving face super-resolution.

The qualitative comparisons are shown in Fig. 8. It can be observed that our proposed C-Face network outperforms previous methods in both global and local details. For instance, C-Face network recovers faithful details in mouth, teeth, and mustache, etc., while other methods tend to generate blurry results or incorporate unnatural noise. C-Face is capable of restoring accurate and reasonable mouth direction even the face images are accompanied

Table 4: Comparison of face verification rates evaluated by SphereFace for various face hallucination models on CelebA.

Models	Scale=2	Scale=4	Scale=8
HR (128×128)	0.9188	0.9188	0.9188
Bicubic	0.9137	0.8597	0.6203
VDSR	0.9177	0.8802	0.7313
Wavelet-SRNet	0.9188	0.9022	0.8438
EDSR	0.9178	0.8965	0.8097
SRGAN	0.9155	0.8947	0.8062
DBPN	0.9188	0.9018	0.8357
RDN	0.9197	0.8995	0.8477
SICNN	0.9178	0.8753	0.7070
ESRGAN	0.9172	0.8985	0.8343
SRFBN	0.9165	0.8970	0.8307
C-Face (Ours)	**0.9202**	**0.9043**	**0.8538**

by pose variations (e.g., the third row in Fig. 8b) while other methods introduce undesired textures or fail to generate enough details.

5. Ablation Study

In this section, we conduct an ablation study to estimate the effectiveness of our method. The experimental settings in Section 4.1 and 4.2 remain unchanged in the following ablation experiments.

5.1 The Choice of the Compare Images

For each input image I_o during the training, we will choose a compare image I_c for it to calculate the C-Face loss. The selecting rule of I_c for each I_o is described in Section 3.3 and contains a hyperparameter k. For previous experiments in Section 4, we set $k = 3$. However, to show that $k = 3$ is the reasonable choice for our model, we implement the experiments on $k = 1$ to $k = 5$ as well as randomly choosing an I_c from all the photos belonging to the same class with I_o in the dataset. Table 6 shows the results of each selection method for the C-Face. It can be observed that $k = 3$ or $k = 4$ brings the better performance of identity preserving and recognizability. If the k is too small, the compare image is too similar to the original HR image, and it is impossible to obtain additional identity information from the compare face images. On the contrary, if k is too large, the compare image will be very different from the original image, which is unable to capture the identity consistency between the compare face image I_c and the original HR face image I_o.

Table 5: Comparison of PSNR and SSIM of SR faces reconstructed by various face hallucination models on CelebA.

Method	Scale=2		Scale=4		Scale=8	
	PSNR (dB)	SSIM	PSNR (dB)	SSIM	PSNR (dB)	SSIM
Bicubic	34.2380	0.9449	28.4892	0.8407	23.7447	0.6410
VDSR	35.0572	0.9520	30.0858	0.8770	25.1269	0.6987
Wavelet-SRNet	35.7770	0.9564	31.7079	0.9073	27.2819	0.7918
EDSR	35.9918	0.9569	31.6853	0.9038	26.8454	0.7750
SRGAN	35.5009	0.9518	31.0906	0.8924	26.3385	0.7477
DBPN	**36.1076**	**0.9579**	31.9608	0.9078	27.3514	0.7917
RDN	36.0980	0.9579	**32.0567**	**0.9089**	**27.5843**	**0.8013**
SICNN	34.0152	0.9381	29.3030	0.8569	24.6323	0.6885
ESRGAN	35.1837	0.9502	31.8892	0.9075	27.3576	0.7924
SRFBN	36.0469	0.9576	31.6743	0.9043	27.0613	0.7839
C-Face (Ours)	35.5605	0.9540	31.3789	0.8961	26.8520	0.7691

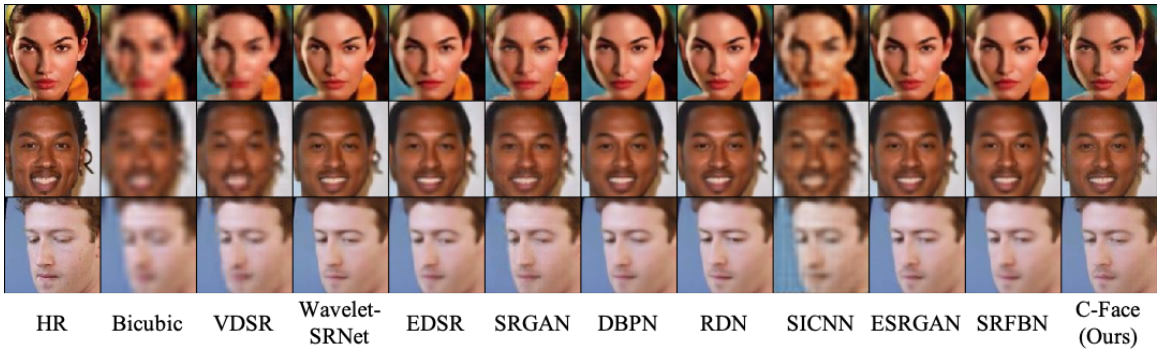
Table 6: Ablation study results on the choice of the compare face images.

C-Face	VR@Standard	VR@BLUFR	DIR	PSNR	SSIM
Random	0.9350	62.7250	24.6500	33.0463	0.9148
k=1	0.9380	62.7045	24.4298	32.3248	0.9040
k=2	0.9392	62.6514	25.6149	32.8596	0.9122
k=3	0.9403	63.2833	26.0716	32.9545	0.9132
k=4	0.9392	63.9404	27.0118	32.9347	0.9133
k=5	0.9380	63.2339	26.7296	33.0028	0.9140

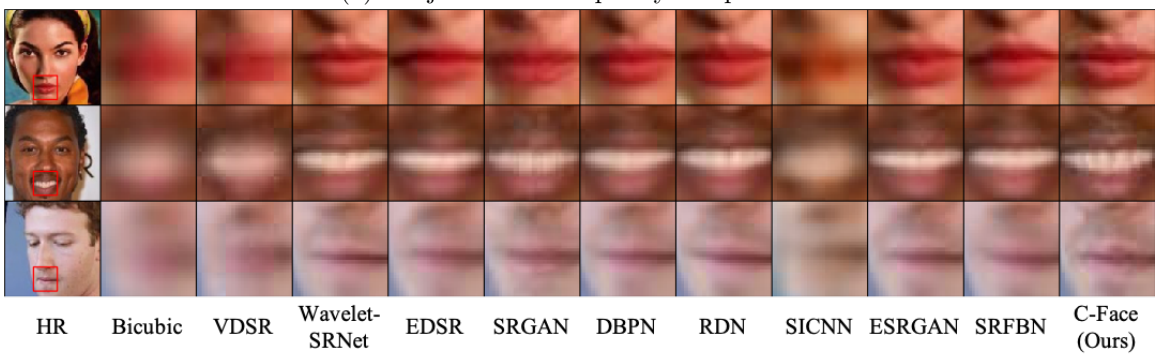
5.2 The Effectiveness of the Training Algorithm

In Section 3.4, we propose a novel training algorithm that will fine-tune the CNN^R model as well as our GAN-based model. In our training algorithm, we combine the reconstruction set $\mathbb{D}^{SR}$ and the original set $\mathbb{D}^{HR}$ and use the combined dataset to fine-tune the CNN^R. In order to demonstrate the superiority of our proposed training approach, we evaluate different training methods as follows:

- Version 1: Only use stage 1 of Algorithm 1 to train our model (C-Face-v1);
- Version 2: Use Algorithm 1 to train the model (C-Face-v2);



(a) Subjective visual quality comparison.



(b) Details of zoomed in regions.

Figure 8: Qualitative comparison on the CelebA for 8× face super-resolution. Our C-Face network produces realistic details in identity facial components, such as mouth, teeth, and mustache.

- Version 2-1: Only use $\mathbb{D}^{SR}$ to fine-tune the CNN^R in the stage 3 of Algorithm 1;
- Version 2-2: Train a CNN^R with $\mathbb{D}^{SR}$ from scratch in the stage 3 of Algorithm 1;
- Version 2-3: Do nothing in stage 3, use the original SphereFace in the stage 4 of Algorithm 1.

Table 7: Ablation study results on different training approaches.

Methods	VR@Standard	VR@BLUFR	DIR	PSNR	SSIM
version 1	0.9337	60.3485	22.1485	33.4452	0.9217
version 2	0.9403	63.2833	26.0716	32.9545	0.9132
version 2-1	0.9382	63.3349	24.7008	32.8952	0.9123
version 2-2	0.9352	61.6216	24.1057	33.0494	0.9137
version 2-3	0.9388	63.7920	24.8658	32.8045	0.9102

Table 7 shows the results of each method. As can be seen, the Version 2, which is also the C-Face-v2 model, has a better performance than the other training methods. Especially on the open-set identity test (DIR), our proposed training algorithm achieves the best performance. The reason that the C-Face-v2 model has a lower accuracy than version 2-1 and 2-3 on the LFW-BLUFR protocol is as follows. Directly using the C-Face loss to train the GAN model will make the identity features of images, which belong to the same person, coming closer to each other. This will be more helpful for face verification. In stage 3 and stage 4 of Algorithm 1, the fine-tuned CNN^R will make every I^{SR} find a better place in the feature space from each other, which will be helpful to the open-set identity test. However, those stages will not further decrease the distance between the images of the same person, so they do not help the verification test and may have a little disturbance to the original results.

6. Conclusion

In this paper, we propose a novel model named the C-Face network for face hallucination. Our GAN-based model can generate high-resolution face images while preserving identity information. A novel C-Face loss is employed to make sure that the output images belonging to the same person have similar identity features. We also propose a new training algorithm to overcome the convergence problem. Experimental results demonstrate that our method can significantly improve the face recognition/verification rate of SR face images. Our C-Face network achieves superior performance on both of the two test datasets that are not used for training, showing its remarkable generalization capability. In the future, we will simplify the training strategy which now has multiple stages. We also plan to apply the idea of our model to other similar tasks such as heterogeneous face recognition.

Acknowledgments

This work was supported in part by the National Key R&D Program of China under Grant 2021ZD0201300, and by the National Science Foundation of China under Grant 61876076.

References

Bulat, A., & Tzimiropoulos, G. (2018). Super-fan: Integrated facial landmark localization and super-resolution of real-world low resolution faces in arbitrary poses with gans. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 109–117.

Chen, Y., Tai, Y., Liu, X., Shen, C., & Yang, J. (2018). Fsrnet: End-to-end learning face super-resolution with facial priors. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2492–2501.

Deng, J., Guo, J., Xue, N., & Zafeiriou, S. (2019). Arcface: Additive angular margin loss for deep face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4690–4699.

Dong, C., Loy, C. C., He, K., & Tang, X. (2014). Learning a deep convolutional network for image super-resolution. In *Proceedings of the European Conference on Computer Vision*, pp. 184–199.

Farrugia, R. A., & Guillemot, C. (2017). Face hallucination using linear models of coupled sparse support. *IEEE Trans. Image Process.*, *26*(9), 4562–4577.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. C., & Bengio, Y. (2014). Generative adversarial nets. In *Advances in Neural Information Processing Systems*, pp. 2672–2680.

Grm, K., Scheirer, W. J., & Struc, V. (2020). Face hallucination using cascaded super-resolution and identity priors. *IEEE Trans. Image Process.*, *29*, 2150–2165.

Haris, M., Shakhnarovich, G., & Ukita, N. (2018). Deep back-projection networks for super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1664–1673.

Hsu, C., Lin, C., Su, W., & Cheung, G. (2019). Sigan: Siamese generative adversarial network for identity-preserving face hallucination. *IEEE Trans. Image Process.*, *28*(12), 6225–6236.

Huang, G. B., Mattar, M., Berg, T., & Learned-Miller, E. (2008). Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. In *Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition.*

Huang, H., He, R., Sun, Z., & Tan, T. (2017). Wavelet-srnet: A wavelet-based CNN for multi-scale face super resolution. In *IEEE International Conference on Computer Vision*, pp. 1698–1706.

Huang, H., He, R., Sun, Z., & Tan, T. (2019). Wavelet domain generative adversarial network for multi-scale face hallucination. *Int. J. Comput. Vis.*, *127*(6-7), 763–784.

Jiang, J., Chen, C., Ma, J., Wang, Z., Wang, Z., & Hu, R. (2017a). SRLSP: A face image super-resolution algorithm using smooth regression with local structure prior. *IEEE Trans. Multim.*, *19*(1), 27–40.

Jiang, J., Ma, J., Chen, C., Jiang, X., & Wang, Z. (2017b). Noise robust face image super-resolution through smooth sparse representation. *IEEE Trans. Cybern.*, *47*(11), 3991–4002.

Jiang, J., Wang, C., Liu, X., & Ma, J. (2021). Deep learning-based face super-resolution: A survey. *ACM Comput. Surv.*, *55*(1), 1–36.

Johnson, J., Alahi, A., & Fei-Fei, L. (2016). Perceptual losses for real-time style transfer and super-resolution. In *Proceedings of the European Conference on Computer Vision*, pp. 694–711.

Kim, J., Lee, J. K., & Lee, K. M. (2016a). Accurate image super-resolution using very deep convolutional networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1646–1654.

Kim, J., Lee, J. K., & Lee, K. M. (2016b). Deeply-recursive convolutional network for image super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1637–1645.

Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A. P., Tejani, A., Totz, J., Wang, Z., & Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 105–114.

Li, Z., Yang, J., Liu, Z., Yang, X., Jeon, G., & Wu, W. (2019). Feedback network for image super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3867–3876.

Liao, S., Lei, Z., Yi, D., & Li, S. Z. (2014). A benchmark study of large-scale unconstrained face recognition. In *IEEE International Joint Conference on Biometrics*, pp. 1–8.

Lim, B., Son, S., Kim, H., Nah, S., & Lee, K. M. (2017). Enhanced deep residual networks for single image super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1132–1140.

Liu, W., Wen, Y., Yu, Z., Li, M., Raj, B., & Song, L. (2017). Sphereface: Deep hypersphere embedding for face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6738–6746.

Liu, Z., Luo, P., Wang, X., & Tang, X. (2015). Deep learning face attributes in the wild. In *IEEE International Conference on Computer Vision*, pp. 3730–3738.

Mechrez, R., Talmi, I., Shama, F., & Zelnik-Manor, L. (2018). Maintaining natural image statistics with the contextual loss. In *Proceedings of the Asian Conference on Computer Vision*, pp. 427–443.

Schroff, F., Kalenichenko, D., & Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 815–823.

Taigman, Y., Yang, M., Ranzato, M., & Wolf, L. (2014). Deepface: Closing the gap to human-level performance in face verification. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1701–1708.

Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., & Loy, C. C. (2018). ESRGAN: enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision workshops*.

Yi, D., Lei, Z., Liao, S., & Li, S. Z. (2014). Learning face representation from scratch. *arXiv, 1411.7923*.

Yu, X., Fernando, B., Ghanem, B., Porikli, F., & Hartley, R. (2018a). Face super-resolution guided by facial component heatmaps. In *Proceedings of the European Conference on Computer Vision*, pp. 219–235.

Yu, X., Fernando, B., Hartley, R., & Porikli, F. (2018b). Super-resolving very low-resolution face images with supplementary attributes. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 908–917.

Yu, X., Fernando, B., Hartley, R., & Porikli, F. (2020). Semantic face hallucination: Super-resolving very low-resolution face images with supplementary attributes. *IEEE Trans. Pattern Anal. Mach. Intell.*, *42*(11), 2926–2943.

Yu, X., & Porikli, F. (2016). Ultra-resolving face images by discriminative generative networks. In *Proceedings of the European Conference on Computer Vision*, pp. 318–333.

Yu, X., & Porikli, F. (2017). Face hallucination with tiny unaligned images by transformative discriminative neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 4327–4333.

Yu, X., Porikli, F., Fernando, B., & Hartley, R. (2020). Hallucinating unaligned face images by multiscale transformative discriminative networks. *Int. J. Comput. Vis.*, *128*(2), 500–526.

Zangeneh, E., Rahmati, M., & Mohsenzadeh, Y. (2020). Low resolution face recognition using a two-branch deep convolutional neural network architecture. *Expert Syst. Appl.*, *139*, 112854.

Zhang, K., Zhang, Z., Cheng, C., Hsu, W. H., Qiao, Y., Liu, W., & Zhang, T. (2018a). Super-identity convolutional neural network for face hallucination. In *Proceedings of the European Conference on Computer Vision*, pp. 196–211.

Zhang, R., Isola, P., Efros, A. A., Shechtman, E., & Wang, O. (2018b). The unreasonable effectiveness of deep features as a perceptual metric. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 586–595.

Zhang, Y., Tsang, I., Luo, Y., Hu, C., Lu, X., & Yu, X. (2021). Recursive copy and paste gan: Face hallucination from shaded thumbnails. *IEEE Trans. Pattern Anal. Mach. Intell.*, *1*(1), 1–18.

Zhang, Y., Tsang, I. W., Luo, Y., Hu, C., Lu, X., & Yu, X. (2020). Copy and paste GAN: face hallucination from shaded thumbnails. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7353–7362.

Zhang, Y., Tian, Y., Kong, Y., Zhong, B., & Fu, Y. (2018). Residual dense network for image super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2472–2481.