



# Blind motion deblurring with cycle generative adversarial networks

Quan Yuan<sup>1</sup> · Junxia Li<sup>1,2</sup> · Lingwei Zhang<sup>1</sup> · Zhefu Wu<sup>1</sup> · Guangyu Liu<sup>1</sup>

Published online: 4 October 2019  
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

## Abstract

Blind motion deblurring is one of the most basic and challenging problems in image processing and computer vision. It aims to recover a sharp image from its blurred version knowing nothing about the blurring process. Many existing methods use the maximum a posteriori or expectation maximization framework to tackle this problem, but they cannot handle well the natural images with high-frequency features. Most recently, deep neural networks have been emerging as a powerful tool for image deblurring. In this paper, we show that encoder–decoder architecture gives better results for image deblurring tasks. In addition, we propose a novel end-to-end learning model that refines the generative adversarial network by many novel strategies to tackle the problem of image deblurring. Experimental results show that our model can capture high-frequency features well, and achieve the competitive performance.

**Keywords** Image processing · Blind deblurring · Motion deblurring · Cycle consistency

## 1 Introduction

With the increasing digital cameras and mobile phones, a huge amount of high-resolution images are taken every day [2,8,11,18,51], e.g., the latest Huawei Mate20 series mobile phones have over 60 megapixels. However, sensor shake is often inevitable, resulting in undesirable motion blurring. Although sharp images might be obtained by fixing devices or taking the images again, in many occasions, however, we have no chance to fix the devices or take the images again, for example, in remote sensing [15], video surveillance [44], medical imaging [14] and some other related fields. Therefore, how to obtain sharp images from blurred ones has been noticed by researchers in many fields for many years, but the problem still cannot be well-solved due to the complexity of the motion blur process and, most importantly, the high-resolution natural images often have rich details. Most

existing methods may not produce satisfactory results, as shown in Fig. 1.

Image deblurring problems are a kind of image degradation problems, which can be expressed as

$$I^{\text{blur}} = A(I^{\text{sharp}}) + n, \quad (1)$$

where  $I^{\text{blur}}$  is the given blurred image and  $I^{\text{sharp}}$  is the sharp image.  $A$  is a degradation function, and  $n$  denotes possible noise. In this work, we shall focus on the cases where the degradation process is shift invariant; thereby, the generation process of a blurred image is given by

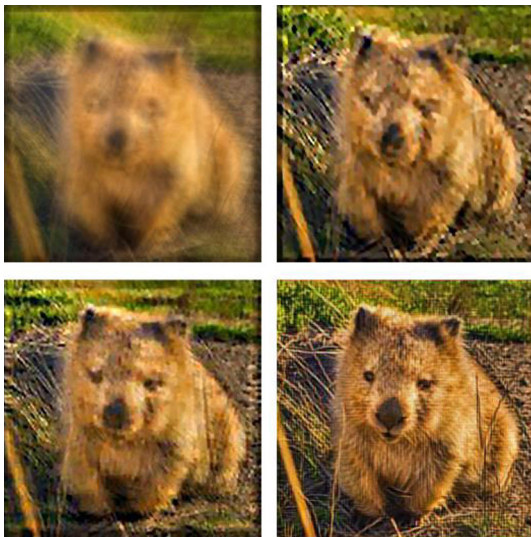
$$I^{\text{blur}} = I^{\text{sharp}} * k + n, \quad (2)$$

where  $*$  denotes 2D convolution and  $k$  is the blur kernel. To obtain the sharp image and the blur kernel simultaneously, some commonly used approaches are MAP [3,52] and variational Bayes [9,28]. Lots of methods have been proposed and explored in the literature. For example, Chan et al. [3] proposed total variation to regularize the gradients of the sharp image. Zhang et al. [52] used a sparse coding method for sharp image recovering. Cai et al. [1] applied sparse representation to estimate the sharp image and blur kernel at the same time. Although these methods obtained moderate good results, they cannot apply to real applications and most importantly, cannot handle well high-frequency features.

✉ Junxia Li  
junxiali99@163.com

<sup>1</sup> Jiangsu Key Lab of Big Data Analysis Technology, Nanjing University of Information Science and Technology, Nanjing 210044, China

<sup>2</sup> Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology (CICAET), Nanjing University of Information Science and Technology, Nanjing 210044, China



**Fig. 1** Results on a challenging motion blurred image. From left to right and top to bottom : blurred image, the result by Pan et al. [35], the results by Xu et al. [48] and Ours. The results show that our model distinctly outperforms the two competing methods

To achieve fast image deblurring, it is straightforward to consider the idea of deep learning that pre-trained network models by plenty of training examples. Although the training process is computationally expensive, deep learning methods can process testing images very efficiently, as they only need to pass an image through the learned network. Most existing deep learning-based methods are built upon the well-known convolution neural network (CNN) [12,34]. However, CNN tends to suppress the high-frequency details in images. To relieve this issue, generative adversarial network (GAN) [10] is one of the promising choices. Kupyn et al. [25] used a GAN-based method that used ResBlocks architecture as the generator. Pan et al. [34] used a GAN-based method to extract intrinsic physical features in images.

In this paper, we proposed a cycle GAN-based method for image deblurring. Specifically, we utilize an encoder–decoder network as a generator and a classification network for the discriminator. It uses a cycle-consistent training strategy that requires training two different generators (one for blurring a sharp image and the other one for sharpening a blurred image) and two discriminators (one for classifying the blurred images and the other for sharp). Besides, we proposed a novel loss function. For cycle loss, which aims to make the reconstructed images and the input images as close as possible under some measurements, there are some classical choices for evaluation, L1 loss, mean square loss, least square loss and perceptual loss.

By some comparison experiments, we demonstrated that perceptual loss can capture high-frequency features in image deblurring. So perceptual loss is used for evaluation in all experiments. Then, we show that U-net-based architec-

ture with L2 norm and perceptual objective performs better in image deblurring problem. Besides, we found that during training, using unpaired images with cycle consistency training strategy can improve the performance in image deblurring tasks. In summary, our contributions of this paper are as follows:

1. A novel cycle GAN-based architecture is presented for image deblurring.
2. We proposed a new loss function in our architecture.
3. A novel training strategy is proposed to tackle the image deblurring problem.

## 2 Related works

Image deblurring is a classical problem in image processing and computer vision. We can divide it into learning-based methods and learning-free methods.

In learning-free methods, most existing works suppose that blur is shift invariant and caused by motion [4,21,29], which can be treated as a deconvolution problem [22,29,46,53]. There are many ways to solve this; Liu et al. [29] used Bayesian estimation, that is,

$$p(I^{\text{sharp}}, k | I^{\text{blur}}) \propto P(I^{\text{blur}} | I^{\text{sharp}}, k) P(I^{\text{sharp}}) P(k). \quad (3)$$

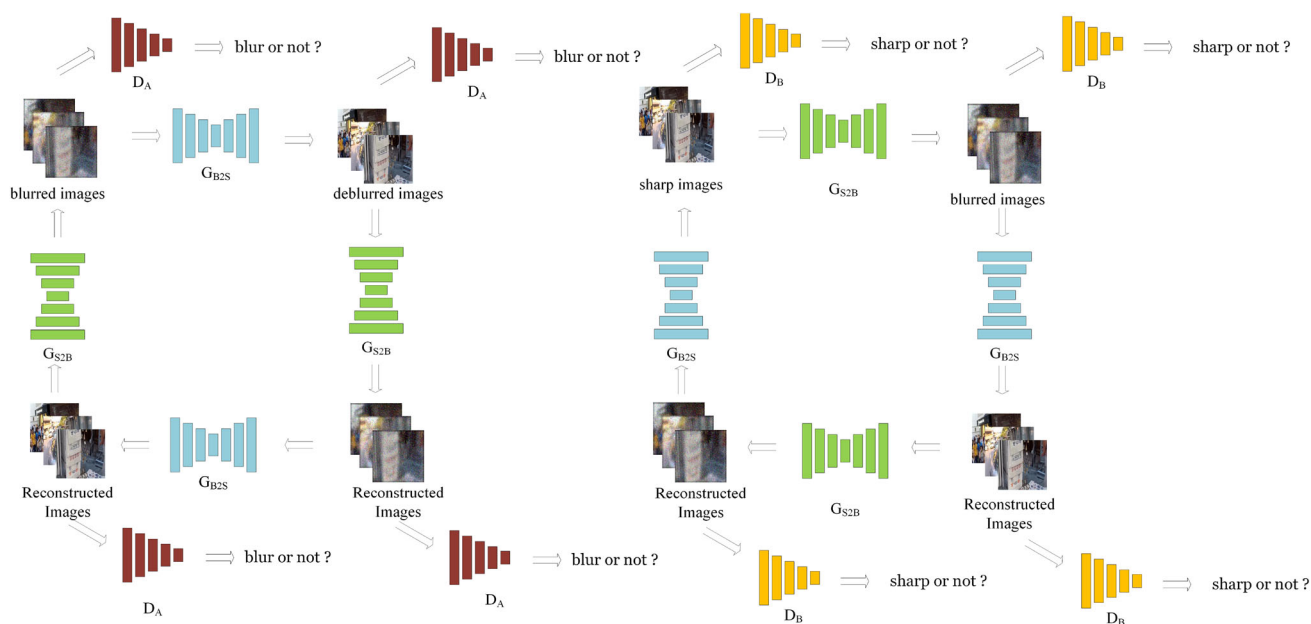
One commonly used deblurring method is the maximum a posteriori (MAP) framework, where the latent sharp image  $I^{\text{sharp}}$  and the blur kernel  $k$  can be obtained by [7],

$$\arg \max_{I^{\text{sharp}}, k} P(I^{\text{blur}} | I^{\text{sharp}}, k) P(I^{\text{sharp}}) P(k). \quad (4)$$

Chan et al. [3] used a robust total variation minimization method which is effective for regularizing the edge of the sharp images. Zhang et al. [52] used a sparse coding method for sharp image recovering, which assumed that the natural image patch can be sparsely represented by an over-complete dictionary. Cai et al. [1] applied sparse representation to estimate sharp image and blur kernel at the same time. Krishnan et al. [23] found that the minimum of their loss function in many existing methods does not correspond to their real sharp images, so they used a normalized sparsity prior to tackle this problem. Michaeli et al. [31] found that multi-scale properties can also be used for blind deblurring problems, and they regard self-similarity as an image prior. Ren et al. [38] used a low rank prior for both raw pixels and their gradients.

Another common approach to estimate motion blur process is to maximize the marginal distribution:

$$\begin{aligned} p(k, I^{\text{blur}}) &= \int p(k, I^{\text{sharp}} | I^{\text{blur}}) dI^{\text{sharp}} \\ &= \int p(I^{\text{blur}} | k) p(k) dI^{\text{sharp}}. \end{aligned} \quad (5)$$



**Fig. 2** Network architecture with cycle-consistent training strategy, including two cycle generators and two discriminators

Fergus et al. [9] proposed a motion deblurring method based on variational Bayes. Levin et al. [28] used an expectation–maximization (EM) method to estimate blur process. These two approaches have some drawbacks: it is hard to optimize, time-consuming and cannot handle high-frequency features well.

Learning-based methods use deep learning techniques, which aim to find intrinsic features through the learning process by themselves. Deep learning [27] has boosted the research in related fields such as image recognition [24] and image segmentation [13]. For deblurring problems using deep learning techniques, [25] trained a CNN architecture to learn the mapping function from blurred images to sharp ones. [34] used a CNN architecture with a physics-based image prior to learn the mapping function.

One of the novel deep learning techniques is generative adversarial networks, usually known as GANs, introduced by Goodfellow et al. [10], inspired by the zero-sum game in game theory proposed by Nash et al. [33], which has achieved many exciting results in image in-painting [50], style transfer [16,17,54], and it can even be used in other fields such as material science [40]. The system includes a generator and a discriminator. The generator tries to capture the latent real data distribution, and output a new data sampled from the real data distribution, while discriminator tries to discriminate whether the input data are from real data distribution or not. Both the generator and the discriminator can be built based on convolutional neural networks [27], and trained based on the above ideas.

Instead of input a random noise in origin generative adversarial nets [10], conditional GAN [6] inputs random noise

with discrete labels or even images [16]. Zhu et al. [54] take a step further, using conditional GAN with unlabeled data, which gives more realistic images in style transfer tasks. Inspired by this idea, Isola et al. [16] proposed one of the first image deblurring models based on generative adversarial nets [10].

While numerous learning-based methods have been proposed, most of the works need paired training data [25,32], which is hard to collect in practice, and strong supervision of these methods may cause over-fitting.

### 3 Proposed method

The goal of blind image deblurring is to recover the sharp images given only the blurred images, with no information about the blurring process. We introduce a GAN-based model with a novel objective and training strategy to tackle this problem. The whole model architecture is shown in Fig. 2.

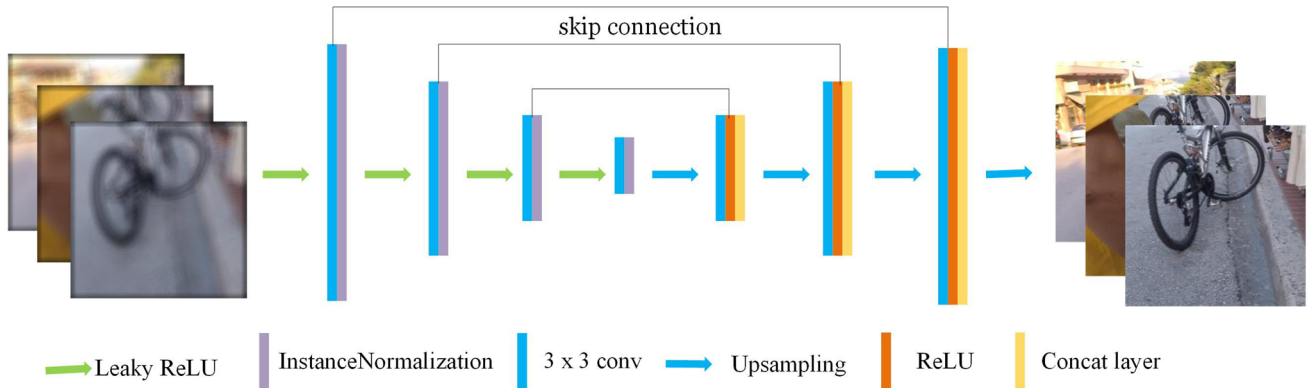
#### 3.1 Model architecture

For discriminator architecture, we use slightly modified version of PatchGAN architecture [16], and the model parameters are shown in Table 1. Instead of classifying the whole image as sharp or not sharp, PatchGAN-based discriminator tries to classify each image patch from the whole images, which gives better results in image deblurring problems. Experiments show that PatchGAN-based architecture can achieve good results if the image patches are a quarter size

**Table 1** Model parameters of discriminators

Layer	Type	Input shape	Output shape	Params
1	Input	(None, 256, 256, 3)	(None, 256, 256, 3)	0
2	conv2d	(None, 256, 256, 3)	(None, 128, 128, 64)	3136
3	conv2d	(None, 128, 128, 64)	(None, 64, 64, 128)	131,200
4	conv2d	(None, 64, 64, 128)	(None, 32, 32, 256)	524,544
5	conv2d	(None, 32, 32, 256)	(None, 16, 16, 512)	2,097,664
6	conv2d	(None, 16, 16, 512)	(None, 16, 16, 1)	8193

Each convolution layer is activated with leaky ReLU ( $\alpha = 0.2$ ) and normalized with instance normalization except the first layer



**Fig. 3** Generator network built in this work. We use encoder–decoder network architecture with skip connection and cycle consistency objective training strategy, which gives comparative results in image deblurring tasks

of the input images [16]; so in the work, we choose image patch =  $70 \times 70$  in all experiments.

As the sharp images and corresponding blurred images are similar in pixel values, it is efficient to distinguish whether the input is from blur domain or sharp domain separately, so we build two discriminators as shown in Table 1. We also report quantitative results (see in Sect. 4—Experiments for more detail).

For generator, Ronneberger et al. [39] used encoder–decoder architecture and Kupyn et al. [25] used ResBlock architecture for image deblurring. The generator architecture is shown in Fig. 3. Hereby, the network only consists of convolution and transpose convolution with instance normalization. For the convolution layer, we apply leaky ReLU activation. For the transpose convolution layer, we apply ReLU activation. In the encoder part, each block consists of a downsampling convolution layer, which halves the height and width with stride 2 and doubles the number of channels  $[H, W, C] \rightarrow [H/2, W/2, C \times 2]$ . In the decoder part, each block inverts the effect of downsampling  $[H, W, C] \rightarrow [H \times 2, W \times 2, C/2]$ . We use filter size of  $4 \times 4$  in all convolution and deconvolution blocks.

### 3.2 Training

Our goal is to learn the mapping function between blur domain  $B$  and sharp domain  $S$  given samples  $\{\text{blur}_i\}_{i=1}^M$  where  $\text{blur}_i \in B$  and  $\{\text{sharp}_j\}_{j=1}^N$  where  $\text{sharp}_j \in S$ . A combination of the following losses is used as objectives:

$$\mathcal{L}(D_A, D_B, G_{B2S}, G_{S2B}) = \mathcal{L}_{\text{adv}} + \alpha \mathcal{L}_{\text{cycle}}, \tag{6}$$

where  $\mathcal{L}, \mathcal{L}_{\text{adv}}, \mathcal{L}_{\text{cycle}}, \alpha$  is the total loss function, adversarial loss, cycle loss and their parameters, respectively. The adversarial loss tries to ensure the deblurred images as realistic as possible; cycle loss tries to ensure that the deblurred images can transfer back to the blur domain, which can also make the deblurred images as realistic as possible.

For the two mapping functions  $G_{S2B} : I^{\text{sharp}} \rightarrow I^{\text{blur}}, G_{B2S} : I^{\text{blur}} \rightarrow I^{\text{sharp}}$  aims to transfer the sharp images to the blur domain and transfer the blurred images to the sharp domain, respectively. The adversarial loss is as follows:

$$\mathcal{L}_{\text{adv}} = \mathcal{L}_{\text{adv}1} + \mathcal{L}_{\text{adv}2}, \tag{7}$$

where

$$\mathcal{L}_{\text{adv}1} = \mathbb{E}_{I^{\text{blur}} \sim p_{\text{data}}(I^{\text{blur}})} [\log D_A(G_{B2S}(I^{\text{blur}}))], \tag{8}$$

$$\mathcal{L}_{adv2} = \mathbb{E}_{I^{sharp} \sim p_{data}(I^{sharp})} [\log(1 - D_B(G_{S2B}(I^{sharp})))] \quad (9)$$

where the two discriminators  $D_A, D_B$  tries to distinguish whether the input images are blur or not, sharp or not, respectively. Generators  $G_{S2B}$  and  $G_{B2S}$  try to fool the discriminators and generate the images from specific domain as realistic as possible.

Isola et al. [16] and Zhu et al. [54] showed that least square loss [30] can perform better than mean square loss in style transfer task, and Kupyn et al. [25] used least square loss [30] for image deblurring tasks. So far, we do not know which loss objective performs better in image deblurring problems, mean square loss or least square loss [30]; we have done some experiments, see in Sect. 4—Experiments for more detail.

For cycle loss, which aims to make the reconstructed images and the input images as close as possible under some measurements, there are two classical choices for evaluation, L1 loss or mean square loss, least square loss [30] or perceptual loss [17]. The experiments show that perceptual loss [17] can capture high-frequency features in image deblurring task, which gives more texture and details. So perceptual loss is used in all experiments. Cycle loss is as follows:

$$\mathcal{L}_{cycle} = \mathcal{L}_{cycle1} + \mathcal{L}_{cycle2}, \quad (10)$$

where

$$\mathcal{L}_{cycle1} = \frac{1}{N^{(i,j)} M^{(i,j)}} \sum_{x=1}^{N^{(i,j)}} \sum_{y=1}^{M^{(i,j)}} (\sigma_{i,j}(I^{sharp})_{x,y} - \sigma_{i,j}(G_{B2S}(I^{blur}))_{x,y})^2, \quad (11)$$

$$\mathcal{L}_{cycle2} = \frac{1}{N^{(i,j)} M^{(i,j)}} \sum_{x=1}^{N^{(i,j)}} \sum_{y=1}^{M^{(i,j)}} (\sigma_{i,j}(I^{blur})_{x,y} - \sigma_{i,j}(G_{S2B}(I^{sharp}))_{x,y})^2, \quad (12)$$

where  $\sigma_{i,j}$  is the feature map which obtains from the  $i$ -th max pooling layer after the  $j$ -th convolution layer from VGG-19 network, and  $N^{(i,j)}, M^{(i,j)}$  are the dimensions of the corresponding feature map; the perceptual loss can capture high level intrinsic features which has been proved to work well

in image deblurring [25], and some other image processing task [16,54].

So in summary, we aim to optimize the following objective function:

$$G_{B2S}^* = \arg \max_{D_A, D_B} \min_{G_{B2S}, G_{S2B}} \mathcal{L}(D_A, D_B, G_{B2S}, G_{S2B}) \quad (13)$$

We train the network with a batch size of 2, and give 100 epochs over the training data. The reconstructed images are regularized with cycle-consistent objective with a strength of 10. No dropout technique is used since the model does not overfit within 100 epochs.

For the optimization procedure, we perform ten steps on  $G_{S2B}$  and  $G_{B2S}$ , and then one step on  $D_A$  and  $D_B$ . We use Adam [19] optimizer with a learning rate of  $2 \times 10^{-3}$  in the first 80 epochs, and then linearly decay the learning rate to zero in the following epochs to ensure the convergence. The whole training process is shown in Fig. 2.

The key point is to train the model in one scene for given epochs, and then move to another scene. When the input is from the blur domain, the starting point and the cycle training process are the left-hand side. When the input is from the sharp domain, the starting point and cycle training process are the right-hand side. Notice that we just have one model during training; for different input from a different domain, the starting point of the model can be a little different.

## 4 Experiments

We implement our model with Keras [5] library. All the experiments are performed on a workstation with NVIDIA Tesla K80 GPU.

### 4.1 Network analysis

*Cycle consistency* Cycle consistency ensures that the deblurred images can transfer to the blur domain, and blur images can transfer back to the sharp domain, which can make sure that our model learns what is “blur” mean, and give more realistic results. We report a quantitative result in Table 2 to demonstrate the advantage of using cycle consistency.

**Table 2** A quantitative evaluation on the effectiveness of cycle consistency

Methods	PSNR	SSIM	MS-SSIM	IFC	VIF	NIQE
Ours (without cycle)	22.74	0.77	0.85	3.78	0.82	4.96
Ours (with cycle)	24.60	0.83	0.90	4.24	0.91	5.23

**Table 3** A quantitative evaluation on the effectiveness of two discriminators

Methods	PSNR	SSIM	MS-SSIM	IFC	VIF	NIQE
Ours (1 discriminator)	21.51	0.80	0.86	3.56	0.79	5.22
Ours (2 discriminators)	24.60	0.83	0.90	4.24	0.91	5.23

**Table 4** Model parameters of generators

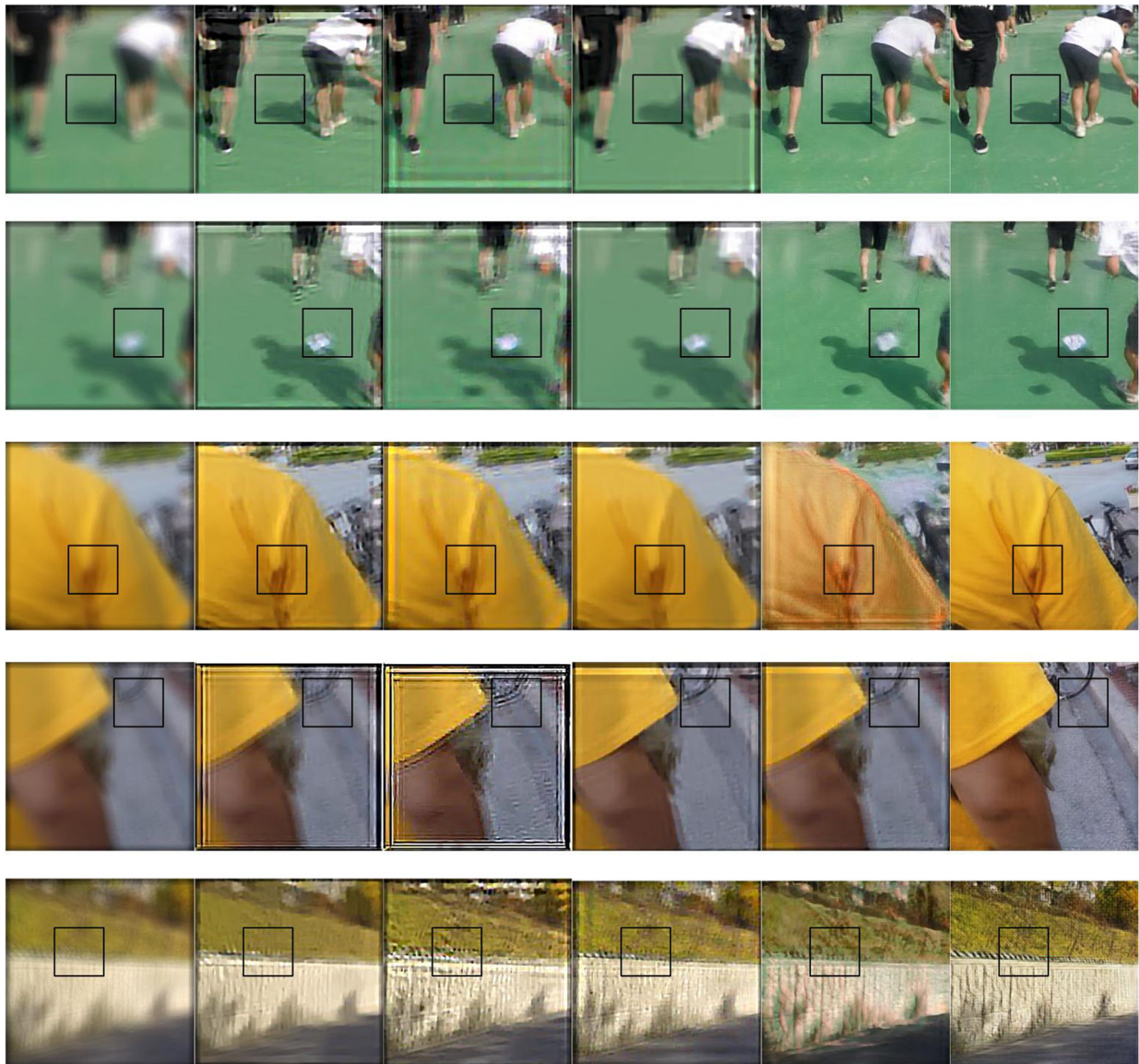
Layer	Type	Input shape	Output shape	Params
1	Input	(None, 256, 256, 3)	(None, 256, 256, 3)	0
2	conv2d	(None, 256, 256, 3)	(None, 128, 128, 32)	1568
3	conv2d	(None, 128, 128, 32)	(None, 64, 64, 64)	32,832
4	conv2d	(None, 64, 64, 64)	(None, 32, 32, 128)	131,200
5	conv2d	(None, 32, 32, 128)	(None, 16, 16, 256)	524,544
6	conv2d	(None, 16, 16, 256)	(None, 32, 32, 128)	524,416
7	concat	(None, 32, 32, 128)	(None, 32, 32, 128)	0
8	conv2d	(None, 32, 32, 128)	(None, 64, 64, 64)	262,208
9	concat	(None, 64, 64, 64)	(None, 64, 64, 64)	0
10	conv2d	(None, 64, 64, 64)	(None, 128, 128, 32)	65,568
11	concat	(None, 128, 128, 32)	(None, 128, 128, 32)	0
12	conv2d	(None, 128, 128, 32)	(None, 256, 256, 3)	3075

Every convolution layer is activated with leaky ReLU layer( $\alpha = 0.2$ ) and normalized with instance normalization. Each deconvolution layer is activate with ReLU and normalized with instance normalization

**Table 5** Comparisons of our model with other four methods on the five images shown in Fig. 4

Results	Methods	PSNR	SSIM	MS-SSIM	IFC	VIF	NIQE
img1	Pan et al. [37]	19.50	0.80	0.98	0.22	0.25	<b>29.45</b>
	Pan et al. [35]	18.55	0.77	0.97	0.19	0.30	28.10
	Xu et al. [48]	19.29	0.80	0.97	0.27	0.44	25.57
	Kupyn et al. [25]	25.98	0.89	0.99	0.67	<b>0.67</b>	26.38
	Ours	<b>26.56</b>	<b>0.91</b>	<b>0.99</b>	<b>0.77</b>	0.66	28.86
img2	Pan et al. [37]	18.62	0.72	0.95	0.24	0.33	<b>25.93</b>
	Pan et al. [35]	18.60	0.71	0.93	0.40	0.54	25.29
	Xu et al. [48]	21.08	0.80	0.96	0.65	0.54	25.29
	Kupyn et al. [25]	24.76	0.87	<b>0.97</b>	1.03	0.64	23.56
	Ours	<b>27.11</b>	<b>0.90</b>	0.97	<b>1.51</b>	<b>0.90</b>	25.42
img3	Pan et al. [37]	17.27	0.60	0.96	0.24	0.16	20.20
	Pan et al. [35]	16.19	0.56	0.96	0.30	0.22	21.80
	Xu et al. [48]	17.25	0.56	0.97	0.42	0.26	<b>22.43</b>
	Kupyn et al. [25]	13.16	0.67	0.96	0.94	0.33	19.27
	Ours	<b>19.45</b>	<b>0.69</b>	<b>0.98</b>	<b>0.95</b>	<b>0.38</b>	21.06
img4	Pan et al. [37]	16.51	0.66	0.96	0.20	0.36	27.05
	Pan et al. [35]	14.16	0.57	0.93	0.22	0.54	25.15
	Xu et al. [48]	19.31	0.73	0.97	0.29	0.49	26.37
	Kupyn et al. [25]	19.65	0.74	0.97	0.27	0.41	<b>27.71</b>
	Ours	<b>21.94</b>	<b>0.82</b>	<b>0.98</b>	<b>0.76</b>	<b>0.74</b>	23.76
img5	Pan et al. [37]	19.25	0.82	0.97	0.25	0.22	28.03
	Pan et al. [35]	18.54	0.80	0.96	0.23	0.26	<b>28.91</b>
	Xu et al. [48]	19.48	0.83	0.97	0.37	0.33	27.12
	Kupyn et al. [25]	19.69	0.82	0.74	0.58	0.45	23.09
	Ours	<b>25.26</b>	<b>0.93</b>	<b>0.98</b>	<b>1.36</b>	<b>0.67</b>	25.94

Bold value indicates that the according methods achieve the state-of-the-art performance comparing with the other methods



**Fig. 4** Some visual comparison results of our model and other four approaches. From left to right: blurred images, the results by Pan et al. [37], Pan et al. [35], Xu et al. [48], Kupyn et al. [25] and the results produced by the proposed method

*Skip connections* Skip connection is widely used to combine the different levels of information which can also benefit back propagation. Inspired by Ronneberger et al. [39] and the success of skip connections, we use U-net-based architecture with skip connections as shown in Fig. 3.

*Motion blur generation* For motion blur generation, Kupyn et al. [25] proposed a method which can generate realistic random motion trajectory. We use this method to generate blur kernels to blur images.

*Architecture selection* We do some experiments in Table 3 to demonstrate the effectiveness of using two discriminators. The results show that using two discriminators can significantly improve performance.

We also do some experiments to find the optimal choice for generator and training objectives in Table 4. The results show that for image deblurring task, the optimal choice for generator architecture is U-net-based architecture, and the optimal evaluation for optimization objective is a least square loss. The generator architectures are shown in Fig. 3 and Table 4

## 4.2 Results analysis

### 4.2.1 GoPRO dataset

The proposed GoPRO dataset consists of 21,000 images, including 11,000 blurred images and 10,000 sharp images.

**Table 6** Quantitative evaluations for our model in GoPRO dataset using different evaluations

Methods	PSNR	SSIM	MS-SSIM	IFC	VIF	NIQE
Pan et al. [37]	18.70	0.77	0.94	0.23	0.23	25.92
Pan et al. [35]	17.92	0.73	0.95	0.28	0.36	<b>26.05</b>
Xu et al. [48]	19.29	0.73	<b>0.97</b>	0.44	0.42	24.13
Kupyn et al. [25]	20.91	0.82	0.91	0.65	0.50	23.07
Ours	<b>24.29</b>	<b>0.86</b>	0.95	<b>1.14</b>	<b>0.66</b>	25.20

Bold value indicates that the according methods achieve the state-of-the-art performance comparing with the other methods

**Table 7** Quantitative evaluations of three methods in terms of PSNR and SSIM on the Kohler et al. [20] dataset

Methods	PSNR	SSIM
Nah et al. [32]	29.00	0.91
Tofighi et al. [45]	25.67	0.88
Ours	26.13	0.90

We use 20,000 images for training, and 1000 images for testing.

For image evaluation, most of the works used full reference measurements PSNR and SSIM in all their experiments [34–37]. Tofighi et al. [45] used SNR and ISNR for evaluation. For other image assessments, VIF [41] captures wavelet features which focus on high-frequency features, and IFC [42] puts more weights on edge features. Lai et al. [26] pointed out that the full reference image assessments VIF and IFC are better than PSNR and SSIM.

For the fair comparison, we choose different learning-free methods proposed by Pan et al. [37], Pan et al. [35] and learning-based methods proposed by Xu et al. [48] and Kupyn et al. [25].

Some result examples are shown in Table 5 and Fig. 4. All the salient regions are pointed out in each image. We also report the quantitative results of average evaluations on this dataset in Table 6. We observe that our method outperforms many competitive methods in PSNR, SSIM, MS-SSIM, IFC and VIF. We also observe that our methods recovered more textures, sharp edges, background and fewer artifacts.

#### 4.2.2 Köhler dataset

This dataset consists of four ground-truth images and 12 blurry images for each of them. These blurs are caused by replaying recorded 6D camera motion. We report the quantitative results on this dataset comparing with Tofighi et al. [45] in Table 7. From this table, we can see that our method performs better than Tofighi et al. [45].

**Table 8** Quantitative evaluations for our model in dataset of Sun et al. [43]

Methods	PSNR
Xu et al. [49]	28.01
Ours	26.79

**Table 9** Quantitative evaluations for our model in dataset of Wieschollek et al. [47]

Methods	SSIM
Wieschollek et al. [47]	0.86
Ours	0.72

#### 4.2.3 Dataset of Sun et al.

This dataset consists of 640 images generated from 80 natural images and eight blur kernels. We add 1% Gaussian noise as done in Xu et al. [49]. We report the quantitative results on this dataset comparing with Xu et al. [49] in terms of average PSNR in Table 8. The method proposed by Xu et al. [49] suppresses extraneous textures and meanwhile enhances salient edges during training, which gives better results in PSNR performance.

#### 4.2.4 Dataset of Wieschollek et al.

In the method of Wieschollek et al. [47], the authors use a 720p high-resolution video from Youtube to generate the dataset. For a fair comparison, here we report the quantitative results on this dataset comparing with Wieschollek et al. in Table 9. The method proposed by Wieschollek et al. [47] used recurrent neural network architecture with multi-scale paired input, which can achieve state-of-the-art performance when dealing with video blurring and burst blurring problems.

## 5 Conclusions

In this paper, we provide an unsupervised method for blind motion deblurring problem. We build the network with a cycle training strategy. We use two discriminators to distinguish whether the input is blur or not, sharp or not, separately,



**Table 10** Performance of different generator architectures dealing with deblurring problem in GoPRO dataset

Methods	PSNR	SSIM	MS-SSIM	IFC	VIF	NIQE
Ours (Resblock based + L1)	18.16	0.73	0.81	3.00	0.80	4.24
Ours (Resblock based + L2)	23.48	0.80	0.88	3.88	0.85	5.01
Ours (U-net based + L1)	21.32	0.76	0.88	3.17	0.84	<b>5.60</b>
Ours (U-net based + L2)	<b>24.60</b>	<b>0.83</b>	<b>0.90</b>	<b>4.24</b>	<b>0.91</b>	5.23

We evaluate on different full reference and non-reference evaluations. The results show that U-net based + L2 architecture performs better than the other choices

Bold value indicates that the according methods achieve the state-of-the-art performance comparing with the other methods

which can perform better in image deblurring tasks. We show that encoder–decoder-based architecture gives better results. For optimization objective, least square loss performs better than mean square loss.

During training, the experiments show that this model can deal with image deblurring task well without giving any domain-specific knowledge. It can recover more high-frequency textures and details, which not only outperform many competitive methods in many different image quality assessments but also in human visualization evaluation (Table 10).

We show that the key point is to train the model in one scene for given epochs, and then move to another scene, which can ensure that the model learns exactly what “blur” and “sharp” mean.

We also show that our model can handle blur caused by motion or camera shake; the recovered image has fewer artifacts compared to many existing methods. We conduct extensive experiments on three other datasets and report quantitative results.

**Acknowledgements** We thank Guangan Liu and Yubao Sun for their helpful discussions and advices. This work was supported by the National Natural Science Foundation of China (NSFC) under Grant Nos. 61702272, 61773219, 61771249 and 61802199, and the Startup Foundation for Introducing Talent of NUIST (2243141701034).

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

- Cai, J.F., Ji, H., Liu, C., Shen, Z.: Blind motion deblurring from a single image using sparse approximation. In: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pp. 104–111. IEEE (2009)
- Cambra, A.B., Murillo, A.C., Muñoz, A.: A generic tool for interactive complex image editing. *Vis. Comput.* **34**, 1493–1505 (2018)
- Chan, T.F., Wong, C.K.: Total variation blind deconvolution. *IEEE Trans. Image Process.* **7**, 370–375 (1998)
- Chandramouli, P., Jin, M., Perrone, D., Favaro, P.: Plenoptic image motion deblurring. *IEEE Trans. Image Process.* **27**, 1723–1734 (2018)
- Chollet, F.: Keras (2015)
- Dai, B., Fidler, S., Urtasun, R., Lin, D.: Towards diverse and natural image descriptions via a conditional gan. *arXiv preprint arXiv:1703.06029* (2017)
- Duda, R.O., Hart, P.E., Stork, D.G.: *Pattern Classification*. Wiley, London (2012)
- Fan, Q., Shen, X., Hu, Y.: Detail-preserved real-time hand motion regression from depth. *Vis. Comput.* **34**(9), 1145–1154 (2018)
- Fergus, R., Singh, B., Hertzmann, A., Roweis, S.T., Freeman, W.T.: Removing camera shake from a single photograph. In: *ACM Transactions on Graphics (TOG)*, vol. 25, pp. 787–794. ACM (2006)
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: *Advances in Neural Information Processing Systems*, pp. 2672–2680 (2014)
- Guan, H., Cheng, B.: How do deep convolutional features affect tracking performance: an experimental study. *Vis. Comput.* **34**, 1701–1711 (2018)
- Guo, S., Yan, Z., Zhang, K., Zuo, W., Zhang, L.: Toward convolutional blind denoising of real photographs. *arXiv preprint arXiv:1807.04686* (2018)
- He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. In: *Computer Vision (ICCV), 2017 IEEE International Conference on*, pp. 2980–2988. IEEE (2017)
- Hobbs, J.B., Goldstein, N., Lind, K.E., Elder, D., Dodd III, G.D., Borgstede, J.P.: Physician knowledge of radiation exposure and risk in medical imaging. *J. Am. Coll. Radiol.* **15**, 34–43 (2018)
- Ineichen, P.: High turbidity solis clear sky model: development and validation. *Remote Sens.* **10**(3), 435 (2018)
- Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. *arXiv preprint* (2017)
- Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: *European Conference on Computer Vision*, pp. 694–711. Springer, Berlin (2016)
- Khmag, A., Al Haddad, S., Ramlee, R., Kamarudin, N., Malallah, F.L.: Natural image noise removal using nonlocal means and hidden markov models in transform domain. *Vis. Comput.* **34**, 1661–1675 (2018)
- Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014)
- Kohler, R., Hirsch, M., Mohler, B.J., Scholkopf, B., Harmeling, S.: Recording and playback of camera shake: benchmarking blind deconvolution with a real-world database, pp. 27–40 (2012)
- Kotera, J., Šroubek, F.: Motion estimation and deblurring of fast moving objects. In: *2018 25th IEEE International Conference on Image Processing (ICIP)*, pp. 2860–2864. IEEE (2018)
- Krishnan, D., Fergus, R.: Fast image deconvolution using hyper-laplacian priors. In: *Advances in Neural Information Processing Systems*, pp. 1033–1041 (2009)

23. Krishnan, D., Tay, T., Fergus, R.: Blind deconvolution using a normalized sparsity measure. In: *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pp. 233–240. IEEE (2011)
24. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*, pp. 1097–1105 (2012)
25. Kupyn, 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. 2261–2269 (2017)
26. Lai, W.S., Huang, J.B., Hu, Z., Ahuja, N., Yang, M.H.: A comparative study for single image blind deblurring. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1701–1709 (2016)
27. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature* **521**, 436 (2015)
28. Levin, A., Weiss, Y., Durand, F., Freeman, W.T.: Efficient marginal likelihood optimization in blind deconvolution. In: *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pp. 2657–2664. IEEE (2011)
29. Liu, G., Chang, S., Ma, Y.: Blind image deblurring using spectral properties of convolution operators. *IEEE Trans. Image Process.* **23**, 5047–5056 (2014)
30. Mao, X., Li, Q., Xie, H., Lau, R.Y., Wang, Z., Smolley, S.P.: Least squares generative adversarial networks. In: *Computer Vision (ICCV), 2017 IEEE International Conference on*, pp. 2813–2821. IEEE (2017)
31. Michaeli, T., Irani, M.: Blind deblurring using internal patch recurrence. In: *European Conference on Computer Vision*, pp. 783–798. Springer, Berlin (2014)
32. Nah, S., Kim, T.H., Lee, K.M.: Deep multi-scale convolutional neural network for dynamic scene deblurring. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017)
33. Nash, J.: Non-cooperative games. *Ann. Math.* **54**(2), 286–295 (1951)
34. Pan, J., Liu, Y., Dong, J., Zhang, J., Ren, J., Tang, J., Tai, Y.W., Yang, M.H.: Physics-based generative adversarial models for image restoration and beyond. *arXiv preprint arXiv:1808.00605* (2018)
35. Pan, J., Sun, D., Pfister, H., Yang, M.H.: Blind image deblurring using dark channel prior. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1628–1636 (2016)
36. Pan, J., Sun, D., Pfister, H., Yang, M.H.: Deblurring images via dark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell.* **40**(10), 2315–2328 (2018)
37. Pan, J., Zhe, H., Su, Z., Yang, M.H.: Deblurring text images via l0-regularized intensity and gradient prior. In: *IEEE Conference on Computer Vision and Pattern Recognition* (2014)
38. Ren, W., Cao, X., Pan, J., Guo, X., Zuo, W., Yang, M.H.: Image deblurring via enhanced low-rank prior. *IEEE Trans. Image Process.* **25**, 3426–3437 (2016)
39. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-assisted Intervention*, pp. 234–241. Springer, Berlin (2015)
40. Sanchez-Lengeling, B.: Inverse molecular design using machine learning: generative models for matter engineering. *Science* **361**, 360–365 (2018)
41. Sheikh, H.R., Bovik, A.C.: Image information and visual quality. In: *Acoustics, Speech, and Signal Processing, 2004. Proceedings (ICASSP'04)*. IEEE International Conference on, vol. 3, pp. iii–709. IEEE (2004)
42. Sheikh, H.R., Bovik, A.C., De Veciana, G.: An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Trans. Image Process.* **14**(12), 2117–2128 (2005)
43. Sun, L., Cho, S., Wang, J., Hays, J.: Edge-based blur kernel estimation using patch priors. In: *Proceedings of the IEEE International Conference on Computational Photography (ICCP)*. IEEE (2013)
44. Sun, Z., Zhang, Q., Li, Y., Tan, Ya.: Dppdl: a dynamic partial-parallel data layout for green video surveillance storage. *IEEE Trans. Circuits Syst. Video Technol.* **28**(1), 193–205 (2018)
45. Tofghi, M., Li, Y., Monga, V.: Blind image deblurring using row-column sparse representations. *IEEE Signal Process. Lett.* **25**(2), 273–277 (2018)
46. Wang, R., Tao, D.: Training very deep CNNs for general non-blind deconvolution. *IEEE Trans. Image Process.* **27**, 2897–2910 (2018)
47. Wieschollek, P., Hirsch, M., Scholkopf, B., Lensch, H.P.A.: Learning blind motion deblurring. In: *International Conference on Computer Vision*, pp. 231–240 (2017)
48. Xu, L., Jia, J.: Two-phase kernel estimation for robust motion deblurring. In: *European Conference on Computer Vision*, pp. 157–170. Springer, Berlin (2010)
49. Xu, X., Pan, J., Zhang, Y., Yang, M.: Motion blur kernel estimation via deep learning. *IEEE Trans. Image Process.* **27**(1), 194–205 (2018)
50. Yeh, R.A., Chen, C., Lim, T.Y., Schwing, A.G., Hasegawa-Johnson, M., Do, M.N.: Semantic image inpainting with deep generative models. In: *CVPR*, vol. 2, p. 4 (2017)
51. Yu, Z., Liu, Q., Liu, G.: Deeper cascaded peak-piloted network for weak expression recognition. *Vis. Comput.* **34**, 1691–1699 (2018)
52. Zhang, H., Yang, J., Zhang, Y., Huang, T.S.: Sparse representation based blind image deblurring. In: *Multimedia and Expo (ICME), 2011 IEEE International Conference on*, pp. 1–6. IEEE (2011)
53. Zhang, K., Zuo, W., Gu, S., Zhang, L.: Learning deep CNN denoiser prior for image restoration. In: *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2 (2017)
54. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv preprint* (2017)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Quan Yuan** received his bachelor's degree in atmospheric science from Nanjing University of Information Science and Technology, Nanjing, China, in 2017, where he is pursuing the M.S. degree since 2017. His research interests include image processing and machine learning.



**Junxia Li** received the Ph.D. degree from Nanjing University of Science and Technology in 2017. Now, she is a lecturer in the School of Automation, Nanjing University of Information Science and Technology, Nanjing, China. Her research interests include image segmentation, visual saliency detection and computer vision.



**Zhefu Wu** received his bachelor's degree from the School of Binjiang, Nanjing University of Information Science and Technology, Nanjing, China, in 2017, where he is pursuing the M.S. degree since 2017. His research interests include video prediction.



**Lingwei Zhang** received his bachelor's degree from the School of Automation, Nanjing University of Information Science and Technology, Nanjing, China, in 2017, where he is pursuing the M.S. degree since 2017. His research interests include video prediction.



**Guangyu Liu** received his bachelor's degree from the School of Automation, Nanjing University of Information Science and Technology, Nanjing, China, in 2014, where he is pursuing the M.S. degree since 2017. His research interests include image deblurring.