Supplementary Material to "GDPO-SR: Group Direct Preference Optimization for One-Step Generative Image Super-Resolution"

The following materials are provided in this supplementary file (unless otherwise specified, GDPO-SR adopts $t_{add} = 250$ and $t_{diff} = 100$ during inference):

A. The average performance comparison between the baseline and GDPO-SR on the RealSR dataset;
B. Controlling the generative capability of GDPO-SR through the timestep t_{add};
C. Ablation study;
D. More visual comparisons (referring to Sec. 5.2 in the main paper);
E. Comparisons with GAN-based methods.

A. Average performance comparison

The model exhibits varying performance under different sampling noise. To ensure a fair and reliable comparison, we conduct 50 independent experiments on the RealSR dataset and evaluate the average performance of NAOSD and GDPO-SR. The results, presented in Table 1, show that GDPO-SR achieves superior average performance compared to NAOSD, indicating that the overall performance of the model is enhanced after reinforcement learning.

B. Control of Generative Capability

The generative capability of GDPO-SR can be controlled by adjusting the diffusion timestep t_{add} during inference. This adjustment allows the model to balance fidelity and realism according to different requirements. The quantitative results on the RealSR dataset are presented in Table 2. Note that the GDPO-SR model is fixed, only t_{add} is adjusted during inference, and all variants are fed with the identical noise for the same input. As can be seen, t_{add} provides an effective way to control the model's generative capability: larger t_{add} lead to higher no-reference metric scores, indicating stronger generative capability.

C. Ablation Study

C.1. Ablation Study on the group size.

We conducted four experiments to investigate the effect of group size (G), with G set to 4, 6, and 8, respectively. The results are presented in Table 3. It can be observed that as G increases, the model's generative capability improves, as indicated by the higher NR metric. This suggests that a larger G leads to more diverse sample generation, thereby enhancing the model's overall generative performance. In contrast, when G is too small, the limited diversity of samples restricts effective preference learning, resulting in minimal performance gains.

Table 1. Performance comparison (averaged over 50 stochastic runs) on the RealSR dataset. Arrows denote if higher (↑) or lower (↓) values represent better performance. The best results are highlighted in **red**.

Method	PSNR↑	SSIM↑	LPIPS↓	FID↓	DISTS↓	MANIQA↑	MUSIQ↑	CLIPIQA↑	AFINE↓
NAOSD	25.27	**0.7354**	0.2685	116.67	0.1997	0.6458	68.85	0.6620	19.28
GDPO-SR	**25.49**	0.7336	**0.2673**	**116.03**	**0.1978**	**0.6607**	**69.21**	**0.6766**	**18.19**

Table 2. The impact of t_{add} on the generation capability of GDPO-SR. Arrows denote if higher (↑) or lower (↓) values represent better performance. The best and second best results are highlighted in **red** and **blue**, respectively.

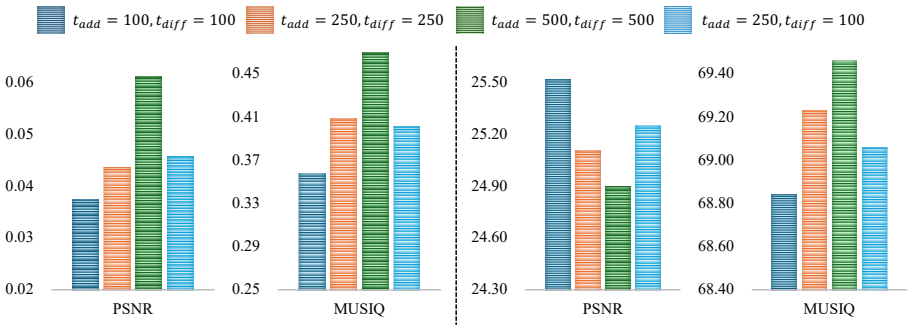
Method	PSNR↑	SSIM↑	LPIPS↓	DISTS↓	MANIQA↑	MUSIQ↑	CLIPIQA↑	AFINE↓
GDPO-SR (t_{add}=150)	**25.85**	**0.7499**	**0.2541**	**0.1938**	0.6521	68.09	0.6331	19.89
GDPO-SR (t_{add}=250)	**25.48**	**0.7328**	**0.2675**	**0.1980**	**0.6615**	**69.42**	**0.6760**	**17.73**
GDPO-SR (t_{add}=350)	24.84	0.7074	0.2900	0.2088	**0.6646**	**70.14**	**0.6994**	**17.41**

Table 3. Ablation studies on Group Size on the RealSR dataset.

G	PSNR↑	DISTS↓	MANIQA↑	CLIPIQA↑
NAOSD	25.25	0.2001	0.6459	0.6617
GDPO-SR (G=4)	25.45	0.1996	0.6491	0.6610
GDPO-SR (G=6)	25.48	0.1980	0.6615	0.6760
GDPO-SR (G=8)	25.34	0.1955	0.6619	0.6788

Table 4. Ablation studies on FR metrics on the RealSR dataset.

FR	PSNR↑	DISTS↓	MANIQA↑	CLIPIQA↑
NAOSD	25.25	0.2001	0.6459	0.6617
PSNR	25.48	0.1980	0.6615	0.6760
LPIPS	25.28	0.1997	0.6605	0.6835
PSNR+LPIPS	25.33	0.1970	0.6548	0.6701



(a) The impact of time on the fluctuation range of the metrics.

(b) Performance comparison of different time.

Figure 1. Ablation studies on the timestep setting in NAOSD.

C.2. Ablation Study on the timestep in NAOSD.

To investigate the impact of t=(t_{add}, t_{diff}) on NAOSD, we conduct four ablation experiments by settings t as (100,100), (250,250), (500,500) and (250,100). Fig. 1(a) shows the fluctuation range, which is defined as the difference between the maximum and minimum values for each metric, under the four settings. For each setting, we randomly sample 50 outputs per input on the RealSR dataset, yielding 50 values for each metric. As can be seen, when the timestep increases from (100,100) to (500,500), both PSNR and MUSIQ exhibit larger fluctuation ranges. This suggests that a larger t introduces more noise into the generation process, which increases randomness and enhances sample diversity. Fig. 1(b) presents the performance comparison across different t. As observed, increasing t enhances the generative ability but compromises fidelity. To address this trade-off, we adopt an unequal-time strategy ($t_{add} = 250$, $t_{diff} = 100$) to balance fidelity and generative capacity while maintaining a reasonable fluctuation range.

C.3. Ablation Study on FR reward metrics

We conducted four experiments on the RealSR dataset to investigate the effect of different FR metrics as reward function, including PSNR, LPIPS, and their combination PSNR+LPIPS. As shown in Table 4, using these FR metrics as rewards consistently improves performance. Since LPIPS is a semantic perceptual metric that is less sensitive to pixel-wise errors, using it alone as the reward results in slight PSNR gains. The combination PSNR+LPIPS alleviates this limitation, achieving a better balance.

C.4. Ablation Study on Sample Generation Methods in RL

There are various ways to generate multiple samples, such as adjusting the classifier-free guidance (CFG) scale or modifying the diffusion timestep t_{add}. In this paper, GDPO-SR generates multiple samples by altering the injected noise, enabling diverse outcomes from a single input. To investigate the effects of different sample generation methods, we conduct an ablation study on the Real-ISR dataset, as shown in Table 5. GDPO-SR-CFG denotes the variant that generates samples by changing the CFG, while GDPO-SR-t_{add} represents the variant by varying t_{add}. As shown in Table 5, both the CFG-based and timestep-based sampling methods lead to notable improvements in no-reference metrics; however, they tend to degrade full-reference metrics. In contrast, generating multiple samples with different noises yields more consistent improvements across both no-reference and full-reference metrics.

This is mainly because, when generating multiple samples, changing the CFG or t_{add} introduces an inherent trade-off between fidelity and perceptual quality rather than improving both simultaneously. As demonstrated in Table 2, increasing t_{add} enhances perceptual quality (as reflected by higher no-reference metric scores) but degrades fidelity (as indicated by

Table 5. Ablation study on sample generation methods on the Real-ISR dataset. Arrows denote if higher (↑) or lower (↓) values represent better performance.

Method	PSNR↑	SSIM↑	LPIPS↓	DISTS↓	MANIQA↑	MUSIQ↑	CLIPIQA↑	AFINE↓
NAOSD	25.25	0.7346	0.2689	0.2001	0.6459	69.06	0.6617	18.90
GDPO-SR (CFG)	25.06	0.7174	0.2782	0.2049	0.6543	69.47	0.6815	19.61
GDPO-SR (t_{add})	25.29	0.7294	0.2871	0.2133	0.6477	69.46	0.6766	17.12
GDPO-SR	25.48	0.7328	0.2675	0.1980	0.6615	69.42	0.6760	17.73

Table 6. Quantitative comparison between GDPO-SR and the state-of-the-art GAN-based Real-ISR methods on synthetic and real-world datasets. The best and second best results are highlighted in **red** and **blue**, respectively. Arrows denote if higher (↑) or lower (↓) values represent better performance.

Datasets	Methods	PSNR↑	SSIM↑	LPIPS↓	FID↓	DISTS↓	MANIQA↑	MUSIQ↑	CLIPIQA↑	AFINE ↓
DIV2K-val	RealESRGAN	**24.29**	**0.6371**	**0.3112**	**37.64**	**0.2141**	**0.5501**	61.06	**0.5277**	56.20
	BSRGAN	**24.58**	0.6269	0.3351	44.23	0.2275	0.5247	**61.20**	0.5071	62.21
	LDL	23.83	**0.6344**	0.3256	42.29	0.2227	0.5350	60.04	0.5180	**53.19**
	GDPO-SR	23.92	0.6117	**0.2897**	**26.44**	**0.1965**	**0.6423**	**68.80**	**0.6929**	**41.56**
RealSR	RealESRGAN	**25.69**	**0.7616**	0.2727	**135.18**	**0.2063**	**0.5487**	60.18	0.4449	39.53
	BSRGAN	**26.39**	**0.7654**	**0.2670**	141.28	0.2121	0.5399	**63.21**	**0.5001**	45.69
	LDL	25.28	0.7567	0.2766	142.71	0.2121	0.5485	60.82	0.4477	**38.43**
	GDPO-SR	25.48	0.7328	**0.2675**	**112.13**	**0.1980**	**0.6615**	**69.42**	**0.6760**	**17.73**
DrealSR	RealESRGAN	**28.64**	**0.8053**	**0.2847**	**147.62**	**0.2089**	0.4907	54.18	0.4422	39.97
	BSRGAN	**28.75**	0.8031	0.2883	155.63	0.2142	0.4878	**57.14**	**0.4915**	44.87
	LDL	28.21	**0.8126**	**0.2815**	155.53	0.2132	**0.4914**	53.85	0.4310	**39.06**
	GDPO-SR	28.18	0.7839	0.2851	**138.87**	**0.2112**	**0.6180**	**65.63**	**0.7020**	**18.72**

lower full-reference metric scores), and vise versa. Empirically, adjusting the CFG exhibits a similar trend: higher CFG values tend to improve perceptual quality while reducing fidelity. In contrast, altering the noise can generate richer and more natural variations among samples, including rare but valuable cases that simultaneously achieve high fidelity and high perceptual quality. Due to the existence of these rare samples, generating multiple samples with different noises can simultaneously enhance the model's generative capability and fidelity.

D. More Visual Comparisons

We provide more visual comparisons in Fig. 2 to demonstrate the effectiveness of GDPO-SR. Firstly, compared with the base model NAOSD, the post-training method GDPO-SR reconstructs sharper and clearer textures (as shown in the first and second cases) and generates finer details (as seen in the third and fourth cases). Secondly, compared to other advanced SD-based methods, GDPO-SR also exhibits notable advantages. For instance, in the second case, GDPO-SR successfully restores the characters "a" and "y", whereas other methods struggle to produce accurate shapes. Overall, GDPO-SR delivers sharper structures and more natural details, demonstrating strong robustness and generalization in real-world scenarios.

E. Comparisons with GAN-based Methods

We compare GDPO-SR with three representative GAN-based Real-ISR methods: RealESRGAN [7], BSRGAN [10] and LDL [5]. The quantitative results are summarized in Table 6. As observed, GDPO-SR achieves the best performance on most no-reference metrics (MANIQA [9], MUSIQ [4], CLIPIQA [6], and AFINE [3]) across all the three test datasets (DIV2K-val [1], RealSR [2], and DRealSR [8]). For full-reference metrics (*e.g.*, LPIPS, DISTS, and FID), GDPO-SR also delivers competitive results. The visual comparisons are illustrated in Fig. 3. It can be clearly found that the proposed GDPO-SR method can generate more realistic details than those GAN-based methods. These results demonstrate that GDPO-SR effectively balances fidelity and perceptual quality, surpassing GAN-based counterparts in overall visual realism and quantitative performance.

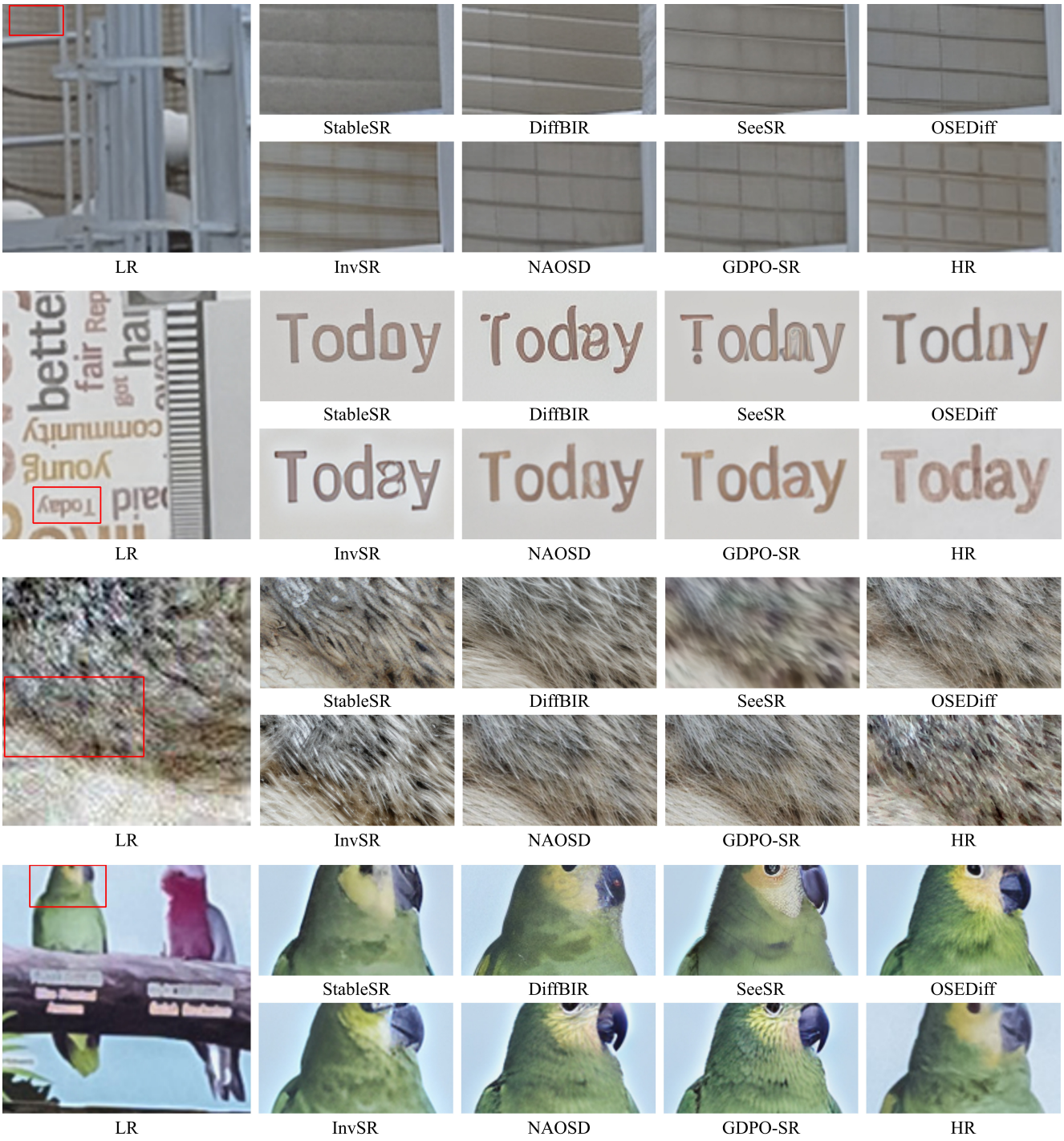


Figure 2. Visual comparison with SD-based Real-ISR methods. Please zoom in for a better view.

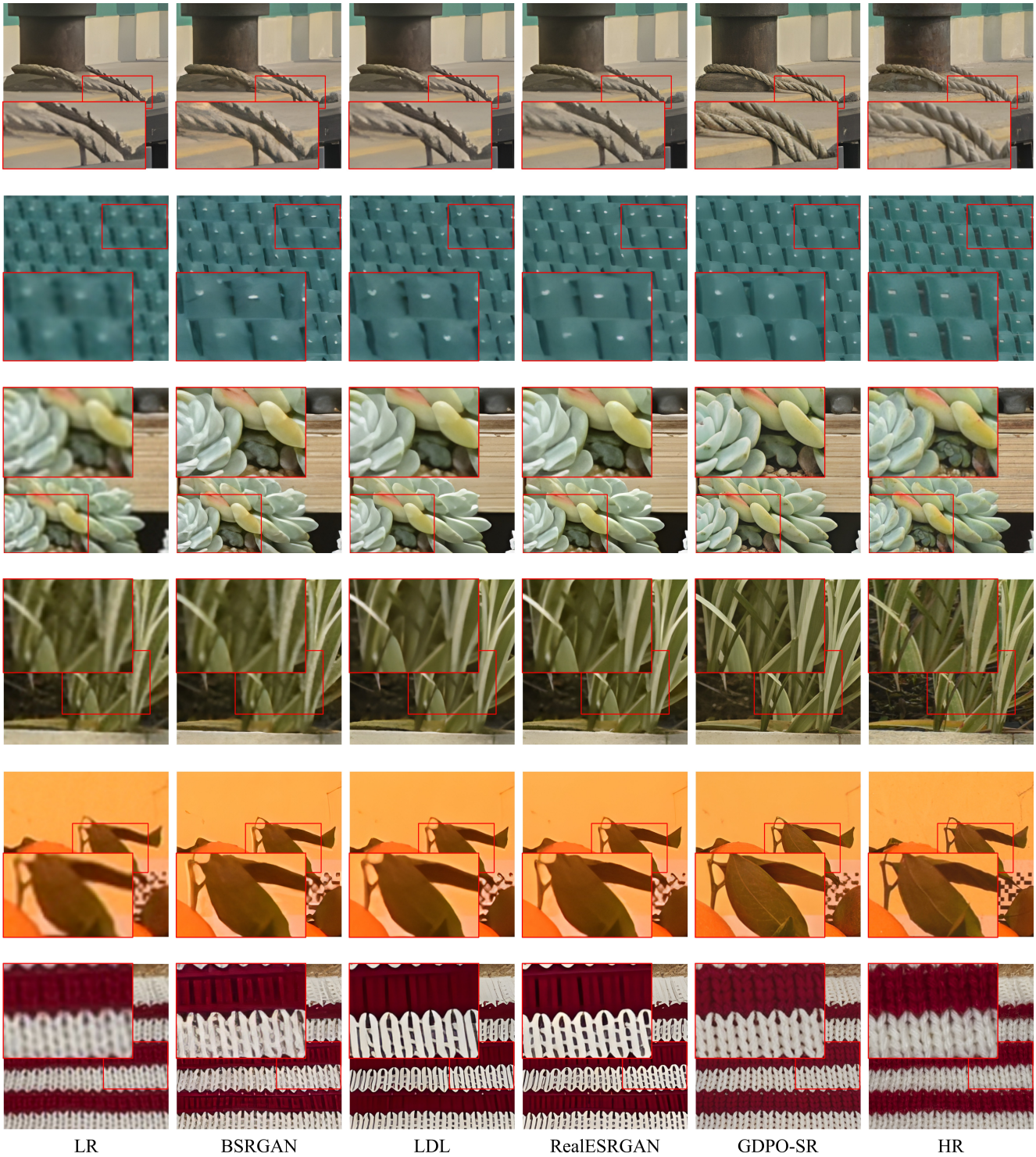


Figure 3. Visual comparison with GAN-based Real-ISR methods. Please zoom in for a better view.

References

[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 126–135, 2017. 3

[2] 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 *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3086–3095, 2019. 3

[3] Du Chen, Tianhe Wu, Kede Ma, and Lei Zhang. Toward generalized image quality assessment: Relaxing the perfect reference quality assumption. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 12742–12752, 2025. 3

[4] Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. Musiq: Multi-scale image quality transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5148–5157, 2021. 3

[5] Jie Liang, Hui Zeng, and Lei Zhang. Details or artifacts: A locally discriminative learning approach to realistic image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5657–5666, 2022. 3

[6] Jianyi Wang, Kelvin CK Chan, and Chen Change Loy. Exploring clip for assessing the look and feel of images. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 2555–2563, 2023. 3

[7] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1905–1914, 2021. 3

[8] Pengxu Wei, Ziwei Xie, Hannan Lu, Zongyuan Zhan, Qixiang Ye, Wangmeng Zuo, and Liang Lin. Component divide-and-conquer for real-world image super-resolution. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VIII 16*, pages 101–117. Springer, 2020. 3

[9] Sidi Yang, Tianhe Wu, Shuwei Shi, Shanshan Lao, Yuan Gong, Mingdeng Cao, Jiahao Wang, and Yujiu Yang. Maniqa: Multi-dimension attention network for no-reference image quality assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1191–1200, 2022. 3

[10] Kai Zhang, Jingyun Liang, Luc Van Gool, and Radu Timofte. Designing a practical degradation model for deep blind image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4791–4800, 2021. 3