Learnable Blur Kernel for Single-Image Defocus Deblurring in the Wild

Jucai Zhai[1], Pengcheng Zeng[1], Chihao Ma[1], Jie Chen[1,2], Yong Zhao[1]*

[1] Shenzhen Graduate School, Peking University
[2] Peng Cheng Laborator
{jucaizhai, zpceng, machihao}@stu.pku.edu.cn, yongzhao@pkusz.edu.cn, chenj@pcl.ac.cn

Abstract

Recent research showed that the dual-pixel sensor has made great progress in defocus map estimation and image defocus deblurring. However, extracting real-time dual-pixel views is troublesome and complex in algorithm deployment. Moreover, the deblurred image generated by the defocus deblurring network lacks high-frequency details, which is unsatisfactory in human perception. To overcome this issue, we propose a novel defocus deblurring method that uses the guidance of the defocus map to implement image deblurring. The proposed method consists of a learnable blur kernel to estimate the defocus map, which is an unsupervised method, and a single-image defocus deblurring generative adversarial network (DefocusGAN) for the first time. The proposed network can learn the deblurring of different regions and recover realistic details. We propose a defocus adversarial loss to guide this training process. Competitive experimental results confirm that with a learnable blur kernel, the generated defocus map can achieve results comparable to supervised methods. In the single-image defocus deblurring task, the proposed method achieves state-of-the-art results, especially significant improvements in perceptual quality, where PSNR reaches 25.56 dB and LPIPS reaches 0.111.

Introduction

Defocus blur occurs when a scene point outside the depth-of-field (DoF) of the lens is out-of-focus (OoF) during a camera capture (Abuolaim and Brown 2020a). As shown in Figure 1, objects located at different depths have different degrees of blur. During the shooting process of the camera, the light from the scene point on the focal plane of the camera's object side is focused on the image plane, and no blur occurs. As the distance between the scene point and the focal plane of the object side of the camera gets farther, the projection of the scene point on the image plane also presents a larger circle of confusion (CoC), resulting in defocus blur. The spatial extent of the CoC can be described by a point spread function (PSF).

Most existing methods (Karaali and Jung 2018; Lee et al. 2021) start with single-image itself to reduce defocus blur. However, according to the PSF, defocus deblurring is related to monocular depth estimation. Both obtain depth clues from an image, which is an unreliable estimation. Therefore, it is a challenge to remove the defocus blur and restore the all-in-focus (AiF) image. Recent work in (Abuolaim and Brown 2020a) proposed a method to remove defocus blur using the left and right views of a dual-pixel (DP) sensor as input. The idea comes from the way DP sensors work, which is similar to stereo views to provide defocus clues. However, extracting real-time DP views is troublesome and complex in algorithm deployment(Abuolaim et al. 2021a). Existing single-image deblurring networks generate images of poor quality and cannot properly handle defocus blur due to the lack of reliable defocus clues. At the same time, there is also a lack of high-frequency (HF) information, which is unsatisfactory in human perception. It can be seen from the MDP-Net (Abuolaim, Afifi, and Brown 2022a) in Figure 1.

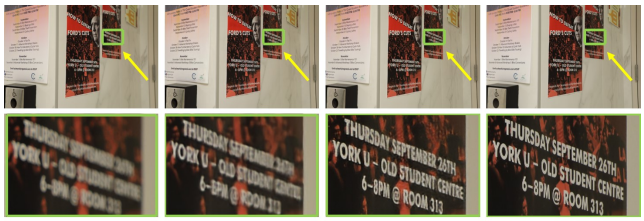

Figure 1: Qualitative comparison on the DPDD dataset. This image has obvious depth information. The first and last columns show defocused input images and their ground truth (GT). MDPNet is the best single-image defocus deblurring network. As we can see, compared to MDPNet, our generated images can handle large-scale defocus blur, recovering good structure and texture.

Based on these findings, we propose a new method to alleviate the defocus blur problem. We first generate a defocus map using DP views to obtain defocus clues. Since the current defocus map has no ground truth, we propose a learnable blur kernel (BK) to estimate the defocus map in an unsupervised way. We then propose DefocusGAN guided by defocus map. Since the defocus blur area is regular, we design a defocus adversarial loss to focus on the learning of blurred regions. Due to the use of an annealing strategy, the defocus map can be removed during inference to achieve

*corresponding author.

single-image deblurring.

The learnable blur kernel can simulate the real blur process, which simplifies the blur kernel calibration process (Xin et al. 2021) and achieves better defocus map estimation. The defocus clues brought by the defocus map can guide the network to deal with the amount of blur in different regions. GAN can enrich the HF information of images, bringing more realistic structure and details. The proposal of defocus loss allows the model to concentrate on learning defocused areas.

Our main contributions are summarized as follows:

- We propose a learnable blur kernel that uses DP views to estimate defocus maps via a self-supervised learning method that does not require calibration of the blur kernel. The defocus map generated by the proposed method is comparable to the current advanced supervised learning method.
- We propose DefocusGAN for the first time, a defocus map guided multi-scale defocus deblurring GAN. Compared with previous methods, the proposed method can maintain the information of clear areas, recover the texture and details of blurred areas, and generate more realistic images.
- The experiment results show that the proposed method is effective, with a small number of parameters, and achieves state-of-the-art performance in the single-image defocus deblurring task, where PSNR reaches 25.56 dB and LPIPS reaches 0.111.

Related Works

Defocus Deblurring

The methods of defocus deblurring are generally divided into two categories. One class of methods is a two-stage cascade method, which first estimates the defocus map and then deblurs the blurred image through non-blind deconvolution (Fish et al. 1995) guided by the defocus map. Another class of methods, such as DPDNet (Abuolaim and Brown 2020a), restore the AiF image directly from the blurred image.

In the two-stage method, the difference between blurred and sharp images is used for defocus map estimation, and then deconvolution is used to restore the defocus regions. (D'Andrès et al. 2016; Yi and Eramian 2016) used handcrafted features to estimate defocus maps from edge differences between sharp and blurred images. (Park et al. 2017) estimated the amount of edge blur by combining deep features and handcrafted features. (Lee et al. 2019) proposed a large dataset to estimate densely defocus maps. (Xin et al. 2021) estimated the defocus map using a calibrated BK in an unsupervised way. (Liang et al. 2022) used the DP views to estimate the defocus map for the first time. However, the above methods either require handcrafted features, ground truth (GT), or the calibration of BK. Using a single blurred image does not reliably estimate the defocus map. This makes the estimation of the defocus map very difficult. Thanks to recent work, (Abuolaim et al. 2021a) proposed a modeling method for BK. Based on the model, we use the learning method to estimate the BK, hoping to obtain better blur characteristics, and estimate the defocus map to obtain a more effective deblurring performance.

In another class of methods, (Abuolaim and Brown 2020a) first introduced a large DP dataset, DPDD, and proposed DPDNet, which was the first deep learning solution to the defocus deblurring problem using the defocus clues provided by DP views. Compared with manual feature design, it achieved better performance. Although DP views were initially used in autofocus tasks (Abuolaim, Punnappurath, and Brown 2018; Abuolaim and Brown 2020b), more applications have been discovered, including defocus deblurring (Vo 2021), depth estimation (Garg et al. 2019; Punnappurath et al. 2020; Wu et al. 2021; Kang et al. 2021), stereo matching (Zhang et al. 2020; Pan et al. 2021), reflection removal (Punnappurath and Brown 2019), synthetic DoF (Wadhwa et al. 2018) and motion synthesis (Abuolaim, Afifi, and Brown 2022b). Subsequently, Recurrent Neural Networks (Abuolaim et al. 2021a) and Multiplane Image (Xin et al. 2021) were also used in DP defocus deblurring task. (Liang et al. 2022) proposed BaMBNet, a blur-aware network, which achieved better results. Since the DP views are difficult to obtain, reasoning and deploying the network are cumbersome. Some recent works have turned to utilizing DP views to assist single-image defocus deblurring. (Lee et al. 2021) proposed IFAN, (Son et al. 2021) proposed KPAC, (Abuolaim, Afifi, and Brown 2022a) proposed a multi-task defocus deblurring framework. We found that none of the above networks considered the restoration of image details. Although they get an acceptable PSNR, HF details and textures are somehow missing. The purpose of deblurring is to restore clear image details, the existing work does not seem to achieve the desired fidelity.

GAN

GAN contains two models, namely the discriminator D and the generator G, which constitutes a minimax game (Goodfellow et al. 2014; Motwani and Parmar 2020). (Ledig et al. 2017) observes that tasks driven by L1 or L2 loss achieve high PSNR, but tend to lack HF details and are unsatisfactory in perceptual quality. GAN usually works well on details and textures. In the field of image deblurring, GAN has a very wide range of applications (Zhang et al. 2022), DeblurGAN (Kupyn et al. 2018) and DeblurGAN-V2 (Kupyn et al. 2019) are the most famous methods of them. While these current works were not suitable for defocus deblurring task, which achieved weak performance. Current deblurring-related GANs are based on local or global learning, whereas defocus blur is regularly regional, which has not been noticed before.

Method

Overall Architecture

Our model consists of two modules: one for defocus map estimation and the other for the defocus map guided multi-scale defocus deblurring (DefocusGAN) module. Figure 2 shows the illustration of our proposed framework. Defocus clues are especially important in single-image defocus deblurring task, therefore, the first step is to estimate the defo-

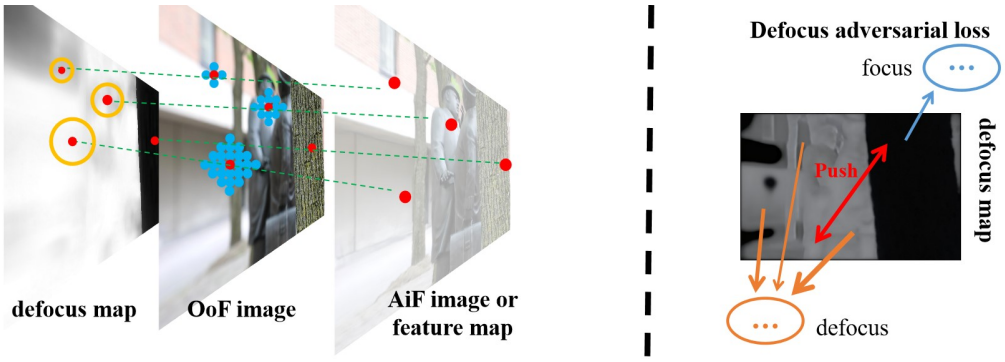

Figure 2: Illustration of our framework, DefocusGAN. Our framework consists of two main modules, the generator G and discriminator D. The G network takes a single input image and outputs an estimated AiF image after a multi-scale defocus deblurring block guided by a defocus map. The discriminator judges the difference with the GT. We propose a defocus adversarial loss to distinguish defocus regions within the image. The left is an illustration of G; the whiter the pixels in the defocus map, the larger the defocus parallax. The right is an illustration of defocus adversarial loss, which applies weights to different blurred areas under the guidance of the defocus map to distinguish out-of-focus and in-focus areas within the image.

cus map using DP images. Unlike BaMBNet, we propose a learnable BK and a blur reconstruction function. Then, we construct DefocusGAN. We train the network using the annealing algorithm. During the inference phase, the guide part of the defocus map can be removed and only a single-image can be used for inference.

Learnable Blur Kernel for Defocus Map Estimation

Obtaining the spatial variation of the CoC and estimating the defocus map can effectively guide the blurred image to defocus deblur. Since the GT of the defocus map is not available, we propose an unsupervised method to estimate the defocus map. As mentioned earlier, the DP views can provide reliable defocus clues. We use the DP views for defocus parallax estimation to generate the defocus map. The value of each pixel on the defocus map represents the radius of the CoC at the current position, which is half of the defocus parallax.

$$\begin{matrix} I_{DPleft} \\ I_{DPright} \end{matrix} \xrightarrow{f(I_l, I_r, \theta)} I_{DM} \xrightarrow{g(I_{AiF}, I_{DM}, \varphi)} I^*_{OoF} \tag{1}$$

Let DP views $(I_{DPleft}, I_{DPright})$ as input, learn a network f to estimate the defocus map I_{DM}, and then guide the AiF image to blur to obtain the blurred image I^*_{OoF}. This blur reconstruction process can be learned using a network g. θ, φ represent the learnable parameters of the network f and g, respectively. Therefore, we propose a learnable blur kernel (BK) for blur operations. Then build a reblur geometric loss, which is the L1-Loss of the estimated blurred image I^*_{OoF} and GT blurred image I_{OoF}:

$$L_{gem} = |I^*_{OoF} - I_{OoF}| \tag{2}$$

Re-blurring requires building BK, while calibrated BK is difficult to obtain (Xin et al. 2021). Based on this intuition, we propose to learn the BK to make the BK more realistic, which does not require a calibration process. The learnable BK is constructed in this way. We need a BK that is similar to the real BK as the initial parameter and train the defocus map estimation network. After a certain stage of training, fix the parameters of the defocus map estimation network, train the defocus map estimation network and the BK alternately, and simultaneously improve the performance of the defocus map and the BK. (Abuolaim et al. 2021a) sampled the BK of the camera, and proposed a method to construct the BK, which is similar to the sampled BK. It is observed that the defocus BK of the camera is a high-pass filter:

$$\begin{aligned} B(x, y) &= (1 + (\frac{D_0}{\sqrt{(x-x_0)^2 + (y-y_0)^2}})^{2n})^{-1} \circ C(x_0, y_0) \\ B_{init} &= G(\kappa, \kappa) * B \end{aligned} \tag{3}$$

Where B is a Butterworth high-pass filter centered at (x_0, y_0), n is the filter order, D_0 controls the 3 dB cut-off frequency, C is a circular function with (x_0, y_0) as the center, and $\circ$ represents the Hadamard product. Then use a gaussian kernel with a standard deviation of $\kappa \times \kappa$ to perform convolution smoothing. This BK is used as the initial parameter.

For the reconstructed model, pixels with different radii of the CoC, we divide the space to blur. There are:

$$I^*_{OoF} = I_{AiF}[c(d)] * H(B(\theta); [c(d)]) \tag{4}$$

Where d is the depth, c is the estimated CoC radius, I_{AiF} is the AiF image, H is the blur reconstruction function g, and B is the learnable BK. The initial parameters are B_{init}, and the optimal BK can be learned according to the loss function.

Since it is an unsupervised estimation, the above loss can't well reflect the size of the CoC of pixels and the influence of noise. To alleviate this problem, we add a prior regularization term to penalize the gradient of the network output and estimate a smooth defocus map. Finally, the total loss is:

$$L_{DM} = L_{gem} + \lambda \|\nabla(I_{DM})\| \tag{5}$$

Where I_{DM} represents the estimated defocus map, λ is the balance factor between the geometric loss term and the regularization term. For the defocus map estimation network f, for simplicity, a network similar to the defocus deblurring network is used, which is introduced in the next section.

Method	Indoor			Outdoor			Indoor & Outdoor				
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	MAE↓	LPIPS↓	Params (M)
JNB	25.52	0.784	0.188	21.16	0.632	0.274	23.28	0.706	0.049	0.232	-
EBDB	25.83	0.790	0.326	21.21	0.631	0.407	23.47	0.708	0.049	0.368	-
DMENet	25.70	0.789	0.315	21.51	0.655	0.402	23.55	0.720	0.049	0.360	26.94
DPDNet (S)	26.52	0.828	0.179	22.08	0.689	0.229	24.25	0.757	0.044	0.204	35.25
IFAN	27.80	0.856	0.131	22.70	**0.719**	0.179	25.18	**0.786**	0.041	0.156	10.48
KPAC	28.02	0.852	0.129	22.64	0.702	0.190	25.26	0.774	0.041	0.161	2.06
MDPNet	28.02	0.840	0.186	22.82	0.689	0.261	25.35	0.763	0.040	0.225	46.86
Ours	**28.31**	**0.857**	**0.086**	**22.94**	0.718	**0.135**	**25.56**	**0.786**	**0.039**	**0.111**	4.59

Table 1: Quantitative comparisons with single-image defocus deblurring methods. The best results are indicated in boldface. Results are on the DPDD dataset (the test set consists of 37 indoor and 39 outdoor scenes).

(a) Input (b) JNB (c) DPDNetS (d) IFAN (e) MDPNet (f) Ours (g) GT

Figure 3: Qualitative comparison on the DPDD dataset. The first and last columns show defocused input images and their GT, respectively. In the columns, we show the deblurring results of different methods.

Defocus Deblurring GAN (DefocusGAN)

DefocusGAN consists of a generator G and a discriminator D. We design a specialized generator network and loss function for the defocus deblurring task. Its architecture and loss will be introduced here, respectively.

DefocusGAN Generator. In the single-image defocus deblurring task, due to the lack of reliable defocus clues without using DP views, the performance on large-area blurring is poor. Inspired by this, we propose a defocus map guided multi-scale defocus deblurring network that utilizes the defocus map to provide defocus clues, constructs multi-scale layers to deal with large area blur, and uses the GAN to restore image details.

With the defocus map, a targeted defocus deblurring operation can be implemented according to the radius of the circle of confusion (CoC) corresponding to the pixel. Use the defocus mask to treat pixels with different circle sizes separately:

$$I^*_{AiF} = \sum (K(c_i) f(I_{OoF}, c_j, \alpha)) \quad s.t. \quad K(c) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \tag{6}$$

Where K is a binary mask function, f is the defocus deblurring function, c is the radius of the CoC, and α are learnable parameters. Areas with the same radius of CoC can be deblurred with the same defocus deblurring function. We set N deblurring branches according to the range of the radius of CoC in the defocus map. Using multiple branches to extract features, adaptive deblurring.

An illustration of our framework is shown in Figure 2. The model takes a single-image as input, passes it through a defocus map guide block (DGB), and obtains preliminary deblurred features. Then downsample to $\frac{1}{4}$ size of the input and go through the same operation. Then downsample to $\frac{1}{8}$ size of the input and repeat the operation. In this way, the cascade refinement can obtain a larger receptive field and obtain the relationship of the large-area blur range. DGB divides multiple branches according to the characteristics of the defocus map. We design 4 sets of defocus masks according to the characteristics of the defocus map and divide DGB into 4 branches. It is then multiplied by the defocus mask to obtain the features of the corresponding area. We assign a weight to the defocus mask, use the simulated annealing algorithm to reduce the weight during the training process to gradually remove the guidance of the defocus mask, and then use the prior learned by the network to continue training.

When the radius of CoC is small, the pixel where it is located can be recovered without aggregating the features

of surrounding pixels. As the radius increases, the distance between the center of CoC and the current pixel increases, and a larger receptive field is required to aggregate the features of the surrounding pixels to achieve the effect of deblurring. For branches with a small radius of CoC, it is only necessary to pay attention to the pixel itself and surrounding features, and a fully convolutional network can meet the needs. For branches with a large radius, it is necessary to aggregate blurred information over a larger receptive field. Using the U-Net-like as the backbone of these branches, we replace the convolutional layer with the Residual Channel Attention Module (RCAB), further improving the ability to obtain global information. Finally, DGB is a 4-branch network where each branch is a Unet-like structure with 8 convolutional layers. The convolutional layers of 4 branches are as follows: 1 fully convolutional layer, 1 RCAB, 2 RCAB, and 3 RCAB, respectively. For $\frac{1}{4}$ and $\frac{1}{8}$ scale features, they have a larger receptive field after downsampling, so we appropriately reduce the parameters in DGB.

DefocusGAN Discriminator. The discriminator D is used to judge the gap between the input image and the real image, we take the output of G as the input of D and use the D like patchGAN (Isola et al. 2017), which uses a 3-layer fully convolutional network.

Overall Loss Function. The overall loss function used for training, especially the proposed defocus adversarial loss, has been investigated in this section.

Defocus adversarial loss. For the defocus deblurring task, we design a defocus adversarial loss that focuses on defocus regions. We reweigh the discriminator response for the first time on defocus deblurring. Due to the regularity of the defocus distribution, based on the defocus clues provided by the defocus map, we can easily get the blurred area. According to the radius of the CoC in the defocused areas, we assign different weights to the image features output by the discriminator. The larger the radius of the CoC in the defocused areas, the greater the defocus weight of the corresponding area. With this loss, the D can increase the ability to distinguish different blurred areas in the image and strengthen the learning ability of G for the blurred areas.

$$L_{defocusadv} = \frac{1}{n}\sum_{n=1}^{N} -\varphi(I_{DM}) D_{\theta D}(G_{\theta G}(I_{OoF})) \quad (7)$$

Where I_{DM} represents the defocused image, I_{OoF} represents the input defocused image, $G_{\theta G}$ represents the G network, and $D_{\theta D}$ represents the D network, $\varphi(I_{DM})$ represents the operation of assigning weights according to the defocus map, and $L_{defocusadv}$ refers to calculating the loss after assigning defocus weights to the AiF image output by discriminator D.

Similar to WGAN-GP (Gulrajani et al. 2017), we add a gradient regularization term. While stabilizing GAN training, push the generative distribution towards a more realistic distribution. Instead of indiscriminately learning pictures, we focus on learning defocus regions. Similar to DeblurGANV2 (Kupyn et al. 2019), we use L1-Loss as the content loss L_c. Compared to previous methods for defocus deblurring, we use a perceptual loss (Johnson, Alahi, and Fei-Fei 2016) L_p to update the model, which computes the Euclidean loss on the VGG19 (Simonyan and Zisserman 2015) $conv3\_3$ features maps.

We use the three losses above weighted to get L_G to train the model, where α and β are hyperparameters to balance different types of loss.

$$L_G = L_c + \alpha \times L_p + \beta \times L_{defocusadv} \quad (8)$$

Method	PSNR↑	SSIM↑	MAE↓	LPIPS↓
DPDNet (DP)	25.13	**0.786**	0.041	0.223
RDPDNet	25.39	0.772	0.040	0.179
Ours	**25.56**	**0.786**	**0.039**	**0.111**

Table 2: Quantitative comparisons with some methods using DP views as input. Results are on the DPDD dataset.

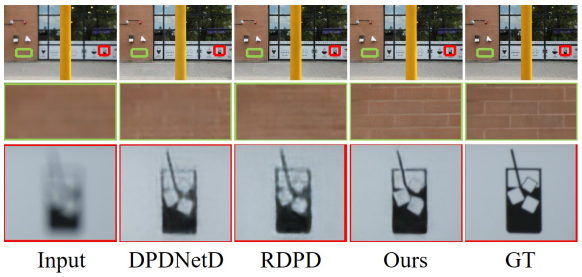

Figure 4: Qualitative comparison with methods using DP views as input on the DPDD dataset.

Experiments

Datasets

We use the dataset DPDD provided by (Abuolaim and Brown 2020a) for training and testing. This dataset has 500 sets of images, and each set of images includes a defocus blurred image, a pair of DP views, and an all-in-focus (AiF) image with a resolution of 1680 × 1120. Here, like most methods (Abuolaim et al. 2021b), flowing the settings, 500 groups have been divided into 350, 74, and 76 groups according to the training set, validation set, and test set. We also use the CUHK dataset (Shi, Xu, and Jia 2015) and the Google PixelDP dataset (Abuolaim and Brown 2020a) to verify the generalization of the network.

Implementation Details

We first train the defocus map estimation network, taking the DP views as input, to estimate the defocus map. The hyperparameter λ is set to 10^{-5}, and the learning rate is set to 2×10^{-5}. First, the blur reconstruction network is fixed, and the defocus map estimation network is trained for 10 epochs. Then fix the parameters of the defocus map estimation network, and then train the blur reconstruction network. Use the method of alternating training, alternating every 5 epochs until 30 epochs. Referring to (Liang et al. 2022), we set the upper limit of the radius of the CoC to 25 pixels.

Then, we train the defocus deblurring network. Here, the 512×512 single-image and the defocus map are fed into

Method	PSNR↑	SSIM↑	MAE↓	LPIPS↓
BaMBNet	22.56	0.687	0.054	**0.319**
DMENet	23.55	**0.720**	**0.049**	0.360
Ours	**23.56**	0.711	**0.049**	0.361

Table 3: Quantitative comparisons with the defocus map used for recovering AiF images on the DPDD dataset. Results are on the DPDD dataset.

Method						Metrics
Base	RCAB	DGB	MS	L_p	GAN	PSNR↑
✓						24.73
✓	✓					25.10
✓	✓	✓				25.30
✓	✓	✓	✓			25.47
✓	✓	✓	✓	✓		25.44
✓	✓	✓	✓	✓	✓	**25.56**

Table 4: Quantitative results of the ablation experiments on the DPDD dataset.

the network. The number of iterations of the simulated annealing algorithm is set to 2×10^4, the hyperparameters α, β are set to 0.012 and 0.002, respectively. The initial learning rate is set to 2×10^{-4}, which decreases by half every 30 epochs. After about 15 epochs, the model no longer relies on the guidance of the defocus map, and it gradually converges after 90 epochs. When inferring, we can discard the defocus map and use only a single-image as input to complete the defocus deblurring operation.

The batch size of both networks is set to 4 and optimized using the Adam optimizer, where b1=0.9, b2=0.999. We implemented the method using Pytorch and trained it on an NVIDIA RTX 3090 GPU.

Performance Evaluation

Like many works, to evaluate the performance of defocus deblurring, we use the test set provided by (Abuolaim and Brown 2020a) for testing. We compare the results with recent single-image defocus deblurring works. JNB (Shi, Xu, and Jia 2015), EBDB (Karaali and Jung 2018) and DMENet (Lee et al. 2019) are methods based on defocus maps. After they estimate the defocus map, they use non-blind deconvolution to defocus deblurring. DPDNet (single) (Abuolaim and Brown 2020a), IFAN (Lee et al. 2021), KPAC (Son et al. 2021), and MDPNet (Abuolaim, Afifi, and Brown 2022a) are direct estimation methods that can directly restore AiF images. For the above methods, we use the code and weights provided by the authors for testing (IFAN uses data augmentation. So we remove this and retrain according to the code and training method provided by the authors). For JNB, EBDB, and DMENet, following the advice of (Abuolaim and Brown 2020a), we use the deconvolution method (Krishnan and Fergus 2009) to recover the AiF image using the estimated defocus map. We also evaluate the number of network parameters in the inference stage to characterize the size of the model.

We use the commonly used metrics PSNR, SSIM, MAE, and LPIPS for defocus deblurring to evaluate the quality of the images. Table 1 shows the quantitative results of our method and other methods. Our method shows higher quality, outperforms all current methods with few model parameters, and restores image details to a great extent, improving the realism of images. Figure 3 shows a qualitative comparison. Traditional methods based on defocus maps and deconvolution have large blur areas. The performance of MDPNet, KPAC, and IFAN is greatly improved compared with the previous results, but often produces unnatural textures such as artifacts. For example, the texture of red walls and bronze figures. In particular, compared with this method, our method can better handle the texture of the image and recover the contours of objects that conform to human subjective perception, such as text in magazines. From Figure 3, we can see that our method can better recover large-area blur, image details, and texture.

For completeness, we also report some methods that take DP views as input. Table 2 shows this comparison. It can be seen that, compared with DPDNet (DP views) and RDPDNet (Abuolaim et al. 2021a), our method has a good performance. As shown in Figure 4, compared to these methods, we perform better, recovering images with texture details and human perception. Models are smaller and more functional. In the inference stage, only a single-image is required, but DP-based methods require access to 2 DP views. Because of the difficulty of obtaining DP views, our method has great advantages.

Since defocus maps have many practical applications (Lee et al. 2019), we compare our method with current deep learning-based methods for recovering defocus maps. We use a non-blind deconvolution method to recover AiF images with defocus maps on the DPDD dataset. It can be seen in Table 3 that, compared with BAMBNet, which is also based on the unsupervised method to recover defocus maps, we achieve great improvement. Compared with the supervised learning-based method DMENet, we achieve competitive results.

BK function	PSNR↑	SSIM↑	MAE↓
Gaussian	22.56	0.687	0.054
Butterworth	23.31	0.702	**0.049**
Ours	**23.56**	**0.711**	**0.049**

Table 5: Ablation study to demonstrate the effectiveness of our learnable blur kernel. Results are on the DPDD dataset.

Ablation Studies

Effects of each module. To demonstrate the effectiveness of each part of the module, we conduct ablation experiments in which all models are trained under the same conditions (e.g., optimizer, learning rate, random seed, etc). Specifically, we take a single-scale network as the baseline model. The components of the network have a defocus map guide (DG) part, RCAB, multi-scale (MS) module, perceptual loss (L_p), and GAN. Used components are recovered gradually from the baseline model. As can be seen in Table 4, the DG part is important for defocusing the image, providing defocusing clues, and improving the performance of the model, but the

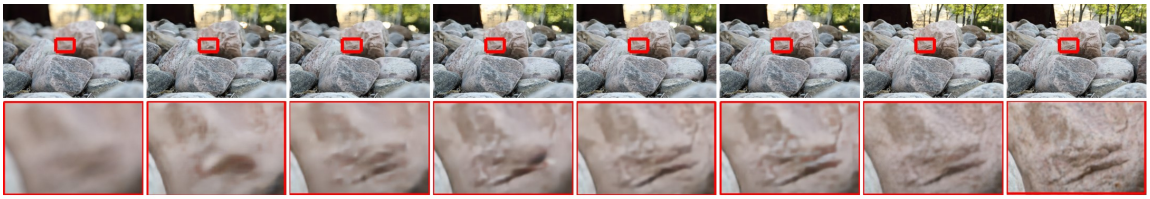

Figure 5: Qualitative results of an ablation study on the DPDD dataset.

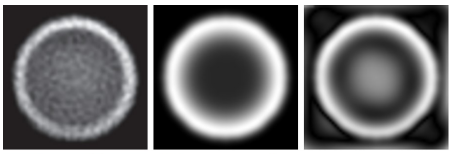

Figure 6: Qualitative comparison between different blur kernels.

.

Method		Metrics		
generator	discriminator	PSNR↑	SSIM↑	LPIPS↓
FPN	w original loss	24.23	0.746	0.150
Ours	w original loss	25.31	0.780	0.149
Ours	w dbGAN loss	24.92	0.763	**0.132**
Ours	w defocus loss	**25.42**	**0.784**	0.160

Table 6: Ablation study to demonstrate the effectiveness of GAN. Results are on the DPDD dataset.

removal effect of large-area blur is still insufficient, as shown in Figure 5. Therefore, we introduce the MS module, which improves the performance of large-area defocus deblurring. Using L1-Loss will cause the image to be too smooth. For the deblurring task, it is important to restore the details and texture of the image. We introduce GAN and perceptual loss to further restore the structure and texture of the image, making the image closer to human perception.

Effectiveness of learnable blur kernel. To demonstrate the effectiveness of the proposed learnable BK, we analyzed the effects of different BKs on the recovered defocus map, as shown in Table 5. It can be seen that the defocus map generated using the learnable BK gets better performance. Since (Abuolaim et al. 2021a) does not provide specific parameters for sampling blur kernels, we can only perform qualitative comparisons. As can be seen from Figure 6, the real blur kernel approximates a band-stop filter, and we have learned this feature well.

The impact of different GANs. We also experiment with different GANs, as shown in Table 6. We use the adversarial loss without defocus weights as the original loss. Each pixel has the same weight, with indiscriminate attention to images. For example, using FPN (Lin et al. 2017) in deblurGANV2 as G, the performance of PSNR 24.23 and LPIPS 0.150 is obtained. This shows that for the defocus deblurring task, to design a network for its characteristics, it is especially necessary to provide defocus clues. We also experiment with doubleGAN (dbGAN) (Kupyn et al. 2019), D that fuses global and local, but it does not work well. We think that the defocus blur is regional, not global, so D with global properties does not work well. While our proposed defocus adversarial loss increases the focus on defocused areas and achieves better performance.

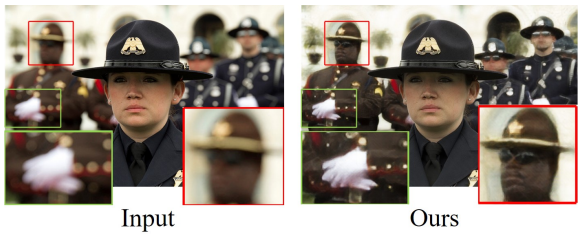

Figure 7: Defocus deblurring results of the proposed method on the CUHK dataset.

Generalization Ability

Generalizability is very important for a model. After training on the DPDD dataset, we test it directly on the Google PixelDP dataset and the CUHK dataset. Since there is no GT, we only show the qualitative results of the model. Figure 7 shows the results of the CUHK dataset. The results on the Google PixelDP dataset can be viewed in the supplementary materials. As we can see, our perception is very good compared to the input. For the CUHK dataset, which mainly contains portraits, we can significantly recover details and textures, such as the soldier's face and hand contours.

Conclusion

We propose a single-image defocus deblurring GAN and an unsupervised method for estimating defocus maps with a learnable blur kernel. The learned defocus map is used to guide the network for defocus deblurring. The proposed network can effectively handle large-area blur and effectively reconstruct image details and textures. The recovered defocus maps are comparable to current supervised methods. The effect of each component is verified experimentally, and it accompanies a great performance with fewer parameters.

Acknowledgements

This work is supported by the Shenzhen Science and Technology Research and Development Fund (No. JCYJ20180503182133411, JSGG20201022153800002, KQTD20200820113105004). Jucai Zhai would like to thank Dr. Emad Iranmanesh for proofreading the article and for his constructive feedback.

References

Abuolaim, A.; Afifi, M.; and Brown, M. S. 2022a. Improving Single-Image Defocus Deblurring: How Dual-Pixel Images Help Through Multi-Task Learning. In *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 82–90.

Abuolaim, A.; Afifi, M.; and Brown, M. S. 2022b. Multi-View Motion Synthesis via Applying Rotated Dual-Pixel Blur Kernels. In *2022 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW)*, 701–708.

Abuolaim, A.; and Brown, M. S. 2020a. Defocus Deblurring Using Dual-Pixel Data. In *Computer Vision – ECCV 2020*, 111–126. Cham: Springer International Publishing.

Abuolaim, A.; and Brown, M. S. 2020b. Online Lens Motion Smoothing for Video Autofocus. In *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 147–155.

Abuolaim, A.; Delbracio, M.; Kelly, D.; Brown, M. S.; and Milanfar, P. 2021a. Learning to Reduce Defocus Blur by Realistically Modeling Dual-Pixel Data. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2269–2278.

Abuolaim, A.; Punnappurath, A.; and Brown, M. S. 2018. Revisiting Autofocus for Smartphone Cameras. In *Computer Vision – ECCV 2018*, 545–559. Cham: Springer International Publishing.

Abuolaim, A.; Timofte, R.; Brown, M. S.; Zhang, D.; Wang, X.; Zamir, S. W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F. S.; Shao, L.; Liu, S.; Lei, L.; Feng, C.; Xiong, Z.; Xiao, Z.; Xu, R.; Zhu, Y.; Liu, D.; Vo, T.; Miao, S.; Shah, N. A.; Liang, P.; Zhong, Z.; Hu, X.; Chen, Y.; Li, C.; Bai, X.; Zhang, C.; Yao, Y.; Gang, R.; Nathan, S.; Ragavendran, T.; Srinija, V.; and Srivatsav, V. 2021b. NTIRE 2021 Challenge for Defocus Deblurring Using Dual-pixel Images: Methods and Results. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 578–587.

D'Andrès, L.; Salvador, J.; Kochale, A.; and Süsstrunk, S. 2016. Non-Parametric Blur Map Regression for Depth of Field Extension. *IEEE Transactions on Image Processing (TIP)*, 25(4): 1660–1673.

Fish, D.; Brinicombe, A.; Pike, R.; and Walker, J. 1995. Blind deconvolution by means of the Richardson–Lucy algorithm. *JOSA A*, 12: 58–65.

Garg, R.; Wadhwa, N.; Ansari, S.; and Barron, J. 2019. Learning Single Camera Depth Estimation Using Dual-Pixels. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 7627–7636.

Goodfellow, I. J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2014. Generative Adversarial Nets. In *Proceedings of the 27th International Conference on Neural Information Processing Systems (NeurIPS)*, 2672–2680. Cambridge, MA, USA: MIT Press.

Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; and Courville, A. 2017. Improved Training of Wasserstein GANs. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS)*, 5769–5779. Red Hook, NY, USA: Curran Associates Inc.

Isola, P.; Zhu, J.-Y.; Zhou, T.; and Efros, A. A. 2017. Image-to-Image Translation with Conditional Adversarial Networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 5967–5976.

Johnson, J.; Alahi, A.; and Fei-Fei, L. 2016. Perceptual Losses for Real-Time Style Transfer and Super-Resolution. In *Computer Vision – ECCV 2016*, 694–711. Cham: Springer International Publishing.

Kang, M.; Choe, J.; Ha, H.; Jeon, H.-G.; Im, S.; and Kweon, I. S. 2021. Facial Depth and Normal Estimation using Single Dual-Pixel Camera. arXiv:2111.12928.

Karaali, A.; and Jung, C. R. 2018. Edge-Based Defocus Blur Estimation With Adaptive Scale Selection. *IEEE Transactions on Image Processing (TIP)*, 27(3): 1126–1137.

Krishnan, D.; and Fergus, R. 2009. Fast Image Deconvolution Using Hyper-Laplacian Priors. In *Proceedings of the 22nd International Conference on Neural Information Processing Systems (NeurIPS)*, 1033–1041. Red Hook, NY, USA: Curran Associates Inc.

Kupyn, O.; Budzan, V.; Mykhailych, M.; Mishkin, D.; and Matas, J. 2018. DeblurGAN: Blind Motion Deblurring Using Conditional Adversarial Networks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 8183–8192.

Kupyn, O.; Martyniuk, T.; Wu, J.; and Wang, Z. 2019. DeblurGAN-v2: Deblurring (Orders-of-Magnitude) Faster and Better. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 8877–8886.

Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; and Shi, W. 2017. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 105–114.

Lee, J.; Lee, S.; Cho, S.; and Lee, S. 2019. Deep Defocus Map Estimation Using Domain Adaptation. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 12214–12222.

Lee, J.; Son, H.; Rim, J.; Cho, S.; and Lee, S. 2021. Iterative Filter Adaptive Network for Single Image Defocus Deblurring. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2034–2042.

Liang, P.; Jiang, J.; Liu, X.; and Ma, J. 2022. BaMBNet: A Blur-Aware Multi-Branch Network for Dual-Pixel Defocus Deblurring. *IEEE/CAA Journal of Automatica Sinica*, 9(5): 878–892.

Lin, T.-Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; and Belongie, S. 2017. Feature Pyramid Networks for Object Detection. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 936–944.

Motwani, T.; and Parmar, M. 2020. A Novel Framework for Selection of GANs for an Application. arXiv:2002.08641.

Pan, L.; Chowdhury, S.; Hartley, R.; Liu, M.; Zhang, H.; and Li, H. 2021. Dual Pixel Exploration: Simultaneous Depth Estimation and Image Restoration. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 4338–4347.

Park, J.; Tai, Y.-W.; Cho, D.; and Kweon, I. S. 2017. A Unified Approach of Multi-scale Deep and Hand-Crafted Features for Defocus Estimation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2760–2769.

Punnappurath, A.; Abuolaim, A.; Afifi, M.; and Brown, M. S. 2020. Modeling Defocus-Disparity in Dual-Pixel Sensors. In *2020 IEEE International Conference on Computational Photography (ICCP)*, 1–12.

Punnappurath, A.; and Brown, M. S. 2019. Reflection Removal Using a Dual-Pixel Sensor. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1556–1565.

Shi, J.; Xu, L.; and Jia, J. 2015. Just noticeable defocus blur detection and estimation. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 657–665.

Simonyan, K.; and Zisserman, A. 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. *CoRR*, abs/1409.1556.

Son, H.; Lee, J.; Cho, S.; and Lee, S. 2021. Single Image Defocus Deblurring Using Kernel-Sharing Parallel Atrous Convolutions. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2622–2630.

Vo, T. 2021. Attention! Stay Focus! In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 479–486.

Wadhwa, N.; Garg, R.; Jacobs, D. E.; Feldman, B. E.; Kanazawa, N.; Carroll, R.; Movshovitz-Attias, Y.; Barron, J. T.; Pritch, Y.; and Levoy, M. 2018. Synthetic Depth-of-Field with a Single-Camera Mobile Phone. *ACM Trans. Graph.*, 37(4).

Wu, X.; Zhou, J.; Liu, J.; Ni, F.; and Fan, H. 2021. Single-Shot Face Anti-Spoofing for Dual Pixel Camera. *IEEE Transactions on Information Forensics and Security*, 16: 1440–1451.

Xin, S.; Wadhwa, N.; Xue, T.; Barron, J. T.; Srinivasan, P. P.; Chen, J.; Gkioulekas, I.; and Garg, R. 2021. Defocus Map Estimation and Deblurring from a Single Dual-Pixel Image. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2208–2218.

Yi, X.; and Eramian, M. 2016. LBP-Based Segmentation of Defocus Blur. *IEEE Transactions on Image Processing (TIP)*, 25(4): 1626–1638.

Zhang, K.; Ren, W.; Luo, W.; Lai, W.-S.; Stenger, B.; Yang, M.-H.; and Li, H. 2022. Deep Image Deblurring: A Survey. arXiv:2201.10700.

Zhang, Y.; Wadhwa, N.; Orts-Escolano, S.; Häne, C.; Fanello, S.; and Garg, R. 2020. Du2Net: Learning Depth Estimation from Dual-Cameras and Dual-Pixels. In *Computer Vision – ECCV 2020*, 582–598. Cham: Springer International Publishing.