Structure-Aware Motion Deblurring Using Multi-Adversarial Optimized CycleGAN

Yang Wen, Jie Chen, Bin Sheng, *Member, IEEE*, Zhihua Chen, Ping Li, *Member, IEEE*, Ping Tan, *Senior Member, IEEE*, and Tong-Yee Lee, *Senior Member, IEEE*

***Abstract*— Recently, Convolutional Neural Networks (CNNs) have achieved great improvements in blind image motion deblurring. However, most existing image deblurring methods require a large amount of paired training data and fail to maintain satisfactory structural information, which greatly limits their application scope. In this paper, we present an unsupervised image deblurring method based on a multi-adversarial optimized cycle-consistent generative adversarial network (CycleGAN). Although original CycleGAN can handle unpaired training data well, the generated high-resolution images are probable to lose content and structure information. To solve this problem, we utilize a multi-adversarial mechanism based on CycleGAN for blind motion deblurring to generate high-resolution images iteratively. In this multi-adversarial manner, the hidden layers of the generator are gradually supervised, and the implicit refinement is carried out to generate high-resolution images continuously. Meanwhile, we also introduce the structure-aware mechanism to enhance the structure and detail retention ability of the multi-adversarial network for deblurring by taking the edge map as guidance information and adding multi-scale edge constraint functions. Our approach not only avoids the strict need for paired training data and the errors caused by blur kernel estimation, but also maintains the structural information better with multi-adversarial learning and structure-aware mechanism. Comprehensive experiments on several benchmarks have shown that our approach prevails the state-of-the-art methods for blind image motion deblurring.**

***Index Terms*— Unsupervised image deblurring, multi-adversarial, structure-aware, edge refinement.**

Manuscript received December 12, 2020; revised May 15, 2021; accepted June 15, 2021. Date of publication July 2, 2021; date of current version July 9, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 61872241 and Grant 61572316; in part by The Hong Kong Polytechnic University under Grant P0030419, Grant P0030929, and Grant P0035358; and in part by the Ministry of Science and Technology, Taiwan, under Grant 108-2221-E-006-038-MY3. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Jiaying Liu. *(Corresponding authors: Bin Sheng; Zhihua Chen.)*

Yang Wen and Bin Sheng are with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: shengbin@sjtu.edu.cn).

Jie Chen is with the Samsung Electronics (China) Research and Development Centre, Nanjing 210012, China (e-mail: ada.chen@samsung.com).

Zhihua Chen is with the Department of Computer Science and Engineering, East China University of Science and Technology, Shanghai 200237, China (e-mail: czh@ecust.edu.cn).

Ping Li is with the Department of Computing, The Hong Kong Polytechnic University, Hong Kong (e-mail: p.li@polyu.edu.hk).

Ping Tan is with the School of Computing Science, Simon Fraser University, Burnaby, BC V5A 1S6, Canada (e-mail: pingtan@sfu.ca).

Tong-Yee Lee is with the Department of Computer Science and Information Engineering, National Cheng Kung University, Tainan 70101, Taiwan (e-mail: tonylee@mail.ncku.edu.tw).

This article has supplementary downloadable material available at https://doi.org/10.1109/TIP.2021.3092814, provided by the authors.

Digital Object Identifier 10.1109/TIP.2021.3092814

I. Introduction

MOTION blur is a painful problem during the process of taking photos by lightweight devices like mobile phones. Because of these inevitable factors in the image acquisition process especially under poor lighting conditions, the image quality will be degraded to undesired blurry images. Image motion deblurring problem is to restore the sharp image from a given blurry image [2]–[5]. There are mainly two types of image deblurring methods: blind and non-blind deblurring. Many works have been focused on non-blind deblurring in recent years, which are based on the assumption that the blur function is known before, like blur caused by camera shake, etc. However, it is a severely ill-posed problem to find the blur kernel for every pixel. Aiming at the problem of non-blind image deblurring, some methods are intended to parameterize the blur model according to the assumed blur source. In [6], Whyte *et al.* assume that the blurs are only caused by the movement of 3D cameras. While this assumption is not always true in practice. Recently, CNNs have shown strong semantic analysis ability and have been widely used in blind image deblurring. In [7], Madam *et al.* propose an architecture that consists of an autoencoder to learn the data prior and an adversarial network to generate and discriminate between the sharp and blurred features. In [8], Schuler *et al.* describe how to use a trainable model to learn blind deconvolution. In [9], Xu *et al.* propose a model that contains two stages, suppressing extraneous details and enhancing sharp edges. In [10], Nah *et al.* propose a multi-scale convolutional neural network (CNN) for blind image deblurring.

Although significant improvements have been made by the emergence of deep learning, three major challenges still stand in the way of the blind motion deblurring problem. (1) Missing handcrafted yet critical prior features: Deep CNNs often ignore the traditional manual features based on statistical prior knowledge for image deblurring. Previous studies [11], [12] have shown that the traditional manual features are very important for image deblurring. (2) Obsolete disposing of multi-scale deblurring: Although the multi-scale architecture has long been used to solve the deblurring problem [10], it may emphasize high-level semantic information and underestimate the key role of underlying features in deblurring. (3) Limited training data: The traditional motion deblurring methods always aim to find the cause of blurring and estimate the approximate blur kernel so as to obtain the training data. The estimation method often has a certain error which leads to the blurred training data generated can only contain

several general specific categories. In addition, training data must contain both pairs of blurred and sharp images [10], [13]–[15], which are often quite difficult to obtain in reality. Otherwise, there is a large distribution difference between the synthesized and real blurred images, so the universality of the network model trained by sharp image and its corresponding synthesized blurred data needs to be further improved.

For the paired training data requirements, various unsupervised CNNs-based methods have been proposed. Nimisha *et al.* [16] propose an unsupervised generative adversarial network (GAN) based method with additional reblur loss and multi-scale gradient loss. Although this method shows good performance on the synthetic data set, it is only for the special blurred type and cannot achieve a satisfactory effect on the real blurred images. Other existing unsupervised methods based on GAN for the image-to-image translation mainly involve learning the mapping of blurred image domain to the sharp image domain, such as CycleGAN [1] and discover generative adversarial network (DiscoGAN) [17]. In this paper, we choose CycleGAN [1] that is well known for its unpaired image-to-image translation to instead the previous network model. We take the advantage of CycleGAN to treat blurred images and sharp images as two different data distributions to overcome the paired data training problem about deblurring mentioned above. Based on CycleGAN, a more flexible deblurring effect can be achieved with an unpaired image dataset than other methods that can only be trained with pairs of sharp and blurred images. For the obsolete disposing of multi-scale deblurring problem, we utilize a multi-adversarial architecture that includes a series of slightly modified dense-blocks [13] to improve the deblurring performance. The multi-adversarial strategy can iteratively generate the sharp images from the low-resolution to the high-resolution. For the missing handcrafted yet critical prior features problem, since previous studies have shown that sharp edge restoration plays a very important role in the structural maintenance of deblurring [9], [11], [12], we use a structure-aware strategy that includes edge guidance by adding the edge map as part of the input and structure enhancement by minimizing the edge loss. Moreover, our architecture can avoid the introduction of other noise factors (such as color and texture) into the generated deblurred images, which is easy to occur in the original CycleGAN, and keep the structure and detail information consistent with the corresponding sharp image as much as possible. Combing with the perceptual loss [18] and multi-scale structural similarity (MS-SSIM [19]) loss, we obtain significantly better image deblurring results than most of the existing methods. As shown in Fig. 1, compared to the classical unsupervised methods, our results in Fig. 1(c) are more satisfying. Our work makes the following three main contributions:

- We introduce an unsupervised approach based on CycleGAN [1] for blind motion deblurring without assuming any restricted blur kernel model. It can avoid the errors caused by blur kernel estimation and overcome the drawback that other methods require pairwise images as the training data [10], [14]. In addition, our model can also automatically generate blurred images from sharp images simultaneously to provide more available data for subsequent studies.
- We propose a multi-adversarial architecture to solve the artifact problem in high-resolution image generation. Different from the traditional multi-scale methods [10] and [20], the multi-adversarial constraints can promote the network to produce the results closest to the sharp images at different resolutions. Although the multi-adversarial structure is somewhat more burdensome than the original CycleGAN, it can effectively suppress the artifacts in the high-resolution image generation.
- We present a structure-aware mechanism based on edge clues for motion deblurring. Since how to effectively restore sharp edges is vital for deblurring effect based on the previous research [9], [12], [21], blurred image and its edge map are fused as the input. Besides, multi-scale edge constraints are introduced in the multi-adversarial architecture to make the adversarial network generate persuasive structural information at different resolutions.

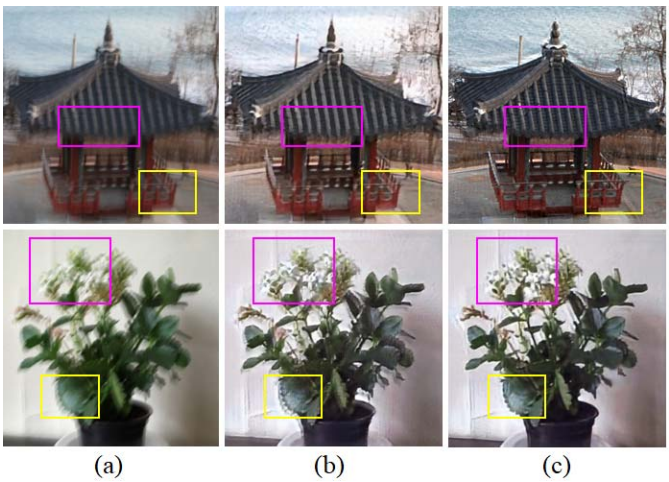


Fig. 1. Comparison of deblurred images by our method and the original CycleGAN on the real images. (a) Blurred images. (b) Deblurring results using original CycleGAN [1]. (c) Deblurring results by our method. It shows our method is more satisfying, especially in the pink and yellow rectangles.

II. Related Work

In recent years, blind image motion deblurring has attracted considerable research attention in the field of computer vision [22]–[24]. In general, image motion deblurring tasks are based on the assumption that the blur is uniform and spatially invariant [9] and the endless number of solutions [10], [14], [15] have been proposed. According to the need for blur kernel estimation, image deblurring methods can be divided into with kernel and kernel-free two categories.

A. Kernel Estimation Method for Motion Deblurring

1) Traditional Kernel Estimation Method for Deblurring: Commonly, diverse methods tend to take advantage of the sharp edges to estimate the blur kernel. Some kernel estimation approaches [25]–[28] rely on implicit or explicit extraction of edge information to detect and enhance the image edges

through a variety of technologies, such as bilateral filtering and gradient amplitude. In [29], Xu *et al.* propose an L0-regularized gradient prior based on the sharp edge information for blind image deblurring. In [12], Pan *et al.* develop an optimization method based on L_0-regularized intensity and gradient prior to generate reliable intermediate results for blur kernel estimation. In [30], Sun *et al.* use dictionary learning to predict the sharp edges with the sharp edge patches of clear images for deblurring. In [31], Pan *et al.* describe a blind image deblurring method with the dark channel prior. In [32], Kim *et al.* propose to estimate the motion flow and the latent sharp image simultaneously based on the total variation (TV)-L1 model. In [33], Bai *et al.* propose a multi-scale latent structure prior and gradually restore the sharp images from the coarse-to-fine scales on a blurry image. Recently, thanks to the powerful semantic analysis and deep mining ability of CNNs, more works tend to use large-scale samples to solve the blind image deblurring problems.

2) CNNs Based Kernel Estimation Method for Deblurring: In recent years, CNNs have played an unparalleled advantage in solving computer vision problems including image deblurring and achieved many promising results [7]–[9]. Some methods use CNNs to estimate the blur kernel to achieve the deblurring task. For instance, Sun *et al.* mainly estimate the probabilistic distribution of the unknown motion blur kernel based on CNN for deblurring [34]. However, these methods have strict requirements for paired training data and cannot directly realize the transformation from the blurred image to the sharp image, and still cannot avoid errors in the process of blur kernel estimation based on CNNs [35], [36]. In contrast, our approach can avoid these errors, since our method is based on the unsupervised image-to-image translation with unpaired training data and can directly realize the transformation from blurred images to sharp images without kernel estimation process. In this paper, we show a comparison with [12], [31], [34] to verify our advantages in Session IV-E.

B. Kernel-Free Learning for Motion Deblurring

Since the popularity of GAN, which is originally designed to solve different image-to-image translation problems [37], [38], more people try to generate deblurred images directly from the blur images with GAN to avoid the distortion caused by kernel estimation. In [39], Xu *et al.* use CNN to learn the deconvolution operation guided by traditional deconvolution schemes. In [7], Nimisha *et al.* propose a novel deep filter based on GAN architecture integrated with global skip connection and dense architecture to tackle this problem. In [13], a special GAN with a densely connected generator and a discriminator is used to generate a deep filter for deblurring. Kupyn *et al.* [14] propose the DeblurGAN method based on the conditional adversarial network and a multi-component loss function for blind motion deblurring. In [40], Li *et al.* propose a depth guided network which contains a deblurring branch and a depth refinement branch for dynamic scene deblurring. Although breakthroughs have been made in these methods, the problems of missing structure information and demanding paired training data still need to be solved. Even the subsequent methods [16], [41], [42] can realize the deblurring task by unsupervised use of unpaired training data, [16], [41] only target at the specific image domain deblurring problem, while [42] will encode other factors (color, texture, etc., instead of blurred information) into the generated deblurred image. Different from these previous methods, our unsupervised method can solve the demand of paired training data problems for the image deblurring. Meanwhile, we utilize the multi-adversarial architecture and structure-aware mechanism to further remove the unpleasant artifacts and maintain structure information effectively.

III. Proposed Method

Our overall flowchart is shown in Fig. 2. In Fig. 2, G_B and G_S are two generator sub-networks which transform from the sharp image to the blurred image and from the blurred image to the sharp image, respectively. D_B and D_S are the discriminators to distinguish the real images and generated images, and give feedback to the generators. Different from the traditional CycleGAN [1], we use the form of multi-adversarial in different resolution constraints to gradually improve the quality of the generated images and use skip connections to make the low-level information better guide the high-level generation structure. Meanwhile, we design a structure-aware mechanism by introducing the multi-scale edge constraints in the multi-adversarial architecture to make the adversarial network generate persuasive structural information at different resolutions, and edge map is also used as part of the input to facilitate the network's retention of structural information. Besides, we add a variety of loss functions (structural loss MS-SSIM and perceptual loss obtained by VGG16) to further strengthen the constraints to reduce the generated false information. Compared with other methods, our method can not only solve the demand of paired data problem, but also can maintain more structural information and achieve a better deblurring effect.

A. Original CycleGAN-Based Deblurring Method

Inspired by the success of the unsupervised method CycleGAN [1], we try to handle the demand of paired training data problem by the unsupervised image-to-image translation manner. Based on the original CycleGAN for deblurring, the architecture includes two generator sub-networks G_B and G_S that transform from the blurred image b to the deblurred (sharp) image s and from the sharp (deblurred) image s to the blurred image b, respectively. D_B and D_S are the discriminators for the blurred image and the sharp (deblurred) image, respectively. The loss function of CycleGAN contains two parts: adversarial loss and cycle-consistency loss. On one hand, the adversarial loss aims to match the distribution of generated images to the data distribution in the target domain. On the other hand, the cycle consistency loss ensures that the cyclic transformation can bring the image back to its original state. Based on the traditional CycleGAN, we can successfully transform from the blurred image domain to the sharp image domain with unpaired training data. However, some annoying artifacts (such as color and texture) will be encoded into the generated results and some structure information also

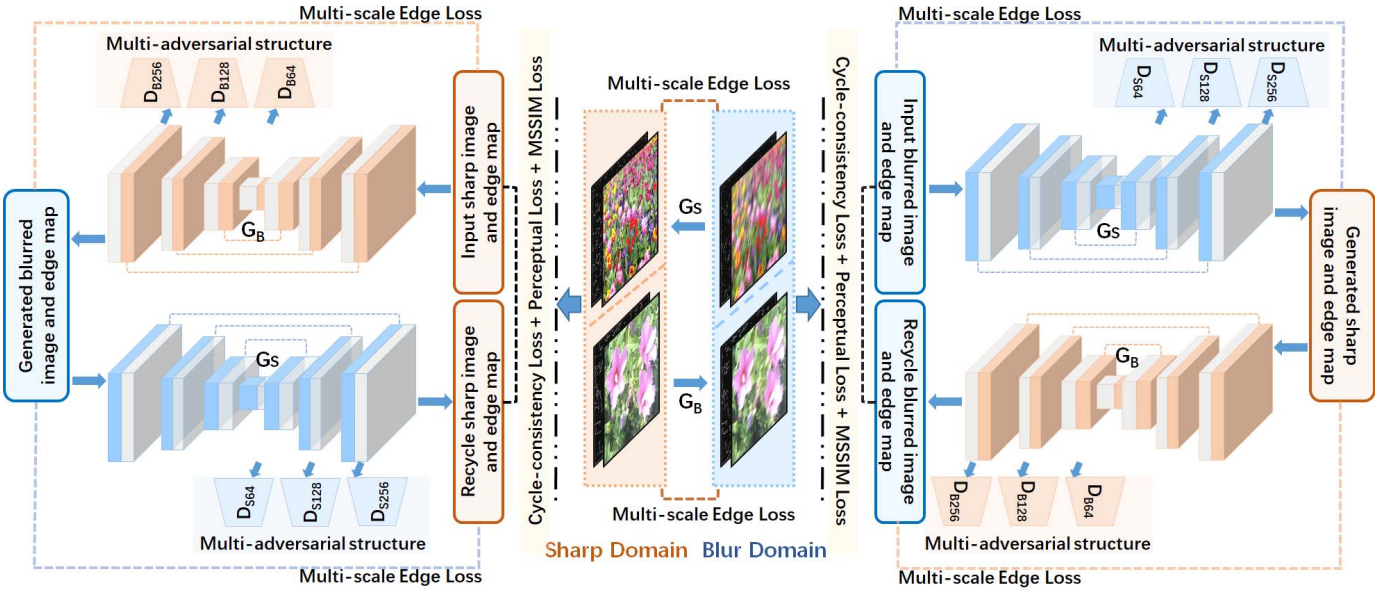


Fig. 2. The flowchart of our structure-aware multi-adversarial optimized CycleGAN. Our architecture relies on the unsupervised image-to-image translation to learn the mapping between blurred images and deblurred (sharp) images with unpaired training data. G_S and G_B are two generator sub-networks for translating blurred image to sharp image and translating sharp image to blurred image, respectively. D_{S64}, D_{S128} and D_{S256} are the discriminators to determine whether the image generated by G_S is real or fake at three resolutions. D_{B64}, D_{B128} and D_{B256} are the discriminators to determine whether the image generated by G_B is real or fake at three resolutions. We restore sharp images by this multi-adversarial manner to iteratively generate high-resolution from low-resolution images. In addition, we introduce the structure-aware mechanism by adding edge input to guide the generation procedure and multi-scale edge losses to maintain more structure details at different resolutions. Besides, we utilize cycle-consistency loss, perceptual loss and MS-SSIM loss to enforce constraints on the structure generation.

sometimes lost [16], [43]. In order to solve these problems, we expect to improve the generation effect step by step with multi-adversarial architecture and structure-ware mechanism.

B. Multi-Adversarial Generative Network

As discussed in Section II-B, the classical GAN-based structure often introduces artifacts when generating realistic images, especially with the increase of resolution. To solve this problem, a multi-scale way is preferred to improve the quality of the generated images [10]. Ideally, a mature multi-scale approach not only can significantly improve the network performance but also need to minimize parameters to reduce time consumption and hardware burden. However, the parameters in some multi-scale approaches at each scale are still independent of each other in some multi-scale methods [10], [20]. Given this, we introduce the multi-adversarial architecture in our unsupervised deblurring model to make full use of the input information and avoid the problem of false information increasing with the increase of resolution.

Inspired by the traditional encoder-decoder network structure [44], the generator G_S in our proposed multi-adversarial network is shown in Fig. 3. The input of the generator sub-network G_S is the blurred image and the corresponding edge map obtained by Sobel operator. The edge map used as part of the input can provide additional structural information to the network. G_S contains a series of convolution layers, deconvolution layers and upper sampling layers. Feature maps are generated from each deconvolution layer through a 3×3 convolution forward layer with output images at different resolutions. From Fig. 3, generator G_S can produce the output images with three resolution levels. Then, three independent discriminators will judge the authenticity of the generated images on different resolutions and feed information to the generators. The hidden layers with different resolutions in the network are constrained and the feature maps are iteratively optimized to generate higher quality results. Additionally, the generated edge maps at three different resolutions are used for multi-scale edge constraints to improve the structure retention performance of the network. We also use skip connections to take full advantage of the low-level information to guide the deconvolution process.

For a blurred image b, generator G_S generates synthesized sharp image s_{b_1}, s_{b_2}, s_{b_3} as outputs. The s_{b_3}, which presents the output of the last deconvolution layer, is sent as the input of G_B to generate three reconstructions $\widehat{b}_1$, $\widehat{b}_2$ and $\widehat{b}_3$. Similarly, for a deblurred (sharp) image s as input, G_B will output synthesized blurred images b_{s_1}, b_{s_2} and b_{s_3}. And with b_{s_3} as the input, the generator G_S will produce three reconstructions $\widehat{s}_1$, $\widehat{s}_2$ and $\widehat{s}_3$. We then supervise these different outputs to force them closer to the target at different resolutions. D_{S64}, D_{S128} and D_{S256} are defined for G_S. D_{B64}, D_{B128} and D_{B256} are defined for G_B. Three resolutions of 64×64, 128×128 and 256×256 are applied on the corresponding deconvolution layers, respectively. The adversarial losses can be written as Eq. (1) and Eq. (2):

$$\begin{aligned} L_{adv}(G_S, D_{S_i}) &= E_{b\sim p(b)}\left[\log(1 - D_{S_i}(G_S(b)_i))\right] \\ &\quad + E_{s_i\sim p(s_i)}\left[\log(D_{S_i}(s_i))\right] \qquad (1) \\ L_{adv}(G_B, D_{B_i}) &= E_{s\sim p(s)}\left[\log(1 - D_{B_i}(G_B(s)_i))\right] \\ &\quad + E_{b_i\sim p(b_i)}\left[\log(D_{B_i}(b_i))\right] \qquad (2) \end{aligned}$$

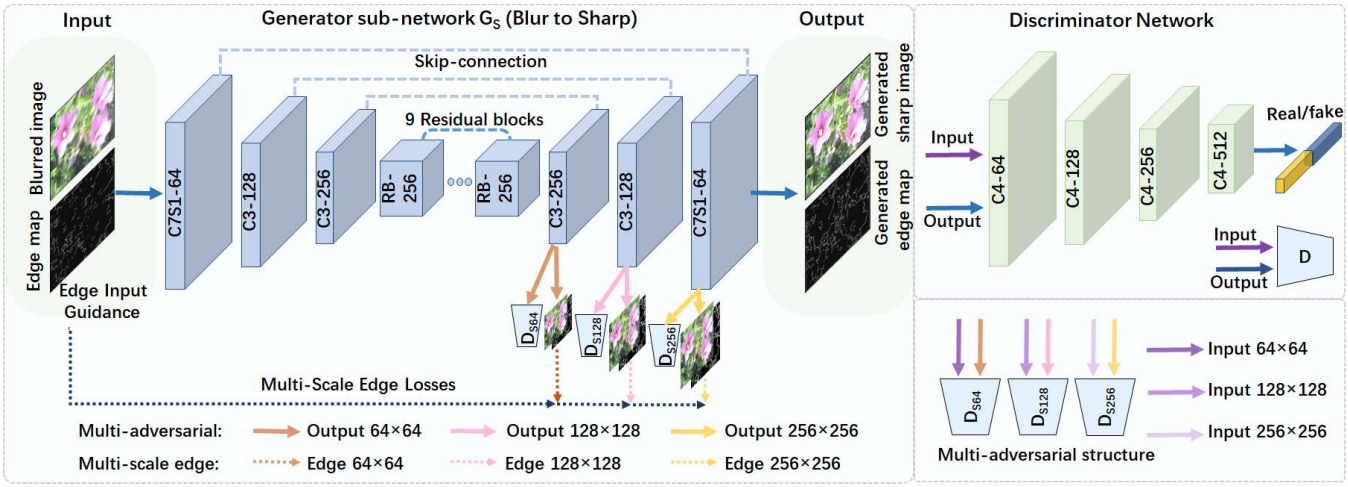


Fig. 3. Network structure of the proposed multi-adversarial generator. G_S is the generator sub-network for the translation from the blurred image to the deblurred (sharp) image. The input of the generator sub-network G_S is the blurred image and the corresponding edge map obtained by Sobel operator. By the multi-adversarial manner, G_S can produce three different resolution outputs (64×64, 128×128 and 256×256). Multi-adversarial supervision is achieved through multiple discriminators in the hidden layers. Discriminators D_{S64}, D_{S128}, D_{S256} are defined for G_S at three different resolutions, respectively. In addition, the generated edge maps at three different resolutions are used for multi-scale edge constraints to improve the structure retention performance of the network. The specific parameters of the generator sub-network are shown in the figure so that we can train our multi-adversarial model with a specific size and test the image of any size.

where $G_S(b)_i = s_{b_i}$, $G_B(s)_i = b_{s_i}$ and $i = 1, 2, 3$ corresponds to the three different resolutions. b_i and s_i are the blurred image and sharp image at i^{th} resolution, respectively. D_{B_i} and D_{S_i} are the discriminators corresponding to G_B and G_S at i^{th} scale, respectively.

As for the cycle-consistency loss in the traditional CycleGAN, it can be improved to multiple resolutions:

$$L_{cyc_{b_i}} = \left\|\widehat{b_i} - b_i\right\|_1 = \|G_B(G_S(b)_3)_i - b_i\|_1 \quad (3)$$

$$L_{cyc_{s_i}} = \|\widehat{s_i} - s_i\|_1 = \|G_S(G_B(s)_3)_i - s_i\|_1 \quad (4)$$

where $G_S(b)_3 = s_{b_3}$ and $G_B(s)_3 = b_{s_3}$. The final multi-adversarial objective function is defined as:

$$\begin{aligned} &L_{MultiGAN}(G_S, G_B, D_S, D_B) \\ &= \sum_{i=1}^{3} (L_{adv}(G_S, D_{S_i}) + L_{adv}(G_B, D_{B_i}) \\ &\quad + \mu_i (L_{cyc_{b_i}} + L_{cyc_{s_i}})) \end{aligned} \quad (5)$$

Simplified as:

$$L_{MultiGAN} = \sum_{i=1}^{3} (L_{adv_i} + \mu_i L_{cyc_i}) \quad (6)$$

where μ_i is the weight parameter at i^{th} resolution to balance the different components. $L_{cyc_i} = L_{cyc_{s_i}} + L_{cyc_{b_i}}$, and $L_{adv_i} = L_{adv}(G_S, D_{S_i}) + L_{adv}(G_B, D_{B_i})$

C. Structure-Aware Mechanism for Deblurring

The high-frequency details of the image are weakened to some extent due to the blurring process, how to restore the structure and details as much as possible in the image deblurring task is very important. Previous studies [11], [16], [45] prove that image edge is of great significance in subjective image quality assessment and image restoration tasks. In [16], an unsupervised network for deblurring with a reblurring cost and a scale-space gradient cost is proposed. In [11], Vasu *et al.* first investigate the relationship between the edge profiles and the camera motion, and then incorporate the edge profiles into an existing blind deblurring framework. In [45], a two-stage edge-aware network is proposed to improve image deblurring according to the feature that human eyes pay more attention to edge sharpening. Although several structure-aware strategies have been successively applied to deblurring problems, it is still difficult to maintain structure information and reduce inherent ambiguity in unsupervised deblurring tasks.

In order to preserve the structural information of the deblurred image to the maximum extent, we introduce the structure-aware mechanism by taking the corresponding edge map as part input and adding multi-scale edge constraint functions in the multi-adversarial architecture. Different from the structure-aware mechanism in other image processing tasks, the structure-aware mechanism in our unsupervised deblurring model not only includes the input edge clues for structural information assistance but also includes multi-scale edge constraints for generating the deblurring with different resolutions. Besides, the multi-scale edge constraints can be organically combined with the multi-adversarial strategy to promote the generation of structural information in unsupervised networks. We have verified that both of them can effectively promote the structure retention ability of the network and generate a more satisfactory deblurring effect through the ablation experiments.

The proposed structure-aware mechanism can emphasize the protection of image geometry to alleviate the important ambiguity problem of the original CycleGAN. In this paper, the proposed structure-aware mechanism network is shown in Fig. 3. Due to the input edge guidance, the Eq. (1) and

Fig. 4. Comparative experiment of structure maintenance effect. (a) The original blurred image. (b) Deblurring result using CycleGAN [1]. (c) Deblurring result with edge map as input. (d) Deblurring result with edge loss. (e) Deblurring result with both edge map as input and edge loss. It shows our method is more satisfying, especially in the yellow rectangles.

Eq. (2) can be revised as Eq. (7) and Eq. (8):

$$L_{adv}(G_S, D_{S_i}) = E_{b\sim p(b)}\left[\log(1 - D_{S_i}(G_S(b, b^e)_i))\right] + E_{s_i\sim p(s_i)}\left[\log(D_{S_i}(s_i, s_i^e))\right] \quad (7)$$

$$L_{adv}(G_B, D_{B_i}) = E_{s\sim p(s)}\left[\log(1 - D_{B_i}(G_B(s, s^e)_i))\right] + E_{b_i\sim p(b_i)}\left[\log(D_{B_i}(b_i, b_i^e))\right] \quad (8)$$

where b^e and s^e are the edge maps of the image b and image s obtained by Sobel operator, respectively. b_i^e and s_i^e are the responding edge maps at i^{th} resolution. By this edge guidance manner, we can take the advantage of the additional edge information to make the generated images in the target domain contain similar edge structure information of the source domain and better guide the discriminator to distinguish the generated images from the real images. However, even the edge guidance can improve the accuracy of discrimination, we find that the generated deblurred image still exits the problems of ringing and oversharp.

In order to handle the existing problems and force the structure of the generated deblurred image to match its corresponding sharp image, we introduce the multi-scale edge losses in the multi-adversarial structure. Since our unsupervised method has no access to the corresponding reference image and it is difficult to generate an accurate corresponding edge map, we follow the heuristic from [16], [46] and utilize the fact that the resized image b^η which is obtained by shrinking a blurred image b with a factor of η is sharper than the image b itself. Thus, we introduce the multi-scale edge losses to enforce the edge of the generated deblurred image to match its corresponding sharp image. The factor of η in our model is set to 0, 1/2 and 1/4 for three different scales respectively. Then, the introduced multi-scale edge losses are defined as:

$$L_{Grad_{b_i}} = \left\|\nabla s_{b_i} - \nabla b_i\right\|_1 = \|\nabla(G_S(b)_i) - \nabla b_i\|_1 \quad (9)$$

$$L_{Grad_{s_i}} = \left\|\nabla b_{s_i} - \nabla s_i\right\|_1 = \|\nabla(G_B(s)_i) - \nabla s_i\|_1 \quad (10)$$

where ∇ is the Sobel operator to calculate the gradient map of an image, and $L_{Grad_i} = L_{Grad_{b_i}} + L_{Grad_{s_i}}$.

Fig. 4 shows the effect of just using the edge loss and adding edge as an input to the generator. From Fig. 4, most structure information can be migrated to the target domain with edge input in Fig. 4(c), and most artificial noise can be effectively eliminated through multi-scale edge losses in Fig. 4(d). The combination can better improve the motion deblurring performance as shown in Fig. 4(e).

D. The Network Structure

1) Generator: The generator in our architecture is shown in Fig. 3. It contains a series of convolution layers and residual blocks. Specific as follows: $C7S1-64$, $C3-128$, $C3-256$, $RB256\times 9$, $TC64$, $TC32$, $C7S1-3$, where, $C7S1-k$ represents a 7×7 ConvBNReLU (Convolution+BatchNorm+ReLU) block with stride 1 and k filters, $C3-k$ represents a 3×3 ConvBNReLU block with stride 2 and k filters. $RBk\times n$ denotes k filters and n residual blocks which contain two 3×3 convolution layers, TCk represents a 3×3 TConvBNReLU (Transposed Convolution+BatchNorm+ReLU) block with stride $1/2$ and k filters. In addition, we introduce the structure-aware architecture (including edge input guidance and multi-scale edge constrains) in G_S and G_B during training process.

2) Discriminator: The discriminator is also shown in Fig. 3. Classic PatchGANs [47] is used as a discriminator to classify overlapping image blocks and determine whether they are real or false. All the discriminator networks at three resolutions mainly include: $C64-C128-C256-C512$, here Ck presents a 4×4 ConvBNLeakyReLU (Convolution + BatchNorm + LeakyReLU) block with stride 2 and k filters. The parameter of LeakyReLU is set to 0.2 in our experiment. According to the specific parameters of the generator and discriminator, we can train our multi-adversarial model with a specific size and test the images of any size.

E. Loss Functions

1) Multi-Scale SSIM Loss: The perceptually motivated metric Structural SIMilarity index (SSIM) [48] has often been used to measure the similarity of two images. To preserve the information of contrast, luminance, structure in the generated images and alleviate the ambiguity problem of CycleGAN, we use the Multi-scale SSIM loss (MS-SSIM) based on SSIM between $\widehat{b_i}$ and b_i in our model. The MS-SSIM we used is defined as:

$$L_{MSSIM_{b_i}} = 1 - \left[l_M(b_i, \widehat{b_i})\right]^{\alpha_M} \prod_{j=1}^{M} [c_j(b_i, \widehat{b_i})]^{\beta_j} [m_j(b_i, \widehat{b_i})]^{\gamma_j} \quad (11)$$

where $l(b_i, \widehat{b_i}) = \frac{2\mu_{b_i}\mu_{\widehat{b_i}}+C_1}{\mu_{b_i}^2+\mu_{\widehat{b_i}}^2+C_1}$, $c(b_i, \widehat{b_i}) = \frac{2\sigma_{b_i}\sigma_{\widehat{b_i}}+C_2}{\sigma_{b_i}^2+\sigma_{\widehat{b_i}}^2+C_2}$ and $m(b_i, \widehat{b_i}) = \frac{\sigma_{b_i\widehat{b_i}}+C_3}{\sigma_{b_i}\sigma_{\widehat{b_i}}+C_3}$. $(b_i, \widehat{b_i})$ denotes the image pair of input image and the reconstructed image, respectively. μ_{b_i}, $\mu_{\widehat{b_i}}$, σ_{b_i}, $\sigma_{\widehat{b_i}}$, $\sigma_{b_i\widehat{b_i}}$ indicate the means, standard deviations and cross-covariance of the image pair $(b_i, \widehat{b_i})$, respectively. C_1, C_2 and C_3 are the constants determined according to reference [48]. $l(b_i, \widehat{b_i})$, $c(b_i, \widehat{b_i})$ and $m(b_i, \widehat{b_i})$ denote the comparison components of luminance, contrast and structure between b_i and $\widehat{b_i}$, respectively. α, β and γ are the hyper-parameters set according to [48], which are used to control the relative weight of the three comparison components.

Similarly, the MS-SSIM loss function $L_{MSSIM_{s_i}}$ between $\widehat{s_i}$ and s_i is defined as the same way, and the total MS-SSIM loss at i^{th} resolution is $L_{MSSIM_i} = L_{MSSIM_{b_i}} + L_{MSSIM_{s_i}}$.

2) Perceptual Loss: Previous work [38] shows that cyclic perceptual-consistency losses have the ability to preserve original image structure by investigating the combination of high-level and low-level features extracted from the second and fifth pooling layers of VGG16 [49] architecture. According to [38], the formulation of cyclic perceptual-consistency loss is given below, where $(b_i, \widehat{b_i})$ refers to the blurred and ground truth image set, ϕ is a VGG16 [38], [49] feature extractor from the second and fifth pooling layers:

$$L_{Perceptual_{b_i}} = \left\|\phi(\widehat{b_i}) - \phi(b_i)\right\|_2^2 \tag{12}$$

Similarly, $L_{Perceptual_{s_i}}$ between $\widehat{s_i}$ and s_i is defined as the same way, and the total perceptual loss at i^{th} resolution is $L_{Perceptual_i} = L_{Perceptual_{s_i}} + L_{Perceptual_{b_i}}$.

3) Identity Preserving Loss: In addition, we use an identity preserving loss to reinforce the identity information of the input image during the unpaired image-to-image translation. Thus, information such as the color of the input and output images can be mapped as accurately as possible. The identity preserving loss between the source domain and target domain can be defined as:

$$L_{Id_{b_i}} = \|G_B(b)_i - b_i\|_1 \tag{13}$$

$$L_{Id_{s_i}} = \|G_S(s)_i - s_i\|_1 \tag{14}$$

The total identity preserving loss at i^{th} resolution is $L_{Id_i} = L_{Id_{b_i}} + L_{Id_{s_i}}$. From the above loss functions described in Eq. (1) $\sim$ Eq. (14), the total loss for our deblurring model is:

$$L = \sum_{i=1}^{3} (L_{adv_i} + \omega_1 L_{cycle_i} + \omega_2 L_{Grad_i} + \omega_3 L_{MSSIM_i} + \omega_4 L_{Id_i} + \omega_5 L_{Perceptual_i}) \tag{15}$$

where, ω_1, ω_2, ω_3, ω_4 and ω_5 are non-negative constants to adjust different influence on overall deblurring effects. i denotes the component at i^{th} resolution. Similar to other previous methods [1], [10], parameters ω_1, ω_2, ω_3, ω_4 and ω_5 in Eq. (15) are set according to the data characteristics for different cases and we weight each loss empirically to balance the importance of each component.

IV. Experimental Results

A. Implementation Details

We conduct our training and testing experiments on a workstation with Intel Xeon E5 CPU and NVIDIA 2080ti GPU. The model we used is implemented with Pytorch platform [50]. For fairness, all the experiments are set in the same data set and environment except for special instructions. Throughout our experiments, we use ADAM [51] solver for model training with parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$. Limited by the memory, the batch-size is set to 2 for all the methods. The initial learning rate is fixed to 0.0002 for the first 30 epoches and then decay to one-tenth every 30 epoches. Totally, 200 epoches already satisfy the convergence condition.

B. Datasets and Metrics

For the blurred text images, we use the dataset BMVC_TEXT [52] which totally contains $66K$ text images with the size 300×300. This dataset contains both defocus blur generated by anti-aliased disc and motion blur generated by a random walk. The blurred images in BMVC_TEXT are divided into two parts: the training set and the test set (50% of the total, and no crossover is ensured), and the corresponding sharp images are divided in the same way. During the training process, we crop the image into 128×128 image blocks in both the blur set and the sharp set. The parameter ω_1 is set to 5, parameters ω_2 and ω_3 are set to 0.5, ω_4 is set to 10 and ω_5 is set to 0 in Eq. (15) because we find that the perceptual loss $L_{Perceptual}$ has little impact on overall performance. To compare with other classical deblurring methods, we choose the algorithms given by Pan *et al.* [12], [31], Xu *et al.* [29], Sun *et al.* [34], MS-CNN [10], DeblurGAN [14]. We also choose other unsupervised methods CycleGAN [1], Madam *et al.* [16] and UID-GAN [43] that trained on the same text training dataset with our unpaired data.

For the blurred face images, the CelebA dataset [53] which mainly includes more than $200K$ face images with size 178×218 are used. We first select $200K$ data from the data set, where $100K$ is the sharp images and the other $100K$ is the blurred images. In addition, we select 2000 images from the remaining images for testing. We scale all the images to 128×128 and ensure that there is no paired data during the unsupervised algorithm training. The method of generating blurred images by sharp images is consistent with the method proposed in UID-GAN [43]. The parameters $\omega_1 \sim \omega_4$ are set in the same way as BMVC_TEXT [52] dataset, and the parameter ω_5 is set to 5.

For the motion blurred images, the same as [10], we firstly use the GoPro dataset proposed in [10] to train our model. Since our model is based on the unsupervised image-to-image translation, during the training process, we firstly segregate the GoPro dataset into two parts. We just use the blurred images from one part and the clean (sharp) image from the second part so that there are no corresponding pairs while the training process. 2103 blurred/clear unpaired images in GoPro dataset are used for training and the remaining 1111 images are used for evaluation. We ensure no overlap in the training pairs and randomly crop the image into 256×256 image blocks in both

TABLE I

ABLATION STUDY ON THE EFFECTIVENESS OF DIFFERENT COMPONENTS IN OUR MODEL. ALL THE RESULTS ARE TESTED ON THE GOPRO DATASET [10]. G_S MEANS THE TRANSLATION FROM THE BLUR DOMAIN TO THE SHARP DOMAIN, AND G_B MEANS THE TRANSLATION FROM THE SHARP DOMAIN TO THE BLUR DOMAIN

With different components	G_S(blur-sharp)		G_B(sharp-blur)	
	PSNR	SSIM	PSNR	SSIM
original CycleGAN method	23.9956	0.8076	24.8028	0.8437
with multi-adversarial structure	25.2630	0.8524	25.4488	0.8618
with edge map input	25.0725	0.8538	25.3111	0.8616
with multi-scale edge constrains	25.4148	0.8530	26.2150	0.8620
with multi-scale SSIM Loss	24.9598	0.8533	26.1492	0.8614
with all above components	**26.2473**	**0.8673**	**26.3428**	**0.8689**

the blur set and the sharp set. The parameter ω_1 is set to 5, parameters ω_2 and ω_3 are set to 0.5, ω_4 is set to 10 and ω_5 is set to 1 in Eq. (15). We use PSNR and SSIM two metrics to show quantitative comparisons with other deblurring algorithms.

C. Ablation Study

To analyze the effectiveness of each important component or loss (perceptual etc.), we perform an ablation study in this section. Both quantitative and qualitative results on the GoPro dataset are presented for the following six variants of our method by adding each component gradually: 1) original CycleGAN method [1]; 2) adding the multi-adversarial structure; 3) adding edge map input component; 4) adding multi-scale edge constraints; 5) adding multi-scale SSIM loss; 6) adding all the above components.

We present the PSNR and SSIM for each variant in Table I. G_S (blur-sharp) means the translation from the blurred domain to the sharp domain, and G_B (sharp-blur) means the translation from the sharp domain to the blurred domain. From Table I, we can see that the multi-adversarial structure significantly improves the deblurring performance because of the multi-resolution constraints. Meanwhile, the structure-aware mechanism (with the edge as input and multi-scale edge constraints) can also preserve the structure and details because of the additional edge information and edge constraints. Even the original CycleGAN basically implements the unsupervised translation from blurred to sharp and from sharp to blurred, it introduces the unpleasant noise information (colors, textures, etc.). In contrast, with adding the multi-adversarial structure, discriminators are able to determine whether the resulting clear image is true or false from multiple resolutions and then feedback to the generators. With the edge map as part of the input, more structure-guided information can be transferred to the target domain. With the multi-scale edge constraints to guide the deblurring process, some unwanted ringing artifacts at the boundary of the generated images can be removed effectively. With the multi-scale SSIM loss, the generated image can preserve the luminance, contrast and structure information effectively. The overall deblurring performance in Table I also shows that there is a close relationship between our multi-adversarial learning and the structure-aware mechanism.

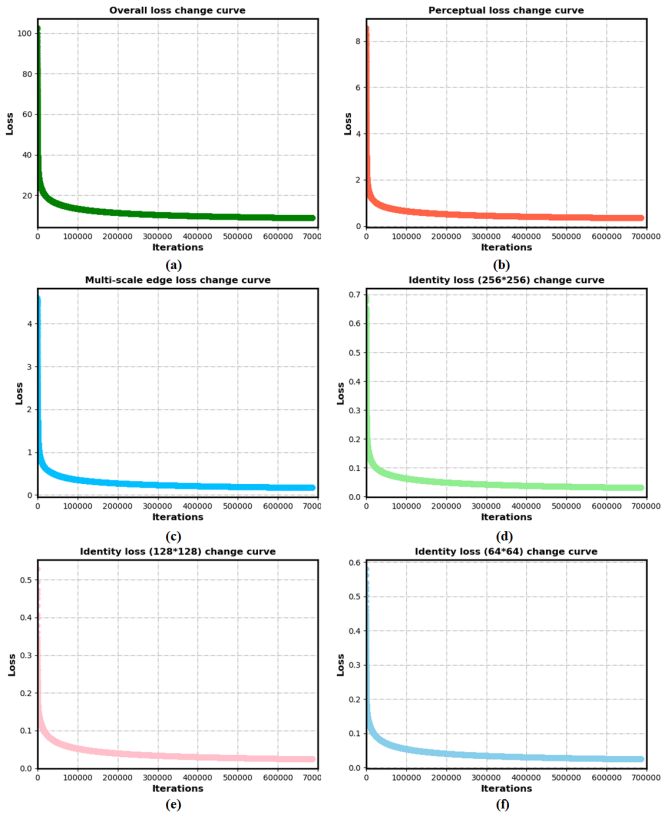


Fig. 5. Stability analysis for our proposed model. (a) The overall loss variation. (b) The perceptual loss variation. (c) The multi-scale edge losses variation of our method at resolution 256×256. (d), (e) and (f) are the identity loss variation at resolution 64×64, 128×128 and 256×256, respectively. (a), (b), (c) and (d) show that different losses of our model can steadily decrease with the increase of iteration times during the training process. (d), (e) and (f) indicate the identity preserving loss of our model decrease steadily with the increase of iteration times at different resolutions.

To illustrate the stability of the proposed model, Fig. 5 shows the different loss change curves of our proposed methods. Fig. 5(a) is the overall loss variation curve. Fig. 5(b) is the perceptual loss variation curve. Fig. 5(c) is the multi-scale edge losses variation of our method at resolution 256×256. Fig. 5(d), Fig. 5(e) and Fig. 5(f) indicate that the identity preserving loss of our model can decrease steadily with the increase of iteration times at different resolutions (64×64, 128×128 and 256×256, respectively). As seen from the change curve of all losses, different types of losses and losses with different resolutions can steadily decline with the increase of iteration times during the training process, which fully indicates that our model is relatively stable.

D. Parameter Sensitivity

As we mentioned in Section III-E, the weight ω_1 for cycle-consistency loss L_{cycle}, ω_4 for identity preserving loss L_{Id}, ω_5 for perceptual loss $L_{Perceptual}$ need to be tuned so that the deblurred image neither stays too close to the original blurred image, nor contains many artifacts. The quantitative performance is shown in Fig. 6. From Fig. 6, we can see that parameter ω_4 setting for L_{Id} is greatly different from the traditional CycleGAN based task (such as for Photo-Sketch). As our method is based on multi-resolution

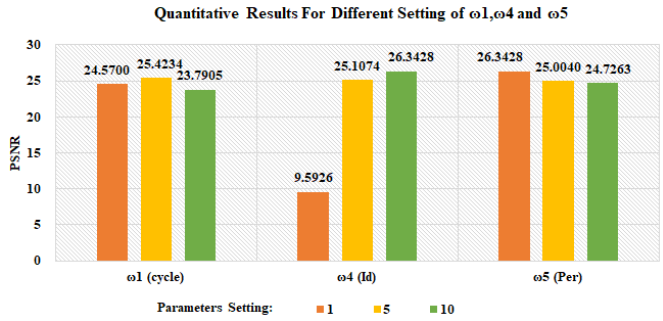


Fig. 6. Quantitative results for different setting of ω_1 for cycle-consistency loss L_{cycle}, ω_4 for identity preserving loss L_{Id}, ω_5 for perceptual loss $L_{Perceptual}$. The orange bar chart represents the average PSNR value on the GoPro test set when parameter ω_1, ω_4 and ω_5 are set to 1, respectively. Correspondingly, the yellow bar represents the average PSNR value on the GoPro test set when parameters ω_1, ω_4 and ω_5 are set to 5, respectively. The green bar represents the average PSNR value on the GoPro test set when represents ω_1, ω_4 and ω_5 are set to 10, respectively. We can see that different parameter settings have a certain influence on the final deblurring effect.

Fig. 7. Visualizations of sample image in GoPro dataset with different sets of ω_5 for perceptual loss $L_{Perceptual}$. As shown in (d), when the ω_5 is set to 0.1, the generated deblurred image is very blurred. As shown in (e) and (f), when the ω_5 is set too high ($\omega_5 = 5$ and $\omega_5 = 10$), vast artifacts will be introduced to cause quality degradation.

adversarial learning, L_{Id} loss has a great impact on the overall deblurring effect, when ω_4 is set to 10, the deblurring effect is the best. For parameter ω_1 is set too high ($\omega_1 = 10$), the deblurred image generated by G_S becomes very blurred and the quantitative performance is poor. In contrast, if the ω_1 is set too low ($\omega_1 = 1$), vast artifacts will be introduced. ω_5 for perceptual loss $L_{Perceptual}$ also has a certain influence on the overall deblurring effect. We set the parameters as $\omega_1 = 5$, $\omega_4 = 10$ and $\omega_5 = 1$ on the GoPro test set. As shown in Fig. 6, many experiments have proved that relatively good results can be obtained when $\omega_5 = 1$. Fig. 7 also shows the visualizations of sample image in GoPro dataset with different setting of ω_5 for perceptual loss $L_{Perceptual}$. From Fig. 7(d), when the ω_5 is set to 0.1, the generated deblurred image is very blurred. In contrast, Fig. 7(e) and Fig. 7(f) show that if the ω_5 is set too high, vast artifacts will be introduced to the generated images, especially in the colored rectangular area. In real experiments, the parameters $\omega_1 \sim \omega_5$ are set according to the data characteristics for different cases.

E. Comparison With State-of-the-arts

1) BMVC_TEXT Dataset [52] and Face Dataset [53]: In order to compare the performance of different algorithms on the text images and face images, we use the same training data (described in Section IV-B) to retrain the CNN-based methods. We randomly select 100 samples from the test set in the BMVC_TEXT dataset and 2000 samples from face dataset [53] (as described in Section IV-B) for evaluation. The quantitative results are presented in Table II. The last column of Table II shows the quality metrics of our deblurred method. From Table II, we could conclude that our method significantly outperforms other state-of-the-art supervised (Pan *et al.* [12], Pan *et al.* [31], Xu *et al.* [29], Sun *et al.* [34], MS-CNN [10] and DeblurGAN [14]) and unsupervised methods (CycleGAN [1], UID-GAN [43] and Madam *et al.* [16]) for text images and face images deblurring. Fig. 8 presents several examples from the BMVC_TEXT dataset [52] to illustrate the qualitative comparisons of other methods with ours. In Fig. 8, especially in the central character part, the deblurring results by our method can achieve the clearest characters. These examples are sufficient to prove that our method can achieve quite effective results on BMVC_TEXT dataset [52].

2) GoPro Dataset: Table III shows the quantitative comparison results with other state-of-the-art deblurring methods on GoPro dataset [10]. The average PSNR and SSIM for image quality assessment show our significant improvement in deblurring effect compared with other popular methods. From Table III we can see that, compared with almost all the classical conventional deblurring algorithms (Xu *et al.* [29], Whyte *et al.* [6] and Kim *et al.* [32]) and the latest unsupervised CNN-based deblurring approaches (CycleGAN [1], DiscoGAN [17], UID-GAN [43], Madam *et al.* [16]), our algorithm shows quite attractive deblurring effect. Meanwhile, compared with most supervised CNN-based deblurring methods (Pix2Pix [47] and Sun *et al.* [34]), we can still achieve relatively satisfactory results. Although our method is slightly inferior to the supervised CNN-based method [10] and DeblurGAN [14] on GoPro, the reason is that it is more difficult to learn unpaired data compared with paired data and CycleGAN itself has performance flaws in handling the generation of high-resolution images. Meanwhile, our method can also achieve better performance on multiple other databases (such as BMVC_TEXT dataset [52] and face dataset [53]). Additionally, methods [10] and [14] require a large amount of paired training data, unlike our unsupervised learning, which can greatly reduce the strong need for paired training data. Fig. 9 shows some visual examples from the GoPro [10] test set. It shows that even in some cases of the GoPro, our approach is as desirable as method [10]. From Fig. 9, it is obvious that the classical conventional deblurring algorithm cannot keep structure information well and most unsupervised

TABLE II

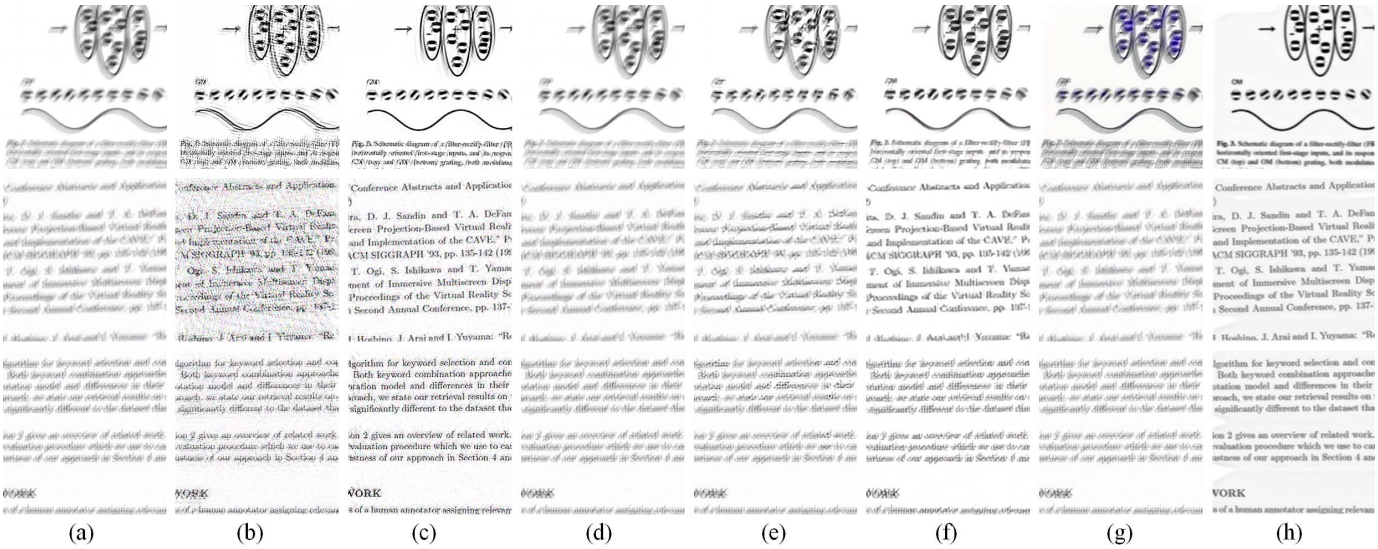
PEAK SIGNAL-TO-NOISE RATIO AND STRUCTURAL SIMILARITY MEASURE, MEAN ON THE BMVC_TEXT [52] AND FACE DATASETS [53]

DateSet	Method	Pan *et al.* [12]	Pan *et al.* [31]	Xu *et al.* [29]	Sun *et al.* [34]	MS-CNN [10]	DeblurGAN [14]	CycleGAN [1]	UID-GAN [43]	Madam *et al.* [16]	Our
Text	PSNR	16.26	17.48	14.26	18.62	18.86	18.69	13.23	18.96	23.22	**23.68**
	SSIM	0.73	0.77	0.54	0.70	0.73	0.74	0.57	0.79	0.86	**0.88**
Face	PSNR	17.34	17.59	16.84	18.26	18.29	18.64	19.40	20.81	20.88	**20.96**
	SSIM	0.52	0.54	0.47	0.55	0.57	0.59	0.56	0.65	0.66	**0.68**

TABLE III

PEAK SIGNAL-TO-NOISE RATIO AND STRUCTURAL SIMILARITY MEASURE, MEAN ON THE GOPRO DATASET [10]

Method	Xu *et al.* [29]	Whyte *et al.* [6]	Kim *et al.* [32]	Sun *et al.* [34]	MS-CNN [10]	DeblurGAN [14]	CycleGAN [1]	DiscoGAN [17]	Pix2Pix [47]	UID-GAN [43]	Madam *et al.* [16]	Our
PSNR	25.184	25.093	23.640	24.689	**28.930**	28.702	25.009	24.827	24.994	25.594	25.787	26.247
SSIM	0.896	0.887	0.824	0.856	0.910	**0.927**	0.851	0.786	0.794	0.851	0.860	0.867



(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 8. Comparison of deblurred images by our method and other popular approaches on some images from BMVC_TEXT dataset [52]. (a) Blurred images. (b) Deblurring results using Pan *et al.* [12]. (c) Deblurring results using Pan *et al.* [31]. (d) Deblurring results using Xu *et al.* [29]. (e) Deblurring results using Sun *et al.* [34]. (f) Deblurring results using MS-CNN [10]. (g) Deblurring results using CycleGAN [1]. (h) Our results. It shows the characters in our results are much clearer.

(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 9. Comparison of deblurred images by our method and other popular approaches on one sample from GoPro Dataset [10]. (a) Blurred image. (b) Deblurring results using Pan *et al.* [12]. (c) Deblurring results using Xu *et al.* [29]. (d) Deblurring results using Sun *et al.* [34]. (e) Deblurring results using MS-CNN [10]. (f) Deblurring results using CycleGAN [1]. (g) Deblurring result using DiscoGAN [17]. (h) Our results. It shows our results are more satisfying, especially in the pink and yellow rectangles.

methods will introduce new artifacts, while our method can better maintain the structure in the areas such as the girl's head flower or arm. We also provide the visual contrast effect on Köhler dataset in Fig. 10, which also verifies our better performance compared with both supervised and unsupervised methods.

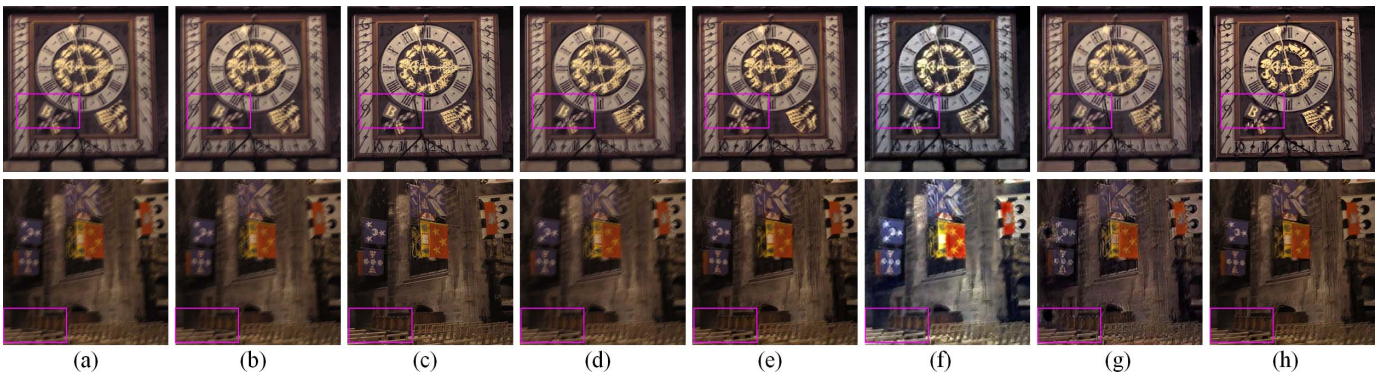


Fig. 10. Comparison of deblurred images by our method and other popular approaches on one sample taken from Köhler Dataset [55]. (a) Blurred image. (b) Deblurring result using Pan *et al.* [12]. (c) Deblurring result using Xu *et al.* [29]. (d) Deblurring result using Sun *et al.* [34]. (e) Deblurring result using MS-CNN [10]. (f) Deblurring result using CycleGAN [1]. (g) Deblurring result using DiscoGAN [17]. (h) Our results. It shows our results are more satisfying, especially in the pink and yellow rectangles.

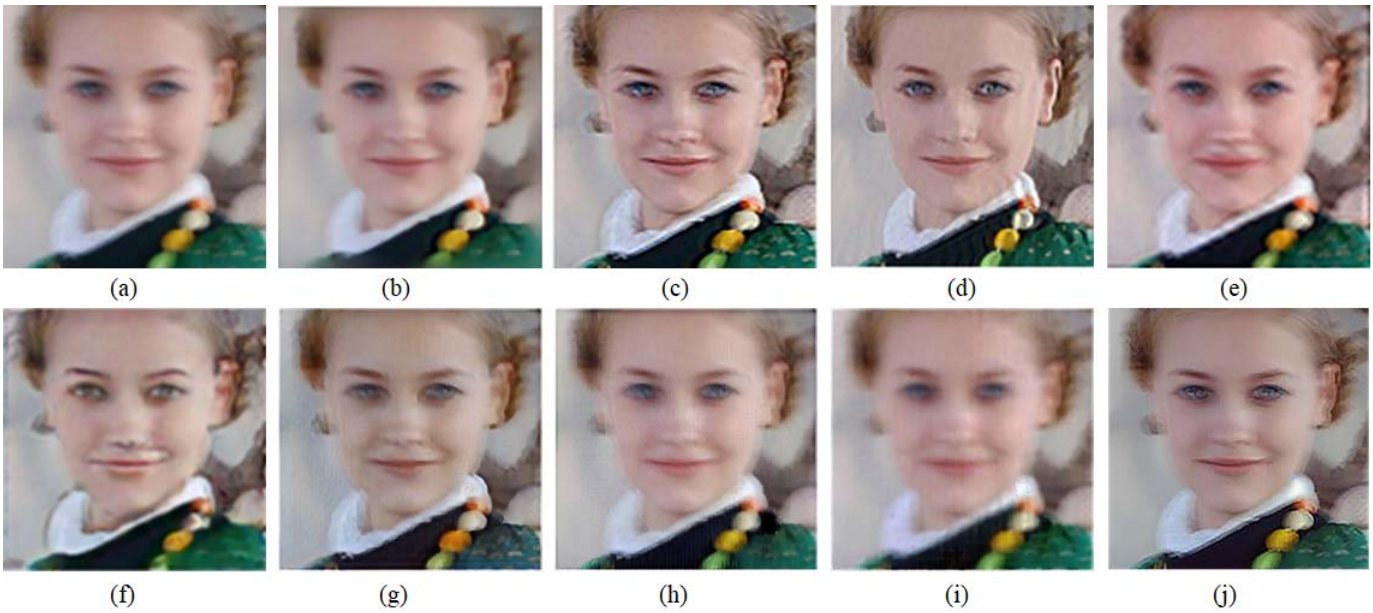


Fig. 11. Comparison of deblurred images by our method and other popular approaches on one real image taken from Lai Dataset [54]. (a) Blurred image. (b) Deblurring result using [31]. (c) Deblurring result using [29]. (d) Deblurring result using [12]. (e) Deblurring result using [34]. (f) Deblurring result using [16]. (g) Deblurring result using CycleGAN [1]. (h) Deblurring result using [17]. (i) Deblurring result using [47]. (j) Deblurring result by our method.

3) Real Dataset: In order to compare the effects of different deblurring algorithms on real blurred images, we use the model trained on GoPro data set to test the real blurred images in the real set of Lai dataset [54]. Since the real blurred images do not provide the corresponding sharp images, it is impossible to evaluate the deblurring effect with the full reference image quality evaluation methods (Such as SSIM and PSNR). Therefore, we compare the deblurring performance of different algorithms in the real blurred images with the help of subjective user analysis. Inspired by [56], we use the Bradley-Terry model to estimate the subjective score. Each blurred image is processed with the deblurring methods Pan *et al.* [12], Xu *et al.* [29], Whyte *et al.* [6], Sun *et al.* [30], MS-CNN [10], CycleGAN [1] and DeblurGAN [14]. We test all these methods with corresponding models trained on GoPro. Together with the original blurred images, all these results are sent for pairwise comparison (22 human raters are involved) to form the winning matrix. The quantitative results in Table IV show that the methods based on CNNs usually have better effect than the convolutional methods, and our method can achieve a more satisfied deblurring effect in real blurred images compared with most existing methods. From Fig. 11, our method shows superior performance compared with other methods, especially in the girl's eyes and mouth.

According to the above experiments, we can conclude that our method has obvious advantages in solving the deblurring task on all the test datasets when comparing with the most existing unsupervised deblurring methods [1], [16], [43]. We can also infer that our unsupervised deblurring method can achieve competitive results with the supervised deblurring algorithm [10], [12], [14], [29] in most datasets except for the GoPro dataset. We believe this is mainly due to CycleGAN's lack of ability to generate high-resolution images and the difficulty for unpaired data learning compared

TABLE IV
AVERAGE SUBJECTIVE EVALUATION SCORES OF DEBLURRING PERFORMANCE ON THE REAL DATASET [54]

Mehtod	Blurry	Xu *et al.* [29]	Whyte *et al.* [6]	Pan *et al.* [12]	Sun *et al.* [30]	MS-CNN [10]	DeblurGAN [14]	CycleGAN [1]	UID-GAN [43]	Madam *et al.* [16]	Our
PSNR	1	0.85	0.64	0.95	0.71	1.10	1.08	0.93	1.13	1.14	**1.18**

TABLE V
THE AVERAGE RUNNING TIME COMPARISONS OF OUR METHOD WITH OTHER SEVERAL CLASSICAL METHODS ON BMVC_TEXT DATASET [52]

Method	Xu *et al.* [29]	Sun *et al.* [34]	MS-CNN [10]	CycleGAN [1]	UID-GAN [43]	Our
Time(s)	377	202	0.18	0.22	0.25	0.29
PSNR	14.26	18.62	18.86	13.63	18.96	23.68

with paired data. Since our deblurring method is based on unsupervised learning and can be trained with finite unpaired training data. Compared with other supervised-based methods, our unsupervised deblurring method has a wider application value.

F. Evaluation of the Running Time

Table V shows the average running time per image comparisons of several classical deblurring methods with 512×512 on the test dataset of BMVC_TEXT dataset [52]. According to Table V, we can see that the proposed unsupervised method achieves the state-of-the-art deblurring quality, while maintains relatively high and competitive speed in comparison to most existing supervised and unsupervised methods on BMVC_TEXT dataset [52]. Even though the time used is slightly longer than CycleGAN [1] and MS-CNN [10] due to the multi-adversarial and multiple constraints structure, we get a better deblurring effect. In future work, we are committed to further streamlining the network and improving its operational efficiency.

V. CONCLUSION AND FUTURE WORK

In this paper, we propose a structure-aware motion deblurring method based on a multi-adversarial optimized CycleGAN model. Unlike previous work, our CycleGAN based method can avoid the error of the kernel estimation and does not need the paired training data to make the training more flexible. In addition, the multi-adversarial constraints in the generator of CycleGAN we used are different from the traditional multi-scale manner to ensure that the results closest to sharpening images are generated at different resolutions. Besides, we introduce a structure-aware method based on edge clues so that the generated deblurred image can keep more structural information as much as possible. Extensive experiments on the different benchmark datasets demonstrate the effectiveness of the method we proposed. In the future, we are committed to solving the problem of significant target deblurring and further reducing the complexity of the network. Besides, we will further explore an unsupervised motion blur method with better performance and apply the proposed network model to the video deblurring problem.

REFERENCES

[1] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2242–2251.

[2] V. Papyan and M. Elad, "Multi-scale patch-based image restoration," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 249–261, Jan. 2016.

[3] M. Temerinac-Ott, O. Ronneberger, P. Ochs, W. Driever, T. Brox, and H. Burkhardt, "Multiview deblurring for 3-D images from light-sheet-based fluorescence microscopy," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1863–1873, Apr. 2012.

[4] A. Danielyan, V. Katkovnik, and K. Egiazarian, "BM3D frames and variational image deblurring," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1715–1728, Apr. 2012.

[5] W. Dong, L. Zhang, G. Shi, and X. Wu, "Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization," *IEEE Trans. Image Process.*, vol. 20, no. 7, pp. 1838–1857, Jul. 2011.

[6] O. Whyte, J. Sivic, A. Zisserman, and J. Ponce, "Non-uniform deblurring for shaken images," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 491–498.

[7] T. M. Nimisha, A. K. Singh, and A. N. Rajagopalan, "Blur-invariant deep learning for blind-deblurring," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4762–4770.

[8] C. J. Schuler, M. Hirsch, S. Harmeling, and B. Scholkopf, "Learning to deblur," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 7, pp. 1439–1451, Jul. 2016.

[9] X. Xu, J. Pan, Y.-J. Zhang, and M.-H. Yang, "Motion blur kernel estimation via deep learning," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 194–205, Jan. 2018.

[10] S. Nah, T. H. Kim, and K. M. Lee, "Deep multi-scale convolutional neural network for dynamic scene deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 257–265.

[11] S. Vasu and A. N. Rajagopalan, "From local to global: Edge profiles to camera motion in blurred images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 558–567.

[12] J. Pan, Z. Hu, Z. Su, and M.-H. Yang, "Deblurring text images via L0-regularized intensity and gradient prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2901–2908.

[13] S. Ramakrishnan, S. Pachori, A. Gangopadhyay, and S. Raman, "Deep generative filter for motion deblurring," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2017, pp. 2993–3000.

[14] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, "DeblurGAN: Blind motion deblurring using conditional adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8183–8192.

[15] X. Tao, H. Gao, X. Shen, J. Wang, and J. Jia, "Scale-recurrent network for deep image deblurring," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8174–8182.

[16] N. T. Madam, S. Kumar, and A. N. Rajagopalan, "Unsupervised class-specific deblurring," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2018, pp. 353–369.

[17] T. Kim, M. Cha, H. Kim, J. K. Lee, and J. Kim, "Learning to discover cross-domain relations with generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 1857–1865.

[18] J. Johnson, A. Alahi, and F.-F. Li, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 694–711.

[19] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. 37th Asilomar Conf. Signals, Syst. Comput.*, vol. 2, 2003, pp. 1398–1402.

[20] Y. Gan, X. Xu, W. Sun, and L. Lin, "Monocular depth estimation with affinity, vertical pooling, and label enhancement," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 232–247.

[21] C. J. Schuler, H. C. Burger, S. Harmeling, and B. Schölkopf, "A machine learning approach for non-blind image deconvolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 1067–1074.

[22] S. Oh and G. Kim, "Robust estimation of motion blur kernel using a piecewise-linear model," *IEEE Trans. Image Process.*, vol. 23, no. 3, pp. 1394–1407, Mar. 2014.

[23] P. Chandramouli, M. Jin, D. Perrone, and P. Favaro, "Plenoptic image motion deblurring," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 1723–1734, Apr. 2018.

[24] Y. Wen, B. Sheng, P. Li, W. Lin, and D. D. Feng, "Deep color guided coarse-to-fine convolutional network cascade for depth image super-resolution," *IEEE Trans. Image Process.*, vol. 28, no. 2, pp. 994–1006, Feb. 2019.

[25] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman, "Removing camera shake from a single photograph," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 787–794, Jul. 2006.

[26] Q. Shan, J. Jia, and A. Agarwala, "High-quality motion deblurring from a single image," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 73:1–73:10, 2008.

[27] L. Xu and J. Jia, "Two-phase kernel estimation for robust motion deblurring," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 157–170.

[28] D. Krishnan, T. Tay, and R. Fergus, "Blind deconvolution using a normalized sparsity measure," in *Proc. CVPR*, Jun. 2011, pp. 233–240.

[29] L. Xu, S. Zheng, and J. Jia, "Unnatural L0 sparse representation for natural image deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 1107–1114.

[30] L. Sun, S. Cho, J. Wang, and J. Hays, "Edge-based blur kernel estimation using patch priors," in *Proc. IEEE Int. Conf. Comput. Photography (ICCP)*, Apr. 2013, pp. 1–8.

[31] J. Pan, D. Sun, H. Pfister, and M.-H. Yang, "Deblurring images via dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 10, pp. 2315–2328, Oct. 2018.

[32] T. H. Kim and K. M. Lee, "Segmentation-free dynamic scene deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2766–2773.

[33] Y. Bai, H. Jia, M. Jiang, X. Liu, X. Xie, and W. Gao, "Single-image blind deblurring using multi-scale latent structure prior," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 7, pp. 2033–2045, Jul. 2020.

[34] J. Sun, W. Cao, Z. Xu, and J. Ponce, "Learning a convolutional neural network for non-uniform motion blur removal," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 769–777.

[35] D. Ren, W. Zuo, D. Zhang, J. Xu, and L. Zhang, "Partial deconvolution with inaccurate blur kernel," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 511–524, Jan. 2018.

[36] D. Gong *et al.*, "From motion blur to motion flow: A deep learning solution for removing heterogeneous motion blur," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3806–3815.

[37] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 105–114.

[38] D. Engin, A. Genç, and H. K. Ekenel, "Cycle-Dehaze: Enhanced CycleGAN for single image Dehazing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 938–946.

[39] L. Xu, J. S. J. Ren, C. Liu, and J. Jia, "Deep convolutional neural network for image deconvolution," in *Proc. Neural Inf. Process. Syst.*, 2014, pp. 1790–1798.

[40] L. Li, J. Pan, W.-S. Lai, C. Gao, N. Sang, and M.-H. Yang, "Dynamic scene deblurring by depth guided model," *IEEE Trans. Image Process.*, vol. 29, pp. 5273–5288, 2020.

[41] B. Lu, J.-C. Chen, and R. Chellappa, "Unsupervised domain-specific deblurring via disentangled representations," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 10217–10226.

[42] Q. Yuan, J. Li, L. Zhang, Z. Wu, and G. Liu, "Blind motion deblurring with cycle generative adversarial networks," *Vis. Comput.*, vol. 36, no. 8, pp. 1591–1601, Aug. 2020.

[43] B. Lu, J.-C. Chen, and R. Chellappa, "UID-GAN: Unsupervised image deblurring via disentangled representations," *IEEE Trans. Biometrics, Behav., Identity Sci.*, vol. 2, no. 1, pp. 26–39, Jan. 2020.

[44] L. Wang, V. Sindagi, and V. Patel, "High-quality facial photo-sketch synthesis using multi-adversarial networks," in *Proc. 13th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, May 2018, pp. 83–90.

[45] Z. Fu, Y. Zheng, H. Ye, Y. Kong, J. Yang, and L. He, "Edge-aware deep image deblurring," *CoRR*, vol. abs/1907.02282, pp. 1–9, Jul. 2019.

[46] Y. Bahat and M. Irani, "Blind Dehazing using internal patch recurrence," in *Proc. IEEE Int. Conf. Comput. Photography (ICCP)*, vol. 8691, May 2016, pp. 783–798.

[47] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5967–5976.

[48] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[49] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent.*, 2015, pp. 1–14.

[50] A. Paszke *et al.*, "Automatic differentiation in PyTorch," in *Proc. Neural Inf. Process. Syst. Workshop*, 2017, pp. 1–4.

[51] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent.*, 2015, pp. 1–15.

[52] M. Hradiš, J. Kotera, P. Zemčík, and F. Šroubek, "Convolutional neural networks for direct text deblurring," in *Proc. Brit. Mach. Vis. Conf.*, 2015, pp. 6:1–6:13.

[53] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 3730–3738.

[54] W.-S. Lai, J.-B. Huang, Z. Hu, N. Ahuja, and M.-H. Yang, "A comparative study for single image blind deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1701–1709.

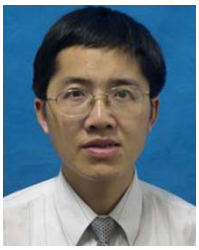
[55] R. Köhler, M. Hirsch, B. Mohler, B. Schölkopf, and S. Harmeling, "Recording and playback of camera shake: Benchmarking blind deconvolution with a real-world database," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 27–40.

[56] O. Kupyn, T. Martyniuk, J. Wu, and Z. Wang, "DeblurGAN-v2: Deblurring (orders-of-magnitude) faster and better," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2019, pp. 8877–8886.

Yang Wen received the M.Eng. degree in computer science from Xidian University, Xi'an, China, in 2015. She is currently pursuing the Ph.D. degree in computer science with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China. Her current research interests include motion deblurring, convolutional neural networks, image/video processing, and computer graphics.

Jie Chen received the B.Eng. degree in computer science from Nanjing University, Nanjing, China. She is currently a Senior Chief Engineer and Senior Architect with Samsung Electronics (China) Research and Development Centre, Nanjing. She is also the Head of the AI Department. Her current research interests include computer vision and big data.

Bin Sheng (Member, IEEE) received the B.A. degree in English and the B.Eng. degree in computer science from the Huazhong University of Science and Technology, Wuhan, China, in 2004, the M.Sc. degree in software engineering from the University of Macau, Taipa, Macau, in 2007, and the Ph.D. degree in computer science and engineering from The Chinese University of Hong Kong, Sha Tin, Hong Kong, in 2011. He is currently a Full Professor with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China. His current research interests include virtual reality and computer graphics. He is an Associate Editor of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY.



Zhihua Chen received the Ph.D. degree in computer science from Shanghai Jiao Tong University, Shanghai, China, in 2006. He is currently a Full Professor with the Department of Computer Science and Engineering, East China University of Science and Technology, Shanghai. His current research interests include image/video processing and computer vision.

Ping Li (Member, IEEE) received the Ph.D. degree in computer science and engineering from The Chinese University of Hong Kong, Sha Tin, Hong Kong, in 2013. He is currently a Research Assistant Professor with The Hong Kong Polytechnic University, Kowloon, Hong Kong. He has one image/video processing national invention patent, and has excellent research project reported worldwide by *ACM TechNews*. His current research interests include image/video stylization, colorization, artistic rendering and synthesis, and creative media.

Ping Tan (Senior Member, IEEE) received the Ph.D. degree in computer science and engineering from The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, in 2007. He is currently an Associate Professor with the School of Computing Science, Simon Fraser University, Burnaby, BC, Canada. His current research interests include computer vision and computer graphics. He has served as an Area Chair for IEEE CVPR, ACM SIGGRAPH, and ACM SIGGRAPH Asia. He has served as an Editorial Board Member of IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE and the *International Journal of Computer Vision*.

Tong-Yee Lee (Senior Member, IEEE) received the Ph.D. degree in computer engineering from Washington State University, Pullman, in May 1995. He is currently a Chair Professor with the Department of Computer Science and Information Engineering, National Cheng Kung University, Tainan, Taiwan. He leads the Computer Graphics Group, Visual System Laboratory, National Cheng Kung University (http://graphics.csie.ncku.edu.tw). His current research interests include computer graphics, non-photorealistic rendering, medical visualization, virtual reality, and media resizing. He is a member of the ACM.