Supplementary Material for
Self-Adaptive Reality-Guided Diffusion for Artifact-Free Super-Resolution

Qingping Zheng[1], Ling Zheng[3], Yuanfan Guo[2], Ying Li[1*], Songcen Xu[2], Jiankang Deng[2], Hang Xu[2]

[1]Northwestern Polytechnical University [2]Huawei Noah's Ark Lab [3]Tsinghua-Fuzhou Institute for Data Technology

1. Comparison with More State-of-the-Art

In this section, we expand our comparison by including additional diffusion-based SR methods to further validate the effectiveness of our proposed Self-Adaptive Reality-Guided Diffusion (SARGD) method. Table 1 presents a comparison of more state-of-the-art Diffusion-SR techniques, including LDM [2], StableSR [3], ResShift [5], PASD [4], and DiffBIR [1], across benchmark datasets for super-resolution at a scale of $\times 4$. The results demonstrate that our training-free SARGD method consistently outperforms these competing methods in terms of both PSNR and SSIM metrics across all datasets.

The SARGD method achieves the highest PSNR values, indicating superior image quality with fewer artifacts and distortions when enlarging images. Specifically, it reaches a PSNR of 32.27 on Set5, 30.01 on Set14, 30.23 on B100, 27.93 on Urban100, and 30.23 on Manga109, marking it as the leader in generating high-fidelity images across diverse types of content, ranging from natural scenes in B100 to the intricate details found in Manga109.

Similarly, the SARGD's performance in SSIM, a metric that measures the perceptual quality of images and their structural similarity to the original, further solidifies its advantage. With scores of 0.871 on Set5, 0.778 on Set14, 0.763 on B100, 0.771 on Urban100, and 0.851 on Manga109, SARGD proves its efficacy in maintaining the structural integrity and textural details of the original images while upscaling, which is crucial for applications requiring high visual fidelity.

The quantitative results underscore the effectiveness of the SARGD approach in delivering high-quality, artifact-free super-resolution images. This superiority is particularly noteworthy considering the evaluation was conducted on a 32G GPU, which reflects a significant computational efficiency alongside its superior performance. These outcomes not only demonstrate the practicality of SARGD in resource-constrained environments but also highlight its potential for widespread adoption in applications requiring high-resolution image processing.

*Corresponding author

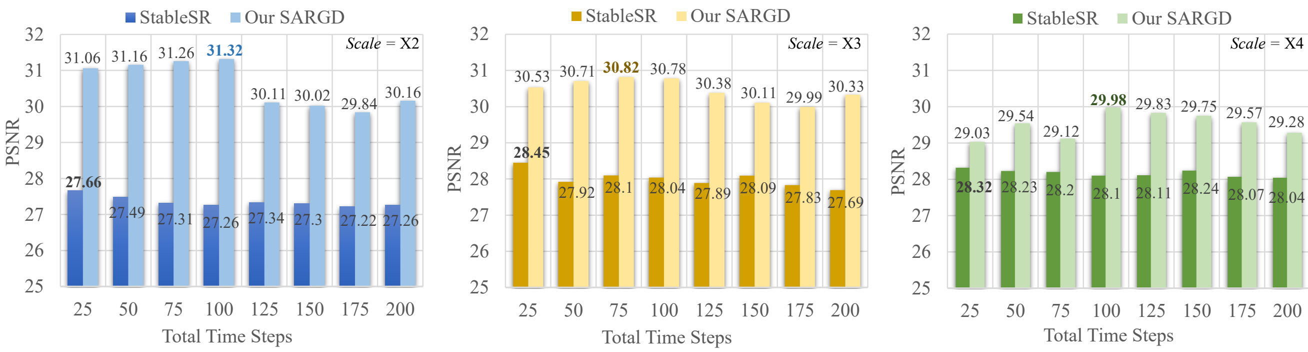
Table 1. **Quantitative comparison with the state-of-the-art Diffusion-SR methods across benchmark datasets for super-resolution at $\times 4$ scale. Bold** highlights the best performance. All evaluations are conducted on a 32G GPU. Our training-free SARGD attains the most favorable results.

Dataset	Metric	LDM	StableSR	ResShift	PASD	DiffBIR	Ours
Set5	**PSNR** ↑	28.30	29.36	31.74	32.04	30.16	**32.27**
	SSIM ↑	0.769	0.812	0.863	0.860	0.817	**0.871**
Set14	**PSNR** ↑	27.21	28.04	29.18	29.60	28.72	**30.01**
	SSIM ↑	0.696	0.733	0.760	0.770	0.719	**0.778**
B100	**PSNR** ↑	27.67	28.40	29.40	29.79	29.50	**30.23**
	SSIM ↑	0.679	0.717	0.732	0.747	0.718	**0.763**
Urban100	**PSNR** ↑	25.91	26.55	27.87	26.70	26.54	**27.93**
	SSIM ↑	0.719	0.744	0.769	0.707	0.703	**0.771**
Manga109	**PSNR** ↑	27.04	27.94	29.71	28.38	27.47	**30.23**
	SSIM ↑	0.815	0.838	0.848	0.828	0.791	**0.851**

2. More Ablation Study

In this section, we conduct additional ablation studies to thoroughly analyze the effectiveness of our proposed training-free SARGD method in achieving artifact-free super-resolution.

Performance of Different Time Steps. Figure 1 presents a comparative analysis of PSNR performance at scales $\times 2$, $\times 3$, and $\times 4$ between our SARGD and StableSR [3], employing various inference time steps. The results depicted in Figure 1(c) show that StableSR's performance decreases as the number of total time steps increases, likely due to a potential loss of detail from extended denoising processes. In contrast, our SARGD demonstrates a gradual improvement in super-resolution performance, boosting from a PSNR of 29.03 at 25 steps to 29.98 at 100 steps. Despite a minor decrease in performance observed between 100 to 200 steps, the PSNR at each interval remains higher than the score at 25 steps, indicating consistent enhancement over time. This is attributed to its self-refinement mechanism, which actively adjusts the middle latent representations to preserve and improve fidelity throughout the super-resolution process. Furthermore, our SARGD surpasses the performance of StableSR in terms of image quality across different enlargement scales while also achieving significantly more efficient processing with a reduction in inference steps from



(a) ×2 (b) ×3 (c) ×4

Figure 1. Performance comparison between our proposed SARGD and StableSR across various time steps and scales (×2, ×3, and ×4).

200 to just 25. Specifically, the comparative performance gains of SARGD over StableSR are evident at magnifications of ×2 (31.06 vs. 27.26), ×3 (30.53 vs. 27.69), and ×4 (29.03 vs. 28.04), indicating a notable improvement in image quality at these respective scales.

Compares Various Artifact Detections. To verify the effectiveness of the artifact detection model used in our proposed SAGD method, we compare the performance of different artifact detection methods, including PAL4VST [7] and PAL4Inpaint [6]. Based on the result presented in Table 2, we can draw several observations regarding the performance of different artifact detection methods including for super-resolution at a scale of ×4 across the Set14 and B100 datasets.

Firstly, both PAL4Inpaint and PAL4VST demonstrate significant improvements in PSNR, SSIM, and DISTS scores compared to the baseline method across both datasets. PAL4VST particularly stands out with the highest PSNR and SSIM scores, as well as the lowest DISTS score, indicating its superior performance in artifact detection for super-resolution.

Moreover, the effectiveness of artifact detection methods appears consistent across different datasets, as evidenced by similar trends in performance across Set14 and B100. This consistency suggests the robustness of these methods across various image types and characteristics.

Overall, PAL4VST emerges as the most effective artifact detection method, consistently outperforming both the baseline and PAL4Inpaint in terms of PSNR, SSIM, and DISTS scores across both datasets. This highlights its potential for improving the quality and fidelity of super-resolved images, making it a promising choice for practical applications in image enhancement.

Table 2. **Comparing the performance of different artifact detection methods across Set14 and B100 datasets for super-resolution at ×4 scale. Bold** highlights the best performance.

Method	Set14			B100		
	PSNR ↑	SSIM ↑	DISTS ↓	PSNR ↑	SSIM ↑	DISTS ↓
Baseline	27.21	0.696	0.236	27.67	0.679	0.252
PAL4Inpaint ($\mathcal{A}$)	29.65	0.776	0.183	30.01	0.760	0.163
PAL4VST ($\mathcal{A}$)	**30.01**	**0.778**	**0.156**	**30.23**	**0.763**	**0.156**

References

[1] Xinqi Lin, Jingwen He, Ziyan Chen, Zhaoyang Lyu, Ben Fei, Bo Dai, Wanli Ouyang, Yu Qiao, and Chao Dong. Diffbir: Towards blind image restoration with generative diffusion prior. *arXiv preprint arXiv:2308.15070*, 2023. 1

[2] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10684–10695, 2022. 1

[3] Jianyi Wang, Zongsheng Yue, Shangchen Zhou, Kelvin CK Chan, and Chen Change Loy. Exploiting diffusion prior for real-world image super-resolution. *arXiv preprint arXiv:2305.07015*, 2023. 1

[4] Tao Yang, Peiran Ren, Xuansong Xie, and Lei Zhang. Pixel-aware stable diffusion for realistic image super-resolution and personalized stylization. *arXiv preprint arXiv:2308.14469*, 2023. 1

[5] Zongsheng Yue, Jianyi Wang, and Chen Change Loy. Resshift: Efficient diffusion model for image super-resolution by residual shifting. *Advances in Neural Information Processing Systems*, 36, 2024. 1

[6] Lingzhi Zhang, Yuqian Zhou, Connelly Barnes, Sohrab Amirghodsi, Zhe Lin, Eli Shechtman, and Jianbo Shi. Perceptual artifacts localization for inpainting. In *European Conference on Computer Vision*, pages 146–164. Springer, 2022. 2

[7] Lingzhi Zhang, Zhengjie Xu, Connelly Barnes, Yuqian Zhou, Qing Liu, He Zhang, Sohrab Amirghodsi, Zhe Lin, Eli Shechtman, and Jianbo Shi. Perceptual artifacts localization for image synthesis tasks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 7579–7590, 2023. 2