Supplementary Material for Uncertainty Learning in Kernel Estimation for Multi-Stage Blind Image Super-Resolution

Zhenxuan Fang[1], Weisheng Dong[1](✉), Xin Li[2], Jinjian Wu[1], Leida Li[1], and Guangming Shi[1]

[1] School of Artificial Intelligence, Xidian University, Xi'an, China
zxfang@stu.xidian.edu.cn, {wsdong, jinjian.wu}@mail.xidian.edu.cn
{ldli, gmshi}@xidian.edu.cn
[2] Lane Dep. of CSEE, West Virginia University, Morgantown WV, USA
xin.li@mail.wvu.edu

In this supplementary material, we provide more details of the DCNN in kernel estimation network, more visualization results of the learned regularization parameters $\boldsymbol{w}$ and more visual comparison results on both synthetic and real-world images. Besides, we present the unsatisfactory result of the spatially variant degradation in blind SR.

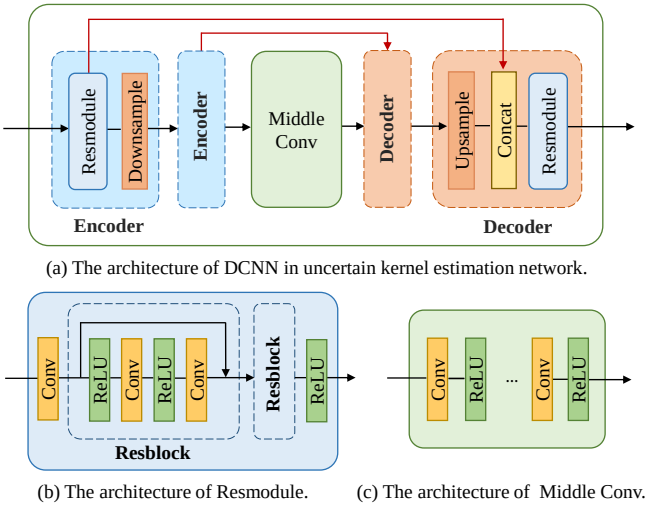
The code is available at https://github.com/Fangzhenxuan/UncertaintySR.

1 DCNN in Kernel Estimation Network

In the proposed uncertain kernel estimation network, we use a DCNN to extract the features of the underlying blur kernel, the architecture of the DCNN is shown in Fig. 1. We adopt a U-net structure in [2]. 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 Conv layer with stride of 2, the upsampling layer in decoding block is an inverse transpose Conv layer. Feature maps produced by 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 More Visual results of the learned regularization parameters $\boldsymbol{w}$

Fig. 2 shows the ground truth images and their corresponding regularization parameters $\boldsymbol{w}$ (with normalization) estimated in 4 stages. We can see that $\boldsymbol{w}$ is rather sparse and the values clearly reflect the edges and textures of images. With the help of the well-learned $\boldsymbol{w}$, the proposed SR network will pay more attention to reconstruct the high-frequency edges and textures.



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

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

Fig. 1. (a) The the architecture of the DCNN in the proposed uncertain kernel estimation network. (b) The architecture of the Resmodule. (c) The architecture of the Middle Conv, which contains 5 convolution layers with ReLU activation function.

3 More visual comparison results

We provide more visual comparisons on synthetic images in Fig. 3, the specific blur kernel used for generating the LR image is displayed on the upper left. We have compared our method with several recent state-of-the-art blind SR methods, including IKC [1], DASR [5], KOALAnet [2] and MANet [4]. We further conduct experiments on real-world historical images [3] to demonstrate the generalization property and effectiveness of our method in Fig. 4. From Fig. 3 and Fig. 4, it can be observed that the proposed method can achieve higher reconstruction quality and recover more details of the textures and edges than the other methods.

4 Unsatisfactory results on spatially variant degradation

As illustrated in Fig. 2(b) of the paper, a single convolution operation is adopted in layer $\mathbf{A}$ and $\mathbf{A}^\top$, the same convolution kernel ($\boldsymbol{k}$) is used to traverse the whole feature map. Such a simplified model corresponds to the assumption with spatially invariant blur in the degradation process. Therefore, the proposed network is not suitable for spatially variant degradation in the blind SR problem, as shown in Fig. 5. How to extend this work to handle the situation of blind image SR with spatially varying kernel is left for future work.

Table 1. SSIM results of the estimated kernels by kernel estimation network with or without Uncertainty Learning (UL).

Kernel estimation	Noise	1	2	3	4	5	6	7	8	9	Average
w/o UL	0	0.9634	0.9832	0.9558	0.9506	0.9515	0.9465	0.9407	0.9238	0.9064	0.9469
	10	0.9590	0.9773	0.9485	0.9456	0.9469	0.9398	0.9389	0.9186	0.8949	0.9410
	20	0.9543	0.9764	0.9376	0.9375	0.9390	0.9326	0.9317	0.9112	0.8877	0.9342
w/ UL	0	**0.9661**	**0.9873**	**0.9683**	**0.9690**	**0.9699**	**0.9585**	**0.9591**	**0.9515**	**0.9365**	**0.9629**
	10	**0.9623**	**0.9869**	**0.9649**	**0.9667**	**0.9675**	**0.9573**	**0.9582**	**0.9493**	**0.9321**	**0.9605**
	20	**0.9598**	**0.9856**	**0.9570**	**0.9594**	**0.9598**	**0.9501**	**0.9453**	**0.9426**	**0.9291**	**0.9543**

5 Similarity metric of the estimated kernel

In Table 1 of the paper, we compare the mean absolute error (MAE) of the blur kernels estimated by our uncertain network and the modified deterministic network. Here we also provide the SSIM metric of kernel similarity, as shown in Table 1.

References

1. Gu, J., Lu, H., Zuo, W., Dong, C.: Blind super-resolution with iterative kernel correction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1604–1613 (2019) 2
2. Kim, S.Y., Sim, H., Kim, M.: Koalanet: Blind super-resolution using kernel-oriented adaptive local adjustment. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10611–10620 (2021) 1, 2
3. Lai, W.S., Huang, J.B., Ahuja, N., Yang, M.H.: Deep laplacian pyramid networks for fast and accurate super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 624–632 (2017) 2
4. Liang, J., Sun, G., Zhang, K., Van Gool, L., Timofte, R.: Mutual affine network for spatially variant kernel estimation in blind image super-resolution. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4096–4105 (2021) 2
5. Wang, L., Wang, Y., Dong, X., Xu, Q., Yang, J., An, W., Guo, Y.: Unsupervised degradation representation learning for blind super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10581–10590 (2021) 2

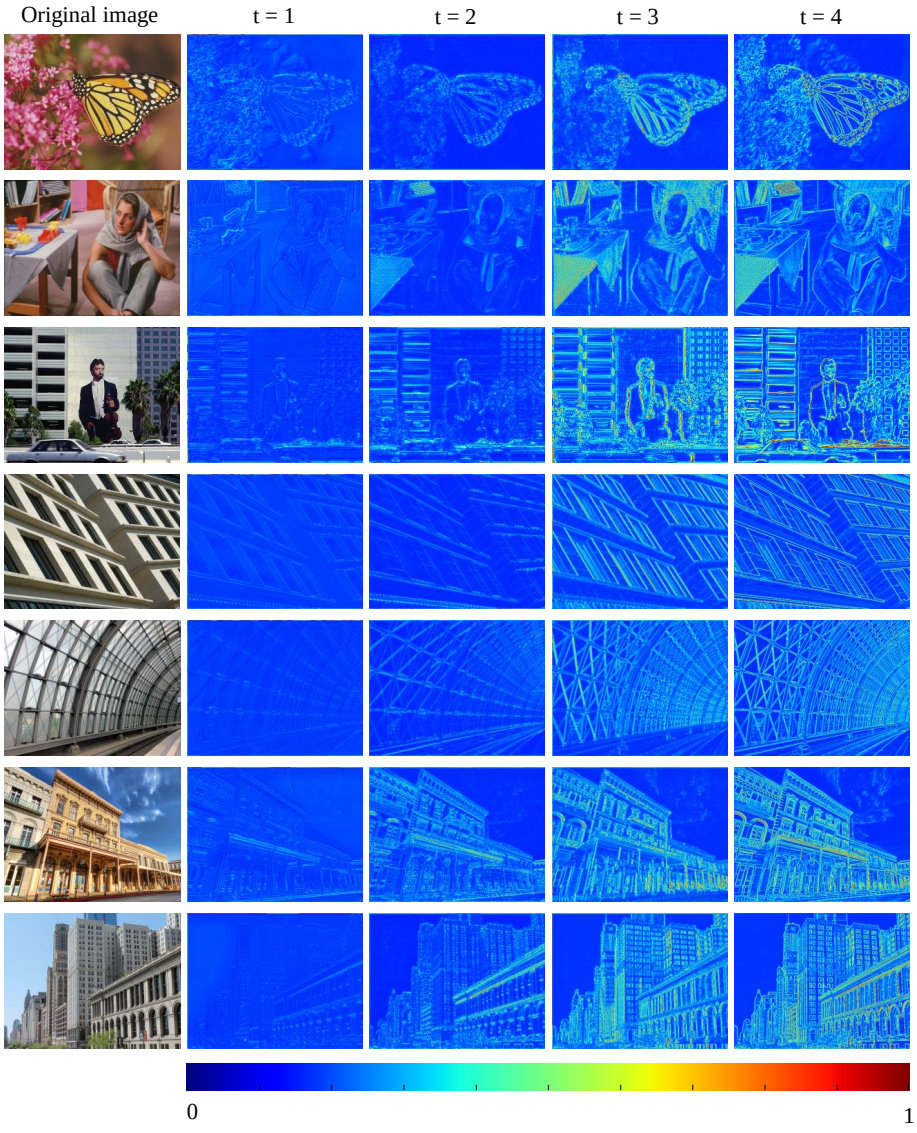
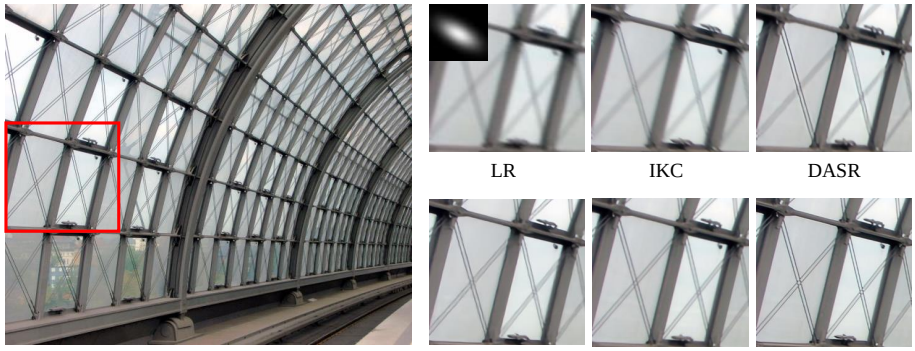


Fig. 2. The original HR images and the visualization of the learned regularization parameters $\boldsymbol{w}$ estimated in 4 stages (with normalization).

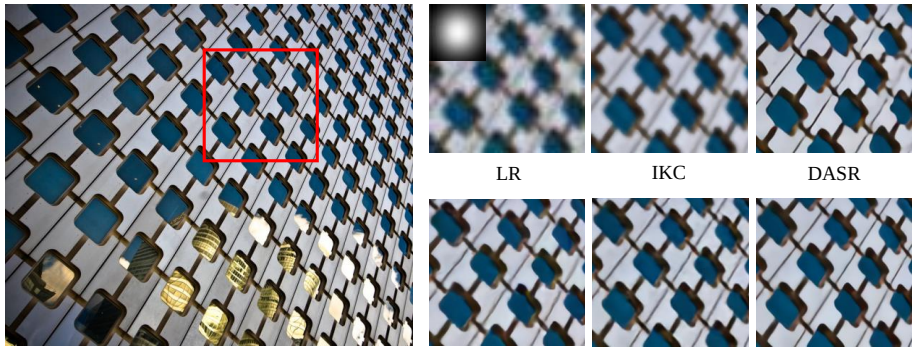


Img_002 from Urban100

(a) Scale factor ×2

Img_001 from Urban100

(b) Scale factor ×3



Img_041 from Urban100

(c) Scale factor ×4

Fig. 3. More visual results of different methods on synthetic images for scale factor ×2, ×3 and ×4. Noise levels are set to 0, 5 and 10 for three images, respectively.

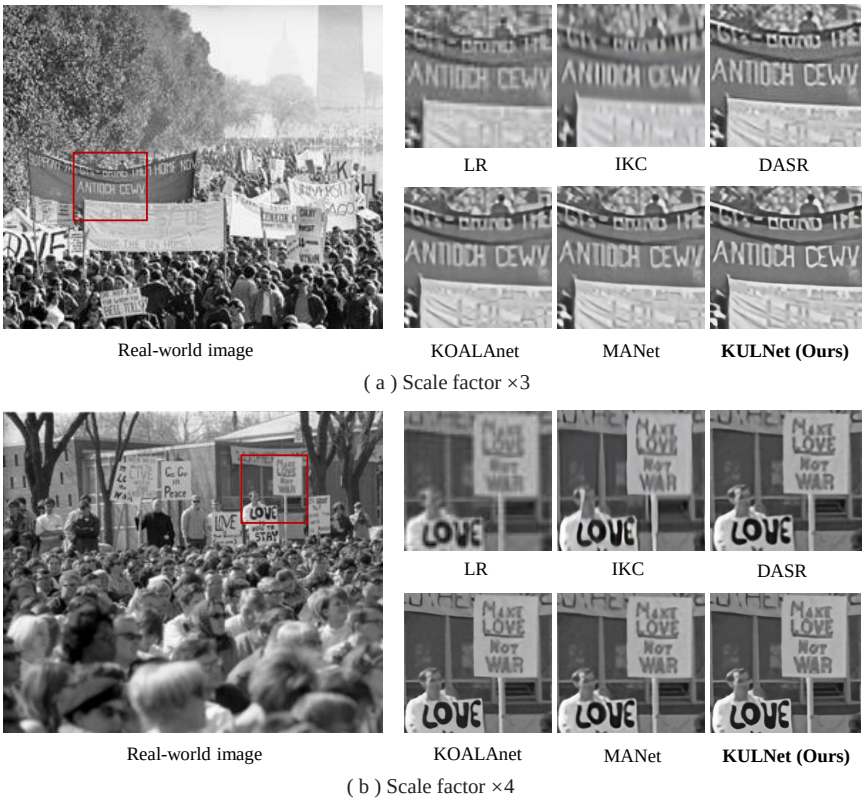


Fig. 4. More visual results of different methods on real-world images for scale factor ×3 and ×4.

Fig. 5. Unsatisfactory result on spatially variant degradation image.