# Learning to Deblur Face Images via Sketch Synthesis

**Songnan Lin,[1*] Jiawei Zhang,[2†] Jinshan Pan,[3] Yicun Liu,[2]**
**Yongtian Wang,[1] Jing Chen,[1†] Jimmy Ren[2]**
[1]Beijing Institute of Technology, Beijing, China, [2]SenseTime Research, Shenzhen, China
[3]Nanjing University of Science and Technology, Nanjing, China
{linsongnan2015, wyt, chen74jing29}@bit.edu.cn, {zhjw1988, sdluran, yicunliu96, jimmy.sj.ren}@gmail.com

**Abstract**

The success of existing face deblurring methods based on deep neural networks is mainly due to the large model capacity. Few algorithms have been specially designed according to the domain knowledge of face images and the physical properties of the deblurring process. In this paper, we propose an effective face deblurring algorithm based on deep convolutional neural networks (CNNs). Motivated by the conventional deblurring process which usually involves the motion blur estimation and the latent clear image restoration, the proposed algorithm first estimates motion blur by a deep CNN and then restores latent clear images with the estimated motion blur. However, estimating motion blur from blurry face images is difficult as the textures of the blurry face images are scarce. As most face images share some common global structures which can be modeled well by sketch information, we propose to learn face sketches by a deep CNN so that the sketches can help the motion blur estimation. With the estimated motion blur, we then develop an effective latent image restoration algorithm based on a deep CNN. Although involving the several components, the proposed algorithm is trained in an end-to-end fashion. We analyze the effectiveness of each component on face image deblurring and show that the proposed algorithm is able to deblur face images with favorable performance against state-of-the-art methods.

## 1 Introduction

Single-image motion deblurring has long been an active research topic in computer vision and image processing. It aims to recover a latent clear image and the corresponding blur kernel from a single blurred input which is usually caused by camera shake or object motion. The blurring process is usually modeled as a convolution when the blur is spatially invariant:

$$y = k * x + n, \tag{1}$$

where $y, x, k, n$ denote blurred image, latent image, blur kernel and noise, respectively; $*$ denotes the convolution operator.

To estimate the blur kernels, most state-of-the-art methods (Cho and Lee 2009; Xu and Jia 2010; Krishnan, Tay, and Fergus 2011; Xu, Zheng, and Jia 2013) either implicitly or explicitly restore sharp edges from blurry images. Although using the sharp edges is able to help motion blur estimation for generic natural images, the sharp edges by these methods consider only local edges rather than structural information of a particular object class, which leads to less effectiveness for the blurry face images as they contain scarce textures (see Figure 1(c)).

In order to extract more reliable edges, Pan et al. (2014) search similar exemplars from an external dataset and generate robust contours (Figure 1(k)) as the informative structures for kernel estimation. However, this method needs manually labeled contours and the exemplar selection procedure is computationally inefficient.

Recently, convolutional neural networks (CNNs) have been developed for image deblurring and shown promising results. Schuler et al. (2016) use a deep CNN to estimate blur kernels. However, this method is less effective for the large motion blur. Xu et al. (2018) and Pan et al. (2018) use a deep CNN to predict local sharp edges from blurry images for blur kernel estimation and employ the conventional deblurring algorithm (Pan et al. 2014) to estimate blur kernels and latent images. These methods are effective for the images with rich textures but less effective for the blurry face images with scarce textures (Figure 1(e)). In addition, they use the iterative optimization framework for kernel estimation and need non-blind deconvolution, e.g. (Zoran and Weiss 2011), to restore the sharp images, which makes them computationally expensive. Several algorithms (Nah, Hyun Kim, and Mu Lee 2017; Kupyn et al. 2018; Zhang et al. 2018; Tao et al. 2018) exploit the advantage of high capacity of deep learning models and design end-to-end trainable networks to restore the sharp images directly. However, these methods mainly rely on the large model capacity and pay less attention to analyzing the adopted architectures, which makes them less effective for face image deblurring as shown in Figure 1(b) and (i).

To develop efficient deep models for face image deblurring, Shen et al. (2018) directly concatenate face parsing results with blurry face images as the input of their network. However, this method is limited by the face parsing results as the face parsing and the image deblurring are handled sep-

*This work was done when Songnan Lin was an intern at SenseTime.

†Corresponding author.

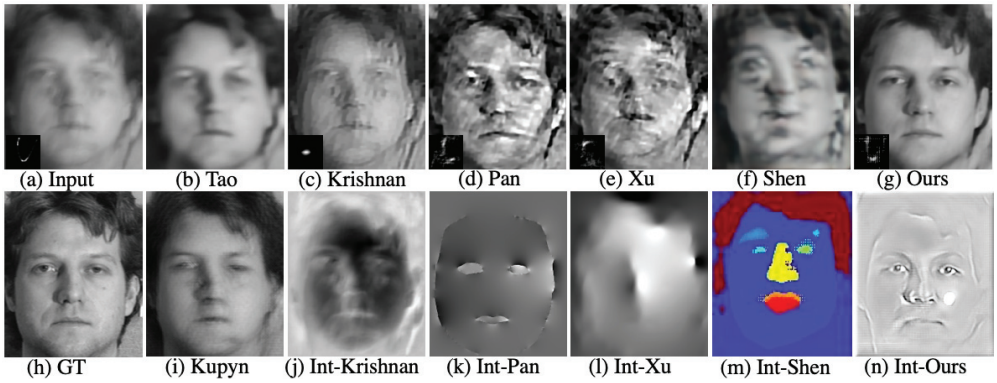


Figure 1: A challenging case for face image deblurring. "Int" denotes the intermediate structure or face parsing of its corresponding algorithm. The bottom left corners of (a) and (c)(d)(e)(g) are the ground truth blur kernel and the estimated ones, respectively. The proposed method generates a sketch (n) with semantic structures and restores a higher-quality image (g) via an end-to-end network based on the conventional deblurring process.

arately.

As human faces are an interesting object class with numerous applications, and existing deep learning-based methods do not make full use of properties of face images, it is a great need to develop an efficient algorithm to solve the face image deblurring problem.

In this paper, we propose an effective face image deblurring algorithm based on deep CNNs and sketch information of face images. The proposed algorithm is motivated by the conventional deblurring process which involves a motion blur estimation step and a latent clear image restoration step. To better exploit the properties of face images for blur estimation, we propose to learn sketches of face images so that they can guide the blur estimation. As the sketches of face images usually model the global structures, such as the lower contour, mouth, eyes and nose, our sketch-based method is more effective than those based on local sharp edges for face image deblurring. Although the blur kernel is estimated, restoring latent clear images is not a trivial task. Different from the deep CNN-based methods (Schuler et al. 2016; Xu et al. 2018) which only use deep CNNs to estimate blur kernels, we develop an effective deep model to restore clear images. To further remove artifacts and restore images with finer details, we use the learned sketches to constrain the network for the final image restoration. Figure 2 shows the framework of the proposed algorithm. Our method integrates the sketch synthesis, kernel estimation, image restoration and image de-artifacts into a unified framework and can be trained in an end-to-end manner, so that it can incorporate the domain knowledge of image deblurring neatly and make the whole network more compact for image deblurring. As the proposed network is designed based on the conventional deblurring process, it is able to handle blurred face images with significant blur and generate physically correct results.

The main contributions of this paper are summarized as follows:

- We propose to learn sketch synthesis for face image deblurring. The sketch information acts as the semantic prior which is able to model the global structures of face images and thus facilitates the blur estimation.
- We propose an end-to-end trainable neural network to restore latent clear face images. The proposed network design is based on the conventional deblurring process which includes the motion blur estimation module and the latent clear image restoration module. The blur estimation module can be efficiently solved by using the learned sketch information. The latent clear image restoration module is further constrained by the learned sketches of face images, which is able to restore latent clear images.
- We quantitatively and qualitatively evaluate the proposed approach and show that it outperforms the state-of-the-art deblurring methods on both synthetic and real data.

## 2 Proposed Method

To better motivate our algorithm, we first revisit the conventional image deblurring process. As image deblurring is an ill-posed problem, conventional algorithms (Fergus et al. 2006; Cho and Lee 2009; Krishnan, Tay, and Fergus 2011; Xu, Zheng, and Jia 2013; Pan et al. 2016) usually impose priors on blur kernels and latent images to constrain the solution space and then estimate the blur kernel as well as the latent clear image by iteratively solving:

$$\min_{k} \|\nabla y - k * \nabla s\|_2^2 + \phi(k), \tag{2}$$

and

$$\min_{x} \|y - k * x\|_2^2 + \rho(x), \tag{3}$$

where $\phi(k)$ and $\rho(x)$ are image priors w.r.t. $k$ and $x$; $\nabla$ denotes the image gradient operator; $\nabla s$ denotes the sharp edges extracted from the intermediate latent image $x$ or blurred image $y$. Based on (2), numerous algorithms either implicitly or explicitly restore salient edges $\nabla s$ for kernel estimation. As mentioned above, the local sharp edges are less effective for face images (Figure 1 (j) and (l)) while using face contours or exemplars (Figure 1 (k)) leads to time-consuming algorithms. In addition, solving (3) is not a trivial task. To overcome these problems, we propose to learn

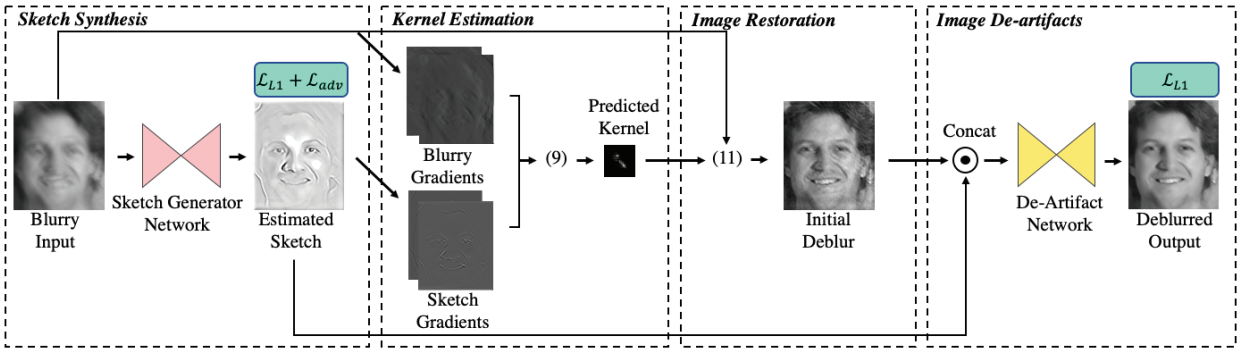


Figure 2: Overview of our network. Blurry inputs are fed into a sketch generator network to predict sketches supervised by adversarial and content losses. Then, given the gradients of blurry inputs and sketches, blur kernels are estimated by (9). Afterward, initial deblurred images are restored via (11). To further suppress artifacts, the final deblurred results are obtained by a de-artifact network.

sketches of face images for blur kernel estimation by a deep CNN. With the estimated blur kernels, we then use a deep CNNs to estimate the latent clear images. To this end, the proposed network consists of four parts:

- *Sketch synthesis:* it aims to extract sketches, which model the global structures of face images, such as the lower contour, mouth, eyes and nose, from blurry face images.
- *Kernel estimation:* it utilizes the learned sketches to estimate the blur kernels according to (2).
- *Image restoration:* it generates the latent clear images based on the estimated kernels.
- *Image de-artifacts:* it uses the sketches of face images to remove the artifacts from the deblurred images and restore better-deblurred images with finer details.

**Sketch Synthesis.** We exploit global structure priors of face images via a sketch generator network by taking advantage of its semantic abstraction capacity. Specifically, given a blurry image $y$, we aim to learn a mapping to transform $y$ into a sketch $s$ via the proposed sketch generator network. Similar to $pix2pix$ (Isola et al. 2017), we supervise the network with a combination of adversarial and content losses:

$$\mathcal{L} = \mathcal{L}_{adv}(G, D, s, y) + \lambda \mathcal{L}_{L1}(G, s, y), \tag{4}$$

in which $\lambda$ is a hyperparameter that controls the relative importances of two terms. The adversarial loss is used to generate more realistic sketches, modeled as $\hat{s} = G(y), y \sim \mathrm{P}_y$, where $y$ is a sample from the blurred image distribution. The discriminator $D(s)$ is based on WGAN (Hörmander, Totaro, and Waldschmidt 2006) and $\mathcal{L}_{adv}(G, D, s, y)$ is formulated as:

$$\mathcal{L}_{adv}(G, D, s, y) = \mathop{\mathrm{E}}_{s\sim \mathrm{P}_r}[D(s)] - \mathop{\mathrm{E}}_{\hat{s}\sim \mathrm{P}_g}[D(\hat{s})], \tag{5}$$

where $\mathrm{P}_r$ and $\mathrm{P}_g$ denote the distribution of real sketches and the distribution of generated sketches, respectively. As the sketch synthesis module not only needs to generate realistic sketch-style images but also inherently learns to preserve structure alignment between latent images and generated sketches, we utilize $L_1$ distance as a content loss:

$$\mathcal{L}_{L1}(G, s, y) = \mathop{\mathrm{E}}_{\hat{s}\sim \mathrm{P}_g}[\| \hat{s} - s \|_1]. \tag{6}$$

The final objective function is defined as:

$$G^* = \min_G \max_{D\in\mathcal{D}} \mathcal{L}, \tag{7}$$

where $\mathcal{D}$ represents the set of 1-Lipschitz functions.

**Kernel Estimation.** With the predicted sketch $\hat{s}$, we can estimate blur kernel according to (2). Similar to state-of-the-art methods (Xu, Zheng, and Jia 2013; Pan et al. 2016), we adopt the commonly used prior $\|k\|_2^2$ to constrain $k$. Thus, the blur kernel can be estimated by

$$\hat{k} = \arg\min_k \|\nabla y - k * \nabla \hat{s}\|_2^2 + \beta \|k\|_2^2, \tag{8}$$

in which $\beta$ is a positive weight parameter. As (8) is a least squares problem, its solution can be obtained by

$$\hat{k} = \mathcal{F}^{-1}\left(\frac{\overline{\mathcal{F}(\nabla_h \hat{s})}\mathcal{F}(\nabla_h y) + \overline{\mathcal{F}(\nabla_v \hat{s})}\mathcal{F}(\nabla_v y)}{|\mathcal{F}(\nabla_h \hat{s})|^2 + |\mathcal{F}(\nabla_v \hat{s})|^2 + \beta}\right), \tag{9}$$

where $\mathcal{F}$ denotes the discrete Fourier transform matrix, $\mathcal{F}^{-1}$ and $\overline{\mathcal{F}(\cdot)}$ are the inverse and the complex conjugate of $\mathcal{F}$, respectively; $\nabla_h$ and $\nabla_v$ denote the horizontal and vertical gradient operators. After obtaining the blur kernel $k$, we use the same method as (Pan et al. 2014; Gong et al. 2016; Schuler et al. 2016; Xu et al. 2018) to normalize $k$ so that the sum of its elements is 1.

**Image Restoration.** Given a blurred input $y$ and the estimated kernel $\hat{k}$, restoring the latent image $\hat{x}_0$ is a classic non-blind deconvolution problem. To efficiently get the latent clear image, we estimate the latent clear image by

$$\hat{x}_0 = \arg\min_x \|\hat{k} * x - y\|_2^2 + \gamma \|x\|_2^2, \tag{10}$$

where we use $\gamma\|x\|_2^2$ as the prior $\rho(x)$ and $\gamma$ is a positive parameter. Based on (10), we can get $\hat{x}_0$ by

$$\hat{x}_0 = \mathcal{F}^{-1}\left(\frac{\overline{\mathcal{F}(\hat{k})}\mathcal{F}(y)}{|\mathcal{F}(\hat{k})|^2 + \gamma}\right). \tag{11}$$

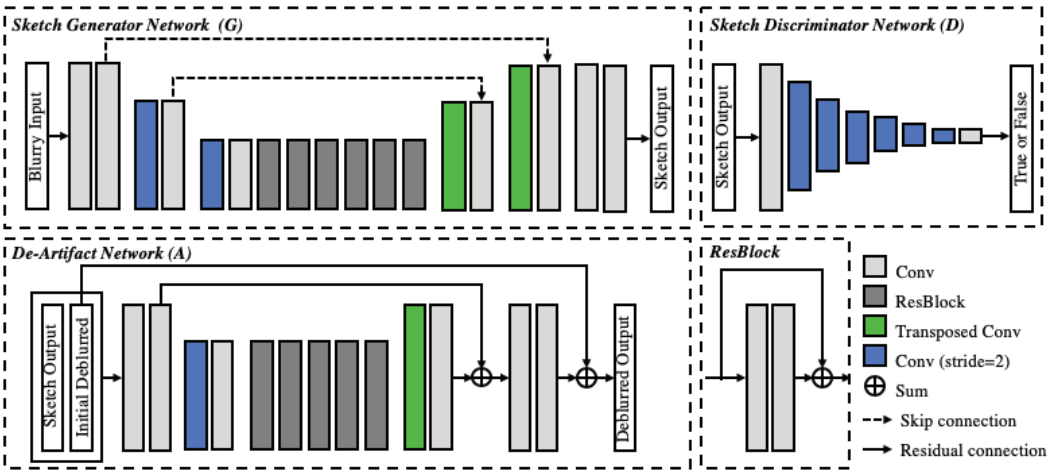


Figure 3: Structure of the sub-networks. Sketch Generator Network contains 10 convolutional layers, 2 transposed convolutional layers, 6 residual blocks, and 2 skip connections. Sketch Discriminator Network is identical to $PatchGAN$. De-Artifacts Network contains 7 convolutional layers, 1 transposed convolutional layer, 5 residual blocks, and 2 residual connections.

**Image De-artifacts.** Although using (10) can efficiently obtain latent clear images, it usually leads to the results with significant artifacts. To remove artifacts and restore a sharp image with finer details, we propose a deep CNN to restore the final latent clear image guided by the synthesized sketch $\hat{s}$. Let $\mathcal{A}$ denote this network, we use the following loss function to constrain the network $\mathcal{A}$.

$$\hat{x}_f = \arg\min_{\hat{x}} \mathrm{E}[\| \hat{x} - x \|_1], \tag{12}$$

in which $\hat{x} = \mathcal{A}(\hat{x}_0, \hat{s}), \hat{x}_0 \sim \mathrm{P}_{\hat{x}_0}, \hat{s} \sim \mathrm{P}_g$, where $\mathrm{P}_{\hat{x}_0}$ denotes the distribution of the initial deblurred images.

Based on the above considerations, we develop a deep CNN which formulates the above components in a unified framework so that the network can be trained in an end-to-end fashion and effectively solve face image deblurring.

## 2.1 Network Architecture

**Sketch Generator Network.** The U-Net based architecture of the generator $G$ is shown in Figure 3. It consists of 10 convolutional layers, 2 transposed convolutional layers, and 6 residual blocks. An instance normalization (IN) layer and a Rectified Linear Unit (ReLU) activation layer are added after every convolutional layer and transposed convolutional layer. Except that the first and last convolutional layers take $7 \times 7$ filters to capture large spatial information, we apply $3 \times 3$ convolutional layers in this network. Moreover, two skip connections are used between the encoder and the decoder.

**Sketch Discriminator Network.** During the training phase, the WGAN (Hörmander, Totaro, and Waldschmidt 2006) based discriminator $D$ is used. The architecture of the sketch discriminator network is identical to $PatchGAN$ (Isola et al. 2017).

**De-Artifacts Network.** Inspired by the artifact removal network (Son and Lee 2017), we propose a de-artifact network $A$ as shown in Figure 3. Different from (Son and Lee 2017), the proposed network introduces the synthesized sketches to encourage the restoration to preserve details and edges. The sharp sketches and the initial deblurred results are concatenated and fed into this network. The network is downsampled once to enlarge the receptive field and save the computational cost. Afterward, we use 5 residual blocks which have proven effective in (Son and Lee 2017) for image restoration. The following transposed convolutional layer reconstructs the features with full resolution. A Parametric Rectified Linear Unit (PReLU) activation layer is added after every convolutional layer and transposed convolutional layer to add non-linearity into the network. Furthermore, we adopt two residual connections to make the network effective to handle changes between blurry-sharp image pairs and maintain intensity consistency.

# 3 Experimental Results

In this section, we quantitatively and qualitatively evaluate the proposed network against state-of-the-art algorithms on both synthetic and real data.

**Training Dataset.** We evaluate the proposed methods on four synthetic datasets: CMU PIE (Gross et al. 2010), Helen (Le et al. 2012), CelebA (Liu et al. 2015) and PubFig (Kumar et al. 2009). Blurred inputs, sharp images, and ground truth sketches are required during training. We collect 2,184 images from CMU PIE, 2,000 ones from Helen, 2,000 ones from CelebA and 2,400 ones from PubFig as the corresponding training datasets. As for Helen and PubFig that may contain multiple faces in one image, we use Dlib (King 2009) to detect the facial key points and align the face images with (Kazemi and Sullivan 2014).

As for the blurred inputs, we synthesize 20,000 motion

Table 1: Average PSNR, SSIM and time cost (sec) with image size of $240 \times 200$.

| Dataset | Criterion | Pan (2016) | Xu (2018) | Pan (2014) | Tao (2018) | Kupyn (2018) | Shen (2018) | Ours |
|---|---|---|---|---|---|---|---|---|
| MultiPIE | PSNR | 23.67 | 20.18 | 23.30 | 26.51 | 23.43 | 19.90 | 28.57 |
| | SSIM | 0.717 | 0.521 | 0.607 | 0.774 | 0.715 | 0.496 | 0.844 |
| Helen | PSNR | 20.46 | 17.47 | 18.84 | 24.89 | 20.93 | 18.62 | 26.12 |
| | SSIM | 0.599 | 0.416 | 0.469 | 0.711 | 0.626 | 0.702 | 0.751 |
| CelebA | PSNR | 20.62 | 17.18 | 19.34 | 23.73 | 20.89 | 20.84 | 24.92 |
| | SSIM | 0.601 | 0.408 | 0.502 | 0.727 | 0.644 | 0.794 | 0.859 |
| PubFig | PSNR | 20.76 | 17.74 | 19.04 | 25.70 | 20.95 | 21.04 | 27.09 |
| | SSIM | 0.644 | 0.436 | 0.493 | 0.782 | 0.707 | 0.753 | 0.813 |
| Time(sec) | | 82.12 | 9.972 | 20.414 | 0.032 | 0.014 | 0.053 | 0.029 |

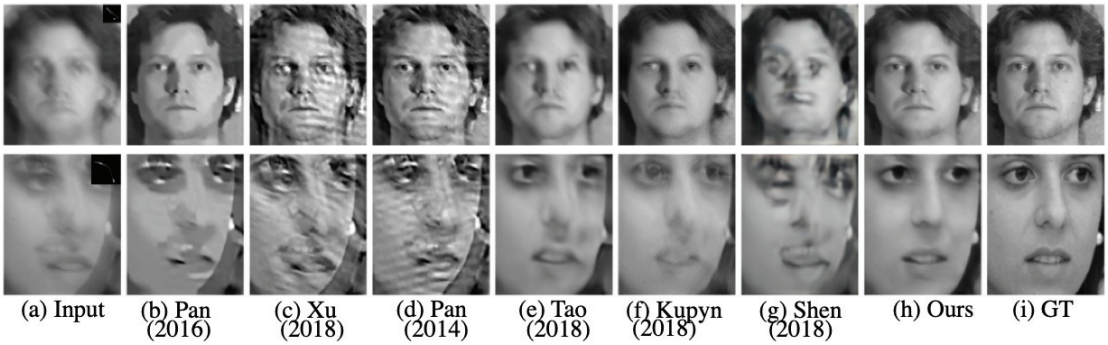


Figure 4: Visual comparisons with state-of-the-art on the synthetic dataset. Blurred inputs (a) are generated by convolving sharp images (i) with blur kernels. Kernel sizes from top to bottom are $31 \times 31$, $51 \times 51$ and kernels are shown in the top right corners of (a). (b)-(g) are the deblurred results from other methods, which contain artifacts and residual blurs. The proposed method performs favorably on different sizes of blur kernels in (h).

blur kernels similar to (Boracchi and Foi 2012) with sizes ranging from $21 \times 21$ to $51 \times 51$ and generate the blurred images by convolving latent images. To add diversity to the training data, we apply several data augmentations. We perform geometric transformations including randomly rotating within $\pm 5^o$, translating within $\pm 5\%$ of the image size, scaling within $\pm 0.1$ and cropping to $240 \times 200$. Besides, the brightness is uniformly sampled within $\pm 0.2$.

As for the ground truth of the sketches, we use two public face sketch dataset (Tang and Wang 2003) and (Wang and Tang 2008). As they only have 188 and 123 images respectively, we apply a sketch synthesis algorithm (Chen et al. 2018), which presents strong generalization to face photos in the wild, to transfer ground-truth latent images into the sketches and treat them as additional sketch ground truth.

**Implementation Details.** The training procedure is divided into three stages. First, we train the sketch generator network from the proposed face sketch dataset. The network is supervised by adversarial and content losses (4), which learns to capture the inherent facial properties and generates realistic sketches. And then, the de-artifacts network is trained with the fixed sketch generator network. Finally, the whole network is finetuned in an end-to-end manner, supervised by (12) alone. We use Pytorch (Paszke et al. 2017) to train the network on a GeForce GTX 1080 GPU. Adam (Kingma and Ba 2014) is adopted to optimize the network with momentum and momentum2 as 0.5 and 0.999 for sketch synthesis in the first training stage, 0.9 and 0.999 for the rest stages. Each stage contains 60 epochs and the learning rate is 0.0002. In all experiments, the parameters $\lambda$, $\beta$ and $\gamma$ are set as 10, 0.01 and 0.01. Besides, we adopt boundary adjustment (Kruse, Rother, and Schmidt 2017) for image restoration to alleviate the problematic circular convolution assumption imposed by FFT-based inference.

**Results on Synthetic Dataset.** For quantitative and qualitative evaluations, we collect 342 sharp face images in CMU PIE (Gross et al. 2010), 330 ones in Helen (Le et al. 2012), 300 ones in CelebA (Liu et al. 2015) and 300 ones in PubFig (Kumar et al. 2009). And then, we generate a test set with 8 random generated blur kernels. The kernel sizes are $51 \times 51$, $41 \times 41$, $31 \times 31$, $21 \times 21$ and every kernel size has two kernels. These large kernel sizes are used to validate the effectiveness of the proposed network for large blur. We compare the proposed network with 6 state-of-the-art algorithms including a conventional image deblurring method (Pan et al. 2016), a conventional face image deblurring method (Pan et al. 2014), a CNN-based edge selection method (Xu et al. 2018), end-to-end CNN-based methods (Tao et al. 2018; Kupyn et al. 2018) and a CNN-based method with face parsing prior (Shen et al. 2018).

We evaluate average PSNR and SSIM on the synthetic dataset in Table 1. The proposed algorithm performs favorably against state-of-the-art methods for face image deblurring. Figure 4 shows qualitative comparisons of different sizes of blur kernels. Conventional image deblurring method (Pan et al. 2016) and the edge selection-based

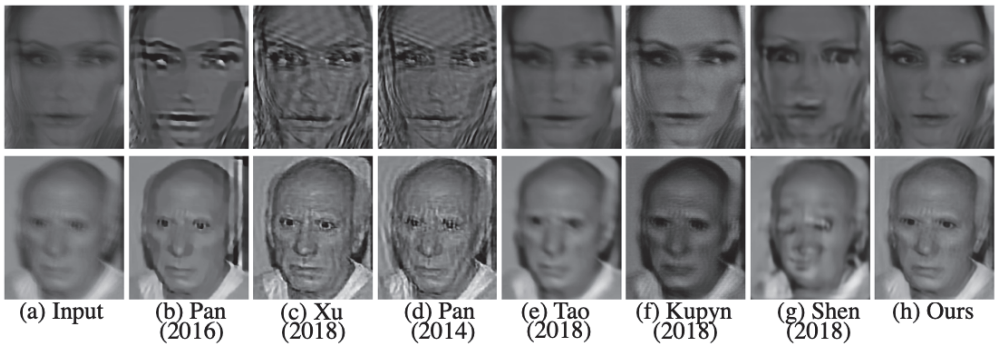


Figure 5: Visual comparisons with state-of-the-art on real blurry face images. The deblurred results of the proposed method have fewer artifacts and more details (best viewed on high-resolution displays).

CNN method (Xu et al. 2018) are not specially designed for the domain knowledge of face images, thus are less effective on blurry face images and introduce noticeable ringing artifacts. The face deblurring method (Pan et al. 2014) highly relies on the similarity of the corresponding exemplar, which leads to an unstable performance when the blur is large. The end-to-end CNN-based methods (Tao et al. 2018; Kupyn et al. 2018; Shen et al. 2018) neglect to analyze the adopted architecture and thus is less effective, especially on large motion blur. Shen et al. (2018) produce more artifacts because they conduct face parsing and image deblurring separately, which makes the approach sensitive to face parsing. On the contrary, the proposed method exploits semantic priors of face images and hinges on the conventional deblurring process simultaneously via an end-to-end architecture. The restored face images present more fine details and fewer visual artifacts.

**Results on Real Images.** We also evaluate the proposed method using real blurry images as shown in Figure 5. Our method restores more visually pleasing face images than the state-of-the-art.

**Run-time.** Table 1 shows the average time cost based on 10 images with a image size of $240 \times 200$. Pan et al. (2016) and Pan et al. (2014) run on a 4.2 GHz Intel i7 CPU, while the other algorithms run on a GeForce GTX 1080 GPU. The proposed method is more efficient than most of the state-of-the-art algorithms except (Kupyn et al. 2018).

## 4 Ablation Study

The proposed method is designed based on the conventional deblurring process. Firstly, it extracts global structures from blurry images by the sketch generator network. And then it estimates blur kernels using sketches. We restore latent clear images based on these kernels, and remove artifacts finally. In this section, we further analyze the effectiveness of each component in face image deblurring.

**Effectiveness of Sketch.** As stated in Sec. 2, the sketch information plays an important role in blur estimation. To demonstrate the effectiveness of sketch information in blur estimation, we compare the proposed method without using sketch constraints (4) (denoted as "Direct" ). We train this baseline method using the same settings as the proposed method for fair comparisons. The results in Figure 7 (c)(h) and Table 2 demonstrate that the proposed method without sketch constraints makes the sketch synthesis less effective for extracting global structures, which accordingly affects the kernel estimation and the latent image restoration. In contrast, the proposed network can better estimate structures from blurry face images (Figure 7(j)), estimate more accurate kernels using sketches (top right of Figure 7(e)) and restore much clearer and sharper images (Figure 7(e)).

Table 2: Ablation study. "Direct" trains the network without using sketch constraints. "$L_0$" supervises the sketch generator network by the structures extracted by the $L_0$ smoothing filter. "Naive E2E" restores latent images without explicit kernel estimation. "Intensity" estimates kernels in the intensity space. "$A$ w/o s" only inputs the initial deblurred results into the de-artifact network. The proposed method, including a sketch synthesis, a gradient-space-based kernel estimation and a sketch guided de-artifact network, achieves the highest quantitative results. Please see manuscripts for more details.

| Methods | Direct | $L_0$ | Naive E2E | Intensity | $A$ w/o s | Ours |
|---|---|---|---|---|---|---|
| PSNR | 23.85 | 27.68 | 25.72 | 27.37 | 28.13 | 28.57 |
| SSIM | 0.716 | 0.826 | 0.731 | 0.812 | 0.839 | 0.844 |

**Relations with Local Structure Extraction Methods.** We note that several algorithms (Xu and Jia 2010; Krishnan, Tay, and Fergus 2011; Xu, Zheng, and Jia 2013) use local structures of images for blur kernel estimation. The local structures are usually extracted by some edge-preserving filter methods (e.g., $L_0$ smoothing filter (Xu et al. 2011; Pan et al. 2018)). As stated in Sec. 2, local structures do not always help the face image deblurring as the local structures are usually not preserved during the blurring process. In contrast, the sketch information of face images models the global structures. Such global structures model the common structures of face images. Using global structures is


Figure 6: Visual comparisons with the local structure-based experiment on $21 \times 21$ motion blur kernel (top) and $51 \times 51$ motion blur kernel (bottom). "BInt" and "Int" denote the outputs of the sketch generator network before and after end-to-end fine-tuning, respectively. "Ori" denotes the result of image restoration. The experiment "$L_0$" (b)-(f) supervises the sketch generation network by the local structure extracted by the $L_0$ smoothing filter. The proposed method (g)-(j) predicts a finer structure, especially on key facial components, and restores a sharp image with fewer visual artifacts and more details. Although after end-to-end fine-tuning, there is a large intensity difference in predicted sketches, the positions of the primary salient edges are consistent, which is effective for face image deblurring. Please see manuscripts for more details.

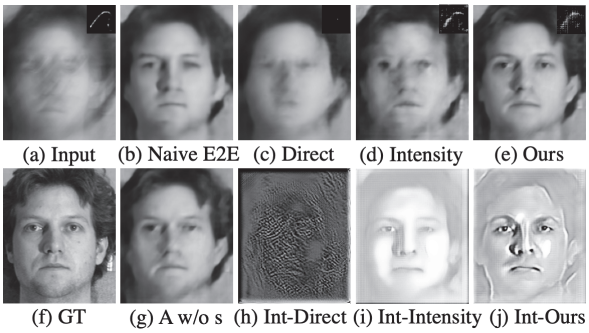


Figure 7: Ablation study. "Int" denotes the output of the sketch generator network. The top right corners of (a) and (c)(d)(e) are the ground truth kernel and the estimated ones, respectively. "Naive E2E" restores latent images without explicit kernel estimation. "Direct" trains the network without using sketch constraints. "Intensity" estimates kernels in the intensity space. "$A$ w/o s" only inputs the initial deblurred results into the de-artifact network. The proposed method predicts a sketch (j) and restores a higher-quality image (e). Please see manuscripts for more details.

able to help blur kernel estimation. To verify the above discussions, we use the sketch generator network to learn the local structures where the network outputs are supervised by the structures which are extracted by the $L_0$ smoothing filter (Xu et al. 2011; Pan et al. 2018). The results in Table 2 show that using sketch information achieves higher performance than the local structures extracted by $L_0$ smoothing filter. Figure 6 shows the visual comparisons on intermediate and final results. We note that the sketch-based and local structure-based networks can both perform well on small motion blur (top). However, due to limited salient edges in severely-blurred face images (bottom), the local structure-based network fails to extract sufficient structures, such as eyes and nose in Figure 6(d), which introduces severe ringing artifacts in the initial deblurred images, as shown in Figure 6(e). Even with a de-artifact network, these artifacts cannot be removed which is shown in Figure 6(f). As for the sketch-based network, although after end-to-end finetuning, there is a large intensity difference between Figure 6(g) and Figure 6(h), the positions of the primary salient edges are consistent, which are sufficient to achieve accurate kernel estimation and image restoration.

**Effectiveness of the Proposed Deblurring Process.** Although our method is trained in an end-to-end manner, it involves the blur kernel estimation and the latent image restoration. To demonstrate the effectiveness of the proposed framework, we compare the method that directly estimates the latent clear images without explicitly involving kernel estimation (denoted as "Naive E2E"). For fair comparisons, "Naive E2E" shares the same architecture with that of the sketch generator network, but it is only supervised by the latent ground truth. The comparisons shown in Table 2 and Figure 7 demonstrate that "Naive E2E" is less effective and provides overly smooth results on large motion blur. In contrast, as the proposed method hinges on the conventional deblurring process to deblur face images with kernel estimation and image restoration sequentially, it is robust to severely-blurred images and restores face images with more details.

**Kernel Estimation in Intensity v.s. Gradient Spaces.** The proposed kernel estimation step is performed in the gradient space. We also conduct an experiment about the kernel estimation using the intensity space. The results in Table 2("Intensity") and Figure 7(d)(i) indicate that estimating blur kernels in the gradient space is more effective for face image deblurring.

**Effectiveness of the Sketch in the De-Artifact Network.** To better preserve the structure information from the generated sketches as shown in Figure 7(j), the sketches and the initial deblurred results are concatenated together and fed

into the proposed de-artifact network. In order to validate the effectiveness of this concatenation, we compare the method only using the initial deblurred results as the input of the de-artifact network (denoted as "$A$ w/o s"). The final deblurred results shown in Table 2 and Figure 7(g) indicate that using sketches is able to facilitate the image restoration and thus generates the images with finer details and edges.

## 5 Conclusions

In this work, we have proposed to learn the sketches of face images to solve face image deblurring. We have shown that using the sketches of face images is able to facilitate the motion blur kernel estimation. The whole framework is motivated by the conventional deblurring process, which explicitly involves the kernel estimation and latent clear image restoration steps. Benefiting from this design, the proposed method is able to handle the face images with significant blur. By training the proposed network in an end-to-end manner, the proposed algorithm is able to restore clear images. Experiments on the synthetic dataset and real images demonstrate that the proposed method performs favorably against state-of-the-art deblurring approaches.

## 6 Acknowledgments

This project was supported by the National Natural Science Foundation of China (Nos. 61271375, 61872421, 61922043) and the Natural Science Foundation of Jiangsu Province (No. BK20180471).

## References

Boracchi, G., and Foi, A. 2012. Modeling the performance of image restoration from motion blur. *TIP*.

Chen, C.; Liu, W.; Tan, X.; and Wong, K.-Y. K. 2018. Semi-supervised learning for face sketch synthesis in the wild. *ACCV*.

Cho, S., and Lee, S. 2009. Fast motion deblurring. *TOG*.

Fergus, R.; Singh, B.; Hertzmann, A.; Roweis, S. T.; and Freeman, W. T. 2006. Removing camera shake from a single photograph. In *TOG*.

Gong, D.; Tan, M.; Zhang, Y.; Van den Hengel, A.; and Shi, Q. 2016. Blind image deconvolution by automatic gradient activation. In *CVPR*.

Gross, R.; Matthews, I.; Cohn, J.; Kanade, T.; and Baker, S. 2010. Multi-pie. *Image and Vision Computing*.

Hörmander, F. H. N. H. L.; Totaro, N. S. B.; and Waldschmidt, A. V. M. 2006. Grundlehren der mathematischen wissenschaften 332.

Isola, P.; Zhu, J.-Y.; Zhou, T.; and Efros, A. A. 2017. Image-to-image translation with conditional adversarial networks. In *CVPR*.

Kazemi, V., and Sullivan, J. 2014. One millisecond face alignment with an ensemble of regression trees. In *CVPR*.

King, D. E. 2009. Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research* 1755–1758.

Kingma, D. P., and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Krishnan, D.; Tay, T.; and Fergus, R. 2011. Blind deconvolution using a normalized sparsity measure. In *CVPR*.

Kruse, J.; Rother, C.; and Schmidt, U. 2017. Learning to push the limits of efficient fft-based image deconvolution. In *ICCV*.

Kumar, N.; Berg, A. C.; Belhumeur, P. N.; and Nayar, S. K. 2009. Attribute and simile classifiers for face verification. In *ICCV*.

Kupyn, O.; Budzan, V.; Mykhailych, M.; Mishkin, D.; and Matas, J. 2018. Deblurgan: Blind motion deblurring using conditional adversarial networks. In *CVPR*.

Le, V.; Brandt, J.; Lin, Z.; Bourdev, L.; and Huang, T. S. 2012. Interactive facial feature localization. In *ECCV*.

Liu, Z.; Luo, P.; Wang, X.; and Tang, X. 2015. Deep learning face attributes in the wild. In *ICCV*.

Nah, S.; Hyun Kim, T.; and Mu Lee, K. 2017. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *CVPR*.

Pan, J.; Hu, Z.; Su, Z.; and Yang, M.-H. 2014. Deblurring face images with exemplars. In *ECCV*.

Pan, J.; Sun, D.; Pfister, H.; and Yang, M.-H. 2016. Blind image deblurring using dark channel prior. In *CVPR*.

Pan, J.; Ren, W.; Hu, Z.; and Yang, M.-H. 2018. Learning to deblur images with exemplars. *TPAMI*.

Paszke, A.; Gross, S.; Chintala, S.; Chanan, G.; Yang, E.; DeVito, Z.; Lin, Z.; Desmaison, A.; Antiga, L.; and Lerer, A. 2017. Automatic differentiation in pytorch.

Schuler, C. J.; Hirsch, M.; Harmeling, S.; and Schölkopf, B. 2016. Learning to deblur. *PAMI*.

Shen, Z.; Lai, W.-S.; Xu, T.; Kautz, J.; and Yang, M.-H. 2018. Deep semantic face deblurring. In *CVPR*.

Son, H., and Lee, S. 2017. Fast non-blind deconvolution via regularized residual networks with long/short skip-connections. In *ICCP*.

Tang, X., and Wang, X. 2003. Face sketch synthesis and recognition. In *ICCV*.

Tao, X.; Gao, H.; Shen, X.; Wang, J.; and Jia, J. 2018. Scale-recurrent network for deep image deblurring. In *CVPR*.

Wang, X., and Tang, X. 2008. Face photo-sketch synthesis and recognition. *PAMI*.

Xu, L., and Jia, J. 2010. Two-phase kernel estimation for robust motion deblurring. In *ECCV*.

Xu, L.; Lu, C.; Xu, Y.; and Jia, J. 2011. Image smoothing via l 0 gradient minimization. In *TOG*.

Xu, X.; Pan, J.; Zhang, Y.-J.; and Yang, M.-H. 2018. Motion blur kernel estimation via deep learning. *TIP*.

Xu, L.; Zheng, S.; and Jia, J. 2013. Unnatural l0 sparse representation for natural image deblurring. In *CVPR*.

Zhang, J.; Pan, J.; Ren, J.; Song, Y.; Bao, L.; Lau, R. W.; and Yang, M.-H. 2018. Dynamic scene deblurring using spatially variant recurrent neural networks. In *CVPR*.

Zoran, D., and Weiss, Y. 2011. From learning models of natural image patches to whole image restoration. In *ICCV*.