WILEY | Hindawi

Research Article

Structures Guided Dynamic Scene Deblurring Method

Qing Qi [ID]

Department of Physics and Electronics Information Engineering, Qinghai Minzu University, Xining, China

Correspondence should be addressed to Qing Qi; qiqing@tju.edu.cn

In this paper, we integrate image gradient priors into a generative adversarial networks (GANs) to deal with the dynamic scene deblurring task. Even though image deblurring has progressed significantly, the deep learning-based methods rarely take advantage of image gradients priors. Image gradient priors regularize the image recovery process and serve as a quantitative evaluation metric for evaluating the quality of deblurred images. In contrast to previous methods, the proposed model utilizes a data-driven way to learn image gradients. Under the guidance of image gradient priors, we permeate it throughout the design of network structures and target loss functions. For the network architecture, we develop a GradientNet to compute image gradients via horizontal and vertical directions in parallel rather than adopt traditional edge detection operators. For the loss functions, we propose target loss functions to constrain the network training. The proposed image deblurring strategy discards the tedious steps of solving optimization equations and taking further advantage of learning massive data features through deep learning. Extensive experiments on synthetic datasets and real-world images demonstrate that our model outperforms state-of-the-art (SOAT) methods.

1. Introduction

In the field of computer vision, image deblurring is a crucial and challenging task. Object motion, camera shake, and other complicated circumstances invariably result in blurry observations during the image acquisition process. Clean and fine image details are required for postprocessing applications such as traffic surveillance, object recognition, and image segmentation. Therefore, the importance of image deblurring is particularly prominent. Blind image deblurring aims at recovering deblurred images from a known blurry degraded image. The mathematical expression of the image deblurring modeling process can be expressed as

$$b = s * k + n, \tag{1}$$

where s, k, n, and b denote clean images, kernels, noise, and blurry images, respectively.

According to Equation (1), we can derive that estimating kernels and computing clean images is a tough task. In traditional image deblurring methods [1–7], the blurry kernel is usually determined by estimating significant structures of the image, and then nonblind deconvolution is computed to obtain the deblurred image. However, these methods have the following limitations: (1) these methods can only extract features from a limited number of images. It may not stimulate realistic dynamic blurry scene; (2) these methods obtain deblurred images by computing optimization equations, which is time-consuming, and the algorithm is less real-time.

With the development of deep learning, convolutional neural networks (CNNs) are exploited in methods [8–11] to learn blur kernels at the pixel level. Subsequently, nonblind deconvolution operations are employed to generate latent images. Although this strategy combines classic and learning-based methods, accumulated errors are derived by blending separately estimated kernels and deblurred images. To overcome the limitations of the isolated image deblurring strategy of "kernel estimation-deconvolution," methods [12–14] followed an end-to-end manner to directly investigate the underlying relationship between blurry and deblurred images. The multiscale deblurring methods [15, 16] followed the image deblurring strategy of "coarse-to-fine" to achieve deblurring. However, it has some limitations: first, at some

scales, the multiscale network structure tends to be overfitting. Second, multiscale networks focus on the relationships between scales rather than the relationship between the original input and the clean counterpart. Although deep learning-based methods alleviate the limitations of handcraft feature extraction methods, the lack of image prior guidance makes the network difficult to optimize and converge.

According to the mechanism of the human visual system, human eyes are the most sensitive to image structures compared to other components [17]. In other words, the human eye is the most intuitive in observing whether the image has clean and significant structures. On the one hand, structures of the image can be exploited as a priori for regularizing the image recovery process. On the other hand, image structures can be served as a quantitative evaluation metric for evaluating the quality of deblurred images. Few works focus on integrating image structure priors to CNN-based methods. Qi et al. [18] designed an edge adversarial mechanism for tackling dynamic scene deblurring. However, the processing of adaptively learning image structures is neglected.

Inspired by the image super-resolution method of [19], we propose an image deblurring method based on image gradients priors for tacking the dynamic scene deblurring task. Meanwhile, image gradient priors permeate the design of network structure and target loss function. For the loss functions, we propose objective loss functions aim to supervise the generator such that generated images have significant structure information. For the network topology, we propose a subnetwork of GradientNet. First, we introduce a recurrent gradient convolutional layer (RGCL) to localize and represent image gradient features rather than a conventional edge detector. Second, multibranch reuse blocks (MBRBs) are proposed to learn high-dimension local features in a multipath reuse way; third, recalibrated channels of these accumulated and enhanced nonlocal features are highlighted by the non-local SENet module [20]. Finally, we arrange several MBRBs in a cascaded manner to enhance and aggregate the relationship of structural features. Extensive experiments on synthetic datasets and real-world images confirm the effectiveness of the proposed model, which achieves decent performance and is comparable to or better than SOTA methods.

The main contributions of this paper are presented as follows:

First, we propose a GAN, which has decent performance in restoring clean images for tackling the dynamic scene deblurring task.

Second, we introduce a GradientNet to compute image gradients via horizontal and vertical directions in parallel.

Third, we develop multiterm target loss functions to drive the generator to generate images with salient structures.

2. Related Works

In recent years, many image deblurring methods have been proposed. We mainly introduce image deblurring methods

from aspects of traditional segmentation-based methods and learning-based methods.

2.1. Traditional Segmentation-Based Methods. Dynamic blurry scenes are spatially varied in pixel level, and solutions for the uniform blurry task may not fit the complicated dynamic scene task. According to blurry regions in a dynamic scene, Kim et al. [21] adopted an image segmentation method to separate blurry regions and then deal with each of them, respectively. However, motion segmentation cannot accurately separate blurry regions. Kim and Lee [22] proposed an alternative segmentation-free method by exploiting the deblurring strategy to solve this challenging task. This scheme avoids drawbacks brought by inaccurate segmentation. However, the dynamic blurry scene is spatially varied in pixel level in extreme cases. Especially, when object motion and camera shake simultaneously occur in the imaging process. As a compromise, Pan et al. [23] proposed a method based on the segmentation confidence map for enhancing the segmentation accuracy of different regions of the degraded images.

2.2. Learning-Based Methods. CNNs are commonly used in image processing fields [24–28], due to their extremely robust feature extraction capabilities. Schuler et al. [11] and Xu et al. [29] employed multiple CNNs to implement image features extraction, blur kernels estimation, and deblurred image reconstruction separately. Chakrabarti [10], Sun et al. [8], and Gong et al. [9] exploited CNNs to estimate nonuniform blur kernels; then, they used existing nonblind deconvolution methods to generate deblurred images, following the image deblurring strategy that involves kernel estimation and nonblind deconvolution algorithms. Li et al. [30] and Ren et al. [31] adopted image priors with the benefits of CNN to achieve image deblurring. These methods make use of deconvolution algorithms and CNNs. Nonetheless, this isolated deblurring strategy causes cumulative errors, resulting in blurred details in the recovered images.

To overcome the limitations of the above algorithms, researchers further propose deblurring methods [12–14] to directly construct the essential relationship between blurry and clean images in an end-to-end manner. Nah et al. [15] followed the "coarse-to-fine" image deblurring strategy and propose a multiscale deblurring method. Although the multiscale network reduces the difficulty of image deblurring in a divide-and-conquer manner, the weight parameters of multiple scales are independent. And each scale of the network only deals with the image of the current resolution. Based on the literature [15], Tao et al. [16] proposed a multiscale CNN with shared parameters. On the one hand, this network takes advantage of the dependence of weight parameters between multiscales; on the other hand, it can reduce network parameters and stabilize network optimization. To acquire high-dimensional feature representations, Gao et al. [32] proposed an image deblurring method based on selective parameter sharing and nested connections, which can effectively extract high-order nonlinear features. Furthermore, Zhang et al. [33] introduced an image

deblurring method based on a recurrent neural network to learn the high-dimensional features by indirectly expanding receptive fields for image reconstruction and deblurring. However, image deblurring methods based on CNN do not consider the semantic information between blurry images and clean images.

GAN [34] is a machine learning architecture proposed by Goodfellow et al. in 2014. GAN has been applied in the computer vision community. Inspired by CycleGAN [35], image deblurring can be considered as an image translation task by translating blurry degraded input to the blurry-free one. Since the highly unstable property of GAN, it is difficult to simultaneously train two pairs of GAN models. Furthermore, directly transferring this recycling framework to image deblurring task unsurprisingly generates poor results. Nimisha et al. [36] proposed a GAN for tackling class-specific image deblurring task in an unsupervised fashion. Due to a lack of ground truths, they utilize blurry images themselves to guide the network to acquire image color information. Kupyn et al. [37] developed a conditional GAN named DeblurGAN. They propose a content loss to capture the semantic correspondence difference between blurry images and corresponding ground truths. Lately, to satisfy the different requirements of real-time processing and deblurring performance, Kupyn et al. [38] proposed three models with feature pyramid network architecture. Qi et al. [18] introduced a method based on an edge adversarial mechanism to narrow the difference in image structure between deblurred images and ground truths. To further facilitate the quality of deblurred image structures, we propose an image gradient-driven dynamic scene deblurring method to constrain image recovery. Specifically, we design a GradientNet to investigate the relationship between image gradients and dynamic scene deblurring in a data-driven way rather than adopt classic image edge detectors.

3. Proposed Method

In this section, we introduce specific illustrations of the overall network architecture, GradientNet, and target loss functions.

3.1. Network Architecture. We tailor a deep learning framework specifically for the challenging dynamic scene deblurring task. Figure 1 depicts the overall pipeline of the proposed network. The generator is designed to generate deblurred images with clean appearances and salient structures in an end-to-end fashion, while the discriminator is designed to assign correct labels to the fake and real images. Furthermore, the discriminator supervises the generator to ensure that it produces images infinitely close to being clean.

3.1.1. Generator. As shown in Figure 2, the generator is employed to map deteriorated blurry images to their clean counterparts. At the stage of the encoder, the blurry degraded inputs are spatially compressed and encoded. At the bottle of the U-Net, we introduce a GradientNet to investigate and preserve significant structures in a multi-branch reuse fashion. Correspondingly, at the stage of the

decoder, decoded feature representations of blurry images are accessible for recovering deblurred images. Specific network framework and parameters configuration of the developed model are marked in Figure 2. Furthermore, skip connections are exploited to bridge the semantic gap between encoded features and their decoded counterparts. Following that, we introduce the backbones of GradientNet in detail.

3.1.2. GradientNet. In this paper, we propose a GradientNet to investigate the relationship between gradient representations and dynamic scene deblurring in a data-driven way. As shown in Figure 2, GradientNet consists of a prepositional feature transition module (PRFTM) and a multi-branch reuse unit (MBRU). Next, we introduce PRFTM and MBRU in detail, respectively.

(1) PRFTM. PRFTM is employed to extract higher-dimensional features by acting as partitions and buffers. In particular, PRFTM consists of two cascaded convolutional layers with kernels of 3×3. The following is the mathematical expression:

$$U_{-1} = f_{FE1}(U_{n-1}), \tag{2}$$

$$U_0 = f_{FE2}(U_{-1}), \tag{3}$$

where f_{FE1} and f_{FE2} denote the first and second convolutional layer in PRFTM, respectively. U_{n-1} implicates the input feature that delivers to PRFTM. U_{-1} represents the output feature from f_{FE1}. U_0 implies the output feature from f_{FE2}. Convolutional layers increase the number of channels of the input image and maps it to a space with more potential features.

(2) MBRU. As shown in Figure 2, MBRU consists of 10 MBRBs and a CA module [20]. Features of an image are usually obtained from local receptive fields that reflect spatial relationships within the image domain. For the dynamic scene deblurring task, it is essential to capture features that represent the overall data distributions of images. Therefore, we first learn high-order local features from MBRBs; then, recalibrated channels of these accumulated and enhanced nonlocal features are highlighted by the CA module. Indirectly, the proposed MBRU constructs the interdependencies between pixels and channels. Assuming there are n MBRBs, the n-th output U_n can be expressed as

$$U_n = MBRBs_n(U_0), \tag{4}$$

where U_0 denotes input features from PRFTM, and U_n implicates output features processed by MBRU.

(3) MBRBs. It is essential to tightly correlate the learned features among the proposed network for addressing the image deblurring task. Therefore, we adopt the multipath reuse manner to enhance gradient features obtained by RGCL.
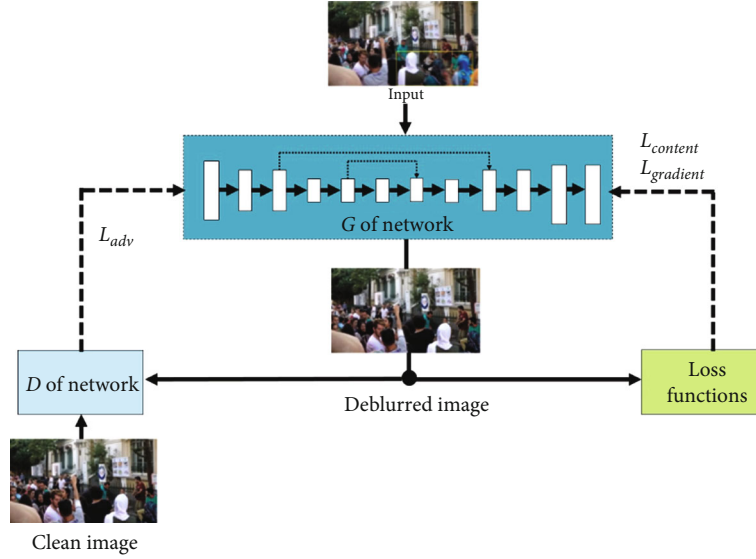
FIGURE 1: The entire framework of the proposed model comprises target loss functions and the network training process. The generator aims to generate deblurred images with clear appearances and discernible structures, while the discriminator is designed to assign correct labels to the fake and real images. Solid lines represent forward propagations, while dashed lines implicate back propagations.
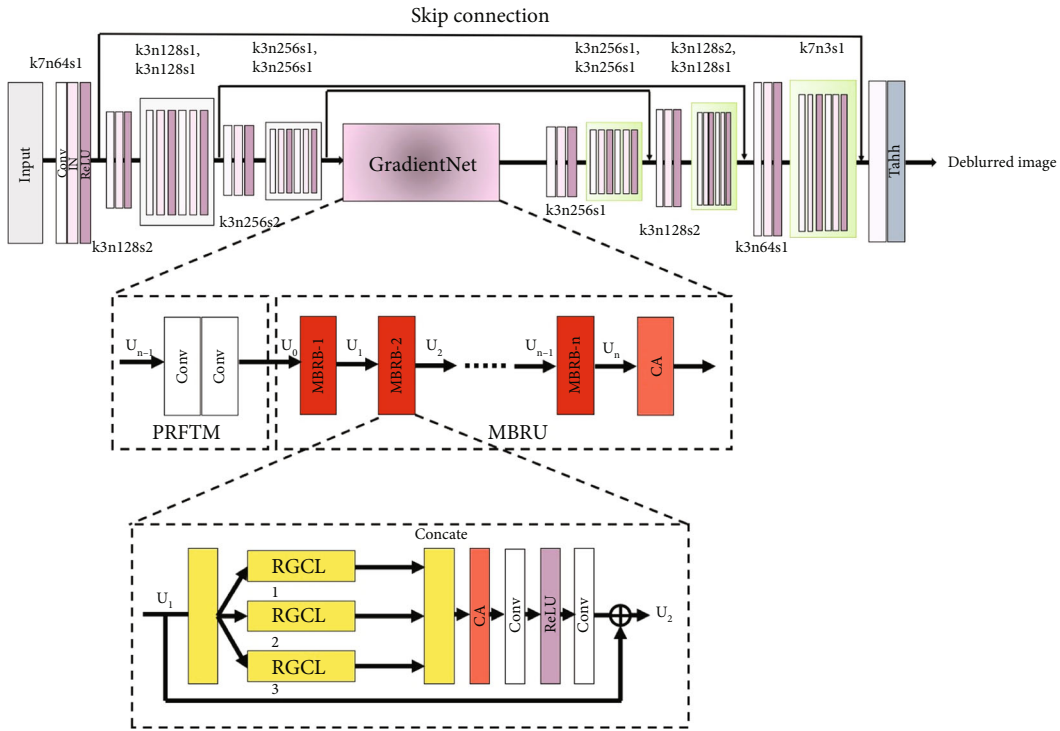

FIGURE 2: The overall structure of the generator in the proposed network, where IN standards for instance normalization, k implies kernel size, n expresses the number of feature maps, s indicates stride, Tanh represents Tanh function, and CA denotes channel attention module.

Reasons for not adopting a multiscale structure are (1) multiscale methods of [15, 16] follow the image deblurring strategy of "coarse-to-fine" to generate deblurred image gradually. Unlike [15, 16], we capture image gradient features in a multipath reuse manner simultaneously; (2) for blurry images, the images at multiple scales have different degrees of blurry. When the image features at different scales are fused, it is easy to produce blurry appearances; (3) this is not conducive to achieve gradient features enhancement. The specific network architecture of MBRBs is displayed in Figure 2. Each MBRB contains following steps: first, we arrange three RGCL in multipath reuse mode; second, these reused features are fused by a concatenate operation; third, the CA module is also employed to recalibrate essential channel features; fourth, two convolutional layers and ReLU activation function [39] are followed as a buffer to obtain

processed features; finally, building a residual connection between the input and output. The mathematical expression of MBRB can be formulated as

$$f_n = \text{Con}((\text{RGCL}_1(U_0) + \text{RGCL}_2(U_0) + \text{RGCL}_3(U_0))),\tag{5}$$

$$f_n = f_{\text{FE4}}(\text{ReLU}(f_{\text{FE3}}(\text{CA}(f_n)))) + U_0,\tag{6}$$

where RGCL denotes recurrent gradient convolutional layer, Con implicates a concatenate operation, f_{FE3} and f_{FE4} denote two convolutional layers in MBRB, and f_n implies processed features.

(4) RGCL. Classic edge detectors, such as Sobel and Laplace, can only calculate edges of a certain intensity. Given the above limitations, we introduce a RGCL that deals with image gradients in horizontal and vertical directions in parallel, as shown in Figure 3. Compared to plain convolutional layers, the idea of this rectangle convolution kernel design is based on the following: first, reducing network parameters and algorithmic complexity of the proposed network; second, dedicating to dealing with a certain dimension (vertical or horizontal) structures information. The parallel investigation can conserve more information than sequential design. In general, RGCL has the following advantages: (1) RGCL emphasizes image structures information by using the vertical and horizontal gradient information in parallel; (2) kernels of plain convolutional layers with the same receptive field have more computing complexity and parameters; (3) the recurrent mechanism is adopted to continuously enhance significant image structures obtained by gradient convolutional layers.

In order to implement the "parallel" processing strategy, inspired by cascaded several convolutional layers, we reconsider the traditional convolutional layer as two convolutional layers with asymmetrical kernels in horizontal and vertical directions, as shown in Figure 3. We specifically develop asymmetrical convolutional kernels with sizes of 3×1 and 1×3 rather than the generally square kernels. Furthermore, the recurrent number of gradient convolutional layers is configured as 3. The mathematical expression of GCL is as follows:

$$U_{\text{RGCL}} = \text{CA}\big(\text{Sigmoid}\big(\text{Conv}\big(\text{LR}(k_{1 \times n} \otimes U_0 + k_{n \times 1} \otimes U_0)_r\big)\big)\big) + \text{Input},\tag{7}$$

where k implies convolution kernels, n standards for kernel sizes, $\otimes$ denotes convolutional operations, r denotes the recurrent times, and U_{RGCL} expresses gradients information processed by RGCL.

Figure 4 depicts a sample of structural textures obtained using various filters. Compared to the Sobel operator, RGCL identifies edges in both vertical and horizontal directions. The results of Sobel and RGCL are shown in Figures 4(b) and 4(e), respectively. Gradient maps obtained by the filters
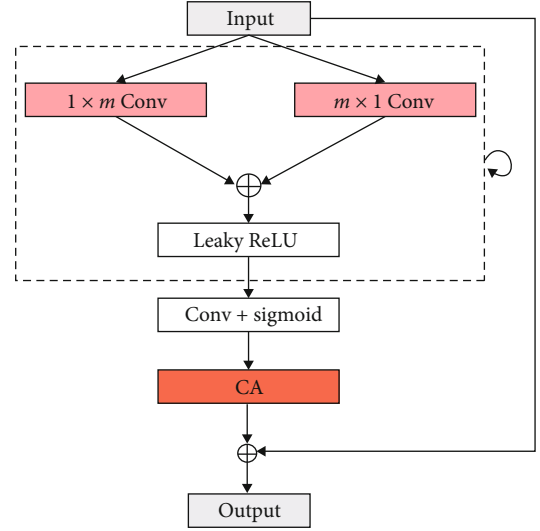

FIGURE 3: The pipeline of the proposed RGCL. Unlike normal convolution, RGCL aims at processing images parallel in vertical and horizontal directions by the rectangular kernel size.

of Sobel are rough outlines. In contrast, our method is able to extract and preserve significant structural features of the image, thus validating the possibility of the proposed RGCL to successfully investigate image structures. In particular, in Figures 4(c) and 4(d), we compare the produced edge maps with horizontal and vertical filters demonstrating that RGCL may extract more edge information than other edge detect operators.

3.1.3. Discriminator. The discriminator receives a generated image or a real image as input. The discriminator returns a probability in the range of [0,1] that indicates how real the input image is. Unlike the high-level task, deblurred image clarity is heavily influenced by local attributes rather than overall assessment. As a result, we use PatchGAN [40] as the discriminator in this paper.

3.2. Loss Functions. Throughout the design of network structures and target loss functions, we use image gradient priors. The proposed network's objective loss functions include (1) the semantic content loss L_{content}, which drives the image semantic coherence between generated images and clean ones; (2) the structure loss L_{gradient}, which facilitates generated images to retain salient structures; and (3) the adversarial loss L_{adv}, which supervises generated results to toward sharp counterparts.

3.2.1. Content Loss Function. Johnson et al. [41] propose a VGG19-based [42] semantic loss that has been pretrained on the ImageNet dataset to meet the visual perception of human eyes. For network optimization, we apply content loss [41] to drive perceptual similarity between generated images and corresponding ground truths at the feature level. Accordingly, the following is the mathematical expression of L_{content}:

(a) Input

(b) Sobel
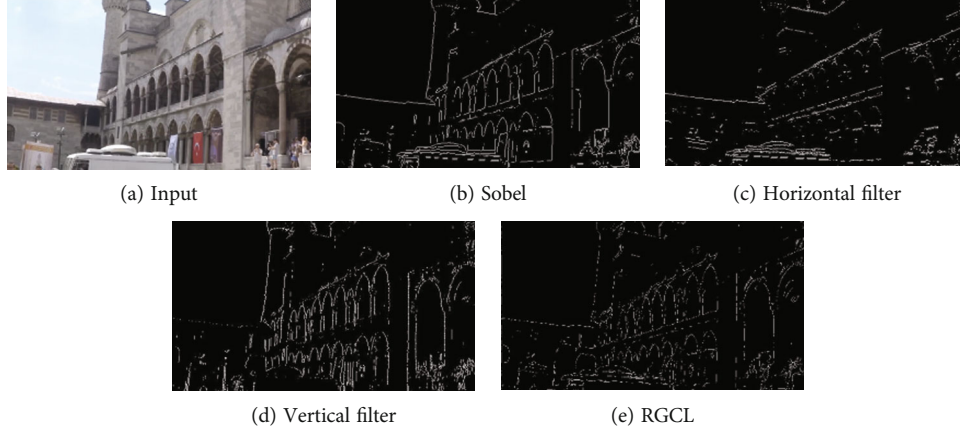
(c) Horizontal filter

(d) Vertical filter

(e) RGCL

FIGURE 4: Gradient maps are visualized. Gradient maps of input have distinct structures, whereas gradient maps obtained by filters of Sobel are rough outline. Our method is capable of regenerating gradient maps with distinct outlines.

$$L_{\text{content}} = \frac{1}{CWH} \sum_{x=1}^{W} \sum_{y=1}^{H} \left\| \phi_{i,j}(s_k)_{x,y} - \phi_{i,j}(G(b_k))_{x,y} \right\|_2, \quad (8)$$

where s_k represents the ground truth, $G(b_k)$ denotes the generated image, C, W, and H imply the number, height, and width of feature maps, respectively. Following [41], we restrict the content difference of s_k and $G(b_k)$ via $\left\| \phi_{i,j}(s_k)_{x,y} - \phi_{i,j}(G(b_k))_{x,y} \right\|_2$. Let $\phi_{i,j}$ be the feature map obtained by the j-th convolutional after activation and before the i-th pooling layer within the pretrained VGG19 network ϕ. Here, we experimentally select the "ReLU4-3" layer in the pretrained VGG19 model to extract semantic feature representations of generated images and sharp ones, respectively.

3.2.2. Gradient Loss Function. We introduce a gradient loss function to supervise the generator that learns image structure information because gradient maps are capable of reflecting salient structures of images. We adopt a gradient loss to narrow the gap between blurry images and ground truths in horizontal and vertical gradient directions. Accordingly, the following is the mathematical expression of L_{gradient}:

$$L_{\text{gradient}} = \frac{1}{WH} \sum_{x=1}^{W} \sum_{y=1}^{H} \left[\left\| \nabla_h(s_k)_{x,y} - \nabla_h(G(b_k))_{x,y} \right\| \right]$$
$$+ \left[\left\| \nabla_v(s_k)_{x,y} - \nabla_v(G(b_k))_{x,y} \right\| \right], \quad (9)$$

where ∇_h and ∇_v represent gradient operations in the horizontal and vertical directions, respectively. We supervise the structure difference of s_k and $G(b_k)$ along the horizontal direction and vertical direction via $\left\| \nabla_h(s_k)_{x,y} - \nabla_h(G(b_k))_{x,y} \right\|_1$ and $\left\| \nabla_v(s_k)_{x,y} - \nabla_v(G(b_k))_{x,y} \right\|_1$.

3.2.3. Adversarial Loss. The purpose of adversarial loss optimization is to lead image feature maps from the blurry image domain B to the clean image domain S. We also employ WGAN-GP [43] as a network optimization critique

function. L_{adv} can be expressed mathematically as follows of L_{adv}:

$$L_{\text{adv}} = E_{G(b_k) \sim P_{\text{data}}(b)}[D(G(b_k))] - E_{s_k \sim P_{\text{data}}(s)}[D(s_k)]$$
$$- \lambda E_{\hat{x} \sim P_{\hat{x}}} \left[\left(\left\| \nabla_{\hat{x}} D(\hat{x}) \right\|^p \right) \right], \quad (10)$$

where $\lambda E_{\hat{x} \sim P_{\hat{x}}}[(\left\| \nabla \hat{x} D(\hat{x}) \right\|_2 - 1)^2]$ represents a sample uniformly sampled on the line between s_k and b_k, $G(b_k)$ signifies the generated image, s_k expresses the clean one, $E_{G(b_k) \sim P_{\text{data}}(b)}[D(G(b_k))]$ and $E_{s_k \sim P_{\text{data}}(s)}[D(s_k)]$ denote the expectation of D allocate the correct label to $G(b_k)$ and s_k, respectively.

Finally, we use multiterm loss functions to optimize the proposed model from the perspectives of L_{content}, L_{gradient}, and L_{adv}. The following is the definition of our network's overall loss function:

$$L(G, D) = \alpha L_{\text{content}} + \beta L_{\text{gradient}} + \gamma L_{\text{adv}}, \quad (11)$$

we adopt the following weight coefficients for each constraint item: $\alpha = 10$, $\beta = 1$, and $\gamma = 1$, respectively.

4. Experiments

In this part, we first introduce datasets and implementation details. The results of the proposed method are compared with SOTA methods subjectively and objectively on synthetic and real datasets. Then, we perform user study and computational complexity. Finally, we conduct ablation investigations to evaluate the relationship between image deblurring performance and loss functions as well as building blocks.

4.1. Datasets. In this paper, standard benchmarks of GOPRO [15], Köhler et al. [44], and Lai et al. [45] are utilized to train and test the proposed network. To train the network model, we select the training dataset of GOPRO [15]. The datasets of Köhler et al. [44] and Lai et al. [45] are adopted as test datasets to evaluate the deblurring performance of the

FIGURE 5: The blurry sample images of GOPRO [15].

FIGURE 6: The blurry sample images of [44].

FIGURE 7: The blurry sample images of the real-world subdataset of Lai [45].

proposed model. Next, we further present these datasets as follows:

4.1.1. GOPRO. The GOPRO dataset was introduced by Nah et al. in 2017 as a standard dataset for training neural networks for image deblurring. A GoPro4 Hero Black camera is used to shoot the video sequence of the dataset, and blurry images are processed by averaging clear frames. The GOPRO collection has 3214 pairs of blurry and clear images with a total resolution of 1280×720 pixels. The 2103 pairs of images are employed for training, and the remaining 1111 pairs of images are utilized for testing. An example of the dataset of GOPRO is shown in Figure 5.

4.1.2. Köhler. Köhler et al. [44] put up an experimental environment for recording and sampling the six-dimensional camera's motion trajectory to capture a series of clear images. Köhler consists of four clean images that correspond to 12 blurry images of a different blur. The dataset contains a total of 48 blurry images. An example of the dataset of Köhler is shown in Figure 6.

4.1.3. Lai. In 2016, Lai et al. presented a dataset [45] for evaluating image deblurring algorithms. The dataset contains two synthetic datasets and one real dataset. The dataset includes natural images, text images, face images, and other

images. An example of the dataset of Lai is shown in Figure 7.

4.2. Training Strategy and Implementation Details. This section delves into the training process for the entire framework. We use the GOPRO dataset to train the proposed model in order to obtain deblurred images in a data-driven manner. The generator is fed blurry inputs that have been randomly cropped to 256×256 pixel sizes for training. To optimize the network, target loss functions reduce the distance between the learned picture and the label image. The discriminator receives the deblurred image or a real image at random, and it generates a probability in the range of [0, 1] that shows how real the input image is. The discriminator is used to supervise the training of the generator, which is used to learn deblurred images with distinct structures and clean appearances.

The experimental software configurations for our network are as follows: the operating system is Ubuntu 14.04, and the deep learning framework is Pytorch. The graphics card is an NVIDIA 1080Ti, and the processor is an Intel(R) Core(TM) I7 (16 GB RAM). The network is trained using 150 epochs in total. We set the generator and discriminator learning rates to 0.0001 and the batch size to 4. The Adam optimizer [46] is used for optimization training, with the parameters $\beta_1 = 0.5$ and $\beta_2 = 0.999$. During training, the generator is updated once, and the discriminator is upgraded

| (a) Blurry input | (b) Qi et al. [18] | (c) Gong et al. [9] | (d) Nah et al. [15] |

| (e) Kupyn et al. [37] | (f) Mustaniemi et al. [47] | (g) Kupyn et al. [38] | (h) Ours |

FIGURE 8: The visual effect of comparison methods and the proposed model on the synthetic dataset of GOPRO [15]. Our method has satisfactory visual effects and good details.

5 times. When the network has converged, original blurry inputs are conveyed to the pretrained generator to obtain deblurred images.

4.3. Experiments on Synthetic Data. Subjective and objective comparison studies using simulated blurry datasets are used to validate the proposed network's efficiency and effectiveness. Several typical learning-based methods are selected for comparing image deblurring performance on the test datasets of GOPRO [15] and Köhler et al. [44]. In the conventional maximum a posteriori framework, Gong et al. [9] adopted a data-driven method to replace the kernel estimate operator. Nah et al. [15] is a multiscale CNN in the traditional sense. GAN-based methods of [37, 38] are also considered. In addition, the image deblurring method based on the edge adversarial mechanism [18] is also compared. To be fair, we recurrent these methods by conducting their official implementations with default settings and parameters. Deblurred results on these synthetic test datasets are shown in Figures 8–10.

Figure 8 displays some deblurred outcomes as a result of severe camera shake. From degraded observations with substantial blur, the developed network is capable of recovering clean appearances and significant structures. CNN is used by Gong et al. [9] to estimate blur kernels in an end-to-end manner. Due to the operator's erroneous kernel estimation, the outcomes are indiscriminately blurry. Even though receptive fields are enlarged on 120 convolution layers, the method of [15] is disabled to solve blur caused by significant

camera motion. The nonuniform image deblurring difficulty is insurmountable for GAN-based methods [37, 38]. Using only a semantic content loss function, for example [37], the image deblurring method may not be penalized.

Figure 9 depicts a case caused by object motion. Methods [9, 37, 38, 47] fail to recover blurry local areas even though the fraction of blurry regions in the overall image is minimal. The deblurred image generated by the proposed method contains distinct structures as shown in Figure 9(h).

Figure 10 shows some deblurred results from complex blurring settings involving considerable camera shake and object motion. Methods [9] do not yield good outcomes. Although stacking numerous convolutional layers can increase receptive fields, Nah et al. [15] produce underlying results without sharp edges, such as person motion outlines. Because content loss functions can only attempt to capture semantic correspondences between high-level representations, semantic content loss functions may not be able to deal with complex non-uniform scenarios. Therefore, methods of [37, 38, 47] are unable to restore substantial edges and fine details. To some extent, the results of methods [18] can recover image appearances. However, the background and pedestrian movement trajectory are not clean enough. The proposed network has a robust capacity for producing deblurred images with conspicuous structure and fine features when compared to SOTA methods.

We conduct several experiments on the synthetic datasets of Köhler et al. [44] to further demonstrate the generalization of the proposed network. Figure 11 shows some

(a) Blurry input (b) Qi et al. [18] (c) Gong et al. [9] (d) Nah et al. [15]

(e) Kupyn et al. [37] (f) Mustaniemi et al. [47] (g) Kupyn et al. [38] (h) Ours

FIGURE 9: The visual effect of comparison methods and the proposed model on the synthetic dataset of GOPRO [15]. Our method has satisfactory visual effects and good details.

(a) Blurry input (b) Qi et al. [18] (c) Gong et al. [9] (d) Nah et al. [15]

(e) Kupyn et al. [37] (f) Mustaniemi et al. [47] (g) Kupyn et al. [38] (h) Ours

FIGURE 10: The visual effect of comparison methods and the proposed model on the synthetic dataset of GOPRO [15]. Our method has satisfactory visual effects and good details.
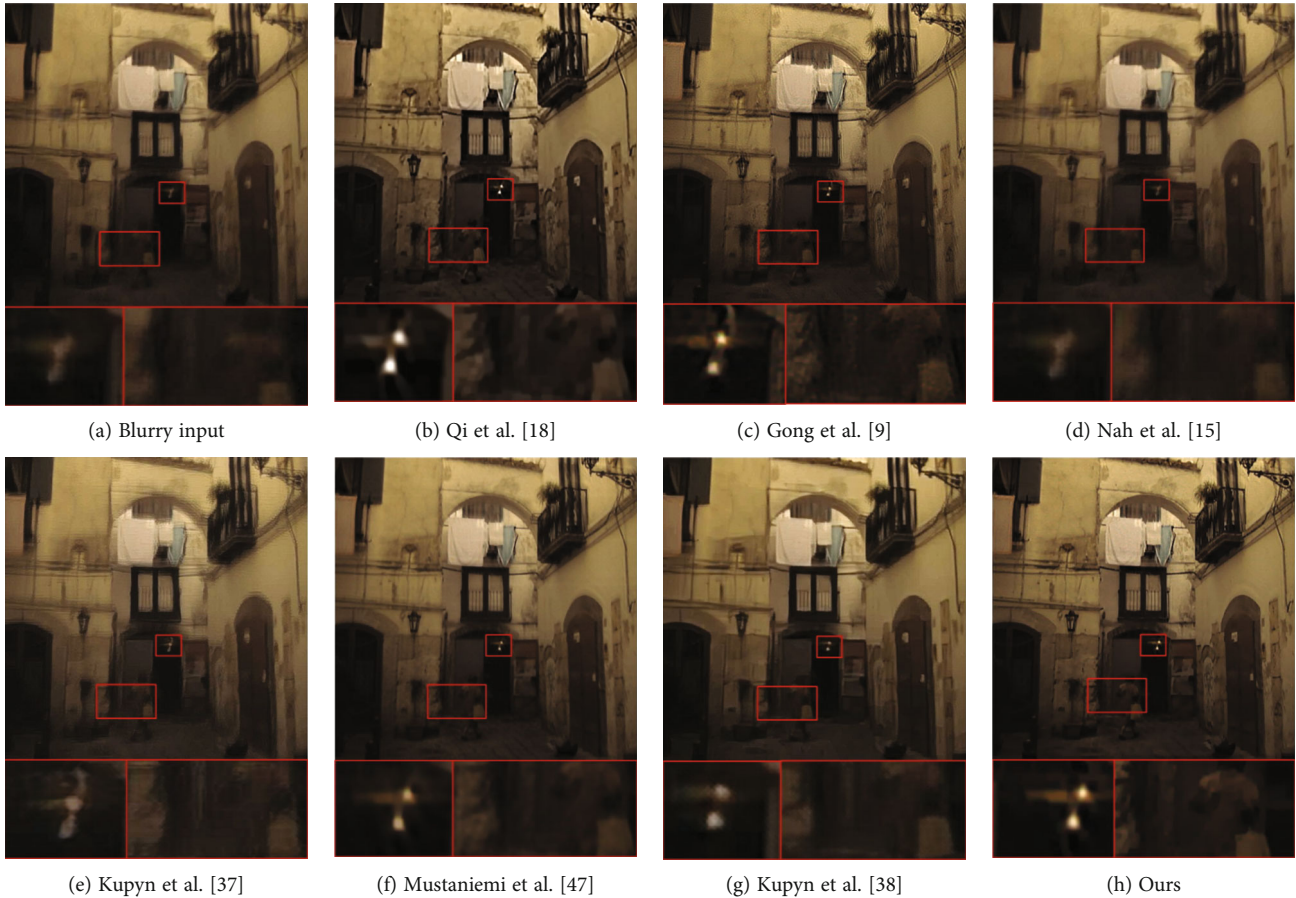
FIGURE 11: Visual comparison on the dataset of Köhler et al. [44], zoom in for the best view. The proposed method has a photo-realistic effect and generates much clearer details.

recovered results from SOTA methods and our model. Because methods [15, 18, 37, 38] are trained on the GOPRO dataset, they have some deblurring performance on the same test dataset. Their methods exhibit weak robustness on different datasets according to their over smoothed results on the Köhler dataset. In comparison to SOTA results, the proposed network has good generalization and generates deblurred images with significant structure and fine details.

To objectively evaluate the proposed method and SOTA methods on synthetic datasets, we use two quantitative assessment metrics: peak signal to noise ratio (PSNR) and structural similarity index (SSIM) [48]. Table 1 provides average PSNR and SSIM values from GOPRO and Köhler test datasets. Our network outperforms the competition in terms of PSNR and SSIM. The highest PSNR score indicates that there are the fewest similarities in content between the deblurred image and the matching ground truth. The SSIM value increases as the structural similarity difference between the deblurred image and the matching ground truth decreases. Because (1) the developed network directly narrows differences by various restrictions, resulting in greater pixel-wise coherence; and (2) image appearances are aided by primary building blocks, which contribute to structural similarity.

4.4. Experiments on Real Datasets. The evaluation of image deblurring performance is carried out on synthetic datasets, as mentioned in the preceding subsection, and satisfactory results are obtained. Real-world blurry images are frequently the result of more convoluted scenarios. This section uses the dataset of Lai [45] as a test dataset to confirm the effectiveness and generalization of the proposed network. Figures 12–14 show three groups of results from our network and comparison methods, respectively.

Figure 12 illustrates a blurry image under low-light situations. Gong et al. [9] adopt a data-driven way to estimate kernels and then adopt existing nonblind deconvolution algorithms to yield deblurred images. Because the kernel estimation and deblurred image computation are carried out independently, the "two-step" method rarely yields satisfactory results. Kupyn et al. [37] have low compatibility with real-world blurry images. The method [38] is an enhanced version of DeblurGAN that can restore appearances on synthetic blurry images. Because synthetic datasets cannot stimulate the true blur imaging process, methods of [15, 38] have limited generalization on real datasets, as seen in Figures 12(d) and 12(g). Although the developed network does not use low-light training samples, it still performs well on synthetic datasets.

TABLE 1: Performance comparison with SOTA methods evaluated by quantitative assessment of PSNR (dB) and SSIM. We calculate average values on synthetic datasets of GOPRO [15] and Köhler et al. [44].

Methods	GOPRO PSNR	GOPRO SSIM	Köhler PSNR	Köhler SSIM
Qi et al. [18]	28.9019	0.8694	21.3521	0.6521
Gong et al. [9]	27.2778	0.8187	21.2371	0.6490
Nah et al. [15]	28.3225	0.8588	20.8507	0.6340
Kupyn et al. [37]	25.2363	0.7773	19.0843	0.5838
Mustaniemi et al. [47]	25.9563	0.8285	20.4833	0.6442
Kupyn et al. [38]	27.8086	0.8664	21.2987	0.6544
Ours	29.2252	0.8714	21.3728	0.6602

(a) Blurry input (b) Qi et al. [18] (c) Gong et al. [9] (d) Nah et al. [15]

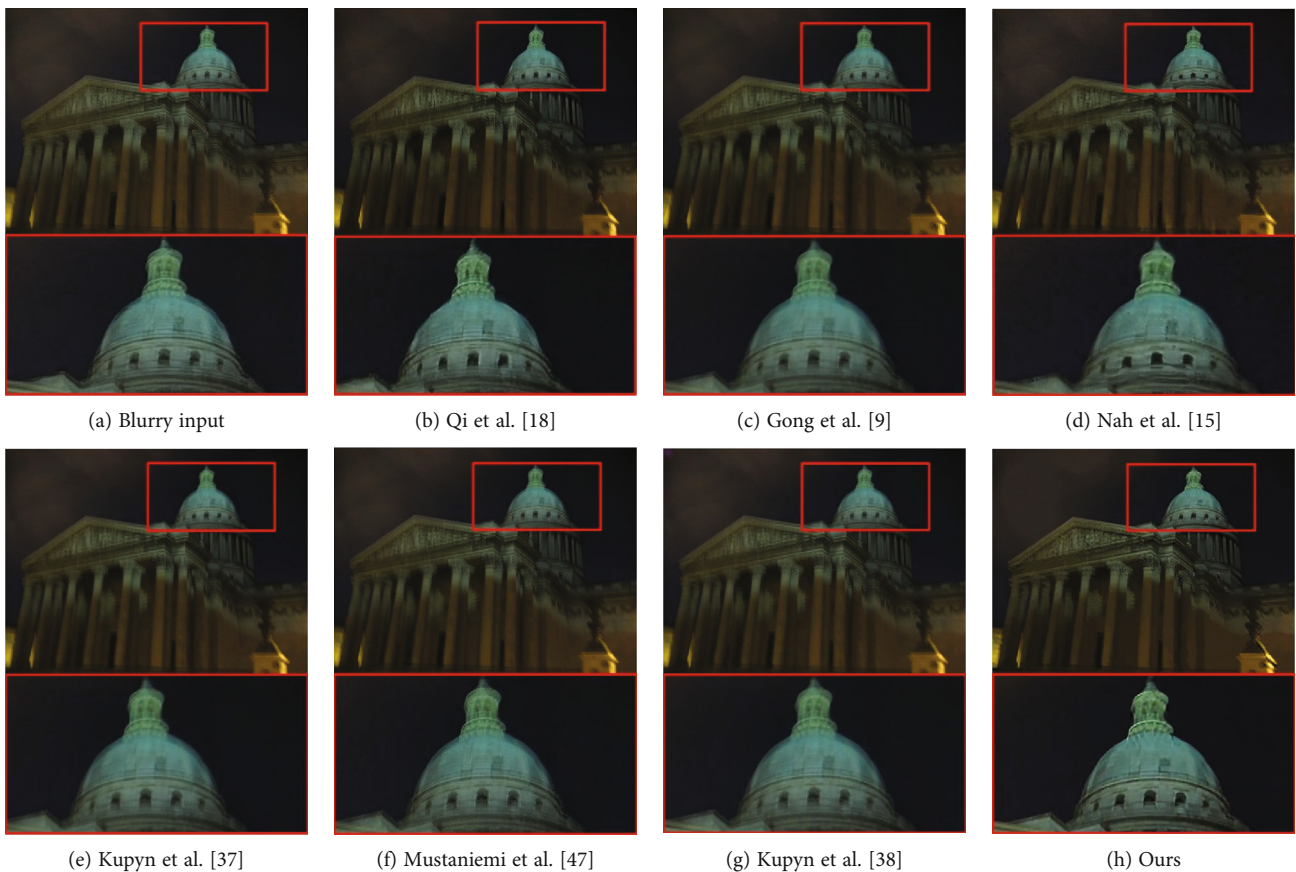(e) Kupyn et al. [37] (f) Mustaniemi et al. [47] (g) Kupyn et al. [38] (h) Ours

FIGURE 12: Visual comparison on the real-world images of Lai et al. [45]. Our network has satisfactory visual effects and good details.

Figures 13, 13(d), 13(e), 13(f), and 13(g) show that the class-specific blurry image has a light blur effect, but no high-quality deblurred results. This is because the methods of [9, 15, 37, 38, 47] fail to stimulate the real-world image blurring process. Figure 14 displays a blurry image under low-light conditions. Our method can restore blurred text and image background decently.

In conclusion, the proposed network is trained on simulated datasets and then tested on real-world images, confirming the model's efficiency and generalization that surpass SOTA methods. It takes advantage of the training data as well as the elaborately constructed network.

4.5. User Study. We randomly select 10 blurry images from datasets of Lai et al. [45] and GOPRO [15] for user study to perform an objective visual comparison. The results processed by various methods are displayed randomly and compared to the corresponding blurry input. Following that, we invite 5 participants with image processing experience to score the results. Furthermore, the participants have no idea which results are generated by ours. The scores range from 1 (worst) to 10 (best). We set the raw underwater image scores to 5 as a baseline. We anticipated that the good result has visual pleasant effects and abundant details, particularly the edge information.

(a) Blurry input   (b) Qi et al. [18]   (c) Gong et al. [9]   (d) Nah et al. [15]

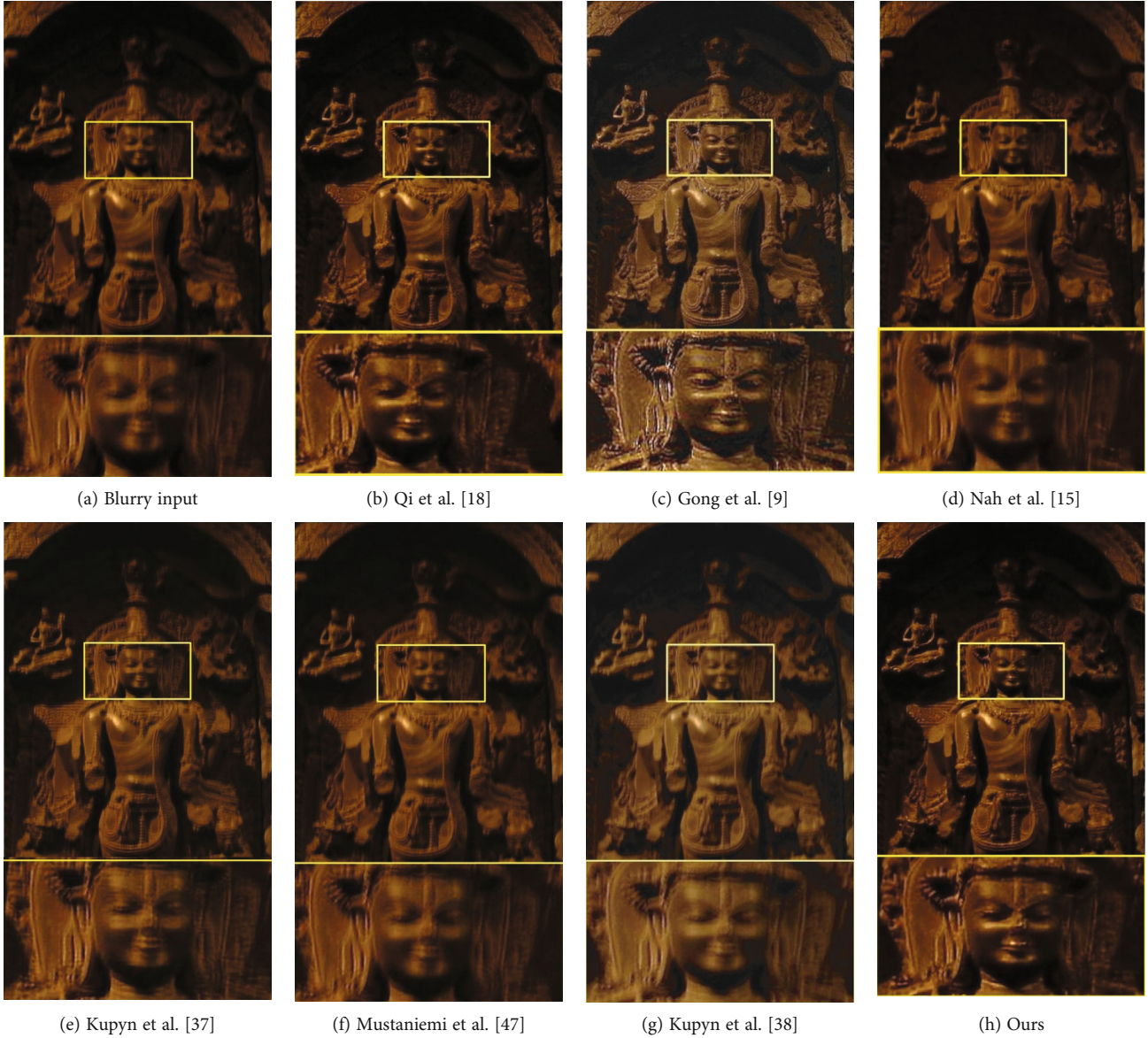(e) Kupyn et al. [37]   (f) Mustaniemi et al. [47]   (g) Kupyn et al. [38]   (h) Ours

FIGURE 13: Visual comparison on the real-world images of Lai et al. [45]. Our network has satisfactory visual effects and good details.

Table 2 displays the average visual quality scores of 10 test images. In Table 2, our results get the highest score. It indicates that our results are more consistent with human visual perception. In addition, the image deblurring methods also achieve decent scores. For blurry input, Qi et al. [18], Gong et al. [9], Nah et al. [15], Kupyn et al. [37], Mustaniemi et al. [47], Kupyn et al. [38], and ours, the average visual quality scores for the selected 20 blurry images are 5, 6.3, 5.3, 6.4, 5.4, 6.0, 5.6, and 7.2.

4.6. Execution Time and Computational Complexity. We show the developed method's efficiency by comparing it to SOTA methods in terms of execution time and computational complexity. The results of calculating floating point operations per second (FLOPs) and average execution time on images with a resolution of 1280×720 pixels are shown in Table 3. Because the deblurring method [9] involves a nonblind deconvolution procedure, they encounter compu-

tational difficulties. Nah et al. [15] use a multiscale CNN to expand the receptive fields for image deblurring performance. Because multiscale networks are independent between scales, the method [15] takes longer than Kupyn et al. [37] and Kupyn et al. [38]. Mustaniemi et al. [47] construct a U-Net-based CNN to optimize the network for image recovery. Mustaniemi et al. [47] increase the receptive fields by constructing the feature pyramid network. Qi et al. [18] propose a partial weight sharing network by building several blocks for achieving image deblurring. Unlike the previously described methods for image deblurring, the proposed method involves the formation of GradientNet, which saves a little more time than Kupyn et al. [38].

4.7. Investigation of the Impact of Loss Functions. To demonstrate the efficiency of each loss function in our model, we conduct ablation experiments, which include the following three experiments:

FIGURE 14: Visual comparison on the real-world images of Lai et al. [45]. Our network has satisfactory visual effects and good details.

TABLE 2: The average visual quality scores of test images.

Methods	I1	I2	I3	I4	I5
Inputs	5	5	5	5	5
Qi et al. [18]	6.3	6.5	6.9	6.0	6.0
Gong et al. [9]	5.0	5.2	5.6	5.9	5.0
Nah et al. [15]	6.8	6.9	6.3	5.5	6.5
Kupyn et al. [37]	5.1	5.5	5.3	5.4	5.6
Mustaniemi et al. [47]	5.5	6.1	6.3	5.5	6.5
Kupyn et al. [38]	5.4	5.7	5.3	6.5	5.0
Ours	7.5	7.7	7.2	6.9	6.8

TABLE 3: The efficiency of comparison image deblurring methods and the proposed method. We calculate algorithmiccomplexity and the average execution time on images with the size of 1280×720 of the test dataset of GOPRO.

Methods	FLOPs	Seconds
Qi et al. [18]	468.54G	0.8
Gong et al. [9]	4.12G	1300
Nah et al. [15]	1760.04G	8.1
Kupyn et al. [37]	678.29G	1.1
Mustaniemi et al. [47]	128.13G	4.5
Kupyn et al. [38]	411.34G	0.6
Ours	446.98G	0.7

(1) The proposed network without $L_{content}$ (w/o cont)

(2) The proposed network without $L_{gradient}$ (w/o gradient)

(3) The proposed network

Parameter setting and training methods consistent with the proposed network are used in the ablation experiments. The quantitative evaluation is performed on the dataset of GORPO. Figure 15 exhibits the visual results obtained from the ablation experiments performed on each module of the proposed network. The blurred image is presented in Figure 15(a). As illustrated in Figure 15(b), the network without $L_{content}$ generates visually unpleasant results. Such findings demonstrate that the performance of the proposed network has a positive correlation with $L_{content}$. In comparison, $L_{content}$ measures image semantic content coherence more robust. As shown in Figure 15(c), without the restrictions of $L_{gradient}$, the visual quality of the deblurred image significantly declined. Finally, an ablation investigation of proposed loss functions demonstrates that $L_{content}$ and
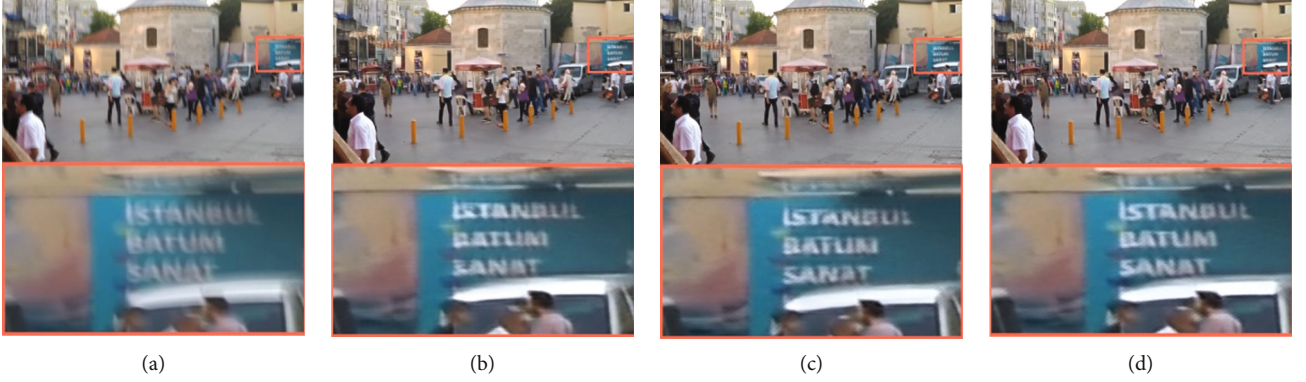
| (a) | (b) | (c) | (d) |

FIGURE 15: Visual results for the ablation study of loss functions on the nonuniform dataset GOPRO. (a) Blurry input, (b, c) are the results of the components of w/o content loss function and w/o gradient loss function, respectively. (d) The result of the proposed method. The proposed method generates clearer images with fine details and salient structures.

TABLE 4: Quantitative evaluations on the dataset of GOPRO for the ablation study of loss functions.

Loss	PSNR	SSIM
W/o content	26.3751	0.7921
W/o gradient	28.7011	0.8617
Ours	29.2252	0.8714

$L_{gradient}$ make significant improvements to deblurring performance. This confirms that it is feasible to exploit image structures as the target loss function to optimize network training.

In addition, the PSNR and SSIM objective evaluations are employed to assess the outcomes achieved from each ablated part, and the average results of the above networks tested on the GOPRO dataset are shown in Table 4. We observed that (1) when $L_{content}$ is removed, quantitative metrics of PSNR and SSIM sharply decrease; (2) when $L_{gradient}$ is ablated, the deblurring performance of our model also decreases. The quantitative evaluation results are compatible with the subjective visual effects, validating the efficiency of the image deblurring method based on image gradient priors.

4.8. Ablation Study on the Effectiveness of Building Blocks. To validate that the proposed network's image deblurring performance has a positive correlation with the constructed building blocks of RGCL, MBRBs, and GradientNet, we conduct ablation experiments, which involve the following four experiments:

(1) Our model without RGCL (w/o RGCL)

(2) Our model without (w/o MBRBs)

(3) Our model without GradientNet (w/o GradientNet)

(4) Our model

The first experiment denotes we replace RGCL with normal convolution layers in MBRBs. The purpose of the second experiment is to replace the multipath reuse module with a single path. The third experiment is to ablate GradientNet and validate how important the GradientNet is to generating salient structures.

We adopt the identical training strategy, parameter values, loss functions, and training datasets as in Section 4 in these experiments unless otherwise stated. Figure 16 displays the visual results obtained from the ablation experiments performed on each block of the proposed network. As exhibited in Figure 16(b), without the proposed RGCL, the letters that could not be recovered are dense with each other and difficult to identify. As shown in Figure 16(c), without the assistance of MBRBs, the visual quality of the deblurred image has a slight improvement to Figure 16(b). As illustrated in Figure 16(d), without GradientNet, the visual quality of the deblurred image significantly declined. Such finding validates that the performance of the proposed network has a positive correlation with GradientNet. Finally, an ablation investigation of proposed building blocks validates that RGCL, MBRBs, and GradientNet make considerable improvements to deblurring performance.

To evaluate the effectiveness of building blocks in our model, quantitative evaluation indicators of PSNR and SSIM are also employed. As displayed in Table 5, the average results of the above building blocks are tested on the GOPRO dataset. We can indicate that our model achieves the best quantitative assessment results. When RGCL, MBRBs, and GradientNet are removed, respectively, the quantitative metrics of the corresponding networks have different degrees of decrease. The quantitative evaluation results match the subjective visual effects, indicating that the image deblurring method based on image gradient priors is effective.

In conclusion, (1) instead of using plain convolutional layers, RGCL is developed to investigate and preserve the majority of image structures in a parallel manner; (2) MBRBs achieve structural feature correlations in a multipath reuse manner, then, recalibrated channels of these accumulated and enhanced nonlocal features are highlighted by implementing SENet module; (3) the structure-preserving subnetwork named GradientNet reinforces structural feature correlations by arranging many MBRBs in a cascade
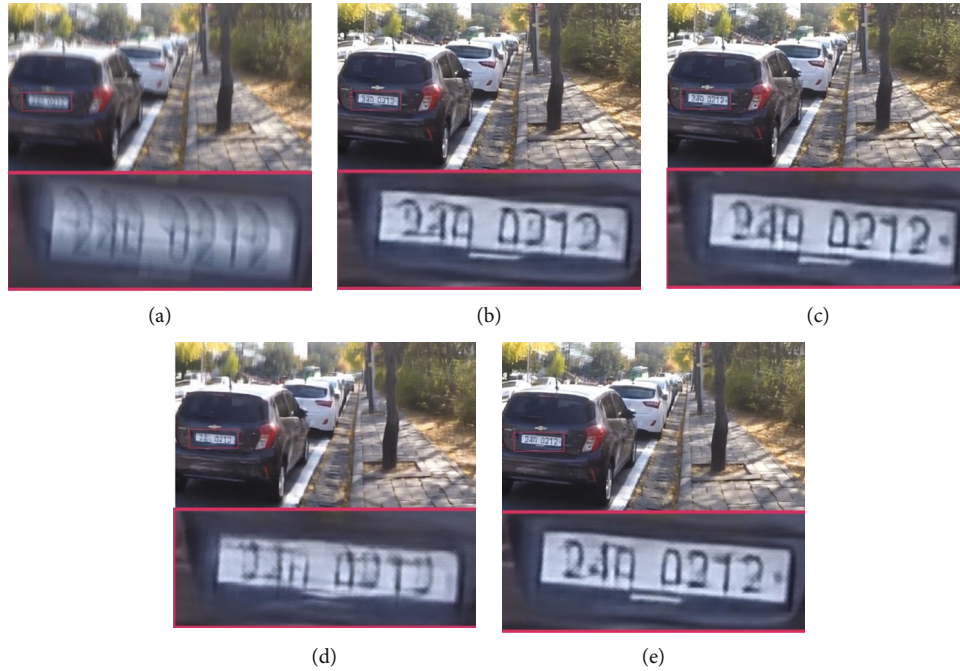
FIGURE 16: Visual results for ablation studies of building blocks on the dataset GOPRO. (a) Blurry input. (b–d) The results of network w/o RGCL, w/o MBRBs, and w/o GradientNet, respectively. (e) The result of the proposed method. Our network has satisfactory visual effects and good details.

TABLE 5: Quantitative evaluations on the dataset of GOPRO for the ablation study of building blocks.

Loss	PSNR	SSIM
W/o RGCL	29.0429	0.8691
W/o MBRBs	29.0859	0.884
W/o GradientNet	27.8708	0.8324
Ours	29.2252	0.8714

manner. It achieves the dynamic scene deblurring task under the guidance of image gradient features.

5. Conclusions

In this paper, we integrate image gradients in the proposed GAN-based network to generate images with clean appearances and conspicuous structures dynamically. We permeate image gradient priors to the design of network structures and loss functions. We develop a structure-preserving subnetwork named GradientNet, which contains the building blocks of RGCL and MBRBs. It investigates the relationship between image gradients and dynamic scene deblurring in a data-driven way. Subjective and objective comparison experiments on several synthetic datasets and real images are conducted to demonstrate the efficiency of the developed GradientNet. However, the proposed method has weak generalization on blurry images containing fewer structures and fruitful textures. Recovered images of those inputs are not smooth and have a strong coarse-grained visual perception. In the future, we consider extending and upgrading our model to mitigate the above limitations.

Data Availability

No data were used to support this study.

Conflicts of Interest

The author declares no conflicts of interest.

Acknowledgments

References

[1] S. Cho and S. Lee, "Fast motion deblurring," in *ACM SIG-GRAPH Asia 2009 papers*, pp. 1–8, Yokohama, Japan, 2009 Dec 1.

[2] L. Xu and J. Jia, "Two-phase kernel estimation for robust motion deblurring," in *European conference on computer vision*, pp. 157–170, Springer, Berlin, Heidelberg, 2010.

[3] L. Xu, S. Zheng, and J. Jia, "Unnatural L_0 sparse representation for natural image deblurring," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1107–1114, Portland, OR, USA, 2013.

[4] J. Pan, D. Sun, H. Pfister, and M. H. Yang, "Blind image deblurring using dark channel prior," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1628–1636, Las Vegas, Nevada, USA, 2016.

[5] W. Ren, X. Cao, J. Pan, X. Guo, W. Zuo, and M. H. Yang, "Image deblurring via enhanced low-rank prior," *IEEE Transactions on Image Processing*, vol. 25, no. 7, pp. 3426–3437, 2016.

[6] L. Sun, S. Cho, J. Wang, and J. Hays, "Edge-based blur kernel estimation using patch priors," in *IEEE International Conference on Computational Photography (ICCP)*, pp. 1–8, Cambridge, MA, USA, 2013.

[7] H. Zhang, J. Yang, Y. Zhang, and T. S. Huang, "Sparse representation based blind image deblurring," in *2011 IEEE International Conference on Multimedia and Expo*, pp. 1–6, Barcelona, Spain, 2011.

[8] J. Sun, W. Cao, Z. Xu, and J. Ponce, "Learning a convolutional neural network for non-uniform motion blur removal," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 769–777, Boston, MA, USA, 2015.

[9] D. Gong, J. Yang, L. Liu et al., "From motion blur to motion flow: a deep learning solution for removing heterogeneous motion blur," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3806–3815, Honolulu, Hawaii, USA, 2017.

[10] A. Chakrabarti, "A neural approach to blind motion deblurring," in *14th European Conference*, pp. 221–235, Amsterdam, The Netherlands, 2016.

[11] C. J. Schuler, M. Hirsch, S. Harmeling, and B. Schölkopf, "Learning to deblur," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 7, pp. 1439–1451, 2016.

[12] M. Hradiš, J. Kotera, P. Zemck, and F. Šroubek, "Convolutional neural networks for direct text deblurring," in *Proceedings of BMVC*, vol. 10, Swansea, UK, 2015no. 2.

[13] M. Jin, M. Hirsch, and P. Favaro, "Learning face deblurring fast and wide," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 745–753, Salt Lake City, Utah, USA, 2018.

[14] X. Mao, C. Shen, and Y. B. Yang, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," *Advances in neural information processing systems*, vol. 29, pp. 2802–2810, 2016.

[15] S. Nah, T. H. Kim, and K. M. Lee, "Deep multi-scale convolutional neural network for dynamic scene deblurring," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 257–265, Honolulu, Hawaii, USA, 2017.

[16] X. Tao, H. Gao, X. Shen, J. Wang, and J. Jia, "Scale-recurrent network for deep image deblurring," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 8174–8182, Salt Lake City, Utah, USA, 2018.

[17] X. Ran and N. Farvardin, "A perceptually motivated three-component image model-part i: description of the model," *IEEE Transactions on Image Processing*, vol. 4, no. 4, p. 401, 1995.

[18] Q. Qi, J. Guo, and W. Jin, "EGAN: non-uniform image deblurring based on edge adversarial mechanism and partial weight sharing network," *signal processing: image communication*, vol. 88, p. 115952, 2020.

[19] Y. Liu, Q. Jia, X. Fan, S. Wang, S. Ma, and W. Gao, "Cross-SRN: structure-preserving super-resolution network with cross convolution," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 8, pp. 4927–4939, 2022.

[20] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7132–7141, Salt Lake City, Utah, USA, 2018.

[21] T. Kim, B. Ahn, and K. Lee, "Dynamic scene deblurring," in *Proceedings of the IEEE international conference on computer vision*, pp. 3160–3167, Sydney, Australia, 2013.

[22] T. Kim and K. Lee, "Segmentation-free dynamic scene deblurring," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2766–2773, Columbus, OH, USA, 2014.

[23] J. Pan, Z. Hu, Z. Su, H. Y. Lee, and M. H. Yang, "Soft-segmentation guided object motion deblurring," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 459–468, Las Vegas, Nevada, USA, 2016.

[24] C. Li, R. Cong, J. Hou, S. Zhang, Y. Qian, and S. Kwong, "Nested network with two-stream pyramid for salient object detection in optical remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 11, pp. 9156–9166, 2019.

[25] C. Li, J. Guo, and C. Guo, "Emerging from water: underwater image color correction based on weakly supervised color transfer," *IEEE Signal processing letters*, vol. 25, no. 3, pp. 323–327, 2018.

[26] C. Li, C. Guo, J. C. Guo, P. Han, H. Z. Fu, and R. Cong, "PDR-net: perception-inspired single image dehazing network with refinement," *IEEE Transactions on Multimedia*, vol. 22, no. 3, pp. 704–716, 2020.

[27] C. Li, S. Anwar, and F. Porikli, "Underwater scene prior inspired deep underwater image and video enhancement," *Pattern Recognition*, vol. 98, article 107038, 2020.

[28] C. Li, C. Guo, J. Guo, P. Han, H. Fu, and R. Cong, "Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior," *IEEE Transactions on Image Processing*, vol. 25, no. 12, pp. 5664–5677, 2016.

[29] L. Xu, J. S. Ren, C. Liu, and J. Jia, "Deep convolutional neural network for image deconvolution," in *Proceedings of the Advances in Neural Information Processing Systems (NIPS)*, pp. 1790–1798, Montreal, Canada, 2014.

[30] L. Li, J. Pan, W. S. Lai, C. Gao, N. Sang, and M. H. Yang, "Learning a discriminative prior for blind image deblurring," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 6616–6625, Salt Lake City, USA, 2018.

[31] D. Ren, K. Zhang, Q. Wang, Q. Hu, and W. Zuo, "Neural blind deconvolution using deep priors," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3341–3350, 2020.

[32] H. Gao, X. Tao, X. Shen, and J. Jia, "Dynamic scene deblurring with parameter selective sharing and nested skip connections," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 3848–3856, Long Beach, USA, 2019.

[33] J. Zhang, J. Pan, J. Ren et al., "Dynamic scene deblurring using spatially variant recurrent neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2521–2529, Salt Lake City, USA, 2018.

[34] I. Goodfellow, J. Pouget-Abadie, M. Mirza et al., "Generative adversarial nets," in *NIPS*, pp. 2672–2680, Montreal, Canada, 2014.

[35] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, pp. 2223–2232, Venice, Italy, 2017.

[36] T. M. Nimisha, K. Sunil, and A. Rajagopalan, "Unsupervised class-specific deblurring," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 353–369, Munich, Germany, 2018.

[37] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, "Deblurgan: blind motion deblurring using conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 8183–8192, Salt Lake City, Utah, USA, 2018.

[38] O. Kupyn, T. Martyniuk, J. Wu, and Z. Y. Wang, "Deblur-GAN-v2: deblurring (orders-of-magnitude) faster and better," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 8878–8887, Seoul,Korea, 2019.

[39] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proceedings of International Conference on International Conference on Machine Learning*, vol. 30no. 1, p. 3, Atlanta, USA, 2013.

[40] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1125–1134, Honolulu, Hawaii, USA, 2017.

[41] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *14th European Conference*, pp. 694–711, Amsterdam, The Netherlands, 2016.

[42] K. Simonyan and A. Zisserman, "A very deep convolutional networks for large-scale image recognition," 2014, https://arxiv.org/abs/1409.1556.

[43] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of wasserstein gans," *Advances in neural information processing systems*, vol. 30, 2017.

[44] R. Köhler, M. Hirsch, B. Mohler, B. Schölkopf, and S. Harmeling, "Recording and playback of camera shake: benchmarking blind deconvolution with a real-world databaseEuropean conference on computer vision," Springer, Berlin, Heidelberg, 2012.

[45] W. S. Lai, J. B. Huang, Z. Hu, N. Ahuja, and M. H. Yang, "A comparative study for single image blind deblurring," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1701–1709, Las Vegas, Nevada, USA, 2016.

[46] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," 2014, https://arxiv.org/abs/1412.6980.

[47] J. Mustaniemi, J. Kannala, S. Särkkä, J. Matas, and J. Heikkila, "Gyroscope-aided motion deblurring with deep networks," in *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1914–1922, Waikoloa, HI, USA, 2019 Jan 7.

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