# Supplementary: Tackling the Ill-Posedness of Super-Resolution through Adaptive Target Generation

Younghyun Jo[1] Seoung Wug Oh[2] Peter Vajda[3] Seon Joo Kim[1]
[1]Yonsei University [2]Adobe Research [3]Facebook

## A. Justification for Using Affine Transformation

The key ideas of our work are the new training scheme and the concept of the adaptive target. We chose affine transformation as a way to validate our idea since it is a simple and basic method in handling geometric deformation. There are other options such as projective transformation or a deformation field. However, as shown in Table 5 in the main paper, Tr + Rot model already shows better performance than Tr + Rot + Shear + Scale model. We infer that above a certain level of flexibility in transformation, the transformation alters the original contents in an undesirable way for SR, and we chose affine transformation for its simplicity and effectiveness.

We would also like to point out that we use $7 \times 7$ patch-level affine transformation, not a single transformation for the whole image for flexibility. We conducted additional empirical analysis on the effectiveness of the patch-level affine transformation. We gathered 10 HR patches $y^j$ which are approximately mapped to the same LR patch $x$ using corresponding kernels $k^j$ as follows:

$$x \approx (y^j * k^j) \downarrow_s, \tag{1}$$

where $j \in [1, 10]$ denotes the index of different HRs and $s = 4$. By using our ATG, we generated $y^{1\to2}$, ..., $y^{1\to10}$ (*i.e.* $y^{1\to m}$ = ATG$(y^1, y^m)$ where $m \in [2, 10]$). To check how accurate the generated results are, we compared PSNR values before and after the transformation by computing PSNR($y^1$, $y^m$) and PSNR($y^{1\to m}$, $y^m$) respectively. We computed this for 130 texture regions (total 1300 samples) and the mean PSNR improves from 23.80 to 28.82. In comparison, RANSAC based optimization with SIFT features for image-level affine and perspective transformation shows the PSNR values of 24.06 and 24.08 respectively which are much lower than ours. This experiment empirically shows that our approach is effective for 1) modeling the small difference in HR images which all map to the same LR image and 2) preserving the original contents. Note that it was difficult to apply the conventional optimization to estimate the patch-level transformation due to the difficulty of extracting features in our small $7\times7$ input. On the other hand, ATG estimates affine matrix in a single-shot manner and was fine-tuned together with SR net to maximize the performance.

## B. User Study

We conducted user studies on both blind and non-blind SR (bicubic downsampling) to verify the perceptual satisfaction of our results. In the user studies, we show 4 patches of images from different methods to participants and ask them to determine the ranking according to the visual quality. Our user interface for the user study is shown in Fig. 1, and we provide a movable box to let the participants see everywhere in a given image.

For the blind SR survey, IKC [4], KG [1] + ZSSR [9], and SRMD [14] are used for comparisons on Gaussian8 [4] testset for isotropic Gaussian kernels and DIV2KRK [1] testset for random kernels. Note that the results of SRMD are reproduced using GT kernel information. For the Gaussian8 testset, we select 10 images from Urban100 [5] for $\sigma = 1.8$ and $\sigma = 3.2$, therefore, total 20 images are used. We also select 20 images for the DIV2KRK testset. For non-blind SR survey, ESRGAN [12], NatSR [11], and RankSRGAN [15] are used for comparisons on selected 20 images from Urban100 testset.

The comparison is performed 30 times for each image. The surveys are conducted with Amazon Mechanical Turk, and the results are shown in Fig. 2. For the confidence of the results, we do a sanity check using questions with an obvious answer. All responses from the participants who failed to pass the sanity check are excluded for the final results. Rank 1 is for the best quality, and rank 4 is for the worst one. In all comparisons, people mostly prefer our results on average.

## C. More Results

More visual results of blind SR on Gaussian8 and DIV2KRK testsets are shown in Fig. 3 and Fig. 4 respectively. Also, more visual results of non-blind SR (bicubic downsampling) are shown in Fig. 5. The overall clarity is improved by applying our method.

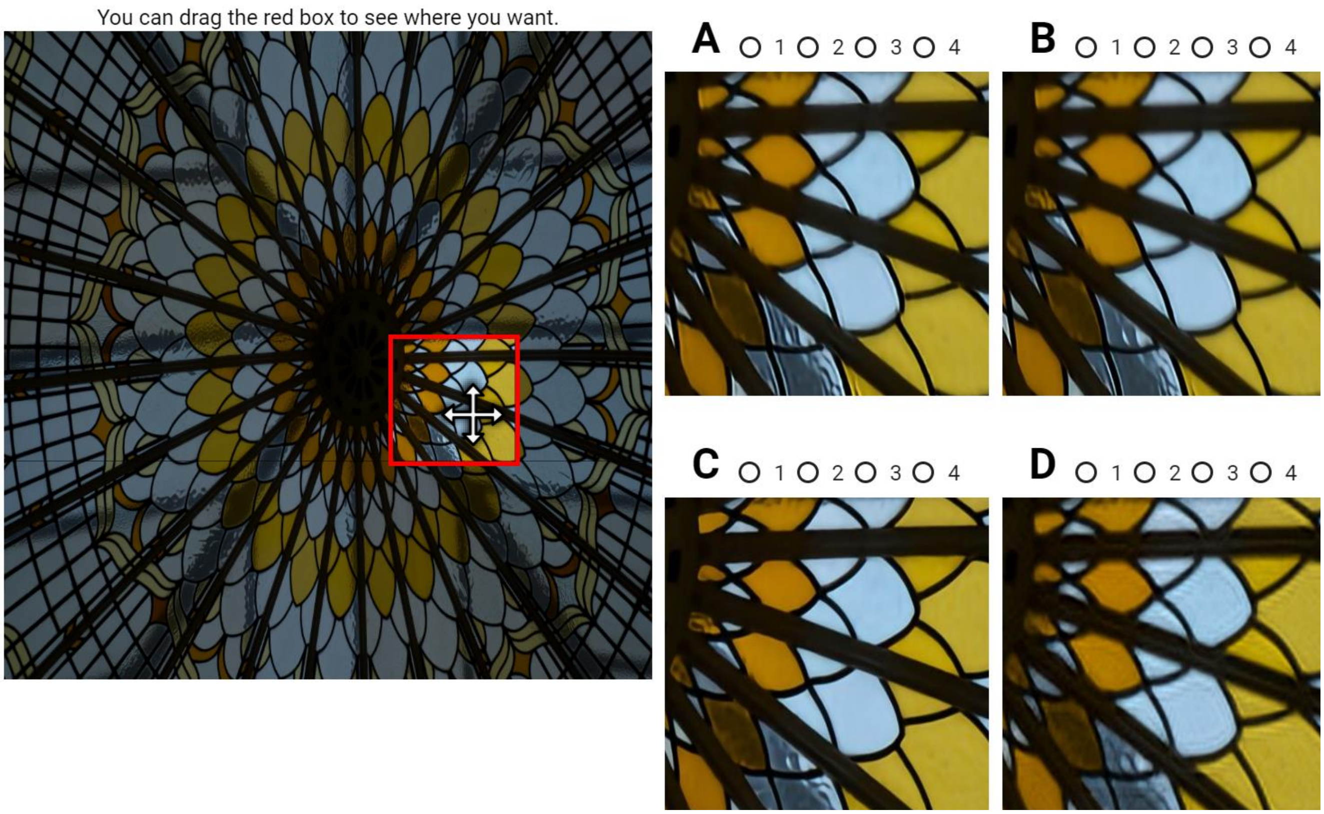
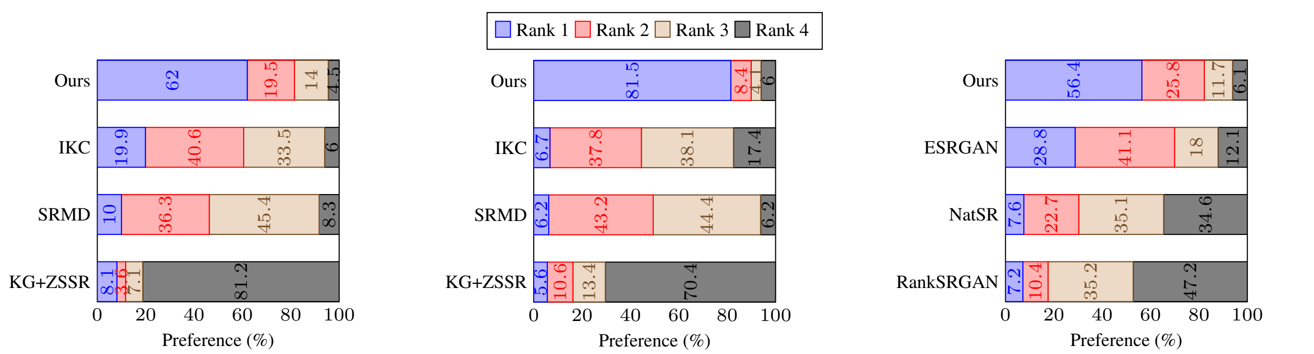


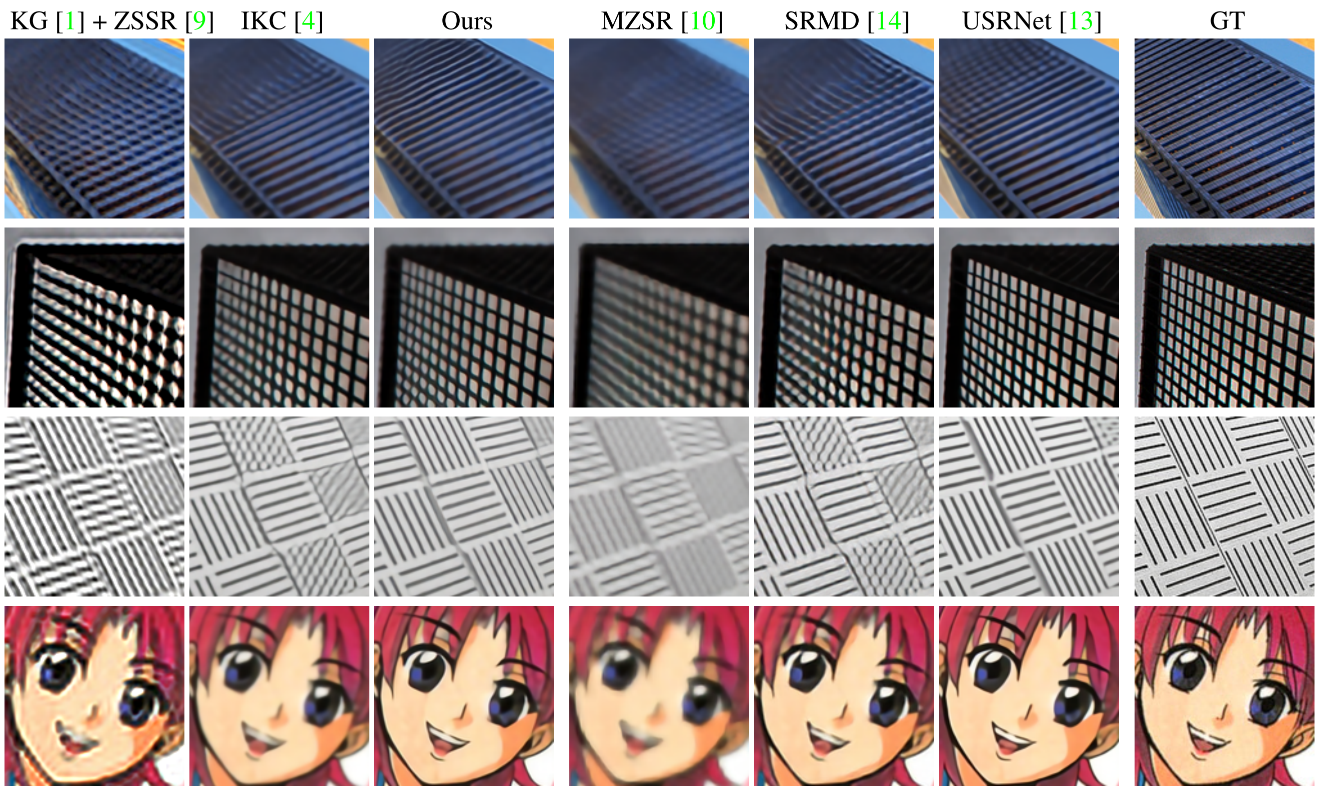

Figure 1. Our user interface for the user study. Participants are asked to rank among 4 images. Users can watch different areas by dragging the red box.



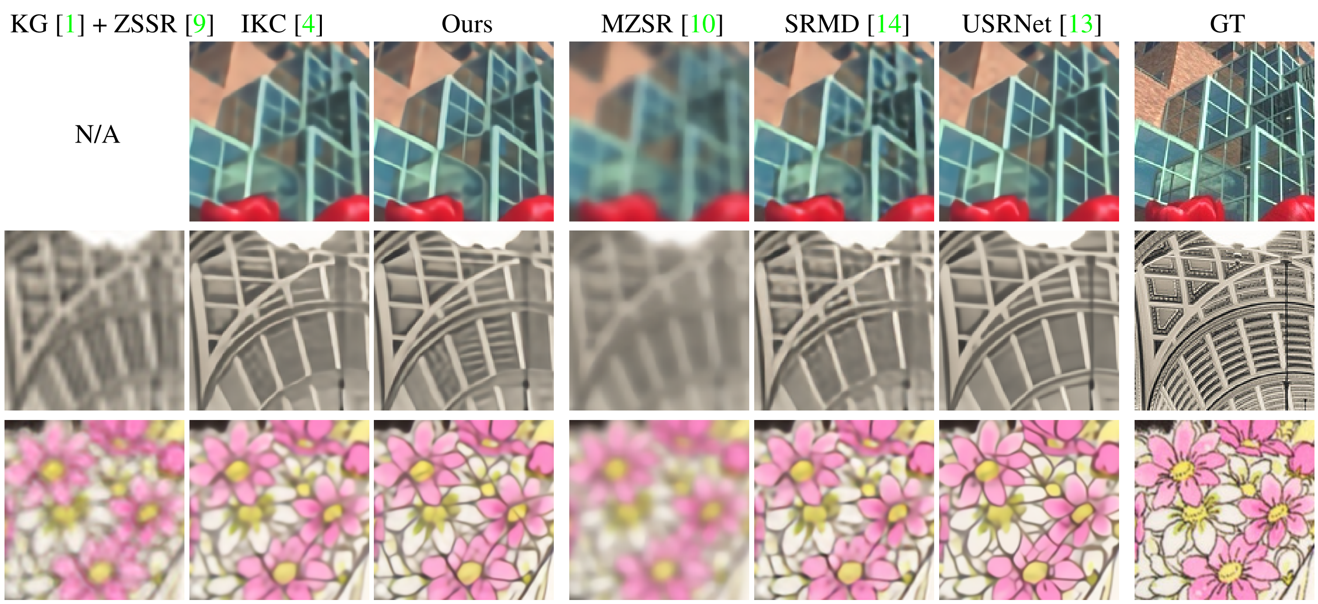

(a) Isotropic Gaussian kernels (b) Random kernels (c) Bicubic downsampling

Figure 2. The results of user studies. (a) Blind SR results on Gaussian8 testset. (b) Blind SR results on DIV2KRK testset. (c) Non-blind SR (bicubic downsampling) results on Urban100 testset. Our results mostly get rank 1 for the all surveys.

In addition, we verified that our method is also effective for real-world single-image SR. We follow the same experimental setting as in the main paper, specifically, we trained RRDB [12] network with our ATG by using a dataset proposed in [3]. For the upscaling factor 4, we achieved 29.03 PSNR value for the 100 test images, and it is better than the winner of the NTIRE 2019 challenge [2].




(a) $\sigma = 1.8$




(b) $\sigma = 3.2$

Figure 3. More visual results on Gaussian8 testset for isotropic Gaussian kernels. (a) The results for the kernel width 1.8 and (b) 3.2. Our results successfully restore sharp edges and look closer to the GT.

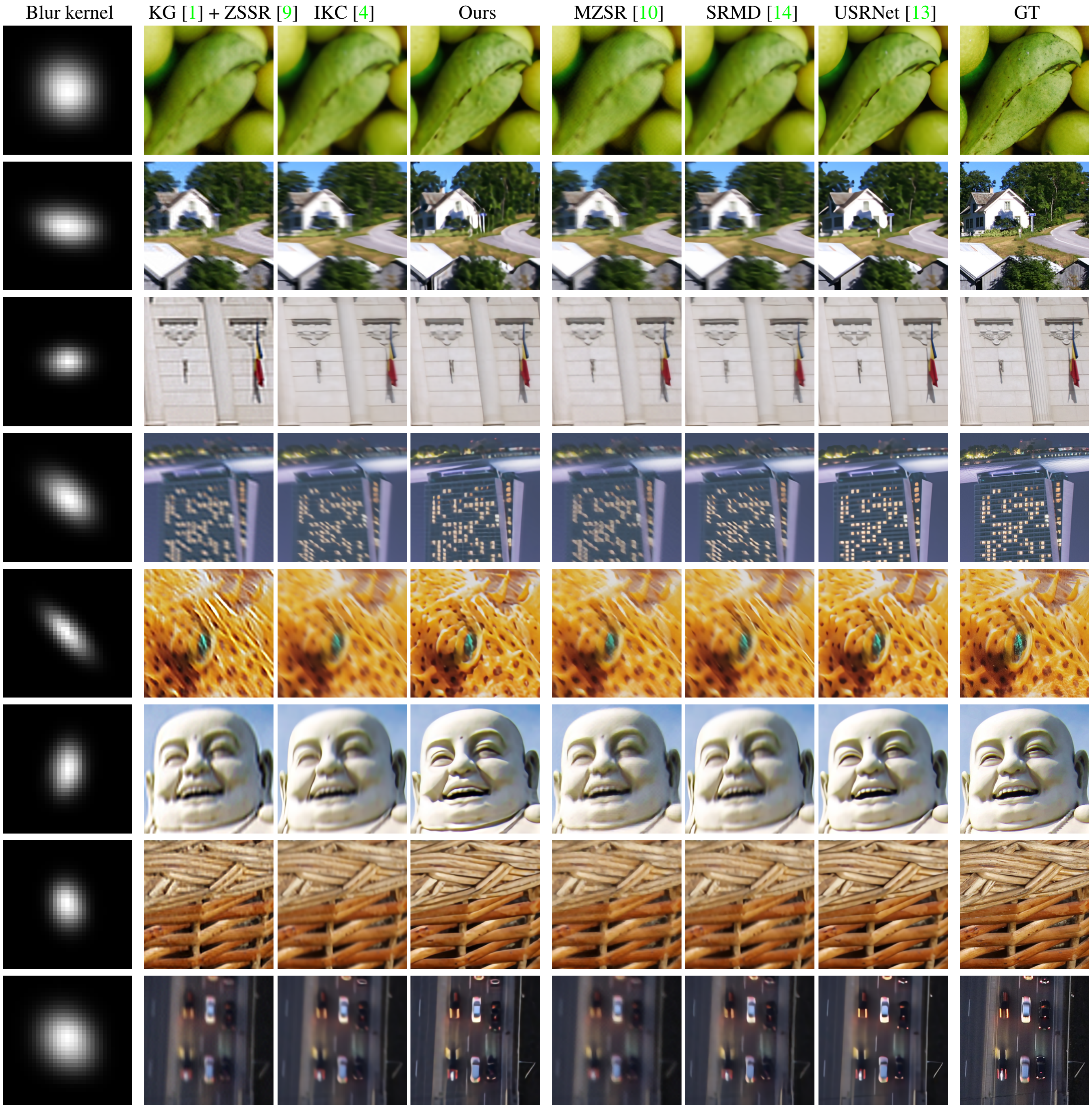



Figure 4. More visual results on DIV2KRK testset for random kernels. Each low-resolution test image is simulated by different blur kernels shown on the left. Our results show obvious sharpness and clarity compared to other results regardless of the kernel shapes.

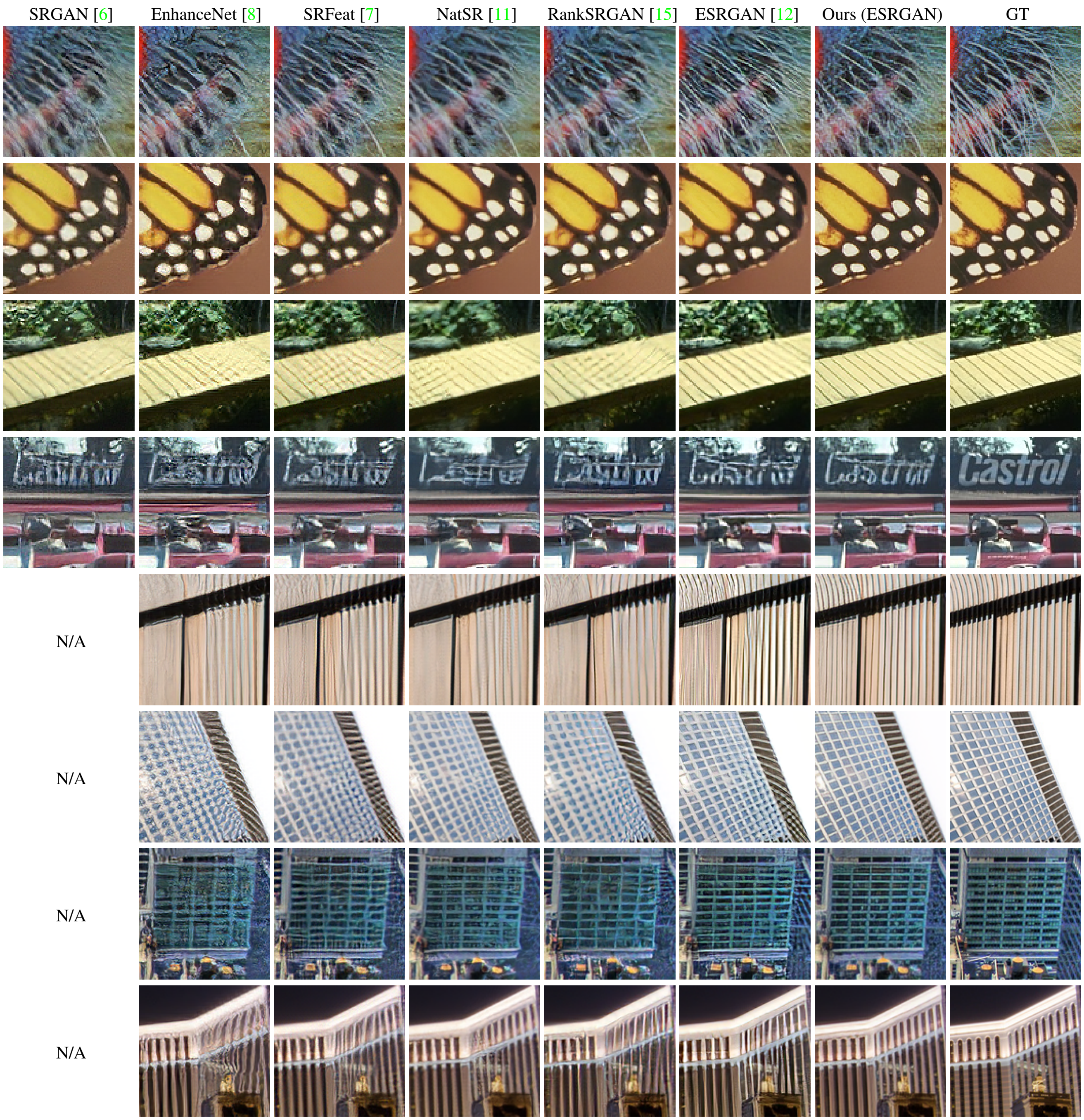



Figure 5. More visual results for non-blind SR (bicubic downsampling). Our method generates consistent details that correspond to the GT.

## References


[1] Sefi Bell-Kligler, Assaf Shocher, and Michal Irani. Blind super-resolution kernel estimation using an internal-gan. In *NeurIPS*, pages 284–293, 2019. 1, 3, 4

[2] Jianrui Cai, Shuhang Gu, Radu Timofte, and Lei Zhang. Ntire 2019 challenge on real image super-resolution: Methods and results. In *CVPR Workshops*, 2019. 2

[3] Jianrui Cai, Hui Zeng, Hongwei Yong, Zisheng Cao, and Lei Zhang. Toward real-world single image super-resolution: A new benchmark and a new model. In *ICCV*, 2019. 2

[4] Jinjin Gu, Hannan Lu, Wangmeng Zuo, and Chao Dong. Blind super-resolution with iterative kernel correction. In *CVPR*, pages 1604–1613, 2019. 1, 3, 4

[5] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *CVPR*, pages 5197–5206, 2015. 1

[6] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo-realistic single image super-resolution using a generative adversarial network. In *CVPR*, July 2017. 5

[7] Seong-Jin Park, Hyeongseok Son, Sunghyun Cho, Ki-Sang Hong, and Seungyong Lee. Srfeat: Single image super-resolution with feature discrimination. In *ECCV*, pages 439–455, 2018. 5

[8] Mehdi S. M. Sajjadi, Bernhard Scholkopf, and Michael Hirsch. Enhancenet: Single image super-resolution through automated texture synthesis. In *ICCV*, Oct 2017. 5

[9] Assaf Shocher, Nadav Cohen, and Michal Irani. "zero-shot" super-resolution using deep internal learning. In *CVPR*, pages 3118–3126, 2018. 1, 3, 4

[10] Jae Woong Soh, Sunwoo Cho, and Nam Ik Cho. Meta-transfer learning for zero-shot super-resolution. In *CVPR*, pages 3516–3525, 2020. 3, 4

[11] Jae Woong Soh, Gu Yong Park, Junho Jo, and Nam Ik Cho. Natural and realistic single image super-resolution with explicit natural manifold discrimination. In *CVPR*, June 2019. 1, 5

[12] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *ECCV Workshops*, September 2018. 1, 2, 5

[13] Kai Zhang, Luc Van Gool, and Radu Timofte. Deep unfolding network for image super-resolution. In *CVPR*, pages 3217–3226, 2020. 3, 4

[14] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Learning a single convolutional super-resolution network for multiple degradations. In *CVPR*, pages 3262–3271, 2018. 1, 3, 4

[15] Wenlong Zhang, Yihao Liu, Chao Dong, and Yu Qiao. Ranksrgan: Generative adversarial networks with ranker for image super-resolution. In *ICCV*, October 2019. 1, 5