




Masked Scale-Recurrent Network for Incomplete Blurred Image Restoration

Jingzhou Zhu^(✉) , Wentao Chao, and Dong Yang

School of Artificial Intelligence, Beijing Normal University, Beijing, China
{zhujz, chaowentao, yd}@mail.bnu.edu.cn

Abstract. Multi-scale learning has been demonstrated to be an excellent deblurring approach in image restoration according to recent studies. It makes the optimization of the function easier to achieve the global optimum. In order to restore an image that is both incomplete and blurry, we propose a Masked Scale-Recurrent Network (MSRN) in this paper, a restoration method based on multi-scale learning and an asymmetric autoencoder. It implements restoration in an end-to-end manner without any prior knowledge or other given conditions. Firstly, we process the GoPro dataset and obtain a dataset of incomplete images. And then, we perform a self-supervised reconstruction pre-training on the autoencoder, with a series of resblocks that increase the quality of the input image and improve the representation learning in the latent space. Finally, on the processed data, we train the model and finish the adjustment of the entire network. Compared with classical multi-scale learning, we introduce masks to help the model train more efficiently by focusing on essential regions of the image. It is also shown that MSRN has successful image restoration capability as well as robustness, as demonstrated in our experiments.

Keywords: image restoration · masked autoencoder · residual network · scale-recurrent network · multi-scale learning

1 Introduction

Image restoration technology is a vital component in image processing, where the input image may contain various interference. In this article, we will propose a new restoration method for incomplete blurred images. Additionally, this method can also be applied to blurry images with severe local damage. Extremely severe image interference is essentially worthless and does not contribute to image restoration or even worse. Therefore, interference that cannot be recovered by existing methods can be deleted directly and only the blurred images that can be restored should be retained. All images with irreversible interference or severe data loss should be converted into incomplete blurred images, unifying the format of the input pictures. When the input formats are unified, the first section of our task-incomplete blurry image restoration-is to reconstruct the images.

Image reconstruction, also known as inpainting, is a classic problem in computer vision. Popular approaches for image inpainting include foreground-aware image inpainting [2] and pluralistic image completion [4]. However, these methods have limitations when the missing parts are too large, or there is not enough labeled data. Recently, Alexey Dosovitskiy et al. [5] applied Transformer [6] from natural language processing (NLP) to computer vision (CV), opening up new possibilities for Transformer in this field. Among them, the mask reconstruction technique represented by MAE [8] (CV) and BEiT [10] (NLP) is a self-supervised training method that can effectively address the need for labeled data. The image mask reconstruction model MAE is an autoencoder with ViT [5] as its backbone. Its training efficiency is significantly higher than a simple ViT model, and its generalization ability is competitive in image reconstruction tasks.

The second section of the task is deblurring. Researchers in various fields have been conducting in-depth research to restore blurred or defocused [1] images to clear images as much as possible. The goal is to recover various details that should exist in a sharp image and facilitate subsequent picture processing. In the past five years, some methods, including conditional methods, have achieved quite remarkable results [3, 9, 12, 15, 16, 18, 20, 21, 25]. However, blurry images are usually accompanied by irreversible distortion, large-scale dense interference, and data loss due to various factors. This means the input image may change from a simple blurred image to a blurred image with several other interference, and in the worst case, it may become an incomplete blurred image. Therefore, common image restoration algorithms may only play a limited role in such circumstances. Among these algorithms, the multi-scale learning network is easy to expand and has a simple framework with a comparatively small number of parameters, making it easier to train. Therefore, we use SRN [3], which is designed based on scale-recurrent structure, as the backbone of our network and extend SRN to make the entire network more suitable for the restoration task of incomplete blurred images.

Based on the analysis above, we propose a new masked scale-recurrent network (MSRN), which is based on MAE and SRN. This network utilizes the visual representation learning ability of MAE for image reconstruction and exploits SRN’s multi-scale recurrent network for fuzzy image restoration. On this basis, to adapt MSRN to new learning tasks, we have made the following three contributions:

Deep MAE. We introduce a deep masked autoencoder (DMAE) which puts a series of resblocks [22] in front of the autoencoder to preprocess the input image, resulting in a more accurate latent representation of images. This allows the autoencoder to reconstruct more valuable information on the missing parts of images for deblurring.

A New Scaling Method. We propose a new scaling method that adds a new learning scale and utilizes masks in scaling images. This approach forces the model to learn more accurate feature representation compared to training with

complete photos at different scales and ensures that images learned at different scales do not exhibit pixel distortion.

Shortcut Connection. We build a shortcut connection between DMAE and SRN, which shares the position of invisible patches with the multi-scale network. With this connection, SRN learns more about the location of incomplete parts and achieves better performance in restoring those patches restored by DMAE.

2 Related Work

In this section, we will recap the background of image restoration concisely and briefly introduce related structures including their basic concepts and characteristics.

2.1 Image Deblurring

Image blur can be divided into motion blur, defocus blur, Gaussian blur, and mixed blur. Image deblurring with known fuzzy kernels is called non-blind image deblurring. The image whose fuzzy kernel is unknown is called blind image deblurring. The deep learning neural network based deblurring method, which is also a commonly used approach nowadays, has achieved quite good results on complex images with multiple unknown fuzzy kernels. After the publication of using deep learning networks to predict the direction and width of blur for deblurring [7], Su et al. [11] proposed a deep learning method for video deblurring, which uses an autoencoder with shortcut connections. Furthermore, the deep multi-scale network proposed by Nah et al. [23] obtained fruitful achievement in deblurring tasks.

In multi-scale deblurring networks, the scale-recurrent network (SRN) proposed by Tao et al. [3] is simpler and has fewer parameters. Compared to [23], it is easier to train while achieving better results. SRN is inspired by the very successful “coarse-to-fine” scheme for single-image deblurring. The input is a series of images sampled from the original image at different scales and the corresponding network outputs an intermediate image according to the given resized image. Next, the upsampled processed image and the image from the next scale are given to the network of the next scale. At the same time, there is a channel between different scales transferring hidden parameters to other scales. It is used to share implicit state parameters with other scale networks to improve convergence performance. It is precisely because of this that the total number of parameters can be significantly reduced. SRN can be described as:

$$I^i, h^i = SRN(B^i, UP(I^{i+1}), UP(h^{i+1}); \theta_{SRN}) \quad (1)$$

where i represents the index of different scales. B^i and I^i are blurry images and intermediate images at i scale. h^i is hidden state features at the i -th scale. UP is the operation that upsamples or converts input from $i + 1$ to i scale. θ_{SRN} is other parameters of SRN.

Each scale’s network of SRN uses a symmetric CNN architecture with skip-connections. Similar to U-net [19], the first half of the network gradually converts the input i -th scale image into latent spatial features with smaller resolution and multiple channels. The second half of the network gradually restores the features to the original i -th scale image. Encoder and decoder of the same scale are connected by a channel, which also can accelerate convergence. Based on [22, 23], SRN employs ResBlocks and adds several ResBlocks between the encoding and decoding convolution layers. The loss function applied by SRN is the simple Euclidean loss and has achieved sufficiently good qualitative and quantitative results.

2.2 Image Inpainting

Image inpainting is to restore the damaged part of the image by algorithms and keep the restoration as consistent as possible with the original image. Image inpainting is not only a crucial task in computer vision but also a basic task for other subsequent processing of images. Classical image inpainting includes inpainting methods based on partial differential equations [34] and samples [32], but they usually consume a lot of computing time and have limited recovery effects. Later, with the success of deep learning in image processing, researchers tried to introduce different deep networks to achieve inpainting and proposed a large number of methods for inpainting such as autoencoder, U-Net [19], GAN [33], and Transformer [6]. Numerous experiments have shown that these methods have achieved better performance than traditional methods in inpainting.

2.3 CNN/Transformer for Image Processing

In deep learning networks, as the depth increases, the network becomes more difficult to train [17]. ResNet [22] makes training for deep networks much easier than before. It reconstructs the network and learns residual functions with reference from the output of the former layer during learning. Specifically, it creates a shortcut from input to output. This simple but effective approach prevents the model from gradient disappearing when the network is deep.

The masked autoencoder designed by He et al. [8] is a more general denoising autoencoder [14]. It is self-supervised and has two core innovations on the basis of ViT-Large/-Huge [5]. The first one is the asymmetric encoder-decoder architecture. The encoder only encodes visible patches while the decoder can see all patches. Experiments proved that this lightweight structure not only increases recovery accuracy but also drastically reduces floating point operations. The second core design is to mask the image with a high proportion, in that a small amount of masking impedes the model’s attempt to acquire worthy knowledge, which will be no different from simple interpolation algorithms. Results showed that this design forces the model to learn better representations and the model achieves the best performance at a mask rate of around 75%.

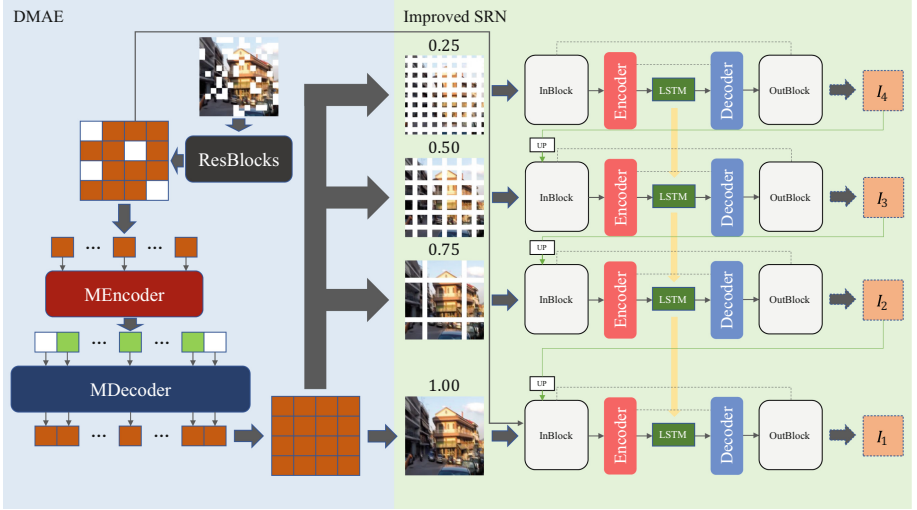


Fig. 1. The architecture of our proposed MSRN

3 Approach

Our goal is to design a universal, end-to-end network to complete the learning of incomplete blurry image restoration. Therefore, relying solely on multi-scale learning of SRN cannot complete this task. We subsequently introduced MAE and, to achieve the best recovery effect, image reconstruction should precede image deblurring. In addition, this task is more complicated than simply deblurring, so it is necessary to enhance the overall learning ability of the network. To check the effectiveness of our method, we tried to train SRN alone and MAE plus SRN to recover an incomplete and blurry image. When we were training these two networks, however, we found that neither using the SRN network alone nor simply connecting MAE and SRN can achieve good recovery results, introduced in our experiment. Our proposed MSRN is an image restoration network that combines mask reconstruction and multi-scale learning. It implements the task of using a single network to complete reconstruction and deblurring. Figure 1 illustrates the overall structure of our proposed network. It takes an incomplete blurry image as the input and completes the image inpainting through DMAE. And the reconstructed image enters the improved SRN network for deblurring. The output is a corresponding complete and unambiguous picture. Thus, our network is able to recover in an end-to-end manner. The whole network can be described as:

$$I_i = \begin{cases} ISRN[\mathcal{C}(\mathcal{M}_i(\mathcal{B}), UP(I_{i+1})), UP(h_{i+1}); \theta], & i \neq 1 \\ ISRN[\mathcal{C}(\mathcal{B}, UP(I_{i+1}), IB^*), UP(h_{i+1}); \theta], & i = 1 \end{cases} \quad (2)$$

where i denotes the scale index. IB^* represents incomplete and blurry images after ResBlocks. I_i is the intermediate representation at the i -th scale, with

$i = 1$ is the output of MSRN. h_i stands for hidden parameters in LSTM [24] of different scale networks. Operation \mathcal{C} is concatenation while \mathcal{M} is scaling by mask. \mathcal{B} is the complete blurry image after DMAE and the process in DMAE can be described as:

$$\begin{aligned}
 IB^* &= Res(IB) \\
 \mathcal{B} &= MAE(IB^*)
 \end{aligned}
 \tag{3}$$

where IB is the input image.

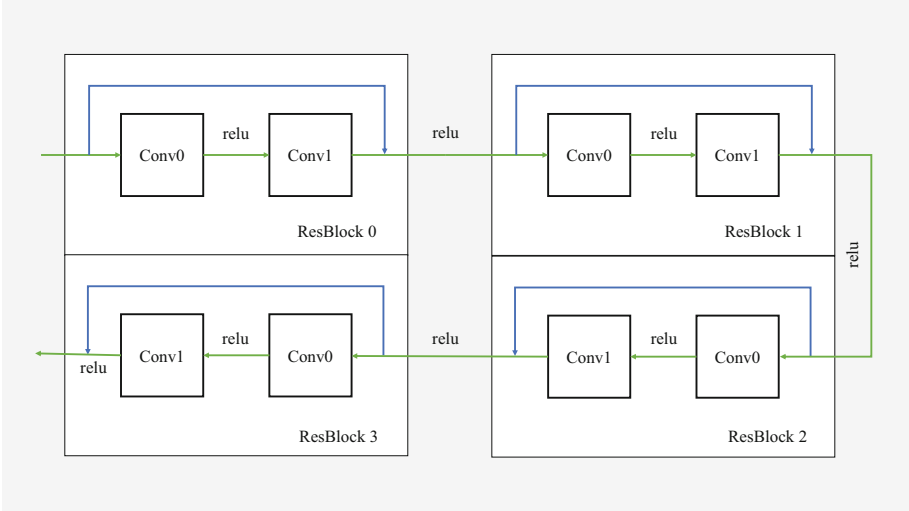


Fig. 2. The ResBlocks in DMAE

3.1 DMAE

Since MAE is a flexible autoencoder that exhibits excellent performance in representation learning on images and is suitable for multiple tasks [8], including image classification, target detection, and image segmentation, it is necessary to make some improvements on MAE to improve its performance on a single task. We use DMAE in the first half of MSRN to restore incomplete blurred images to complete blurred images. During the pre-training, we introduced a series of ResBlocks before the MAE, as shown in Fig. 1. The whole of these ResBlocks is a simple pre-trained network that improves the quality of blurry images, which allows MAE to generate clearer patches on the missing parts. We only keep the uncovered parts of the result. The structure of ResBlocks is shown in Fig. 2. Because the model will become increasingly difficult to train if the network deepens [26,28], we use residual connection [22] in each block, which

skips one or more layers. It maintains the learning gradient of the network at a trainable level without additional parameters and computational complexity.

Based on [8], the image is segmented using the method in [5] that divides the image into patches. Next, we mask 75% of the image by “random sampling” [8] when training MAE. In the encoder, visible patches are mapped to the latent space through linear projection with position information. The input of the decoder consists of both visible and invisible patches. The decoder then restores the information of the incomplete parts.

3.2 Improved SRN

In the scale-recurrent network [3], each scale is composed of a symmetric and skip-connected autoencoder [27] with an LSTM in the middle of it. The output of each scale’s network is the feature map learned by this scale network. After being upsampled, it will be superimposed into the image of the next scale along the depth direction and they are input to the next scale network. At the same time, the parameters in the LSTM [24,30] will also be transferred to the LSTM of the next scale. On this basis, we have made the following improvements.

The first improvement we made in SRN was the application of masks to scale images, as shown in Fig. 1. In image convolution, images are in the form of large matrices, while encoders encode discrete patches. Therefore, we have considered the following two issues when performing masking. The first is that using random sampling will cause the relative positions between various patches to be disrupted, which hinders the scaled image from presenting the correct shape. The network consequently can’t learn valuable image features. The second is that the image will also exhibit significant distortion when the mask scale is too large, which can also cause the aforementioned problems. Therefore, we mask the image by grid-wise sampling [8], with a width of 16 pixels. It reduces the redundancy of the image while ensuring that the image is recognizable, allowing the network to learn more content. And this approach has been verified in subsequent experiments.

In order to improve the overall fitting ability of the network, we put forward another improvement which is the addition of a scale network. In [3] and our experiments, it has been confirmed that increasing the number of scales can improve image restoration performance. However, in our experiments, we found another issue which is the distortion on restored patches is still very distinct. We believe the reason is that there is a notable difference between the blurred portions recovered by DMAE and other blurred portions. In response to this distortion, we add the output result of ResBlocks in DMAE to the last-scale network, transferring the location of the missing patch on the original image to the back-end network. Our experiments have shown that our model with these modifications is more effective than the baseline.

3.3 Loss Function

The loss function in DMAE is the mean squared error (MSE) and we only calculate it on the missing part of the image. In the improved SRN, we use Euclidean loss [3]. The Euclidean loss on each scale network is

$$\mathcal{L}_E = \sum_{i=1}^n \frac{\|I_i - I_i^*\|_2^2}{\mathcal{N}_i} \quad (4)$$

where \mathcal{N}_i is the total number of elements in i -th scaled image. I_i and I_i^* are the output image and sharp image respectively at the i -th scale. In fact, our loss functions are simple since we learned that the mean square error has sufficient capability to train a good network during our work.

4 Experiments

The training and testing of MAE, SRN and MSRN were completed on the GoPro [23] dataset. To reasonably shorten training time and meet MAE input requirements, we adjusted the resolution of all images from 1280×720 to 640×640 using bilinear interpolation, and the GPU in the training was RTX 3060 Ti. The incomplete ratio is 0.25. For a fair comparison, the methods in each experiment were completed under the same training images. In the experiment, we compared MSRN with other methods and tested the performance of the network under different conditions.

4.1 Dataset

The GoPro dataset is generated by the GOPRO4 Hero Black camera [23]. Blur images are obtained by averaging several consecutive latent frames and the corresponding clear images are the intermediate frames of these consecutive frames. This dataset is publicly available, with 3214 pairs of blurred and clear images with a resolution of 1280×720 . There are 1111 pairs of test images, accounting for about one-third of the total dataset. Both Nah et al. and SRN tested on this dataset and successively proposed dynamic deblurring SOTA models.

4.2 Model Training

For the pre-training of ResBlocks, we used complete images for training. Because the purpose of introducing ResBlocks is only to pre-process the input image, we conducted 400 epochs during training. In the pre-training of MAE, we conduct self-supervised training for the autoencoder. The size of masks is 16×16 and other settings are the same as the default configuration in [8]. In improved SRN, the training data are original masked images, images restored by DMAE, and clear images. We did not change much in parameters. The parameters of Adam solver [29] are $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 1 \times 10^{-8}$. The variables in our

model are initialized by Xavier method [28]. The learning rate is exponentially attenuated, with an initial value of 1×10^{-4} . The batch size is 16 while the training epoch is 3000 and it is enough in that we noticed the models had already converged very well at around the 2600th epoch.

4.3 Comparisons

To the best of our knowledge, due to the lack of a specific restoration model for incomplete blurred images in recent work, we mainly compared our methods with other applications of SRN, which is state-of-the-art in dynamic deblurring. In addition, we also compared ours with previous state-of-the-art [23] in dynamic deblurring. The results are shown in Fig. 3. Note that the images in the first row are from the training set, and the subsequent images are from the testing set.



Fig. 3. Visual comparison. (a) Input. (b) Results of Nah *et al.* [24]. (c) Vanilla SRN. (d) Results of M-SRN. (e) Our results. (f) Ground Truth

Initially, we tried to use Vanilla SRN to complete reconstruction and deblurring, but, as shown in Fig. 3, SRN’s learning ability is not enough and the effect of restoration is limited. Later, we used MAE for image reconstruction after SRN, but there were very obvious edges around the restored patch which means that restoration should be in front of deblurring. Based on former experiments, we

decided to adopt MAE first and SRN next. However, simply connecting these two models does not have a good recovery effect on the patches restored by MAE, and even the deblurring effect on the other visible parts becomes poor. Therefore, our proposed MSRN is based on the experience above. It enhances the overall learning ability of the model and makes some improvements for the restoration of incomplete parts. In the test, MSRN achieves good results in both image deblurring and image reconstruction, as shown in Fig. 3.

Table 1. Quantitative comparison between ours and other state-of-the-art methods of dynamic scene deblurring on test data.

Methods	Metrics					
	PSNR \uparrow	SSIM \uparrow	MS-SSIM \uparrow	LPIPS \downarrow	SRER \uparrow	RMSE \downarrow
Input	10.5701	0.6696	0.5532	0.5160	52.7439	0.0185
Nah <i>et al.</i>	22.3217	0.7082	0.8347	0.4020	56.2137	0.0048
Vanilla SRN	23.0606	0.7720	0.8306	0.2534	56.0942	0.0049
Mask SRN	23.4083	0.7883	0.8479	0.2538	56.6596	0.0042
SRN_M	23.2217	0.7386	0.8682	0.3481	56.6372	0.0043
M_SRN	23.7817	0.7924	0.8697	0.2543	57.2761	0.0037
Ours	24.4951	0.8004	0.8637	0.2383	57.2984	0.0037

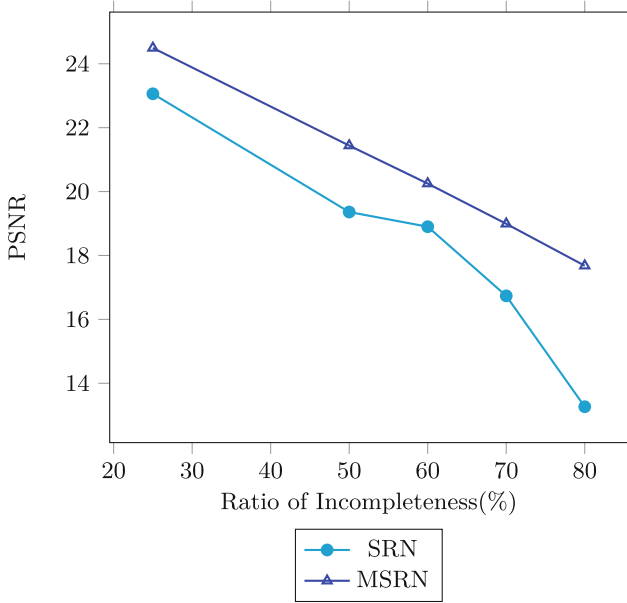
In order to verify the effectiveness of our model, we have designed a series of baseline models, among which SRN and the method of Nah et al. are popular images deblurring models. We use PSNR, SSIM, MS-SSIM [31], LPIPS [13], Signal to Reconstruction Error Ratio (SRER) and Root-MSE (RMSE) as metrics to evaluate the quality of image restoration. The results are shown in Table 1. Mask SRN is the method that scales images by grid mask and the results prove that the learning ability of SRN with the mask is stronger. SRN_M means adding MAE to the back of the original SRN. It deblurs the mask image first using SRN and then reconstructs the image using MAE. M_SRN is the model in which MAE is inserted in front of SRN, and the result indicates that the latter method has a better effect in our experiment. Finally, with the exception of MS-SSIM, our model achieves the best results. Note that in Fig. 3, we can more clearly demonstrate the excellent recovery effect of our model for restoring those invisible patches while deblurring.

4.4 Performance of Different Strategy

To verify that increasing the number of scales in our network can improve the learning ability of the network, we trained models under different scales. Similarly, we use the same data for training. The results are shown in Table 2 and they show that increasing the number of scales can improve the performance of the model. It is worth noting that when the number of scales increases from 2 to 3, the performance of the model still has a significant improvement. However,

Table 2. Results of different scales.

Number of scales	PSNR \uparrow	SSIM \uparrow	MS-SSIM \uparrow	LPIPS \downarrow	SRER \uparrow	RMSE \downarrow
1 Scale	21.3659	0.6547	0.7935	0.4669	55.8059	0.0053
2 Scales	22.1812	0.7120	0.8093	0.3909	56.0156	0.0048
3 Scales	22.7207	0.7639	0.8251	0.3423	56.2244	0.0045
4 Scales	23.3764	0.7693	0.8357	0.3336	56.6743	0.0042

**Fig. 4.** Performance of MSRN and SRN under different degrees of incompleteness

as the number of scales increases from 3 to 4, the improvement of the model has become very limited.

We further examine our model in different deficiency conditions. As shown in Fig. 4, The horizontal axis is the ratio of incompleteness, and the vertical axis is the PSNR indicator of the test results. In order to explore the recovery effect of the model under a high rate of incompleteness. Besides the original 25% mask rate, we conducted experiments with mask rates of 50%, 60%, 70%, and 80% to evaluate the performance of the model under extreme conditions. When the mask rate gradually increases, our model mainly exhibits a linear decline. Significantly, when the mask rate is greater than 60%, the gap between the two gradually widens. And the downward trend of SRN shows a parabolic decline, while our model still maintains a linear decline. From this point, it shows that our model still has a certain recovery effect for severely damaged blurred images, revealing robustness. Furthermore, we have also attempted to change other parts of the network. For instance, we tried different loss functions for

training and different activation function in resblocks. However, we have found that our current MSRN performs best.

5 Conclusion

In this paper, we proposed a novel and robust model, called MSRN, for restoring incomplete and blurry images without any prior knowledge. MSRN utilizes a new scaling method and adds a scale to the model to improve the restoration effect of the missing parts, along with DMAE and a shortcut connection to address the characteristics of incomplete images. The key feature of MSRN is the “coarse-to-fine” scheme for image restoration and the use of an asymmetric autoencoder structure for image reconstruction. Our extensive experiments demonstrate that MSRN outperforms existing methods, including DeepDeblur [23] and SRN. Furthermore, we confirm that CNN and Transformer can be effectively combined to provide new solutions for solving novel problems. In future work, we plan to introduce object detection techniques to further improve our design and achieve better image restoration on more diverse and practical incomplete and blurry images.

Acknowledgements. This work is supported by the National Key Research and Development Project Grant, grant/award number: 2018AAA01008-02.

References

1. Ban, Y., et al.: Depth estimation method for monocular camera defocus images in microscopic scenes. *Electronics* **11**(13), 2012 (2022)
2. Xiong, W., et al.: Foreground-aware image inpainting. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5840–5848 (2019)
3. Tao, X., Gao, H., Shen, X., Wang, J., Jia, J.: Scale-recurrent network for deep image deblurring. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8174–8182 (2018)
4. Zheng, C., Cham, T.J., Cai, J.: Pluralistic image completion. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1438–1447 (2019)
5. Dosovitskiy, A., et al.: An image is worth 16x16 words: transformers for image recognition at scale. *arXiv preprint [arXiv:2010.11929](https://arxiv.org/abs/2010.11929)* (2020)
6. Vaswani, A., et al.: Attention is all you need. In: *Advances in Neural Information Processing Systems*, vol. 30 (2017)
7. Sun, J., Cao, W., Xu, Z., Ponce, J.: Learning a convolutional neural network for non-uniform motion blur removal. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 769–777 (2015)
8. He, K., Chen, X., Xie, S., Li, Y., Dollár, P., Girshick, R.: Masked autoencoders are scalable vision learners. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16000–16009 (2022)
9. Wu, C., Wo, Y., Han, G., Wu, Z., Liang, J.: Non-uniform image blind deblurring by two-stage fully convolution network. *IET Image Process.* **14**(11), 2588–2596 (2020)

10. Bao, H., Dong, L., Piao, S., Wei, F.: BEiT: BERT pre-training of image transformers. arXiv preprint [arXiv:2106.08254](https://arxiv.org/abs/2106.08254) (2021)
11. Su, S., Delbracio, M., Wang, J., Sapiro, G., Heidrich, W., Wang, O.: Deep video deblurring for hand-held cameras. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1279–1288 (2017)
12. Chen, L., Zhang, J., Pan, J., Lin, S., Fang, F., Ren, J.S.: Learning a non-blind deblurring network for night blurry images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10542–10550 (2021)
13. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 586–595 (2018)
14. Vincent, P., Laroche, H., Bengio, Y., Manzagol, P.A.: Extracting and composing robust features with denoising autoencoders. In: Proceedings of the 25th International Conference on Machine Learning, pp. 1096–1103 (2008)
15. Kupy, O., Budzan, V., Mykhailych, M., Mishkin, D., Matas, J.: Deblurgan: blind motion deblurring using conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8183–8192 (2018)
16. Mei, J., Wu, Z., Chen, X., Qiao, Y., Ding, H., Jiang, X.: Deepdeblur: text image recovery from blur to sharp. *Multimedia Tools Appl.* **78**, 18869–18885 (2019)
17. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556) (2014)
18. Gong, G., Zhang, K.: Local blurred natural image restoration based on self-reference deblurring generative adversarial networks. In: 2019 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), pp. 231–235. IEEE (2019)
19. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
20. Kupy, O., Martyniuk, T., Wu, J., Wang, Z.: DeblurGAN-v2: deblurring (orders-of-magnitude) faster and better. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 8878–8887 (2019)
21. Fan, J., Wu, L., Wen, C.: Sharp processing of blur image based on generative adversarial network. In: 2020 5th International Conference on Advanced Robotics and Mechatronics (ICARM), pp. 437–441. IEEE (2020)
22. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
23. Nah, S., Hyun Kim, T., Mu Lee, K.: Deep multi-scale convolutional neural network for dynamic scene deblurring. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3883–3891 (2017)
24. Hochreiter, S., Schmidhuber, J.: Long short-term memory. *Neural Comput.* **9**(8), 1735–1780 (1997)
25. Zou, Q., Luo, L., Wang, X., Chen, D., Liu, F.: Improved image deblurring algorithm to generate antagonistic neural network. In: 2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP) (2020)
26. Bengio, Y., Simard, P., Frasconi, P.: Learning long-term dependencies with gradient descent is difficult. *IEEE Trans. Neural Networks* **5**(2), 157–166 (1994)

27. Mao, X., Shen, C., Yang, Y.B.: Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. In: *Advances in Neural Information Processing Systems*, vol. 29 (2016)
28. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pp. 249–256. *JMLR Workshop and Conference Proceedings* (2010)
29. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) (2014)
30. Shi, X., et al.: Convolutional LSTM network: a machine learning approach for precipitation nowcasting. In: *Advances in Neural Information Processing Systems*, vol. 28 (2015)
31. Wang, Z., Simoncelli, E.P., Bovik, A.C.: Multiscale structural similarity for image quality assessment. In *The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers*, vol. 2, pp. 1398–1402. *IEEE* (2003)
32. Efros, A.A., Leung, T.K.: Texture synthesis by non-parametric sampling. In: *Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 2, pp. 1033–1038. *IEEE* (1999)
33. Goodfellow, I., et al.: Generative adversarial networks. *Commun. ACM* **63**(11), 139–144 (2020)
34. Shen, J., Chan, T.F.: Mathematical models for local nontexture inpaintings. *SIAM J. Appl. Math.* **62**(3), 1019–1043 (2002)