



Blind Image Deblurring via Deep Discriminative Priors

Lerenhan Li^{1,2} · Jinshan Pan³ · Wei-Sheng Lai² · Changxin Gao¹ · Nong Sang¹ · Ming-Hsuan Yang² 

Received: 22 June 2018 / Accepted: 24 December 2018 / Published online: 9 January 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

We present an effective blind image deblurring method based on a data-driven discriminative prior. Our work is motivated by the fact that a good image prior should favor sharp images over blurred ones. In this work, we formulate the image prior as a binary classifier using a deep convolutional neural network. The learned prior is able to distinguish whether an input image is sharp or not. Embedded into the maximum a posterior framework, it helps blind deblurring in various scenarios, including natural, face, text, and low-illumination images, as well as non-uniform deblurring. However, it is difficult to optimize the deblurring method with the learned image prior as it involves a non-linear neural network. In this work, we develop an efficient numerical approach based on the half-quadratic splitting method and gradient descent algorithm to optimize the proposed model. Furthermore, we extend the proposed model to handle image dehazing. Both qualitative and quantitative experimental results show that our method performs favorably against the state-of-the-art algorithms as well as domain-specific image deblurring approaches.

Keywords Blind image deblurring · Deep learning · Discriminative prior

Communicated by Svetlana Lazebnik.

✉ Nong Sang
nsang@hust.edu.cn

Lerenhan Li
lrhli@hust.edu.cn

Jinshan Pan
sdluran@gmail.com

Wei-Sheng Lai
wlai24@ucmerced.edu

Changxin Gao
cgao@hust.edu.cn

Ming-Hsuan Yang
mhyang@ucmerced.edu

¹ National Key Laboratory of Science and Technology on Multispectral Information Processing, School of Automation, Huazhong University of Science and Technology, Wuhan, China

² Electrical Engineering and Computer Science, University of California, Merced, CA, USA

³ School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China

1 Introduction

Blind image deblurring aims to recover a latent image from a blurred input, which is a classical problem in computer vision. When the blur is spatially invariant, the blurring process is usually modeled by

$$B = I \otimes k + n, \quad (1)$$

where \otimes denotes the convolution operator, B , I , k and n denote the blurred image, latent sharp image, blur kernel, and noise, respectively. Recovering I and k from a blurred image B is an ill-posed problem as there exist infinite solutions to satisfy (1). Therefore, additional constraints and prior knowledge on both blur kernels and images are required.

The main success of the recent blind image deblurring methods mainly stems from the development of effective image priors and edge-prediction strategies. Edge-prediction based methods (Cho and Lee 2009; Xu and Jia 2010) select strong edges for estimating blur kernels, which do not perform well when such visual cues are not available in the input image. To alleviate the issues with heuristic edge selections, numerous algorithms based on natural image priors have been proposed, including normalized sparsity (Krishnan et al. 2011), L_0 gradients (Xu et al. 2013) and dark channel

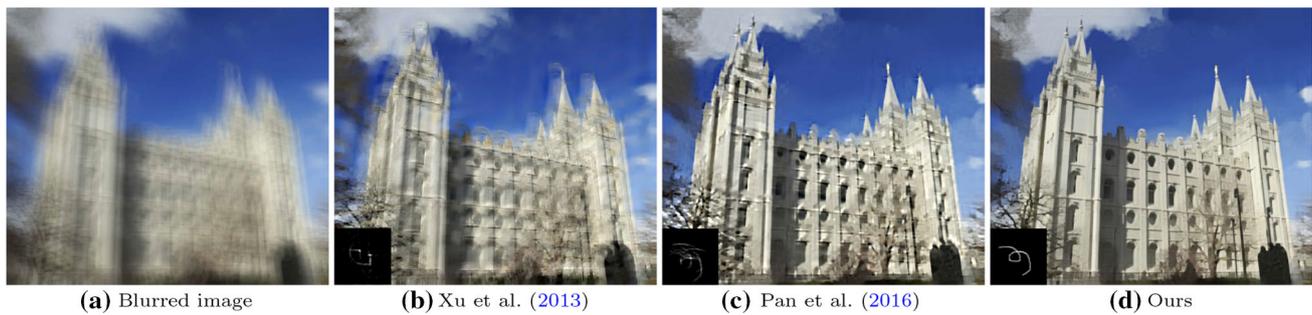


Fig. 1 A challenging blurred image with significant blur. We propose a discriminative image prior learned by a deep binary classification network for image deblurring. For the blurred image B in (a) and its corresponding sharp image I , we obtain $\frac{\|\nabla I\|_0}{\|\nabla B\|_0} = 0.85$, $\frac{\|D(I)\|_0}{\|D(B)\|_0} = 0.82$ and $\frac{f(I)}{f(B)} = 0.03$, where ∇ , $D(\cdot)$, $\|\cdot\|_0$ and $f(\cdot)$ denote the gradient

prior (Pan et al. 2016). These algorithms perform effectively on generic natural images but do not generalize well to specific scenarios, such as text (Pan et al. 2014b), face (Pan et al. 2014a) and low-illumination images (Hu et al. 2014). As most priors are hand-crafted and based on limited observations of specific scene statistics, these approaches do not generalize well to handle various types of images. Therefore, it is of great interest to develop a general image prior to deal with different scenarios in the wild.

In this work, we propose a blind deblurring algorithm based on a data-driven image prior. We formulate the image prior as a binary classifier to distinguish sharp images from blurred ones. Specifically, we first train a deep convolutional neural network (CNN) to classify blurred (labeled as 1) and sharp (labeled as 0) images. To handle arbitrary image sizes in the coarse-to-fine framework, we adopt a global average pooling layer (Lin et al. 2013) in the CNN. In addition, we use a multi-scale training strategy to ensure the classifier is robust to different input image sizes. We then take the learned CNN classifier as a regularization term with respect to latent images in the optimization framework. Figure 1 shows an example that the proposed image prior is more discriminative (i.e., has a lower ratio between the response of blurred and sharp images) than the state-of-the-art hand-crafted prior (Pan et al. 2016).

While the intuition of this work is straightforward, in practice it is difficult to optimize the deblurring method with the learned image prior as a non-linear CNN is involved. To address this issue, we develop an efficient numerical algorithm based on the half-quadratic splitting method and gradient descent approach. The proposed algorithm converges quickly in practice and can be applied to different scenarios as well as non-uniform deblurring.

The main contributions of this work are summarized as follows:

operator (Xu et al. 2013), the dark channel (Pan et al. 2016), L_0 norm (Pan et al. 2016; Xu et al. 2013) and the proposed classifier, respectively. The proposed prior is more discriminative than the hand-crafted priors, thereby leading to better deblurred results. A larger ratio indicates that the prior responses for B and I are closer and cannot be well separated

- We propose an effective discriminative image prior for blind image deblurring, which can be learned via a deep CNN classifier. To ensure that the proposed prior (i.e., classifier) can handle images of different sizes, we use the global average pooling and multi-scale training strategy to train the proposed CNN.
- We use the learned classifier as a regularization term of the latent image and develop an efficient optimization algorithm to solve the deblurring model.
- We demonstrate that the proposed algorithm performs favorably against the state-of-the-art methods on both the natural image deblurring benchmarks and domain-specific deblurring tasks.
- We show the proposed method can be directly extended to the non-uniform deblurring and image dehazing.

2 Related Work

Recent years have witnessed significant advances in single image deblurring. In this section, we focus our discussion on the most relevant deblurring methods based on optimization and learning.

2.1 Optimization-Based Methods

Optimization-based approaches can be categorized into implicit and explicit edge enhancement methods. The implicit edge enhancement approaches focus on developing effective image priors to favor sharp images over blurred ones. Representative image priors include sparse gradients (Fergus et al. 2006; Levin et al. 2009; Xu and Jia 2010), normalized sparsity (Krishnan et al. 2011), color-line (Lai et al. 2015), L_0 gradients (Xu et al. 2013), patch priors (Sun et al. 2013), and self-similarity (Michaeli and Irani 2014).

Although these image priors have been successfully applied to deblur natural images, these methods are less effective for specific scenarios, such as text, face, and low-illumination images. As the statistics of these domain-specific images are significantly different from those of natural scenes, Pan et al. (2014b) propose the L_0 -regularized prior on both intensity and gradient for deblurring text images. Hu et al. (2014) detect the light streaks from the low-light images for estimating blur kernels. Recently, Pan et al. (2016) propose a dark channel prior for deblurring natural images, which can be applied to face, text and low-illumination images as well. However, the dark channel prior is less effective when no dark pixels can be extracted from an image. Yan et al. (2017) propose to incorporate a bright channel prior with the dark channel prior to improve the robustness of the deblurring algorithm.

Image priors can be considered as non-reference image quality measurements, which can distinguish whether an image is blurry or sharp based on the prior response. Zhao et al. (2015) conduct a user study to measure image sharpness and evaluate the performance of existing image quality measurements. Leclaire and Moisan (2013) propose a simplified sharpness index based on the Fourier transform and use it for blind image deblurring.

While those algorithms demonstrate the state-of-the-art performance, most priors are hand-crafted and designed under limited observations. In this work, we propose to learn a discriminative prior using a deep CNN. The data-driven prior is able to distinguish blurred and sharp inputs for various types of images and serves as an effective prior in the deblurring framework.

2.2 Learning-Based Methods

With the success of deep CNNs on high-level vision problems (He et al. 2016; Long et al. 2015), several approaches have adopted deep CNNs in image restoration tasks, including super-resolution (Dong et al. 2014; Kim et al. 2016; Lai et al. 2017), denoising (Mao et al. 2016) and JPEG deblocking (Dong et al. 2015). A few methods (Schuler et al. 2016; Sun et al. 2015; Yan and Shao 2016; Chakrabarti 2016) first adopt deep networks to predict blur kernels and then use conventional non-blind deconvolution algorithms to recover sharp images. However, due to the complexity and variety of blur kernels in the wild, estimating blur kernels from deep networks is less robust and not effective compared to conventional optimization-based methods.

Recently, several end-to-end methods are proposed for handling dynamic scenes (Nah et al. 2017; Noroozi et al. 2017; Kupyń et al. 2018; Tao et al. 2018), text images (Hradiš et al. 2015), and face images (Jin et al. 2018; Shen et al. 2018). These approaches bypass the blur kernel estimation step and directly restore a sharp image from a blurred input. However,

the state-of-the-art end-to-end CNN-based methods do not perform competitively against conventional optimization-based algorithms when the images suffer from significant motion blur.

We note several approaches have been developed to train deep CNNs as image priors or denoisers for non-blind deconvolution (Sreehari et al. 2016; Zhang et al. 2017a, b; Bigdeli et al. 2017). However, these methods focus on restoring fine details and do not perform well on reconstructing strong edges for blind deconvolution. In this work, we take advantage of both the conventional optimization-based framework and the discriminative strength of deep CNNs. We embed the learned CNN prior into the coarse-to-fine optimization framework for solving the blind image deblurring problem.

2.3 Single Image Dehazing

Several single image dehazing methods (He et al. 2011; Berman et al. 2016) estimate the transmission maps and atmospheric light based on hand-crafted image priors, while recent approaches (Ren et al. 2016; Li et al. 2017a) train deep networks for the estimation. However, existing methods recover the sharp image directly from an element-wise division between the input image, atmospheric light, and estimated transmission map, which tends to introduce color distortion. To address this problem, we first learn an image prior by training our classifier on a set of hazy/clean images. We then apply an optimization method to restore the clean image based on the estimated transmission map, atmospheric light, and the learned image prior.

3 Learning a Data-Driven Image Prior

In this section, we describe the motivation, network design, loss function, and implementation details of the proposed image prior.

3.1 Motivation

Based on the image formation model (1), the optimization-based blind image deblurring methods solve the following problem:

$$\min_{I, k} \|I \otimes k - B\|_2^2 + \gamma \|k\|_2^2 + \lambda p(I), \quad (2)$$

where $p(I)$ is the regularization term (also known as image prior), and $\|\cdot\|_2^2$ denotes an L_2 norm. In addition, γ and λ are regularization weights. The key to the success of this framework lies in the latent image prior $p(I)$. An effective image prior should favor sharp images over blurred ones. As such, an effective prior should have lower energies to sharp images

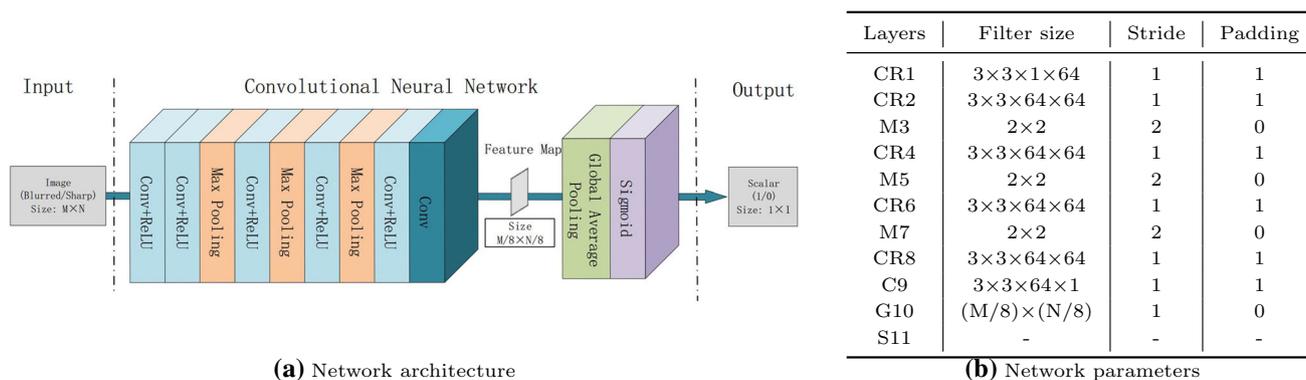


Fig. 2 Architecture and parameters of the proposed binary classification network. We adopt a global average pooling layer instead of a fully-connected layer to handle different sizes of input images. “CR” denotes the convolutional layer followed by a ReLU non-linear func-

tion, “M” denotes the max-pooling layer, “C” denotes the convolutional layer, “G” denotes the global average pooling layer and “S” denotes the non-linear sigmoid function

than blurred ones when used in solving (2). This observation motivates us to learn a data-driven discriminative prior via binary classification. To this end, we train a deep CNN by predicting blurred images as positive (labeled as 1) and sharp images as negative (labeled as 0) samples. Compared with the state-of-the-art image priors (Xu et al. 2013; Pan et al. 2016), the assumption of our prior is simple and straightforward without using any hand-crafted features or functions.

3.2 Binary Classification Network

We train a deep CNN to determine the probability of an input being a blurry image patch. As we aim to embed the network as a prior into the coarse-to-fine optimization framework, this model should be able to handle images of different sizes. Instead of using fully-connected layers, we adopt the global average pooling layer (Lin et al. 2013) which converts feature maps of different sizes into a single scalar before the sigmoid layer. In addition, there is no additional parameter in the global average pooling layer, which alleviates the overfitting problem. Figure 2 shows the architecture and detail parameters of our binary classification network.

3.3 Loss Function

We denote the input image by x and the network parameters to be optimized by θ . The deep network learns a mapping function $f(x; \theta) = P(x \in \text{Blurred}|x)$ that predicts the probability of the input image being a blurred one. We optimize the network via the binary cross entropy loss function:

$$L(\theta) = -\frac{1}{N} \sum_{i=1}^N \hat{y}_i \log(y_i) + (1 - \hat{y}_i) \log(1 - y_i), \quad (3)$$

where N is the number of training samples in a batch, $y_i = f(x_i)$ is the output of the classifier, and \hat{y}_i is the label of the input image. We assign $\hat{y} = 1$ for blurred images and $\hat{y} = 0$ for sharp images.

3.4 Training Details

We sample 500 sharp images from the dataset of Huiskes and Lew (2008), which contains natural, manmade scene, face, low-illumination and text images. We then use the method of Boracchi and Foi (2012) to generate 200 random blur kernels with the size ranging from 7×7 pixels to 51×51 pixels, and synthesize blurred images by convolving the sharp images with blur kernels and adding the Gaussian noise with $\sigma = 0.01$. In total, we generate 100,000 blurred images as the positive samples (labeled as 1) for training. We also sample another 50,000 sharp images as the negative samples (labeled as 0). During training, we randomly crop 200×200 patches from the training images. In order to make the classifier more robust to different image size, we adopt a multi-scale training strategy by randomly resizing the input images between one quarter and full scale.

We implement our network using the MatConvNet (Vedaldi and Lenc 2015) toolbox. We use the Xavier method (Glorot and Bengio 2010) to initialize the network parameters and use the Stochastic Gradient Descent (SGD) method for optimizing the network. In this work, we use batch size of 50, momentum of 0.9 and weight decay of 10^{-4} . The learning rate is set to 0.001 and decreased by a factor of 5 for every 50 epochs.

3.5 Effectiveness of the Binary Classifier

We train the binary classification network to predict the probability of an input image being blurred (from 0 to 1). We first

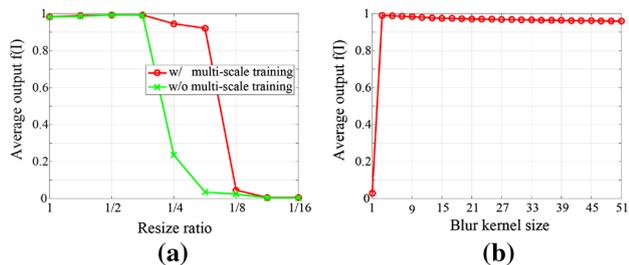


Fig. 3 Robustness of the proposed CNN prior as the classifier. **a** Average output of different resize ratios on the dataset (Köhler et al. 2012), **b** average output of different sizes of blur kernels. The multi-scale training strategy encourages the classifier to be more robust to different sizes of images

use images of 200×200 pixels for training. We evaluate the classification accuracy using the images from the dataset by Köhler et al. (2012), where the image size is 800×800 pixels. To analyze the performance of the classifier on different sizes of images, we downsample each image by a ratio between $[1, 1/16]$ and plot the average output in Fig. 3a (green curve). We use bicubic interpolation as the downsampling operator. When the size of test images is large or close to the size of training images, the output is near 1. However, the average

output drops significantly when images are downsampled by more than 4 times. As the downsampling operation reduces the blur effect, it is increasingly difficult for the classifier to distinguish blurred and sharp images.

To address this issue, we use a multi-scale training strategy by randomly downsampling each batch of images after the blur process between a factor of 1 and 4. As shown in the red curve of Fig. 3a, the classifier performs more robustly to different size of input images with this multi-scale training strategy, and fits well in the proposed coarse-to-fine optimization framework.

We further analyze whether the proposed prior performs robustly to the blur magnitude (i.e., blur kernel size). We synthesize 240 motion blur kernels with the size ranging from 1×1 pixels to 51×51 pixels (i.e., 10 kernels for each size, and blur kernels with the size 1×1 are delta function such that the convolved images are sharp). We then plot the average output of our binary classifier on an image of 800×800 pixels. Figure 3b shows that the proposed discriminative prior performs robustly to a wide range of the blur kernel size.

Figure 4 shows the activation of one feature map from the C9 layer (i.e., the last convolutional layer before the global average pooling) in our classification network. The output of the C9 layer is a single-channel feature map and is equivalent

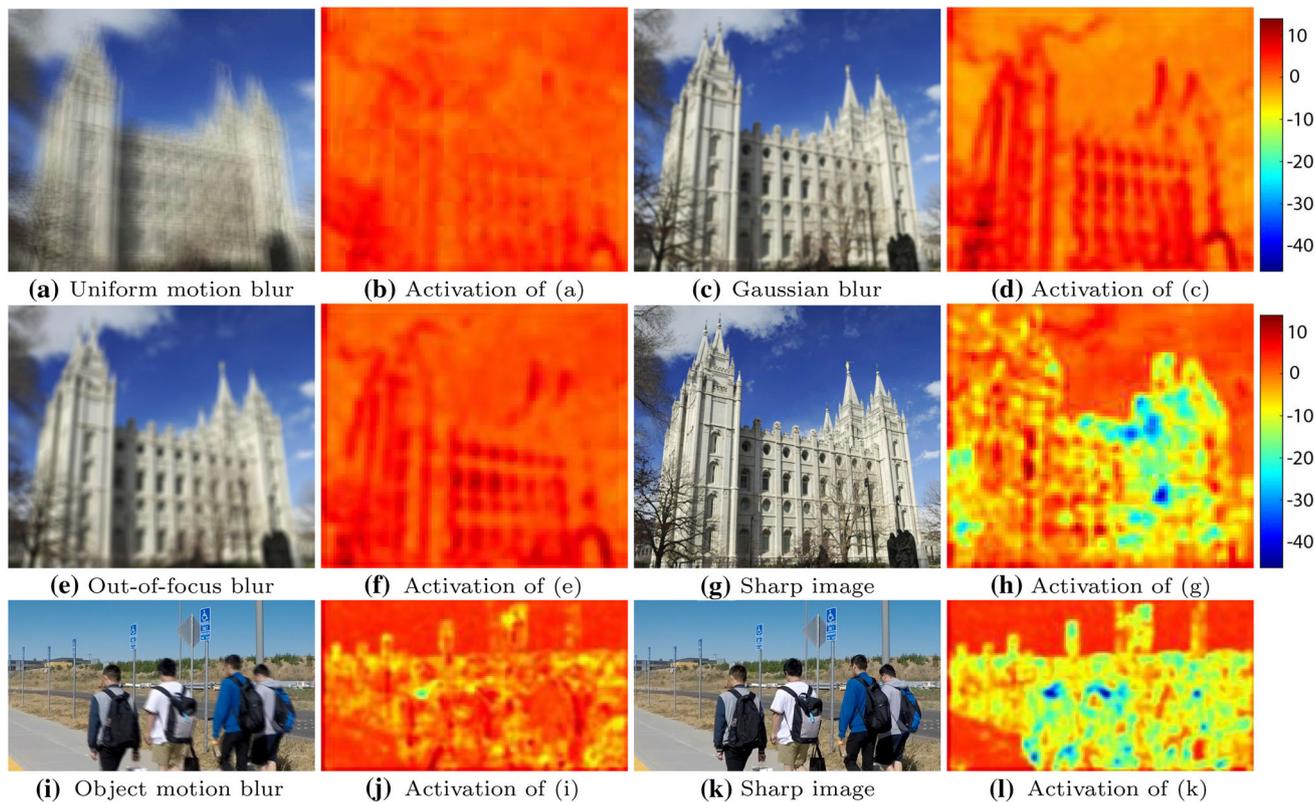


Fig. 4 Activation of feature maps in our binary classification network. We show the activation from the C9 layer. The sharp image (or the sharp region in a non-uniform blurred image) has much lower responses than that of the blurred images (region). The proposed network is effective to all types of image blur

Table 1 Average classifier response to different types of blurred images

Blur type	Blurred image	Sharp image
Gaussian	0.99	0.01
Out-of-focus	0.99	0.01

lent to the response of the CNN image prior. To evaluate the robustness of the classifier to different blur types, we generate Out-of-focus and Gaussian blur images by convolving the sharp image with a disk and Gaussian kernel, respectively. We also provide a non-uniform blurred image, which is caused by object motion, as input. All of these types of blurred images do not appear in the training dataset. As shown in Figure 4, the proposed classifier is effective to classify different types of blurred images. We also perform quantitative evaluations on Gaussian and Out-of-focus images. We generate 100 blurred images by convolving 100 sharp images with randomly chosen kernels for each type and output the response of the classifier to both blurred and sharp images. We randomly set the standard variation from 0.5 to 5 to generate the Gaussian blur kernel and the radius from 2 to 5 pixels to generate the Out-of-focus blur kernel (i.e., disk kernel). Table 1 shows the quantitative evaluations of the classifier on Gaussian and Out-of-focus blurred images. For a uniform blurred image, while the blurred image has a high response on the entire image, the activation of the sharp image has a much lower response except for smooth regions (e.g., sky). For a non-uniform blurred image (e.g., caused by object motion blur), the blurred image has a high response on the moving objects but lower response on the non-smooth background (i.e., sharp region).

Figure 4i–l demonstrate that the proposed prior is also effective on non-uniform blurred images under the assumption that non-uniform blurred images are locally-uniform blur. We discuss the extension to non-uniform deblurring in details in Sect. 5.

4 Blind Image Deblurring

We use the trained network as the latent image prior $p(\cdot)$ in (2). In addition, we adopt the L_0 gradient prior (Xu et al. 2013; Pan et al. 2016) as a regularization term. The proposed model for uniform deblurring is

$$\min_{I,k} \|I \otimes k - B\|_2^2 + \gamma \|k\|_2^2 + \mu \|\nabla I\|_0 + \lambda f(I), \quad (4)$$

where γ , μ , and λ are the weight parameters. We optimize (4) by solving the latent image I and blur kernel k alternatively.

Thus, we divide the problem (4) into I sub-problem:

$$\min_I \|I \otimes k - B\|_2^2 + \mu \|\nabla I\|_0 + \lambda f(I), \quad (5)$$

and k sub-problem:

$$\min_k \|I \otimes k - B\|_2^2 + \gamma \|k\|_2^2. \quad (6)$$

4.1 Solving I Sub-problem

In (5), both $f(I)$ and $\|\nabla I\|_0$ are non-convex, which make minimizing (5) computationally intractable. To tackle this issue, we adopt the half-quadratic splitting method (Xu et al. 2011) to introduce the auxiliary variables u and $g = (g_h, g_v)$ with respect to the image and corresponding gradients in horizontal and vertical directions, respectively. The energy function (5) can be rewritten as

$$\min_{I,g,u} \|I \otimes k - B\|_2^2 + \alpha \|\nabla I - g\|_2^2 + \beta \|I - u\|_2^2 + \mu \|g\|_0 + \lambda f(u), \quad (7)$$

where α and β are the penalty parameters. When α and β approach infinity, the solution of (7) becomes equivalent to that of (5). We solve (7) by minimizing I , g , and u alternatively and thus avoid directly minimizing the non-convex functions $f(I)$ and $\|\nabla I\|_0$.

We solve the latent image I by fixing g as well as u , and optimizing:

$$\min_I \|I \otimes k - B\|_2^2 + \alpha \|\nabla I - g\|_2^2 + \beta \|I - u\|_2^2, \quad (8)$$

which is a least squares optimization problem. The closed-form solution is:

$$I = F^{-1} \left(\frac{\overline{F(k)}F(B) + \beta F(u) + \alpha F_G}{\overline{F(k)}F(k) + \beta + \alpha F_D} \right), \quad (9)$$

where $F(\cdot)$ and $F^{-1}(\cdot)$ denote the Fourier and inverse Fourier transforms; and $\overline{F(\cdot)}$ is the complex conjugate operator. In (9),

$$F_G = \sum_{d \in \{h,v\}} \overline{F(\nabla_d)} F(g_d), \quad (10)$$

$$F_D = \sum_{d \in \{h,v\}} \overline{F(\nabla_d)} F(\nabla_d), \quad (11)$$

where $\nabla_h = [1, -1]$ and $\nabla_v = [1, -1]^T$ are the horizontal and vertical differential operators, respectively.

Given the latent image I , we solve g and u by:

$$\min_g \alpha \|\nabla I - g\|_2^2 + \mu \|g\|_0, \quad (12)$$

Algorithm 1 Blur Kernel Estimation

Input: Blurred Image B
Output: Intermediate latent image I and blur kernel k .
 1: initialize k with results from the coarser level
 2: **while** $i < \text{iter}_{\max}$ **do**
 3: solve for I by (7).
 4: solve for k by (16).
 5: $i \leftarrow i + 1$
 6: **end while**

$$\min_u \beta \|I - u\|_2^2 + \lambda f(u). \tag{13}$$

Since (12) is an element-wise minimization problem, we adopt the strategy of Pan et al. (2014b) to solve it:

$$g = \begin{cases} \nabla I, & \|\nabla I\|^2 \geq \frac{\mu}{\alpha}, \\ 0, & \text{otherwise.} \end{cases} \tag{14}$$

To solve (13), we use the back-propagation approach to compute the derivative of $f(\cdot)$. We update u using the gradient descent method:

$$u^{(s+1)} = u^{(s)} - \eta \left[\beta (u^{(s)} - I) + \lambda \frac{df(u^{(s)})}{du^{(s)}} \right], \tag{15}$$

where η is the step size and s denotes the iteration index.

4.2 Solving k Sub-problem

We estimate the blur kernel using the image gradients (Cho and Lee 2009; Pan et al. 2014b, 2016), which has been shown to be more effective than using the image intensity (Pan et al. 2018). We re-formulate (6) to:

$$\min_k \|\nabla I \otimes k - \nabla B\|_2^2 + \gamma \|k\|_2^2, \tag{16}$$

which can also be efficiently solved by the Fast Fourier Transform (FFT). We then set the negative elements in k to 0 and normalize k so that the sum of all elements is equal to 1.

4.3 Non-blind Deconvolution

We use an effective ringing suppression non-blind deconvolution algorithm (Pan et al. 2014b) to recover the latent image after the kernel is estimated. We first (Krishnan and Fergus 2009) to recover the latent image I_l . Then, we restore another latent image I_0 by solving the following minimization problem:

$$\min_I \|I \otimes k - B\|_2^2 + \mu \|\nabla I\|_0. \tag{17}$$

The final deblurred image \hat{I} is obtained by

$$\hat{I} = I_l - BF(I_l - I_0), \tag{18}$$

where BF represents the bilateral filter. As I_l may contain ringing artifacts and I_0 only contain the main structure of the image, applying the bilateral filter on the difference $I_l - I_0$ can be viewed as a step to suppress the ringing artifacts. Unless mentioned otherwise, we use the same non-blind deconvolution method to restore the final sharp images in all the experiments.

4.4 Implementation Details

We use the coarse-to-fine strategy with an image pyramid (Pan et al. 2014b, 2016) to optimize (4). At each pyramid level, we alternatively solve (5) and (16) for iter_{\max} iterations. We then upsample the estimated k and I as the initial solution at the next pyramid scale. We summarize the main steps for the blur kernel estimation in Algorithm 1. When performing the convolution and discrete Fourier transform, we pad the image periodically to handle the boundary condition. In all the experiments, we set $\lambda = \mu = 0.004$, $\gamma = 2$, and $\eta = 0.1$. To balance the accuracy and speed, we empirically set $\text{iter}_{\max} = 5$ and $s_{\max} = 10$.

5 Extension to Non-uniform Deblurring

The proposed discriminative image prior can be easily extended for non-uniform motion deblurring. Based on the geometric model of camera motion (Tai et al. 2011; Whyte et al. 2012), we represent a blurred image as the weighted sum of a latent sharp image under a set of geometry transformations:

$$B = \sum_t k_t H_t I + n, \tag{19}$$

where B , I and n are the blurred image, latent image and noise in the vector forms, respectively; t denotes the index of camera pose samples; k_t is the weight of the t th camera pose satisfying $k_t \geq 0$, $\sum_t k_t = 1$; and H_t denotes a matrix derived from the homography. We use the bilinear interpolation when applying H_t on a latent image I . The blur model (19) is then reduced to:

$$B = KI + n = Ak + n, \tag{20}$$

where $K = \sum_t k_t H_t$, $A = [H_1 I, H_2 I, \dots, H_t I]$ and $k = [k_1, k_2, \dots, k_t]^T$. We solve the non-uniform deblurring problem by alternatively minimizing:

Table 2 Datasets for performance evaluation of single image deblurring algorithms in this paper

Dataset	Synthetic/real	Image type	Latent images	Blur kernels	Blurred images	Non-blind deconvolution	Evaluation metric
Levin et al. (2009)	Synthetic	natural	4	8	32	Pan et al. (2014b)	ERSR
Köhler et al. (2012)	Real	Natural	4	12	48	Pan et al. (2014b)	PSNR
Sun et al. (2013)	Synthetic	Natural	80	8	640	Zoran and Weiss (2011)	ERSR
Pan et al. (2014a)	Synthetic	Face	60	8	480	Pan et al. (2014b)	PSNR
Pan et al. (2014b)	Synthetic	Text	15	8	120	Pan et al. (2014b)	PSNR

PSNR and ERSR denote peak signal-to-noise ratio and error-ratio success rate, respectively

$$\min_{\mathbf{I}} \|\mathbf{KI} - \mathbf{B}\|_2^2 + \lambda f(\mathbf{I}) + \mu \|\nabla \mathbf{I}\|_0 \quad (21)$$

and

$$\min_{\mathbf{k}} \|\mathbf{Ak} - \mathbf{B}\|_2^2 + \gamma \|\mathbf{k}\|_2^2. \quad (22)$$

Following the strategies to optimize (5), we rewrite (21) as

$$\min_{\mathbf{I}, \mathbf{u}, \mathbf{g}} \|\mathbf{KI} - \mathbf{B}\|_2^2 + \alpha \|\nabla \mathbf{I} - \mathbf{g}\|_2^2 + \beta \|\mathbf{I} - \mathbf{u}\|_2^2 + \mu \|\mathbf{g}\|_0 + \lambda f(\mathbf{u}), \quad (23)$$

where \mathbf{u} and \mathbf{g} denote the vector forms of u and g in (7), respectively. We use the same approach in (12) and (13) to solve \mathbf{u} and \mathbf{g} .

In order to solve (23) with respect to \mathbf{I} , we use the fast forward approximation method (Hirsch et al. 2011) and assume that the latent image \mathbf{I} is locally uniform. The non-uniform blur process is modeld as

$$\mathbf{B} \approx \mathbf{Z}_1^{-1} \sum_{p=1}^P \zeta_p^{-1} (F^{-1}(\text{diag}(F(\mathbf{Z}_a a_p)))) F(\zeta_p \text{diag}(m_p)) \mathbf{I}, \quad (24)$$

where $\zeta_p(\cdot)$ denotes the operator that crops the p th patch from an image, $\zeta_p^{-1}(\cdot)$ denotes the operator to paste the patch back to the original image, \mathbf{Z}_1 and \mathbf{Z}_a are the zero-padding matrix whose size matches the size of the vector from the summation, and $\text{diag}(\cdot)$ denotes a diagonal matrix with the element of vector x on the main diagonal. In (24), m_p is a window function and the size is the same as I , which satisfies $\sum_p m_p = 1$. In addition, a_p denotes the blur kernel in the p th patch and satisfies $a_p = \sum_t k_t b_t^p$, where t is the index of camera pose samples and b_t^p is the kernel basis in the p th patch. We determine the parameter t by counting the overlapped patches in the whole image. In this work, we sample local patches with a size of 31×31 and a stride of 15

pixels. Based on (24), the solution of (23) can be written as

$$\mathbf{I} = \sum_{p=1}^P \zeta_p^{-1} \left(F^{-1} \left(\frac{F_{KB} + \beta F(\zeta_p(u)) + \alpha F_{G'}}{F_{K'} + \beta + \alpha F_{D'}} \right) \right), \quad (25)$$

where F_{KB} , $F_{G'}$, $F_{K'}$, and $F_{D'}$ are defined by:

$$F_{KB} = \overline{F(\mathbf{Z}_a a_p)} F(\zeta_p \text{diag}(m_p)) \mathbf{B}, \quad (26)$$

$$F_{G'} = \sum_{d \in \{h, v\}} \overline{F(\zeta_p(\nabla_d))} F(\zeta_p(g_d)), \quad (27)$$

$$F_K = \overline{F(\mathbf{Z}_a a_p)} F(\mathbf{Z}_a a_p), \quad (28)$$

$$F_{D'} = \sum_{d \in \{h, v\}} \overline{F(\zeta_p(\nabla_d))} F(\zeta_p(\nabla_d)). \quad (29)$$

We use the steps in Algorithm 1 and replace (25) with (7) to solve (21) and (22).

6 Experimental Results

We evaluate the proposed algorithm on the natural scene datasets (Levin et al. 2009; Köhler et al. 2012; Sun et al. 2013) as well as text (Pan et al. 2014b), and face (Pan et al. 2014a) image datasets. Table 2 shows a list of the evaluated datasets in our experiments. All the experiments are carried out on a desktop computer with an Intel Core i7-3770 processor and 32 GB RAM. The source code and the datasets will be available on the project websites at https://sites.google.com/view/lerenhanli/homepage/learn_prior_deblur.

6.1 Natural Images

We first evaluate the proposed algorithm on natural image dataset of Köhler et al. (2012), which contains 4 latent images and 12 blur kernels. We compare with 5 generic image deblurring methods (Cho and Lee 2009; Xu and Jia 2010; Whyte et al. 2012; Pan et al. 2016; Yan et al. 2017; Kupyn et al. 2018; Tao et al. 2018). Figure 5 shows the deblurred results of one example in which our method generates sharper images with fewer ringing artifacts.

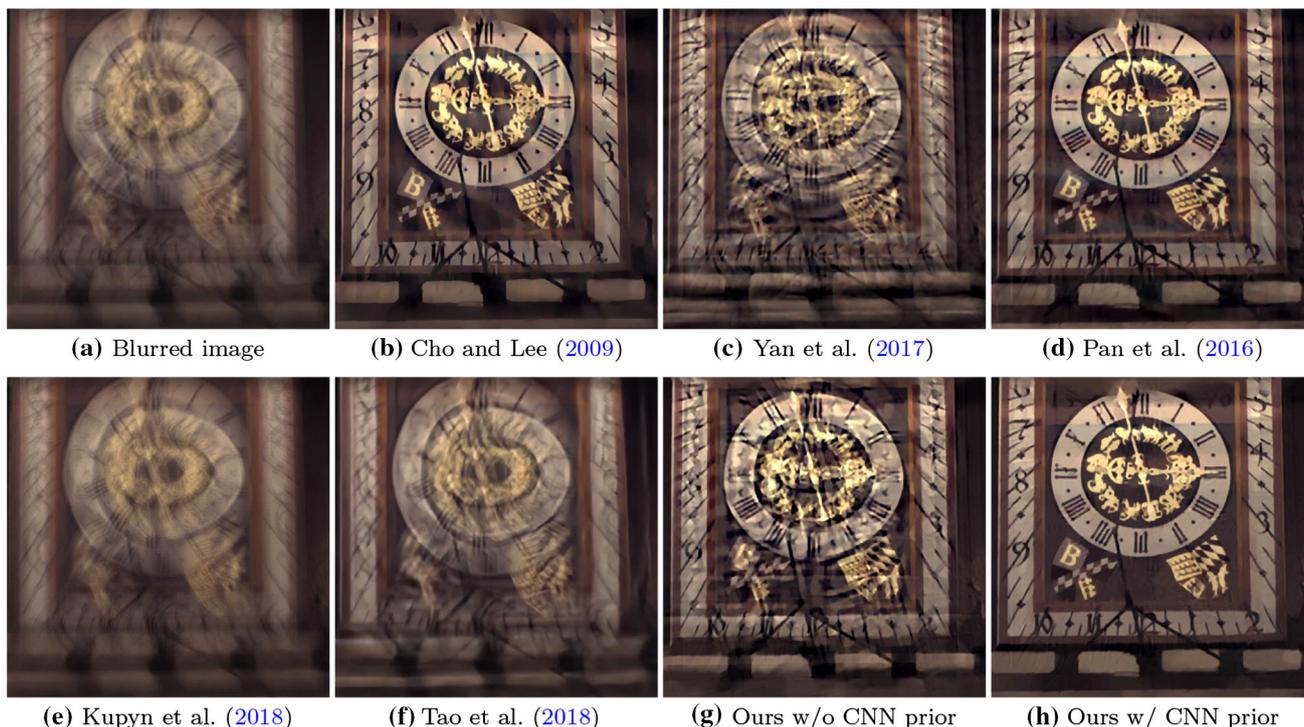


Fig. 5 A challenging example from the dataset (Köhler et al. 2012). The proposed algorithm restores visually more pleasing results with fewer ringing artifacts

We use the same protocol as Köhler et al. (2012) to compute the PSNR by comparing each deblurred image with 199 sharp images captured along the same camera motion trajectory. As shown in Fig. 6a, our method achieves the highest PSNR on average.

Next, we evaluate our algorithm on the deblurring dataset provided by Sun et al. (2013), which consists of 80 sharp images and 8 blur kernels from Levin et al. (2009). We compare with the 6 optimization-based deblurring methods (Levin et al. 2011; Krishnan et al. 2011; Xu et al. 2013; Xu and Jia 2010; Sun et al. 2013; Pan et al. 2016) and one CNN-based method (Chakrabarti 2016). We first compute the Sum of Squared Differences (SSD) between the deblurred results and the ground-truth images. Considering the error is affected by the kernel size, we measure the ratio between the deblurred error with the estimated kernel and that with the ground-truth kernel. We then plot the cumulative histogram in Fig. 6b. The percentage of the results with error ratios below a threshold is defined as the success rate. For fair comparisons, we apply the same non-blind deconvolution (Zoran and Weiss 2011) to restore the latent images. The proposed method performs competitively against the state-of-the-art algorithms.

We also compare our method with existing algorithms (Krishnan et al. 2011; Xu et al. 2013; Pan et al. 2016; Kupyn et al. 2018; Tao et al. 2018) on real-world blurred images. Here we use the same non-blind deconvolution algorithm (Pan et al. 2014b) for fair comparisons. As shown in Fig. 7,

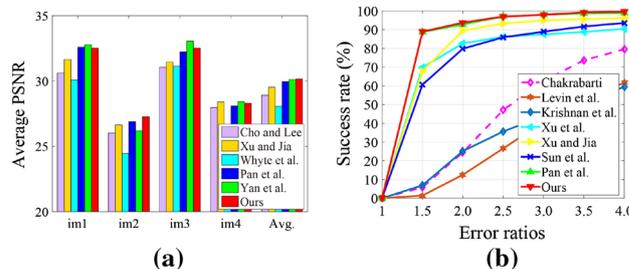


Fig. 6 Quantitative evaluations on the benchmark datasets. **a** Köhler et al. (2012) and **b** Sun et al. (2013)

our method generates sharper images with fewer artifacts and performs comparably to the state-of-the-art method (Pan et al. 2016).

6.2 Domain-Specific Images

We evaluate our algorithm on the text image dataset (Pan et al. 2014b) which consists of 15 sharp text images and 8 blur kernels from Levin et al. (2009). We show the average PSNR in Table 3. While the text deblurring approach (Pan et al. 2014b) achieves the highest PSNR, the proposed method performs favorably against the state-of-the-art generic deblurring algorithms (Cho and Lee 2009; Xu and Jia 2010; Levin et al. 2011; Xu et al. 2013; Pan et al. 2016). Figure 8 shows the deblurred results on a blurred text image.

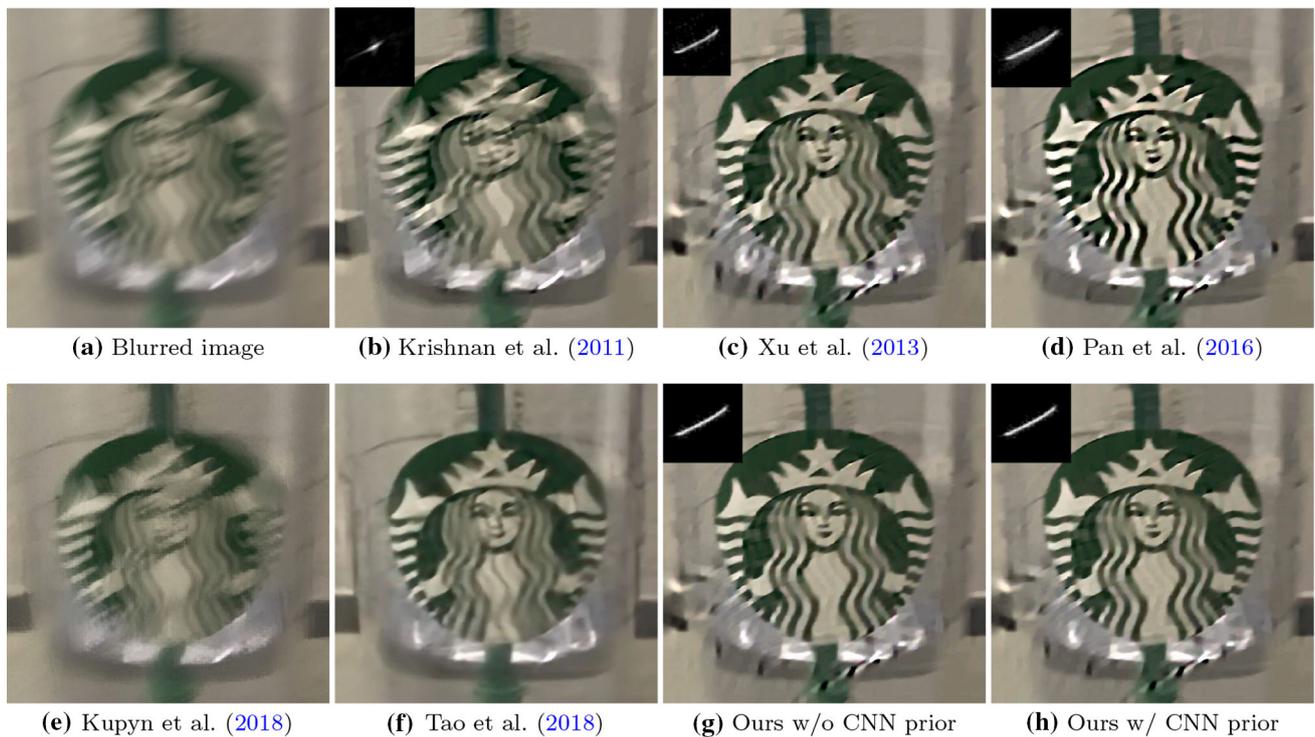


Fig. 7 Deblurred results on a real blurred image. Our result is sharper with fewer artifacts. (e Kupyn et al. (2018) and f Tao et al. (2018) are end-to-end methods that do not have kernel estimation)

Table 3 Quantitative evaluations on the text image dataset (Pan et al. 2014b)

Methods	Average PSNR
Cho and Lee (2009)	23.80
Xu and Jia (2010)	26.21
Levin et al. (2011)	24.90
Xu et al. (2013)	26.21
Pan et al. (2016)	27.94
Pan et al. (2014b)	28.80
Ours	<u>28.10</u>

Our method performs favorably against generic image deblurring approaches and is comparable to the text deblurring method (Pan et al. 2014b)

Bold value indicates the best performance

The proposed method generates a much sharper image with sharper texts.

Figure 9 shows an example of the low-illumination image from the dataset of Pan et al. (2016). Due to the influence of large saturated regions, the generic image deblurring method (Xu et al. 2013) does not generate sharp images. In contrast, our method generates a comparable result with Hu et al. (2014), which is specially designed for the low-illumination images.

Figure 10 shows the deblurred results on a face image. We note that the proposed method learns a *generic* image prior and effectively deblurs images on different domains. Compared with the state-of-the-art methods (Xu et al. 2013; Yan et al. 2017; Shen et al. 2018), the proposed algorithm generates the image with fewer ringing artifacts.



Fig. 8 Deblurred results on a text image. Our method generates sharper deblurred image with more sharper characters than the state-of-the-art text deblurring algorithm (Pan et al. 2014b)

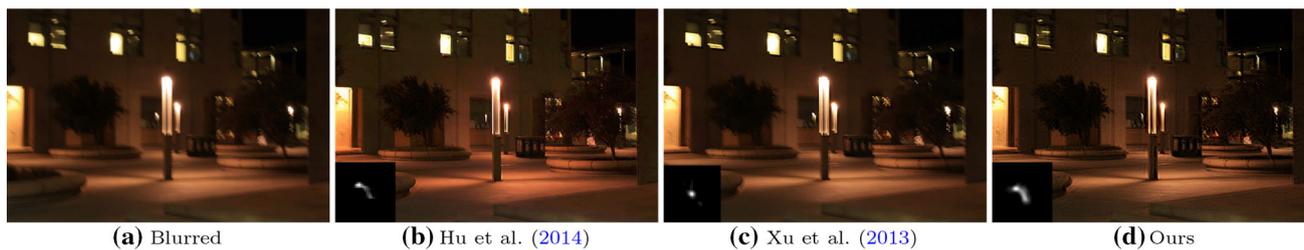


Fig. 9 Deblurred results on a low-illumination image. Our method yields comparable results to Hu et al. (2014), which is specially designed for deblurring low-illumination images

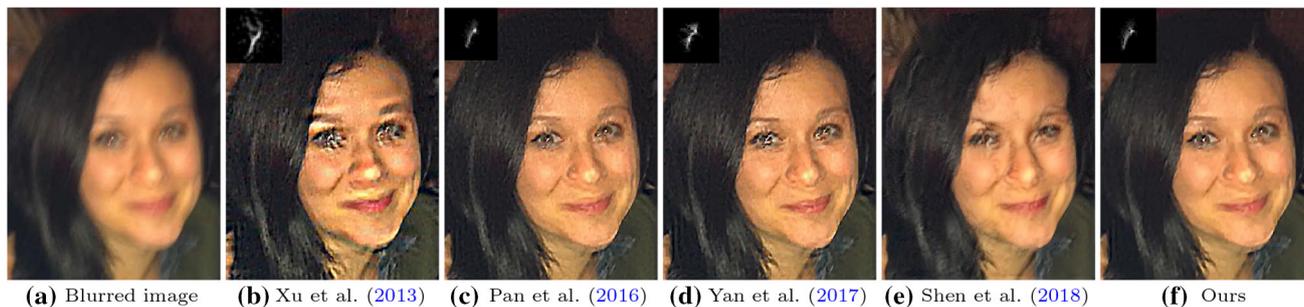


Fig. 10 Deblurred results on a face image. Our method generates visually more pleasing results. (e Shen et al. (2018) is an end-to-end method that does not have kernel estimation)

6.3 Non-uniform Deblurring

We evaluate the performance of our method and existing non-uniform deblurring algorithms on the dataset by Lai et al. (2016), which contains 4 camera motion trajectories and 25 sharp images. As shown in Table 4, the proposed algorithm performs favorably against the state-of-the-art non-uniform image deblurring methods (Whyte et al. 2012; Xu et al. 2013; Pan et al. 2016; Kupyn et al. 2018; Tao et al. 2018) on the dataset by Lai et al. (2016). We show the results of non-uniform deblurring in Fig. 11. The proposed performs well against the state-of-the-art-methods and restores images with sharp edges and sharp textures.

Table 4 Quantitative evaluations on the non-uniform deblurring dataset (Lai et al. 2016)

Methods	Average PSNR
Whyte et al. (2012)	15.72
Xu et al. (2013)	18.51
Pan et al. (2016)	19.64
Kupyn et al. (2018)	16.36
Tao et al. (2018)	16.16
Ours	20.85

Our method performs favorably against the state-of-the-art non-uniform deblurring algorithms
 Bold value indicates the best performance

7 Analysis and Discussion

In this section, we discuss the connection with L_0 -regularized priors and learning-based priors, and analyze the convergence property, robustness to Gaussian noise as well as hyper-parameters, and the limitations of the proposed method.

7.1 Relation with L_0 -Regularized Priors

Several methods (Pan et al. 2014b; Xu et al. 2013) adopt the L_0 -regularized priors for blind image deblurring due to the strong sparsity of the L_0 norm. For example, the deblurring algorithms (Pan et al. 2016; Yan et al. 2017) enforce the L_0

sparsity on the extreme channels (i.e., dark and bright channels) as the blur process affects the distribution of the extreme channels. In addition, the L_0 gradient prior is adopted in both the state-of-the-art approaches and the proposed method to regularize the solution of latent images. We visualize the intermediate results in Fig. 12. The methods based on the L_0 -regularized prior on extreme channels (Pan et al. 2016; Yan et al. 2017) do not recover strong edges when there are not enough dark or bright pixels. The proposed method without the learned image prior (i.e., using L_0 gradient prior only) cannot restore strong edges well for blur kernel estimation. Although the CNN prior alone is not able to recover the latent image well, the approach by combining L_0 gradient and CNN prior restores sharper edges in the early

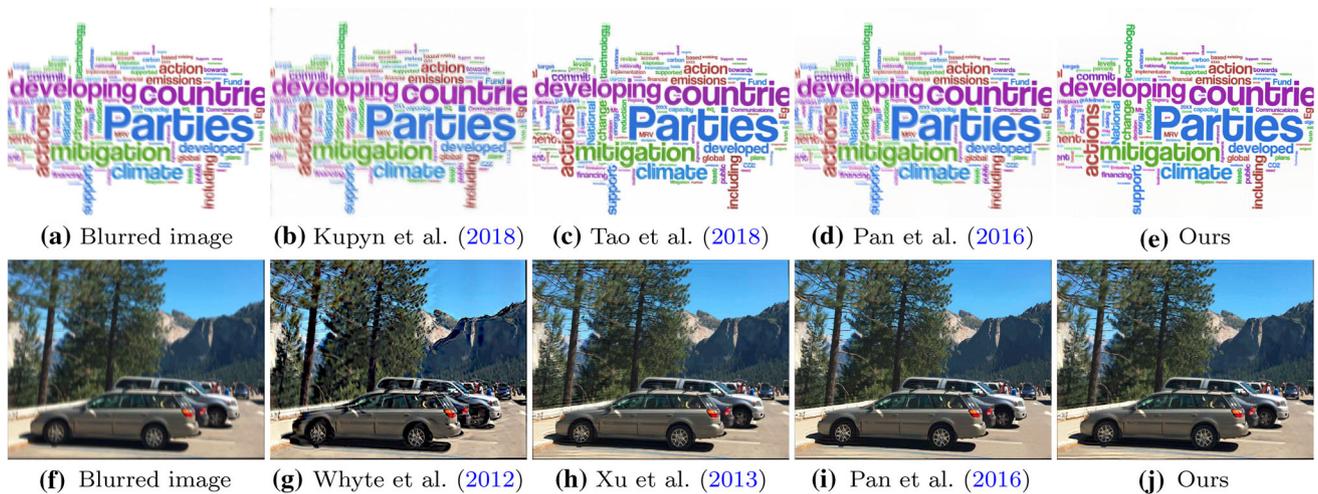


Fig. 11 Deblurred results on non-uniform blurred images (top: a synthetic image from Lai et al. (2016), bottom: a real image). We extend the proposed method for non-uniform deblurring and provide comparative results with the state-of-the-art methods

stage of the optimization process and helps estimate blur kernels.

To better understand the contribution of each term in (4), especially the relation between the L_0 gradient prior and CNN prior, we compare the deblurring results using different sparsity constraints on the gradient and with or without CNN prior on the dataset by Levin et al. (2009). As shown in Fig. 13a, while the L_0 gradient prior helps preserve more image structures, the integration with the proposed CNN prior leads to significant performance gain.

In addition, we apply L_1 -norm and L_2 -norm to image gradients, and compare the performance with and without the CNN prior. Figure 13a demonstrates that L_0 -norm is the most effective approach to constrain the sparsity on image gradients. The proposed CNN prior does not perform well when used alone as the prior learns features only from the image intensity, which cannot preserve structural information well. However, the method integrating the CNN prior and L_0 gradient prior performs better than using each gradient prior solely. Furthermore, the proposed method performs slightly better than that without the blur kernel regularization (i.e., $\|k\|_2^2$).

7.2 Relation with Learning-Based Priors

Numerous non-blind deconvolution methods learn image priors based on the Gaussian mixture model (Zoran and Weiss 2011) or deep CNNs (Zhang et al. 2017a, b). These learned image priors are effective for reducing noise and recovering fine details for non-blind deconvolution. However, we note it is not straightforward to extend these algorithms to effectively handle blind image deblurring.

To demonstrate the effectiveness of the proposed discriminative image prior, we extend existing learning-based priors (Zhang et al. 2017b; Zoran and Weiss 2011) to recover latent images for blind image deblurring. Specifically, we replace our image prior [i.e., $f(u)$ in (4)] with priors by Zhang et al. (2017b) and Zoran and Weiss (2011), and use the same strategy in Algorithm 1 to estimate blur kernels. We compare these learning-based image priors on the dataset by Levin et al. (2009) and present quantitative evaluations in Fig. 13. The method using the proposed discriminative prior performs favorably against those based on other learning-based priors. Figure 14 shows that the methods based on the priors (Zhang et al. 2017b; Zoran and Weiss 2011) do not restore sharp edges for blur kernel estimation. In contrast, the proposed image prior is able to recover strong edges which are useful for estimating blur kernels more accurately.

7.3 Run Time and Convergence Analysis

As the proposed optimization model involves a non-linear and non-convex term, it is of great interest to analyze the run time and convergence rate. We evaluate the state-of-the-art methods on images of different sizes and report the average execution time in Table 5. The proposed method runs competitively with the state-of-the-art approaches (Pan et al. 2016; Yan et al. 2017). In addition, we quantitatively evaluate convergence rate of the proposed optimization using images from the dataset of Levin et al. (2009). We compute values of the objective function (4) at the finest image scale. We also evaluate the average kernel similarity (Hu and Yang 2012) between estimated kernel K and ground-truth \hat{K} :

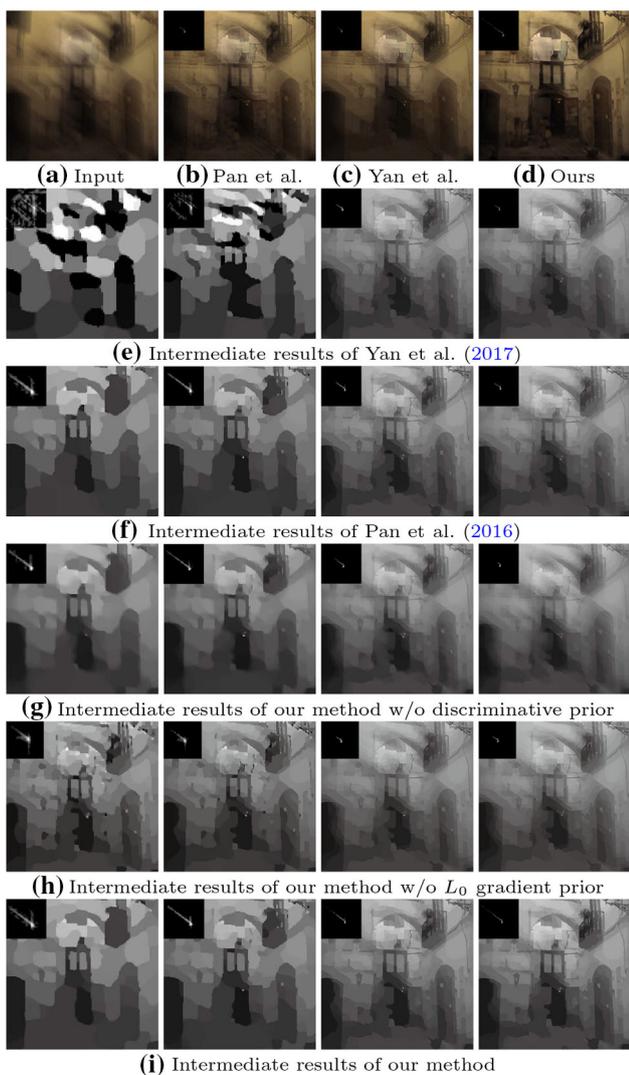


Fig. 12 Deblurred images and intermediate results. We show the intermediate latent images over iterations (from left to right) in (e–h). Our discriminative prior recovers intermediate results with more strong edges for kernel estimation

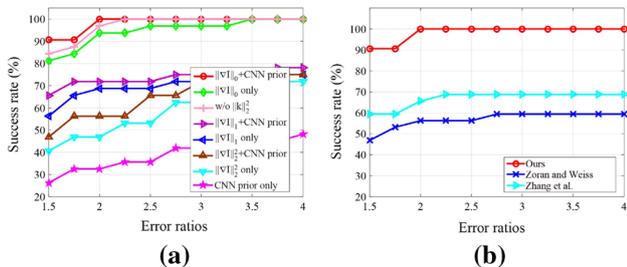


Fig. 13 Quantitative evaluations on proposed CNN prior on the dataset by Levin et al. (2009). **a** Comparisons with different sparsity constraints on image gradient and w/ or w/o the proposed CNN prior, **b** comparisons with the state-of-the-art learning priors

$$S(K, \hat{K}) = \max_{\chi} \frac{\sum_{\tau} K(\tau) \cdot \hat{K}(\tau + \chi)}{\|K\|_2^2 \cdot \|\hat{K}\|_2^2}, \quad (30)$$

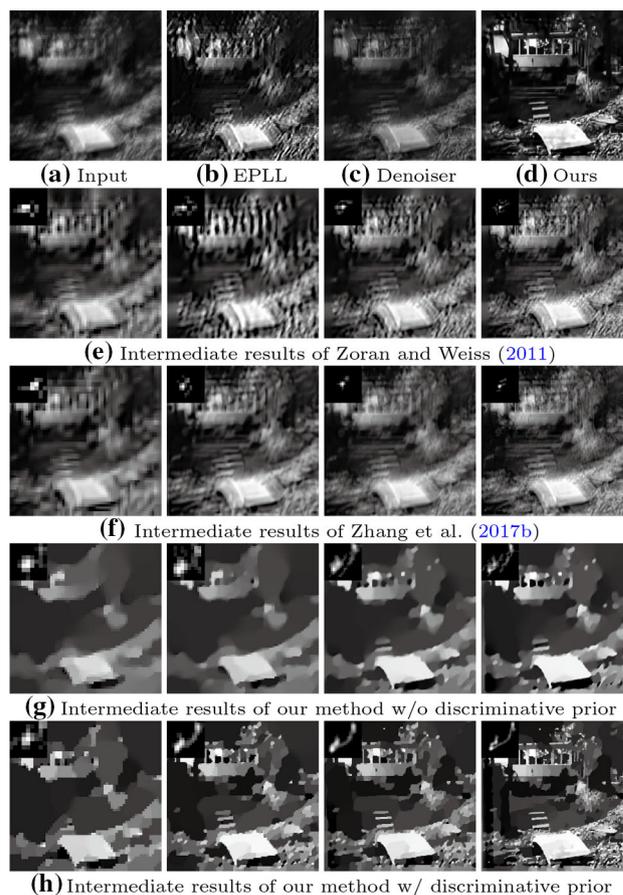


Fig. 14 Deblurred images and intermediate results. We compare the deblurred results with the state-of-the-art learning based methods [Zoran and Weiss 2011 (EPLL); Zhang et al. 2017b (Denoiser)] in (a–d) and illustrate the intermediate latent images over iterations (from left to right) in (e–h). The proposed prior recovers intermediate results with strong edges for kernel estimation while others fail

Table 5 Execution time comparisons

Method	255 × 255	600 × 600	800 × 800
Xu et al. (2013) (C++)	1.11	3.56	4.31
Krishnan et al. (2011)	24.23	111.09	226.58
Levin et al. (2011)	117.06	481.48	917.84
Pan et al. (2014b)	115.23	412.88	767.54
Pan et al. (2016)	134.31	691.71	964.90
Yan et al. (2017)	264.78	996.03	1150.48
Ours	109.27	379.52	654.65

We report the average execution time (in seconds) on three different sizes of images. All the algorithms are implemented in MATLAB except the method of Xu et al. (2013)

where χ is the possible shift between the two kernels, and τ denotes element coordinates in the kernel. In addition, $K(\tau)$ and $\hat{K}(\tau)$ are zeros when τ is out of the kernel range. Figure 15a, b show that our algorithm converges well within 50 iterations.

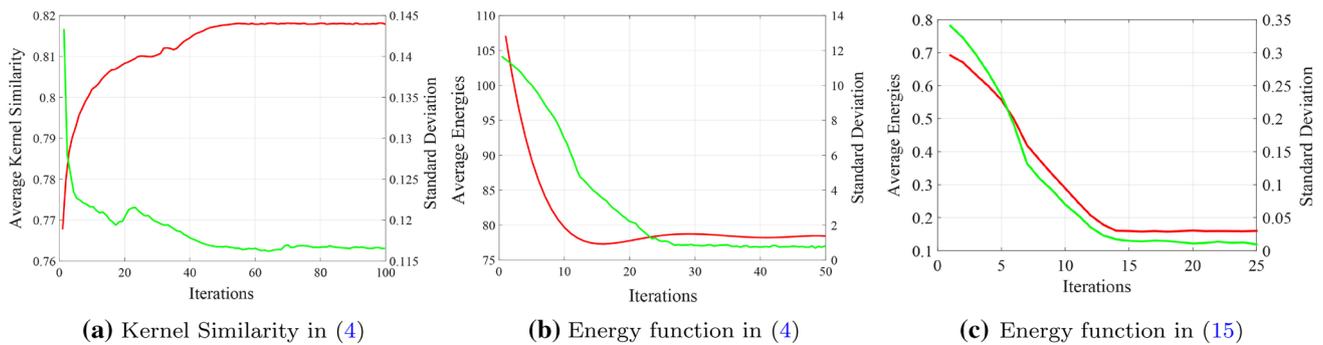


Fig. 15 Convergence analysis. We analyze the kernel similarity and the objective function (4) at the finest image scale. Our method converges well within 50 iterations. We also evaluate the energy function (15).

The function converges within 15 iterations at the step of 0.1. (The red lines denote average Kernel Similarity/Energy function which the green ones denote the standard deviations)

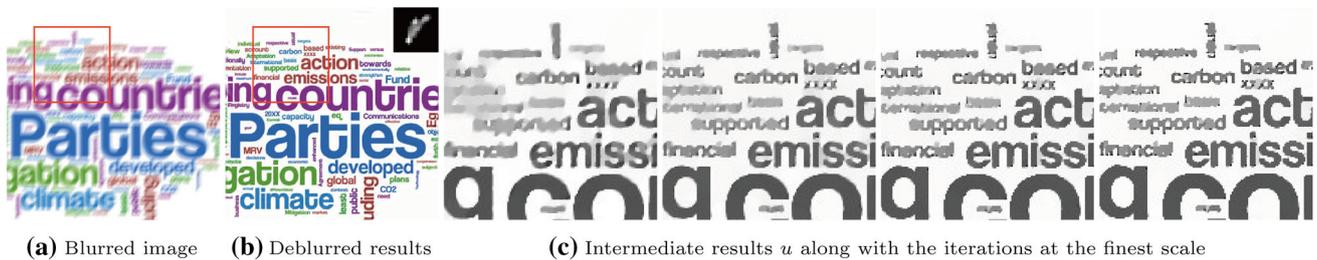


Fig. 16 Analysis on gradient descend method (15). Characters in the intermediate results become clearer long with the gradient descend (Color figure online)

Furthermore, we evaluate the convergence of the gradient descent method (15). As shown in Fig. 15c, the method converges well within 15 iterations at the step of 0.1. To better understand the gradient descent method, we use a text image to illustrate how u changes in (15) along with the iterations. As shown in Fig. 16, characters in the intermediate results (i.e., u) become clearer along with the gradient descent method.

7.4 Robustness to Image Noise

Figure 17 shows the deblurred results on a blurred image which contains Gaussian noise. As the dark channel (Pan et al. 2016) and extreme channel priors (Yan et al. 2017) are based on pixel intensity, these methods do not perform well due to the influence of image noise. In contrast, we add 1% Gaussian noise on blurred images when training our discriminative prior. As such, the proposed method performs more robustly to Gaussian noise and generates images with fewer artifacts.

For quantitative performance evaluation, we add 1–5% Gaussian or salt and pepper noise on five blurred images from the dataset of Levin et al. (2009). For fair comparisons, we use the same non-blind deconvolution method (Zoran and Weiss 2011) to restore latent images. Figure 18a shows that the proposed algorithm performs favorably against the

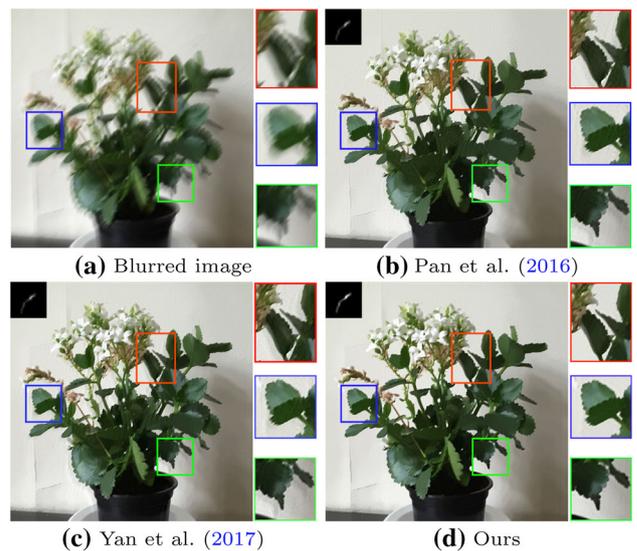


Fig. 17 A blurred image with Gaussian noise. The image priors based on intensity information (Pan et al. 2016; Yan et al. 2017) is less robust to images with Gaussian noise. In contrast, the deblurred results from the proposed algorithm contain fewer artifacts as shown in the zoom-in areas

state-of-the-art methods (Pan et al. 2016; Yan et al. 2017) on different Gaussian noise levels. However, both the state-of-

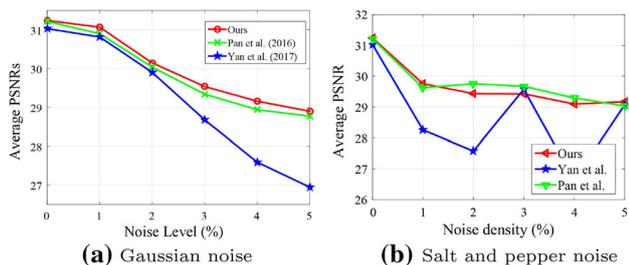


Fig. 18 Evaluations on blurred images with noise. Our method performs favorably against the state-of-the-art methods (Pan et al. 2016; Yan et al. 2017) on handling Gaussian noise. However, both the state-of-the-art approaches and proposed algorithm are less robust to salt and pepper noise

the-art approaches and our algorithm are less effective when images contain salt and pepper noise (Fig. 18b).

7.5 Analysis on Hyper-Parameters

There are four important hyper-parameters in the proposed model [λ , μ , γ in (4), and η in (15)]. For sensitivity analysis, we evaluate the proposed algorithm by varying one hyper-parameter and fixing others. We compute the kernel similarity metric (Hu and Yang 2012) to measure the accuracy of kernel estimation on the dataset of Levin et al. (2009). Figure 19 shows that the proposed method is not sensitive to change of each hyper-parameter within a wide range.

7.6 Limitations

The proposed algorithm is less effective when the linear convolution model does not hold for the blurred input images. We evaluate the proposed method on severe noise images, dynamic scene blurred images, and portrait images with out-of-focus backgrounds.

Severe noise Experimental results in Sect. 7.4 show that our image prior can handle Gaussian noise better than other intensity-based image priors (Pan et al. 2016; Yan et al. 2017). However, the proposed prior is less effective for images containing significant amount of noise. Figure 20 shows examples with 5% Gaussian noise and salt and pepper noise. The algorithm based on (2) cannot handle the denoising and deblurring problems jointly. Thus, the proposed algorithm

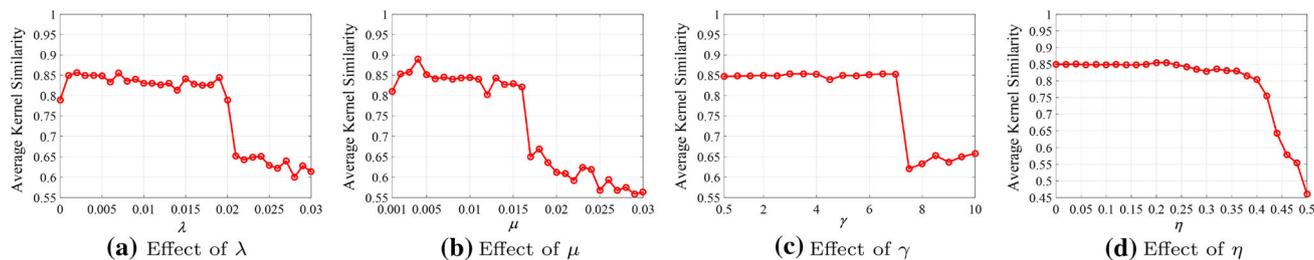


Fig. 19 Sensitivity analysis on the hyper-parameters λ , μ , γ , and η in the proposed model

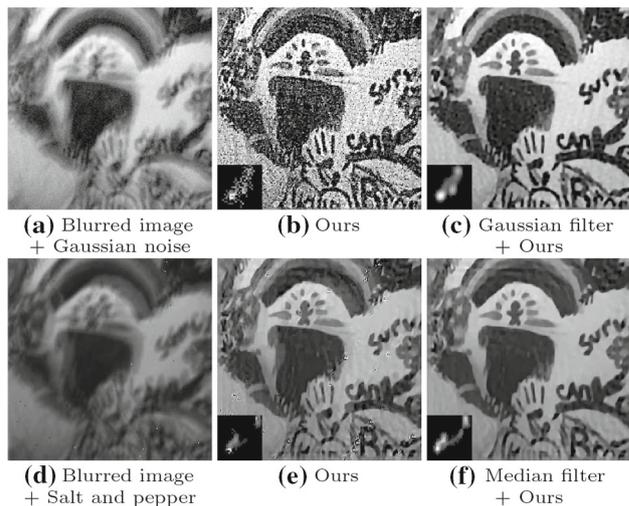


Fig. 20 Limitations on severe noise images of the proposed method. Our learned image prior is not effective on handling images with significant amount of noise

cannot restore the image well as shown in Fig. 20b. A simple solution is to first apply a low-pass filter to reduce image noise before deblurring. We adopt the Gaussian filter and median filter when input images contain the Gaussian noise and salt and pepper noise, respectively. As shown in Fig. 20c, the proposed method with these filters performs better while the details are not preserved well in the restored images. Our future work will consider jointly handling the deblurring and denoising in a principled way.

Object motion and out-of-focus blur The blur caused by out-of-focus or object motion is highly non-uniform. Existing methods often require image segmentation or depth estimation to deblur the image. As the linear convolution model used in the proposed algorithm does not hold for such cases, the proposed algorithm is less effective as shown in Figs. 21 and 22.

8 Extension to Image Dehazing

We demonstrate that the proposed discriminative prior can be extended to solve the image dehazing problem. Image dehazing aims to recover the clean image from a hazy or



Fig. 21 Limitations on dynamic scene blur images of the proposed method. Our method is not effective on dynamic scene deblurring

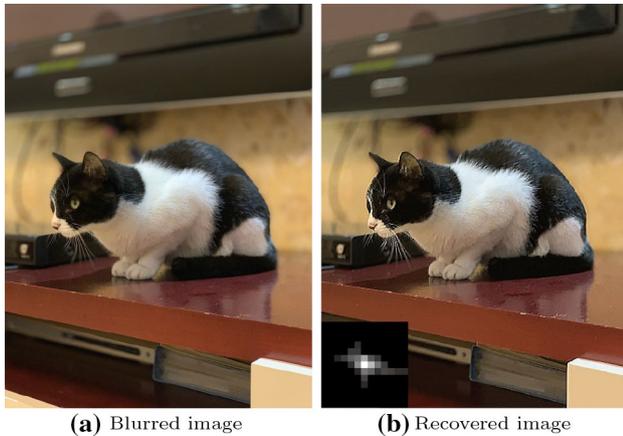


Fig. 22 Limitations on deblurring the background in a portrait image

foggy input. A hazy image can be formulated by the following model:

$$J(x) = I(x)t(x) + A(1 - t(x)), \quad (31)$$

where $J(x)$ denotes the hazy image, $I(x)$ represents the clean image, $t(x)$ is the transmission map, and A describes the atmospheric light. Solving (31) is also an well-known ill-posed problem. He et al. (2011) propose to first estimate $t(x)$ and A and then compute the clean image $I(x)$ by

$$I(x) = \frac{J(x) - A(1 - t(x))}{t(x)}. \quad (32)$$

However, computing $I(x)$ by direct element-wise division tends to introduce artifacts and color distortion when $J(x)$ contains noise or the estimated $t(x)$ and A are not accurate enough. To address this problem, we restore the clean image $I(x)$ by solving the following optimization problem:

$$\min_{I(x)} \|J(x) - I(x)t(x) - A(1 - t(x))\|_2^2 + \theta f(I(x)), \quad (33)$$

where θ is the weight to balance the prior term $f(I(x))$. The transmission map $t(x)$ and atmospheric light A are estimated by the method of He et al. (2011). We set $\theta = 0.01$ and use the half-quadratic and gradient descent methods as described in Sect. 4 to solve (33).

In order to embed the learned image prior into the optimization-based framework (33), we re-train the classifier with hazy images (labeled as 1) and haze-free images (labeled as 0). We use the dataset of Ren et al. (2016), which contains 2413 synthetic hazy images and corresponding clean images. We adopt the same training strategy in Sect. 3.4 to train our classifier. As shown in Fig. 23, the re-trained classifier generates different feature maps for hazy and clean images.

We evaluate the proposed method on the synthetic test set of the image dehazing benchmark dataset (Li et al. 2019), which contains 500 synthetic hazy images. Table 6 shows our method performs favorably against the state-of-the-art single-image dehazing algorithms (He et al. 2011; Berman et al. 2016; Ren et al. 2016; Pan et al. 2018; Li et al. 2017b) in terms of PSNR and SSIM.

As shown in Fig. 24e, the results of our method are visually more pleasing and have less color distortion compared

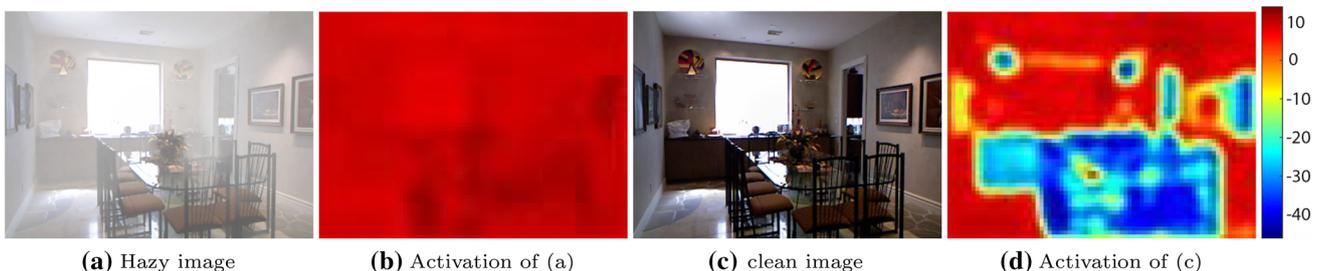


Fig. 23 Activation of feature maps of a pair of hazy and clean images. Our re-trained classifier can distinguish whether an image is hazy or not

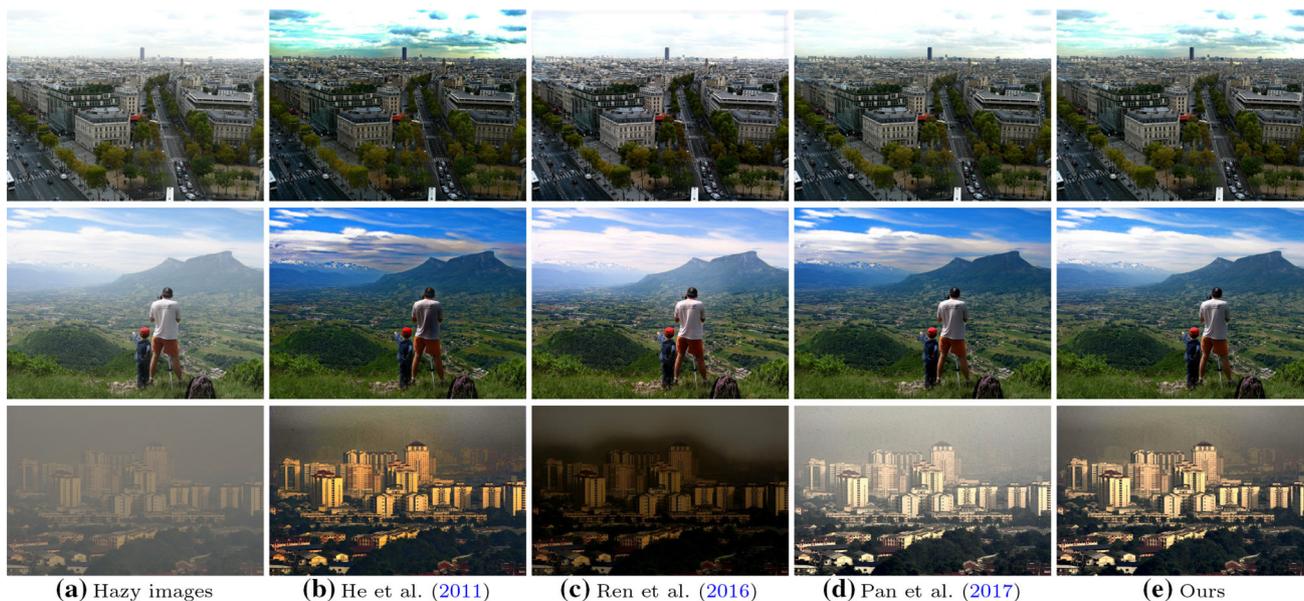


Fig. 24 Dehazing results on real hazy images. Our dehazing approach generates better results compared with the state-of-the-art methods (He et al. 2011; Ren et al. 2016; Pan et al. 2018)

Table 6 Average PSNR/SSIM values on the SOTS dataset from RESIDE (Li et al. 2019)

Methods	PSNR/SSIM
He et al. (2011)	15.49/0.73
Berman et al. (2016)	17.27/0.75
Ren et al. (2016)	17.57/0.81
Pan et al. (2018)	18.84/0.85
Li et al. (2017b)	19.06/0.85
Ours	19.35/0.88

Our method performs favorably against the state-of-the-art image dehazing approaches
 Bold value indicates the best performance

with those by solving (32) (Fig. 24b). The proposed method generates images with clearer details than the state-of-the-art approaches (He et al. 2011; Ren et al. 2016; Pan et al. 2018).

9 Concluding Remarks

In this paper, we propose a data-driven discriminative prior for blind image deblurring. We learn the image prior via a binary classification network based on the criterion that the prior should favor sharp images over blurred ones in various scenarios. We adopt a global average pooling layer and a multi-scale training strategy such that the network performs robustly to different sizes of images. We then embed the learned image prior into a coarse-to-fine optimization framework and develop an efficient half-quadratic splitting

algorithm for estimating blur kernels. Our prior performs well on several types of images, including natural, text, face and low-illumination images, and can be easily extended to the non-uniform deblurring and image dehazing problems. Extensive quantitative and qualitative evaluations demonstrate that the proposed method performs favorably against the state-of-the-art generic and domain-specific blind deblurring algorithms.

Acknowledgements This work is partially supported by the National Natural Science Foundation of China (Nos. 61433007 and 61872421), the National Science Foundation CAREER (No. 1149783), the Natural Science Foundation of Jiangsu Province (No. BK20180471), and gifts from Adobe and Nvidia. Lerenhan Li is supported by a scholarship from China Scholarship Council.

References

Berman, D., Treibitz, T., & Avidan, S. (2016). Non-local image dehazing. In *IEEE conference on computer vision and pattern recognition*.

Bigdeli, S. A., Zwicker, M., Favaro, P., & Jin, M. (2017). Deep mean-shift priors for image restoration. In *Neural information processing systems*.

Boracchi, G., & Foi, A. (2012). Modeling the performance of image restoration from motion blur. *IEEE Transactions on Image Processing*, 21(8), 3502–3517.

Chakrabarti, A. (2016). A neural approach to blind motion deblurring. In *European conference on computer vision*.

Cho, S., & Lee, S. (2009). Fast motion deblurring. *ACM Transactions on Graphics*, 28(5), 145.

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., Deng, Y., Change Loy, C., & Tang, X. (2015). Compression artifacts reduction by a deep convolutional network. In *IEEE international conference on computer vision*.
- Fergus, R., Singh, B., Hertzmann, A., Roweis, S. T., & Freeman, W. T. (2006). Removing camera shake from a single photograph. *ACM Transactions on Graphics*, 25(3), 787–794.
- Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics* (pp. 249–256).
- He, K., Sun, J., & Tang, X. (2011). Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12), 2341–2353.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *IEEE Conference on computer vision and pattern recognition*.
- Hirsch, M., Schuler, C. J., Harmeling, S., & Schölkopf, B. (2011). Fast removal of non-uniform camera shake. In *IEEE International conference on computer vision*.
- Hradiš, M., Kotera, J., Zemčík, P., & Šroubek, F. (2015). Convolutional neural networks for direct text deblurring. In *British machine vision conference*.
- Hu, Z., & Yang, M. H. (2012). Good regions to deblur. In *European conference on computer vision*.
- Hu, Z., Cho, S., Wang, J., & Yang, M. H. (2014). Deblurring low-light images with light streaks. In *IEEE conference on computer vision and pattern recognition*.
- Huiskes, M. J., & Lew, M. S. (2008). The MIR flickr retrieval evaluation. In *ACM international conference on multimedia information retrieval*.
- Jin, M., Hirsch, M., & Favaro, P. (2018). Learning face deblurring fast and wide. In *IEEE conference on computer vision and pattern recognition workshops*.
- Kim, J., Lee, J. K., & Lee, K. M. (2016). Accurate image super-resolution using very deep convolutional networks. In *IEEE conference on computer vision and pattern recognition*.
- Köhler, R., Hirsch, M., Mohler, B., Schölkopf, B., & Harmeling, S. (2012). Recording and playback of camera shake: Benchmarking blind deconvolution with a real-world database. In *European conference on computer vision*.
- Krishnan, D., & Fergus, R. (2009). Fast image deconvolution using hyper-laplacian priors. In *Neural information processing systems*.
- Krishnan, D., Tay, T., & Fergus, R. (2011). Blind deconvolution using a normalized sparsity measure. In *IEEE conference on computer vision and pattern recognition*.
- Kupyn, O., Budzan, V., Mykhailych, M., Mishkin, D., & Matas, J. (2018). Deblurgan: Blind motion deblurring using conditional adversarial networks. In *IEEE conference on computer vision and pattern recognition*.
- Lai, W. S., Ding, J. J., Lin, Y. Y., & Chuang, Y. Y. (2015). Blur kernel estimation using normalized color-line prior. In *IEEE conference on computer vision and pattern recognition*.
- 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*.
- Leclaire, A., & Moisan, L. (2013). Blind deblurring using a simplified sharpness index. In *International conference on scale space and variational methods in computer vision*.
- Levin, A., Weiss, Y., Durand, F., & Freeman, W. T. (2009). Understanding and evaluating blind deconvolution algorithms. In *IEEE conference on computer vision and pattern recognition*.
- Levin, A., Weiss, Y., Durand, F., & Freeman, W. T. (2011). Efficient marginal likelihood optimization in blind deconvolution. In *IEEE conference on computer vision and pattern recognition*.
- Li, B., Peng, X., Wang, Z., Xu, J., & Feng, D. (2017a). Aod-net: All-in-one dehazing network. In *IEEE international conference on computer vision*.
- Li, B., Peng, X., Wang, Z., Xu, J. Z., & Feng, D. (2017b). Aod-net: All-in-one dehazing network. In *IEEE international conference on computer vision*.
- Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., et al. (2019). Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1), 492–505.
- Lin, M., Chen, Q., & Yan, S. (2013). Network in network. [arXiv:1312.4400](https://arxiv.org/abs/1312.4400).
- Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *IEEE conference on computer vision and pattern recognition*.
- Mao, X., Shen, C., & Yang, Y. B. (2016). Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. In *Neural information processing systems*.
- Michaeli, T., & Irani, M. (2014). Blind deblurring using internal patch recurrence. In *European conference on computer vision*.
- Nah, S., Kim, T. H., & Lee, K. M. (2017). Deep multi-scale convolutional neural network for dynamic scene deblurring. In *IEEE conference on computer vision and pattern recognition*. <https://doi.org/10.1109/CVPR.2017.35>.
- Noroози, M., Chandramouli, P., & Favaro, P. (2017). Motion deblurring in the wild. In *German conference on pattern recognition*.
- Pan, J., Hu, Z., Su, Z., Yang, M. H. (2014a). Deblurring face images with exemplars. In *European conference on computer vision*.
- Pan, J., Hu, Z., Su, Z., Yang, M. H. (2014b). Deblurring text images via l0-regularized intensity and gradient prior. In *IEEE conference on computer vision and pattern recognition*.
- Pan, J., Sun, D., Pfister, H., & Yang, M. H. (2016). Blind image deblurring using dark channel prior. In *IEEE conference on computer vision and pattern recognition*.
- Pan, J., Sun, D., Pfister, H., & Yang, M.-H. (2018). Deblurring images via dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(10), 2315–2328.
- Ren, W., Liu, S., Zhang, H., Pan, J., Cao, X., & Yang, M. H. (2016). Single image dehazing via multi-scale convolutional neural networks. In *European conference on computer vision*.
- Schuler, C. J., Hirsch, M., Harmeling, S., & Schölkopf, B. (2016). Learning to deblur. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(7), 1439–1451.
- Shen, Z., Lai, W. S., Xu, T., Kautz, J., & Yang, M. H. (2018). Deep semantic face deblurring. In *IEEE conference on computer vision and pattern recognition*.
- Sreehari, S., Venkatakrishnan, S. V., Wohlberg, B., Buzzard, G. T., Drummy, L. F., Simmons, J. P., et al. (2016). Plug-and-play priors for bright field electron tomography and sparse interpolation. *IEEE Transactions on Computational Imaging*, 2(4), 408–423.
- Sun, J., Cao, W., Xu, Z., & Ponce, J. (2015). Learning a convolutional neural network for non-uniform motion blur removal. In *IEEE conference on computer vision and pattern recognition*.
- Sun, L., Cho, S., Wang, J., & Hays, J. (2013). Edge-based blur kernel estimation using patch priors. In *IEEE international conference on computational photography*.
- Tai, Y. W., Tan, P., & Brown, M. S. (2011). Richardson-lucy deblurring for scenes under a projective motion path. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(8), 1603–1618.
- Tao, X., Gao, H., Shen, X., Wang, J., & Jia, J. (2018). Scale-recurrent network for deep image deblurring. In *IEEE conference on computer vision and pattern recognition*.
- Vedaldi, A., & Lenc, K. (2015). Matconvnet: Convolutional neural networks for matlab. In *ACM international conference on multimedia*.

- Whyte, O., Sivic, J., Zisserman, A., & Ponce, J. (2012). Non-uniform deblurring for shaken images. *International Journal of Computer Vision*, 98(2), 168–186.
- Xu, L., & Jia, J. (2010). Two-phase kernel estimation for robust motion deblurring. In *European conference on computer vision*.
- Xu, L., Lu, C., Xu, Y., & Jia, J. (2011). Image smoothing via l_0 gradient minimization. *ACM Transactions on Graphics*, 30(6), 174.
- Xu, L., Zheng, S., & Jia, J. (2013). Unnatural l_0 sparse representation for natural image deblurring. In *IEEE conference on computer vision and pattern recognition*.
- Yan, R., & Shao, L. (2016). Blind image blur estimation via deep learning. *IEEE Transactions on Image Processing*, 25(4), 1910–1921.
- Yan, Y., Ren, W., Guo, Y., Wang, R., & Cao, X. (2017). Image deblurring via extreme channels prior. In *IEEE conference on computer vision and pattern recognition*.
- Zhang, J., Pan, J., Lai, W. S., Lau, R., & Yang, M. H. (2017a). Learning fully convolutional networks for iterative non-blind deconvolution. In *IEEE conference on computer vision and pattern recognition*.
- Zhang, K., Zuo, W., Gu, S., & Zhang, L. (2017b). Learning deep CNN denoiser prior for image restoration. In *IEEE conference on computer vision and pattern recognition*.
- Zhao, P., Cheng, Y., & Pedersen, M. (2015). Objective assessment of perceived sharpness of projection displays with a calibrated camera. In *Colour and visual computing symposium*.
- Zoran, D., & Weiss, Y. (2011). From learning models of natural image patches to whole image restoration. In *IEEE international conference on computer vision*.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.