

Self-supervised image blind deblurring using deep generator prior*

LI Yuan^{1,2}, WANG Shasha², and CHEN Lei^{1**}

1. School of Software, Henan University, Kaifeng 475004, China

2. School of Electronic and Electrical Engineering, Shangqiu Normal University, Shangqiu 476000, China

(Received 8 July 2021; Revised 3 October 2021)

©Tianjin University of Technology 2022

Deep generative prior (DGP) is recently proposed for image restoration and manipulation, obtaining compelling results for recovering missing semantics. In this paper, we exploit a general solution for single image deblurring using DGP as the image prior. To this end, two aspects of this object are investigated. One is modeling the process of latent image degradation, corresponding to the estimation of blur kernels in conventional deblurring methods. In this regard, a Reblur2Deblur network is proposed and trained on large-scale datasets. In this way, the proposed structure can simulate the degradation of latent sharp images. The other is encouraging deblurring results faithful to the content of latent images, and matching the appearance of blurry observations. As the generative adversarial network (GAN)-based methods often result in mismatched reconstruction, a deblurring framework with the relaxation strategy is implemented to tackle this problem. The pre-trained GAN and pre-trained ReblurNet are allowed to be fine-tuned on the fly in a self-supervised manner. Finally, we demonstrate empirically that the proposed model can perform favorably against the state-of-the-art methods.

Document code: A **Article ID:** 1673-1905(2022)03-0187-6

DOI <https://doi.org/10.1007/s11801-022-1111-0>

Image deblurring problem has recently attracted considerable attention in the imaging community, where the blur is caused by camera shake or object motions. Many researches have been devoted to addressing this classical problem in the last decade. The progress in this field can be attributed to the advancement of efficient inference algorithms, various natural image priors, and more general motion blur models^[1]. However, it remains a challenging computer vision problem, since blind image deblurring is a typical ill-conditioned inverse problem. Our goal in this paper is to propose a more general solution for single image deblurring using a generative adversarial network (GAN) trained on large-scale natural images as the image prior.

Although lots of attempts have been presented for image deblurring, there are still gaps toward an image prior that captures rich images semantics, such as the sparse priors, the total variation (TV) prior, the edges-based priors and so on^[2]. Recently, the priors based on deep learning have been used for image deblurring. The deep image prior (DIP) is proposed in Refs.[5] and [6] to capture low-level image statistics and show powerful image denoising, super-resolution, inpainting, etc. Subsequently, the deep generative prior (DGP) is presented in Ref.[7] that provides compelling results to restore missing semantics, e.g., color, patch, resolution of various degraded

images.

However, these deep priors mentioned above cannot be directly applied for single image deblurring due to the DIP or DGP network may be designed to generate natural images. Still, it is limited to capture the prior of blur kernels. In Ref.[6], the authors developed a fully connected network (FCN) to capture the prior of uniform motion blur. Nevertheless, DIP is intrinsically limited by the current statistics of the input images. It is impossible that Ref.[6] could perform on images acquired "in the wild" as a general deblurring solution. In Ref.[7], DGP is presented for image restoration and manipulation as the image prior. Although compelling results have been achieved for image restoration (e.g., super-resolution, inpainting), some limitations lead to mismatched reconstruction. The reason is that the training distribution of sharp images inevitably limits the GAN. Moreover, we cannot obtain the complete training dataset for image deblurring due to the unknown image degradation process. Therefore, DGP cannot be directly applied for single image deblurring. Otherwise, it may produce latent images' unfaithful content.

For presenting a general solution for single image blind deblurring by using DGP as the image prior, two problems should be tackled, namely, modeling the process of latent image degradation, encouraging deblurring

* This work has been supported by the National Natural Science Foundation of China (No.61273251).

** E-mail: chenleinj@henu.edu.cn

results faithful to the content of latent images, and matching the appearance of blurry observations. The GAN-based methods tend to produce excellent results due to their ability to capture sharp images' natural distribution. As shown in Fig.1, we select test images from CelebA^[8] to verify the effectiveness of the proposed solution.

To tackle these two problems, we first design a Reblur2Deblur network to model the process of image degradation, training it on the ImageNet dataset^[9]. It is worth pointing out that the Reblur2Deblur should be applied to deblur images "in the wild" at the cost of reducing the performance of deblurring results, not just for domain-specific deblurring like Ref.[10]. It can be observed that Reblur2Deblur network performs worse compared with our method, and there are some apparent artifacts on the deblurring results due to its general applicability. Then we implement the single image deblurring by using DGP and pre-trained ReblurNet networks through relaxing the assumption of existing GAN-inversion methods. Primarily, we allow the generator and pre-trained ReblurNet to be fine-tuned on the fly in a self-supervised manner. It can be observed from Fig.2 that the deblurred result generated by pre-trained GAN in which the generator is fixed has some differences of color and texture compared with the blurry observation. In contrast, our method's deblurring result performs better, having the almost same color and texture.

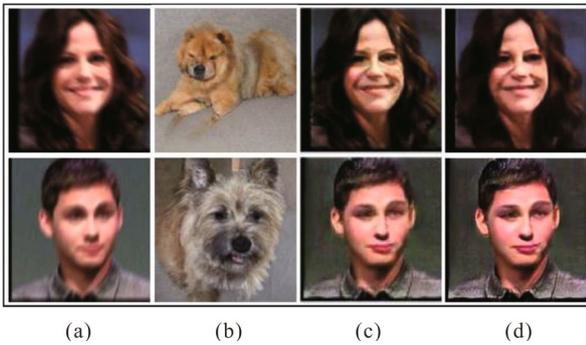


Fig.1 Comparison of deblurring results on the dataset of CelebA^[8]: (a) Blurry image; (b) Image generated by pre-trained GAN; (c) Deblurring results by the Reblur2Deblur network; (d) Deblurring results by our model

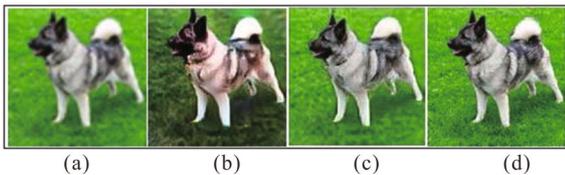


Fig.2 Deblurring results generated by the pre-trained GAN: (a) Blurry image; (b) Deblurred result by the pre-trained GAN from Ref.[7] where the GAN generator is fixed; (c) Deblurred result by our model where the GAN generator is allowed to be fine-tuned; (d) Ground truth

In this paper, we adopt a self-supervised manner to enforce the generator to produce latent sharp images. The diagram of the proposed model is illustrated in Fig.3.

Our contributions are summarized as below.

(1) A specific solution for modeling the process of latent image degradation is presented, corresponding to the estimation of blur kernels in conventional deblurring methods.

(2) A relaxation strategy is adopted while implementing the single image deblurring using the pre-trained GAN and pre-trained ReblurNet. The generator and ReblurNet are allowed to be fine-tuned on the fly in a self-supervised manner.

(3) We verify empirically that our method can perform favorably against state-of-the-art methods on several different datasets, indicating the proposed model's general applicability for single image deblurring.

The proposed deblurring network uses DGP as the image prior, and the overall restoration scheme is illustrated in Fig.3. Two components are the key, ReblurNet and the backward with relaxation strategy. The detailed description will be given as follows.

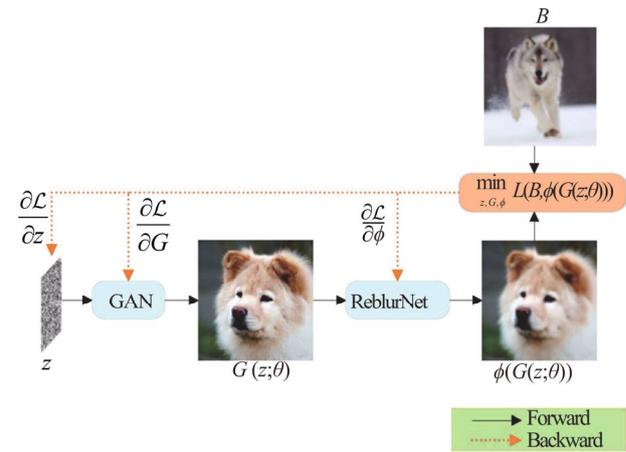


Fig.3 Illustration of the proposed image deblurring network

Motivated by Refs.[10] and [11], Reblur2Deblur is presented using discrete disentangle representation in a supervised manner. As shown in Fig.4, the content and blur features are achieved from the blurred images using the content encoder E_B^c and blur encoder E_B^a . Meanwhile, the content features of sharp images are obtained by content encoders E_S^c . The deblurring results are generated by passing content features and blur features of blurred images through DeblurNet. The re-blurred results are generated by taking the content features of sharp images and the blur features of blurred images as the input of ReblurNet. (B, S) and (\hat{B}, \hat{S}) are pairs of blurred and sharp images used for training. The training image pairs are produced by the method of Ref.[12], and the sharp images are randomly selected from the dataset of ImageNet.

Based on the above analysis, the content encoders ex-

tract the content information, as the blur encoders extract the blur information from blurred images. However, since the sharp images don't have any blur information, the content encoder E_S^c should be a perfect content extractor. Therefore, for learning how to disentangle the content features from blurred images effectively, we share the weights of the last layers of these two content encoders so that the content encoders can project the content features of both domains onto a shared space. However, this design of sharing weights by itself does not guarantee that the blur encoder extracts only blur features. It may encode content or other features as well. Inspired by Refs.[10] and [11], a Kullback-Leibler (KL) divergence loss is added to ensure that the blur encoder extracts as little content information as possible. The training details are described as follows.

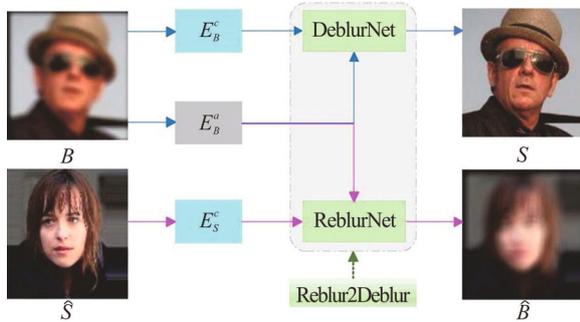


Fig.4 Overview of the proposed Reblur2Deblur framework

To ensure the general applicability and ease to implementation of the proposed Reblur2Deblur, mean square error (*MSE*) is used as the loss function in the training process. Since the pairs of blurred images and sharp images exist, the training is supervised. Thus, the loss function of self-reconstruction and cross-reconstruction can be formulated as

$$\mathcal{L}_{\text{Rec}} = \mathcal{L}_{\text{MSE}}(B, B') + \mathcal{L}_{\text{MSE}}(S, S') + \mathcal{L}_{\text{MSE}}(\hat{S}, \hat{S}') + \mathcal{L}_{\text{MSE}}(\hat{B}, \hat{B}'), \quad (1)$$

where $z_B^c = E_B^c(B)$ and $z_B^a = E_B^a(B)$ represent the blurred image's content and blur features, $B' = R(z_B^c, z_B^a)$ represents the self-reconstruction, $S' = D(z_B^c, z_B^a)$ represents the deblur-reconstruction, R denotes the blurring process by using ReblurNet, and D indicates the deblurring process using DeblurNet. Correspondingly, the self-reconstruction of the sharp image \hat{S} can be represented as $\hat{S}' = D(z_S^c, z_B^a)$, and the re-blur reconstruction as $\hat{B}' = R(z_S^c, z_B^a)$.

As mentioned above, the KL divergence loss can be represented as

$$\mathcal{L}_{\text{KL}} = \frac{1}{2} \sum_{i=1}^N [\mu_i^2 + \sigma_i^2 - \log(\sigma_i^2) - 1], \quad (2)$$

where μ and σ are the mean and standard deviation of the blur features z_B^a , and N is the dimension of z_B^a . In the end,

combining Eq.(1) and Eq.(2), the full objective function can be written as

$$\mathcal{L} = \lambda \mathcal{L}_{\text{KL}} + \mathcal{L}_{\text{Rec}}, \quad (3)$$

where λ is the parameter to balance the importance of \mathcal{L}_{KL} and \mathcal{L}_{Rec} .

The architecture and implementation details refer to Ref.[10], and the same network architectures are used. Each content encoder consists of three strided convolution layers and four residual blocks. The blur encoder is composed of four strided convolution layers and a fully connected layer. For the ReblurNet and DeblurNet, their architectures are symmetric to the content encoders, consisting of four residual blocks and three transposed convolution layers. For training, sharp images are selected randomly from the dataset of ImageNet^[9] and blurred by the method of Ref.[12]. The training process is implemented in the PyTorch framework on a personal computer (PC) equipped with one NVIDIA RTX 2080Ti graphics processing unit (GPU).

In this subsection, we conduct a blurred image restoration using pre-trained GAN (DGP as the image prior) and pre-trained ReblurNet (as shown in Fig.4) with a relaxation strategy. It allows the pre-trained GAN and ReblurNet to be fine-tuned on the fly in a self-supervised manner by this relaxation strategy (the data flow of the updating process is shown in Fig.3). A detailed description of DGP and the relaxation strategy is shown below.

The restoration of the latent image is realized by using GAN-inversion, which can be formulated as

$$\begin{aligned} z^* &= \underset{z \in \mathbb{R}^i}{\text{argmin}} E(\hat{x}, G(z; \theta)) = \underset{z \in \mathbb{R}^i}{\text{argmin}} \mathcal{L}(\hat{x}, R(G(z; \theta))), \\ x^* &= G(z^*; \theta), \end{aligned} \quad (4)$$

where \hat{x} is the blurred observation, x^* is the restoration result, G represents the GAN generator, θ is the parameter of the GAN generator, z is the latent vector as the input of G , $R(\cdot)$ is an image degradation transform, and $\mathcal{L}(\cdot)$ is a distance metric such as *MSE*. Ideally, the degradation transform $R(\cdot)$ is known and G is powerful enough to capture natural images' significant features. In that case, Eq.(4) can force z^* into the latent space, and then produce the reconstruction result in high-performance by using z^* . However, in practice, it is not always the case.

Firstly, the degradation transform $R(\cdot)$ is unknown, and the pre-trained ReblurNet has no capacity for accurately modeling the process of image degradation. Then, the approximated manifold of the restoration result generated by the GAN-inversion methods may not follow the actual one if the GAN generator is fixed in Eq.(4). If the generator is fixed, its inversion cannot faithfully reconstruct the unseen and complex images.

The relaxation strategy is used for reconstruction. It allows the pre-trained GAN and pre-trained ReblurNet to

be fine-tuned on the fly in a self-supervised manner. To this end, the restoration of a blurred observation can be reformulated as follows

$$\begin{aligned} \theta^*, z^*, \mathcal{G}^* &= \underset{\theta, z, \mathcal{G}}{\operatorname{argmin}} \mathcal{L}(\hat{\mathbf{x}}, R(G(z; \theta), \mathcal{G})), \\ \mathbf{x}^* &= G(z^*; \theta^*), \end{aligned} \quad (5)$$

where \mathcal{G} is the parameter of the ReblurNet R . With the relaxation strategy's help, a better deblurring result can be obtained compared with Eq.(4), and the detailed experiments will be demonstrated below.

It is worth pointing out that distance metric \mathcal{L} plays an essential role in the optimization process of Eq.(5). As discussed in Ref.[7], we adopt a BigGAN to reconstruct latent sharp images progressively, and this BigGAN is pre-trained on the ImageNet training set for conditional image synthesis. Accordingly, we adopt the discriminator-based distance metric^[7] as \mathcal{L} in Eq.(5), which is trained along with the BigGAN. In this way, we can preserve the parameter structure of the BigGAN during the optimization process as soon as possible. Furthermore, to improve the performance of latent image restoration, we use the deblurring results generated by the DeblurNet (shown in Fig.4) to constrain the solution space of Eq.(5), which can be rewritten as follows

$$\begin{aligned} \theta^*, z^*, \mathcal{G}^* &= \underset{\theta, z, \mathcal{G}}{\operatorname{argmin}} \mathcal{L}(\hat{\mathbf{x}}, R(G(z; \theta), \mathcal{G})) + \eta \mathcal{L}(\tilde{\mathbf{s}}, G(z; \theta)), \\ \mathbf{x}^* &= G(z^*; \theta^*), \end{aligned} \quad (6)$$

where $\tilde{\mathbf{s}}$ is the deblurring result generated by the DeblurNet, and η is used to balance the importance of the deblurring results generated by the DeblurNet. In our following experiments, we empirically set $\eta=1$. As shown in Fig.1, the performance of the deblurring results produced by DeblurNet is worse than the one generated by Eq.(6).

We use the standard image deblurring benchmark CelebA^[8] to evaluate our method's effectiveness. The CelebA is composed of more than 202 000 face images. Most of the faces are of good quality and at near-frontal poses. We randomly select 1 000 sharp face images as the test set resized and cropped to the size of 128×128, then these selected images are blurred by the method in Ref.[12]. To give a fairer comparison of results, we conduct all the following experiments in the PyTorch framework and use peak signal to noise ratio (*PSNR*) and structural similarity (*SSIM*) as the image quality matrices. The experiments are implemented on a PC equipped with one NVIDIA RTX 2080Ti GPU. The batch size is set to 16, and the learning rate is set to 0.000 1. Adam optimizer is used for back-propagation.

As shown in Eq.(6), the deblurring results produced by DeblurNet are used further to improve the deblurring performance of the proposed method. To validate the necessity of adding the constraint of DeblurNet into the proposed framework, we first implement extensive abla-

tion experiments by removing this constraint on the dataset of LAI et al^[13]. That is, we will compare the restoration performance of Eq.(5) and Eq.(6). Then, we conduct experiments on the benchmark dataset by CelebA^[8] for image blind deblurring, and deblurring results are compared to some state-of-the-art methods to verify the effectiveness of the proposed model.

We conduct ablation experiments to verify the effect of the DeblurNet on our model using the dataset of LAI et al^[13]. Mathematically, our purpose is to validate the necessity of adding $\mathcal{L}(\tilde{\mathbf{s}}, G(z; \theta))$ into Eq.(5) as the new deblurring model. Thus, we compare the restoration performance of these three methods (pre-trained DeblurNet, the method with Eq.(5), and the method with Eq.(6)) using the 100 synthetic non-uniform blurred images. An example of deblurring results produced by these three methods is shown in Fig.5.

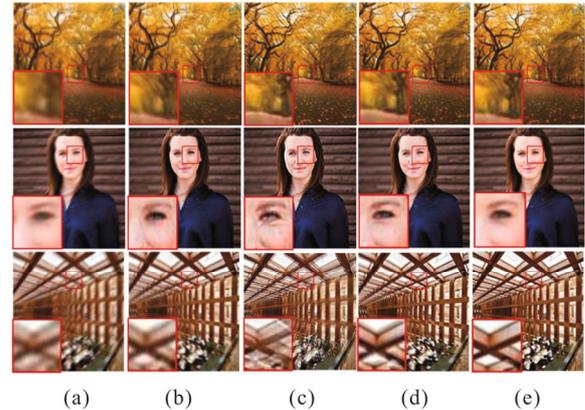


Fig.5 Comparison of deblurring results to verify the effect of the DeblurNet on our model: (a) Blurred images; (b) Deblurring results using the DeblurNet; (c) Deblurring results by Eq.(5); (d) Deblurring results by Eq.(6); (e) Ground truth

It can be observed that despite high-quality deblurring results to be obtained, there are still various artifacts on the deblurring results of DeblurNet and Eq.(5), such as the eye region in the second row. The better deblurring results are achieved by Eq.(6), which combines the essential information of the deblurring products of DeblurNet and Eq.(5). It is easy to observe that the deblurring results of our method with Eq.(6) have a better visual effect.

Using the dataset of CelebA, we compare our method with several state-of-the-art blind deblurring methods, including the ones proposed by KRISHNAN et al^[14], SUN et al^[2], PAN et al^[15], LU et al^[10] and KUPYN et al^[12]. The visual comparison on the dataset of CelebA^[8] is shown in Fig.6. We just offer the deblurring results of Refs.[10] and [14] due to their high performance. It can be observed that there is some incomplete deblurring or artifacts on the products of these state-of-the-art methods.

In contrast, our model can achieve a better visual performance, since a well-trained GAN generator can still capture rich statistics of latent sharp images, which validate our method's effectiveness to a certain extent.

Tab.1 shows the quantitative comparison using average *PSNR* and *SSIM*. As can be observed, our approach takes the second among all methods in terms of average *PSNR* and *SSIM*. Our model performs better than those in Refs.[2] [12], verifying that DGP is more potent than these hand-crafted priors for image deblurring. Ref.[10] performs slightly better than our model, as it is specific for face deblurring, while others are generic deblurring methods.

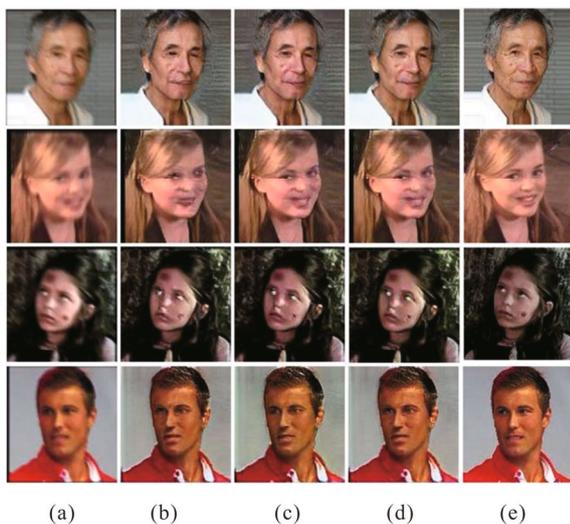


Fig.6 Visual comparison with state-of-the-art methods on the dataset of CelebA^[8]: (a) Blurred images; (b) Deblurring results generated by Ref.[14]; (c) Deblurring results generated by Ref.[10]; (d) Deblurring results generated by our model; (e) Ground truth

Tab.1 Quantitative performance comparison with state-of-the-art methods on the CelebA dataset

Methods	<i>PSNR</i>	<i>SSIM</i>
Ref.[14]	18.51	0.56
Ref.[2]	16.79	0.47
Ref.[15]	17.52	0.54
Ref.[10]	20.79	0.67
Ref.[12]	19.76	0.61
Ours	20.64	0.64

A general solution is proposed for image blind deblurring by using DGP as the image prior, which can be divided into two steps, namely, modeling the process of latent image degradation and deblurring with a relaxation strategy by using a pre-trained GAN and pre-trained ReblurNet. Firstly, the Reblur2Deblur network is proposed to simulate the degradation of sharp images and trained on the ImageNet dataset. Then, the relaxation strategy

allows the parameters of these two pre-trained networks to be fine-tuned in a self-supervised manner. In this way, the proposed model encourages deblurring results faithful to latent images' content and matching the appearance of blurred observations. In the end, we verify empirically that our method can perform favorably against the state-of-the-art methods, indicating the general applicability of the proposed model for single image deblurring.

Statements and Declarations

The authors declare that there are no conflicts of interest related to this article.

References

- [1] NAH S, SON S, LEE S, et al. NTIRE 2021 challenge on image deblurring[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 19-25, 2021, virtual. New York: IEEE, 2021: 149-165.
- [2] SUN L, CHO S, WANG J, et al. Edge-based blur kernel estimation using patch priors[C]//IEEE International Conference on Computational Photography (ICCP), April 19-21, 2013, Cambridge, MA, USA. New York: IEEE, 2013: 1-8.
- [3] WHYTE O, SIVIC J, ZISSERMAN A, et al. Non-uniform deblurring for shaken images[J]. International journal of computer vision, 2012, 98(2): 168-186.
- [4] ZHOU Y, KOMODAKIS N. A map-estimation framework for blind deblurring using high-level edge priors[C]//European Conference on Computer Vision, September 6-12, 2014, Zurich, Switzerland. Berlin, Heidelberg: Springer-Verlag, 2014: 142-157.
- [5] ULYANOV D, VEDALDI A, LEMPITSKY V. Deep image prior[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 9446-9454.
- [6] REN D, ZHANG K, WANG Q, et al. Neural blind deconvolution using deep priors[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 13-19, 2020, Seattle, WA, USA. New York: IEEE, 2020: 3341-3350.
- [7] PAN X G, ZHAN X H, DAI B, et al. Exploiting deep generative prior for versatile image restoration and manipulation[C]//European Conference on Computer Vision, August 23-28, 2020, Glasgow, UK. Berlin, Heidelberg: Springer-Verlag, 2020: 262-277.
- [8] LIU Z, LUO P, WANG X, et al. Deep learning face attributes in the wild[C]//Proceedings of the IEEE International Conference on Computer Vision, December 7-13, 2015, Santiago, Chile. New York: IEEE, 2015: 3730-3738.
- [9] DENG J, DONG W, SOCHER R, et al. Imagenet: a large-scale hierarchical image database[C]//2009 IEEE

- Conference on Computer Vision and Pattern Recognition, June 20-25, 2009, Miami, FL, USA. New York: IEEE, 2009: 248-255.
- [10] LU B, CHEN J C, CHELLAPPA R. Unsupervised domain-specific deblurring via disentangled representations[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 15-20, 2019, Long Beach, USA. New York: IEEE, 2019: 10225-10234.
- [11] DU W, CHEN H, YANG H. Learning invariant representation for unsupervised image restoration[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 13-19, 2020, Seattle, WA, USA. New York: IEEE, 2020: 14483-14492.
- [12] KUPYN O, BUDZAN V, MYKHAILYCH M, et al. Deblurgan: blind motion deblurring using conditional adversarial networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 8183-8192.
- [13] LAI W S, HUANG J B, HU Z, et al. A comparative study for single image blind deblurring[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 1701-1709.
- [14] KRISHNAN D, TAY T, FERGUS R. Blind deconvolution using a normalized sparsity measure[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 20-25, 2011, Colorado Springs, CO, USA. New York: IEEE, 2011: 233-240.
- [15] PAN J, SUN D, PFISTER H, et al. Deblurring images via dark channel prior[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 40(10): 2315-2328.