

# Defocus Blur Detection via Depth Distillation (Supplementary Material)

Xiaodong Cun and Chi-Man Pun

University of Macau, Macau, China  
`{yb87432,cmpun}@umac.mo`

## Comparison Between Reception Field Block and the proposed Selective Reception Field Block

We propose a Selective Reception Field Block (SRFB) for defocus blur detection. Here, we provide the details of the differences between Reception Field Block (RFB) [5] and the proposed block. First, our block is especially designed for DBD while theirs is used in general object detection. Second, their block aims to encode features for detection, while ours is designed as decoder in FCN. Third, as shown in Figure. 1, their block is an inception-like structure, while ours only uses a similar idea(increasing the dilation rate and kernel size at the same time) to build feature pyramids.

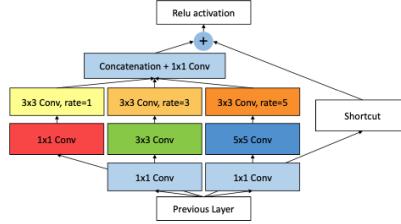


Fig. 1: The architecture of RFB.

## Evaluation of Distilled Depth

Although depth estimation is not our main target and our network can only predict the depth for partial defocus images, we still compare the distilled depth with our teacher network (Chen *et al.* [1]). We plot some visual comparisons in Fig. 2. It is obvious that original results in Chen *et al.* [1] are sensitive to color (first and third examples) because their network only trains on sparse points. Differently, our network can generate more convincing relative depth between the blurry and sharp region. It is because the distilled depth of our framework will consider the sharp, DOF region to have a similar depth. For there is no ground truth depth in the defocus dataset, we conduct the subjective experiments between ours and Chen *et al.* [1]. Similar to the user study experiments in Hu *et al.* [3], we randomly choose 30 samples from each method on DUT dataset and ask 50 people (both male and female, aging 20-28) to identify which result is more convincing. We collect the numbers of positive opinion on

our method from each participant and show the histogram of opinions in Fig. 3, where 66% participants prefer our results to those of Chen *et al.* [1]

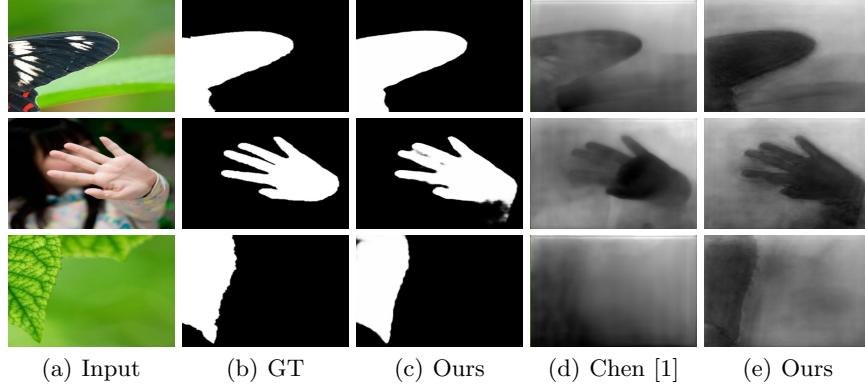


Fig. 2: Comparison of relative depth on DUT dataset. Our depth gain the information from defocus detection which is robust to low-level features (such as color).

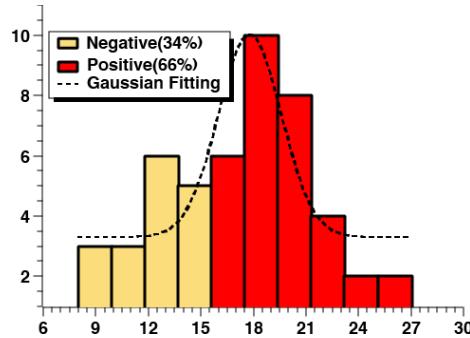


Fig. 3: User study. We plot the votes histogram of each participate. The  $x$ -axis means the count of positive choice for each individual. The positive choice is defined as: in each control group of the questionnaire, they think our depth is more convincing. The  $y$ -axis means the distribution (histogram) of all the participants. There are 30 samples in the subjective study and when the positive choice is larger than 15, we consider this participant prefers our method. Thus, in above figure, 66% of the participant choice our method.

#### More Details of Decoder

As shown in the decoder Figure of the main paper (marking the original feature channels equal to  $x$ ), in the first CONV( $3 \times 3$ )-BN-ReLU block, we reduce the channels of feature map to  $\frac{1}{4}x$  for efficient memory usage. The second CONV( $1 \times 1$ )-BN-ReLU increases the channels of feature map to  $\frac{1}{2}x$  for meeting the requirements of next decoder.

We build 5 branch SRFB (4 different kernels: 1x1 Conv. (dilation=1), 3x3 Conv. (dilation=3), 5x5 Conv. (dilation=5), 7x7 Conv. (dilation=7), the original feature) in all the experiments. In each SRFB, we use the same default parameters for attention as SK-Block.

As for SAB, firstly, two CONV( $1 \times 1$ ) blocks are mapping the features to side depth map and side defocus map. These side outputs need the supervision from the ground truth or depth distillation. Then, we concatenate them into 2 channel features, and feed it into two convolutional blocks for attention map generation. Here, the first CONV( $3 \times 3$ )-BN-ReLU increases the channel number of concatenated side outputs to 32 while the second one reduces it to a single feature map for  $Sigmoid(\cdot)$ .

#### More Results on Defocus Blur Detection

We plot more results of ours and other state-of-the-art DBD methods in Figure. 4. Besides, we show some results of ours compared with state-of-the-art methods from related tasks (Shadow Detection and Salient Object Detection) in Figure. 5.

#### More Results on Depth Estimation

We compare the distilled depth from our network with Chen *et al.* [1] in the last two columns of Figure. 5. It is obvious that our method can predict the depth well in partial defocus images.

#### Side Outputs

We plot final result and all levels of side outputs in Figure. 6. Our prediction fusion block merges the side outputs using a  $1 \times 1$  convolutional block and the final result benefits from all the knowledge from side outputs.

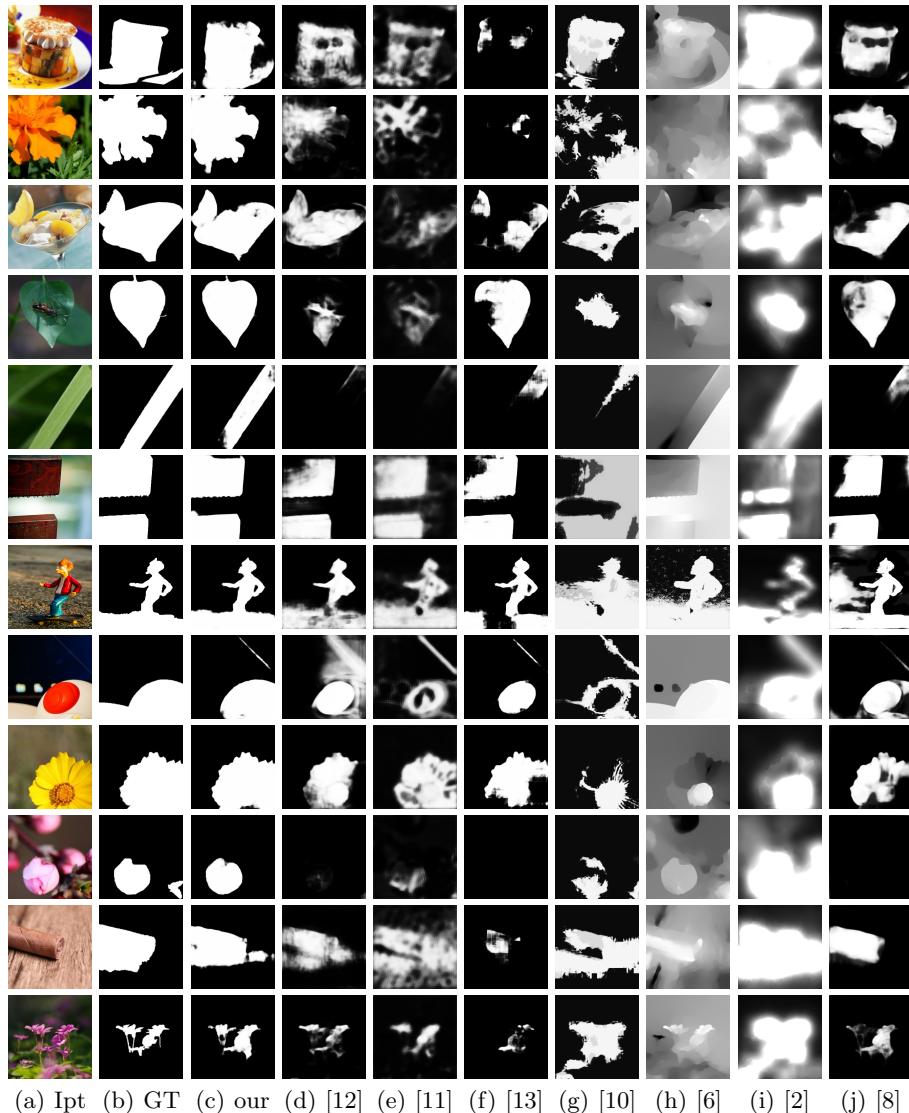


Fig. 4: More comparisons with states-of-the-art DBD methods from DUT and CUHK datasets. From left to right: (a)Input, (b)Target, (c)Ours, (d)BTBC [12], (e)BTBF [12], (f)CE [13], (g)LBP [10], (h)DHCF [6], (i)HiFST [2], (j)DF [8].

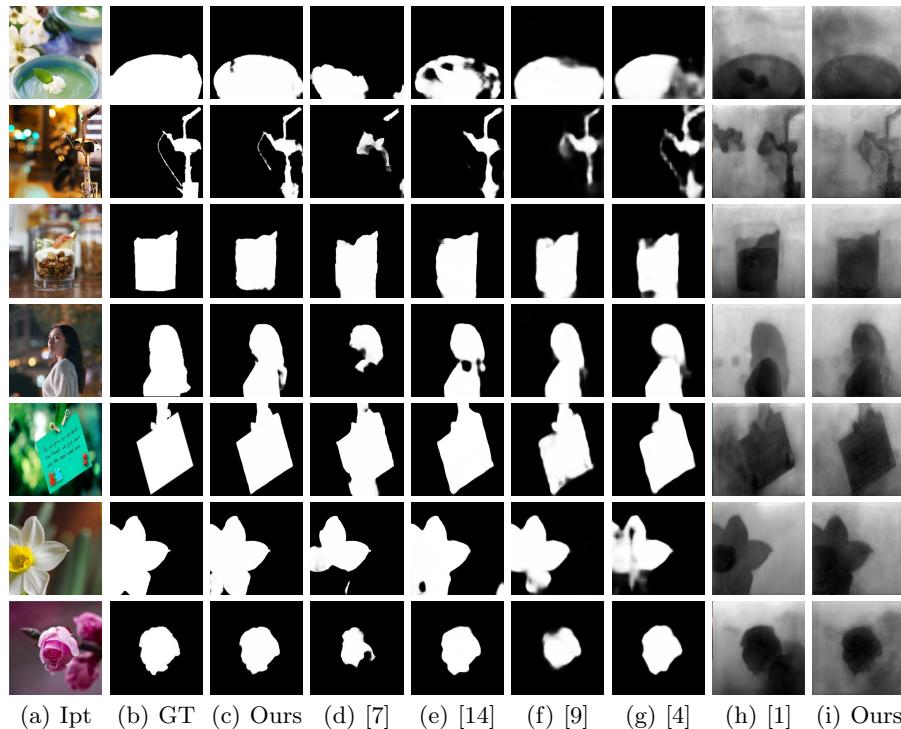


Fig. 5: More comparisons with states-of-the-art methods on relative tasks as illustrated in the main paper. Also, we plot more comparison on depth estimation in the last two columns. From left to right: (a)Input, (b)Target, (c)Our defocus mask, (d)BAS [7], (e)BDRAR [14], (f)CPD [9], (g)DSC [4], (h)Depth in Chen [1],(i)Our depth.

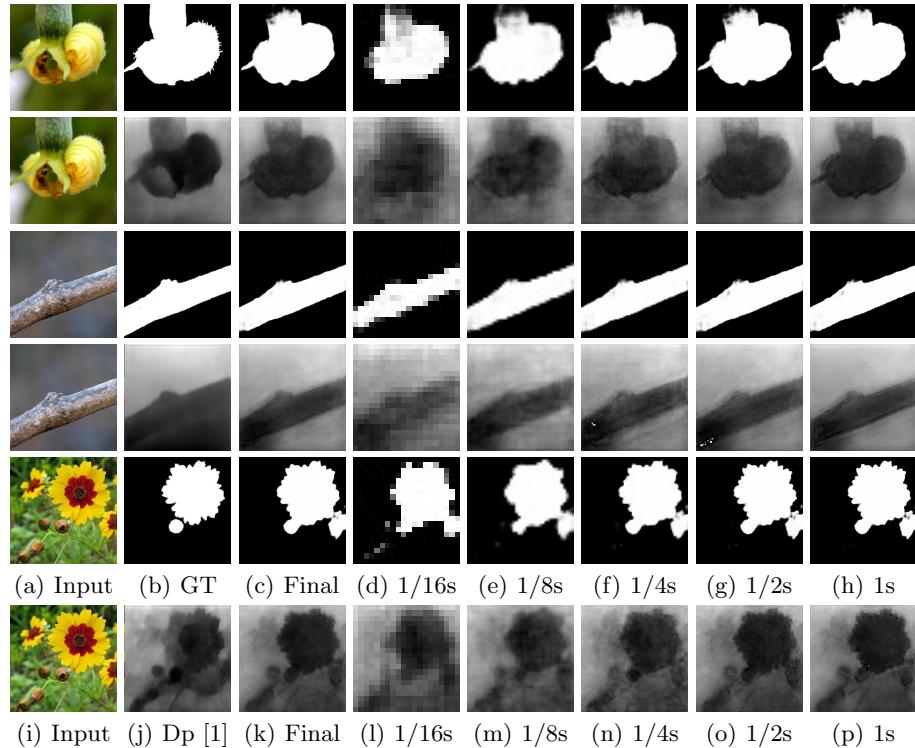


Fig. 6: More visualizations on side outputs. Our network fuses all the side outputs to generate the final result. We plot each level of the side defocus map from finer to coarser in (d)-(h) while each level of the side depth map from finer to coarser in (l)-(p). The  $1/ks$  means the resolution in this side output is  $1/k$  of the original image size.

## References

1. Chen, W., Fu, Z., Yang, D., Deng, J.: Single-image depth perception in the wild. In: NeurIPS. pp. 730–738 (2016)
2. Golestaneh, S.A., Karam, L.J.: Spatially-Varying Blur Detection Based on Multiscale Fused and Sorted Transform Coefficients of Gradient Magnitudes. CVPR (Mar 2017)
3. Hu, X., Jiang, Y., Fu, C.W., Heng, P.A.: Mask-shadowgan: Learning to remove shadows from unpaired data. ICCV (2019)
4. Hu, X., Zhu, L., Fu, C.W., Qin, J., Heng, P.A.: Direction-aware spatial context features for shadow detection. In: CVPR. pp. 7454–7462 (2018)
5. Liu, S., Huang, D., et al.: Receptive field block net for accurate and fast object detection. In: ECCV. pp. 385–400 (2018)
6. Park, J., Tai, Y.W., Cho, D., Kweon, I.S.: A Unified Approach of Multi-scale Deep and Hand-Crafted Features for Defocus Estimation. CVPR **cs.CV** (2017)
7. Qin, X., Zhang, Z., Huang, C., Gao, C., Dehghan, M., Jagersand, M.: Basnet: Boundary-aware salient object detection. In: CVPR (June 2019)
8. Tang, C., Zhu, X., Liu, X., Wang, L., Zomaya, A.: Defusionnet: Defocus blur detection via recurrently fusing and refining multi-scale deep features. In: CVPR. pp. 2700–2709 (2019)
9. Wu, Z., Su, L., Huang, Q.: Cascaded partial decoder for fast and accurate salient object detection. In: CVPR. pp. 3907–3916 (2019)
10. Yi, X., Eramian, M.: Lbp-based segmentation of defocus blur. IEEE transactions on image processing **25**(4), 1626–1638 (2016)
11. Zhao, W., Zhao, F., Wang, D., Lu, H.: Defocus blur detection via multi-stream bottom-top-bottom fully convolutional network. In: CVPR. pp. 3080–3088 (2018)
12. Zhao, W., Zhao, F., Wang, D., Lu, H.: Defocus blur detection via multi-stream bottom-top-bottom network. IEEE TPAMI (2019)
13. Zhao, W., Zheng, B., Lin, Q., Lu, H.: Enhancing diversity of defocus blur detectors via cross-ensemble network. In: CVPR (June 2019)
14. Zhu, L., Deng, Z., Hu, X., Fu, C.W., Xu, X., Qin, J., Heng, P.A.: Bidirectional feature pyramid network with recurrent attention residual modules for shadow detection. In: ECCV (2018)