Multimedia Tools and Applications manuscript No.
(will be inserted by the editor)

DeepDeblur: Text Image Recovery from Blur to Sharp

Jianhan Mei · Ziming Wu · Xiang Chen · Yu Qiao · Henghui Ding · Xudong Jiang

Received: date / Accepted: date

Abstract Digital images could be degraded by a variety of blur during the image acquisition (*i.e.* relative motion of cameras, electronic noise, capturing defocus, and so on). Blurring images can be computationally modeled as the result of a convolution process with the corresponding blur kernel and thus, image deblurring can be regarded as a deconvolution operation. In this paper, we explore to deblur images by approximating blind deconvolutions using a deep neural network. Different deep neural network structures are investigated to evaluate the their deblurring capabilities, which contributes to the optimal design of a network architecture. It is found that shallow and narrow networks are not capable of handling complex motion blur. We thus, present a deep network with 20 layers to cope with text image blur. In addition, a novel network structure with Sequential Highway Connections (SHC) is leveraged to gain superior convergence. The experiment results demonstrate the state-of-the-art performance of the proposed framework with higher visual quality of the delurred images.

Keywords Text Deblurring · Convolutional Neural Network (CNN) · Blind Deconvolution · Short Connection

Jianhan Mei[†], Henghui Ding[‡], Xudong Jiang[*]
School of Electrical and Electronics Engineering, Nanyang Technological University, Singapore
[†]E-mail: jianhan001@e.ntu.edu.sg
[‡]E-mail: ding0093@e.ntu.edu.sg
[*]E-mail: exdjiang@ntu.edu.sg

Ziming Wu
The Hong Kong University of Science and Technology, Hong Kong, China
E-mail: zwual@connect.ust.hk

Xiang Chen
Technische Universität Darmstadt, Darmstadt, Germany
E-mail: xiang.chen@gcc.tu-darmstadt.de

Yu Qiao
Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China
E-mail: yu.qiao@siat.ac.cn

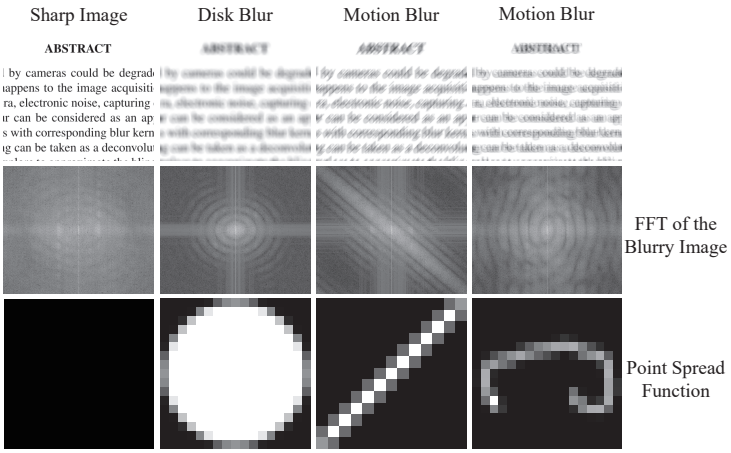

Fig. 1 Different blur kernels produce different blurry images. Blurry function of the last column are generated by [2, 3]

1 Introduction

Image blur has always been a major problem in the field of computer vision. In practice, there are various causes of image blur, *i.e.*, aberrations in the optical system, relative motion during capturing, atmospheric turbulence effects, electronic noise, random noise in the environment, and *etc*. Different kinds of degradations result in a large variety of image blur, which makes the deblurring task challenging and complicated. Fig. 1 shows images degraded by the disk blur and the motion blur, their DFT and Point Spread Function (PSF) respectively. Among them, the presented motion blurs are generated from two different ways. One is from the camera relative movement of single direction while the other is from a more unstable motion trajectory.

We think that image blurring process can be modeled as the convolution operation between a sharp image and a blur kernel, additively merging with some noise (Fig. 1). Thus, sharp image restoration can be viewed as a deconvolution process on a blurry image. Most of the traditional deblurring approaches infer the blur kernels by incorporating the prior of image features. For example, modeling image gradient by the mixed Gaussian distribution [20], making more non-smooth to the output image if the noise is well suppressed [5], constraining high-frequency information for the image [6], from coarse to fine iterative optimization strategy [1, 18], and *etc*. The sharp image can be obtained by a non-blind deconvolution process with the estimated blur kernel.

It is well known that deep convolution neural networks have gained great success in the most of vision tasks due to the superior capability of extracting high-level features. In the field of image deblurring, there are also some amazing achievements. Ruomei *et al*. developed a deep belief network to regress the blur parameters through blurring classification of targeted image [21], Jian *et al*. used deep convolutional networks to find the parameters of unidirectional motion blur for each local region on the image then deblur the image with blind deconvolution method [9], and so on. Comparing with the traditional blind deconvolution

algorithms, Convolutional Neural Networks (CNNs) are more dependent on the learning strategies. In other words, CNN based deblurring frameworks are data-driven approaches that learn blur statistical distribution from data and predict the sharp image. However, the coefficient space of the blur kernel is infinite, making it an ill-posed problem in most cases.

In this paper, we present a novel deep neural network structure to deal with text image blur, which achieves the state-of-the-art performance. Our main contributions are:

- Exploring to approximate the blind deconvolution process by a deep neural network;

- Different network structures are investigated to get insights on network architecture design regarding their deblurring capability;

- By leveraging the proposed Sequential Highway Connection (SHC), our framework achieves superior performance on the text image dataset.

2 Related Work

Kernel Estimation: Li *et al*. claimed that the blurring kernel estimation is not always profited by the edge priors, hence they proposed a kernel estimation method based on the spatial prior and the iterative support detection kernel refinement. It effectively reduces the enforce sparsity adverse effect from the hard threshold of the kernel elements [13]. Anat *et al*. demonstrated the naive MAP limitation and then showed that mostly favors non-blur explanations, they declared that a single MAP estimation of the kernel can be well constrained because the blurring kernel size is often smaller than the image size [1].

Deep Learning: Internal covariate shift often happens during the deep convolutional neural network training because the distribution of each layers inputs changes. This leads to the network sensitive to parameter initialization and slow convergence, Sergey *et al*. addressed this problem by normalizing layer inputs. They integrated the normalization as a part of the model architecture and performed it to each training mini batch [22]. It is well known that deeper neural networks have greater capability but are difficult to train, Kaiming *et al*. proposed the residual function to set up a shortcut to alleviate the vanishing or exploding gradients problem. This makes the network easier to deliver the signal of the loss function in a very deep structure [11].

Deep Learning based Deblurring: With the development of deep learning, CNN based deblurring methods appear. By using CNN, [9] [4] predicted the blur kernel in spatial and Fourier domain respectively. In [7], the fully convolutional network was used to estimate the motion flow of the motion blur. Inspired by the pixel to pixel approaches, [16] [17] used multi-scale CNN to directly generate the sharp image. [19] performed blind kernel-free deblurring based on the pix2pix framework [8]. And [12] combined Generative Adversarial Networks (GAN) with pixel level prediction by using the conditional adversarial networks.

Text Deblurring: Jinshan *et al*. utilized the salient edge selection that can effectively estimate the blurring kernel. Thus they considered a kernel estimation technique based on the L0-regularized prior of image intensity and image gradient [10]. Li *et al*. designed a deep learning framework combines the traditional

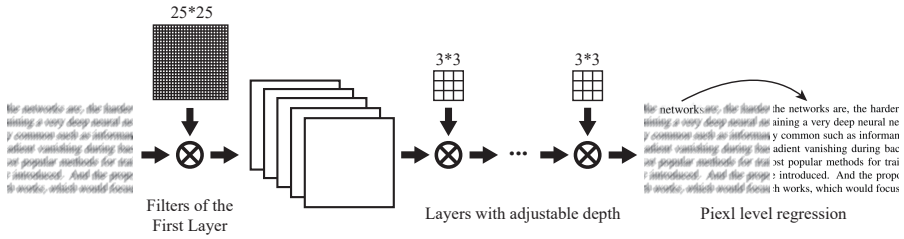

Fig. 2 The brief structure of the proposed network.

optimization-based schemes for image deconvolution. It shows that the deep convolutional neural network can make a decent performance for challenging image deblurring tasks [14]. Michal *et al*. trained a convolutional neural network with blurring inputs without any specific blur and/or noise models, which indicates that deep learning affords the text document deblurring task on combined kernels of defocus and motion blur [15].

3 Approach

3.1 Deconvolution Mapping

Before introducing the proposed method, a basic analysis of using CNN for image deblurring will be demonstrated in this section. As aforementioned, without considering additive noise, the process of image blurring can be approximated as the convolution between the sharp image and a blur kernel:

$$I_B = I_S * k \tag{1}$$

where, I_B is the blurry image, I_S is the sharp image and k represents the blurring kernel that makes the image blurry. And $*$ denotes the sign of convolution operation.

If the blurring kernel is considered as a filter that blurs the sharp image, the process of image blurring can, therefore, be re-formulated as:

$$I_B = K[I_S] \tag{2}$$

where, $K[\cdot]$ can be regarded as the filter in Fourier domain corresponding to its spatial form k.

Meanwhile, the deblurring process is casted as the convolution operation between the blurry image and the inverse filter corresponding to K, which can also be regarded as a deconvolution mapping between the blur and sharp images:

$$I_S = K^{-1}[I_B] \tag{3}$$

where, $K^{-1}[\cdot]$ is the inverse filter corresponding to the blurring filter K.

So far, the problem of deblurring is converted into designing an inverse blurring filter as done by most of the previous works. Considering the deep neural network as a complex non-linear mapping function, we assume that it is capable of learning to approximate the deconvolution mapping between the degraded image and the

corresponding sharp image. First, to validate the assumption, we start with a simple experiment under the case of simple motion blur to explore the deblurring ability of a neural network.

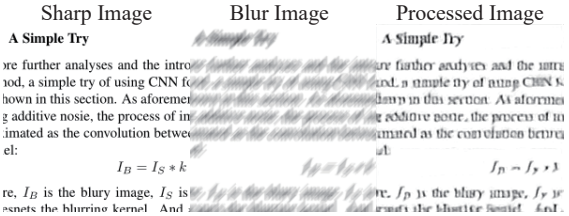

Fig. 3 Result of a simple trial experiment.

In this experiment, the data used for network training and testing are all image patches from electronic documents (more information about the dataset can be found in next section). The patches are blurred by blurring kernels generated automatically by different elementary PSFs.

Following the structure illustration of Fig. 2, a pixel level regression network with only 3 convolution layers is built. The network is trained and tested on a simple dataset in which all the sharp text images are blurred by using single-direction motion kernels but with different kernel scales. The network is trained by directly regressing the blurry image pixel to sharp image pixel. Adopting fully convolutional structure, the input and output of the network are in the same resolution. One visualized example of the simple experiment is illustrated in Fig. 3.

As shown in Fig. 3, the network shows good performance in the task and a PSNR of 15.133dB can be obtained on the testing set. However, due to the lack of diversity of the blurring kernel category, this task is much easier than the situation we meet in the real world. Considering the linear inverse blurring filter, the network may not have many parameters to learn and may not need high level non-linear functions. On the other hand, the limited diversity of blurring kernel patterns makes the network easy to overfit the training set. To further explore the capability of the networks, more diverse and complex blur kernels are required.

3.2 Capability of the Network

To address the problems of blur degradation in real scenarios, more sophisticated blur kernels should be considered. Following the trajectory and PSF generation method in [2, 3], randomly generated PSFs are used in our dataset which tries to approximate real blurring kernels. The more complex of the blur the data contain, the more complicated mapping does our network require. If the above-mentioned shallow network with 3 layers is deployed, the PSNR on the testing set decreases rapidly. In other words, the network needs stronger non-linear mapping ability corresponding to the complexity of the image blur.

When the generated kernels have much more degree of freedom, it is more difficult to learn the underlying patterns. For the network, it needs to solve a problem with a larger solution space while a shallow and simple network may not have such capability.

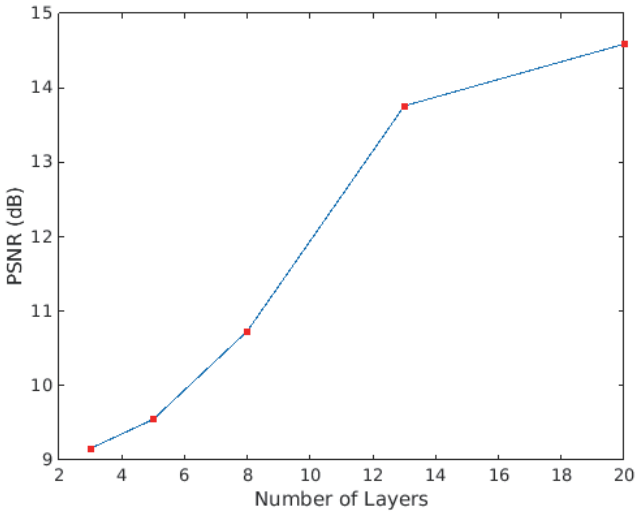

Fig. 4 Deblurring performance with respect to the number of the network layers.

Increasing the number of the layers is an intuitive approach to facilitate the network with strong mapping capability, in which more parameters are involved. Networks with different depth are tested. Results are shown in Fig 4.

From Fig. 4, with the increase of network depth, its performance can be significantly improved. To well train the deep neural network, all the layers are trained with batch normalization [22]. However, during training the deep network, it could be still very hard to ensure a deep network have a good astringency. To obtain better convergence of deep networks, several methods and a new type of highway connection will be introduced in next section.

3.3 Proposed Method

To facilitate the neural network with the higher capability of deconvolution mapping ability, a deep neural network with 20 layers is proposed to deal with the text image blur. Since a much larger number of parameters are involved, it is difficult to make a neural network converge due to problems such as information loss during forward and gradient vanishing during backward. Thus, a typical batch normalization strategy, a residual connection and a new highway connection are introduced into our framework for the purpose of better learning convergence and thus, better deblurring performance.

3.3.1 Batch Normalization and Residual Connection

According to [22], a deeper network is getting more complicated and harder to train as more parameters are involved. At the same time, gradient vanishing problem would commonly happen. By normalizing the mini-batch of each layer,

the aforementioned problem is alleviated to some extent:

$$\widehat{x}^{(k)} = \frac{x^{(k)} - E[x^{(k)}]}{\sqrt{Var[x^{(k)}]}} \tag{4}$$

where, $x^{(k)}$ is the input of the kth layer before the normalization, $E[\cdot]$ and $Var[\cdot]$ denote the mean and variance of its input respectively, and $\widehat{x}^{(k)}$ is the normalized output after batch normalization [22].

Batch normalization is applied through the mini-batch statistics, which means that the mean and variance are computed over each batch. Batch normalization prevents small changes of the parameters from amplifying into larger and sub-optimal changes in activations in gradients so that higher learning rates can be applied during training network with batch normalization. Also, by regularizing the model, batch normalization enhances the generalization ability of network.

Besides batch normalization, to train a deep network well, it is important to keep the information from the original data that preserves most of the details. To better utilize the well-trained features from the previous layers and preserve the original data details, a residual structure is proposed. In a building block, it is defined as:

$$y = F(x, \{W_i\}) + x \tag{5}$$

where, x is the input of the block. The set of W_i contains the parameters of the block. $F(\cdot)$ denotes the mapping of x by the set of W_i, and y is the output of this residual learning block [11].

Considering that a shallow network can be well trained easily, a deeper network based on the well trained shallow network and the additional layers to learn the residual between the output of the shallow network and the label should be better than the shallow one. Such residual structure ensures the deeper network always better than its shallow ones.

3.3.2 Sequential Highway Connection

Motivated by the Residual Network (ResNet) in [11], a the Sequential Highway Connection (SHC) is proposed, as shown in Fig. 5, in which the output of first layer is connected to all subsequent layers to build a highway between it and all subsequent layers.

$$y_n = F(y_{n-1} + y_1) \tag{6}$$

where, y_n is the output of the n_th layer, and $F(\cdot)$ denotes to the convolutional operation of the current layer.

This sequential highway connection is leveraged not only to promote the learning convergence as shown in [11], but also to prevent the information loss by connecting the information from the beginning part of the network to the subsequent layers. Considering the deblurring is performed as a filtering process that is the pixel level regression in the network, spatial structure of the original image is very important. However, to find the inverse blurring filter that can be regarded as the inverse kernel, a large respective field of the network is required. From the above two point, combining the idea of ResNet [11], the Sequential Highway Connection can find a better tradeoff between the information specification and the respective field enlargement. Further experiments will show its merits.

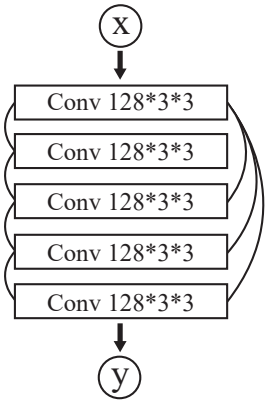

Fig. 5 One block of the proposed Sequential Highway Connection.

The proposed high way connection requires that the two ends of the connection have the same dimensionality. In ResNet [11], residual blocks are defined. For blocks that have the different dimensionalities of input and output, the linear mapping can be used for dimensionality. However, in this work, we find that adding information closer to the raw data to later layers can lead to better performance. Meanwhile, in the pixel level regression task, we empirically found that there is no much difference in using layers with different number of filters. Thus, the proposed network structure has repetitive layers. And all the layers with the same dimensionality of output are structured in the sequential connection, though the proposed structure can be used in several blocks in one network.

4 Experiment

4.1 Dataset and Experiment Setup

We train and test our algorithms with a self-generated dataset. Regarding to [15], over 1000 pages of scientific published articles are downloaded from CiteSeerX[1] repository randomly. After cropping and filtering, more than 0.2 million of 64 by 64 image patches whose pixel variance is larger than a pre-set threshold are generated. Two different kinds of blurring kernels are used in the experiments, which are the kernels generated by PSF [2, 3] and the estimated real kernels [1]. PSFs generated by [2, 3] are used both in training and testing set. Among the PSFs, the size of the templates are set as 21 by 21 and the maximum length of the trajectory is 17.

The network used in the experiments follows the structure of Fig. 2 but with the different number of layers. The first layer of our network uses filters of a very large of size. Padding is set for each convolution layers except for the first layer so that the output has a cropped size of the input. The SHC is applied by dividing the layers into several blocks. We create our network with one SHC block, in which all the convolution layers with the same size of filters are used in the SHC structure.

[1] http://citeseerx.ist.psu.edu/

A 20 layers network is used for comparison in which the size of its first layer is 25 by 25 and those of the rest layers are 3 by 3. The number of the filters of all layers are set as 128.

Several methods are compared with the proposed network. All the compared methods are deployed using their default setting. Our algorithm is tested on one NVIDIA Titan Black GPU using the deep learning programming library Caffe [23].

4.2 Comparison Experiments

4.2.1 Convergency Comparison with ResNet

Before testing the deblurring performance, an experiment is investigated to compare the convergence between the ResNet [11] and our network structure. These two networks are trained on the same dataset built by the same PSFs generation method, and their training losses are compared in Fig. 6.

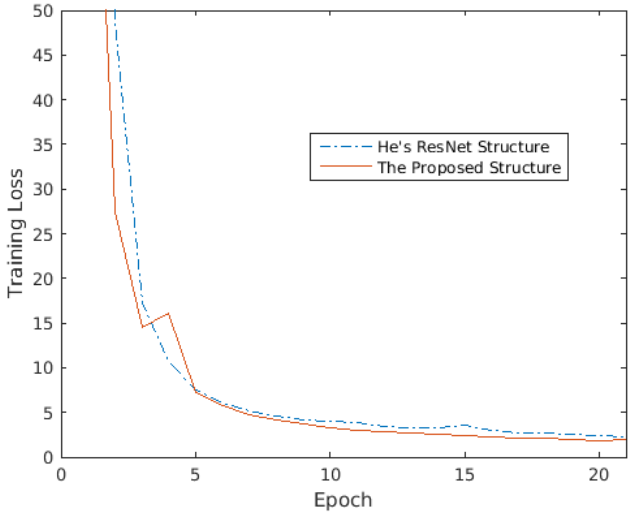

Fig. 6 Training convergency comparison with original ResNet.

As shown in Fig. 6, our experiment result demonstrates that the proposed connection outperforms highway connection in [11] with faster and more stable convergence. The deblur result of our network has higher image quality in terms of PSNR value as well as the visual performance. The deeper of the proposed network benefit from the output of the beginning layer that is not severely corroded yet so that it can be taken as external information during the learning procedure. However, the blur information from the original data will be also included in the meantime. Therefore, the tradeoff needs to be set carefully by deciding which layer is connected to the subsequent layers. In this work, the output of the first layer is used as the complementing source for the subsequent layers.

4.2.2 Performance Comparison

In this section, deblurring performance of different kinds of methods are compared. 5 methods are selected for the comparison, among which the methods from [14, 15] are deep-learning-based, the neural network from [15] are re-trained on our dataset with batch normalization.

Table 1 Performance Comparison Evaluated by PSNR. Two Phase [13], CVPR 2011 [6], CVPR 2014 [10], DCNN [14], BMVC 2015 [15]

Method	Generated Kernels PSNR (dB)	Real Kernels PSNR (dB)
Two Phase	13.184	11.157
CVPR 2011	12.618	10.140
CVPR 2014	11.819	9.757
DCNN	10.081	10.054
BMVC 2015	12.437	11.262
SHC	**15.205**	**11.568**
SHC (Multi-Stages)	**15.197**	**11.583**

As shown in Table 1, evaluated via the average PSNR on the testing dataset, the proposed method outperforms the existed methods. Comparing with non-deep methods that are based on the L0 regularization and total variation model, the results demonstrate that the neural network can learn the distribution without manual regularization. However, from [14], the inverse kernel for deblurring sometimes could be very large, which causes that only a network with the larger respective field can generate such kernels. That is the reason why the first layer of the proposed network has the largest size of convolutional filters.

4.2.3 Multi-Stage Deblurring

Furthermore, by observing the visualized results, the proposed network, even without any artifacts suppression method, create much less artifacts than the others. Such property makes it possible to use multi-stage framework.

As the proposed network introduces little artifact, cascaded networks trained for different kinds of blur can be used on a single image at the same time. Here, a two-stage method is employed on the testing dataset. Firstly, a deep motion deblurring network is applied to the blurry image. Then, a network trained on disk blurring data cascade after that for further removing the disk blur. In Table 1, there is not much difference of the testing PSNRs for the generated kernels dataset because the testing dataset does not contain disk blur. On the other hand, the dataset contains real-world kernels show a slight improvement of the multi-stage delurring, which indicates that the real kernels may always contain disk blur so that the multi-stage can lead to better results.

5 Conclusions

This paper proposes a novel deep learning model for deblurring text images. We train our models efficiently through the proposed Sequential Highway Connection

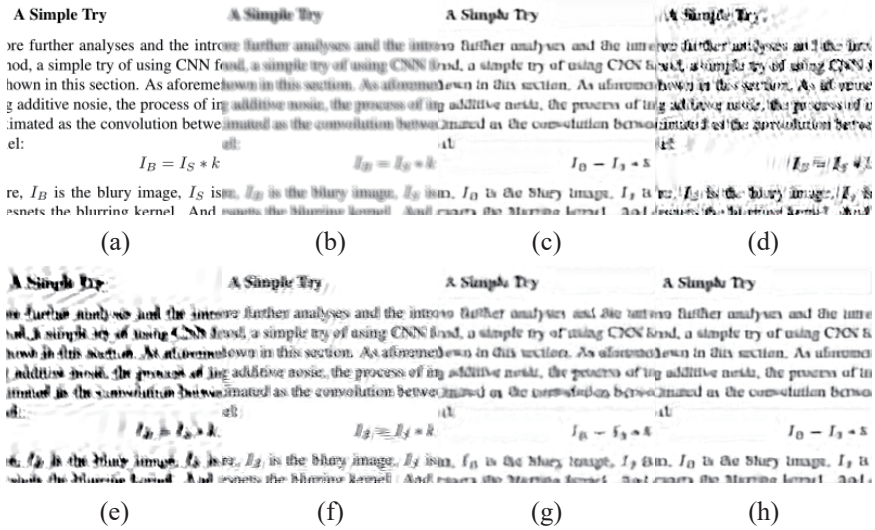

Fig. 7 Some visualized results. (a) Sharp Image, (b) Blurry Image, (c) Our Method, (d) Two Phase [13], (e) CVPR 2011 [6], (f) CVPR 2014 [10], (g) DCNN [14], (h) BMVC 2015 [15].

(SHC). It turns out that the result generated by our method contains less artificial structures. Moreover, we figure out that deeper networks provide better performance for the deblurring task, which indicates the effectiveness of unconstrained regression towards to the ground truth. In comparison with the recent deblurring approaches, our method achieves comparably the state-of-the-art performance.

References

1. Anat, L., Yair, W., Fredo, D., T, F.W.: Understanding and evaluating blind deconvolution algorithms. In: International Conference on Computer Vision and Pattern Recogintion (CVPR) (2009) 2, 3, 8
2. Boracchi, G., Foi, A.: Uniform motion blur in poissonian noise: blur/noise trade-off. In: IEEE Transactions on Image Processing (TIP) (2011) 2, 5, 8
3. Boracchi, G., Foi, A.: Modeling the performance of image restoration from motion blur. In: IEEE Transactions on Image Processing (TIP) (2012) 2, 5, 8
4. Chakrabarti, A.: A neural approach to blind motion deblurring. In: Europeon Conference on Computer Vision (ECCV), pp. 221–235 (2016) 3
5. Dilip, K., Rob, F.: Fast image deconvolution using hyper-laplacian priors. In: Conference and Workshop on Neural Information Processing Systems (NIPS) (2009) 2
6. Dilip, K., Terence, T., Rob, F.: Blind deconvolution using a normalized sparsity measure. In: International Conference on Computer Vision and Pattern Recogintion (CVPR) (2011) 2, 10, 11
7. Gong, D., Yang, J., Liu, L., Zhang, Y., Reid, I.D., Shen, C., van den Hengel, A., Shi, Q.: From motion blur to motion flow: A deep learning solution for removing heterogeneous motion blur. In: International Conference on Computer Vision and Pattern Recogintion (CVPR), pp. 3806–3815 (2017) 3
8. Isola, P., Zhu, J., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: International Conference on Computer Vision and Pattern Recogintion (CVPR), pp. 5967–5976 (2017) 3
9. Jian, S., Wenfei, C., Zongben, X., Jean, P.: Learning a convolutional neural network for non-uniform motion blur removal. In: International Conference on Computer Vision and Pattern Recogintion (CVPR) (2015) 2, 3

10. Jinshan, P., Zhe, H., Zhixun, S., Ming-Hsuan, Y.: Deblurring text images via l0-regularized intensity and gradient prior. In: International Conference on Computer Vision and Pattern Recogintion (CVPR) (2014) 3, 10, 11
11. Kaiming, H., Xiangyu, Z., Shaoqing, R., Jian, S.: Deep residual learning for image recognition. In: International Conference on Computer Vision and Pattern Recogintion (CVPR) (2016) 3, 7, 8, 9
12. Kupyn, O., Budzan, V., Mykhailych, M., Mishkin, D., Matas, J.: Deblurgan: Blind motion deblurring using conditional adversarial networks. arXiv preprint arXiv:1711.07064 (2017) 3
13. Li, X., Jiaya, J.: Two-phase kernel estimation for robust motion deblurring. In: Europeon Conference on Computer Vision (ECCV) (2010) 3, 10, 11
14. Li, X., SJ, R.J., Ce, L., Jiaya, J.: Deep convolutional neural network for image deconvolution. In: Conference and Workshop on Neural Information Processing Systems (NIPS) (2014) 4, 10, 11
15. Michal, H., Jan, K., Pavel, Z., Filip, Š.: Convolutional neural networks for direct text deblurring. In: British Machine Vision Conference (BMVC) (2015) 4, 8, 10, 11
16. Nah, S., Kim, T.H., Lee, K.M.: Deep multi-scale convolutional neural network for dynamic scene deblurring. In: International Conference on Computer Vision and Pattern Recogintion (CVPR), pp. 257–265 (2017) 3
17. Noroozi, M., Chandramouli, P., Favaro, P.: Motion deblurring in the wild. In: Pattern Recognition - 39th German Conference (GCPR), pp. 65–77 (2017) 3
18. Qi, S., Jiaya, J., Aseem, A.: High-quality motion deblurring from a single image. In: ACM Transactions on Graphics (TOG) (2008) 2
19. Ramakrishnan, S., Pachori, S., Gangopadhyay, A., Raman, S.: Deep generative filter for motion deblurring. In: International Conference on Computer Vision (ICCV), pp. 2993–3000 (2017) 3
20. Rob, F., Barun, S., Aaron, H., T, R.S., T, F.W.: Removing camera shake from a single photograph. In: ACM Transactions on Graphics (TOG) (2006) 2
21. Ruomei, Y., Ling, S.: Image blur classification and parameter identification using two-stage deep belief networks. In: British Machine Vision Conference (BMVC) (2013) 2
22. Sergey, I., Christian, S.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167 (2015) 3, 6, 7
23. Yangqing, J., Evan, S., Jeff, D., Sergey, K., Jonathan, L., Ross, G., Sergio, G., Trevor, D.: Caffe: Convolutional architecture for fast feature embedding. arXiv preprint arXiv preprint arXiv:1408.5093 (2014) 9