Natural and Realistic Single Image Super-Resolution with Explicit Natural Manifold Discrimination - Supplementary Material

Jae Woong Soh Gu Yong Park Junho Jo Nam Ik Cho
Department of ECE, INMC, Seoul National University, Seoul, Korea
{soh90815, benkay, jottue}@ispl.snu.ac.kr, nicho@snu.ac.kr

1. Natural Manifold Discrimination

An example of our proposed LR-HR model of the natural manifold is illustrated in Figure 1, which also shows our implementation and example images for training our NMD. Specifically, Figure 1a shows how I_A looks like in real implementation. As shown, I_A is a convex combination of bicubic interpolated image $h(I_{LR})^{\uparrow}$ and ground-truth high-resolution image I_{HR} with $\alpha = 0.8$. When we down-sample I_A using the bicubic kernel, it is very close to the I_{LR} with PSNR about 45 dB in RGB colorspace.

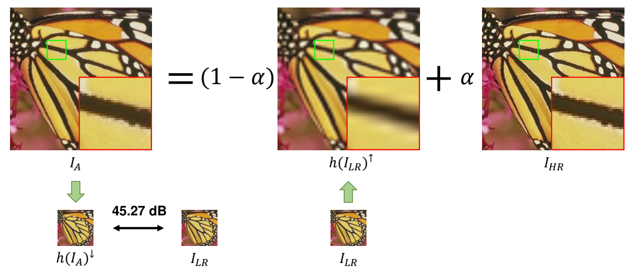
Figure 1b shows how I_B is synthesized. We exaggerate the noise by using $\sigma = 0.5$ for better visualization and explanation. The noise is injected only in the last column and row of 8×8 block-wise DCT coefficients. When we down-sample I_B with bicubic kernel, its similarity with I_{LR} in terms of PSNR is about 60 dB. When we use $\sigma = 0.1$, we obtained the PSNR between $h(I_B)^{\downarrow}$ and I_{LR} about 75 dB.

2. Discriminator Architecture for GAN

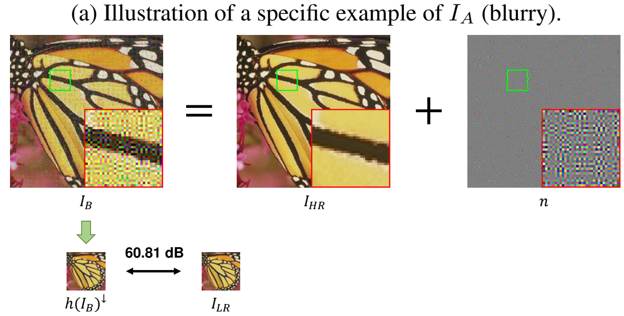
For the discriminator in our GAN-based method, we employ a VGG-like [8] structure shown in Figure 2. Instead of using max-pooling operations, we adopt convolution layers with stride 2. Also, we use one convolution layer followed by a global average pooling operation instead of fully-connected layers. We apply leaky ReLU for all convolution layers except the last one and also adopt spectral normalization [6] for all convolution weights.

3. Mean Opinion Score (MOS)

We conduct MOS test on DIV2K validation set [9] for subjective quality assessment. Specifically, six examples (HR, nearest neighbor (NN), EDSR [4], EnhanceNet [7], SFT-GAN [10], NatSR) are shown to the subjects. Two images, HR and NN is to calibrate human raters as 5 (good quality) and 0 (bad quality) score for each, and other four images are randomly shuffled and shown. We asked 21 human raters to score the images considering the naturalness and perceptual quality. Since DIV2K set is too large to see the effect, the cropped 400×400 center pixels are evaluated.



(a) Illustration of a specific example of I_A (blurry).



(b) Illustration of a specific example of I_B (noisy).

Figure 1: Illustration of examples of I_A and I_B

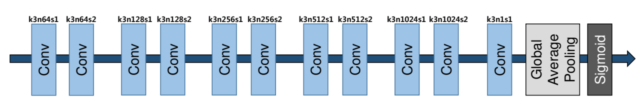


Figure 2: The discriminator network architecture for adversarial training.

Table 1 and Figure 3 shows the overall results. The best MOS result is highlighted in bold. As shown, our NatSR shows the best result and SFT-GAN rates similar to ours. EDSR scores the worst due to its blurriness. We also measured ANOVA (ANalysis Of VAriance) between SFT-GAN and our NatSR. The result is $p = 0.15$ which

Method	MOS
EDSR	2.13
ENet	2.84
SFT-GAN	3.21
NatSR (Ours)	**3.42**

Table 1: Mean Opinion Score result on average. The best result is highlighted in **bold**.

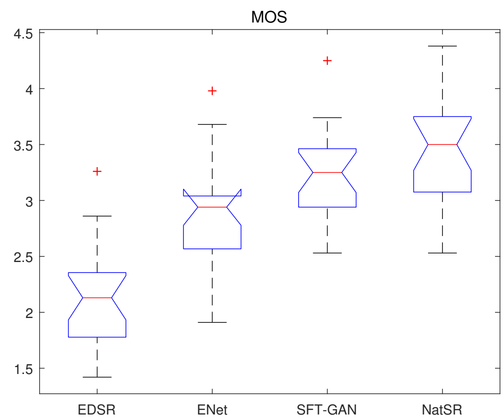


Figure 3: Mean Opinion Score distribution.

shows less significance ($p > 0.05$) between SFT-GAN and NatSR. One more interesting result we found is that SFT-GAN scores mostly higher than NatSR on the images with animals with fur or feathers which is the part of semantic categories whereas our NatSR scores better in other images. As SFT-GAN works in two steps (semantic segmentation and super-resolution with additional super-vision), we believe that our NatSR is more versatile than SFT-GAN since ours works in end-to-end fashion without any supervision to the network.

4. More Visualization Results

We also visualize more results of our NatSR with other comparisons from the next page.

References

[1] M. Bevilacqua, A. Roumy, C. Guillemot, and M. L. Alberi-Morel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In *BMVC*, 2012. 7

[2] J.-B. Huang, A. Singh, and N. 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. 7, 8

[3] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. P. Aitken, A. Tejani, J. Totz, Z. Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017. 6, 7

[4] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee. Enhanced deep residual networks for single image super-resolution. In *The IEEE conference on computer vision and pattern recognition (CVPR) workshops*, 2017. 1, 3, 4, 5, 6, 7, 8

[5] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 2, pages 416–423. IEEE, 2001. 6

[6] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida. Spectral normalization for generative adversarial networks. *arXiv preprint arXiv:1802.05957*, 2018. 1

[7] M. S. Sajjadi, B. Schölkopf, and M. Hirsch. Enhancenet: Single image super-resolution through automated texture synthesis. In *Computer Vision (ICCV), 2017 IEEE International Conference on*, pages 4501–4510. IEEE, 2017. 1, 3, 4, 5, 6, 7, 8

[8] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In *In International Conference on Learning Representations*, 2015. 1

[9] R. Timofte, E. Agustsson, L. Van Gool, M.-H. Yang, L. Zhang, B. Lim, S. Son, H. Kim, S. Nah, K. M. Lee, et al. Ntire 2017 challenge on single image super-resolution: Methods and results. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on*, pages 1110–1121. IEEE, 2017. 1, 3, 4, 5, 6

[10] X. Wang, K. Yu, C. Dong, and C. C. Loy. Recovering realistic texture in image super-resolution by deep spatial feature transform. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018. 1, 3, 4, 5, 6

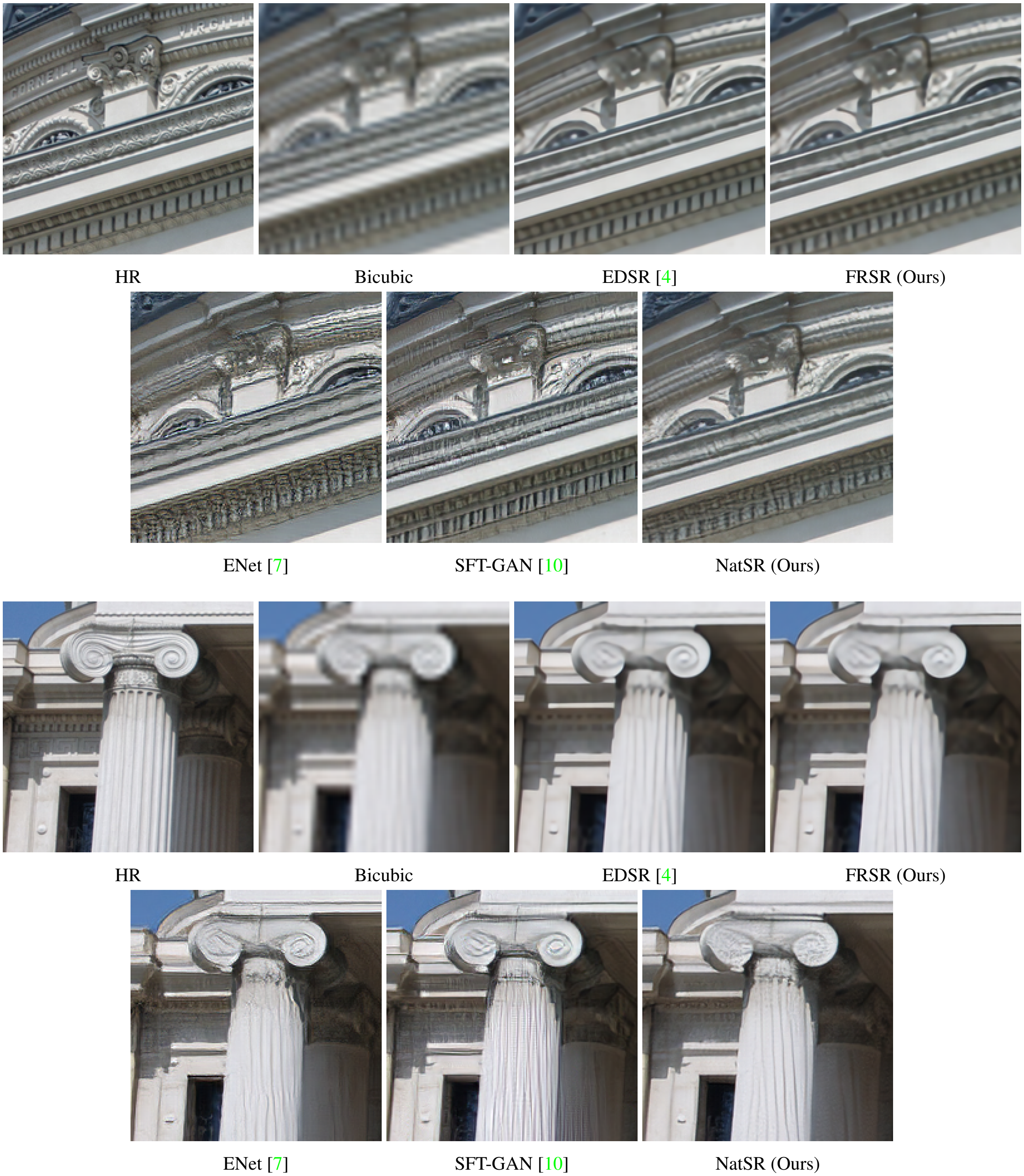
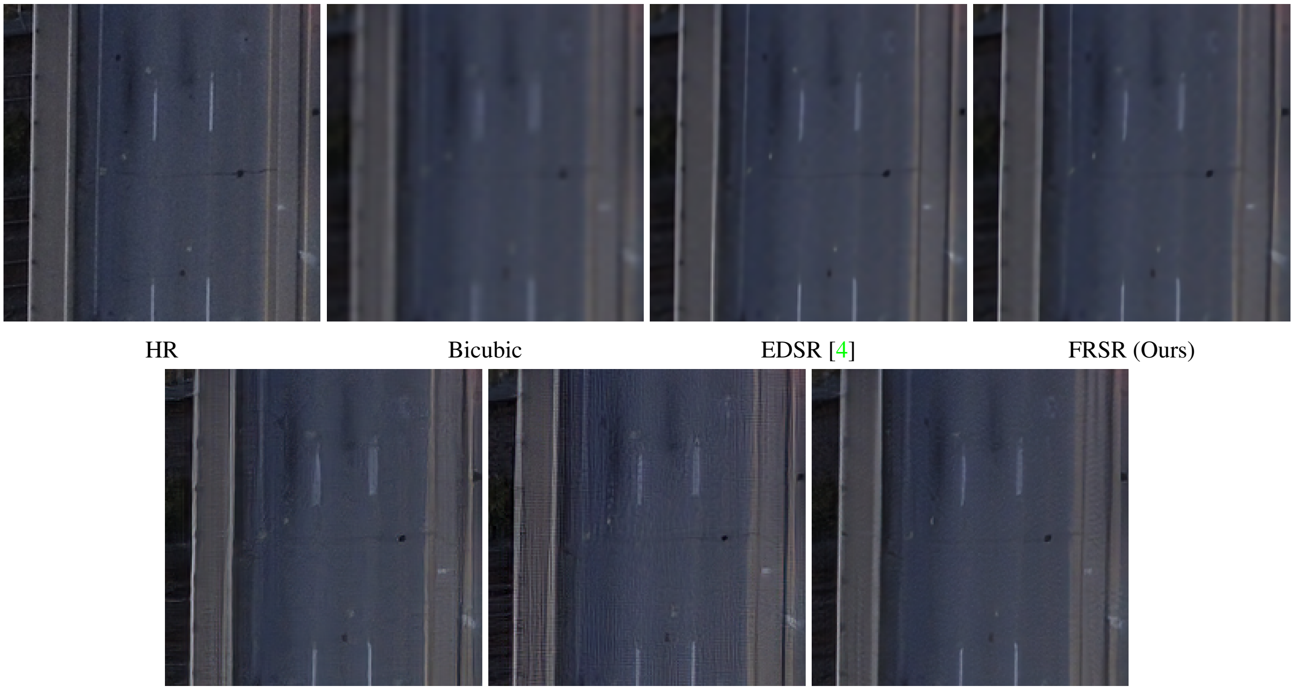


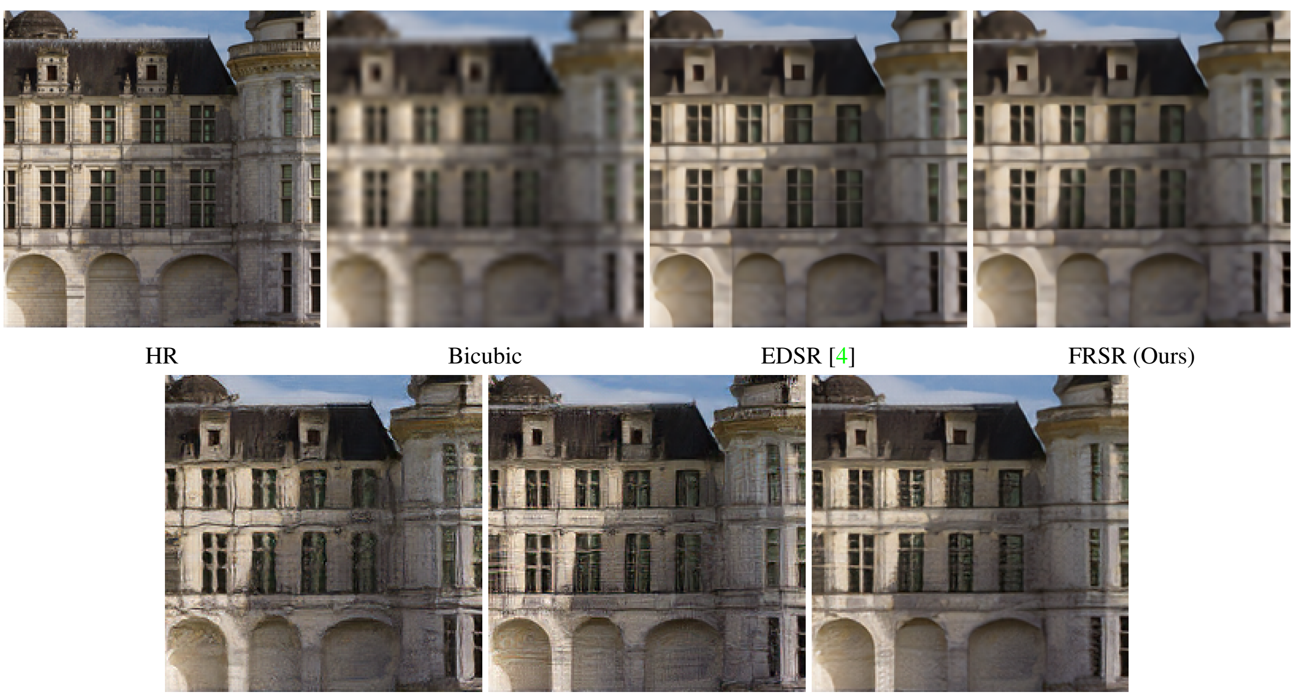
Figure 4: Visualized results on “0812” of DIV2K validation set [9].



HR | Bicubic | EDSR [4] | FRSR (Ours)

ENet [7] | SFT-GAN [10] | NatSR (Ours)

Figure 5: Visualized results on “0827” of DIV2K validation set [9].



HR | Bicubic | EDSR [4] | FRSR (Ours)

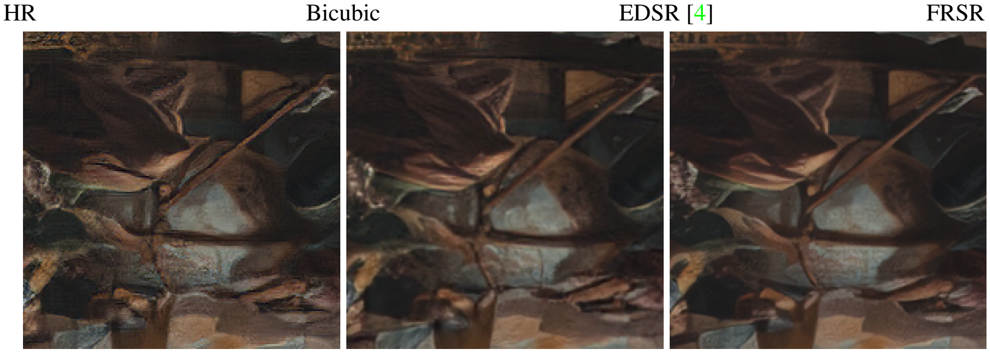
ENet [7] | SFT-GAN [10] | NatSR (Ours)

Figure 6: Visualized results on “0830” of DIV2K validation set [9].

HR Bicubic EDSR [4] FRSR (Ours)

ENet [7] SFT-GAN [10] NatSR (Ours)

HR Bicubic EDSR [4] FRSR (Ours)



ENet [7] SFT-GAN [10] NatSR (Ours)

Figure 7: Visualized results on "0841" of DIV2K validation set [9].

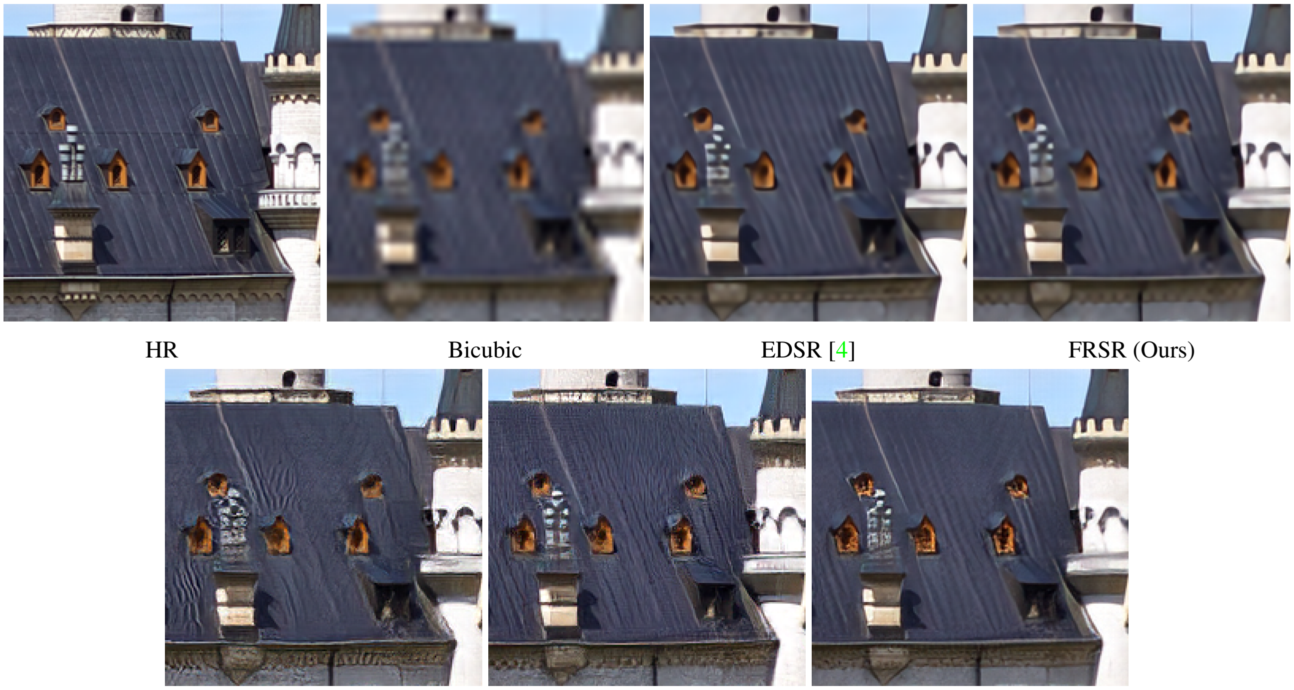


Figure 8: Visualized results on “0887” of DIV2K validation set [9].

Figure 9: Visualized results on “19021” of BSD100 [5].

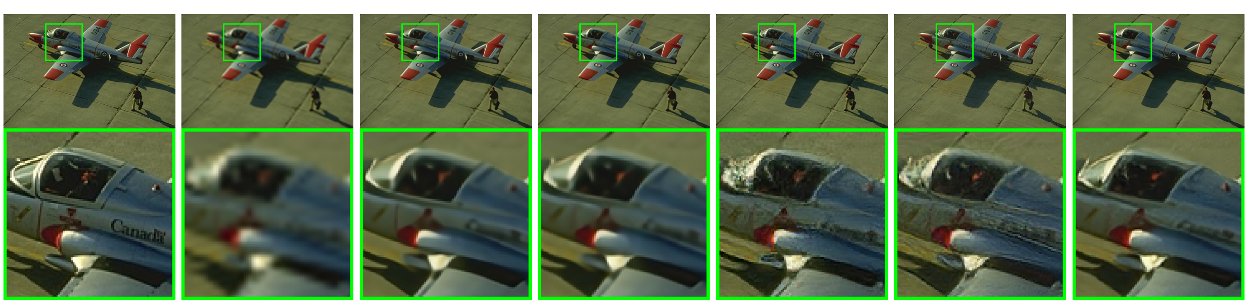


Figure 10: Visualized results on “37073” of BSD100 [5].

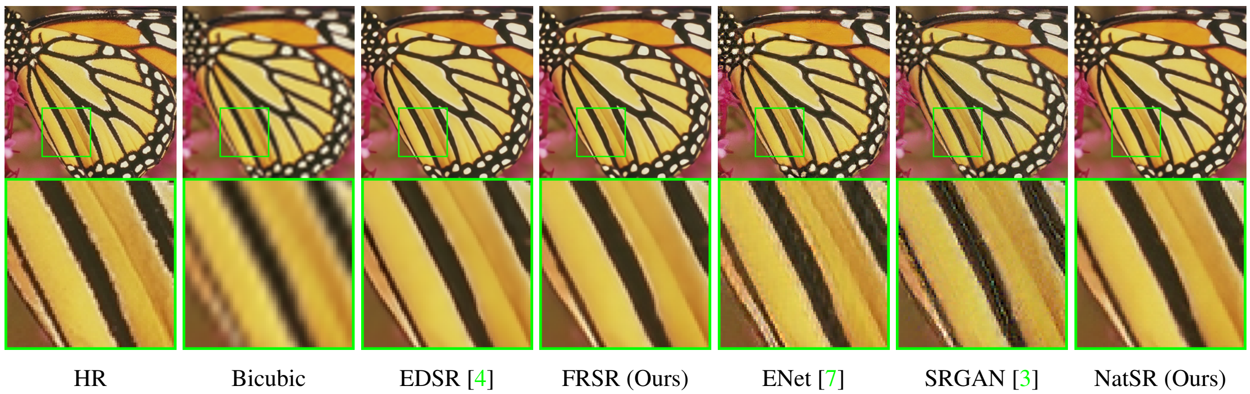


Figure 11: Visualized results on “butterfly” of Set5 [1].

Figure 12: Visualized results on “014” of Urban100 [2].

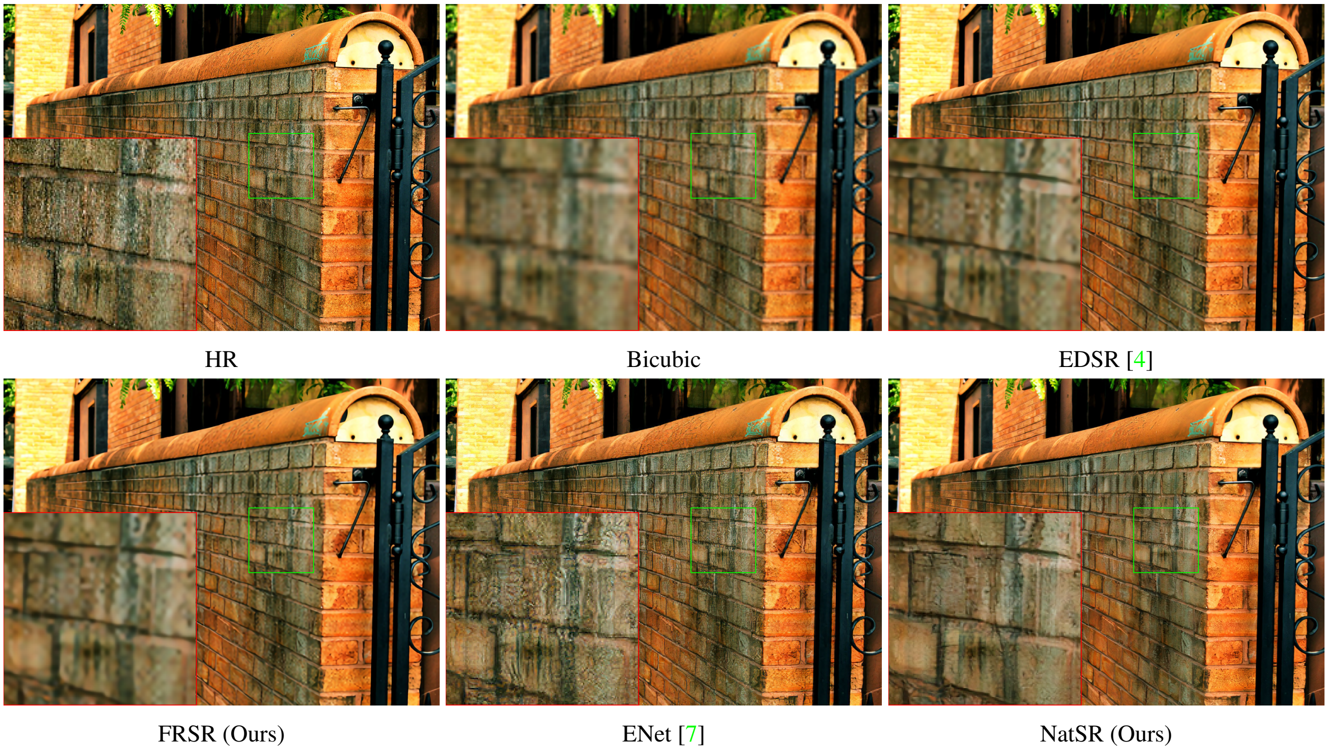


Figure 13: Visualized results on "018" of Urban100 [2].

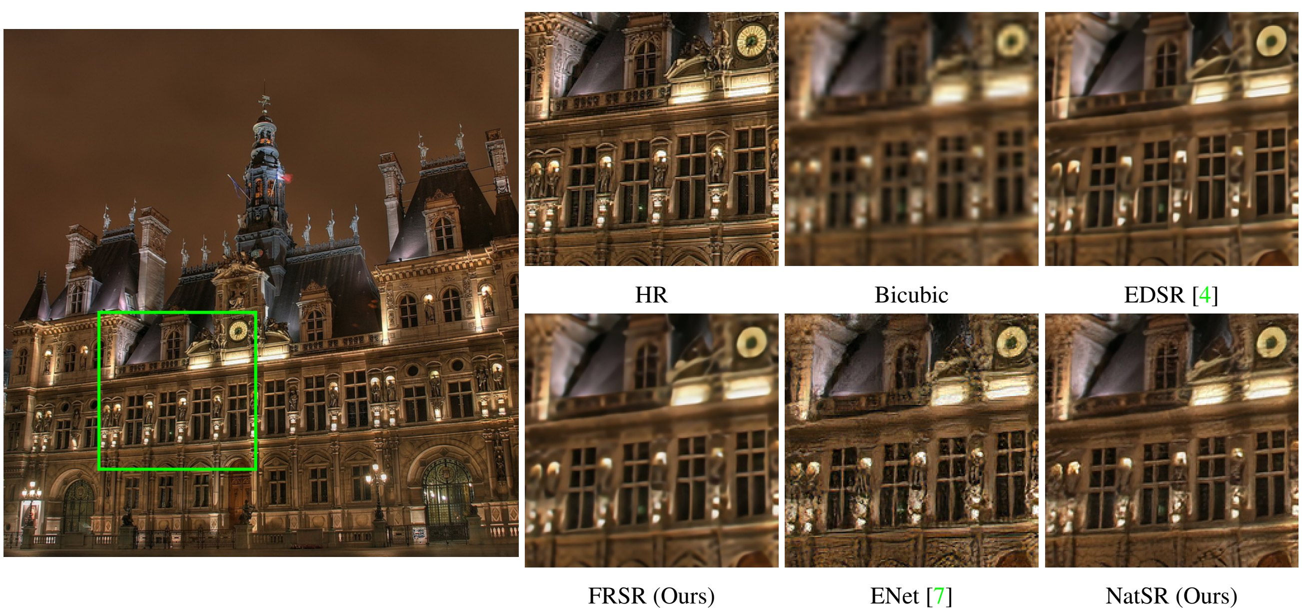


Figure 14: Visualized results on "022" of Urban100 [2].