Supplementry Material for IDENet: Implicit Degradation Estimation Network for Efficient Blind Super Resolution

Asif Hussain Khan
University of Udine
Udine, Italy
khan.asifhussain@spes.uniud.it

Christian Micheloni
University of Udine
Udine, Italy
christian.micheloni@uniud.it

Niki Martinel
University of Udine
Udine, Italy
niki.martinel@uniud.it

Abstract

This supplementary material accompanies the CVPR main paper submission entitled "IDENet: Implicit Degradation Estimation Network for Efficient Blind Super Resolution". It contains additional experiments, analyses and results that were excluded from the main paper due to page limit constraints. In particular, we expanded the ablation study included in the main paper by evaluating the importance of the patch size for model training, and the effects of blur kernel widths in the SR process. We finally compounded the visual results included in the main paper with additional outcomes/comparisons with existing models.
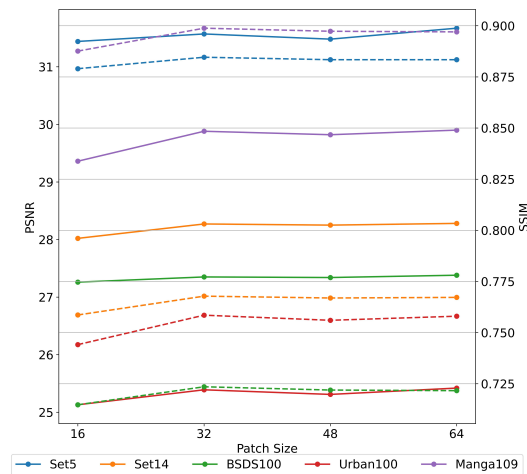
Figure 1. Performance of IDENet under various patch sizes at a scale factor of $\times 4$. The solid lines refer to the PSNR (left y-axis), and the dashed lines represent the SSIM (right y-axis).

1. Impact of Image Patch Size

We conducted an ablation study and explored the impact of various patch sizes on PSNR as shown in Figure 1. While the patch size 64×64 achieved a higher PSNR (a gain of 0.04 dB concerning the 32×32 patch), it introduced computational overhead, which, in our setup, results in 0.0045 more seconds for image inference and in additional 0.84GB of occupied memory. The 32×32 patch size struck a balance on computational efforts and performance, hence making it the optimal choice for our efficient approach.

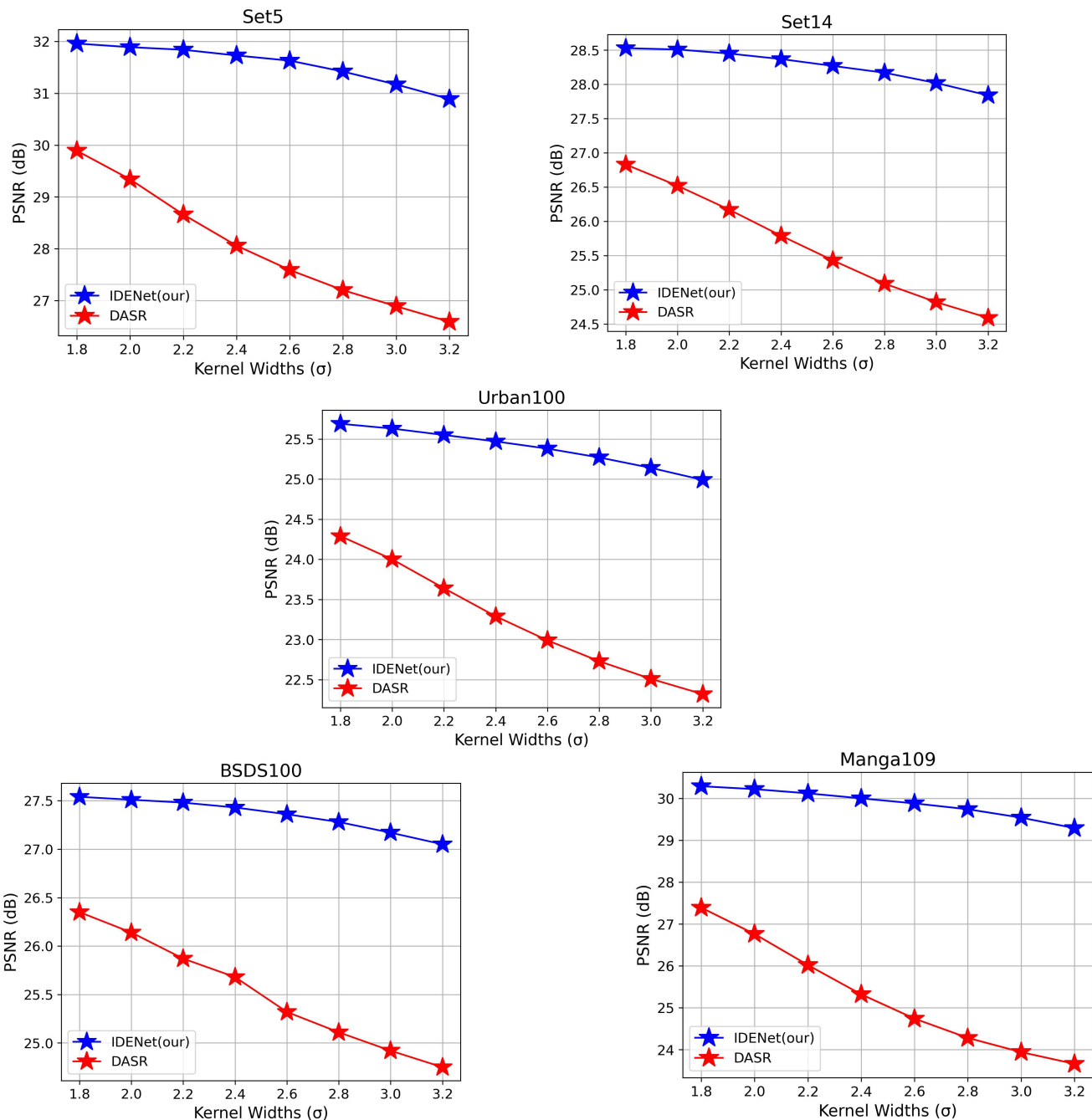Figure 2. Performance curve shows PNSR on different public benchmark datasets for scale factor ×4.'The kernel's width (σ) is set from 1.8 to 3.2.

2. Impact of Blur Kernel Widths

In Figure 2, we report on the comparison between our approach and the state-of-the-art direct implicit blind SR competitor, DASR [5]. In particular, we conducted an analysis on the impact of the different blur kernel widths that are used in the Isotropic Gaussian kernel setting (*i.e.*, the Gaussian8 kernels [3, 7]). We evaluated the performance of our approach and DASR on the same benchmarks we used in the main paper, *i.e.*, Set5 [2], Set14 [10], BSD100 [8], Urban100 [4], and Manga109 [9]. Results show that our approach has significantly more stable performance than DASR when the blur kernel
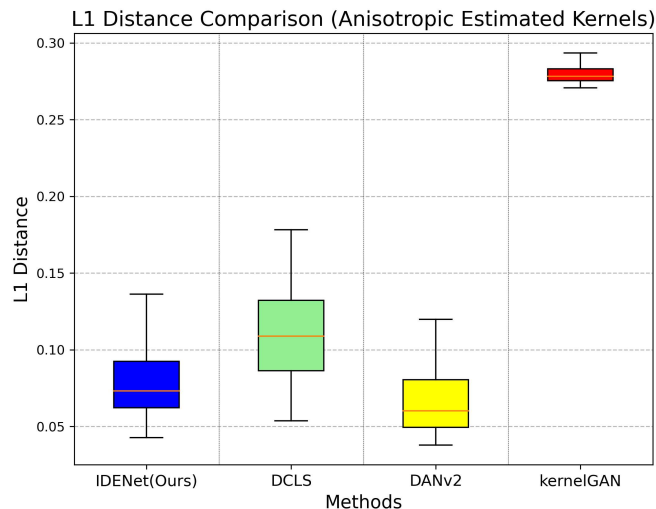
Figure 3. Plot represents the computed L1 Distance between the estimated anisotropic and ground truth kernels for different methods.
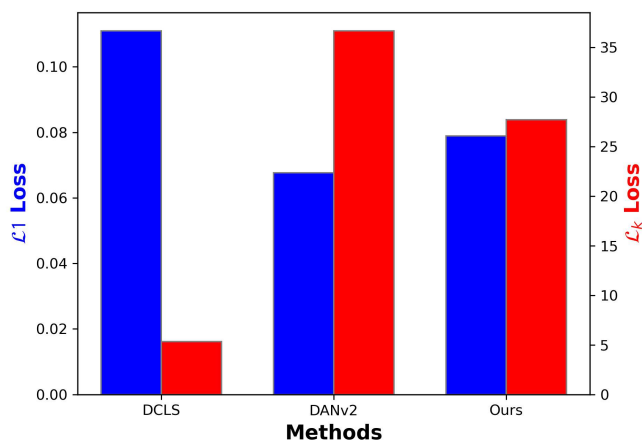

Figure 4. Comparison of $\mathcal{L}_1$ (computed between ground truth kernels and estimated kernels) and $\mathcal{L}_k$ losses between DCLS, DANv2 and IDENet(Ours). The results are computed for the DIV2KRK validation dataset.

width is increased. In particular, on average, our method loses less than 1 dB when the largest Gaussian8 [3, 7] kernel width (*i.e.*, 3.2) is considered. Differently, DASR sees a relevant performance drop (*i.e.*, even more than 3 dB for the Manga109 dataset) when the same setting is considered. These results indicate that our approach can estimate the kernels efficiently exploited by the deconvolution approach and the refinement module to generate the SR output.

3. Blur Kernel Estimation Quality

In Figure 7 of the main submission, we reported the quality and the average L1 distance between our estimated kernels and the ground-truth ones. Figure 3 further analyzes such a result by showing more details about the quality of all the estimations made by our method and existing state-of-the-art works. Results show that our IDENet performs remarkably well in kernel estimation compared to DCLS and kernelGAN [1] with a notable difference in computed L1 distances (*i.e.*,0.0311 and 0.2008 respectively). Furthermore, it performs closely to DANv2 [6] exhibiting only a marginal L1 difference (*i.e.*, of 0.0112).

3.1. Noisy values shown in the estimated kernels.

The noisy values are due to how we implicitly estimate the degradation kernels through our novel loss –without the need for the ground truth degradation kernels. The implicit kernel loss term in (10) minimizes the difference between the GT LR image and the downsampled GT HR image (filtered with the learnable kernel). This does not guarantee that the kernel is smooth: the bilinear downsampling operator (denoted as $\downarrow_S$) smooths neighboring values, thus canceling out the "imperfections" resulting from the noisy filter. To validate our hypothesis, we computed the results in Figure 4. This shows that, having a kernel that is very similar to the ground truth (*i.e.*, a low $\mathcal{L}_1$ distance between ground truth and estimated kernel) does not guarantee that $\mathcal{L}_k$ shows the same behavior. Indeed, while our method has noisy values and a larger $\mathcal{L}_1$ with respect to DANv2, it provides a better result than it when considering the estimated kernel for the degradation loss in (10), which, for implicit SR, is the relevant term.

4. Perceptual Quality Measurement

We extended our evaluation considering the perceptual measure (LPIPS). Values in Table 1 demonstrate that while DASR is trained on pixel, regression, perceptual, and adversarial losses, IDENet achieved better LPIPS results than DASR, with comparable performance than explicit methods.

		LPIPS↓				
Estimation Type	**Methods**	**Set5**	**Set14**	**BSD100**	**Urban100**	**Manga109**
	IKC	0.1607	0.2621	0.3456	0.2431	0.1342
	DANv1	0.1483	0.2560	0.3326	0.2086	0.0924
Explicit BSR††	DANv2	0.1459	0.2497	0.3267	0.2030	0.0910
	DCLS	0.1474	0.2515	0.3242	0.2000	0.0870
	AdaTarget	0.1932	0.3121	0.4029	0.2518	0.1286
Implicit BSR†††	DASR	0.2467	0.3035	0.3962	0.2931	0.2247
	IDENet(Ours)	0.1523	0.2556	0.3384	0.2262	0.0999

Table 1. LPIPS comparison between SOTA method.

5. Large Scale Factors

6. Qualitative Results

In this section, we aim to provide more visualizations for comparing the state-of-the-art explicit method (*i.e.*, DCLS [7]) and state-of-the-art implicit method (*i.e.*, DASR [5]) with our approach (*i.e.*, IDENet). These visual comparisons are based on different benchmark datasets and encompass varying kernel widths.

6.1. Isotropic Kernels

Figure 6-9 showcase the results of our method under different isotropic blur kernels with various widths. In Figure 6, our approach successfully reconstructs the image, preserving details, colors, and textures. This achievement mirrors the comparative visual quality attained by DCLS. In contrast, the image reconstructed by DASR exhibits noticeable blurriness. Figure 7 and Figure 8 further underscore the effectiveness of our proposed approach in achieving visually pleasant results. The restoration of facial details in *Img '189080'* (Figure 7) and preservation of shapes and structures in *Img '025'* (Figure 8) demonstrate the excellence of our method. In Figure 9, we also report on a sample case where low PSNR/SSIM values are achieved by different methods. Our approach excels in reconstructing and preserving details within the image. Conversely, DASR tends to smooth out these details. The visual comparisons across different blur kernels highlight the robust performance of our approach in image reconstruction and preservation of image visual features, showcasing notable advantages over our direct DASR [5] state-of-the-art competitor.

6.2. Anisotropic Kernels

Anisotropic kernels are diverse and complicated kernels. Figure 10-13 Showcase the visual quality of our approach in reconstructing images that are degraded by such anisotropic kernels. The visuals result in Figure 10-12 highlight the effectiveness of our method in maintaining the visual appeal of images, hence achieving comparable performance to that of DCLS despite

Scale Factor	LR Image	PSNR/SSIM SR Image(IDENet trained on x2)	LR Image	PSNR/SSIM SR Image(IDENet trained on x4)
x2		40.09/0.9815		29.74/0.9203
x4		25.10/0.8134		32.65/0.9424
x8		22.28/0.7312		22.52/0.7380
x16		19.84/0.6607		19.85/0.6602
x32		18.44/0.6205		18.44/0.6203

Figure 5. Visual comparison on image *'monarch'* from Set14 dataset on different scale factors.

such a method using ground-truth kernel information. Figure 13 shows the results of an image being reconstructed with low PSNR/SSIM. The LR version of such *Img '007'* sample contains very few details that none of the existing methods is able to exploit towards the generation of a high-quality SR image. This is particularly visible in the tree branches closer to the camera. In such a case, our method achieves comparable results to DCLS, while DASR faces more challenges and introduces noticeable blurriness. Overall, the visual evaluations across anistropic kernels emphasize the robust performance of our efficient IDENet method, particularly in scenarios demanding precision in reconstructing intricate details and textures, where it outperforms the state-of-the-art implicit blind SR approach, *i.e.*, DASR [5].

LR Image

Kernel Width = 2.0

PSNR/SSIM

GT Image

DCLS †† 35.44/0.9492

DASR ††† 31.09/0.8916

IDENet(Ours) ††† 34.26/0.9392

Figure 6. *Img 'bird* from set 5 dataset for ×4.'kernel width is 2.0.

LR Image

Kernel Width = 2.8

GT Image   PSNR/SSIM

DCLS †† 34.84/0.8822

DASR ††† 30.04/0.8306

IDENet(Ours) ††† 34.58/0.8808

Figure 7. *Img '189080'* from BSDS100 dataset for ×4.'kernel width is 2.8.
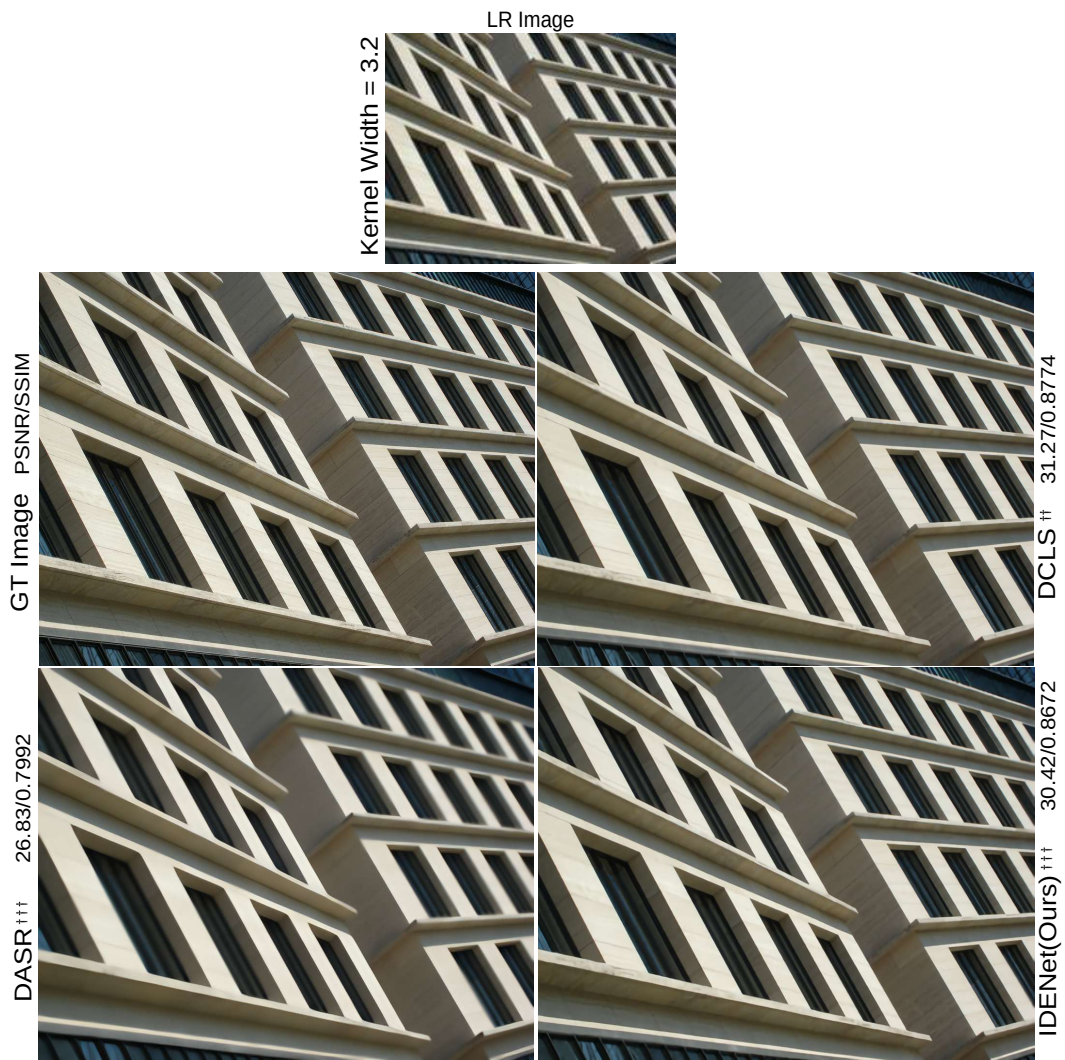
Figure 8. *Img '025'* from Urban100 dataset for ×4.'kernel width is 3.2.

LR Image

Kernel Width = 3.2

GT Image   PSNR/SSIM

DCLS †† 19.69/0.3842

DASR ††† 18.55/0.2049

IDENet(Ours) ††† 19.50/0.3384
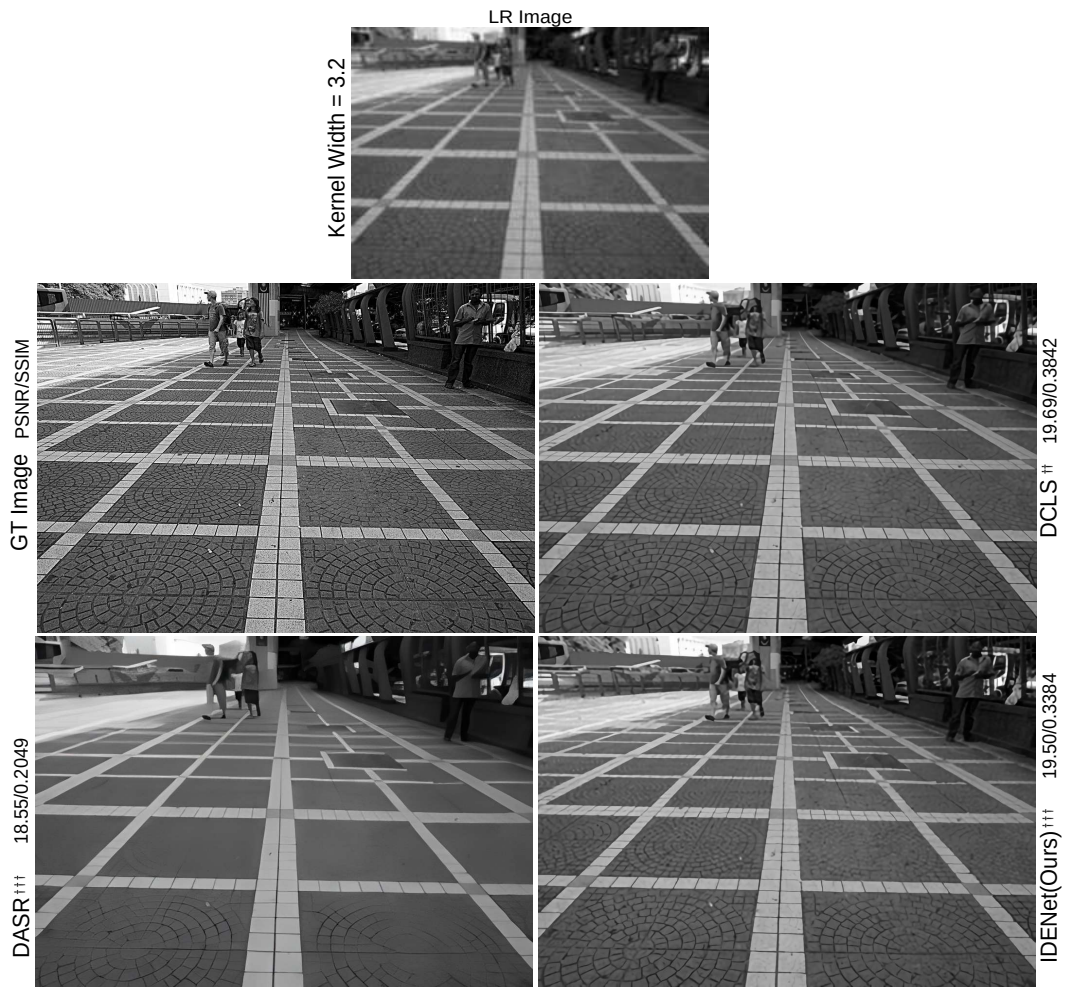
Figure 9. *Img '095'* from Urban100 dataset for ×4.'kernel width is 3.2.

LR Image

GT Image   PSNR/SSIM

DCLS †† 44.40/0.9899

DASR ††† 38.51/0.9785

IDENet(Ours) ††† 43.95/0.9888

Figure 10. *Img '044'* from DIV2KRK dataset for $\times 4$.'

LR Image

GT Image   PSNR/SSIM

DCLS †† 40.48/0.9812

DASR ††† 36.32/0.9722

IDENet(Ours) ††† 39.56/0.9799
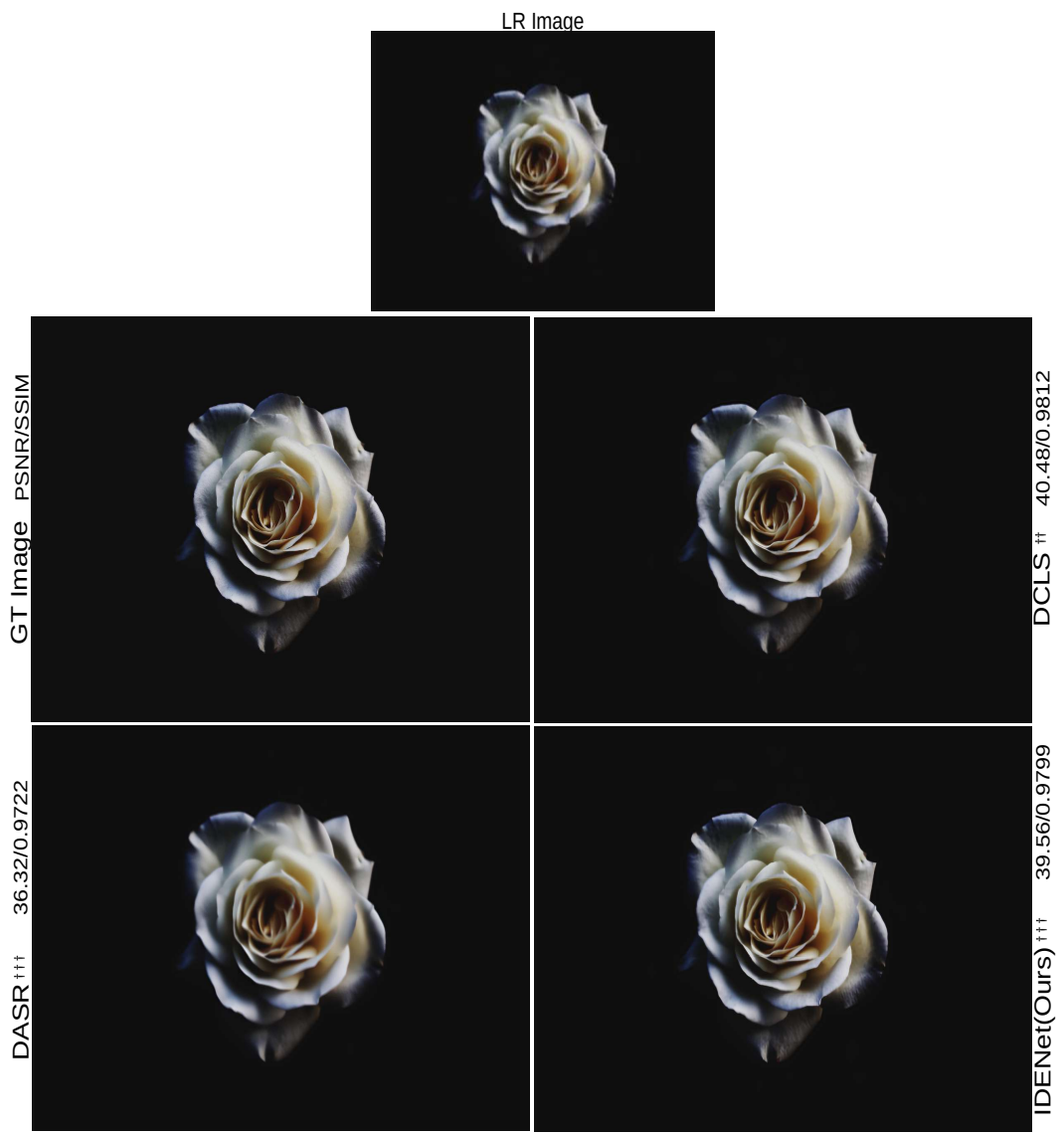
Figure 11. *Img '043'* from DIV2KRK dataset for ×4.'

LR Image

GT Image PSNR/SSIM

DCLS †† 34.06/0.9347

DASR ††† 29.97/0.8813

IDENet(Ours) ††† 33.30/0.9281

Figure 12. *Img '053'* from DIV2KRK dataset for ×4.'

LR Image

GT Image   PSNR/SSIM

DCLS †† 20.95/0.4488

DASR ††† 19.78/0.3381

IDENet(Ours) ††† 20.82/0.4345

Figure 13. *Img '007'* from DIV2KRK dataset for ×4.'

References

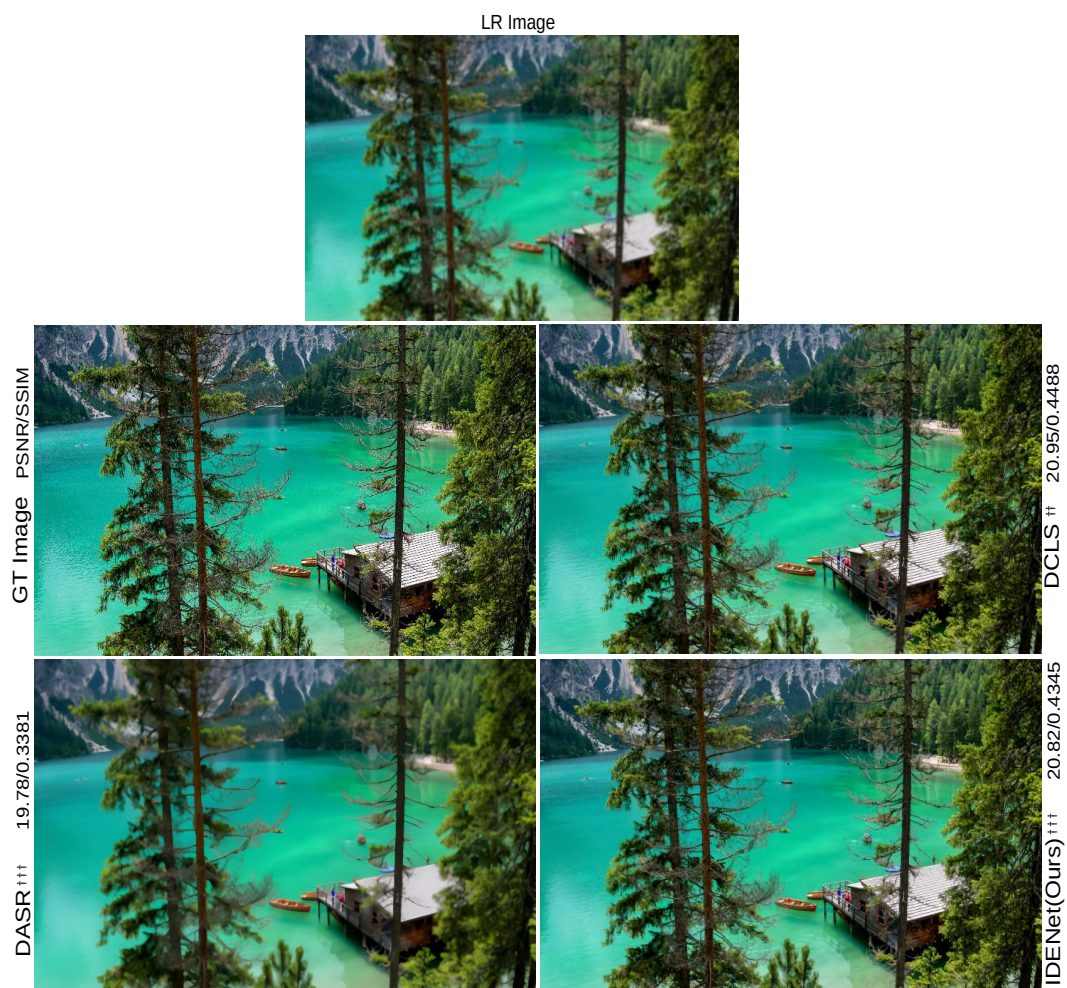[1] Sefi Bell-Kligler, Assaf Shocher, and Michal Irani. Blind super-resolution kernel estimation using an internal-gan. *Advances in Neural Information Processing Systems*, 32, 2019. 3

[2] Marco Bevilacqua, Aline Roumy, Christine Guillemot, and Marie Line Alberi-Morel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. 2012. 2

[3] Jinjin Gu, Hannan Lu, Wangmeng Zuo, and Chao Dong. Blind super-resolution with iterative kernel correction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1604–1613, 2019. 2, 3

[4] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5197–5206, 2015. 2

[5] Jie Liang, Hui Zeng, and Lei Zhang. Efficient and degradation-adaptive network for real-world image super-resolution. In *European Conference on Computer Vision*, pages 574–591. Springer, 2022. 2, 4, 5

[6] Zhengxiong Luo, Yan Huang, Shang Li, Liang Wang, and Tieniu Tan. End-to-end alternating optimization for blind super resolution. *arXiv preprint arXiv:2105.06878*, 2021. 3

[7] Ziwei Luo, Haibin Huang, Lei Yu, Youwei Li, Haoqiang Fan, and Shuaicheng Liu. deep constrained least squares for blind image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17642–17652, 2022. 2, 3, 4

[8] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms. *Department of Electrical Engineering and Computer Sciences University of California, Berkeley*. 2

[9] Yusuke Matsui, Kota Ito, Yuji Aramaki, Azuma Fujimoto, Toru Ogawa, Toshihiko Yamasaki, and Kiyoharu Aizawa. Sketch-based manga retrieval using manga109 dataset. *Multimedia Tools and Applications*, 76:21811–21838, 2017. 2

[10] Roman Zeyde, Michael Elad, and Matan Protter. On single image scale-up using sparse-representations. In *Curves and Surfaces: 7th International Conference, Avignon, France, June 24-30, 2010, Revised Selected Papers 7*, pages 711–730. Springer, 2012. 2