IFCSR: Inference-Free Fidelity-Realism Control for One-Step Diffusion-based Real-World Image Super-Resolution (Supplementary Material)

1. Additional Ablation Studies

Comparison of two-stage training scheme with different orders While we adopt a two-stage scheme that first optimizes the fidelity-specific network and then trains the realism-specific network, one can reverse the order as an alternative (i.e., training the realism-specific network first). Table 1 compares our method with this alternative design. Since enhancing realism is relatively complicated compared to fidelity, training the realism-specific network in Stage 1 tends to produce images less specialized for realism (see the third row), thus resulting in a misaligned fidelity-realism spectrum in Stage 2. In contrast, training the fidelity-specific network in the first stage facilitates clearer fidelity-oriented learning (see the fourth row), leading to a well-aligned controllable space along the fidelity-realism spectrum.

Orders	γ	PSNR ↑	SSIM ↑	MUSIQ ↑	CLIPIQA ↑
real. → fid.	0.0	**25.55**	0.7722	**53.78**	**0.3669**
	0.5	**26.73**	**0.7823**	55.92	0.4209
	1.0	26.17	**0.7417**	**67.49**	**0.6410**
fid. → real. (default)	0.0	**28.00**	**0.7975**	**47.64**	**0.3337**
	0.5	25.14	0.7055	69.68	0.6522
	1.0	**22.10**	**0.6075**	**70.29**	**0.7011**

Table 1. Numerical results comparing the default two-stage training scheme and its alternative design with reversed training order, evaluated on the RealSR [2] dataset. 'fid.' and 'real.' indicate fidelity- and realism-specific networks, respectively. The **blue** and **brown** numbers denote the highest and lowest scores among different values of the controllable parameter γ.

Comparison of our method with an alternative design in Stage 2 During Stage 2, our method optimizes the trainable parameters using our controllable model that linearly combines the fidelity- and realism-specific images, i.e., $\mathbb{E}_{\gamma\sim U(0,1)}[\mathcal{L}_{\text{2nd}}(\hat{\mathbf{x}}_H, \mathbf{x}_H)]$. Another alternative design is to directly minimize the loss between the realism-specific image and the ground truth, i.e., $\mathcal{L}_{\text{2nd}}(\hat{\mathbf{x}}_H^{\text{real}}, \mathbf{x}_H)$. Table 2 shows the results under these two designs, using the same fidelity-specific network and realism-specific loss. As shown in the table, compared to the alternative, optimization with our controllable model strengthens no-reference metrics (MUSIQ [8] and CLIPIQA [12]) while degrading reference-based metrics (PSNR and SSIM [16]), ultimately producing a wider fidelity-realism spectrum.

Settings	γ	PSNR ↑	SSIM ↑	MUSIQ ↑	CLIPIQA ↑
real. only	0.0	**28.00**	**0.7975**	**47.64**	**0.3331**
	0.5	27.53	0.7831	59.59	0.4905
	1.0	**25.86**	**0.7302**	**67.90**	**0.6763**
our model (default)	0.0	**28.00**	**0.7975**	**47.64**	**0.3337**
	0.5	25.14	0.7055	69.68	0.6522
	1.0	**22.10**	**0.6075**	**70.29**	**0.7011**

Table 2. Numerical comparison of our method with an alternative design in Stage 2, evaluated on the RealSR dataset. "real. only" denotes optimization using the realism-specific image, while "our model" indicates optimization using our controllable model. The **blue** and **brown** numbers denote the highest and lowest scores among different values of the controllable parameter γ.

Comparison of our specialized losses with different k Fig. 1 analyzes the effect of the parameter k in our specialized losses. When $k = 1$, the losses behave like LPIPS [28] combined with MSE on unit-normalized inputs, resulting in a narrow fidelity-realism spectrum (see the black-colored lines in the top and middle rows). As k increases, larger weights are assigned to shallow-layer feature distances during fidelity-specific training, and to deep-layer feature distances during realism-specific training. This mechanism promotes stronger specialization in image-space fidelity (e.g., PSNR and SSIM) and realism (e.g., MUSIQ and CLIPIQA), thus extending the spectrum (see the other lines). While the fidelity-realism spectrum broadens, increasing k lowers overall quality across γ (i.e., shifting the curves toward the bottom-left region in the top and middle rows of the figure). To balance controllability and overall quality, we empirically set $k = 7$ as the default.

On the one hand, the bottom row of Fig. 1 indicates that our specialized losses with depth-dependent weighting amplify an inconsistency in the trade-off between feature-space fidelity (e.g., LPIPS) and perceptual realism. Compared to $k = 1$, where per-layer feature distances are equally utilized in our losses, we observe that higher values of k progressively accentuate the non-monotonic trend

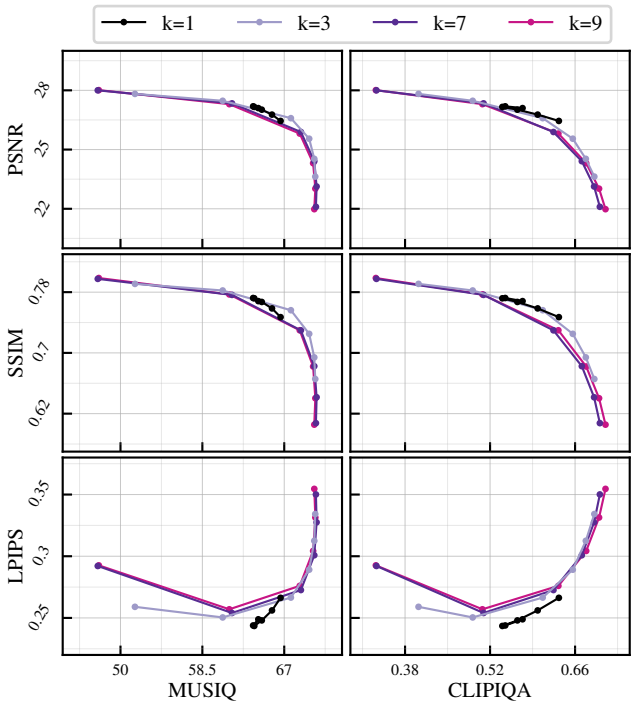

Figure 1. Effect of parameter $k \in \{1, 3, 7, 9\}$ in the specialized losses on the fidelity-realism spectrum, evaluated on the RealSR dataset. The x-axis shows MUSIQ and CLIPIQA, and the y-axis represents PSNR, SSIM, and LPIPS. With monotonically decreasing and increasing weighting, larger k places greater emphasis on shallow-layer feature distances during fidelity-specific training and on deep-layer feature distances during realism-specific training, leading to stronger specialization and a wider controllable spectrum.

of the curve. Nevertheless, the resulting spectra under the default setting of k along image-space fidelity and realism exhibit clear and wide, demonstrating that our design retains effective controllability.

2. Potential of VSD Loss into Our Framework

Adapting score distillation methods for ISR tasks [6, 11, 18] has been widely studied as an effective strategy for enhancing perceptual realism, and such methods may also benefit our controllable framework. To explore their potential benefits, we conduct a simple experiment in which we additionally apply the VSD loss from OSEDiff [18] to the predicted latent of the realism-specific network during Stage 2. As shown in Table 3, even this straightforward integration leads to noticeable improvements in no-reference metrics at the cost of reference-based metrics (see the third and sixth rows). These results indicate increased specialization toward realism via score distillation, and we leave more sophisticated integration strategies as a promising direction for future work.

Settings	γ	PSNR ↑	SSIM ↑	MUSIQ ↑	CLIPIQA ↑
Ours w/ VSD loss	0.0	**28.00**	**0.7975**	**47.64**	**0.3333**
	0.5	24.03	0.6849	70.42	0.6672
	1.0	**20.68**	**0.5833**	**70.99**	**0.7098**
Ours w/o VSD loss (default)	0.0	**28.00**	**0.7975**	**47.64**	**0.3337**
	0.5	25.14	0.7055	69.68	0.6522
	1.0	**22.10**	**0.6075**	**70.29**	**0.7011**

Table 3. Ablation study on incorporating the VSD loss from OSEDiff into our framework during Stage 2, evaluated on the RealSR dataset. The **blue** and **brown** numbers highlight the highest and lowest scores among different values of the parameter γ.

3. Additional Results and Comparisons

Visual results of our method with different γ Fig. 2 shows additional visual results of our method under different values of the controllable parameter γ. As γ increases from zero to one, the outputs gradually shift from prioritizing fidelity to realism, effectively preserving fine structures and sharpening details, e.g., windows (1st row), hairs (2nd row), clothing (3rd row), and leaves (4th row). This smooth control through γ clearly shows the effectiveness of our controllable approach.

Additional comparisons with diffusion-based ISR methods Table 4 and Fig. 3 present additional quantitative and visual comparisons among the state-of-the-art diffusion-based ISR approaches (StableSR [13], DiffBIR [10], SeeSR [19], PASD [21], ResShift [24], SinSR [15], OSEDiff [18], AdcSR [3], TSD-SR [6], InvSR [25], TVT [22], CTMSR [23], and PiSA-SR [11]). Under the realism-oriented setting ($\gamma = 1.0$), our method achieves high scores on no-reference metrics (e.g., the second-best MUSIQ, MANIQA, and CLIPIQA on DIV2K-val [1]) and generates rich textures and visually pleasing details. On the other hand, the fidelity-oriented setting ($\gamma = 0.0$) achieves the highest PSNR and SSIM scores, producing relatively clean yet over-smoothed results. The balanced configuration using $\gamma = 0.5$ offers quality comparable to the recent diffusion-based methods.

Additional comparisons with controllable diffusion-based ISR methods Fig. 4 and Fig. 5 compare our method with existing controllable diffusion-based ISR approaches (StableSR [13], PASD [21], and PiSA-SR [11]). As shown by the visual transitions in Fig. 4 and the blue- and orange-colored curves in Fig. 5, both StableSR and PASD exhibit limited controllability, failing to provide a clear and sufficiently wide fidelity-realism trade-off. PiSA-SR and our method show comparable visual quality and similar trends along the fidelity-realism spectrum. Overall, these results indicate that our approach offers competitive quality to the state-of-the-art controllable method (i.e., PiSA-SR) across the entire controllable spectrum.

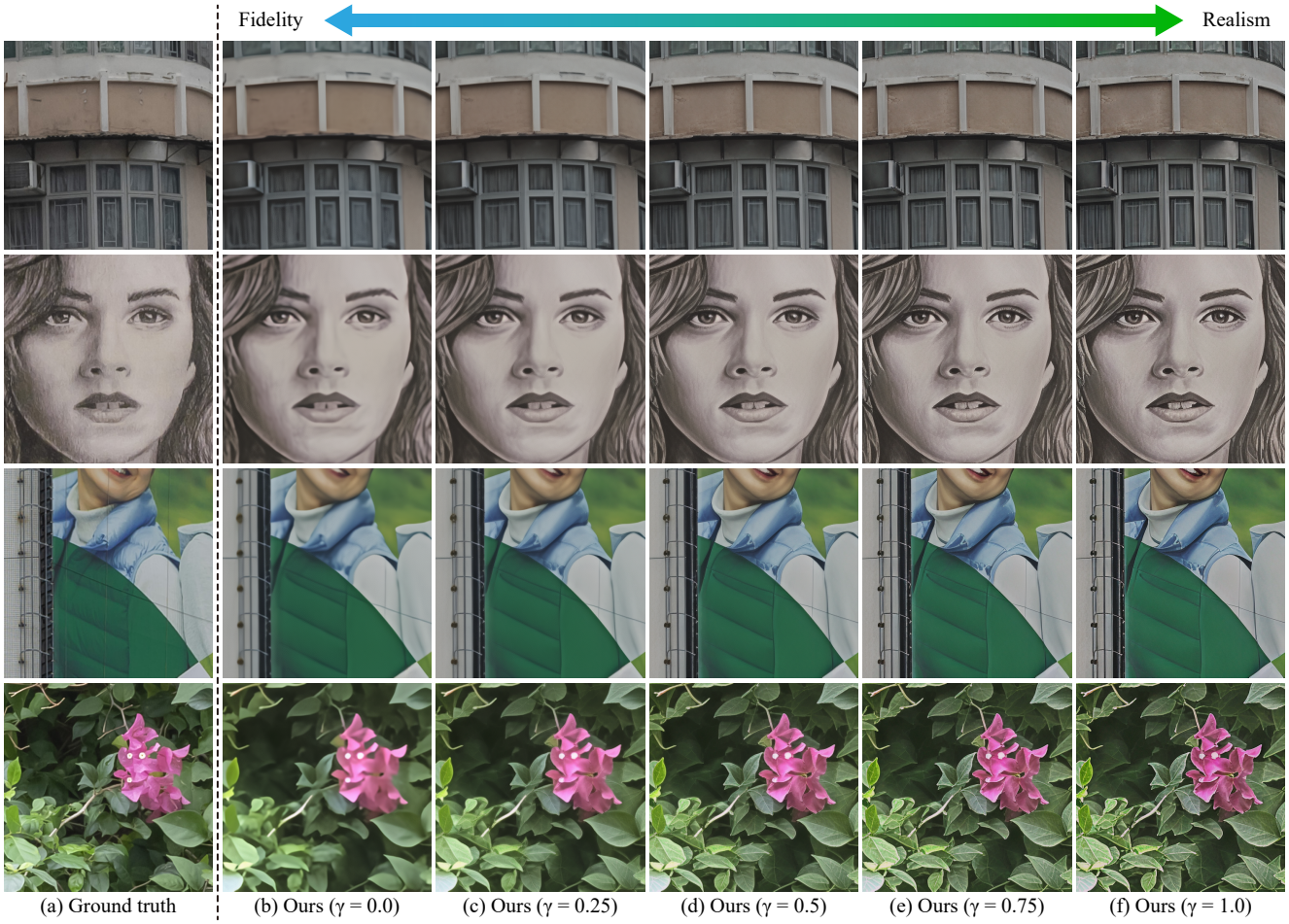

Figure 2. Visual results of our method with different values of the controllable parameter γ. From top to bottom, the tested images are NIKON_008, NIKON_046, CANON_034, and NIKON_007 from the RealSR dataset.

Datasets	Methods	PSNR ↑	SSIM ↑	LPIPS ↓	DISTS ↓	FID ↓	NIQE ↓	MUSIQ ↑	MANIQA ↑	CLIPIQA ↑
DIV2K-val	StableSR-s200	23.27	0.5730	0.3105	0.2043	**24.57**	4.7748	65.81	0.6189	0.6776
	DiffBIR-s50	23.17	0.5451	0.3654	0.2204	32.38	5.0131	69.90	**0.6454**	0.7292
	SeeSR-s50	23.68	0.6043	0.3194	0.1968	25.89	4.8134	68.66	0.6222	0.6934
	PASD-s20	24.42	0.6256	0.3793	0.2218	31.51	5.4259	61.23	0.5397	0.5554
	ResShift-s15	24.71	0.6230	0.3404	0.2244	41.92	6.4573	60.84	0.5338	0.5944
	SinSR-s1	24.29	0.6010	0.3222	0.2045	37.17	5.7917	63.32	0.5413	0.6541
	OSEDiff-s1	23.72	0.6109	0.2942	0.1975	26.33	4.7103	67.96	0.6147	0.6680
	AdcSR-s1	23.74	0.6017	0.2853	0.1899	25.52	**4.3577**	68.00	0.6073	0.6763
	TSD-SR-s1	23.02	0.5808	**0.2673**	**0.1821**	**23.28**	**4.3231**	**71.69**	0.6192	**0.7414**
	InvSR-s1	23.29	0.5921	0.3167	0.2044	30.18	4.6890	68.67	0.6296	0.7125
	TVT-s1	24.23	**0.6292**	**0.2773**	**0.1860**	24.78	5.5997	68.67	0.6060	0.6987
	CTMSR-s1	**24.88**	0.6264	0.3027	0.2040	34.11	5.1193	65.61	0.5165	0.6597
	PiSA-SR-s1	23.87	0.6058	0.2823	0.1934	25.09	4.5560	69.68	0.6401	0.6928
	Ours-s1 ($\gamma = 0.0$)	**25.77**	**0.6712**	0.4340	0.2711	49.05	6.8827	38.44	0.3294	0.3280
	Ours-s1 ($\gamma = 0.5$)	22.90	0.5732	0.2966	0.1961	27.35	4.6584	70.09	0.6165	0.6937
	Ours-s1 ($\gamma = 1.0$)	19.99	0.4894	0.3339	0.2175	29.89	4.7138	**70.53**	**0.6404**	**0.7321**

Table 4. Quantitative comparison with recent diffusion-based ISR methods on the DIV2K-val [1] dataset. We measure reference-based metrics (PSNR, SSIM [16], LPIPS [28], and DISTS [5]), no-reference metrics (NIQE [27], MUSIQ [8], MANIQA [20], and CLIPIQA [12]), and FID [7]. 's' stands for the number of diffusion steps. We highlight the best and the second-best results with **blue** and **green**.

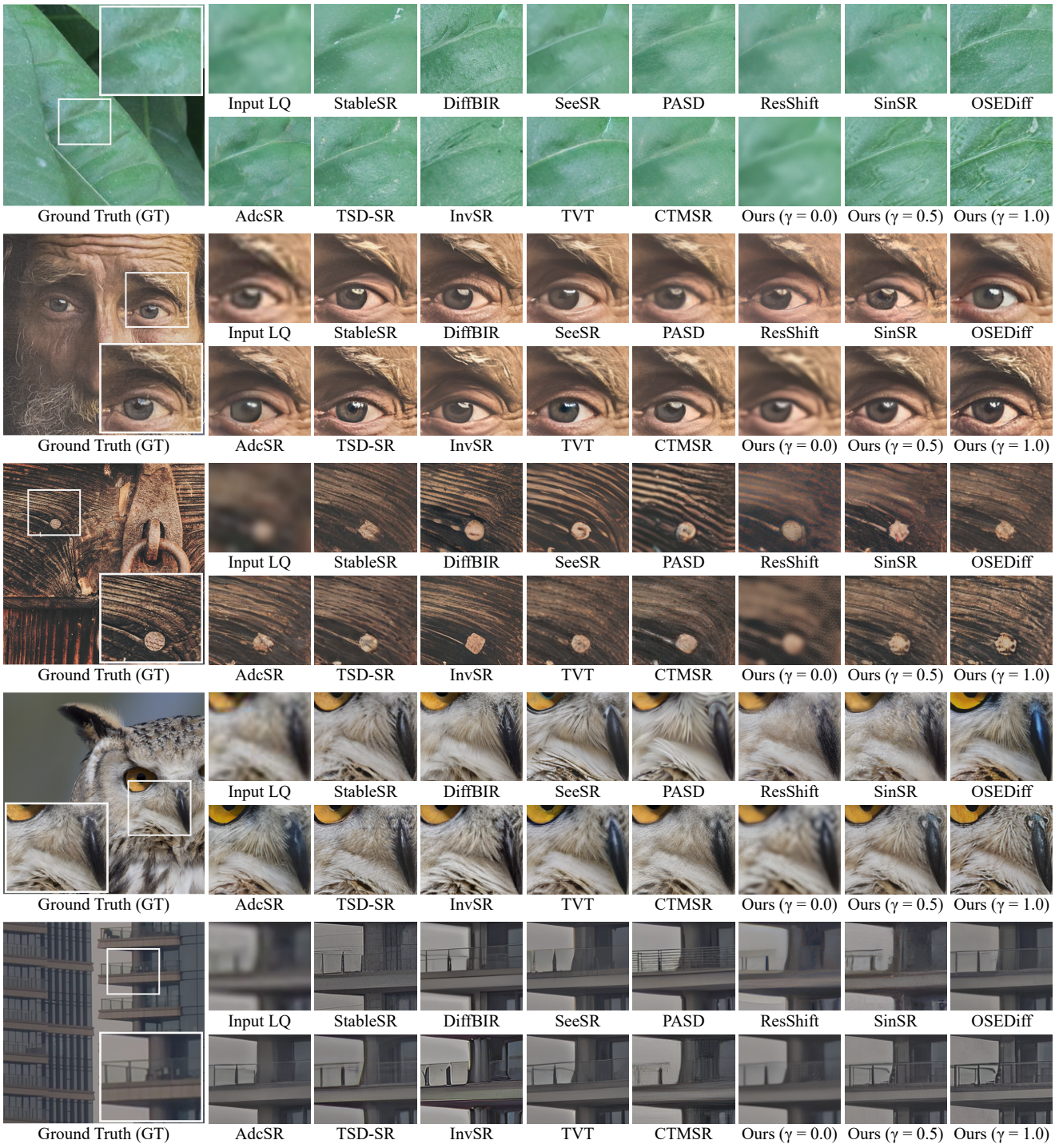

Figure 3. Visual comparison of our approach with recent diffusion-based ISR methods. From top to bottom, the tested images are IMG_190 (DRealSR [17]), NIKON_047 (RealSR [2]), 0885_PCH_00019 (DIV2K-val [1]), 0880_PCH_00007 (DIV2K-val), and P1160646 (DRealSR).

Figure 4. Visual comparison among controllable ISR on P1171031 and DSC_1454 images from the DRealSR dataset. We chose their parameters for fidelity ((a) and (d)) and realism ((c) and (f)) based on the metrics evaluated on the DRealSR, specifically, the parameters achieving the highest PSNR for fidelity and the highest MUSIQ score for realism. The central settings ((b) and (e)) correspond to the central parameter values, i.e., StableSR ($\omega \in \{0.7, 0.4, 0.2\}$), PASD ($n \in \{900, 700, 500\}$), and PiSA-SR ($\lambda_{pix} = 1.0, \lambda_{sem} \in \{0.2, 0.8, 1.5\}$).

Comparisons with non-diffusion-based methods Table 5 and Fig. 6 show quantitative and visual comparisons with four non-diffusion-based ISR methods: BSRGAN [26], Real-ESRGAN [14], FeMaSR [4], and SwinIR [9]. While these methods often produce blurry and less realistic outputs, our method with $\gamma = 1.0$ generates richer details and achieves high no-reference metrics (e.g., the best scores for MUSIQ, MANIQA [20], and CLIPIQA across all benchmarks). The fidelity-oriented setting (i.e., $\gamma = 0.0$) achieves the highest PSNR and SSIM across all benchmarks but exhibits relatively inferior feature-space reference-based metrics (i.e., LPIPS and DISTS [5]) compared to existing methods. The balanced setting, i.e., $\gamma = 0.5$, outperforms no-reference metrics while preserving reference-based metrics (e.g., PSNR and SSIM) comparable to previous methods.

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 (CVPR) Workshops*, 2017. 2, 3, 4

[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 (ICCV)*, 2019. 1, 4

[3] Bin Chen, Gehui Li, Rongyuan Wu, Xindong Zhang, Jie Chen, Jian Zhang, and Lei Zhang. Adversarial diffusion compression for real-world image super-resolution. In *Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR)*, pages 28208–28220, 2025. 2

[4] Chaofeng Chen, Xinyu Shi, Yipeng Qin, Xiaoming Li, Xiaoguang Han, Tao Yang, and Shihui Guo. Real-world blind super-resolution via feature matching with implicit high-resolution priors. In *Proceedings of the 30th ACM Interna-*

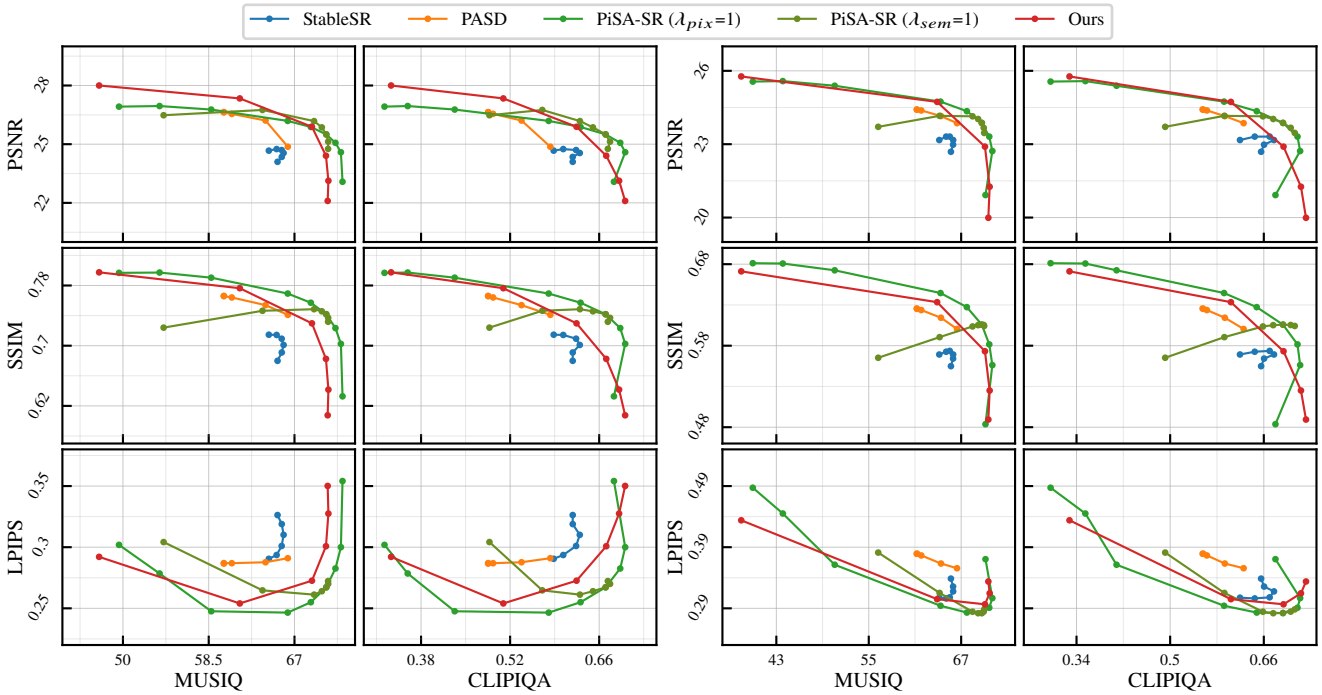

Figure 5. Comparison of controllable ISR methods on the RealSR and DIV2K-val datasets, illustrating the fidelity-realism spectrum. The y-axis represents PSNR, SSIM, and LPIPS, and the x-axis represents MUSIQ and CLIPIQA. StableSR and PASD are evaluated using $\omega \in [0.0, 1.0]$ and $n \in [100, 900]$, respectively. For PiSA-SR, two curves are presented: one varying $\lambda_{sem} \in [0.0, 1.5]$ with $\lambda_{pix} = 1$, and the other varying $\lambda_{pix} \in [0.0, 1.2]$ with $\lambda_{sem} = 1$.

Datasets	Methods	PSNR ↑	SSIM ↑	LPIPS ↓	DISTS ↓	FID ↓	NIQE ↓	MUSIQ ↑	MANIQA ↑	CLIPIQA ↑
RealSR	BSRGAN	**26.38**	0.7651	**0.2656**	0.2124	141.31	5.6360	63.29	0.5417	0.5119
	Real-ESRGAN	25.69	0.7614	0.2709	**0.2060**	135.20	5.7932	60.37	0.5492	0.4485
	FeMaSR	25.07	0.7356	0.2937	0.2286	141.08	5.7661	59.06	0.4862	0.5409
	SwinIR	26.31	**0.7729**	**0.2539**	**0.1946**	129.45	5.6802	58.69	0.5228	0.4366
	Ours ($\gamma = 0.0$)	**28.00**	**0.7975**	0.2921	0.2187	138.88	7.9015	47.64	0.4025	0.3337
	Ours ($\gamma = 0.5$)	25.14	0.7055	0.2866	0.2141	**110.60**	**5.4313**	**69.68**	**0.6293**	**0.6522**
	Ours ($\gamma = 1.0$)	22.10	0.6075	0.3502	0.2367	**115.31**	**4.9010**	**70.29**	**0.6643**	**0.7011**
DRealSR	BSRGAN	**28.70**	0.8028	0.2858	0.2144	155.60	6.5388	57.16	0.4855	0.5093
	Real-ESRGAN	28.61	**0.8051**	**0.2818**	**0.2089**	147.69	6.6899	54.27	0.4900	0.4519
	FeMaSR	26.87	0.7570	0.3157	0.2239	157.82	**5.9085**	53.71	0.4418	0.5638
	SwinIR	28.50	0.8044	**0.2743**	**0.2040**	148.25	6.5591	52.74	0.4729	0.4445
	Ours ($\gamma = 0.0$)	**30.86**	**0.8472**	0.3179	0.2340	156.39	8.5957	42.32	0.3380	0.3471
	Ours ($\gamma = 0.5$)	27.60	0.7371	0.3278	0.2361	**133.84**	6.0570	**66.57**	**0.5940**	**0.7149**
	Ours ($\gamma = 1.0$)	24.63	0.6377	0.3957	0.2587	**144.00**	**5.7044**	**67.14**	**0.6229**	**0.7446**
DIV2K-val	BSRGAN	**24.58**	0.6269	0.3351	0.2275	44.23	4.7510	61.20	0.5070	0.5247
	Real-ESRGAN	24.29	**0.6372**	**0.3112**	0.2141	37.64	**4.6800**	61.06	0.5501	0.5277
	FeMaSR	23.06	0.5887	0.3126	**0.2057**	35.87	4.7415	60.83	0.5074	0.5998
	SwinIR	23.93	0.6286	0.3160	0.2143	36.34	4.7062	60.21	0.5431	0.5337
	Ours ($\gamma = 0.0$)	**25.77**	**0.6712**	0.4340	0.2711	49.05	6.8827	38.44	0.3294	0.3280
	Ours ($\gamma = 0.5$)	22.90	0.5732	**0.2966**	**0.1961**	**27.35**	**4.6584**	**70.09**	**0.6165**	**0.6937**
	Ours ($\gamma = 1.0$)	19.99	0.4894	0.3339	0.2175	**29.89**	4.7138	**70.53**	**0.6404**	**0.7321**

Table 5. Quantitative comparison with non-diffusion-based ISR approaches on the RealSR, DRealSR, and DIV2K-val datasets. We highlight the best and the second-best results with **blue** and **green**, respectively.

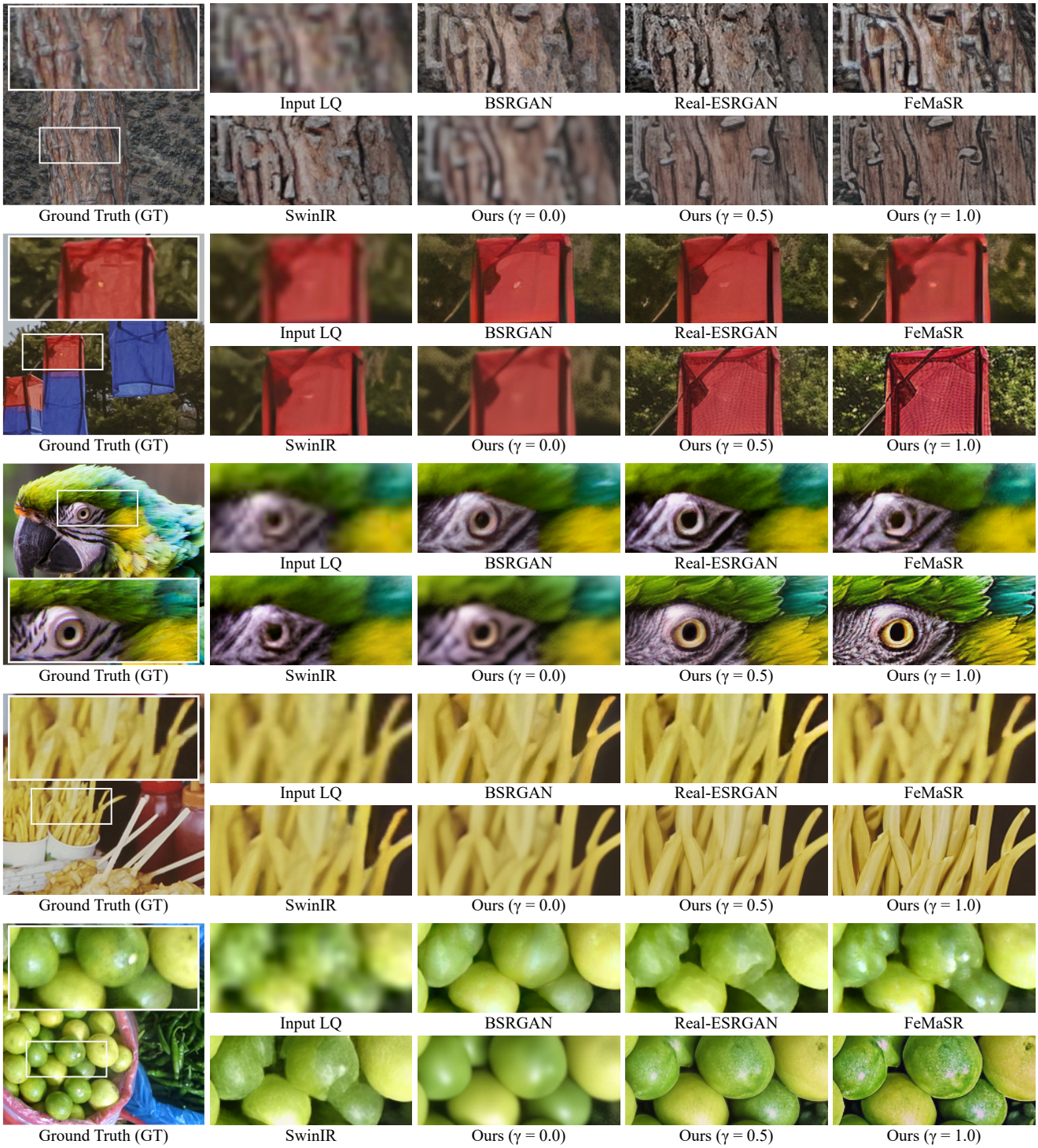

Figure 6. Visual comparison with non-diffusion-based ISR methods. From top to bottom, the tested images are SONY_129 (DRealSR), CANON_011 (RealSR), 0886_PCH_00010 (DIV2K-val), CANON_025 (RealSR), and 0889_PCH_00034 (DIV2K-val).

tional Conference on Multimedia, pages 1329–1338, 2022. 5

[5] Keyan Ding, Kede Ma, Shiqi Wang, and Eero P. Simoncelli. Image quality assessment: Unifying structure and texture similarity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(5):2567–2581, 2022. 3, 5

[6] Linwei Dong, Qingnan Fan, Yihong Guo, Zhonghao Wang, Qi Zhang, Jinwei Chen, Yawei Luo, and Changqing Zou. Tsd-sr: One-step diffusion with target score distillation for real-world image super-resolution. In *Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR)*, pages 23174–23184, 2025. 2

[7] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Advances in Neural Information Processing Systems*, 2017. 3

[8] 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 (ICCV)*, pages 5148–5157, 2021. 1, 3

[9] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, pages 1833–1844, 2021. 5

[10] Xinqi Lin, Jingwen He, Ziyan Chen, Zhaoyang Lyu, Bo Dai, Fanghua Yu, Yu Qiao, Wanli Ouyang, and Chao Dong. Diffbir: Toward blind image restoration with generative diffusion prior. In *European Conference on Computer Vision (ECCV)*, pages 430–448. Springer, 2024. 2

[11] Lingchen Sun, Rongyuan Wu, Zhiyuan Ma, Shuaizheng Liu, Qiaosi Yi, and Lei Zhang. Pixel-level and semantic-level adjustable super-resolution: A dual-lora approach. In *Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR)*, pages 2333–2343, 2025. 2

[12] Jianyi Wang, Kelvin CK Chan, and Chen Change Loy. Exploring clip for assessing the look and feel of images. In *AAAI*, 2023. 1, 3

[13] Jianyi Wang, Zongsheng Yue, Shangchen Zhou, Kelvin CK Chan, and Chen Change Loy. Exploiting diffusion prior for real-world image super-resolution. *International Journal of Computer Vision*, 132(12):5929–5949, 2024. 2

[14] 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 (ICCV) Workshops*, pages 1905–1914, 2021. 5

[15] Yufei Wang, Wenhan Yang, Xinyuan Chen, Yaohui Wang, Lanqing Guo, Lap-Pui Chau, Ziwei Liu, Yu Qiao, Alex C. Kot, and Bihan Wen. Sinsr: Diffusion-based image super-resolution in a single step. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 25796–25805, 2024. 2

[16] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4): 600–612, 2004. 1, 3

[17] 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 *European Conference on Computer Vision (ECCV)*, pages 101–117. Springer, 2020. 4

[18] Rongyuan Wu, Lingchen Sun, Zhiyuan Ma, and Lei Zhang. One-step effective diffusion network for real-world image super-resolution. In *Advances in Neural Information Processing Systems*, pages 92529–92553, 2024. 2

[19] Rongyuan Wu, Tao Yang, Lingchen Sun, Zhengqiang Zhang, Shuai Li, and Lei Zhang. Seesr: Towards semantics-aware real-world image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 25456–25467, 2024. 2

[20] 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 (CVPR) Workshops*, pages 1191–1200, 2022. 3, 5

[21] Tao Yang, Rongyuan Wu, Peiran Ren, Xuansong Xie, and Lei Zhang. Pixel-aware stable diffusion for realistic image super-resolution and personalized stylization. In *European Conference on Computer Vision (ECCV)*, pages 74–91. Springer, 2024. 2

[22] Qiaosi Yi, Shuai Li, Rongyuan Wu, Lingchen Sun, Yuhui Wu, and Lei Zhang. Fine-structure preserved real-world image super-resolution via transfer vae training. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 12415–12426, 2025. 2

[23] Weiyi You, Mingyang Zhang, Leheng Zhang, Xingyu Zhou, Kexuan Shi, and Shuhang Gu. Consistency trajectory matching for one-step generative super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 12747–12756, 2025. 2

[24] Zongsheng Yue, Jianyi Wang, and Chen Change Loy. Resshift: Efficient diffusion model for image super-resolution by residual shifting. In *Advances in Neural Information Processing Systems*, pages 13294–13307, 2023. 2

[25] Zongsheng Yue, Kang Liao, and Chen Change Loy. Arbitrary-steps image super-resolution via diffusion inversion. In *Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR)*, pages 23153–23163, 2025. 2

[26] 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 (ICCV)*, pages 4791–4800, 2021. 5

[27] Lin Zhang, Lei Zhang, and Alan C. Bovik. A feature-enriched completely blind image quality evaluator. *IEEE Transactions on Image Processing*, 24(8):2579–2591, 2015. 3

[28] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 1, 3