International Journal of Computer Vision (2019) 127:785–800
https://doi.org/10.1007/s11263-019-01148-6

Joint Face Hallucination and Deblurring via Structure Generation and Detail Enhancement

Yibing Song[1] · Jiawei Zhang[2] · Lijun Gong[3] · Shengfeng He[4] · Linchao Bao[1] · Jinshan Pan[5] · Qingxiong Yang[6] · Ming-Hsuan Yang[7]

Received: 22 February 2018 / Accepted: 2 January 2019 / Published online: 17 January 2019

Abstract
We address the problem of restoring a high-resolution face image from a blurry low-resolution input. This problem is difficult as super-resolution and deblurring need to be tackled simultaneously. Moreover, existing algorithms cannot handle face images well as low-resolution face images do not have much texture which is especially critical for deblurring. In this paper, we propose an effective algorithm by utilizing the domain-specific knowledge of human faces to recover high-quality faces. We first propose a facial component guided deep Convolutional Neural Network (CNN) to restore a coarse face image, which is denoted as the base image where the facial component is automatically generated from the input face image. However, the CNN based method cannot handle image details well. We further develop a novel exemplar-based detail enhancement algorithm via facial component matching. Extensive experiments show that the proposed method outperforms the state-of-the-art algorithms both quantitatively and qualitatively.

Keywords Face hallucination · Face deblurring · Convolutional Neural Network

1 Introduction

Human faces captured in the real world usually suffer from the imaging process. The large distance between human faces and camera leads to limited pixel sampling on the camera sensor. Meanwhile, relative movement during exposure brings blur to the digital images. As a result, the captured faces are usually in small resolution and contain moderate blur. As camera viewpoints cannot be changed frequently in some cases (e.g., video surveillance scenario), there is a need to restore the face images for further analysis. However, existing algorithms designed for image super-resolution or image deblurring cannot handle this problem well due to the influence of both resolution and blur (Fig. 1). As human faces contain rich details around facial components, i.e., eyes, mouth, etc., it is of great interest to develop an effective algorithm to estimate clear high-resolution (HR) face images by the domain-specific knowledge of faces.

Communicated by Chellappa, Liu, Kim, Torre and Loy.

✉ Jinshan Pan
sdluran@gmail.com

Yibing Song
dynamicstevenson@gmail.com

Jiawei Zhang
zhjw1988@gmail.com

Lijun Gong
gljvivi@gmail.com

Shengfeng He
shengfenghe7@gmail.com

Linchao Bao
linchaobao@gmail.com

Qingxiong Yang
liiton.research@gmail.com

Ming-Hsuan Yang
mhyang@ucmerced.edu

1 Tencent AI Lab, Shenzhen, China

2 Sensetime Research, Shenzhen, China

3 Tencent, Shenzhen, China

4 South China University of Technology, Guangzhou, China

5 Nanjing University of Science and Technology, Nanjing, China

6 University of Science and Technology of China, Hefei, China

7 University of California at Merced, Merced, USA

(a) Input (Bic) **(b)** SFH [57] **(c)** DFE [38] **(d)** RBF [56] **(e)** Ours

Fig. 1 Joint face hallucination and deblurring. The input blurry LR image (with bicubic upsampling) is shown in (**a**). The results generated by existing face hallucination and deblurring methods are shown from (**b**) to (**d**). Different from existing methods which treat face hallucination and deblurring separately, we formulate these two tasks into a single framework for joint prediction. Our method effectively recovers the facial components as shown in (**e**) and performs favorably against the state-of-the-art methods

When handling the restoration problem of face images, the state-of-the-art face hallucination (FH) methods usually learn the mapping function from low-resolution (LR) images to high-resolution HR images either in a regression (Liu et al. 2007; Jia and Gong 2008; Tappen and Liu 2012) or an exemplar-based (Ma et al. 2010; Jia and Gong 2006, 2005) way. Although existing FH algorithms achieve great progress, these algorithms usually assume the blur is intrinsic (e.g., bicubic blur, Gaussian blur). Meanwhile, they are less effective when the inputs contain heavy motion blur. Several algorithms (Pan et al. 2014) have been proposed to deal with blurry face images. However, these algorithms usually assume the high resolution of the blurry input. If the resolution of the blurry input faces is low, these algorithms are not able to generate reliable data for blur kernel estimation (Pan et al. 2014). To solve LR blurry images, recent methods (Park and Mu Lee 2017) jointly estimate image super-resolution and image deblurring. Xu et al. (2017) use Generative Adversarial Nets (GANs) to super-resolve face images. However, without exploiting the unique structures of human faces, these methods are not able to handle face image restoration problem well, especially around facial components. Figure 1 shows an example where the input is a low-resolution face image. Both the state-of-the-art FH algorithm (Yang et al. 2013) and face deblurring (FD) algorithm (Pan et al. 2014) cannot effectively recover clear HR images. Meanwhile, the deep learning based method (Xu et al. 2017) does not effectively recover the detailed structure and reduce blur.

In this paper, we propose a unified framework to super-resolve face images. As this is an ill-posed problem and most information is missing, recovering a face image with both general facial structure and local details using one CNN framework without any domain knowledge is challenging. We formulate the restored image using as a base layer and a detail layer. The base layer is learned by using a CNN guided by facial components. The detail layer is generated by an exemplar-based texture synthesis module. First, our facial structure generation network (FSGN) takes the up-sampled face image and its facial components as the inputs and generates base images. Then we use a patch-wise K-Nearest Neighbor (K-NN) to search between the intermediate face image and exemplar images. In this way, we can accurately establish the correspondences on the HR training images and overcome the limitations of existing feature matching-based methods. The accurate correspondence ensures that the fine-grained facial structures from the HR exemplar images are effectively extracted. Finally, the details from these structures are transferred into the base image through edge-aware image filtering. Figure 1e shows that our algorithm is able to super-resolve blurry face images and generates the face image with much more clear textures. The contributions of this work are summarized as follows:

- We propose a unified framework for joint face hallucination and deblurring by using the special properties of face images. To generate high-quality faces, we develop a face component guided CNN.
- We develop a novel exemplar-based detail transfer algorithm to improve the details and texture estimations of face images.
- We analyze the properties of the proposed algorithm and show that it performs favorably against state-of-the-art face hallucination and deblurring methods on the public benchmarks.

2 Related Work

Face hallucination and deblurring relate closely to generic image super-resolution and deblurring. In this section, we perform a literature review on the most related work of face

hallucination, image super-resolution, and face deblurring and put this work in proper context.

2.1 Face Hallucination

Learning based methods are widely adopted in face analysis approaches including face hallucination (Wang et al. 2014; Song et al. 2014, 2017b) and style transfer (Song et al. 2017a). Face hallucination methods can be categorized as the data-driven framework and the CNN generative framework. In the data-driven framework, various approaches are proposed to learn the transformation between LR and HR to recover the missing details from the input. In Gunturk et al. (2003) and Wang and Tang (2005), generalized approaches on the eigen domain are proposed to map both LR and HR image spaces. The tensor-based methods are proposed by Liu et al. (2015) and Jia and Gong (2008) to well hallucinate multiple model face images across different poses and expressions. In Liu et al. (2007), Principle Component Analysis (PCA) based linear constraints are learned from the training images and a patch-based Markov Random Field (MRF) is used to reconstruct the residues. It can only work on fixed poses and expressions. In Jin and Bouganis (2015), the blurring kernel and transformation of LR faces are jointly estimated by deblurring and registration in PCA subspace. It only works for face region instead of the whole face image. Face hallucination in the compressed domain is proposed in Liu and Yang (2014) and Yang et al. (2018). Image alignment-based methods are adopted for face hallucination where HR face images are matched to LR face images by dense SIFT flow (Tappen and Liu 2012) or feature matching (Yang et al. 2013; Song et al. 2017c). The quality of output HR results depends on image alignment which is less effective when poses and expressions are different between training and input images.

On the other hand, the CNN generative framework predicts HR face images in an end-to-end manner. In Zhou et al. (2015), a Bi-channel CNN is proposed to integrate the input image and face representation for prediction. In Yu and Porikli (2016) and Karras et al. (2018), the GAN framework is applied to hallucinate LR face images. However, the network generates high-resolution images from random noise in Karras et al. (2018) while the face hallucination task is to tackle a specific input image. The transformative discriminative auto-encoders are proposed in Yu and Porikli (2017b), Chen et al. (2017) and Jourabloo et al. (2017) to upsample images and denoise simultaneously during the hallucination. In Yu and Porikli (2017a), a spatial alignment network is proposed for LR and HR matching. A cascaded bi-network is proposed in Zhu et al. (2016) for FH and deep reinforcement learning is applied in Cao et al. (2017) to achieve attention awareness. The CNN generative framework usually handles input face images in an extremely low resolution where the facial components are not able to be distinguished. Even though these methods generate facial structures on the output HR result, these structures are not accurate and lead to incorrect identity. In contrast, our method combines the advantage of both CNN generative and data-driven framework for identity preservations.

2.2 Image Super Resolution

The advancement of CNN has activated a series of investigations on image SR. Starting from SRCNN (Dong et al. 2014, 2016a) where HR images are predicted in an end-to-end manner via several convolutional layers and nonlinear activations, following works are proposed to combine existing models with CNN (Yang et al. 2018b) or improve network capacity. In Wang et al. (2015), a sparse coding model is designed for SR and incarnated as a neural network, which is trained end-to-end in a cascade structure. A convolutional sparse coding method is proposed in Gu et al. (2015) to enforce the pixel consistency during image reconstruction. It exploits the image global correlation to produce a more robust reconstruction of image local structures. In Shi et al. (2016), an efficient sub-pixel convolutional layer is proposed to learn an array of upscaling filters for SR. It will replace the handcrafted upscaling filters for each CNN feature map specifically. A deeply-recursive convolutional network is proposed in Kim et al. (2016b) to involve recursion, whose depth can improve performance without introducing new parameters for additional convolutions. In Kim et al. (2016a), a very deep convolutional network is designed to increase network depth and residual learning is involved to facilitate the training process. An accelerated version of SRCNN is proposed in Dong et al. (2016b) to achieve real-time performance for practical usage. In Johnson et al. (2016), a perceptual loss is introduced for training feed-forward networks in image SR. The generative adversarial network (GAN) is applied to image SR in Ledig et al. (2017) to achieve perceptually satisfaction. In Lai et al. (2017), a deep Laplacian pyramid network is proposed to upsample LR images gradually via deconvolution. Meanwhile, residual learning (Tai et al. 2017) is involved to facilitate the training process. Anchored regression network is proposed in Agustsson et al. (2017) to design a smoothed relaxation of a piecewise linear regressor through the combination of multiple linear regressors over soft assignments to anchor points. In Sajjadi et al. (2017), texture synthesis is proposed in combination with perceptual loss focusing on creating realistic textures rather than optimizing pixel-wise loss function. In Timofte et al. (2017), it has been shown that the method by Lim et al. (2017) performs well against the sate-of-the-art SR algorithms. Different from existing image SR methods, our algorithm is specifically designed for FH and focus on facial structure generation and enhancement.

2.3 Face Deblurring

Most existing deblurring algorithms (Pan et al. 2017a, b; Zhang et al. 2018) focus on the generic image deblurring. As blurry face images contain fewer textures, the generic deblurring algorithms cannot handle this problem well. In Hacohen et al. (2013), a non-rigid dense correspondence is established and blur kernel estimation is performed accordingly. Facial structures are exploited in Pan et al. (2014) to maximum a posteriori deblurring algorithm on an exemplar dataset. These data-driven methods are able to handle blurry images when sharp edges are extracted effectively while requires querying time cost. The deep learning algorithm has also been applied to face deblurring. Xu et al. (2017) use GAN to super-resolve face images. It predicts face images via extremely low-resolution inputs where the facial identity is hard to identify. However, as this method does not consider the specialty of face images, it is less effective to restore some key component of faces. Recently, Shen et al. (2018) propose a face deblurring method which uses several CNNs to generate semantic face labels for guiding the face deblurring process. In this work, we solve the hallucination and deblurring in a unified framework. We note that Shen et al. (2018) use computationally expensive face parsing CNNs to generate pixel-wise semantic face labels. In addition, the prediction accuracy decreases when the input resolution is low, and thus affects the following process. In contrast, we use a computationally efficient facial landmark detector to estimate the facial components to guide the FSGN. Our experimental results show that the proposed facial landmark performs robustly against when the landmarks are not precisely detected. Shen et al. (2018) use two cascaded CNNs to reduce the blurry effect on the face images while we use the FSGN to generate a base image containing the general facial structure and an exemplar-based texture synthesis framework to restore details.

3 Proposed Method

Figure 2 shows the pipeline of our method. It mainly consists of two modules. The first module is a very deep CNN, namely Facial Structure Generation Network, which predicts a base image given the LR input. The base image contains the basic structure of the input face while the facial details are not fully recovered. It is then fed into the second module for detail enhancement. Note that the second module of our method relies on establishing correspondences from the base image to high-resolution exemplar images, which benefits from the first module since major structures of the input face are roughly recovered by the FSGN and thus the establishment of LR-HR correspondences are much easier. We describe the details of the two modules in the following:

3.1 Facial Structure Generation Network

As the unique structure of human faces differs much from the natural images (i.e., textures mostly reside around the facial components), we expect our FSGN to focus on the facial components rather than the remaining regions which are typically flat and less informative. Given an input LR blurry face image, we first upsample it using bicubic interpolation to the same resolution of the output. Then facial landmarks are detected on the upsampled image using the method from Zhu

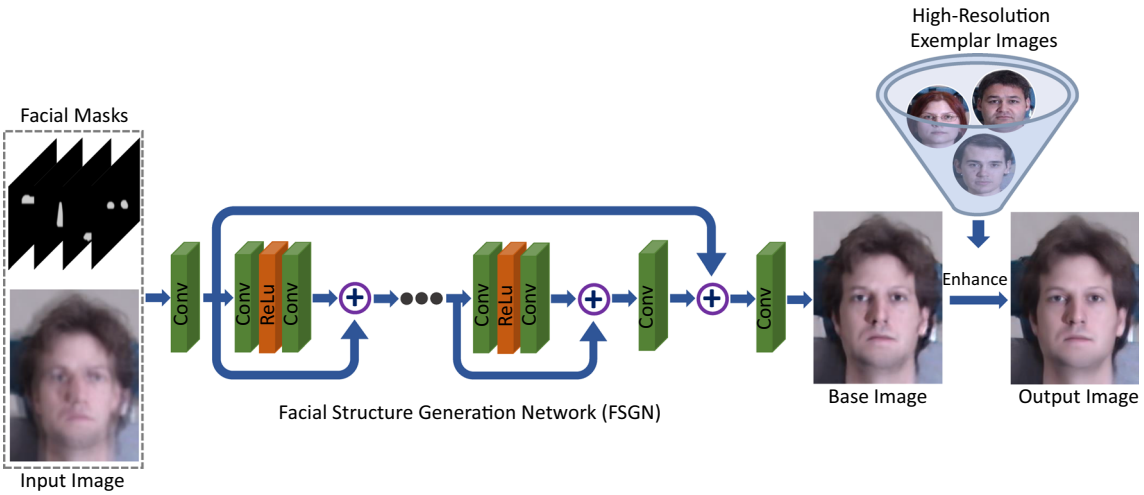


Fig. 2 Pipeline of the proposed method. Given a blurry LR input image, we first upsample it via bicubic interpolation and obtain the facial landmarks for generating facial components. Then we develop a facial structure generation network to generate the base image. Finally, we develop a details enhancement algorithm to estimate the missing details in the base image by HR exemplar images

Table 1 Network parameters of the proposed facial structure generation network

Layer name	FSGN
Conv 1	3×3, 64, pad 1
Conv 2x	$\begin{bmatrix} 3 \times 3, & 64, & \text{pad } 5, & \text{dilation } 5 \\ 3 \times 3, & 64, & \text{pad } 5 \end{bmatrix} \times 5$
Conv 3x	$\begin{bmatrix} 3 \times 3, & 64, & \text{pad } 4, & \text{dilation } 4 \\ 3 \times 3, & 64, & \text{pad } 4 \end{bmatrix} \times 5$
Conv 4x	$\begin{bmatrix} 3 \times 3, & 64, & \text{pad } 3, & \text{dilation } 3 \\ 3 \times 3, & 64, & \text{pad } 3 \end{bmatrix} \times 5$
Conv 5x	$\begin{bmatrix} 3 \times 3, & 64, & \text{pad } 2, & \text{dilation } 2 \\ 3 \times 3, & 64, & \text{pad } 2 \end{bmatrix} \times 5$
Conv 6x	$\begin{bmatrix} 3 \times 3, & 64, & \text{pad } 1, & \text{dilation } 1 \\ 3 \times 3, & 64, & \text{pad } 1 \end{bmatrix} \times 5$
Conv 7	3×3, 64, pad 1
Conv 8	3×3, 64, pad 1

and Ramanan (2012). Facial component masks are then generated using the landmarks. As the input LR images contain blurry pixels, the facial masks may not accurately localize the facial components. Nevertheless, our FSGN performs well when the facial masks cannot be accurately extracted. Following (Yang et al. 2013), we categorize facial components into four types, which are eyebrows, eyes, noses, and mouths, respectively. For each type of facial component, we prepare a mask where the pixels within the component region are marked as 1 and the others as 0. In total, we generate four masks covering all four types of facial components accordingly. Figure 2 shows a direct view of these masks.

3.1.1 Network Architecture

The FSGN consists of 25 residual blocks (He et al. 2016) with 53 convolution layers in total. In addition to the local skip connections in each residue block, we add an additional long-range skip connection from the first convolution layer to the last convolution layer. The network is fully convolutional and we use a dilated 3×3 convolution (Yu et al. 2017) in all the layers. There is no pooling layer in our network and the size of all intermediate feature maps is the same as the input image. The detailed network architecture is shown in Table 1. During training, we create low-resolution blurry images from the ground-truth high-resolution images for input. Meanwhile, these corresponding ground-truth images are used for supervision with the Euclidean loss.

3.2 Detail Enhancement

Although the proposed FSGN is able to restore an image in which major facial components can be effectively recovered, it tends to over-smooth the details of recovered face images (as shown in Fig. 3b). To solve this problem, we propose a detail enhancement method to estimate the missing details by using high-resolution exemplar images. Our detail enhancement method consists of two steps. In the first step, we establish the patch correspondences between the base image generated by FSGN and HR exemplar images, then we use the patches from HR exemplar images to regress the base image and get an intermediate result. In the second step, the details from the intermediate result are transferred to the base image via edge-preserving filtering to obtain the final result. Figure 3 illustrates the effectiveness of our detail enhancement method. The details are presented in the following sections.

3.2.1 Exemplar Regression

Given the base image produced by FSGN, we divide it into local patches. For one patch centered on pixel p, we perform a K nearest neighbor search (K-NN) in the HR exemplar images to find the K most similar patches. Note that the HR exemplar images are face images where the subjects are different from that in the input image. In the K-NN patch search step, we choose a search region in one HR exemplar image for each input patch. The center of each search region is the same as that of the input patch. In the search region, we use a sliding window to select one patch, which is the same size as the input patch. We obtain N patches from N training HR images after patch search and further select K among them.

$$D_p = \alpha \cdot (1 - D_{ncc}) + (1 - \alpha) \cdot D_{abs}, \tag{1}$$

where α is the weight combining the two metrics. It is set as 0.5 in our implementation. We normalize image pixel value to [0, 1] in order to set the two metrics into the same range.

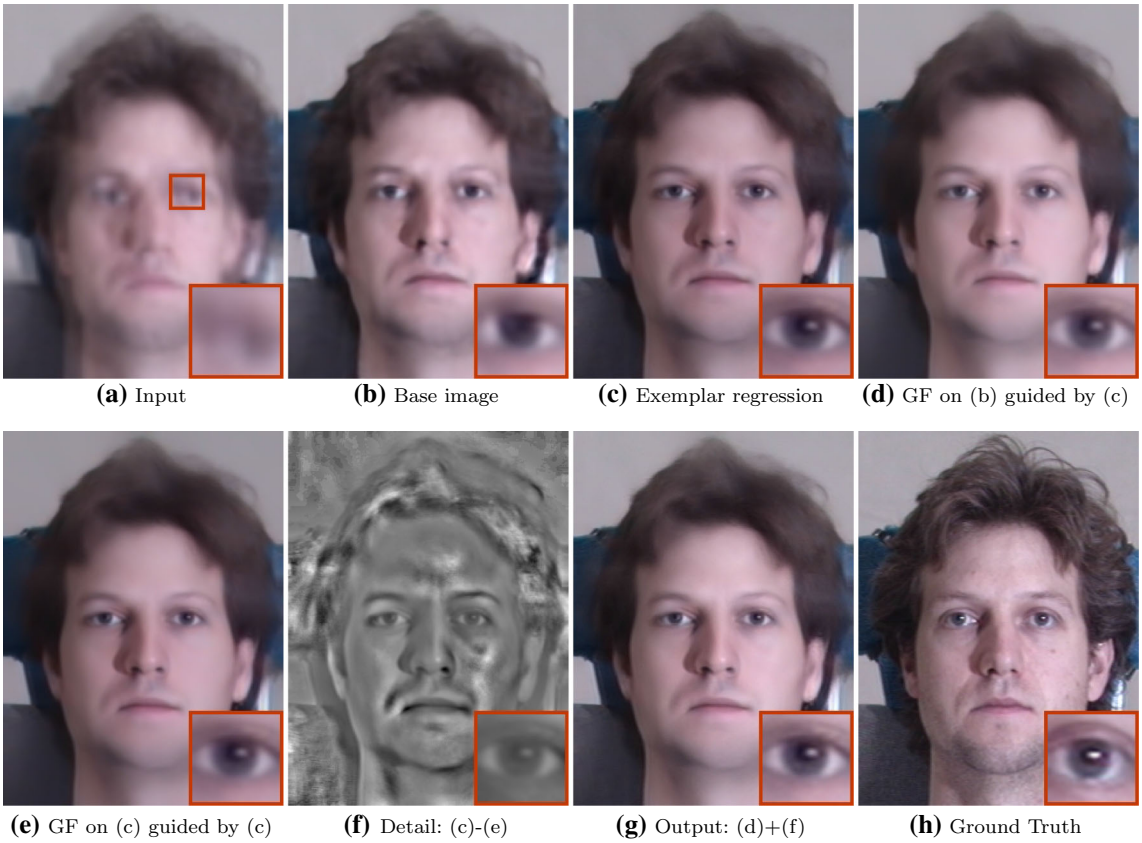
After K-NN search we select K candidate patches from HR exemplar images. Let H_p^i ($i \in [1, \ldots, K]$) denote a vector containing all the pixel values of the ith HR candidate patch, and I_p denote a vector containing the pixel values of the input patch. We also denote the linear regression function as $\mathcal{F}_p = [F_p^1, \ldots, F_p^K]^{\mathrm{T}}$ where F_p^i ($i \in [1, \ldots, K]$) is each coefficient of $\mathcal{F}_p$. The energy function is defined as:

$$E_p^{\text{data}} = ||\mathrm{H}_p \cdot \mathcal{F}_p - \mathrm{I}_p||^2, \tag{2}$$

where $\mathrm{H}_p = [\mathrm{H}_p^1, \mathrm{H}_p^2, \ldots, \mathrm{H}_p^K]$. It is a linear regression form and we can compute $\mathcal{F}_p$ as

$$\mathcal{F}_p = (\mathrm{H}_p^{\mathrm{T}} \cdot \mathrm{H}_p)^{-1} \mathrm{H}_p^{\mathrm{T}} \cdot \mathrm{I}_p. \tag{3}$$

We can efficiently compute $\mathcal{F}_p$ when the patches contain texture (i.e., the pixel values in H_p should not be similar to each other). However, in some cases when p is in the smooth region (e.g., cheek) $\mathrm{H}_p^{\mathrm{T}} \cdot \mathrm{H}_p$ may become a singular matrix and



(a) Input **(b)** Base image **(c)** Exemplar regression **(d)** GF on (b) guided by (c)

(e) GF on (c) guided by (c) **(f)** Detail: (c)-(e) **(g)** Output: (d)+(f) **(h)** Ground Truth

Fig. 3 The process of structure enhancement. The input LR blurry face image is shown in (**a**). The base image generated by FSGN is shown in (**b**). The exemplar regression result is shown in (**c**). We perform guided filtering on (**b**) using (**c**) as guidance to get (**d**). We also filter (**c**) using guided filtering in (**e**). The lost structure details after filtering are shown in (**f**), which is the difference between (**c**) and (**e**). We add the details back to (**d**) to generate the output as shown in (**g**). The ground truth image is shown in (**h**)

thus $\mathcal{F}_p$ is not accurate. We resolve the problem by adding a regularization term as:

$$E_p = E_p^{\text{data}} + E_p^{reg} = ||\mathrm{H}_p \cdot \mathcal{F}_p - \mathrm{I}_p||^2 + \lambda||\mathcal{F}_p||^2, \quad (4)$$

where λ is the weight controlling the influence of regularization term. It is set as the number of pixels in an input patch. We can solve the above energy function as:

$$\mathcal{F}_p = (\mathrm{H}_p^{\mathrm{T}} \cdot \mathrm{H}_p + \lambda \mathbb{1})^{-1} \mathrm{H}_p^{\mathrm{T}} \cdot \mathrm{I}_p, \quad (5)$$

where $\mathbb{1}$ is the identity matrix.

Once we calculate the regression function $\mathcal{F}_p$, we map the HR exemplar patches into the output patch. Let $\bar{\mathrm{H}}_p^i$ ($i \in [1, \ldots, K]$) denote one vector containing the pixel values of the corresponding HR exemplar patches. The output patch R_p can be computed as:

$$\mathrm{R}_p = \sum_{i=1}^{K} \bar{\mathrm{H}}_p^i \cdot F_p^i. \quad (6)$$

We compute the output patch for each pixel. For the overlapping areas between different patches, we perform averaging to generate the final regression result.

3.2.2 Detail Transfer

The regression result contains detailed structures transferred from HR exemplar images. However, it cannot be directly

Algorithm 1 Overview of proposed algorithm

– *Training* –
Train FSGN using face images and masks;
– *Testing (input LR blurry image* $\mathbb{I}$*)* –
Generate base image $\bar{\mathbb{I}}$ using FSGN;
for each pixel p in $\bar{\mathbb{I}}$ **do**
  K-NN patch search using equation 1;
  Calculate regression matrix using equation 5;
  Perform regression R_p using equation 6;
end for
Guided filtering on $\bar{\mathbb{I}}$ using regression image $\mathbb{R}$ to obtain $\bar{\mathbb{I}}^{\mathbb{R}}$;
Guided filtering on $\mathbb{R}$ using regression image $\mathbb{R}$ to obtain $\mathbb{R}^{\mathbb{R}}$;
Compute the output image by $\bar{\mathbb{I}}^{\mathbb{R}} + \mathbb{R} - \mathbb{R}^{\mathbb{R}}$.

adopted as the output. This is because the detailed structures are transferred from exemplar patches which belong to different subjects. The lighting condition of each subject is different from each other, which results in different shading appearances in facial regions between the regressed image and the ground truth (e.g., Fig. 3c, h). We here present an algorithm based on joint edge-preserving filtering (Petschnigg et al. 2004; Eisemann and Durand 2004) to combine the low-frequency appearances of the base image and the high-frequency facial details of the regressed image to generate the final output.

The main steps of our algorithm are shown in Fig. 3. We have a base image shown in (b) and the regressed image shown in (c). We use guided filter (He et al. 2010, 2013) to smooth (b) using (c) as guidance. As such, the facial details of (c) can be transferred into (b). However, the filtered result is likely to be over-smoothed (as shown in Fig. 3d). Nevertheless, we can further extract details from (c) and add them to the filtered result. Specifically, we smooth (c) using guided filtering with itself as guidance shown in (e). Then the smoothed details can be captured by subtracting the smoothed image (e) from (c), as shown in (f). Finally, we add (f) to (d) to get the output image shown in (g). Note that both global appearances and facial details of the output image are similar to the ground truth shown in (h). The pseudo code of our entire algorithm is shown in Algorithm 1.

Our detail enhancement method improves the base image quality by adding identity-specific details. Figure 3 shows that in (b) only the general facial structure is recovered in the base image while the details are still missing. We use the exemplar regression to synthesize the details specifically for the input image. The details are then extracted and transferred to the base image using the guided filter shown in (g). Compared with the base image, our detail enhancement method enriches the local details around facial components while the artifacts are not involved.

4 Experimental Results

We conduct experiments on the Multi-PIE (Gross et al. 2010) and PubFig (Kumar et al. 2009) datasets. The face images in the Multi-PIE dataset are taken in the lab controlled environment while the face images in the PubFig dataset are taken in the real world condition. The resolution of the ground truth images in these two datasets is 320×240. We evaluate our method from two aspects. First, we conduct an ablation study to illustrate the effectiveness of our modules. Second, we compare our method with the state-of-the-art FH methods including FHTP (Liu et al. 2007), SFH (Yang et al. 2013), five image SR methods including bicubic interpolation, SRCSC (Gu et al. 2015), SRCNN (Dong et al. 2016a), VDSR (Kim et al. 2016a), SRResNet (Ledig et al. 2017), and two face deblurring methods DFE (Pan et al. 2014), RBF (Xu et al. 2017). We use PSNR and SSIM (Wang et al. 2004) to quantitatively measure the image quality of the generated results.

Training data and configurations We follow the same setting with that in SFH (Yang et al. 2013) and use 2184 images from the Multi-PIE dataset as training data. To create the input LR blurry images, we first convolve with the ground truth images using random blur kernel and downsample the convolved results. The blur kernel size ranges randomly from 11 to 31, and the Gaussian variance ranges randomly from 1.4 to 1.7. In total, we have generated 200 motion blur kernels without noise and randomly select one to convolve with the ground truth images. The scaling factor is set to 4. To create the input LR images without blur, we convolve with the ground truth images using Gaussian blur kernel and downsample to generate the training inputs. We train our network from scratch using the ADAM solver (Kingma and Ba 2014) with a learning rate of 1e−4. For performance evaluation against the state-of-the-art methods, we follow the original network architecture designs and train them from scratch. The training data is the same as ours and we follow their training configurations to reproduce the results for comparison.

Test data There are 342 test images from the Multi-PIE dataset and 400 images from the PubFig dataset, respectively. We create input images through convolving with test images via random blur kernel and Gaussian kernel. The test images are generated in the same way as training image inputs. Note that there is no identity overlap between the training and test images.

4.1 Ablation Studies

In our detail enhancement step, we use K-NN search on the HR training images to establish the correspondence for exemplar regression. As there exist several parameters, e.g., patch size and the number of candidate patches, in K-NN search, we analyze the effect of these parameters and show how they affect the proposed algorithm. Tables 2 and 3 show the evaluation results. In the detail enhancement step, we set different patch sizes ranging from 10×10 to 30×30 incremented by 5. Table 2 shows that the proposed method performs well when the patch size is 20×20. Table 3 shows the effect of the proposed method with different candidate patch numbers. The proposed method performs well when the patch number is 5.

]

We note that the proposed algorithm still performs well when the facial masks do not align the ground truths. In the deblurring task, we note that the facial masks may deviate

Table 2 Experimental results using different patch sizes during K-NN search on the Multi-PIE dataset

Patch size	PSNR SR/Deblur	SSIM SR/Deblur
10×10	34.14/24.62	0.91/0.83
15×15	34.52/24.87	0.91/0.84
20×20	**34.93/25.75**	**0.92/0.86**
25×25	34.65/25.43	0.92/0.85
30×30	34.42/25.21	0.91/0.85

Bold values indicate the best performance

Table 3 Experimental results using different patch numbers during K-NN search on the Multi-PIE dataset

Patch number	PSNR SR/Deblur	SSIM SR/Deblur
3	34.05/24.97	0.87/0.83
4	34.68/25.34	0.90/0.85
5	**34.93/25.75**	**0.92/0.86**
6	34.91/25.74	0.92/0.86
7	34.92/25.75	0.92/0.86

Bold values indicate the best performance

Table 4 Experimental results with average facial mask deviation on the Multi-PIE dataset

Mask deviation	0	2	4	6	8	10
PSNR	25.54	25.52	25.33	25.10	24.53	24.12
SSIM	0.85	0.85	0.84	0.83	0.81	0.78

The deviation extent is measured with pixels

according to the LR blurry input images. The deviation of the facial masks correlates with the blur kernel size. To analyze how the deviation of the facial masks affect the output result, we use different blur kernels to generate different sets of input images shown in Table 4. For each set , we generate the corresponding facial masks and compute their average deviations with the ground truth masks. The deviation is measured in pixels. Then we generate the output results for each input image and quantitatively evaluate their performance in Table 4. It shows that the performance gradually decreases as there are more deviations of the facial masks. Meanwhile, we observe that there is an obvious degradation of the output quality when the deviation exceeds 6 pixels. Figure 4 shows a visual example of the mask deviation. We use different blur kernel size to produce several input images and generate our results accordingly. The results indicate that when the blur kernel exceeds 31 (i.e., the mask deviation is above 6 pixels) the artifacts occur on the output images. To ensure the facial masks effective for the input images, we set the blur kernel size below 31 pixels to generate the output result.

Our method consists of several modules. Our input is the LR blurry image with four facial masks. Our network structure is FSGN with dilation integration and the details are enriched through detail enhancement step. In this section, we conduct internal analysis to evaluate the performance gain through integrating each module. Our evaluation is conducted on the MultiPIE dataset where we follow the training and evaluation strategies illustrated above. We start to train from scratch using several convolutional layers without a long-range skip connection and empirically find that it does not converge in practice. Inspired by the VDSR (Kim et al. 2016a) method where the output is the combination of the input image and the last layer output, we design a similar structure which is the baseline (denoted as B) of our method. It contains all the 53 convolutional layers and nonlinear activations with a long-range skip connection while the local skip connections are removed. In addition to the baseline configuration, we integrate local residual blocks (i.e., short-range skip connections) to see the performance gain. It is denoted as BL in this study. Note that in this two configurations, we only take the LR blurry face images as input to train the network. In order to evaluate the effectiveness of facial masks, we add them as the input together with the input image. This configuration is denoted as BL + M. Then, we follow the same configuration as BL + M and integrate the dilation into the network, which is denoted as BLD + M. Finally, we add our detail enhancement module and denote it

(a) Mask deviation 0 **(b)** Mask deviation 6 **(c)** Mask deviation 8 **(d)** Mask deviation 10 **(e)** Ground Truth

Fig. 4 Restoration results using different facial mask deviations. The restoration result generated using ground truth facial mask is shown in (**a**). The restoration results generated by less accurate facial masks are shown from (**b**–**d**). The ground truth image is shown in (**e**)

Table 5 Ablation studies on the Multi-PIE dataset with predefined five configurations

	PSNR SR/Deblur	SSIM SR/Deblur
B	33.90/23.97	0.87/0.81
BL	33.95/24.02	0.88/0.82
BL + M	34.22/24.83	0.89/0.84
BLD + M	34.63/25.31	0.91/0.85
BLD + M + DE	**34.93/25.75**	**0.92/0.86**

Bold values indicate the best performance
We denote baseline as B, baseline with local residual blocks as BL, facial masks input as BL + M, dilation integration as BLD + M, and detail enhancement integration as BLD + M + DE

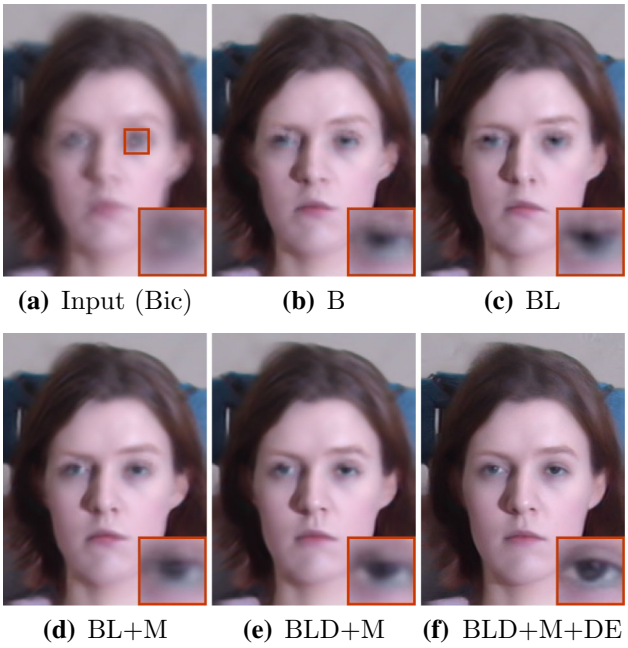


Fig. 5 Visualization of the ablation studies. **a** Bicubic upscaled input LR blurry image. **b** Baseline network performance and **c** local residual blocks integration on (**b**). In (**d**), we retrain (**c**) using facial masks and generate the result. Meanwhile, we involve dilation in (**d**) and generate the result shown in (**e**). The detail enhancement on (**e**) is shown in (**f**)

as BLD + M + DE. These configurations indicate how facial masks, local residual blocks, dilation and detail enhancement affect the image quality of the output results. Moreover, we evaluate the effectiveness of each module when the input is an LR image with and without random motion blur, independently. It corresponds to how our method handles both hallucination and deblurring tasks.

Table 5 shows the quantitative evaluation performance of five configurations under PSNR and SSIM metrics. For each configuration, the output images are generated based on the LR input image with and without motion blur, respectively. Then we compute the numerical results and average them to obtain the listed numbers. The quantitative results show that local residual blocks improve the baseline performance and facial masks improve more when adopted as the input for both SR and Deblur scenarios. Meanwhile, the dilation on our FSGN module is effective to predict the output and detail enhancement makes a further improvement. The performance gain is consistent with both PSNR and SSIM metrics. It indicates that facial masks, local residual blocks, dilation and detail enhancement will contribute to the quality of the face images for hallucination and deblurring. We note that the facial masks further improve the performance on the deblurring tasks compared with the hallucination task. It is because the facial masks enable CNN to attend to facial components containing unique structures, which usually diminish on the blurry inputs.

Figure 5 shows a visual example of these configurations. The input LR blurry face image is shown in (a) and the result generated by the baseline is in (b). The blurry effect still exists around the facial component and little improvement is achieved through local residual blocks integration shown in (c). However, when using facial masks, we notice that the blur around facial component is effectively reduced shown in (d). Furthermore, the dilation integration makes a further improvement (i.e, the right eye region in the close-up) shown in (e). The result generated by all the modules is shown in (f) where local details are transferred from HR exemplar images to (e) in the detail enhancement step.

4.2 Comparisons with the State-of-the-art Methods

We compare our method with the state-of-the-art methods both quantitatively and qualitatively. Table 6 reports the quantitative performance on the Multi-PIE dataset. In addition to the PSNR and SSIM metrics, we also involve the identity similarity (Delac et al. 2005) to measure how the results generated by different methods resemble the ground truth images. We use the ground truth training images to construct a PCA projection matrix which projects the results and corresponding HR images. After projection, we compute the cosine distance between each result and the corresponding ground truth image. This identity similarity is set to quantitatively measure image quality from the perspective of face recognition.

Table 6 shows that the bicubic interpolation achieves higher PSNR values than existing FH methods (i.e., FHTP and SFH) under both SR and deblur tasks. This is because FH methods establish HR correspondences through image alignment which is based on empirical features such as SIFT (Liu et al. 2011). As the resolution of the input image is low, empirical features cannot accurately locate HR correspondences. It leads to the mismatch and incorrect facial details will be transferred. As a result, around facial component areas, we will find the distortion of the shape, shifting of the location or the change of the lightness, as shown in Figs. 6b, 7b and

Table 6 The evaluation of the Multi-PIE dataset with the state-of-the-art methods

	PSNR SR/Deblur	SSIM SR/Deblur	Similarity SR/Deblur
Bicubic	32.43/23.58	0.89/0.81	0.92/0.87
FHTP (Liu et al. 2007)	30.13/23.13	0.82/0.77	0.90/0.86
SFH (Yang et al. 2013)	31.60/23.65	0.86/0.79	0.91/0.86
SRCNN (Dong et al. 2016a)	33.89/23.73	0.90/0.82	0.94/0.90
SRCSC (Gu et al. 2015)	33.95/23.82	0.90/0.82	0.95/0.91
SRResNet (Ledig et al. 2017)	34.10/23.95	0.90/0.81	0.96/0.92
VDSR (Kim et al. 2016a)	34.62/24.33	0.91/0.81	0.97/0.92
DFE (Pan et al. 2014)	31.53/25.26	0.87/0.85	0.94/0.94
RBF (Xu et al. 2017)	30.05/24.73	0.86/0.77	0.93/0.94
Ours	**34.93/25.75**	**0.92/0.86**	**0.98/0.96**

Bold values indicate the best performance

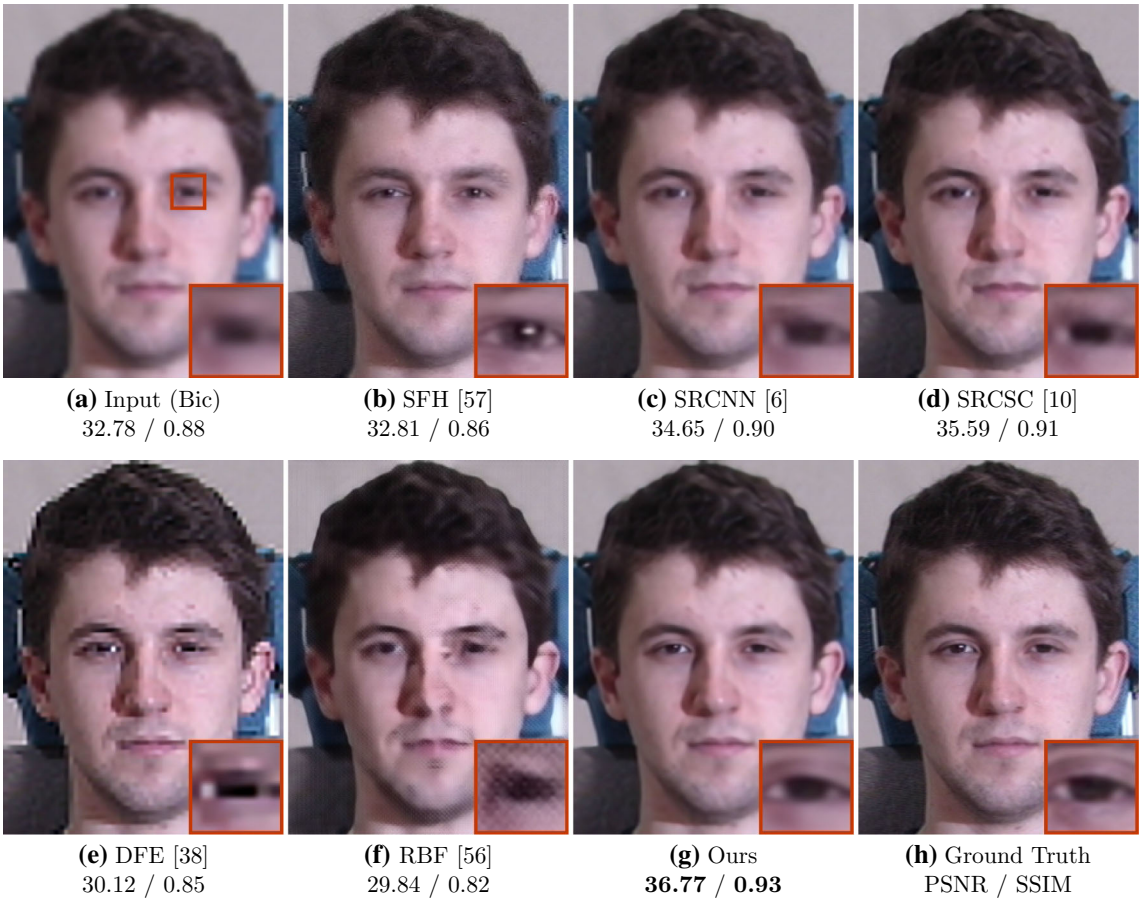


Fig. 6 Qualitative evaluations on the Multiple Dataset. **a** Bicubic upsampled input LR image without motion blur. **b–f** Comparison of the results. **g** Our result. **h** Ground truth image

9b. These artifacts deteriorate the image quality. In the FH task, the SRCNN, SRCSC and SRResNet methods achieve high PSNR values due to their global optimization scheme. However, blur occurs around high-frequency facial structures including eyes, noses, and mouth, which limits the image quality as well. Meanwhile, their performance decreases on the deblurring task. In comparison, the DFE and RBF methods are effective to handle motion blur while limiting their performance in hallucination. Different from existing methods which handle FH and FD independently, our method consists of a unified framework to jointly hallucinate and deblur face images. It recovers the original image content in both low and high frequencies, which enables the similarity of global appearance and local details between the output image and

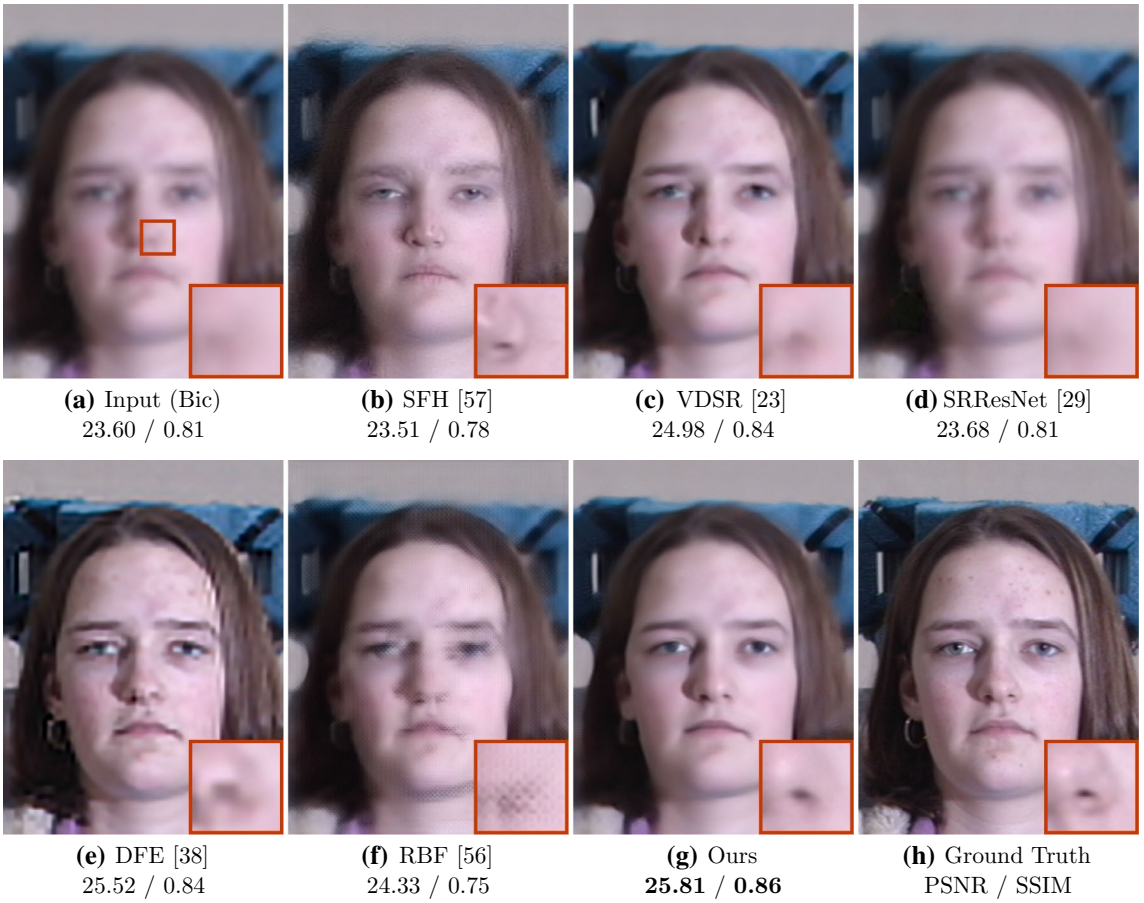


Fig. 7 Qualitative evaluation on the Multiple Dataset. **a** LR input blurry image with bicubic upsampling

Table 7 The evaluation of the PubFig dataset with the state-of-the-art methods

	PSNR SR/Deblur	SSIM SR/Deblur	Similarity SR/Deblur
Bicubic	29.55/22.79	0.86/0.83	0.89/0.84
FHTP (Liu et al. 2007)	26.56/22.51	0.71/0.79	0.87/0.83
SFH (Yang et al. 2013)	28.51/22.70	0.82/0.81	0.88/0.84
SRCNN (Dong et al. 2016a)	31.03/22.85	0.88/0.84	0.91/0.86
SRCSC (Gu et al. 2015)	31.15/23.13	0.88/0.85	0.92/0.87
SRResNet (Ledig et al. 2017)	31.23/23.21	0.88/0.85	0.94/0.87
VDSR (Kim et al. 2016a)	31.67/23.46	0.89/0.86	0.94/0.86
DFE (Pan et al. 2014)	28.74/23.95	0.81/0.87	0.89/0.88
RBF (Xu et al. 2017)	28.43/23.31	0.80/0.86	0.88/0.87
Ours	**31.86/24.12**	**0.90/0.89**	**0.96/0.90**

Bold values indicate the best performance

the ground truth. The evaluation of the PubFig dataset shows the similar performance in Table 7. It indicates our method is effective to overcome real-world input variations. Tables 6 and 7 show that our method performs favorably against the state-of-the-art FH, image SR and FD methods.

Besides quantitative evaluation, we also evaluate our method visually on the benchmarks. We show the qualitative comparison from Figs. 6, 7, 8 and 9. In Fig. 6, we evaluate the proposed algorithm on the Multi-PIE dataset using the input LR image without motion blur. The result generated by SFH is shown in (b) contains light dots on the right eye, which is different from the ground truth. This is because SFH selects the most similar component from the dataset and transfer its gradient to recover high-frequency details. However, the facial component correspondence cannot be well established in LR. In this case, gradient transfer leads

(a) Input (Bic) 31.20 / 0.84 **(b)** SFH [57] 30.39 / 0.81 **(c)** SRCNN [6] 33.37 / 0.88 **(d)** SRResNet [29] 32.99 / 0.88

(e) DFE [38] 28.85 / 0.81 **(f)** RBF [56] 29.05 / 0.78 **(g)** Ours **34.06 / 0.90** **(h)** Ground Truth PSNR / SSIM

Fig. 8 Qualitative evaluation on the PubFig Dataset. **a** LR input image without motion blur which is generated by bicubic upsampling

to the dissimilar generation of the facial structure. Another visual result is shown in Fig. 8b where the lighting, shape and position of the facial components are different from those in Fig. 8h although they look similar. It also indicates that erroneous gradient transfer brings artifacts due to inaccurate correspondence establishment. In addition, noise is included due to incorrect matching around the mouth region. As the PubFig dataset is taken in the real world condition and the training dataset is taken in the lab controlled environment. The component matching is not as accurate as that in Multi-PIE. It brings more artifacts on the generated results.

The results by the representative CNN-based methods are shown in Fig. 6c, d. Although the recovered images have high PSNR and SSIM values compared to that by SFT method, these methods are not effective to capture high-frequency facial details. The structures around facial components (i.e., eyes, nose and mouth) are blurry and details are missing. In addition, the results generated by VDSR and SRResNet shown in Fig. 7c, d show the similar performance. This indicates that CNN methods for general image SR are less effective to preserve details on face images. To solve this problem, we generate the base image through CNN prediction and enhance details via HR exemplar images. The base image contains the low-frequency facial structures similar to the existing CNN based methods. Meanwhile, we synthesize fine-grained structures from HR exemplar images and transfer their high-frequency details back to the base image for enhancement. Our two-stage scheme enables our results are similar to the ground truth in both global appearance and local details shown in Figs. 6g and 8g. The proposed algorithm achieves favorable performance under numerical evaluations as well as visual perception.

Figures 7 and 9 show the qualitative evaluation on the blurry Multi-PIE and PubFig datasets, respectively. The input is an LR face image with random motion blur. It limits the performance of existing FH and image SR results shown in Figs. 7b–d and 9b–d. The exemplar-based face deblurring method DFE selects a suitable exemplar and transfers the gradient into the blurry input image. It is effective to retain the low-frequency structure while limits its performance to restore the facial details. As shown in Figs. 7e and 9e, the artifacts appear on the whole image. Meanwhile, the results generated by the GAN network cannot reduce the artifacts and deteriorates the facial structure as shown in Figs. 7f and 9f. They aim to solve extremely low-resolution face images (e.g., 20×20) with ambiguous facial components, which

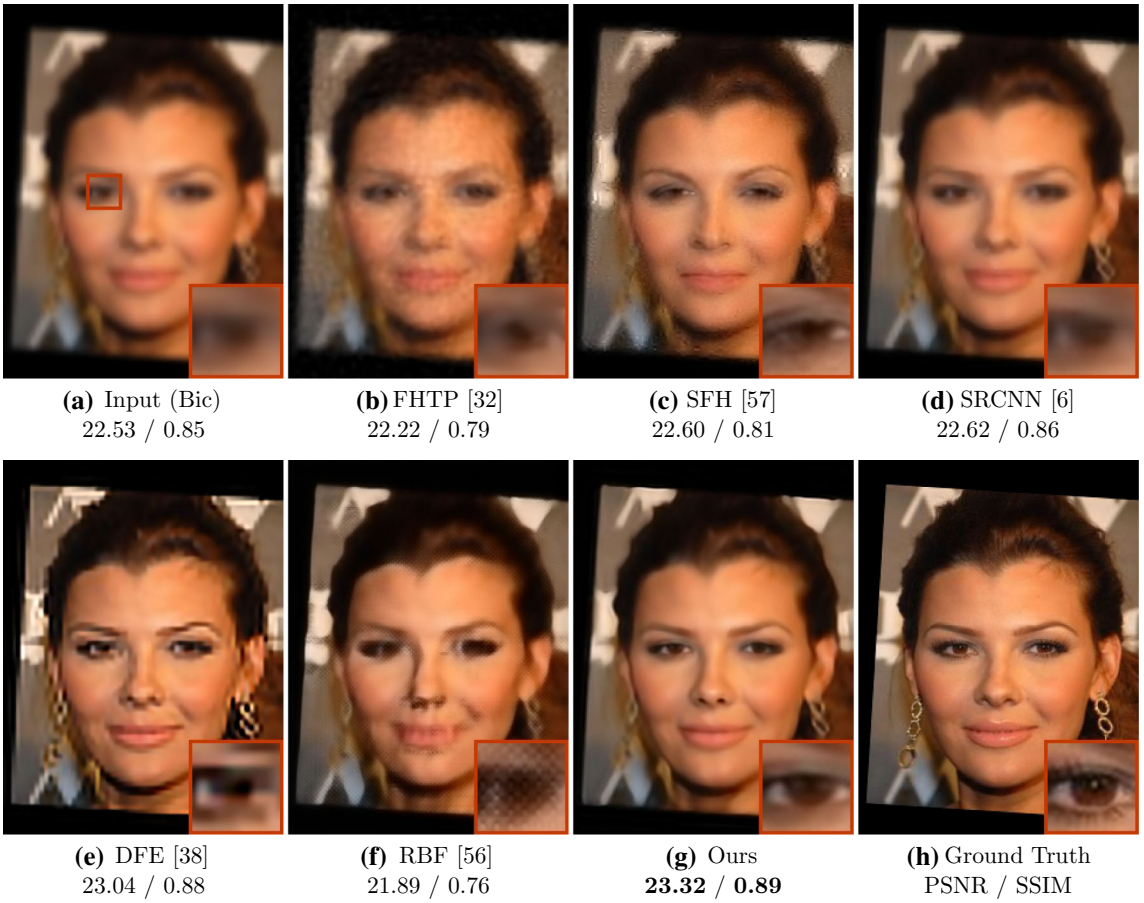


Fig. 9 Qualitative evaluation on the PubFig Dataset. **a** LR input blurry image which is generated by bicubic upsampling

are similar to noise. The GAN loss function will introduce fake details thus degrading the quality of the restored face images. The identity of their output is usually not preserved compared with the ground truth. In comparison, our method first generates the base image to reduce the blurry effect and further enhance the structure details. It can jointly handle the hallucination and deblurring tasks where it performs favorably against existing FH and FD methods quantitatively and qualitatively.

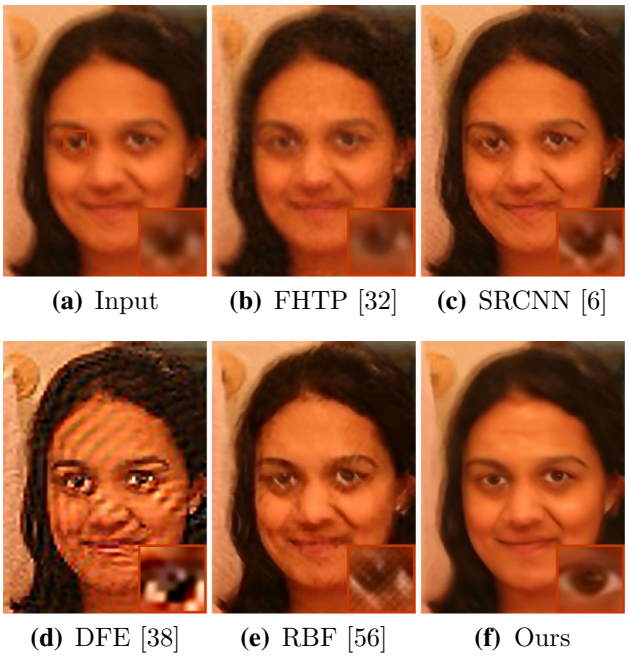
Besides evaluation of the standard benchmarks using synthetic motion blur kernels, we also evaluate on the blurry face images in the real world condition. Figure 10a shows an example where the input is a real blurry face image from (Lai et al. 2016). It fails existing exemplar-based FH and FD methods to establish an accurate correspondence between the input and the exemplar, which brings artifacts shown in (b) and (d). Meanwhile, the CNN based methods are not effective to reduce the blur shown in (c) and (e). Different from existing methods, our method first generates a base image to facilitate exemplar matching and then performs detail enhancement on the base image. It accurately transfers details from the exemplar to the base image shown in (f), which indicates that our method is effective to reduce real blurry face images.

4.3 Computational Cost

We evaluate the time cost of each method to generate an output image. All the evaluations are conducted on a PC with an i7 3.6GHz CPU and a Tesla K40c GPU. Table 8 shows the time cost of each method. We observe that the exemplar-based methods (i.e., FHTP, SFH, DEF) consume much time cost, which is mainly because of the querying on the exemplar dataset. In comparison, the CNN based methods (i.e., SRCNN, SRResNet, VDSR, RBF) take less time for an end-to-end prediction. Our method consists of the CNN prediction and exemplar-based searching, which takes more time than end-to-end CNN prediction while still performs favorably against exemplar-based methods.

4.4 Limitation

The proposed algorithm is less effective when the structures around the facial component are not available or significantly different from the training images. In such cases, the proposed algorithm would reduce to a conventional CNN-based image restoration algorithm as the facial components do not



(a) Input **(b)** FHTP [32] **(c)** SRCNN [6]

(d) DFE [38] **(e)** RBF [56] **(f)** Ours

Fig. 10 Qualitative evaluation on a real blurry face image. **a** LR input blurry image generated by bicubic upsampling

Table 8 The time cost of comparing methods to generate an output 320×240 image on the benchmarks

Methods	Time (s)	Methods	Time (s)
FHTP	98.9	SFH	245.1
SRCNN	4.36	SRCSC	43.56
SRResNet	6.5	VDSR	5.7
DFE	162.4	RBF	4.2
Ours	95.3		

(a) Input **(b)** Ours **(c)** Ground Truth

Fig. 11 Limitations of the proposed method. **a** LR input image which is generated by bicubic upsampling

help the estimation. Figure 11 shows an example where our method is not able to recover clear face images.

5 Concluding Remarks

We propose an effective algorithm to jointly hallucinate and deblur face images. With the guidance of facial components, we develop an FSGN to remove blur and restore the major structures of face images. To recover realistic faces, we develop a detail enhancement algorithm by high-resolution exemplars. Our analysis shows that the proposed method is able to generate high-resolution faces from blurry LR face images. Extensive experimental results demonstrate that our method performs favorably against the state-of-the-art approaches.

Acknowledgements This work has been supported in part by the NSF CAREER (No. 1149783), NSF of China (No. 61872421 and 61572099), NSF of Jiangsu Province (No. BK20180471), and National Science and Technology Major Project (2018ZX04041001-007).

References

Agustsson, E., Timofte, R., & Van Gool, L. (2017). Anchored regression networks applied to age estimation and super resolution. In *IEEE international conference on computer vision*.

Cao, Q., Lin, L., Shi, Y., Liang, X., & Li, G. (2017). Attention-aware face hallucination via deep reinforcement learning. In *IEEE conference on computer vision and pattern recognition*.

Chen, Y., Tai, Y., Liu, X., Shen, C., & Yang, J. (2017). Fsrnet: End-to-end learning face super-resolution with facial priors. arXiv preprint arXiv:1711.10703

Delac, K., Grgic, M., & Grgic, S. (2005). Independent comparative study of PCA, ICA, and LDA on the FERET data set. *International Journal of Imaging Systems and Technology*, *15*(5), 252–260.

Dong, C., Loy, C. C., He, K., & Tang, X. (2014). Learning a deep convolutional network for image super-resolution. In *European conference on computer vision*.

Dong, C., Loy, C. C., He, K., & Tang, X. (2016a). Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *38*(2), 295–307.

Dong, C., Loy, C. C., & Tang, X. (2016b). Accelerating the super-resolution convolutional neural network. In *European conference on computer vision*.

Eisemann, E., & Durand, F. (2004). Flash photography enhancement via intrinsic relighting. In *ACM transactions on graphics (SIGGRAPH)*.

Gross, R., Matthews, I., Cohn, J., Kanade, T., & Baker, S. (2010). Multi-pie. *Image and Vision Computing*, *28*(5), 807–813.

Gu, S., Zuo, W., Xie, Q., Meng, D., Feng, X., & Zhang, L. (2015). Convolutional sparse coding for image super-resolution. In *IEEE international conference on computer vision*.

Gunturk, B. K., Batur, A. U., Altunbasak, Y., Hayes, M. H., & Mersereau, R. M. (2003). Eigenface-domain super-resolution for face recognition. *IEEE Transactions on Image Processing*, *12*(5), 597–606.

Hacohen, Y., Shechtman, E., & Lischinski, D. (2013). Deblurring by example using dense correspondence. In *IEEE international conference on computer vision*.

He, K., Sun, J., & Tang, X. (2010). Guided image filtering. In *European conference on computer vision*.

He, K., Sun, J., & Tang, X. (2013). Guided image filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *6*, 1397–1409.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *IEEE conference on computer vision and pattern recognition*.

Jia, K., & Gong, S. (2005). Multi-modal tensor face for simultaneous super-resolution and recognition. In *IEEE international conference on computer vision*.

Jia, K., & Gong, S. (2006). Multi-resolution patch tensor for facial expression hallucination. In *IEEE conference on computer vision and pattern recognition*.

Jia, K., & Gong, S. (2008). Generalized face super-resolution. *IEEE Transactions on Image Processing*, *17*(6), 873–886.

Jin, Y., & Bouganis, C. S. (2015). Robust multi-image based blind face hallucination. In *IEEE conference on computer vision and pattern recognition*.

Johnson, J., Alahi, A., & Fei-Fei, L. (2016). Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*.

Jourabloo, A., Ye, M., Liu, X., & Ren, L. (2017). Pose-invariant face alignment with a single CNN. In *IEEE international conference on computer vision*.

Karras, T., Aila, T., Laine, S., & Jaakko, L. (2018). Progressive growing of GANs for improved quality, stability, and variation. In *International conference on learning representation*.

Kim, J., Kwon Lee, J., & Mu Lee, K. (2016a). Accurate image super-resolution using very deep convolutional networks. In *IEEE conference on computer vision and pattern recognition*.

Kim, J., Kwon Lee, J., & Mu Lee, K. (2016b). Deeply-recursive convolutional network for image super-resolution. In *IEEE conference on computer vision and pattern recognition*.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. In *International conference on learning representation*.

Kumar, N., Berg, A. C., Belhumeur, P. N., & Nayar, S. K. (2009). Attribute and similar classifiers for face verification. In *IEEE international conference on computer vision*.

Lai, W. S., Huang, J. B., Ahuja, N., & Yang, M. H. (2017). Deep Laplacian pyramid networks for fast and accurate super-resolution. In *IEEE conference on computer vision and pattern recognition*.

Lai, W. S., Huang, J. B., Hu, Z., Ahuja, N., & Yang, M. H. (2016). A comparative study for single image blind deblurring. In *IEEE conference on computer vision and pattern recognition*.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., et al. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In *IEEE conference on computer vision and pattern recognition*.

Lim, B., Son, S., Kim, H., Nah, S., & Lee, K. M. (2017). Enhanced deep residual networks for single image super-resolution. In *The IEEE conference on computer vision and pattern recognition workshops*.

Liu, C., Shum, H. Y., & Freeman, W. T. (2007). Face hallucination: Theory and practice. *International Journal of Computer Vision*, *75*(1), 115–134.

Liu, C., Yuen, J., & Torralba, A. (2011). Sift flow: Dense correspondence across scenes and its applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *33*(5), 978–994.

Liu, S., & Yang, M. H. (2014). Compressed face hallucination. In *IEEE international conference on image processing*.

Liu, W., Lin, D., & Tang, X. (2005). Hallucinating faces: Tensorpatch super-resolution and coupled residue compensation. In *IEEE conference on computer vision and pattern recognition*.

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

Pan, J., Dong, J., Tai, Y. W., Su, Z., & Yang, M. H. (2017). Learning discriminative data fitting functions for blind image deblurring. In *IEEE international conference on computer vision*.

Pan, J., Hu, Z., Su, Z., & Yang, M. H. (2014). Deblurring face images with exemplars. In *European conference on computer vision*.

Pan, J., Sun, D., Pfister, H., & Yang, M. H. (2017). Deblurring images via dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. https://doi.org/10.1109/TPAMI.2017.2753804.

Park, H., & Mu Lee, K. (2017). Joint estimation of camera pose, depth, deblurring, and super-resolution from a blurred image sequence. In *IEEE international conference on computer vision*.

Petschnigg, G., Szeliski, R., Agrawala, M., Cohen, M., Hoppe, H., & Toyama, K. (2004). Digital photography with flash and no-flash image pairs. In *ACM transactions on graphics (SIGGRAPH)*.

Sajjadi, M. S., Scholkopf, B., & Hirsch, M. (2017). Enhancenet: Single image super-resolution through automated texture synthesis. In *IEEE international conference on computer vision*.

Shen, Z., Lai, W. S., Xu, T., Kautz, J., & Yang, M. H. (2018). Deep semantic face deblurring. In *The IEEE conference on computer vision and pattern recognition*.

Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., et al. (2016). Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *IEEE conference on computer vision and pattern recognition*.

Song, Y., Bao, L., He, S., Yang, Q., & Yang, M. H. (2017a). Stylizing face images via multiple exemplars. *Computer Vision and Image Understanding*, *162*, 135–145.

Song, Y., Bao, L., Yang, Q., & Yang, M. H. (2014). Real-time exemplar-based face sketch synthesis. In *European conference on computer vision*.

Song, Y., Zhang, J., Bao, L., & Yang, Q. (2017b). Fast preprocessing for robust face sketch synthesis. In *International joint conference on artificial intelligence*.

Song, Y., Zhang, J., He, S., Bao, L., & Yang, Q. (2017c). Learning to hallucinate face images via component generation and enhancement. In *International joint conference on artificial intelligence*.

Tai, Y., Yang, J., & Liu, X. (2017). Image super-resolution via deep recursive residual network. In *IEEE conference on computer vision and pattern recognition*.

Tappen, M. F., & Liu, C. (2012). A bayesian approach to alignment-based image hallucination. In *European conference on computer vision*.

Timofte, R., Agustsson, E., Van Gool, L., Yang, M. H., Zhang, L., Lim, B., et al. (2017). Ntire 2017 challenge on single image super-resolution: Methods and results. In *IEEE conference on computer vision and pattern recognition workshops*.

Wang, N., Tao, D., Gao, X., Li, X., & Li, J. (2014). A comprehensive survey to face hallucination. *International Journal of Computer Vision*, *106*(1), 9–30.

Wang, X., & Tang, X. (2005). Hallucinating face by eigentransformation. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, *35*(3), 425–434.

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, *13*(4), 600–612.

Wang, Z., Liu, D., Yang, J., Han, W., & Huang, T. (2015). Deep networks for image super-resolution with sparse prior. In *IEEE international conference on computer vision*.

Xu, X., Sun, D., Pan, J., Zhang, Y., Pfister, H., & Yang, M. H. (2017). Learning to super-resolve blurry face and text images. In *IEEE international conference on computer vision*.

Yang, C. Y., Liu, S., & Yang, M. H. (2013). Structured face hallucination. In *IEEE conference on computer vision and pattern recognition*.

Yang, C. Y., Liu, S., & Yang, M. H. (2018a). Hallucinating compressed face images. *International Journal of Computer Vision*, *126*(6), 597–614.

Yang, X., Xu, K., Song, Y., Zhang, Q., Wei, X., & Lau, R. W. (2018b). Image correction via deep reciprocating HDR transformation. In *IEEE conference on computer vision and pattern recognition*.

Yu, F., Koltun, V., & Funkhouser, T. (2017). Dilated residual networks. In *IEEE conference on computer vision and pattern recognition*.

Yu, X., & Porikli, F. (2016). Ultra-resolving face images by discriminative generative networks. In *European conference on computer vision*.

Yu, X., & Porikli, F. (2017a). Face hallucination with tiny unaligned images by transformative discriminative neural networks. In *AAAI conference on artificial intelligence*.

Yu, X., & Porikli, F. (2017b). Hallucinating very low-resolution unaligned and noisy face images by transformative discriminative autoencoders. In *IEEE conference on computer vision and pattern recognition*.

Zhang, J., Pan, J., Ren, J., Song, Y., Bao, L., Lau, R. W., & Yang, M. H. (2018). Dynamic scene deblurring using spatially variant recurrent neural networks. In *IEEE conference on computer vision and pattern recognition*.

Zhou, E., Fan, H., Cao, Z., Jiang, Y., & Yin, Q. (2015). Learning face hallucination in the wild. In *AAAI conference on artificial intelligence*.

Zhu, S., Liu, S., Loy, C. C., & Tang, X. (2016). Deep cascaded bi-network for face hallucination. In *European conference on computer vision*.

Zhu, X., & Ramanan, D. (2012). Face detection, pose estimation, and landmark localization in the wild. In *IEEE conference on computer vision and pattern recognition*.