United Defocus Blur Detection and Deblurring via Adversarial Promoting Learning

Wenda Zhao[1]*, Fei Wei[1], You He[2], and Huchuan Lu[1]

[1] Dalian University of Technology, Dalian, China
{zhaowenda, lhchuan}@dlut.edu.cn; fwei@mail.dlut.edu.cn
[2] Naval Aviation University, Yantai, China
youhe_nau@163.com

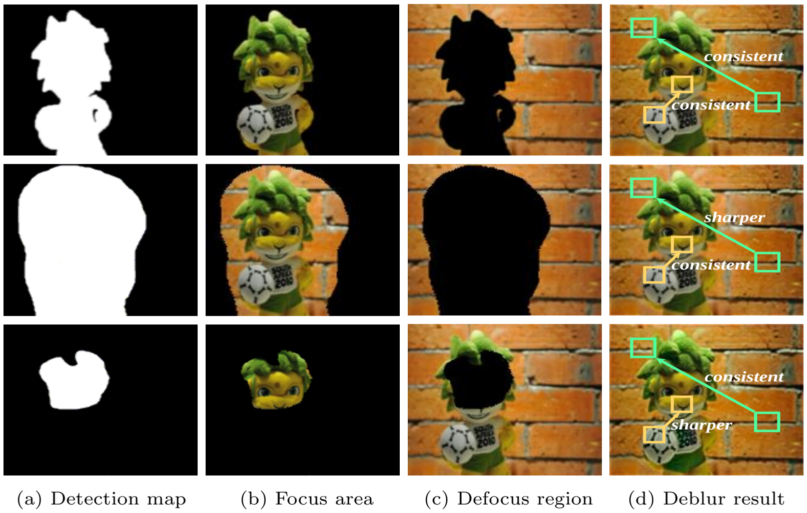
Abstract. Understanding blur from a single defocused image contains two tasks of defocus detection and deblurring. This paper makes the earliest effort to jointly learn both defocus detection and deblurring without using pixel-level defocus detection annotation and paired defocus deblurring ground truth. We build on the observation that these two tasks are supplementary to each other: Defocus detection can segment the focused area from the defocused image to guide the defocus deblurring; Conversely, to achieve better defocus deblurring, an accurate defocus detection as the guide is essential. Therefore, we implement an adversarial promoting learning framework to jointly handle defocus detection and defocus deblurring. Specifically, a defocus detection generator G_{ws} is implemented to represent the defocused image as a layered composition of two elements: defocused image I_{df} and a focused image I_f. Then, I_{df} and I_f are fed into a self-referenced defocus deblurring generator G_{sr} to generate a deblurred image. Two generators of G_{ws} and G_{sr} are optimized alternately in an adversarial manner against a discriminator D with unpaired realistic fully-clear images. Thus, G_{sr} will produce a deblurred image to fool D, and G_{ws} is forced to generate an accurate defocus detection map to effectively guide G_{sr}. Comprehensive experiments on two defocus detection datasets and one defocus deblurring dataset demonstrate the effectiveness of our framework. Code and model are available at: https://github.com/wdzhao123/APL.

Keywords: Defocus blur detection; Defocus deblurring; Adversarial promoting learning

1 Introduction

Defocus blur is common in an image that is captured under optical imaging systems. Detecting defocus blur can provide important clues for various scene understanding, such as salient region detection [9] and depth estimation [8] [21]. Sequentially, defocus deblurring is of great interest for the downstream computer vision tasks, such as object segmentation [17] [16] and tracking [7] [6], *etc.* Thus,

* Corresponding author



(a) Detection map (b) Focus area (c) Defocus region (d) Deblur result

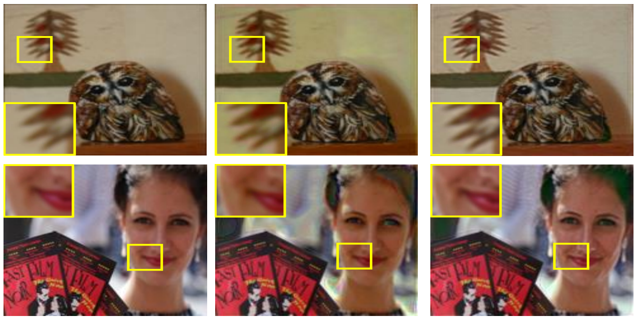
Fig. 1. Correlation illustration of defocus detection and deblurring. First row: An accurate focus detection can effectively segment the focus area and defocus region, thereby generating a natural deblurred image with consistent clarity. Second row: Excessive focus detection makes deblurred image still contain blurred area. Third row: Deficient focus detection makes deblurred image be not consistent in clarity, *e.g.*, over-sharpened focus area.

developing an efficient method for simultaneous defocus detection and deblurring is desirable. However, existing researches handle these two tasks separately.

Previous defocus detection methods can be mainly divided into two categories. One is prior knowledge-based methods [18] [22] [26] [31] [37], *e.g.*, gradient [31] [2] [10] and contrast [26] [18] [27]. Since the priors may dissatisfy some complicated scenes, the performance of defocus detection cannot be guaranteed. The other one is deep learning-based methods [44] [43] [30] [11] [32] [47] [49]. Their effectiveness usually relies on fully-supervised training with pixel-level annotation whose acquisition is time-consuming and expensive.

Existing defocus deblurring researches commonly compute a defocus map to guide the deblurring [41,23], where the defocus map is estimated through synthetic defocus image [13] or utilizing some prior knowledge, *e.g.*, edge [10] [19]. However, synthetic data results in a domain gap and prior knowledge has the scene dependency, which will hinder the performance of defocus deblurring. Recently, methods [1,14,25] propose end-to-end deep learning frameworks for defocus deblurring. Unfortunately, they are limited by the requirement of paired pixel-level ground truth.

Essentially, defocus detection and deblurring are supplementary to each other, as shown in Figure 1. Blur detection can guide deblur generator to achieve defocusing. Sequentially, deblur generator can build the bridge between blur detection generator and discriminator to finetune defocus detection in an adversarial manner. Therefore, we explore the joint learning of defocus detection and defocus deblurring, and propose an adversarial promoting learning frame-



(a) Defocused image (b) Native GAN (c) Self-reference GAN

Fig. 2. Qualitative comparison of different defocus deblurring solutions. Native GAN produces hallucination (*e.g.*, blur and color distortion). We address this through using focused area to provide a clarity reference that guides the defocused area to deblur from the defocused image itself.

work (MPLF) to tackle the problems of training defocus detection model with pixel-level annotation and learning defocus deblurring with paired ground truth. Specifically, MPLF includes three models: a defocus detection generator G_{ws}, a self-referenced defocus deblurring generator G_{sr} and a discriminator D. G_{ws} is implemented to generate a defocus detection map, and then the focused area and unfocused region are segmented from the defocused image to feed into G_{sr}. Two generators of G_{ws} and G_{sr} are optimized alternately in an adversarial manner against a discriminator D with unpaired realistic fully-clear images. Through this adversarial process, G_{sr} will produce a deblurred image to fool D to believe that the deblurred image is a natural fully-clear image, and G_{ws} is forced to produce an accurate defocus detection map without pixel-level supervision.

In particular, a potential solution is to implement generative adversarial networks (GAN) [12] [3] to overcome the dependency on paired data in defocus deblurring. However, purely feeding an unpaired full-focused image to the discriminator often cannot optimize the generator well in the adversarial process, which easily degrades the deblurring image, *e.g.*, blur and color distortion (see Figure 2(b)). Therefore, we design a self-referenced GAN that utilizes the focused area in the defocused image to guide the defocus region to deblur. Specifically, the defocused image is firstly represented as a layered composition of two elements: a defocused image I_{df} and a focused image I_f. Then, we build the self-referenced generator G_{sr} to deblur I_{df} with the reference of I_f. However, directly combining the defocused region with focused area as inputs, or concatenating focused area features with intermediate deep features can hardly make an efficient utilization of the focused area information. To address this problem, we propose an unpaired feature affine transformation model (UFAT) to recursively insert into G_{sr}. In UFAT, the focused area contents are referenced by influencing the feature affine transformation of the defocused region in the process of deblurring, thereby achieving better deblurring performance (see Figure 2(c)).

In short, our contributions are as follows. 1) We make the earliest effort to explore the joint learning of defocus detection and defocus deblurring, fully utilizing their mutual promotion to obtain superior performances on these two tasks. 2) We propose an adversarial promoting learning framework to produce defocus detection in a weakly-supervised fashion, while generating a defocus deblurred image without using paired ground truth. 3) We validate the effectiveness of the proposed method on two defocus detection datasets and one defocus deblurring dataset.

2 Related Work

Fully-supervised defocus detection. Benefitting from defocus detection datasets [22] [48] [47] with pixel-level annotation, deep convolutional neural networks-based methods [29,28,32,30,11,40,15] have been proposed to boost the performance of defocus detection. Among these methods, a main research route is multi-level feature integration. For example, Kim *et al*. [11] adopt long skip connections between encoder features and decoder features to combine multi-level contextual features. Tang *et al*. [32] implement a cross-layer feature fusion strategy to improve performance. Zhao *et al*. [43] design an image-scale-symmetric cooperative network to fuse multi-scale and multi-level features. In addition, some other mechanisms are effectively applied to defocus detection, such as ensemble network [44,49], cut-and-paste strategy [39] and depth distillation [5]. However, these methods are trained with abundant pixel-level ground truth whose acquisition is expensive and time-consuming. Thus, Zhao *et al*. [45] propose a weakly-supervised recurrent constraint network for focus region detection, where bounding box annotations are used.

Unsupervised defocus detection. Unsupervised defocus detection methods are usually concentrated on designing hand-crafted features [42,18,31]. For instance, Shi *et al*. [22] study a few blur feature representations, such as gradient, Fourier domain, and data-driven local filters. Golestaneh *et al*. [2] explore sorted transform coefficients of gradient magnitudes and multiscale fusion strategy. Yi *et al*. [37] adopt local binary patterns (LBP) to measure defocus blur. Hand-crafted features-based methods provide some efficient priors of understanding defocus blur, which may help us further design unsupervised or weakly-supervised deep defocus detection models.

Defocus deblurring. On one hand, defocus deblurring methods [41,10] are concerned on estimating a defocus detection map, and then utilize a non-blind deconvolution technology to achieve deblurring. Shi *et al*. [23] establish the correspondence between sparse edge representation and blur strength to obtain defocus detection map. Park *et al*. [19] combine multi-scale deep and hand-crafted features for defocus estimation. Lee *et al*. [13] build synthetically blurred images with paired ground truth and implement domain adaptation to generate defocus blur maps of real defocused images. On the other hand, works [1,14,25] propose end-to-end defocus deblurring network which are trained with paired pixel-level ground truth.

In contrast, we focus on designing a weakly-supervised defocus deblurring framework without using paired pixel-level ground truth. Particularly, we adopt the focused area directly segmented from the defocused image itself as a reference to guide the defocus to deblur.

GAN-based deblurring. GAN has achieved impressive results in various vision tasks, such as image inpainting [36], shadow removal [34], image denoising [4] and image super-resolution [3]. The main idea is using an adversarial loss with a targeted image that forces the generated image to be high-quality. This provides a potential solution that implements GAN to overcome the dependency on paired data in defocus deblurring. However, purely adopting an unpaired full-focused image as the target in the adversarial process will produce hallucination (see Figure 2(b)).

Different from existing GAN-based low-level image processing methods, we design a self-referenced GAN that utilizes the focused area segmented by defocus detection as a guide to optimize the defocus deblurring generator better, as shown in Figure 2(c). Interestingly, this builds a bridge between defocus detection and defocus deblurring, allowing us to alternately optimize them in an adversarial manner, thereby producing an accurate defocus detection map without using pixel-level annotation.

3 Adversarial Promoting Learning

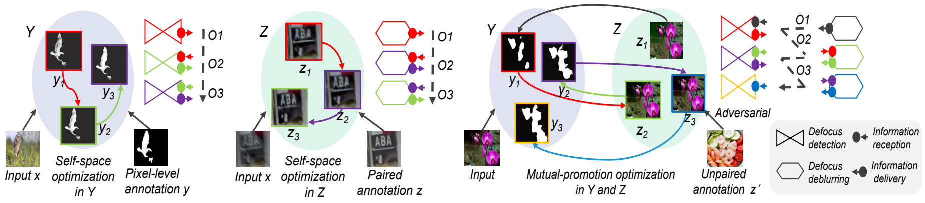
3.1 Motivation and Framework

Existing defocus detection and deblurring methods [48,33] usually train deep networks by recursive strategy in a fully-supervised manner. Let's denote the space of defocused images by $\mathcal{X}$, the space of defocus detection by $\mathcal{Y}$, and the space of deblurring images by $\mathcal{Z}$. Given an input defocused image $x \in \mathcal{X}$, defocus detection or deblurring aims to generate its corresponding detection map $y \in \mathcal{Y}$ or deblurring image $z \in \mathcal{Z}$. Most of the recursive strategy based methods iteratively optimize defocus detection mapping function $\varPhi$ or deblurring mapping function $\varPsi$, *i.e.*,

$$\varPhi(x, y; \phi_1) \rightarrow \varPhi(x, y_1, y; \phi_2) \rightarrow \varPhi(x, y_2, y; \phi_3) \cdot\cdot\cdot, \tag{1}$$

$$\varPsi(x, z; \psi_1) \rightarrow \varPsi(x, z_1, z; \psi_2) \rightarrow \varPsi(x, z_2, z; \psi_3) \cdot\cdot\cdot, \tag{2}$$

where $\{y_1, y_2, \cdots\}$ and $\{z_1, z_2, \cdots\}$ are the subspaces of detection maps and deblurring images, and $\{\phi_1, \phi_2, \cdots\}$ and $\{\psi_1, \psi_2, \cdots\}$ are the parameters of different optimization times of $\varPhi$ and $\varPsi$, respectively. As illustrated in Figure 3(a)-(b), defocus detection and deblurring optimize their mapping functions in their own space, lacking communication with each other. Moreover, they train deep networks in a fully-supervised manner with pixel-level detection annotation or paired deblurring ground truth. The GAN methods [12,3,46] can partially dilute this issue through adversarial training with unpaired data. However, when the scene is complex, GAN methods can hardly produce clear deblurring images with realistic details (Figure 2(b)). Therefore, [46] adds an adversarial discriminator



(a) Recursive detection (b) Recursive deblurring (c) Adversarial promoting framework

Fig. 3. Illustrations of our motivation and framework. Existing defocus detection and deblurring methods [48,33] recursively optimize results (*e.g.*, O1, O2 and O3) in their own space via a fully-supervised fashion (see (a) and (b)). In contrast, we utilize their mutual benefits to propose an adversarial promoting learning framework in a weakly-supervised manner without using pixel-level detection annotation and paired deblurring label (see (c)).

and a classifier to assist in network optimization. But the parameter search space becomes large and the training is difficult to converge.

Different from previous methods [48,46,33,1,14,25], we design an adversarial promoting learning framework to handle defocus detection and deblurring tasks jointly. As is presented in Figure 3(c), a defocus detection generator G_{ws} is implemented to generate defocus detection maps and a self-referenced deblurring generator G_{sr} is built to produce deblurring images. G_{ws} and G_{sr} are optimized alternately in an adversarial manner against a discriminator D with unpaired realistic fully-clear images, *i.e.*,

$$\begin{aligned} &G_{ws}(x, z_1, z'; g^1_{ws}) \to G_{sr}(x, y_1, z'; g^2_{sr}) \to \\ &G_{ws}(x, z_2, z'; g^2_{ws}) \to G_{sr}(x, y_2, z'; g^3_{sr}) \to \\ &\qquad\qquad G_{ws}(x, z_3, z'; g^3_{ws}) \cdots, \end{aligned} \tag{3}$$

where $\{g^1_{ws}, g^2_{ws}, g^3_{ws}, \cdots\}$ and $\{g^1_{sr}, g^2_{sr}, g^3_{sr}, \cdots\}$ are the parameters of different optimization times of G_{ws} and G_{sr}, respectively. z' is the unpaired realistic fully-clear image. Therefore, G_{sr} gradually produces a deblurred image with the guidance of defocus detection map of G_{ws} in the adversarial process with unpaired fully-clear image, and G_{ws} is forced to find the accurate defocus detection map to effectively guide G_{sr}, where pixel-level detection annotation is not used. The network architecture of our model is shown in Figure 4, which will be explained in detail as follows.

3.2 Architecture

Our MPLF is built on the successful applications of GAN, which contains three network models: a defocus detection generator G_{ws}, a self-referenced defocus deblurring generator G_{sr} and a discriminator D.

Self-referenced deblurring generator G_{sr}. Unpaired GANs are easy to produce hallucination (see Figure 2(b)). One underlying reason is that the generator is under-constrained. In practice, we observe that the focused area in a

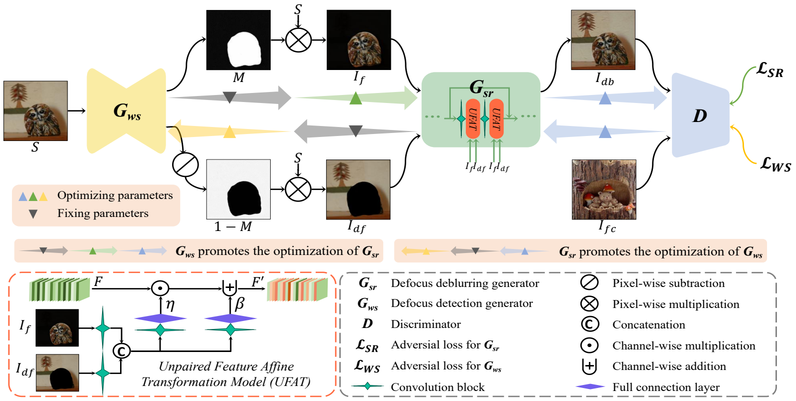

S
G_ws
M
S
I_f
1 − M
S
I_df
G_sr
UFAT
UFAT
I_f I_df I_f I_df
I_db
I_fc
D
ℒ_SR
ℒ_WS
Optimizing parameters
Fixing parameters
G_ws promotes the optimization of G_sr
G_sr promotes the optimization of G_ws
F
η
β
F′
I_f
I_df
Unpaired Feature Affine Transformation Model (UFAT)
G_sr Defocus deblurring generator
G_ws Defocus detection generator
D Discriminator
ℒ_SR Adversial loss for G_sr
ℒ_WS Adversial loss for G_ws
Convolution block
Pixel-wise subtraction
Pixel-wise multiplication
Concatenation
Channel-wise multiplication
Channel-wise addition
Full connection layer

Fig. 4. Architecture of the proposed adversarial promoting learning. Defocus detection generator G_{ws} is encouraged to produce an accurate defocus detection map M that guides G_{sr} to generate better defocus deblurred image. Sequentially, self-referenced generative model G_{sr} is built to utilize the focused area segmented by M from the input image itself for defocus deblurring. Generators G_{ws} and G_{sr} are alternately optimized in an adversarial manner against the discriminator D, where only unpaired realistic fully-clear images are used. Especially, unpaired feature affine transformation model (UFAT) is recursively inserted to G_{sr}, addressing the issue that unpaired GANs produce hallucination.

defocused image contains important information (*e.g.*, clarity degree), which can be adopted to assist defocus deblurring. Therefore, we design a self-referenced generative model G_{sr} to dilute this issue. Consider a defocused image I with size $h \times w \times c$, where h and w are the height and width, and c denotes the number of channels. We define the representation of I as a layered composition of two elements: a defocused image I_{df} and a focused image I_f. I_{df} and I_f are defined through a defocus detection map M with each element belonging to [0,1] as follows

$$I_{df} = (1 - M) \otimes I, I_f = M \otimes I, \tag{4}$$

where $\otimes$ expresses a pixel-wise multiplication operation.

Here, we focus on exploiting I_f to guide the defocus deblurring of I_{df}. However, directly combining I_f and I_{df}, or concatenating intermediate deep features of I_f and I_{df} can hardly obtain good performance (objective analysis is provided in Section 4.2). The potential cause is that their spatial contents are not aligned. Inspired by spatial feature transform [35], we introduce an unpaired feature affine transformation model (UFAT), where we extract feature affine transformation vectors η and β by consulting I_f to influence feature reconstruction of I_{df} in the defocus deblurring process.

$$UFAT(F, I_f, I_{df}) = \eta \odot F \uplus \beta, \tag{5}$$

where F represents a deep feature, $\odot$ and $\uplus$ stand for the channel-wise multiplication and addition operation, respectively.

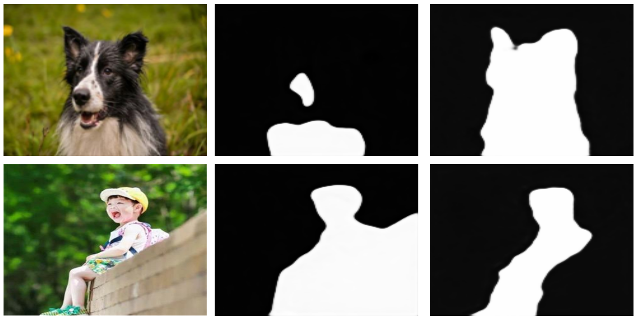


Fig. 5. Qualitative comparison of different optimization solutions for defocus detection.

Figure 4 illustrates the network architecture of UFAT. In particular, UFAT firstly obtains the concatenated features of I_f and I_{df}, and then generates feature affine transformation vectors η and β. Sequentially, η and β are used to help F restructure the feature F'. Further, we utilize UFAT to build our self-referenced generative model G_{sr}. Specifically, G_{sr} is structured with two residual convolution blocks, and UFAT is recursively inserted to help feature reconstruction of I_{df} in the defocus deblurring process. In addition, a global skip connection is implemented to ease the training of the deep network. Notice that the spatial contents of I_f and I_{df} are unaligned, thus UFAT does not generate spatial feature transform. This is different from [35] where spatial-wise transformation is implemented. Our self-referenced generative model achieves better deblurring performance compared with native GAN-based method (see Figure 2(c)). Objective analysis is provided in Section 4.2.

Defocus detection generator G_{ws} and discriminator D. G_{ws} is built to produce a defocus detection map M, which is used to calculate I_f and I_{df} to feed into G_{sr}. Inspired by the U-Net architecture [20], G_{ws} is designed by an encoder-decoder framework with skip connections. The encoder is built with the first four convolution blocks of VGG16 [24] to extract multi-level features. Then, a decoder including four corresponding deconvolution blocks to produce a defocus detection map.

Discriminator D is implemented to distinguish whether the defocus deblurring image is fully-clear, where the first three convolution blocks of VGG16 are used to extract high-level features. Then, three full connection layers are added to output a one-element vector.

3.3 Optimization

Generators G_{ws} and G_{sr} are optimized alternately in an adversarial manner against the discriminator D with unpaired realistic fully-clear images. The optimization loss of G_{sr} can be expressed as

$$\begin{aligned} \mathcal{L}_{SR} = \underset{g_{ws},g_{sr}}{min}\ \underset{d}{max} \mathbb{E}_{I_{fc}\sim P_{fc}} log D(I_{fc}; d) + \\ \mathbb{E}_{I\sim P_{df}} log(1 - D(G_{sr}(G_{ws}(I) \otimes I, \\ (1 - G_{ws}(I)) \otimes I); g_{ws}, g_{sr}))), \end{aligned} \tag{6}$$

where d is weight parameter of D. P_{fc} and P_{df} illustrate empirical distributions of fully-clear images and defocused images, respectively. I_{fc} stands for a fully-clear image.

Optimization starts with G_{sr} which is pretrained using a set of simulated defocus pairs (synthesis strategy is given in Section 4.1). Then, G_{ws} is optimized to produce an accurate defocus detection map to help G_{sr} achieve better defocus deblurring performance. However, only using adversarial supervision to train G_{ws} is extremely under-constrained, and G_{ws} easily produces a trivial solution (see Figure 5(b)). Thus, we utilize blur priors to provide an auxiliary supervision for training G_{ws} as

$$\mathcal{L}_{WS} = \mathcal{L}_{SR} + ||G_{ws}(I) - B(I)||_1, \tag{7}$$

where $B(I)$ is a blur prior model for the input image I. Here, we adopt local contrast knowledge (refer to [37] for details). Using this loss, G_{ws} can generate more accurate defocus detection maps, as shown in Figure 5(c).

3.4 Training Details

Our framework is implemented using Pytorch library on a NVIDIA RTX 2080Ti GPU. Adam with momentum 0.9 is adopted as the optimizer. The mini-batch size is taken to 1, and the learning rate is set to 0.0002. G_{sr}, D_1 and D are initialized with random values. We firstly optimize G_{sr} with simulated defocus pairs for 100 epochs. Then, we alternately optimize G_{ws} and G_{sr}, and each alternation is trained in an adversarial manner with D for 100 epochs. Considering memory capacity, we resize the image to 160 × 160 to verify the effectiveness of the proposed method.

4 Experiments

4.1 Configuration

Datasets. Two widely-used defocus detection datasets of CUHK [22] and DUT [48] are adopted to train and test our framework. The same strategy with [48] is implemented that training images and testing images are divided into 604 and 100 in CUHK, and 600 and 500 in DUT, respectively. It is worth noting that we train our framework without pixel-level defocus detection annotation. Besides, we utilize DP dataset [1] including 76 testing defocus images to evaluate the performance of defocus deblurring.

Table 1. Effect study of defocus detection in the mutual promotion process using F_{max} and MAE scores on both CUHK and DUT datasets. G_{ws}_On stands for the nth optimization of G_{ws}.

Setting	CUHK		DUT	
	F_{max}	MAE	F_{max}	MAE
Single G_{ws}	0.785	0.173	0.679	0.232
G_{ws}_O1	0.790	0.155	0.696	0.213
G_{ws}_O2	0.801	0.133	0.682	0.212
G_{ws}_O3	**0.831**	**0.125**	**0.722**	**0.196**

Table 2. Effect study of defocus deblurring in the mutual promotion process using $PSNR$, $SSIM$ and MAE scores on DP dataset. G_{sr}_On stands for the nth optimization of G_{sr}.

Setting	$PSNR$	$SSIM$	MAE
Single G_{sr}	22.02	0.782	0.063
G_{sr}_O1	24.05	0.785	0.050
G_{sr}_O2	24.99	0.821	0.044
G_{sr}_O3	**25.71**	**0.842**	**0.041**

In addition, to initialize the self-referenced generative defocus deblurring model, we construct a simulated defocus dataset. Specifically, we firstly collect 500 full-focused images with manifold scenes. Then, inspired by [38], we adopt a Gaussian filter with a standard deviation randomly sampled from 0.1 to 10 and window size 15 × 15 to blur a part (60%-70%) of each full-focused image. This process is repeated five times to produce 2500 defocused images.

Evaluations. We utilize two metrics of mean absolute error (MAE) and F-measure score (F_{max}) [48,32,49], to evaluate the performance of defocus detection. A smaller MAE demonstrates a more accurate result. A larger F_{max} indicates a better performance. Besides, peak signal to noise ratio ($PSNR$), structural similarity ($SSIM$) and MAE are adopted to measure defocus deblurring's performance [1]. A larger $PSNR$ or $SSIM$ stands for a better defocus deblurring.

4.2 Ablation Study

Adversarial promotion between defocus detection and deblurring. Our adversarial promoting learning handles defocus detection G_{ws} and deblurring G_{sr} jointly. Defocus detection and deblurring are optimized alternately. Specifically, defocus detection maps of G_{ws} guide G_{sr} to generate deblurred images. Conversely, to effectively guide G_{sr}, G_{ws} is forced to produce accurate defocus detection maps. We implement the following settings to demonstrate the validity of the mutual promotion learning.

Firstly, we study the effects of single detection and single deblurring, *i.e.*, the one task's performance if the other task fails. Single G_{ws}: Training G_{ws} with the supervision of the local contrast prior model [37]. Single G_{sr}: Training G_{sr} in a adversarial manner with D using unpaired full-focused images. Secondly, we

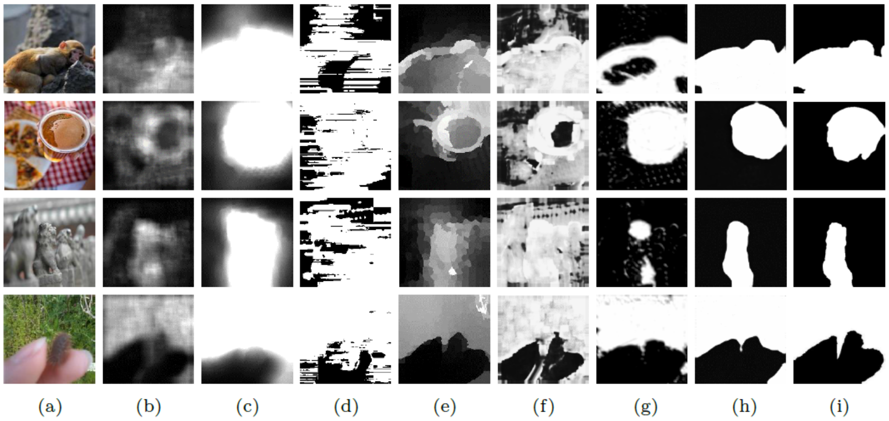


Fig. 6. Visual results of defocus detection produced by different methods. (a)-(i) are source, SVD [26], HiFST [2], KSFV [18], SS [31], DBDF [22], SGNet [46], Ours, and ground truth, respectively.

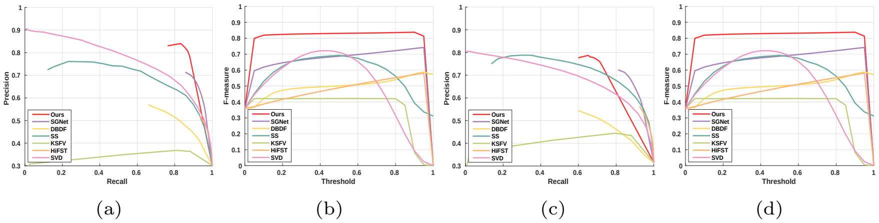
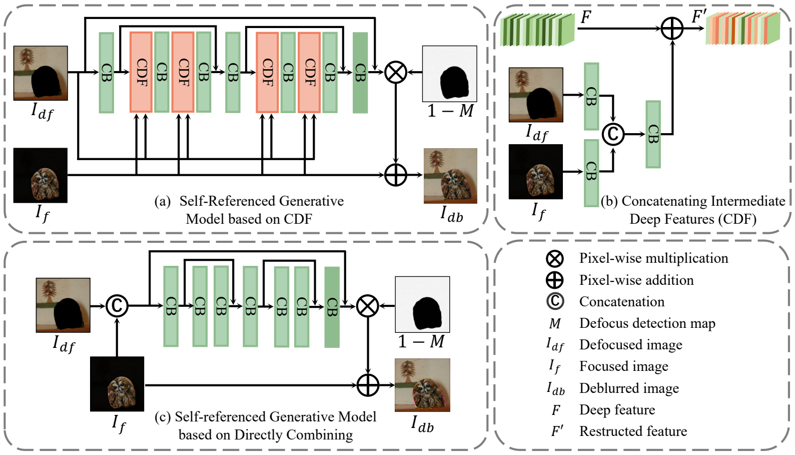


Fig. 7. Comparison of PR curves and F-measure curves of different methods on both CUHK and DUT datasets. (a)-(d) are PR curve on CUHK, F-measure curve on CUHK, PR curve on DUT and F-measure curve on DUT.

implement their adversarial promoting optimization, denoted as G_{ws}_On and G_{sr}_On $(n = 1, 2, 3)$.

Table 1 and Table 2 present objective results. Compared with single detection and single deblurring, our mutual promotion learning improves their performances on all measure scores. With the optimization number increases, the performances of G_{sr} and G_{ws} are promoted. Especially, G_{ws}_O3 and G_{sr}_O3 obtain the best measure scores respectively, improving F_{max}/MAE by 5.9%/27.7% and 6.3%/15.5% than single detection on CUHK and DUT datasets, and raising $PSNR/SSIM/MAE$ by 16.8%/7.7%/34.9% than single deblurring on DP dataset.

Self-referenced deblurring generator. In Section 3.2, we propose a self-referenced deblurring generator to relieve the problem that unpaired GANs are easy to produce hallucination. The core idea is to utilize the focused area I_f in the defocused image itself to guide the defocused region I_{df} to deblur. Structurally, we design UFATs to recursively insert into G_{sr}, making an efficient information utilization of the focused area, named as UFAT-GAN. Here, two aspects are studied to illustrate the effectiveness of the self-referenced deblurring generator.



(a) Self-Referenced Generative Model based on CDF

(b) Concatenating Intermediate Deep Features (CDF)

(c) Self-referenced Generative Model based on Directly Combining

Fig. 8. Illustrations of DC-based and CDF-based self-referenced generative defocus deblurring models. CB stands for a convolution block with three convolution layers.

Table 3. Importance study of self-referenced deblurring generator using $PSNR$, $SSIM$ and MAE scores on DP dataset.

Metric	U-GAN	DC-GAN	CDF-GAN	UFAT-GAN
$PSNR$	22.02	24.16	24.80	**25.71**
$SSIM$	0.782	0.808	0.838	**0.842**
MAE	0.063	0.045	0.043	**0.041**

On one hand, unpaired GAN is implemented for comparing where deblurring generator does not contain UFAT, denoted as U-GAN. On the other hand, U-FAT's two variants are compared: directly making a combination of I_f and I_{df}, and concatenating intermediate deep features of I_f and I_{df}, named as DC-GAN and CDF-GAN. Detailed network structures of DC-GAN and CDF-GAN are shown in Figure 8.

As can be seen in Table 3, implementing our self-referenced mechanism can improve performance, and DC-GAN outperforms U-GAN by 9.7%, 3.3% and 28.6% on $PSNR$, $SSIM$ and MAE, respectively. UFAT-GAN achieves the best performance, especially outperforming DC-GAN and CDF-GAN by 6.4% and 3.7% on $PSNR$. The underlying reason is that UFAT relieves the problem of unpaired spatial contents between I_f and I_{df} through two channel-wise attention vetors η and β.

4.3 Comparison with State-of-the-art Methods

Defocus detection. Our weakly-supervised defocus detection (Ours) is compared with the following five unsupervised state-of-the-art methods: singular value decomposition (SVD) [26], high-frequency multi-scale fusion and sort transform of gradient magnitudes (HiFST) [2], classifying discriminative features (KS-

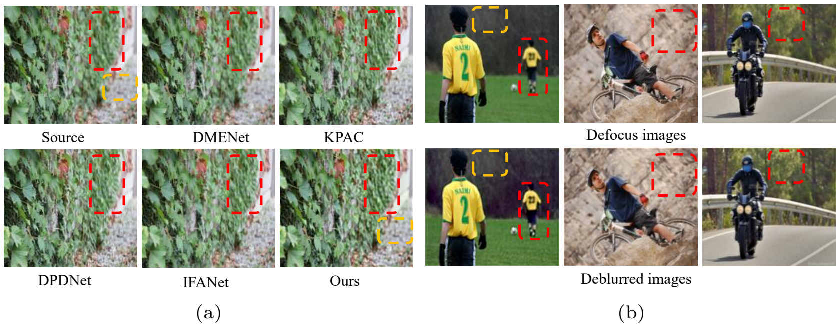


Fig. 9. Qualitative defocus deblurring results. (a) and (b) are visual comparison of different methods on DP and visual results of our method on CUHK and DUT. Our method achieves a uniform clarity in deblurred images.

Table 4. Defocus detection comparison with the state-of-the-art approaches using F_{max} and MAE scores on CUHK and DUT datasets. Average time is calculated on a workstation with a RTX 2080Ti GPU.

	Metric	SVD [26]	HiFST [2]	KSFV [18]	SS [31]	DBDF [22]	SGNet [46]	Ours
CUHK	F_{max}	0.764	0.583	0.420	0.759	0.626	0.732	**0.831**
	MAE	0.267	0.429	0.492	0.316	0.422	0.199	**0.125**
DUT	F_{max}	0.712	0.617	0.489	0.733	0.592	**0.731**	0.722
	MAE	0.288	0.399	0.404	0.320	0.454	0.204	**0.196**
Time	Second	2.153	17.93	4.937	0.336	31.55	0.007	**0.003**

FV) [18], spectral and spatial approach (SS) [31] and discriminative blur detection features (DBDF) [22]. Moreover, one weakly-supervised deep learning method of SGNet [46] is compared. To compare fairly, we utilize the available codes and recommended parameter settings released by authors.

Table 4 shows the quantitative defocus detection results. Ours outperforms the previous state-of-the-art results in general. Especially, ours achieves better performance of MAE by 37.2% and 3.9% than the second-best SGNet on CUHK and DUT, respectively. Besides, ours is highly efficient, which achieves the average testing time of 0.003s. Figure 7 shows their PR curves and F-measure curves, comprehensively verifying the better performance of our method. Qualitative comparison is shown in Figure 6, including of various scenes, such as complex background, unfocused foreground. Our method consistently generates defocus detection results closest to the ground truth.

Defocus deblurring. Our weakly-supervised self-referenced defocus deblurring model is compared with five state-of-the-art methods, including two defocus maps based methods of EBDB [10] and DMENet [13], and three fully-supervised methods DPDNet [1], IFANet [14] and KPAC [25]. For a fair comparison, we implement the networks with their released codes to produce results. Notably, DPDNet is implemented with center view images since its corresponding parameters are not released. The images with resolution of 160 × 160 are used based on the memory capacity.

Table 5. Quantitative comparison of different defocus deblurring methods using $PSNR$, $SSIM$ and MAE scores on DP dataset.

Metric	EBDB	DMENet	Ours	DPDNet	IFANet	KPAC
$PSNR$	23.89	24.99	**25.71**	24.01	24.20	26.51
$SSIM$	0.813	0.767	**0.842**	0.734	0.797	0.861
MAE	0.050	0.044	**0.041**	0.047	0.045	0.038

As shown in Table 5, our model achieves the best performance compared with weakly-supervised EBDB and DMENet. Moreover, our method obtains competitive performance compared with the fully-supervised methods, achieving the gaps of 0.8, 0.019 and 0.003 than the best $PSNR$, $SSIM$ and MAE, respectively. Figure 9 presents qualitative results of different defocus deblurring methods. Our method shows a more uniform clarity in any areas of a deblurred image.

4.4 Limitation

Our framework is implemented in weakly-supervised manner, and can achieve better performances on both defocus detection and deblurring of various defocus blurs. However, it may have limitations in addressing ambiguous boundaries (see the second row in Figure 6) and large blurs (see yellow dashed boxes in Figure 9). Maybe physics-based blur prior can relieve this issue, and we will study it.

5 Conclusion

We present an efficient joint learning framework for defocus detection and deblurring without using pixel-level defocus detection annotation and paired defocus deblurring ground truth. The core idea is utilizing their correlations to build an adversarial promoting learning framework in an adversarial manner. A self-referenced defocus deblurring generator G_{sr} is firstly proposed to obtain the ability of defocus deblurring. In particular, UFAT is designed to recursively insert into G_{sr}, relieving the unpaired spatial contents between the focused area and defocused region through influencing the feature affine transformation of the defocused region. Then, a defocus detection generator G_{ws} is introduced and combines with G_{sr} to jointly learn in an adversarial manner against a discriminator. Thus, G_{ws} is encouraged to produce an accurate defocus detection without pixel-level annotation. Extensive qualitative and quantitative experiment results on three datasets verify the effectiveness of our method.

Acknowledgements. This work is supported by National Natural Science Foundation of China under Grant Nos. 62176038 and U1903215, and Science and Technology Star of Dalian under Grant No. 2021RQ054.

References

1. Abuolaim, A., Brown, M.S.: Defocus deblurring using dual-pixel data. In: European Conference on Computer Vision. pp. 111–126. Springer (2020) 2, 4, 6, 9, 10, 13
2. Alireza Golestaneh, S., Karam, L.J.: Spatially-varying blur detection based on multiscale fused and sorted transform coefficients of gradient magnitudes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 5800–5809 (2017) 2, 4, 11, 12, 13
3. Bulat, A., Yang, J., Tzimiropoulos, G.: To learn image super-resolution, use a gan to learn how to do image degradation first. In: Proceedings of the European conference on computer vision (ECCV). pp. 185–200 (2018) 3, 5
4. Chen, J., Chen, J., Chao, H., Yang, M.: Image blind denoising with generative adversarial network based noise modeling. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3155–3164 (2018) 5
5. Cun, X., Pun, C.M.: Defocus blur detection via depth distillation. In: European Conference on Computer Vision. pp. 747–763. Springer (2020) 4
6. Ding, J., Huang, Y., Liu, W., Huang, K.: Severely blurred object tracking by learning deep image representations. IEEE Transactions on Circuits and Systems for Video Technology **26**(2), 319–331 (2015) 1
7. Guo, Q., Feng, W., Gao, R., Liu, Y., Wang, S.: Exploring the effects of blur and deblurring to visual object tracking. IEEE Transactions on Image Processing **30**, 1812–1824 (2021) 1
8. Gur, S., Wolf, L.: Single image depth estimation trained via depth from defocus cues. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 7683–7692 (2019) 1
9. Jiang, P., Ling, H., Yu, J., Peng, J.: Salient region detection by UFO: Uniqueness, focusness and objectness. In: IEEE International Conference on Computer Vision. pp. 1976–1983 (2013) 1
10. Karaali, A., Jung, C.R.: Edge-based defocus blur estimation with adaptive scale selection. IEEE Transactions on Image Processing **27**(3), 1126–1137 (2017) 2, 4, 13
11. Kim, B., Son, H., Park, S.J., Cho, S., Lee, S.: Defocus and motion blur detection with deep contextual features. In: Computer Graphics Forum. vol. 37, pp. 277–288 (2018) 2, 4
12. Kupyn, O., Martyniuk, T., Wu, J., Wang, Z.: Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 8878–8887 (2019) 3, 5
13. Lee, J., Lee, S., Cho, S., Lee, S.: Deep defocus map estimation using domain adaptation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 12222–12230 (2019) 2, 4, 13
14. Lee, J., Son, H., Rim, J., Cho, S., Lee, S.: Iterative filter adaptive network for single image defocus deblurring. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2034–2042 (2021) 2, 4, 6, 13
15. Li, J., Fan, D., Yang, L., Gu, S., Lu, G., Xu, Y., Zhang, D.: Layer-output guided complementary attention learning for image defocus blur detection. IEEE Transactions on Image Processing **30**, 3748–3763 (2021) 4
16. Luo, B., Cheng, Z., Xu, L., Zhang, G., Li, H.: Blind image deblurring via superpixel segmentation prior. IEEE Transactions on Circuits and Systems for Video Technology (2021) 1

17. Pan, L., Dai, Y., Liu, M., Porikli, F., Pan, Q.: Joint stereo video deblurring, scene flow estimation and moving object segmentation. IEEE Transactions on Image Processing **29**, 1748–1761 (2019) 1
18. Pang, Y., Zhu, H., Li, X., Li, X.: Classifying discriminative features for blur detection. IEEE Transactions on Cybernetics **46**(10), 2220–2227 (2015) 2, 4, 11, 13
19. Park, J., Tai, Y.W., Cho, D., So Kweon, I.: A unified approach of multi-scale deep and hand-crafted features for defocus estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1736–1745 (2017) 2, 4
20. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-assisted Intervention. pp. 234–241. Springer (2015) 8
21. Sakurikar, P., Narayanan, P.: Composite focus measure for high quality depth maps. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1614–1622 (2017) 1
22. Shi, J., Xu, L., Jia, J.: Discriminative blur detection features. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2965–2972 (2014) 2, 4, 9, 11, 13
23. Shi, J., Xu, L., Jia, J.: Just noticeable defocus blur detection and estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 657–665 (2015) 2, 4
24. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556 pp. 1–14 (2014) 8
25. Son, H., Lee, J., Cho, S., Lee, S.: Single image defocus deblurring using kernel-sharing parallel atrous convolutions. In: Proceedings of the IEEE International Conference on Computer Vision (2021) 2, 4, 6, 13
26. Su, B., Lu, S., Tan, C.L.: Blurred image region detection and classification. In: Proceedings of the 19th ACM International Conference on Multimedia. pp. 1397–1400 (2011) 2, 11, 12, 13
27. Tai, Y.W., Brown, M.S.: Single image defocus map estimation using local contrast prior. In: 2009 16th IEEE International Conference on Image Processing (ICIP). pp. 1797–1800. IEEE (2009) 2
28. Tang, C., Liu, X., An, S., Wang, P.: BR^{2}Net: Defocus blur detection via bidirectional channel attention residual refining network. IEEE Transactions on Multimedia (2020) 4
29. Tang, C., Liu, X., Zheng, X., Li, W., Xiong, J., Wang, L., Zomaya, A., Longo, A.: DeFusionNET: Defocus blur detection via recurrently fusing and refining discriminative multi-scale deep features. IEEE Transactions on Pattern Analysis and Machine Intelligence **PP**(99), 1–1 (2020) 4
30. Tang, C., Liu, X., Zhu, X., Zhu, E., Sun, K., Wang, P., Wang, L., Zomaya, A.: R^{2}MRF: Defocus blur detection via recurrently refining multi-scale residual features. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 34, pp. 12063–12070 (2020) 2, 4
31. Tang, C., Wu, J., Hou, Y., Wang, P., Li, W.: A spectral and spatial approach of coarse-to-fine blurred image region detection. IEEE Signal Processing Letters **23**(11), 1652–1656 (2016) 2, 4, 11, 13
32. Tang, C., Zhu, X., Liu, X., Wang, L., Zomaya, A.: Defusionnet: Defocus blur detection via recurrently fusing and refining multi-scale deep features. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 2700–2709 (2019) 2, 4, 10

33. Tao, X., Gao, H., Shen, X., Wang, J., Jia, J.: Scale-recurrent network for deep image deblurring. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 8174–8182 (2018) 5, 6
34. Wang, J., Li, X., Yang, J.: Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1788–1797 (2018) 5
35. Wang, X., Yu, K., Dong, C., Loy, C.C.: Recovering realistic texture in image super-resolution by deep spatial feature transform. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 606–615 (2018) 7, 8
36. Xiong, W., Yu, J., Lin, Z., Yang, J., Lu, X., Barnes, C., Luo, J.: Foreground-aware image inpainting. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5840–5848 (2019) 5
37. Yi, X., Eramian, M.: LBP-based segmentation of defocus blur. IEEE Transactions on Image Processing **25**(4), 1626–1638 (2016) 2, 4, 9, 10
38. Zhang, K., Zuo, W., Zhang, L.: Learning a single convolutional super-resolution network for multiple degradations. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3262–3271 (2018) 10
39. Zhang, N., Yan, J.: Rethinking the defocus blur detection problem and a real-time deep dbd model. In: European Conference on Computer Vision. pp. 617–632. Springer (2020) 4
40. Zhang, S., Shen, X., Lin, Z., Mech, R., Costeira, J.P., Moura, J.M.F.: Learning to understand image blur. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 6586–6595 (2018) 4
41. Zhang, X., Wang, R., Jiang, X., Wang, W., Gao, W.: Spatially variant defocus blur map estimation and deblurring from a single image. Journal of Visual Communication and Image Representation **35**, 257–264 (2016) 2, 4
42. Zhang, Z., Liu, Y., Xiong, Z., Li, J., Zhang, M.: Focus and blurriness measure using reorganized dct coefficients for an autofocus application. IEEE Transactions on Circuits and Systems for Video Technology **28**(1), 15–30 (2016) 4
43. Zhao, F., Lu, H., Zhao, W., Yao, L.: Image-scale-symmetric cooperative network for defocus blur detection. IEEE Transactions on Circuits and Systems for Video Technology **32**(5), 2719–2731 (2021) 2, 4
44. Zhao, W., Hou, X., He, Y., Lu, H.: Defocus blur detection via boosting diversity of deep ensemble networks. IEEE Transactions on Image Processing **30**, 5426–5438 (2021) 2, 4
45. Zhao, W., Hou, X., Yu, X., He, Y., Lu, H.: Towards weakly-supervised focus region detection via recurrent constraint network. IEEE Transactions on Image Processing **29**, 1356–1367 (2019) 4
46. Zhao, W., Shang, C., Lu, H.: Self-generated defocus blur detection via dual adversarial discriminators. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6933–6942 (2021) 5, 6, 11, 13
47. Zhao, W., Zhao, F., Wang, D., Lu, H.: Defocus blur detection via multi-stream bottom-top-bottom fully convolutional network. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 3080–3088 (2018) 2, 4
48. Zhao, W., Zhao, F., Wang, D., Lu, H.: Defocus blur detection via multi-stream bottom-top-bottom network. IEEE Transactions on Pattern Analysis and Machine Intelligence **42**(8), 1884–1897 (2020) 4, 5, 6, 9, 10
49. Zhao, W., Zheng, B., Lin, Q., Lu, H.: Enhancing diversity of defocus blur detectors via cross-ensemble network. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 8905–8913 (2019) 2, 4, 10