# Multi-scale Two-Way Deblurring Network for Non-uniform Single Image Deblurring

Zhongzhe Cheng[1] , Bing Luo[1], Li Xu[2(✉)], Siwei Li[3], Kunshu Xiao[3],
and Zheng Pei[1,2]

[1] The Center for Radio Administration Technology Development, School
of Computer and Software Engineering, Xihua University, Chengdu 610039, China
[2] The School of Science, Xihua University, Chengdu 610039, China
[3] Xihua College of Xihua University, Chengdu 610039, China

**Abstract.** We propose a new and effective image deblurring network
based on deep learning. The motivation of this work is based on tradi-
tional algorithms and deep learning which take an easy-to-difficult app-
roach to image deblurring. In traditional algorithms, a rough blur kernel
is obtained first, and then a precise blur kernel is gradually refined. In
deep learning, the pyramid structure is adopted to restore clear images
from easy to difficult. We hope to recover the clear image by two-way
approximation. One network recovers the roughly clear image from the
blurred image, and the other network recovers part of the structural
information from the blank image, and finally the two networks are
added together to obtain the clear image. Experiments show that since we
decomposed the original deblurring task into two different tasks, the net-
work performance has been effectively improved. Compared with other
latest networks, our network can get clearer images.

**Keywords:** Multi-scale network · Two-way learning · Non-uniform
deblurred

## 1 Introduction

Image deblurring is a traditional computer vision problem. Image blur is mainly
formed by camera shake and object motion. It exists in various scenes in the
world, such as natural images [8], human face images [9], text images [7], etc.
The purpose of traditional deblurring algorithms is to obtain blur kernels and
clear images from blurred images. In low-level vision, this is an ill-posed problem,
because the same blurred image can be corresponded various pairs clear latent
image and blur kernel [1–4]. At the same time, traditional algorithms have gen-
eral effects on non-uniform blur. In real blurred scenes, image blur is often not
affected by a single factor, which makes it difficult for traditional algorithms to
model non-uniform blur, which affects the final deblurring effect.

With the development of deep learning, neural networks are gradually used in image deblurring. Because of the learning ability of neural network, it has good performance on the image with non-uniform blur, which can adaptively deblur each pixel. When the neural network was first used for deblurring, the researchers hoped to estimate the motion blur through the network and obtain the image blur carried by each pixel, which ultimately obtained the blur information of each position of the image [19,21,24]. In recent years, researchers have found that direct estimation of clear images is better than estimation of motion blur. These neural networks are roughly divided into two categories: multi-scale neural networks and generative adversarial networks. The multi-scale neural network deblurs the blurred images of multiple scales to achieve the effect of removing the blur from easy to difficult, which is similar to the pyramid structure in traditional algorithms. Generative adversarial network uses the generation of confrontation mechanism to get closer to the real clear image. Although the pixel difference between the deblurred image and the original image is larger, it is more in line with the human eye's perception of a real clear image.

Multi-scale neural network was first proposed by Nah et al. [10]. They construct a multi-scale network by analogy with the pyramid model in the traditional algorithm, and formed a complete de-blurring network by splicing the de-blurring results between different scales. However, the model has large parameters which results in slower network convergence and longer training time. Then, Tao et al. [11] proposed a multi-scale recurrent neural network based on [10], which greatly reduces model parameters by sharing parameters at different scales, which reduces training time. Zhang et al. [17] construct different parameter sharing and parameter independence method according to the role of each convolutional layer of the network in the network, which further improved the deblurring effect.

Generative adversarial networks are most commonly used for image generation, and then gradually applied to various computer vision tasks. Kupyn et al. [14] use generative adversarial network for image deblurring. Since the deblurred image generated according to the Mean-Square-Error(MSE) loss function does not necessarily in line with the human eye's definition of a clear image, the sharp edges of the image are still partially blurred. The generated confrontation network can use the discriminant model to make the generated image as close to the real image as possible. Although some image information will be lost, it is better than other neural networks in terms of image structure and realism. Then, Kupyn et al. [15] add the pyramid structure and the local-global adversarial loss function to improve the network performance on the basis of the original network.

We hope to propose a new network based on the original multi-scale network and integrate the optimization ideas of generating adversarial network into the network to improve performance. Our main contributions are as follows: (1) We propose a new deblurring network based on a multi-scale framework, which incorporates new optimization ideas and can obtain clearer results. (2) The network we proposed has two branches, which respectively obtain the final

clear image from the two ways of the blurred image and the blank image. (3) We prove that the new network can obtain roughly clear image residual images, and the final deblurred image obtained is better than other networks through experiments.

## 2   Related Work

**Multi-scale Network:** The multi-scale network is similar to the pyramid framework in traditional algorithms, and it is based on the observation: after the blurred image is upsampled, the smaller the image size, the smaller the degree of blur. In other words, the multi-scale network first obtains a rough result by deblurring the low-scale image, and then refines the image through the large-scale network, and finally obtains a clear image. Nah et al. [10] construct a multi-scale deblurring network based on the above principles, but there are problems such as large model and parameters which result in some difficulty in training. Based on the work of [10], Tao et al. [11] use Recurrent Neural Network (RNN), which reduces the model size and the number of parameters. [11] adds connections between feature layers of different scales to obtain a better deblurring result. Gao et al. [12] find that the degree of blur at different scales is different, so using the same network for feature extraction will affect the network's extraction of clear image features. Gao et al. [12] adopt the parameter independence of the feature extraction layer to avoid it. At the same time, Gao et al. [12] also find that the network deblurring process after feature extraction is similar. Therefore, parameter sharing is adopted for part of the convolutional layer, which greatly reduces the number of parameters without reducing performance. However, the multi-scale networks described above are all based on a pyramid structure, and there are only differences in parameter sharing and independence between different networks. In terms of network structure, the three methods have no obvious differences. Cai et al. [25] adds the extreme channel prior to the multi-scale network framework at each level of the network, which improves the network performance by constraining the sparseness of the polar channel of the feature image. In general, the multi-scale network can restore better image content information, but the effect of image edge restoration is general, especially for sharp edges, which still contains some blur information.

**Generative Adversarial Network:** The main purpose of the generation adversarial network in the field of image deblurring is to restore sharp edges, so that the resulting clear image is more in line with human perception. Kupyn et al. [14] propose a new generative confrontation network. The generator network is composed of multiple residual blocks with the same structure, and the discriminator uses Wasserstein distance. The network loss function is constructed by three loss functions which contains MSE, confrontation loss and feature loss. Kupyn et al. [15] make improvements on the basis of the original network of [14]. The pyramid structure is integrated into the generation network, and the global-local discriminator is added to the discriminant network to further improve the network performance. Zhang et al. [17] propose a new optimization idea. Most of
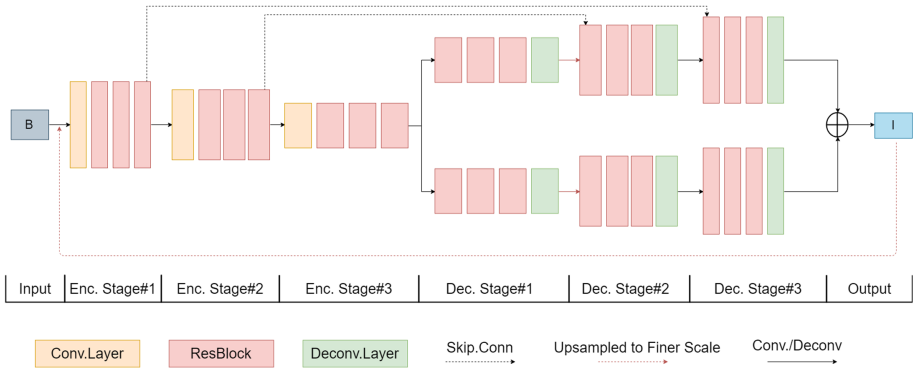
**Fig. 1.** Our proposed Multi-scale two-way learning network.

the discriminators that have been proposed hope that the network will discriminate clear images as 1 and blurred images as 0. Then, by continuously optimizing the generator, the deblurred image will gradually approach from 0 to 1. Zhang et al. [17] hope to move the deblurred image and the clear image closer to 0.5. This optimization result is similar to other methods. In the end, the deblurred image and the clear image will be close, and then the difficulty of optimizing 0 and 1 to 0.5 together is less than other methods, so the effect will be better. However, the above three methods are generative adversarial networks, so they all have a common problem: the discriminator will reduce the consistency of the deblurring result and the original image while optimizing the production plant generator, which leads to the difference in pixel values from the original image (Fig. 1).

## 3   Proposed Method

Most of the network's ideas for image deblurring are from easy to difficult, gradually removing image blur information. Whether it is the pyramid structure in the traditional algorithm, or the multi-scale network and residual learning in the neural network, the improvement of network performance often depends on artificially reducing the difficulty of network learning. We hope to build a network that one part restores the image based on the blurred image and the other part restores the image from blank image, which can complement each other in the process of restoring the image and finally get a clear image.

### 3.1   Multi-scale Two-Way Deblurring Network

As shown in Fig. 3, our proposed network is composed of multiple encoders and decoders. At the same time, the small-scale image is restored to obtain the deblurring results and then passed to the large-scale network after up-sampling. Each encoder is composed of a convolutional layer and three residual blocks, and

(a) Blurry Image        (b) Nah et al.        (c) Tao et al.        (d) Ours

**Fig. 2.** Visual comparison on the dataset of GoPro dataset

the decoder is composed of a deconvolutional layer and three residual blocks. Different from [10,11] network structure, our network has two branches in the decoder part to recover the image content and the remaining information of the image respectively. In order to guide the two branches of the network to recover the corresponding image information, we use the connection between the feature maps to achieve the goal. The up part of the decoder inherits the feature map of the encoder, so the main image content is obtained first. The lower part of the decoder has no feature map skip connection, so the remaining image information is restored from blank image. Finally, we add results from two parts network can get the clear image.

The network we proposed can be expressed by the following formula:

$$I^i, I^i_{up}, I^i_{down} = Net(B^i, I^{i+1}; \theta) \tag{1}$$

where $I^i, I^i_{up}, I^i_{down}$ represents the output of the deblurred image, the up part and the low part of the decoder of the i-th network, and $I^i = I^i_{up} + I^i_{down}$; $B^i$ represents the input of the blurred image; $\theta$ represents network parameters.

### 3.2    Loss Function

We use deblurred images and blurred images on various scales to calculate the mean square error as the loss function of the network. The general multi-scale
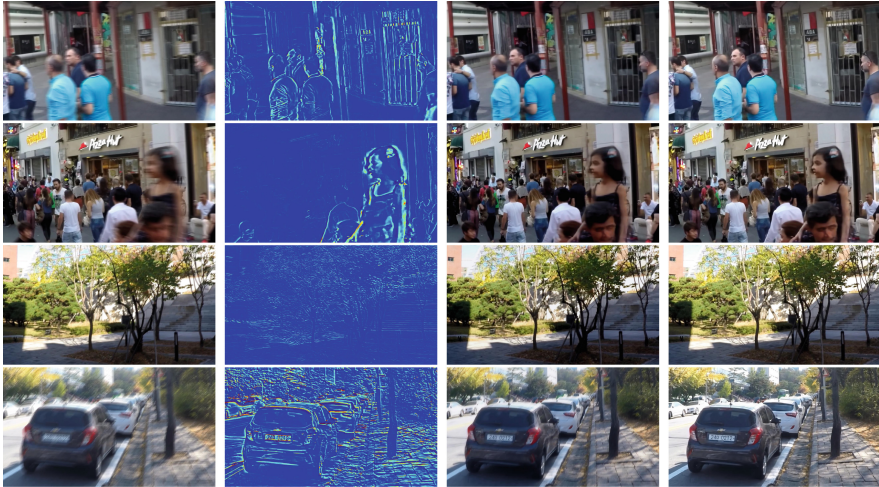
(a) Blurry Image    (b) Down-net    (c) Up-net    (d) Finnal result

**Fig. 3.** Visual comparison of our proposed net

**Table 1.** Table reports the mean PSNR and SSIM obtained over the GoPro dataset [10]

| Method | Nah et al. [10] | Tao et al. [11] | Kupyn et al. [14] | Kupyn et al. [15] | Ours w/o TWNet | Ours |
|--------|-----------------|-----------------|-------------------|-------------------|----------------|------|
| PSNR | 29.08 | 30.10 | 28.70 | 28.17 | 30.18 | **30.56** |
| SSIM | 0.914 | 0.924 | 0.958 | 0.925 | 0.926 | **0.933** |

loss function has the same weight on each scale, but our goal is to get the output on the largest scale, so we increase its weight, hoping that the network will prioritize the final output result. The loss function expression is as follows:

$$L = \sum_{i=1}^{S} \frac{\alpha_i}{T_i} \left\| F_i(B_i; \theta) - I_i \right\|_2^2 \tag{2}$$

where $I_i, B_i$ represents the output of the deblurred image and blurred image of the i-th of scale; $T_i$ represents the number of pixels; $\theta$ represents network parameters; $\alpha_i$ represents the weight of different scales.

## 4    Experiment

We implement our framework on the TensorFlow platform [18]. To be fair, all experiments are performed on the same dataset with the same configuration. For model training, we use Adam solver [26] with $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 10^{-8}$. The learning rate is initially set to 0.0001, exponentially decayed to 0 using power 0.3. We set the convolution kernel size $3 \times 3$, $\alpha_1 = \alpha_2 = 1$, $\alpha_3 = 5$. We randomly

**Table 2.** Table reports the mean PSNR and SSIM obtained over the Köhler dataset [5]

| Method | Xu et al. [6] | Pan et al. [8] | Nah et al. [10] | Tao et al. [11] | Kupyn et al. [14] | Kupyn et al. [15] | Ours |
|--------|---------------|----------------|-----------------|-----------------|-------------------|-------------------|------|
| PSNR | 29.96 | 30.20 | 26.48 | 26.75 | 26.10 | 26.36 | **26.80** |
| SSIM | 0.876 | 0.888 | 0.807 | 0.837 | 0.816 | 0.820 | **0.840** |

crop $256 \times 256$ images from original paired clear and blurred images as training images. We use Xavier method [22] initialization parameters. Our experiments can converge after 4000 epochs.

**Dataset Preparation:** In order to create a large training dataset, methods based on early learning [19–21] synthesize blurred images by convolving clear images with real or generated uniform/uneven blur kernels. Due to the simplified image formation model, the synthesized data is still different from the data captured by the camera. Recently, researchers [10] propose a method to generate blurred images by averaging consecutive short exposure frames in videos taken by high-speed cameras (such as high-speed cameras). For example, GoPro Hero 4 Black, which can approximate long exposure blurry picture. These generated frames are more realistic because they can simulate complex camera shake and object motion, which are common in real photos.

In order to fairly compare the performance differences between different network frameworks, we train our network on the GoPro dataset, which contains 3214 image pairs. Like [10–13], we choose 2103 pairs as the training set and 1111 pairs as the test set.

**Benchmark Dataset:** We first conduct experiments on the test set of the GoPro dataset, which contains many complex blurs caused by camera shake and object motion. Table 1 shows our performance compared with other state-of-the-art methods. We choose Peak-Signal-to-Noise Ratios(PSNR) and Structural Similarity (SSIM) [23] as the evaluation criteria. It can be seen that the generative adversarial network [14,15] have significant advantages in restoring image structure (SSIM), but the result after deblurring is quite different from the original image pixel value (PSNR). At the same time, the multi-scale network is lower than our proposed method in both PSNR and SSIM. In addition, we give the experimental results of the network structure without two-way deblurring. Figure 2 shows the subjective effect of our and other methods. Figure 3 shows the output results of each branch of the bidirectional network. It can be seen that the up part of the network removes part of the image blur and the lower part of the network mainly focuses on the edge of the image. Information was supplemented.

Then, we also conducted experiments on the traditional dataset [5], this dataset consists of 4 images and 12 blur kernels, of which three blur kernels are larger in size and form larger blurs. It can be seen from the Table 2 that since we doesn't train network for large blurred images during the training process, the effect of large blurred images is general. Traditional algorithms directly model

blurred images, so the effect of processing large blurred images is better. However, our network performance is still better than other neural networks, which shows the effectiveness of our proposed network.

## 5    Conclusion

Based on the multi-scale neural network, combined with the optimization idea of two-way approximation, we constructed a new multi-scale two-way deblurring network. This network, like other neural networks, has a significant effect on non-uniform deblurring. Compared with other state-of-the-art multi-scale networks, our network can better restore image edges and get better deblurring images.

## References

1. Shan, Q., Jia, J., Agarwala, A.: High-quality motion deblurring from a single image. ACM Trans. Graph. **27**, 1–10 (2008)
2. Cho, S., Lee, S.: Fast Motion Deblurring. ACM Trans. Graph. **28**, 1–8 (2009)
3. Fergus, R., Singh, B., Hertzmann, A., Roweis, S.T., Freeman, W.T.: Removing camera shake from a single photograph. ACM Trans. Graph. **25**, 787–794 (2006)
4. Xu, L., Jia, J.: Two-phase kernel estimation for robust motion deblurring. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010. LNCS, vol. 6311, pp. 157–170. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15549-9_12
5. Köhler, R., Hirsch, M., Mohler, B., Schölkopf, B., Harmeling, S.: Recording and playback of camera shake: benchmarking blind deconvolution with a real-world database. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012. LNCS, vol. 7578, pp. 27–40. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33786-4_3
6. Xu, L., Zheng, S., Jia, J.: Unnatural l0 sparse representation for natural image deblurring. In: CVPR (2013)
7. J. Pan, Z. Hu, Z. Su, Yang, M.H.: L0-regularized intensity and gradient prior for deblurring text images and beyond. IEEE Trans. Pattern Anal. Mach. Intell. **39**(2), 342C355 (2017)
8. Pan, J., Sun, D., Pfister, H., Yang, M.H.: Blind image deblurring using dark channel prior. In: CVPR (2016)
9. Pan, J., Hu, Z., Su, Z., Yang, M.-H.: Deblurring face images with exemplars. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8695, pp. 47–62. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10584-0_4
10. Nah, S., Kim, T.H., Lee, K.M.: Deep multi-scale convolutional neural network for dynamic scene deblurring. In: CVPR (2017)

11. Tao, X., Gao, H., Shen, X., Wang, J., Jia, J.: Scale-recurrent network for deep image deblurring. In: CVPR (2018)
12. Gao, H., Tao, X., Shen, X., Jia, J.: Dynamic scene deblurring with parameter selective sharing and nested skip connections. In: CVPR (2019)
13. Zhang, H., Dai, Y., Li, H., Koniusz, P.: Deep stacked hierarchical multi-patch network for image deblurring. In: CVPR (2019)
14. Kupyn, O., Budzan, V., Mykhailych, M., Mishkin, D., Matas, J.: DeblurGAN: blind motion deblurring using conditional adversarial networks. In: CVPR (2018)
15. Kupyn, O., Martyniuk, T., Wu, J., Wang, Z.: DeblurGAN-v2: deblurring (Ordersof-Magnitude) faster and better. In: ICCV (2019)
16. 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 (2019)
17. Zhang, K., Luo, W., Zhong, Y., et al.: Deblurring by realistic blurring. arXiv (2020)
18. Abadi, M., et al.: TensorFlow: large-scale machine learning on heterogeneous systems (2015). Software available from tensorflow.org
19. Chakrabarti, A.: A neural approach to blind motion deblurring. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9907, pp. 221–235. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46487-9_14
20. Schuler, C.J., Hirsch, M., Harmeling, S., Scholkopf, B.: Learning to deblur. TPAMI **38**(7), 1439–1451 (2016)
21. Sun, J., Cao, W., Xu, Z., Ponce, J.: Learning a convolutional neural network for non-uniform motion blur removal. In: CVPR (2015)
22. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: AISTATS, pp. 249–256 (2010)
23. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600612 (2004)
24. Bahat, Y., Efrat, N., Irani, M.: Non-uniform Blind Deblurring by Reblurring. In: ICCV (2017)
25. Cai, J., Zuo, W., Zhang, L.: Dark and bright channel prior embedded network for dynamic scene deblurring. IEEE Trans. Image Process. **29**, 6885–6897 (2020)
26. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. In: ICLR (2014)