

# Learning Discriminative Shrinkage Deep Networks for Image Deconvolution

Pin-Hung Kuo<sup>1</sup>   Jinshan Pan<sup>2</sup>   Shao-Yi Chien<sup>1</sup>   Ming-Hsuan Yang<sup>3,4,5</sup>

<sup>1</sup>Graduate Institute of Electronics Engineering, National Taiwan University

<sup>2</sup>Nanjing University of Science and Technology

<sup>3</sup>Google Research

<sup>4</sup>University of California, Merced

<sup>5</sup>Yonsei University

**Abstract.** Most existing methods usually formulate the non-blind deconvolution problem into a maximum-a-posteriori framework and address it by manually designing a variety of regularization terms and data terms of the latent clear images. However, explicitly designing these two terms is quite challenging and usually leads to complex optimization problems which are difficult to solve. This paper proposes an effective non-blind deconvolution approach by learning discriminative shrinkage functions to model these terms implicitly. Most existing methods use deep convolutional neural networks (CNNs) or radial basis functions to learn the regularization term simply. In contrast, we formulate both the data term and regularization term and split the deconvolution model into data-related and regularization-related sub-problems according to the alternating direction method of multipliers. We explore the properties of the Maxout function and develop a deep CNN model with Maxout layers to learn discriminative shrinkage functions, which directly approximates the solutions of these two sub-problems. Moreover, the fast-Fourier-transform-based image restoration usually leads to ringing artifacts. At the same time, the conjugate-gradient-based approach is time-consuming; we develop the Conjugate Gradient Network to restore the latent clear images effectively and efficiently. Experimental results show that the proposed method performs favorably against the state-of-the-art methods in terms of efficiency and accuracy. Source codes, models, and more results are available at <https://github.com/setsunil/DSDNet>.

## 1 Introduction

The single image deconvolution, or deblurring, aims to restore a clear and sharp image from a single blurry input image. Blind image deblurring has attracted interest from many researchers [31, 24, 49, 43]. With the rapid development of deep learning, tremendous progress has been made in blind image deblurring recently [26, 62, 74, 13, 59]. Since the kernel is available via blind methods, how to utilize these kernels well is still an important issue. Therefore, non-blind deconvolution has never lost the attention of researchers over the past decades [50, 14, 8, 10]. Due to the deconvolution problem's ill-posedness, numerous methods explore the statistical properties of clear images as the image priors (e.g., hyper-Laplacian prior [23, 30]) to make this problem tractable. Although using the hand-crafted

image priors facilitates ringing artifacts removal, the fine details are not restored well, as these limited priors may not model the inherent properties of various latent images sufficiently.

To overcome this problem, discriminative image priors are learned from training examples [58,56,9,48]. These methods usually leverage radial basis functions (RBFs) as the shrinkage functions of the image prior. However, the RBFs contain many parameters, leading to complex optimization problems.

Deep convolutional neural network (CNN) has been developed to learn more effective regularization terms for the deconvolution problem [70]. These methods are motivated by [56] and directly estimate the solution to the regularization-related sub-problem by deep CNN models. As analyzed by [56], we can obtain the solution to the regularization-related sub-problem by combining the shrinkage functions. However, as the shrinkage functions are complex (e.g., non-Monotonic), simply using the convolution operation followed by the common activation functions, e.g., ReLU, cannot model the property of the shrinkage functions. Given the effectiveness of the deep features, it is of great interest to learn discriminative shrinkage functions. Therefore, if we can learn more complex shrinkage functions corresponding to the deep features, they shall surpass the hand-crafted ones in solving the regularization-related sub-problem.

We note that image restoration involves an image deconvolution step, usually taking advantage of fast Fourier transform (FFT) [56,43,73] or the Conjugate Gradient (CG) method [2,28,10]. However, the FFT-based approaches usually lead to ringing artifacts, while the CG-based ones are time-consuming. Besides, these two methods suffer from information loss: for FFT, we lose little information when we discard the imaginary parts in the inverse real FFT; for the CG method, the iterations we execute are usually far less than the upper bound. Therefore, developing an effective yet efficient image restoration method is also necessary.

In this paper, we develop a simple and effective model to discriminatively learn the shrinkage functions for non-blind deconvolution, which is called Discriminative Shrinkage Deep Network (DSDNet). We formulate the data and regularization terms as learnable and split the image deconvolution model into the data-related and regularization-related sub-problems. As shrinkage functions can solve both sub-problems, and the learnable Maxout functions can efficiently approximate any complex functions, we directly learn the shrinkage functions of sub-problems via a deep CNN model with Maxout layers [16]. To effectively and efficiently generate clear images from the output of the learned functions, we develop a fully convolutional Conjugate Gradient Network (CGNet) motivated by the mathematical concept of the CG method. Finally, to solve the problem, we formulate our method into an end-to-end network based on the Alternating Direction Method of Multipliers (ADMM) [44]. Experimental results show that the proposed method performs favorably against the state-of-the-art ones.

The main contributions of this work are:

- We propose a simple yet effective non-blind deconvolution model to directly learn discriminative shrinkage functions to implicitly model the data and regularization terms for image deconvolution.
- We develop an efficient and effective CGNet to restore clear images without the problems of CG and FFT.
- The architecture of DSDNet is designed elaborately, which makes it flexible in model size and easy to be trained. Even the smallest DSDNet performs favorably against the state-of-the-art methods in speed and accuracy.

## 2 Related Work

Because numerous image deconvolution methods have been proposed, we discuss those most relevant to this work.

**Statistical Image Prior-Based Methods.** Since non-blind deconvolution is an ill-posed problem, conventional methods usually develop image priors based on the statistical properties of clear images. Representative methods include total variation [52,6,65,46], hyper-Laplacian prior [28,23], and patch-based prior [61,17,75], to name a few. However, these hand-crafted priors may not model the inherent properties of the latent image well; thus, these methods do not effectively restore realistic images.

**Learning-Based Methods.** To overcome the above limitations of the hand-crafted priors, researchers have proposed learning-based approaches, e.g., Markov random fields [51,54], Gaussian mixture models [75], conditional random fields [63,19,57,55], and radial basis functions [56,9].

The learning-based non-blind deconvolution also gets deeper with the development of neural networks. Many methods use deep CNNs to model the regularization term and solve image restoration problems by unrolling existing optimization algorithms. For example, Iterative Shrinkage-Thresholding algorithm [69,67], Douglas-Rachford method [1], Half-Quadratic Splitting algorithm [33,71,21,70,4,32,7,22], gradient descend [15,47] and ADMM [68]. These methods use deep CNN models to estimate the solution to the regularization-related sub-problem. As demonstrated by [56], the solutions are the combinations of the shrinkage functions. Simply using deep CNN models does not model the shrinkage functions well since most activation functions are too simple. Besides, most of them focus on the regularization terms yet ignore the importance of data terms. In addition, the image restoration step in these methods usually depends on an FFT-based solution. However, using FFT may lead to results with ringing artifacts. Even though the edge taper [25] alleviates artifacts, they are still inevitable in many scenes.

To overcome these problems, we leverage Maxout layers to learn discriminative shrinkage functions for regularization and data terms and develop the CGNet to restore images better. Furthermore, we adopt average pooling for noise level estimation and residual block for re-weights computation. In other

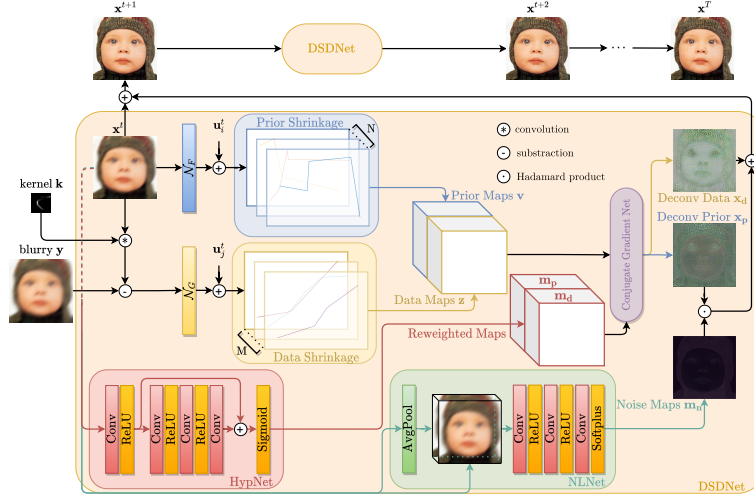


Fig. 1: Overview of the proposed method. The blue blocks and lines are the layers and flow of the regularization terms; the yellow ones are that of data terms. The HypNet is responsible for the reweighted maps the NLNet learns to control the weights of regularization and data terms according to the local noise level.

words, we design each component according to the mathematical characteristics rather than stacking as many convolutional layers as possible, as in most previous works.

**Blind Deblurring Methods.** Numerous end-to-end deep networks [60,40,62,26] have been developed to restore clear images from blurry images directly. However, as demonstrated in [10], when the blur kernels are given, these methods do not perform well compared to the non-blind deconvolution methods. As non-blind deconvolution is vital for image restoration, we focus on this problem and develop a simple and effective approach to restoring high-quality images.

### 3 Revisiting Deep Unrolling-Based Methods

We first revisit deep unrolling-based methods for image deconvolution to motivate our work. Mathematically, the degradation process of the image blur is usually formulated as:

$$y = k * x + n, \quad (1)$$

where  $*$  denotes the convolution operator;  $y$ ,  $k$ ,  $x$  and  $n$  denote the blurry image, the blur kernel, the latent image and noise, respectively. With the known kernel  $k$ , we usually use formulate the deconvolution as a maximum-a-posteriori (MAP) problem:

$$x = \arg \max_x p(x|y, k) = \arg \max_x p(y|x, k)p(x), \quad (2)$$



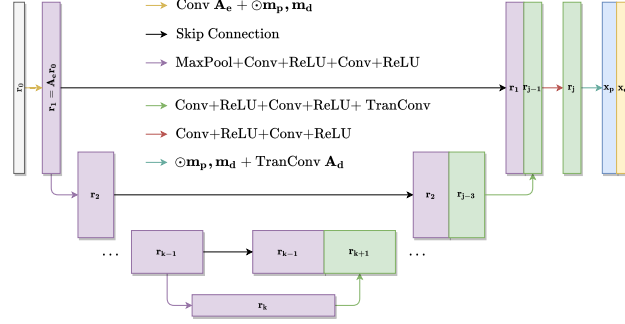


Fig. 2: Network architecture of the proposed CGNet. The residual vector  $\mathbf{r}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}^t$  is the input feature map, the operation  $\mathbf{A}$ , the convolution  $\mathbf{A}_e$  and the transposed convolution  $\mathbf{A}_d$  are composed of the input filters  $\mathbf{F}_i$ ,  $\mathbf{G}_j$  and  $\mathbf{H}$ . The reweighted maps  $\mathbf{m}_p$ ,  $\mathbf{m}_d$  are multiplied between  $\mathbf{A}_e$  and  $\mathbf{A}_d$  as the IRLS.

where  $p(y|x, k)$  is the likelihood of the observation (blurry)  $y$ , while  $p(x)$  denotes an image prior of the latent image  $x$ . This equation is equivalent to

$$\min_x R(x) + D(y - k * x), \quad (3)$$

where  $R(x)$  and  $D(y - k * x)$  denote the regularization term and data term. In addition, the data term is usually modeled in the form of  $\ell_2$ -norm, then (3) can be rewritten as

$$\min_{\mathbf{x}} \frac{1}{2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + R(\mathbf{x}), \quad (4)$$

where  $\mathbf{x}$ ,  $\mathbf{y}$  denote the vector forms of  $x$  and  $y$ , respectively;  $\mathbf{H}$  denotes the Toeplitz matrix of the blur kernel  $k$ . The ADMM method for image deconvolution is usually achieved by solving:

$$\min_{\mathbf{x}, \mathbf{u}} \frac{1}{2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + R(\mathbf{v}) + \mathbf{u}^\top (\mathbf{v} - \mathbf{x}) + \frac{\rho}{2} \|\mathbf{v} - \mathbf{x}\|_2^2, \quad (5)$$

where  $\mathbf{v}$  is an auxiliary variable,  $\mathbf{u}$  is a Lagrangian multiplier,  $\rho$  is a weight parameter.

The solution of (5) can be obtained by alternatively solving:

$$\mathbf{x}^{t+1} = \min_{\mathbf{x}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \rho \|\mathbf{v}^t - \mathbf{x} + \frac{\mathbf{u}^t}{\rho}\|_2^2, \quad (6a)$$

$$\mathbf{v}^{t+1} = \min_{\mathbf{v}} \frac{\rho}{2} \|\mathbf{v} - \mathbf{x}^{t+1} + \frac{\mathbf{u}^t}{\rho}\|_2^2 + R(\mathbf{v}), \quad (6b)$$

$$\mathbf{u}^{t+1} = \mathbf{u}^t + \rho(\mathbf{v}^{t+1} - \mathbf{x}^{t+1}). \quad (6c)$$

Existing methods [70,73,72] usually solve (6a) via fast Fourier transform (FFT) or Conjugate Gradient methods. For (6b), its solution can be represented as a

proximal operator:

$$\mathbf{prox}_{\lambda R}(\mathbf{x}^{t+1} - \mathbf{u}^t) = \arg \min_{\mathbf{v}} \frac{1}{2} \|\mathbf{v} - \mathbf{x}^{t+1} + \mathbf{u}^t\|_2^2 + \lambda R(\mathbf{v}), \quad (7)$$

where  $\lambda = 1/\rho$ . With  $\mathbf{u}^t = \lambda \mathbf{u}^t$ , the multiplier in the (6) and (7) can be absorbed [44]. As demonstrated in [56], (7) can be approximated by shrinkage functions. Existing methods usually use deep CNN model to approximate the solution of (7). However, simply using the convolution operation followed by the fixed activation functions (e.g., ReLU) cannot model the shrinkage functions well as they are far more complex (e.g., non-monotonic) [56]. To better approximate the solution of (6b), we develop a deep CNN model with the Maxout function [16], which can effectively approximate proximal functions. In addition, we note that using FFT to solve (6a) does not obtain better results than the CG method demonstrated by [28,29]. However, the CG method is time-consuming and unstable in deep networks (see Section 5.4 for more detail). To overcome this problem, we learn a differentiable CG network to restore a clear image more efficiently and effectively.

## 4 Proposed Method

Different from existing methods that simply learn the regularization term or the data term [70,4,32], we formulate both the data term and the regularization term as the learnable ones:

$$\begin{aligned} \min_{\mathbf{u}, \mathbf{v}, \mathbf{x}, \mathbf{z}} \quad & \sum_{i=1}^N R_i(\mathbf{v}_i) + \sum_{j=1+N}^{M+N} R_j(\mathbf{z}_j) \\ \text{s.t.} \quad & \mathbf{F}_i \mathbf{x} = \mathbf{v}_i, \quad \mathbf{G}_j (\mathbf{y} - \mathbf{H} \mathbf{x}) = \mathbf{z}_j, \end{aligned} \quad (8)$$

where  $R_i$  denotes the  $i$ -th learnable function;  $\mathbf{v}_i$  and  $\mathbf{z}_j$  are auxiliary variables that correspond to the regularization and data terms;  $\mathbf{F}_i$  and  $\mathbf{G}_j$  are the  $i$ -th and  $j$ -th learnable filters for regularization and data, respectively.

By introducing the Lagrangian multipliers  $\mathbf{u}_i$  and  $\mathbf{u}_j$  corresponding to the regularization and data terms, we can solve (8) using the ADMM method by:

$$\mathbf{v}_i^{t+1} = \mathbf{prox}_{\lambda_i R_i}(\mathbf{F}_i \mathbf{x}^t + \mathbf{u}_i^t), \quad (9a)$$

$$\mathbf{z}_j^{t+1} = \mathbf{prox}_{\lambda_j R_j}(\mathbf{G}_j (\mathbf{y} - \mathbf{H} \mathbf{x}^t) + \mathbf{u}_j^t), \quad (9b)$$

$$\begin{aligned} & \left( \sum_{i=1}^N \rho_i \mathbf{F}_i^\top \mathbf{F}_i + \sum_{j=N+1}^{N+M} \rho_j \mathbf{H}^\top \mathbf{G}_j^\top \mathbf{G}_j \mathbf{H} \right) \mathbf{x}^{t+1} \\ &= \left( \sum_{i=1}^N \rho_i \mathbf{F}_i^\top (\mathbf{v}_i^{t+1} - \mathbf{u}_i^t) + \sum_{j=N+1}^{N+M} \rho_j \mathbf{H}^\top \mathbf{G}_j^\top (\mathbf{G}_j \mathbf{y} - \mathbf{z}_j^{t+1} + \mathbf{u}_j^t) \right) \end{aligned} \quad (9c)$$

$$\mathbf{u}_i^{t+1} = \mathbf{u}_i^t + \mathbf{F}_i \mathbf{x}^{t+1} - \mathbf{v}_i^{t+1}, \quad (9d)$$

$$\mathbf{u}_j^{t+1} = \mathbf{u}_j^t + \mathbf{G}_j (\mathbf{y} - \mathbf{H} \mathbf{x}^{t+1}) - \mathbf{z}_j^{t+1}. \quad (9e)$$

In the following, we will develop deep CNN models with a Maxout layer to approximate the functions of (9a) and (9b). Moreover, we design a simple and effective deep CG network to solve (9c).

#### 4.1 Network Architecture

This section describes how to design our deep CNN models to effectively solve (9a)-(9c).

**Learning Filters  $\mathbf{F}_i$  and  $\mathbf{G}_j$ .** To learn filters  $\mathbf{F}_i$  and  $\mathbf{G}_j$ , we develop two networks ( $\mathcal{N}_F$  and  $\mathcal{N}_G$ ), each containing one convolutional layer. The convolutional layer of  $\mathcal{N}_F$  has  $N$  filters of  $7 \times 7$  pixels, and the convolutional layer of  $\mathcal{N}_G$  contains  $M$  filters of the same size. The  $\mathcal{N}_F$  and  $\mathcal{N}_G$  are applied to  $\mathbf{x}^t$  and  $\mathbf{y} - \mathbf{H}\mathbf{x}^t$  to learn the filters  $\mathbf{F}_i$  and  $\mathbf{G}_j$ , respectively.

**Learning Discriminative Shrinkage Functions for (9a) and (9b).** To better learn the unknown discriminative shrinkage functions of (9a) and (9b), we take advantage of Maxout layers [16]. Specifically, the convolutional Maxout layer consists of two Maxout units. Each Maxout unit contains one convolutional layer followed by a channel-wise Max-pooling layer. Given an input feature map  $\mathbf{X} \in \mathbb{R}^{H \times W \times C}$  and the output feature map  $\mathbf{X}^o \in \mathbb{R}^{H \times W \times KC}$  of the convolutional layer, a Maxout unit is achieved by:

$$o_{h,w,c}(\mathbf{X}) = \max_{j \in [0, K)} x_{h,w,c \times K + j}^o, \quad (10)$$

where  $h \in [0, H)$ ,  $w \in [0, W)$  and  $c \in [0, C)$ ; the  $x_{h,w,c}^o$  is the element of  $\mathbf{X}^o$  at the position  $(h, w, c)$ , and the  $o_{h,w,c}$  is the function to output the  $(h, w, c)$ -th element of the output tensor  $\mathbf{O}$ . In our implementation, we have  $\mathbf{O} \in \mathbb{R}^{H \times W \times C}$  is of the same size as the input  $\mathbf{X}$  and  $K = 4$ .

With two Maxout units, we acquire two output features,  $\mathbf{O}_1$  and  $\mathbf{O}_2$ ; the final output tensor of the Maxout layer is their difference,  $\mathbf{O}_1 - \mathbf{O}_2$ . We note that Maxout networks are universal approximators that can effectively approximate functions. Thus, we use it to obtain the solutions of (9a) and (9b).

**Learning a Differentiable CG Network for (9c).** As stated in Section 3, although using FFT with boundary processing operations (e.g., edge taper and Laplacian smoothing [35]) can efficiently solve (9c), the results are not better than the CG-based solver, which can be observed in Table 5. However, using a CG-based solver is time-consuming. To generate latent clear images better, we develop a differentiable CG network to solve (9c). The CG method is used to solve the linear equation:

$$\mathbf{A}\mathbf{x}^{t+1} = \mathbf{b}, \quad (11)$$

where  $\mathbf{A}$  corresponds to the first term in (9c) and  $\mathbf{b}$  to the last term in (9c). Given  $\mathbf{x} \in \mathbb{R}^d$ , the CG method recursively computes conjugate vectors  $\mathbf{p}_l$  and

find the difference between desired  $\mathbf{x}^{t+1}$  and initial input  $\mathbf{x}^t$  as

$$\mathbf{x}^{t+1} - \mathbf{x}^t = \mathbf{s}_L = \sum_{l=0}^L \alpha_l \mathbf{p}_l,$$

where  $L$  is the iteration number upper-bounded by  $d$ , and  $\alpha_l$  is the weight calculated with  $\mathbf{p}_l$ .

However, if the size of matrix  $\mathbf{A}$  is large, using CG to solve (11) needs high computational costs. To overcome this problem, we develop a differentiable CG network based on a U-Net to compute the  $\mathbf{s}_L$ . The network design is motivated by the following reasons:

- As one of Krylov subspace methods [2], the solution  $\mathbf{s}_L$  can be found in the Krylov subspace  $\mathcal{K}_L(\mathbf{A}, \mathbf{r}_0) = \text{span}\{\mathbf{r}_0, \mathbf{A}\mathbf{r}_0, \dots, \mathbf{A}^{L-1}\mathbf{r}_0\}$ , where  $\mathbf{r}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}^t$  is the residual vector. In other words, the CG method is a function of  $\mathbf{A}$  and  $\mathbf{r}_0$ . Our CGNet takes  $\mathbf{r}_0$  as input and  $\mathbf{A}$  as parts of the network. Its output is  $\mathbf{s}_L$ , which behaves as the CG method.
- For a typical deconvolution problem,  $\mathbf{A}$  is composed of convolution  $\mathbf{A}_e$  and transpose convolution  $\mathbf{A}_d$  pairs as the first term in (9c).  $\mathbf{A}_e$  stands for the operation of  $\mathbf{G}_j\mathbf{H}$  and  $\mathbf{F}_i$ , and  $\mathbf{A}_d$  for  $\mathbf{H}^\top\mathbf{G}_j^\top$  and  $\mathbf{F}_i^\top$ . This observation can be intuitively connected to an encoder-decoder architecture, so we integrate  $\mathbf{A}_e$  into the encoder and  $\mathbf{A}_d$  into the decoder.
- The Conjugate Gradient method is sensitive to noise [34,53]. With an encoder-decoder architecture, U-Net is robust to noise.
- As the Conjugate Gradient method is a recursive algorithm, U-Net computes feature maps in a recursive fashion.

In the practical CG iterations,  $\mathbf{A}_e$  and  $\mathbf{A}_d$  are usually updated with iterative reweighted least squares (IRLS) to utilize the sparsity of priors [29]. We design a simple HypNet to estimate these weights. The network architecture of the HypNet is shown in Fig. 1. In addition, we note that the values of  $\rho_i$  and  $\rho_j$  in (9c) depend on the noise level. We design a simple NLNet to estimate the noise map  $\mathbf{m}_n$ , which plays a role similar to  $\rho_i$  and  $\rho_j$  (see Fig. 1). In contrast to most conventional methods, the NLNet computes the weight for each pixel, which is locally adaptive.

## 5 Experimental Results

### 5.1 Datasets and Implementation Details

**Training Dataset.** Similar to [10,11], the training data is composed of 4,744 images from the Waterloo Exploration dataset [36] and 400 images from the Berkeley segmentation dataset (BSD) [38]. To synthesize blurry images, we first generate 33,333 blur kernels by [55], where the sizes of these generated blur kernels range from  $13 \times 13$  pixels to  $35 \times 35$  pixels. We first crop image patches of  $128 \times 128$  pixels from each image, and then we randomly use generated blur kernels to generate blurry images. Each blurry image is randomly added Gaussian noise with noise levels from 1% to 5%.

**Test Datasets.** We evaluate our method on both synthetic datasets and real ones. For the synthetic case, we use the 100 images from BSD100 and 100 kernels by [55] to generate blurry images, similar to training data. We also use the SET5 [3] dataset with the kernels generated by [5] as our test dataset. In addition, we use the datasets by LEVIN [30] and LAI [27] for evaluation.

For the real-world dataset, we evaluate our model on the data of PAN [43], where 23 blurry images and 23 kernels estimated by their method are contained.

Table 1: Configuration of four models.  $T$  is the number of stages, i.e., how many duplicates are in the whole model.  $M$  and  $N$  denote the number of filters  $\mathbf{F}_i$  and  $\mathbf{G}_i$ , respectively. The PSNR (dB) and execution time are tested on SET5.

	Feather	Light	Heavy	Full
$T$	2	3	3	4
$M, N$	24	24	49	49
PSNR	32.51	32.78	33.11	33.43
Second	1.893	2.191	2.672	3.065

**Implementation Details.** We train the networks using the ADAM [20] optimizer with default parameter settings. The batch size is 8. The total training iterations is 1 million, and the learning rate is from  $1 \times 10^{-4}$  to  $1 \times 10^{-7}$ . We gradually decay the learning rate to  $1 \times 10^{-7}$  every 250,000 iterations and reset it to  $1 \times 10^{-4}$ ,  $5 \times 10^{-5}$  and  $2.5 \times 10^{-5}$  at the iteration of 250,001, 500,001 and 750,001, respectively. To constrain the network training, we apply the commonly used  $\ell_1$ -norm loss to the ground truth and the network output  $\mathbf{x}^T$ . The data augmentation (including  $\pm 90^\circ$  and  $180^\circ$  rotations, vertical and horizontal flipping) is used. In this paper, we train 4 models of different sizes, i.e., Feather, Light, Heavy and Full, whose configurations and simple results on SET5 are shown in Table 1.

It is worth noting that all the stages contain their own parameters and are end-to-end trained rather than share weights [71] or progressively trained [70]. We implement the networks in Pytorch [45] and train on one NVIDIA RTX 3090 GPU.

## 5.2 Quantitative Evaluation

We compare the proposed DSDNet with the state-of-the-art methods including IRCNN [73], SARFL [48], ADM\_UDM [21], CPCr [12], KerUNC [41], VEM [42], DWDN [10], SVMAP [9] and DRUNet [72]. These methods are fine-tuned using the same training dataset as Section 5.1 and choose the better ones from the fine-tuned and the original models for comparison.

Table 2: Average PSNR(dB)/SSIM of the deblurring results with Gaussian noise using different methods. We highlight the **best** and **second** best results. Our Full DSDNet wins first place, while our Light one also performs favorably against these state-of-the-art methods.

Dataset noise		IRCNN [73]		SFARL[48]		ADM_UDM [21]		CPCR [12]		KerUNC [41]		VEM [42]		DWDN [10]		SVMAP [11]		DRUNet [72]		DSDNet(Light)		DSDNet(Full)	
		PSNR	/ SSIM	PSNR	/ SSIM	PSNR	/ SSIM	PSNR	/ SSIM	PSNR	/ SSIM	PSNR	/ SSIM	PSNR	/ SSIM	PSNR	/ SSIM	PSNR	/ SSIM	PSNR	/ SSIM	PSNR	/ SSIM
LEVIN [30]	1%	30.61	/ 0.883	25.41	/ 0.600	31.48	/ 0.922	28.43	/ 0.858	32.02	/ 0.928	32.05	/ 0.927	34.89	/ 0.957	35.24	/ <b>0.962</b>	31.94	/ 0.922	<b>35.48</b>	/ 0.960	<b>36.62</b>	/ <b>0.965</b>
	3%	29.70	/ 0.864	16.82	/ 0.255	28.61	/ 0.812	25.61	/ 0.765	21.72	/ 0.416	29.47	/ 0.867	31.94	/ 0.916	31.20	/ 0.893	30.86	/ 0.905	<b>32.13</b>	/ <b>0.918</b>	<b>32.89</b>	/ <b>0.925</b>
	5%	28.98	/ 0.854	13.07	/ 0.157	27.83	/ 0.827	23.68	/ 0.703	18.25	/ 0.272	27.79	/ 0.819	30.21	/ 0.883	30.12	/ 0.876	29.79	/ 0.880	<b>30.21</b>	/ <b>0.883</b>	30.94	/ 0.893
BSD100 [38]	1%	29.20	/ 0.817	24.21	/ 0.568	29.39	/ 0.836	28.77	/ 0.829	29.23	/ 0.829	29.54	/ 0.848	31.10	/ 0.881	<b>31.52</b>	/ 0.888	30.36	/ 0.872	<b>31.50</b>	/ <b>0.892</b>	<b>32.01</b>	/ <b>0.898</b>
	3%	27.54	/ 0.762	15.80	/ 0.245	26.92	/ 0.722	25.96	/ 0.712	22.10	/ 0.430	27.09	/ 0.746	28.47	/ 0.797	27.94	/ 0.762	28.10	/ 0.798	<b>28.72</b>	/ <b>0.812</b>	<b>29.08</b>	/ <b>0.820</b>
	5%	27.04	/ 0.756	12.56	/ 0.146	26.04	/ 0.697	25.75	/ 0.688	18.99	/ 0.297	26.11	/ 0.698	27.50	/ 0.762	27.59	/ 0.763	27.19	/ 0.767	<b>27.61</b>	/ <b>0.774</b>	<b>27.96</b>	/ <b>0.782</b>
SET5 [9]	1%	30.15	/ 0.853	26.21	/ 0.632	30.52	/ 0.868	30.59	/ 0.875	30.45	/ 0.864	31.00	/ 0.875	32.18	/ 0.893	32.31	/ 0.892	30.84	/ 0.881	<b>32.78</b>	/ <b>0.899</b>	<b>33.43</b>	/ <b>0.905</b>
	3%	28.66	/ 0.813	15.50	/ 0.211	27.64	/ 0.709	27.94	/ 0.799	21.39	/ 0.376	28.40	/ 0.804	29.54	/ 0.838	28.78	/ 0.812	29.21	/ 0.841	<b>29.91</b>	/ <b>0.843</b>	<b>30.40</b>	/ <b>0.851</b>
	5%	27.55	/ 0.789	11.91	/ 0.122	26.75	/ 0.756	26.64	/ 0.754	17.74	/ 0.241	26.46	/ 0.732	28.13	/ <b>0.806</b>	28.02	/ 0.793	27.85	/ 0.805	<b>28.46</b>	/ 0.804	<b>28.89</b>	/ <b>0.814</b>

Table 3: Evaluation on the dataset LAI [27]. The best and second best results are highlighted as Table 2. The SATURATION results of SVMAP [11] are obtained by the model specifically trained for saturation scenes.

Subset	IRCNN [73]		ADM_UDM [21]		KerUNC [41]		VEM [42]		DWDN [10]		SVMAP [11]		DRUNet [72]		DSDNet	
	PSNR	/ SSIM	PSNR	/ SSIM	PSNR	/ SSIM	PSNR	/ SSIM	PSNR	/ SSIM	PSNR	/ SSIM	PSNR	/ SSIM	PSNR	/ SSIM
MANMADE	20.47	/ 0.604	22.43	/ 0.724	22.19	/ 0.725	22.71	/ 0.780	<b>24.02</b>	/ <b>0.836</b>	23.75	/ 0.776	20.62	/ 0.613	<b>25.44</b>	/ <b>0.859</b>
NATURAL	23.26	/ 0.636	25.04	/ 0.733	25.42	/ 0.757	25.29	/ 0.752	25.91	/ <b>0.814</b>	<b>26.23</b>	/ 0.778	23.25	/ 0.630	<b>27.01</b>	/ <b>0.837</b>
PEOPLE	28.04	/ 0.843	28.81	/ 0.866	28.80	/ 0.848	27.19	/ 0.723	30.02	/ <b>0.905</b>	<b>30.88</b>	/ 0.899	28.04	/ 0.838	<b>30.95</b>	/ <b>0.908</b>
SATURATION	16.99	/ 0.642	17.57	/ 0.627	17.70	/ 0.640	17.65	/ 0.600	17.90	/ 0.695	<b>18.75</b>	/ <b>0.733</b>	17.14	/ 0.658	<b>18.38</b>	/ <b>0.734</b>
TEXT	21.37	/ 0.828	25.13	/ 0.883	23.32	/ 0.855	24.92	/ 0.853	25.40	/ 0.877	<b>25.60</b>	/ <b>0.894</b>	21.79	/ 0.829	<b>28.13</b>	/ <b>0.920</b>
Overall	22.03	/ 0.710	23.80	/ 0.767	23.49	/ 0.765	23.55	/ 0.742	24.65	/ <b>0.825</b>	<b>25.04</b>	/ 0.816	22.17	/ 0.714	<b>25.98</b>	/ <b>0.852</b>

PSNR and SSIM [66] are used for quantitative evaluation. All the quantitative evaluations are conducted without border cropping for fair comparisons.

Table 2 shows quantitative evaluation results on the synthetic datasets. The proposed method generates results with higher PSNR and SSIM values. In addition, we note that the proposed light-weighted model generates favorable results against the state-of-the-art, showing the effectiveness of the proposed algorithm. Due to the space limit, we only present the evaluation results of the Full DSDNet hereafter.

We then evaluate our method on the LAI [27] dataset. Because it contains MANMADE, NATURAL, PEOPLE, SATURATION, and TEXT subsets, we present the evaluation accordingly. Table 3 shows that our method performs better than the evaluated methods. Similar to Table 2, our method also achieves the highest PSNR and SSIM in most tests, except for the SATURATION subset. We note that Dong et al. [11] specifically train a model for saturation scenes; thus, this method performs slightly better. However, our model is only trained with common scenes but comparable in terms of PSNR on the SATURATION images. Moreover, the SSIM values of our method are better than SVMAP [11], demonstrating the efficiency and robustness of our approach.

We also quantitatively evaluate our method on real-world blurry images and estimated kernels from PAN [43]. Since the ground truth images are unavailable, we use the no-reference BRISQUE [39] and PIQE [65] metrics for evaluation. Our model achieves the best score in BRISQUE and second place in PIQE, as shown in Table 4. As BRISQUE is a metric based on subject scoring, Table 4

Table 4: Quantitative evaluation of real cases PAN [43]. Non-reference image quality metrics BRISQUE [39] and PIQE [64] are used.

	IRCNN [73]	ADM_UDM [21]	KerUNC [41]	VEM [42]	DWDN [10]	SVMAP [11]	DRUNet [72]	DSDNet
BRISQUE	43.484	36.598	37.816	<a href="#">33.663</a>	34.027	35.508	46.774	<b>33.129</b>
PIQE	78.700	67.605	65.674	<b>44.942</b>	51.348	56.032	81.074	<a href="#">49.788</a>

shows that our model generates more subjectively satisfying results than other state-of-the-art methods.

### 5.3 Qualitative Evaluation

We show visual comparisons of a synthesized and a real-world case in Fig. 3 Fig. 4, respectively.

Fig. 3 shows the results of MANMADE from the dataset LAI [27]. The evaluated methods generate blur results. In contrast, our method reconstructs better images (e.g., the wood texture is better restored, as shown in both red and green boxes).

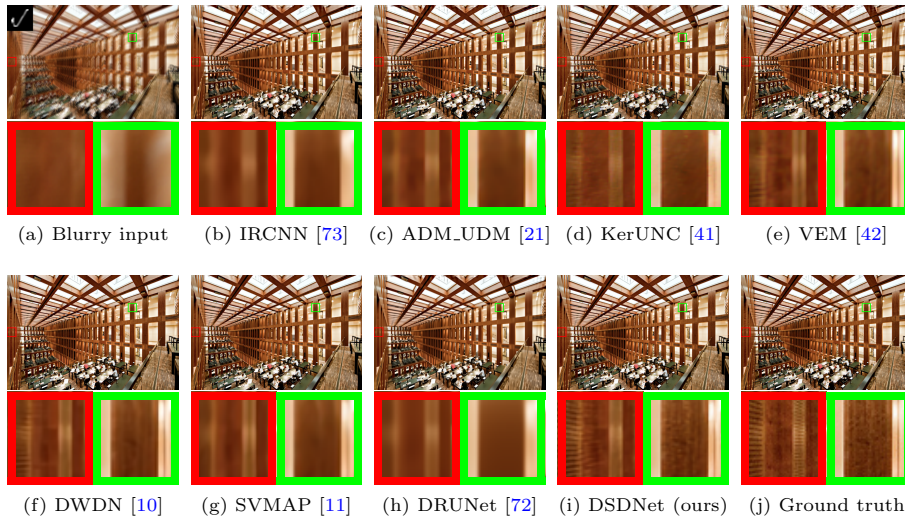


Fig. 3: A synthetic case comes from LAI [27]. Our method restores clearer images with finer details (e.g., the wood texture).

Fig. 4 shows the deblurred results of a pair of real-world blurry images and estimated kernel [43]. Our method restores the text to sharpness in the red boxes and the blackest and sharpest eyebrow in the green boxes. In contrast, other



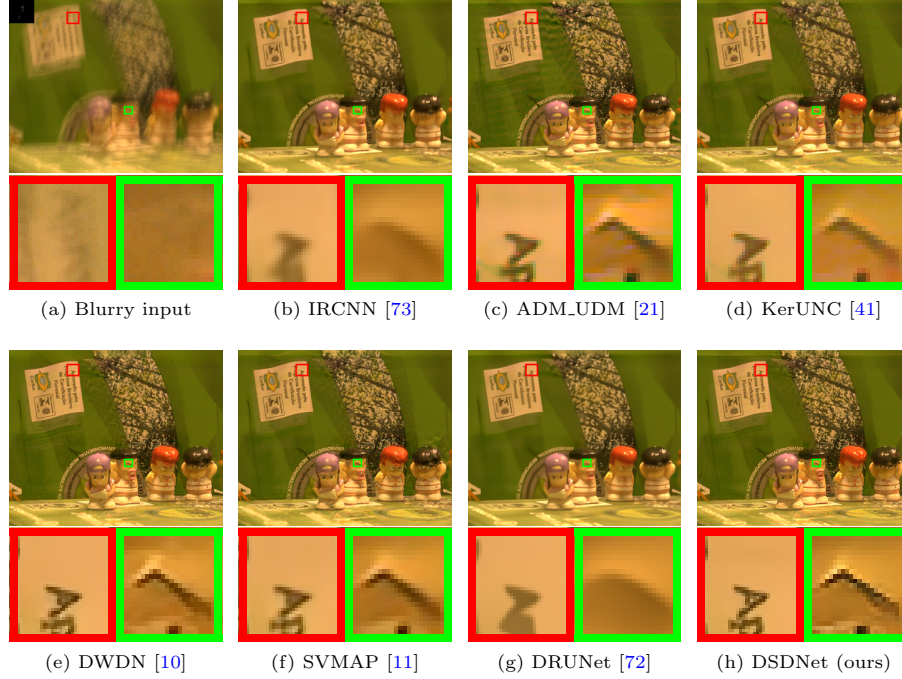


Fig. 4: A real-world case whose kernel is estimated by [43]. Our method generates a better image with clearer text and a sharper eyebrow. In the olive background, our method produces the sharpest result without noise.

methods cannot restore the text well and mix the eyebrow with the skin color. Furthermore, they also generate artifacts and noise in the olive background.

#### 5.4 Ablation Study

In this section, we design experiments to show the efficiency of the proposed discriminative shrinkage functions and the differentiable CGNet. Table 5 shows the ablation results w.r.t. different baselines. In this study, we train 7 models based on the architecture of the Heavy DSDNet. We compare the number of floating point operations (FLOPs) and the parameters in this study.

To validate the effects of the  $\mathbf{F}_i$  and  $\mathbf{G}_j$ , we train a model without estimating these two filters, denoted by “w/o  $\mathbf{F}, \mathbf{G}$ ”. Without the feature maps coming from them, we can only learn the shrinkage functions from RGB inputs. Table 5 shows that the PSNR value of the results by the baseline method without  $\mathbf{F}_i$  and  $\mathbf{G}_j$  is at least 6.45dB lower than that of our approach.

We also evaluate the effect of the Maxout layers by replacing them with ReLU. Table 5 shows that the method using ReLU does not generate better



Table 5: Ablation study on SET5. “w/o  $\mathbf{F}$ ,  $\mathbf{G}$ ” is out of the convolutional layers before the Maxout layers; “ReLU” and “RBF” replace the Maxout layers with ReLU and RBF layers, respectively; “CG” performs conventional Conjugate Gradient method for deconvolution rather than CGNet, “FFT” performs FFT deconvolution with edge taper [25], and the superscript  $\dagger$  means the input is denoised by DRUNet [72] first.

	w/o $\mathbf{F}$ , $\mathbf{G}$	ReLU	RBF	CG	CG $^\dagger$	FFT	FFT $^\dagger$	DSDNet
PSNR(dB)	26.66	32.78	32.98	29.07	32.39	31.30	32.03	33.11
FLOPs(G)	136.26	464.94	468.03	470.73	559.97	288.89	391.57	466.32
Parameters(M)	1.04	237.85	237.85	0.31	32.95	0.27	32.09	237.87
Gain(dB)	-6.45	-0.33	-0.13	-4.04	-0.72	-1.81	-1.08	-/-

results than the proposed method, suggesting the effectiveness of the Maxout layers.

As RBFs are usually used to approximate shrinkage functions, one may wonder whether using them generates better results or not. To answer this question, we replace the Maxout layers with the commonly-used Gaussian RBFs. Table 5 shows that the PSNR value of the method using RBFs is at least 0.13dB lower than that of our method, indicating the effectiveness of the proposed method.

Finally, to demonstrate the efficiency of the proposed CGNet, we train 2 models with the conventional CG method and 2 models with FFT deconvolution. The CG method is unstable with respect to even small perturbations [37,18]. Each optimization step in training and the existence of noise may cause the divergence of the CG method. Hence we have to reduce the learning rate and apply gradient clipping to avoid the gradient exploding during the training. However, it still performs poorly even if the training can be finished. To make the training more feasible, we denoise the inputs first by DRUNet [72], and this model is denoted as “CG $^\dagger$ ”. The training of “CG $^\dagger$ ” is smoother, the learning rate can be set as that of the DSDNet, and the performance is much better than generic “CG”. With another model to denoise, the computational cost is about 26% more FLOPs, and it takes more than twice the time for training compared to the DSDNet.

Similar to CG ones, we also provide results of deconvolution via FFT with the artifact processing operation by [25], i.e., “FFT” and “FFT $^\dagger$ ”. Although FFT ones are much faster than CG ones, the gap between “CG $^\dagger$ ” and “FFT $^\dagger$ ” is considerable, as mentioned in Section 3. As for the test time on SET5, “CG” is 5.6001 seconds, “CG $^\dagger$ ” is 5.8252 seconds, “FFT” is 3.6941 seconds, “FFT $^\dagger$ ” is 4.0330 seconds, and DSDNet is 2.6717 seconds. These result show the efficiency of the proposed CGNet. We include the ablation study on HypNet and NLNet in the supplementary material.

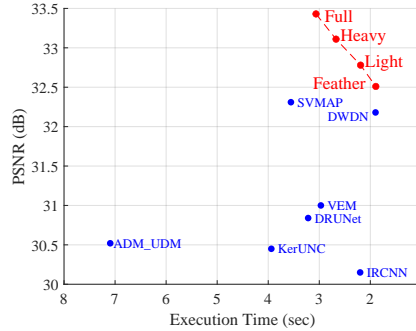


Fig. 5: Speed and accuracy trade-off. The results are evaluated on the SET5 dataset with 1% Gaussian noise. Results of this work are shown in red points; the fastest Feather DSDNet still performs favorably against the state-of-the-art methods (blue points) in PSNR.

### 5.5 Execution Time Analysis

We analyze the execution time of the proposed methods and state-of-the-art ones. All the execution times are evaluated on one nVidia RTX 2080Ti GPU. Fig. 5 shows that our models are faster and more accurate than the state-of-the-art methods. Among these methods, our Feather model is a little bit faster than DWDN by 0.0052 seconds yet outperforms all other methods in PSNR.

### 5.6 Limitations

Although better performance on various datasets has been achieved, our method has some limitations. Our model cannot deal with blurry images containing significant saturation regions, which may lead to overflow. More analysis can be found in the supplementary material.

## 6 Conclusion

In this paper, we present a fully learnable MAP model for non-blind deconvolution. We formulate the data and regularization terms as the learnable ones and split the deconvolution model into data-related and regularization-related sub-problems in the ADMM framework. Maxout layers are used to learn the discriminative shrinkage functions, which directly approximate the solutions of these two sub-problems. We have further developed a CGNet to restore the images effectively and efficiently. With a reasonable design, the size of our model is flexibly adjustable while keeping competitive in performance. Extensive evaluations on the benchmark datasets demonstrate that the proposed model performs favorably against the state-of-the-art non-blind deconvolution methods in terms of quantitative metrics, visual quality, and computational efficiency.

## References

1. Aljadaany, R., Pal, D.K., Savvides, M.: Douglas-rachford networks: Learning both the image prior and data fidelity terms for blind image deconvolution. In: CVPR. pp. 10235–10244 (2019) [3](#)
2. Barrett, R., Berry, M., Chan, T.F., Demmel, J., Donato, J., Dongarra, J., Eijkhout, V., Pozo, R., Romine, C., Van der Vorst, H.: Templates for the solution of Linear Systems: Building Blocks for Iterative Methods. SIAM (1994) [2](#), [8](#)
3. Bevilacqua, M., Roumy, A., Guillemot, C., Alberi-Morel, M.L.: Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In: BMVC (2012) [9](#), [10](#)
4. Bigdeli, S.A., Zwicker, M., Favaro, P., Jin, M.: Deep mean-shift priors for image restoration. In: NeurIPS (2017) [3](#), [6](#)
5. Chakrabarti, A.: A neural approach to blind motion deblurring. In: ECCV. pp. 221–235 (2016) [9](#)
6. Chan, T.F., Wong, C.K.: Total variation blind deconvolution. IEEE TIP **7**(3), 370–375 (1998) [3](#)
7. Chen, L., Zhang, J., Pan, J., Lin, S., Fang, F., Ren, J.S.: Learning a non-blind deblurring network for night blurry images. In: CVPR. pp. 10542–10550 (2021) [3](#)
8. Cho, S., Wang, J., Lee, S.: Handling outliers in non-blind image deconvolution. In: ICCV. pp. 495–502 (2011) [1](#)
9. Dong, J., Pan, J., Sun, D., Su, Z., Yang, M.H.: Learning data terms for non-blind deblurring. In: ECCV. pp. 748–763 (2018) [2](#), [3](#), [9](#)
10. Dong, J., Roth, S., Schiele, B.: Deep wiener deconvolution: Wiener meets deep learning for image deblurring. In: NeurIPS (2020) [1](#), [2](#), [4](#), [8](#), [9](#), [10](#), [11](#), [12](#)
11. Dong, J., Roth, S., Schiele, B.: Learning spatially-variant map models for non-blind image deblurring. In: CVPR. pp. 4886–4895 (2021) [8](#), [10](#), [11](#), [12](#)
12. Eboli, T., Sun, J., Ponce, J.: End-to-end interpretable learning of non-blind image deblurring. In: ECCV. pp. 314–331. Springer (2020) [9](#), [10](#)
13. Gao, H., Tao, X., Shen, X., Jia, J.: Dynamic scene deblurring with parameter selective sharing and nested skip connections. In: CVPR. pp. 3848–3856 (2019) [1](#)
14. Geman, D., Reynolds, G.: Constrained restoration and the recovery of discontinuities. IEEE TPAMI **14**(3), 367–383 (1992) [1](#)
15. Gong, D., Zhang, Z., Shi, Q., van den Hengel, A., Shen, C., Zhang, Y.: Learning deep gradient descent optimization for image deconvolution. IEEE transactions on neural networks and learning systems **31**(12), 5468–5482 (2020) [3](#)
16. Goodfellow, I., Warde-Farley, D., Mirza, M., Courville, A., Bengio, Y.: Maxout networks. In: ICML. pp. 1319–1327 (2013) [2](#), [6](#), [7](#)
17. Gu, S., Zhang, L., Zuo, W., Feng, X.: Weighted nuclear norm minimization with application to image denoising. In: CVPR. pp. 2862–2869 (2014) [3](#)
18. Hadamard, J.: Lectures on Cauchy’s Problem in Linear Partial Differential Equations. Courier Corporation (2003) [13](#)
19. Jancsary, J., Nowozin, S., Rother, C.: Loss-specific training of non-parametric image restoration models: A new state of the art. In: ECCV. pp. 112–125 (2012) [3](#)
20. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014) [9](#)
21. Ko, H.C., Chang, J.Y., Ding, J.J.: Deep priors inside an unrolled and adaptive deconvolution model. In: ACCV (2020) [3](#), [9](#), [10](#), [11](#), [12](#)

22. Kong, S., Wang, W., Feng, X., Jia, X.: Deep red unfolding network for image restoration. *IEEE TIP* **31**, 852–867 (2021) [3](#)
23. Krishnan, D., Fergus, R.: Fast image deconvolution using hyper-laplacian priors. In: *NeurIPS*. pp. 1033–1041 (2009) [1](#), [3](#)
24. Krishnan, D., Tay, T., Fergus, R.: Blind deconvolution using a normalized sparsity measure. In: *CVPR*. pp. 233–240 (2011) [1](#)
25. Kruse, J., Rother, C., Schmidt, U.: Learning to push the limits of efficient fft-based image deconvolution. In: *ICCV*. pp. 4586–4594 (2017) [3](#), [13](#)
26. Kupyn, O., Budzan, V., Mykhailych, M., Mishkin, D., Matas, J.: Deblurgan: Blind motion deblurring using conditional adversarial networks. In: *CVPR*. pp. 8183–8192 (2018) [1](#), [4](#)
27. Lai, W.S., Huang, J.B., Hu, Z., Ahuja, N., Yang, M.H.: A comparative study for single image blind deblurring. In: *CVPR*. pp. 1701–1709 (2016) [9](#), [10](#), [11](#)
28. Levin, A., Fergus, R., Durand, F., Freeman, W.T.: Image and depth from a conventional camera with a coded aperture. *ACM TOG* **26**(3), 70–es (2007) [2](#), [3](#), [6](#)
29. Levin, A., Weiss, Y.: User assisted separation of reflections from a single image using a sparsity prior. *IEEE TPAMI* **29**(9), 1647–1654 (2007) [6](#), [8](#)
30. Levin, A., Weiss, Y., Durand, F., Freeman, W.T.: Understanding and evaluating blind deconvolution algorithms. In: *CVPR*. pp. 1964–1971 (2009) [1](#), [9](#), [10](#)
31. Levin, A., Weiss, Y., Durand, F., Freeman, W.T.: Understanding blind deconvolution algorithms. *IEEE TPAMI* **33**(12), 2354–2367 (2011) [1](#)
32. Li, L., Pan, J., Lai, W.S., Gao, C., Sang, N., Yang, M.H.: Blind image deblurring via deep discriminative priors. *IJCV* **127**(8), 1025–1043 (2019) [3](#), [6](#)
33. Li, Y., Tofghi, M., Geng, J., Monga, V., Eldar, Y.C.: Deep algorithm unrolling for blind image deblurring. *arXiv preprint arXiv:1902.03493* (2019) [3](#)
34. Liu, C.S.: Modifications of steepest descent method and conjugate gradient method against noise for ill-posed linear systems. *Commun. Numer. Anal* **2012** (2012) [8](#)
35. Liu, R., Jia, J.: Reducing boundary artifacts in image deconvolution. In: *ICIP*. pp. 505–508 (2008) [7](#)
36. Ma, K., Duanmu, Z., Wu, Q., Wang, Z., Yong, H., Li, H., Zhang, L.: Waterloo exploration database: New challenges for image quality assessment models. *IEEE TIP* **26**(2), 1004–1016 (2016) [8](#)
37. Marin, L., Háo, D.N., Lesnic, D.: Conjugate gradient-boundary element method for a cauchy problem in the lamé system. *WIT Transactions on Modelling and Simulation* **27** (2001) [13](#)
38. Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: *ICCV*. pp. 416–423 (2001) [8](#), [10](#)
39. Mittal, A., Moorthy, A.K., Bovik, A.C.: No-reference image quality assessment in the spatial domain. *IEEE TIP* **21**(12), 4695–4708 (2012). <https://doi.org/10.1109/TIP.2012.2214050> [10](#), [11](#)
40. Nah, S., Hyun Kim, T., Mu Lee, K.: Deep multi-scale convolutional neural network for dynamic scene deblurring. In: *CVPR*. pp. 3883–3891 (2017) [4](#)
41. Nan, Y., Ji, H.: Deep learning for handling kernel/model uncertainty in image deconvolution. In: *CVPR*. pp. 2388–2397 (2020) [9](#), [10](#), [11](#), [12](#)
42. Nan, Y., Quan, Y., Ji, H.: Variational-em-based deep learning for noise-blind image deblurring. In: *CVPR*. pp. 3626–3635 (2020) [9](#), [10](#), [11](#)
43. Pan, J., Sun, D., Pfister, H., Yang, M.H.: Blind image deblurring using dark channel prior. In: *CVPR*. pp. 1628–1636 (2016) [1](#), [2](#), [9](#), [10](#), [11](#), [12](#)

44. Parikh, N., Boyd, S.: Proximal algorithms. *Foundations and Trends in Optimization* **1**(3), 127–239 (2014) [2](#), [6](#)
45. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S.: Pytorch: An imperative style, high-performance deep learning library. In: Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., Garnett, R. (eds.) *NeurIPS*. pp. 8024–8035. Curran Associates, Inc. (2019), <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf> [9](#)
46. Perrone, D., Favaro, P.: Total variation blind deconvolution: The devil is in the details. In: *CVPR*. pp. 2909–2916 (2014) [3](#)
47. Qiu, H., Hammernik, K., Qin, C., Rueckert, D.: Gradirn: Learning iterative gradient descent-based energy minimization for deformable image registration. *arXiv preprint arXiv:2112.03915* (2021) [3](#)
48. Ren, D., Zuo, W., Zhang, D., Zhang, L., Yang, M.H.: Simultaneous fidelity and regularization learning for image restoration. *IEEE TPAMI* (2019) [2](#), [9](#), [10](#)
49. Ren, W., Cao, X., Pan, J., Guo, X., Zuo, W., Yang, M.H.: Image deblurring via enhanced low-rank prior. *IEEE TIP* **25**(7), 3426–3437 (2016) [1](#)
50. Richardson, W.H.: Bayesian-based iterative method of image restoration. *JoSA* **62**(1), 55–59 (1972) [1](#)
51. Roth, S., Black, M.J.: Fields of experts: A framework for learning image priors. In: *CVPR*. pp. 860–867 (2005) [3](#)
52. Rudin, L.I., Osher, S.: Total variation based image restoration with free local constraints. In: *ICIP*. vol. 1, pp. 31–35 (1994) [3](#)
53. Ryabtsev, A.: The error accumulation in the conjugate gradient method for degenerate problem. *arXiv preprint arXiv:2004.10242* (2020) [8](#)
54. Samuel, K.G., Tappen, M.F.: Learning optimized map estimates in continuously-valued mrf models. In: *CVPR*. pp. 477–484 (2009) [3](#)
55. Schmidt, U., Jancsary, J., Nowozin, S., Roth, S., Rother, C.: Cascades of regression tree fields for image restoration. *IEEE TPAMI* **38**(4), 677–689 (2015) [3](#), [8](#), [9](#)
56. Schmidt, U., Roth, S.: Shrinkage fields for effective image restoration. In: *CVPR*. pp. 2774–2781 (2014) [2](#), [3](#), [6](#)
57. Schmidt, U., Rother, C., Nowozin, S., Jancsary, J., Roth, S.: Discriminative non-blind deblurring. In: *CVPR*. pp. 604–611 (2013) [3](#)
58. Schuler, C.J., Christopher Burger, H., Harmeling, S., Scholkopf, B.: A machine learning approach for non-blind image deconvolution. In: *CVPR*. pp. 1067–1074 (2013) [2](#)
59. Suin, M., Purohit, K., Rajagopalan, A.: Spatially-attentive patch-hierarchical network for adaptive motion deblurring. In: *CVPR*. pp. 3606–3615 (2020) [1](#)
60. Sun, J., Cao, W., Xu, Z., Ponce, J.: Learning a convolutional neural network for non-uniform motion blur removal. In: *CVPR*. pp. 769–777 (2015) [4](#)
61. Sun, L., Cho, S., Wang, J., Hays, J.: Edge-based blur kernel estimation using patch priors. In: *ICCP*. pp. 1–8 (2013) [3](#)
62. Tao, X., Gao, H., Shen, X., Wang, J., Jia, J.: Scale-recurrent network for deep image deblurring. In: *CVPR*. pp. 8174–8182 (2018) [1](#), [4](#)
63. Tappen, M.F., Liu, C., Adelson, E.H., Freeman, W.T.: Learning gaussian conditional random fields for low-level vision. In: *CVPR*. pp. 1–8 (2007) [3](#)

64. Venkatanath, N., Praneeth, D., Bh, M.C., Channappayya, S.S., Medasani, S.S.: Blind image quality evaluation using perception based features. In: National Conference on Communications (NCC). pp. 1–6 (2015) [11](#)
65. Wang, Y., Yang, J., Yin, W., Zhang, Y.: A new alternating minimization algorithm for total variation image reconstruction. *SIAM Journal on Imaging Sciences* **1**(3), 248–272 (2008) [3](#), [10](#)
66. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE TIP* **13**(4), 600–612 (2004) [10](#)
67. Xiang, J., Dong, Y., Yang, Y.: Fista-net: Learning a fast iterative shrinkage thresholding network for inverse problems in imaging. *IEEE TMI* (2021) [3](#)
68. Yang, Y., Sun, J., Li, H., Xu, Z.: Deep admm-net for compressive sensing mri. In: NeurIPS. pp. 10–18 (2016) [3](#)
69. Zhang, J., Ghanem, B.: Ista-net: Interpretable optimization-inspired deep network for image compressive sensing. In: CVPR. pp. 1828–1837 (2018) [3](#)
70. Zhang, J., shan Pan, J., Lai, W.S., Lau, R.W.H., Yang, M.H.: Learning fully convolutional networks for iterative non-blind deconvolution. In: CVPR. pp. 6969–6977 (2017) [2](#), [3](#), [5](#), [6](#), [9](#)
71. Zhang, K., Gool, L.V., Timofte, R.: Deep unfolding network for image super-resolution. In: CVPR. pp. 3217–3226 (2020) [3](#), [9](#)
72. Zhang, K., Li, Y., Zuo, W., Zhang, L., Van Gool, L., Timofte, R.: Plug-and-play image restoration with deep denoiser prior. *IEEE TPAMI* (2021) [5](#), [9](#), [10](#), [11](#), [12](#), [13](#)
73. Zhang, K., Zuo, W., Gu, S., Zhang, L.: Learning deep cnn denoiser prior for image restoration. In: CVPR. pp. 3929–3938 (2017) [2](#), [5](#), [9](#), [10](#), [11](#), [12](#)
74. Zhang, K., Zuo, W., Zhang, L.: Deep plug-and-play super-resolution for arbitrary blur kernels. In: CVPR. pp. 1671–1681 (2019) [1](#)
75. Zoran, D., Weiss, Y.: From learning models of natural image patches to whole image restoration. In: ICCV. pp. 479–486 (2011) [3](#)