Supplementary Material for
Self-supervised Non-uniform Kernel Estimation with Flow-based Motion Prior for Blind Image Deblurring

Zhenxuan Fang[1] Fangfang Wu[1*] Weisheng Dong[1] Xin Li[2] Jinjian Wu[1] Guangming Shi[1]
[1]Xidian University [2]West Virginia University
zxfang@stu.xidian.edu.cn wufangfang@xidian.edu.cn wsdong@mail.xidian.edu.cn
xin.li@mail.wvu.edu jinjian.wu@mail.xidian.edu.cn gmshi@xidian.edu.cn

In this supplementary material, we provide more details of the DCNN in kernel estimation network, the architecture of normalizing flow model, the architecture of the block in the proposed UFPNet, the learned attention features in KAM and more visual comparison results on test images.

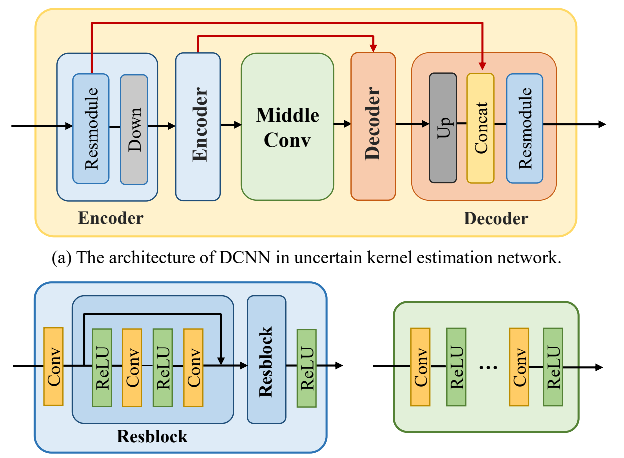
1. The DCNN in Kernel Estimation Network

In the proposed flow-based uncertain kernel estimation network, we use deep convolutional neural networks (DCNN) to estimate the latent code of the underlying blur kernel, the architecture of the DCNN is shown in Fig. 1. We use an encoder-decoder structure like [4, 5], two encoding and decoding blocks are used to reduce and increase the size of feature maps, respectively. The downsampling layer in encoding block is a 3×3 convolution layer with stride of 2, and the upsampling layer in decoding block is a deconvolution layer. The feature maps of encoding blocks are concatenated with the features in decoding blocks using long connections. There are 5 Conv layers with ReLU activation function in the middle convolution module.

2. The Architecture of the Normalizing Flow Model

In section 3.1 of the paper, we propose to establish a bijective mapping between the the complex motion blur kernel and the simple Gaussian distribution by a normalizing flow model. As shown in Fig. 2, the normalizing flow model [7] is built by stacking several invertible flow blocks, which consist of a batch normalization layer, a permutation layer and a affine transformation layer [3]. In affine transformation layer, fully connected neural networks (FCN) are used for scaling and shifting, each FCN stacks fully connected layers and tanh activation layers alternately.

*Corresponding author



(a) The architecture of DCNN in uncertain kernel estimation network.

(b) The architecture of Resmodule. (c) The architecture of Middle Conv.

Figure 1. The architecture of (a) the DCNN in the proposed kernel estimation network. (b) the Resmodule. (c) the Middle Conv.

3. The Architecture of the Block in UFPNet

In the proposed UFPNet, we adopt the block proposed by NAFNet [1], which is a simple and effective basic block for image restoration. As illustrated in Fig. 3, the basic block includes layernorm layer, 1×1 convolution layer, 3×3 deconvolution layer, SimpleGate layer and simplified channel attention layer (SCA). The SimpleGate layer divides the input feature maps into two parts in the channel dimension and element-wise multiply them, therefore the output feature maps have only half of the channels of the input feature maps. The simplified channel attention layer consists of a global average pooling layer and a 1×1 convolution layer.

4. The Learned Attention Features in KAM

We propose a novel kernel attention module (KAM) for the deblurring network to utilize the information from the estimated blur kernel sufficiently. Some learned attention

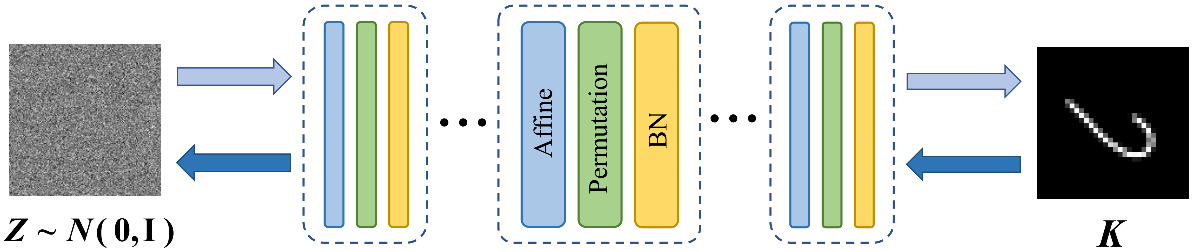
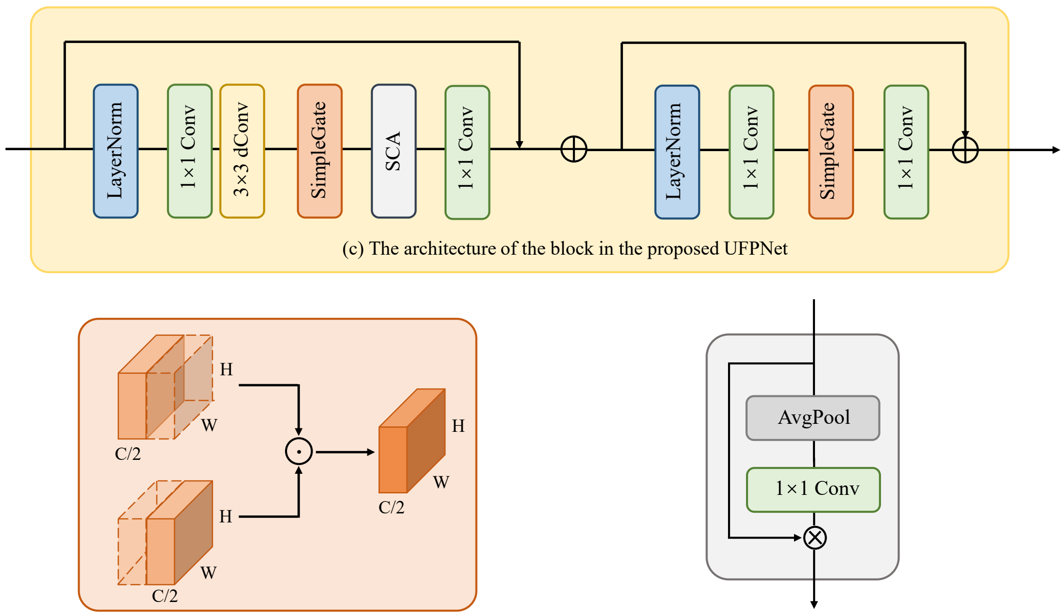


Figure 2. The illustration of the normalizing flow model, each flow block which consists of a batch normalization layer, a permutation layer and a affine transformation layer.



(b) The architecture of SimpleGate layer

(c) The architecture of Simplified Channel Attention (SCA).

Figure 3. The illustration of the block in the proposed UFPNet, and the architecture of SimpleGate layer and simplified channel attention (SCA) layer are illustrated in (b) and (c).

features of Eq. (7) in the paper are visualized in Fig. 4, With the help of the learned attention, image features will pay different attention to areas with different degrees of blurring, achieving better results on non-uniform deblurring.

5. More Visual Comparison Results

We provide more visual comparisons on benchmark test datasets in Fig. 5, 6, 7, 8, 9, 10, 11, 12, 13 and 14. We have compared our UFPNet method with several recent state-of-the-art blind deblurring methods, including HINet [2], DeepRFT [8], Stripformer [9], MSDI-Net [6] and NAFNet [1]. From Fig. 11 and Fig. 12, we can see that our method is effective for removing blur caused by motion trajectory. We also show the visualization results on the RWBI dataset [10] in Fig. 13 and 14, which only contains real-world blurry images without ground truth. It can be observed that the proposed method can achieve higher reconstruction quality and recover more details of the textures and edges than other methods.

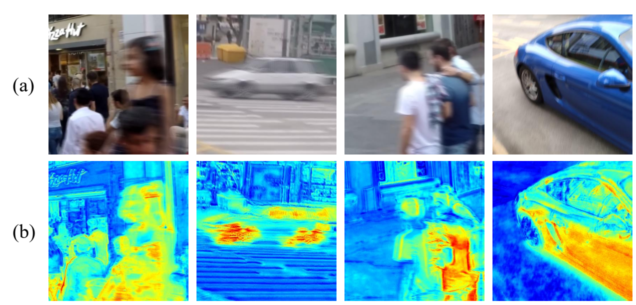


Figure 4. The visualization of the learned attention features in KAM. (a) The blurry image. (b) The learned attention features.

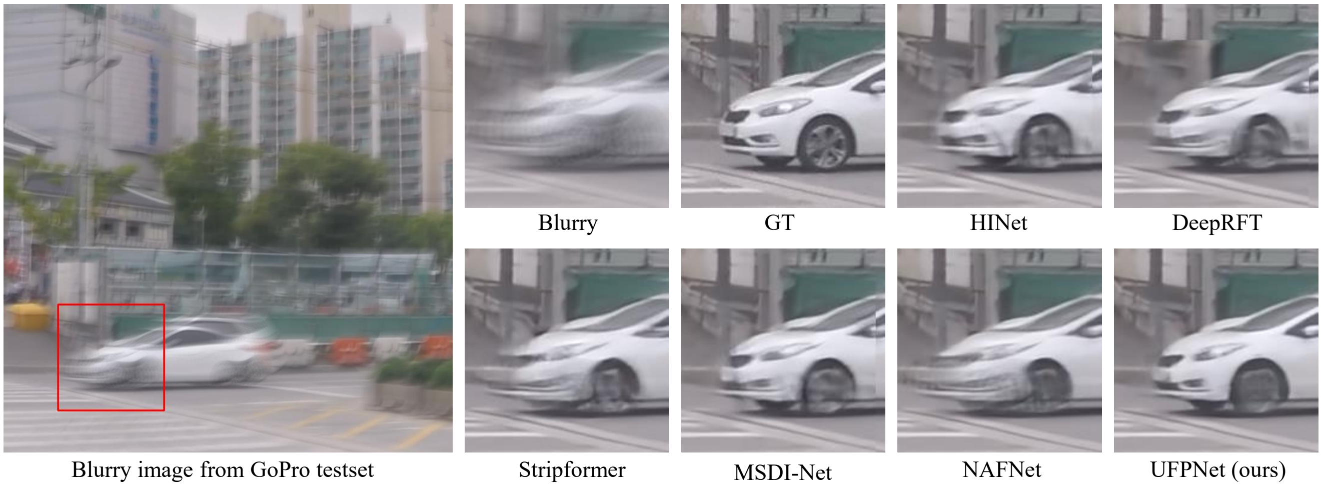


Figure 5. Visual comparisons on the GoPro dataset. From left to right: blurry image, ground-truth, results by HINet [2], DeepRFT [8], Stripformer [9], MSDI-Net [6], NAFNet [1] and UFPNet (ours).

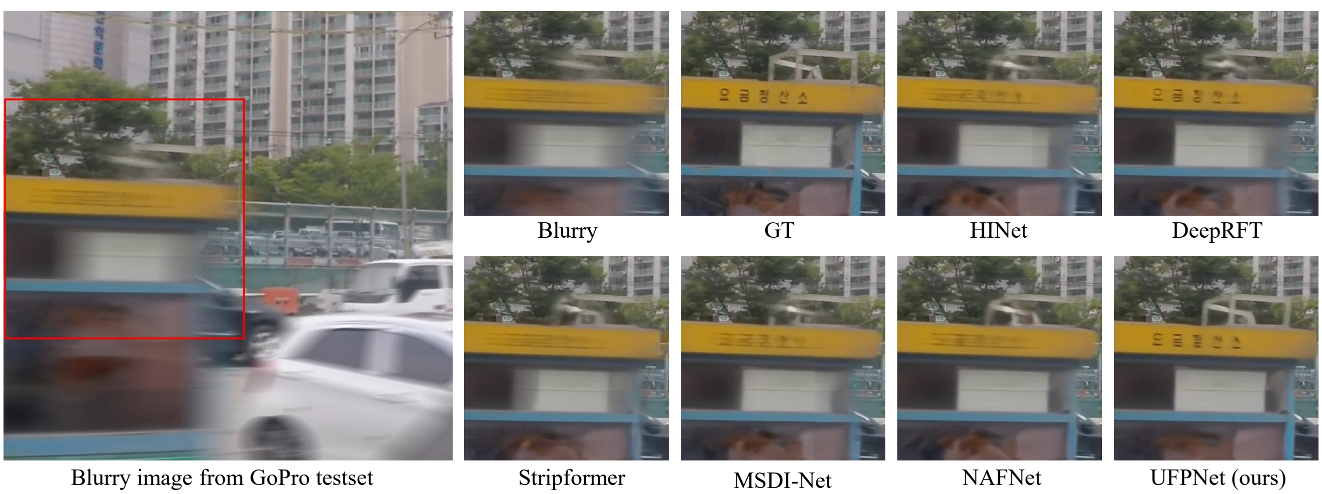


Figure 6. Visual comparisons on the GoPro dataset. From left to right: blurry image, ground-truth, results by HINet [2], DeepRFT [8], Stripformer [9], MSDI-Net [6], NAFNet [1] and UFPNet (ours).

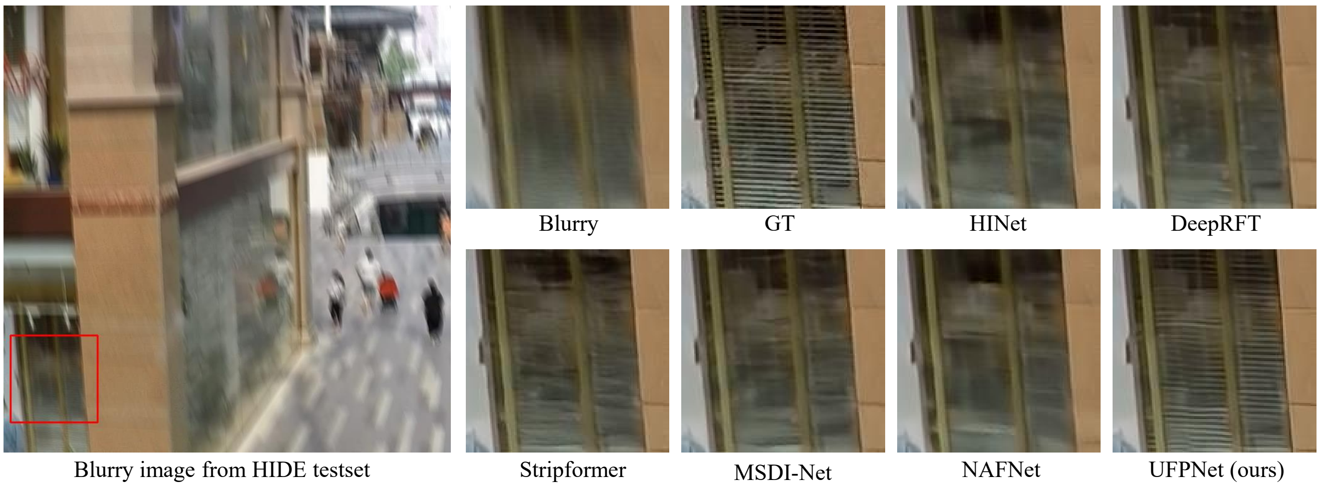


Figure 7. Visual comparisons on the HIDE dataset. From left to right: blurry image, ground-truth, results by HINet [2], DeepRFT [8], Stripformer [9], MSDI-Net [6], NAFNet [1] and UFPNet (ours).

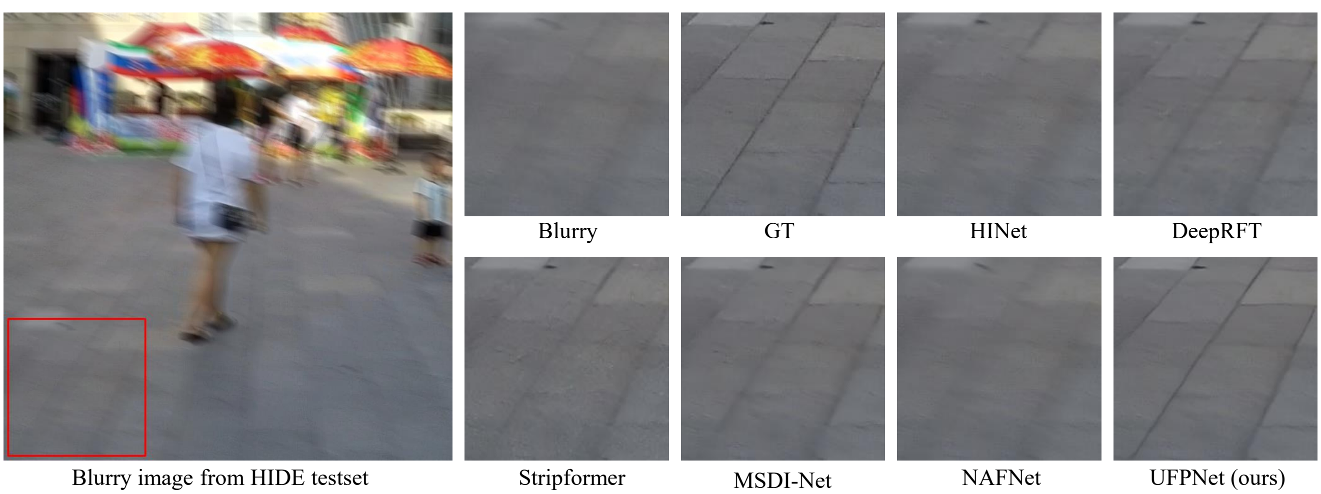


Figure 8. Visual comparisons on the HIDE dataset. From left to right: blurry image, ground-truth, results by HINet [2], DeepRFT [8], Stripformer [9], MSDI-Net [6], NAFNet [1] and UFPNet (ours).

Figure 9. Visual comparisons on the RealBlur-J dataset. From left to right: blurry image, ground-truth, results by HINet [2], DeepRFT [8], Stripformer [9], MSDI-Net [6], NAFNet [1] and UFPNet (ours).

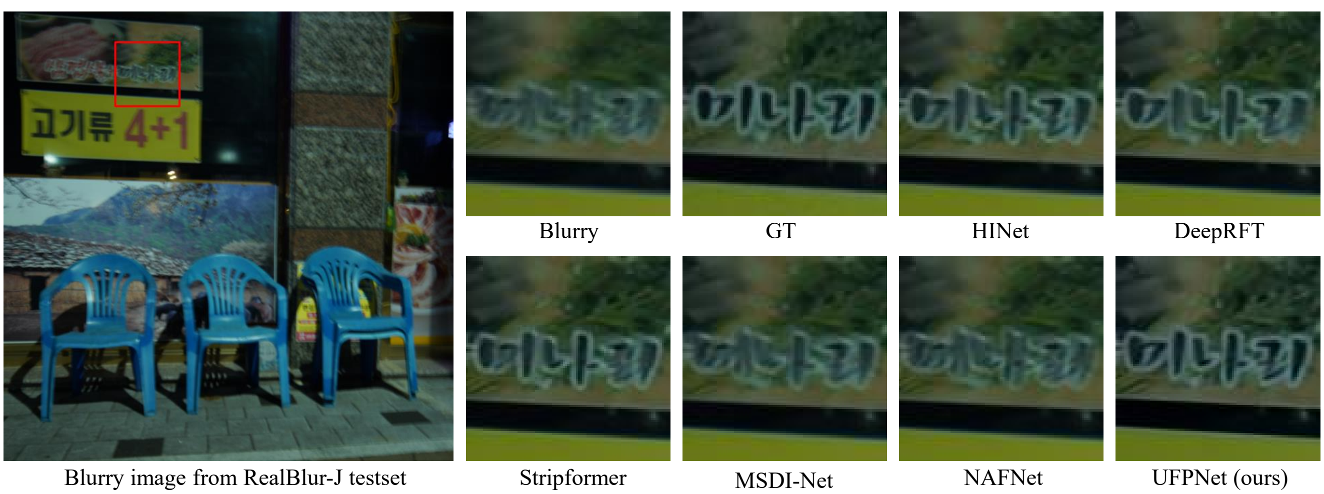


Figure 10. Visual comparisons on the RealBlur-J dataset. From left to right: blurry image, ground-truth, results by HINet [2], DeepRFT [8], Stripformer [9], MSDI-Net [6], NAFNet [1] and UFPNet (ours).

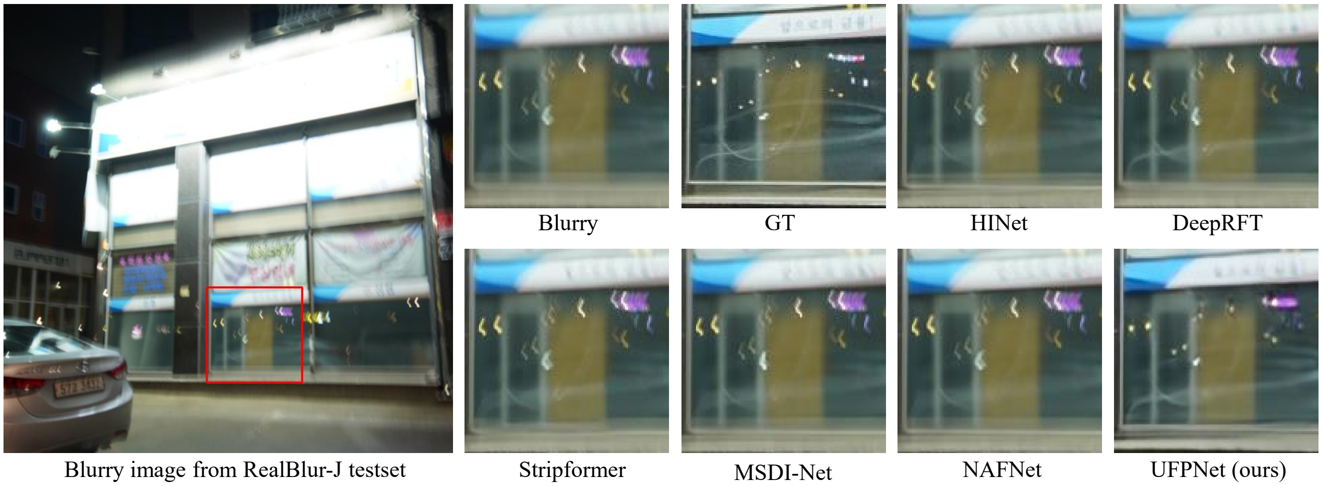


Figure 11. Visual comparisons on the RealBlur-J dataset. From left to right: blurry image, ground-truth, results by HINet [2], DeepRFT [8], Stripformer [9], MSDI-Net [6], NAFNet [1] and UFPNet (ours).

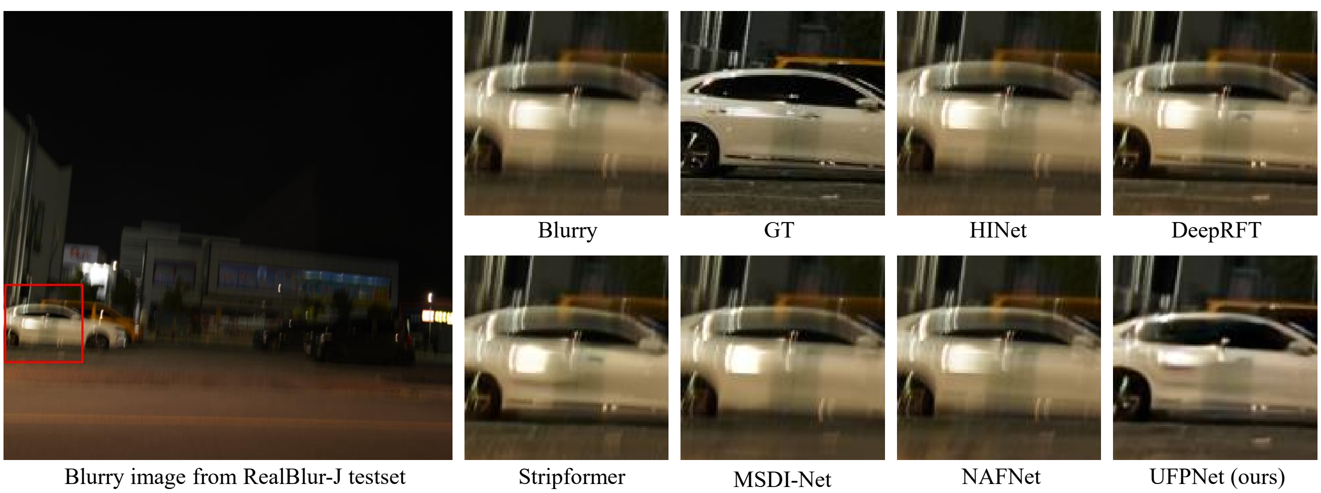


Figure 12. Visual comparisons on the RealBlur-J dataset. From left to right: blurry image, ground-truth, results by HINet [2], DeepRFT [8], Stripformer [9], MSDI-Net [6], NAFNet [1] and UFPNet (ours).

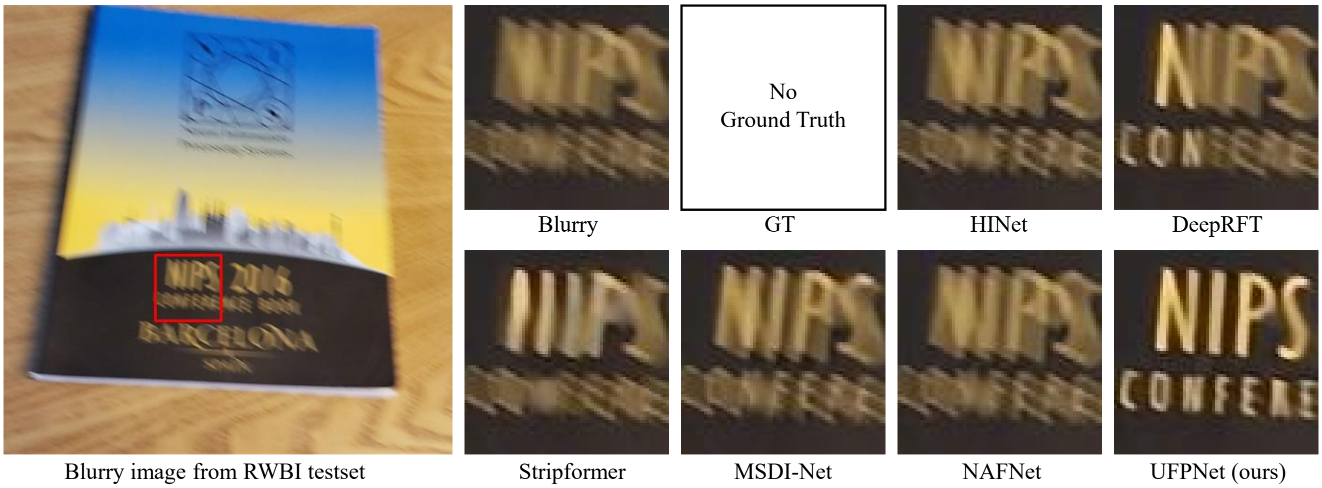


Figure 13. Visual comparisons on the RWBI dataset. From left to right: blurry image, results by HINet [2], DeepRFT [8], Stripformer [9], MSDI-Net [6], NAFNet [1] and UFPNet (ours).

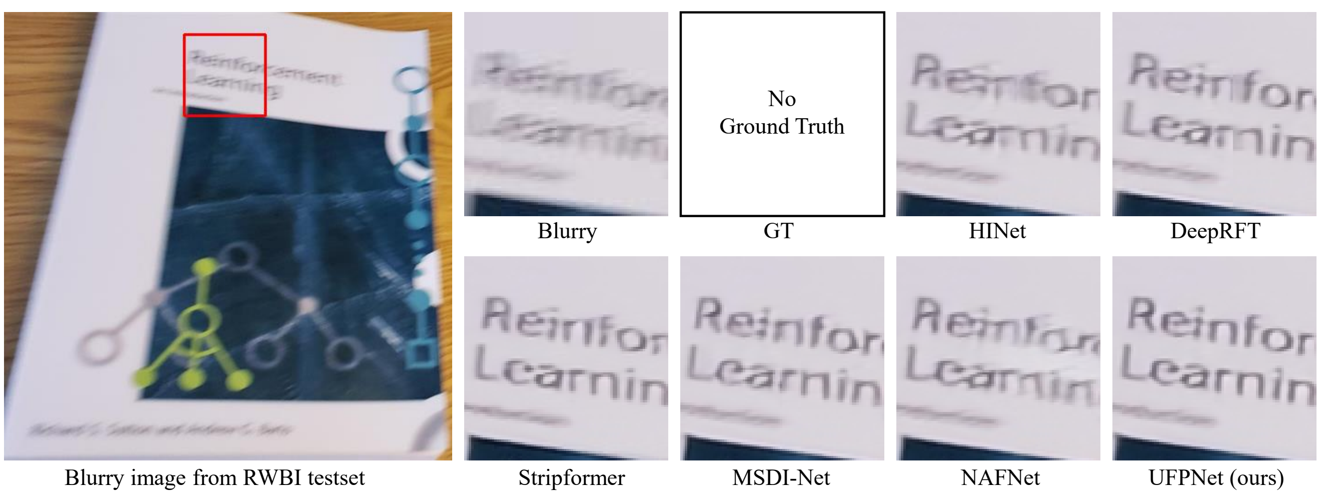


Figure 14. Visual comparisons on the RWBI dataset. From left to right: blurry image, results by HINet [2], DeepRFT [8], Stripformer [9], MSDI-Net [6], NAFNet [1] and UFPNet (ours).

References

[1] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part VII*, pages 17–33. Springer, 2022. 1, 2, 3, 4, 5, 6, 7

[2] Liangyu Chen, Xin Lu, Jie Zhang, Xiaojie Chu, and Chengpeng Chen. Hinet: Half instance normalization network for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 182–192, 2021. 2, 3, 4, 5, 6, 7

[3] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using real nvp. *arXiv preprint arXiv:1605.08803*, 2016. 1

[4] Zhenxuan Fang, Weisheng Dong, Xin Li, Jinjian Wu, Leida Li, and Guangming Shi. Uncertainty learning in kernel estimation for multi-stage blind image super-resolution. In *European Conference on Computer Vision*, pages 144–161. Springer, 2022. 1

[5] Soo Ye Kim, Hyeonjun Sim, and Munchurl Kim. Koalanet: Blind super-resolution using kernel-oriented adaptive local adjustment. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10611–10620, 2021. 1

[6] Dasong Li, Yi Zhang, Ka Chun Cheung, Xiaogang Wang, Hongwei Qin, and Hongsheng Li. Learning degradation representations for image deblurring. In *European Conference on Computer Vision*, pages 736–753. Springer, 2022. 2, 3, 4, 5, 6, 7

[7] Jingyun Liang, Kai Zhang, Shuhang Gu, Luc Van Gool, and Radu Timofte. Flow-based kernel prior with application to blind super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10601–10610, 2021. 1

[8] Xintian Mao, Yiming Liu, Wei Shen, Qingli Li, and Yan Wang. Deep residual fourier transformation for single image deblurring. *arXiv preprint arXiv:2111.11745*, 2021. 2, 3, 4, 5, 6, 7

[9] Fu-Jen Tsai, Yan-Tsung Peng, Yen-Yu Lin, Chung-Chi Tsai, and Chia-Wen Lin. Stripformer: Strip transformer for fast image deblurring. *arXiv preprint arXiv:2204.04627*, 2022. 2, 3, 4, 5, 6, 7

[10] Kaihao Zhang, Wenhan Luo, Yiran Zhong, Lin Ma, Bjorn Stenger, Wei Liu, and Hongdong Li. Deblurring by realistic blurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2737–2746, 2020. 2