Training Generative Image Super-Resolution Models by Wavelet-Domain Losses Enables Better Control of Artifacts

Supplementary Material

1. Further Visual Comparison of SISR

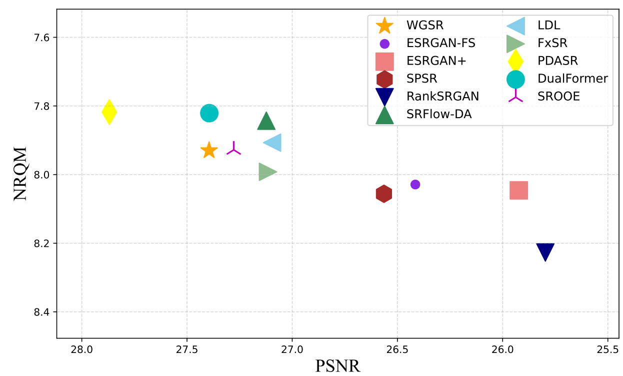
We provide further visual comparisons of 4× SISR between our proposed method WGSR and the other state-of-the-art methods including ESRGAN-FS [3], ESRGAN+ [13], RankSRGAN [15], SPSR [10], SRFlow-DA [6], LDL [7], FxSR [11], PDASR [16], SROOE [12] and DualFormer [8] in Fig. 2 to Fig. 7. From these visual comparisons, one can draw consistent observations in line with the results in the paper. Our proposed method WGSR not only suppresses visual artifacts but also simultaneously restores structural shapes and realistic details.

2. Perception-Distortion Trade-off

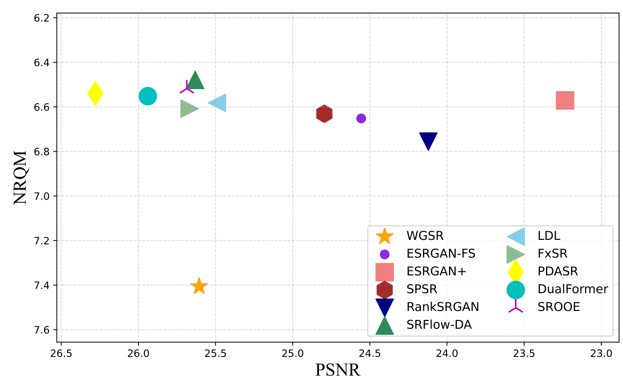
Fig. 1 provides perception-distortion points for our WGSR and other state-of-the-art methods on PSNR-NRQM plane for Set14 [14], Urban100 [4] and DIV2K [1] validation datasets. Our proposed method, WGSR, significantly improves fidelity and perceptual scores across all datasets, resulting in a better PD trade-off point. Specifically, WGSR achieves the highest NRQM [9] scores on Urban100 and DIV2K datasets when compared to other methods with comparable PSNR scores. This also validates the generalization performance of the proposed WGSR to different benchmarks.

3. Further Visual Comparison of Different Wavelet Filters

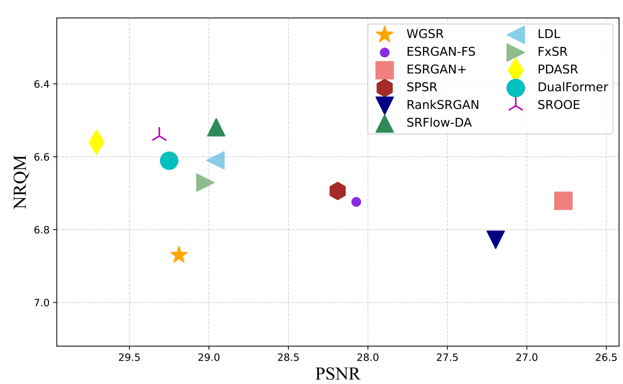
The visual comparison of proposed WGSR method with different wavelet families [5] for 4× SISR is shown in Fig. 8. We observe that visual performance varies according to the choice of wavelet family and the results show that the best perception-distortion trade-off point is achieved by the Symlet "sym7" filter. However, all wavelet filters notably mitigate artifacts while providing photo-realistic SR results.



(a) Set14 [14]

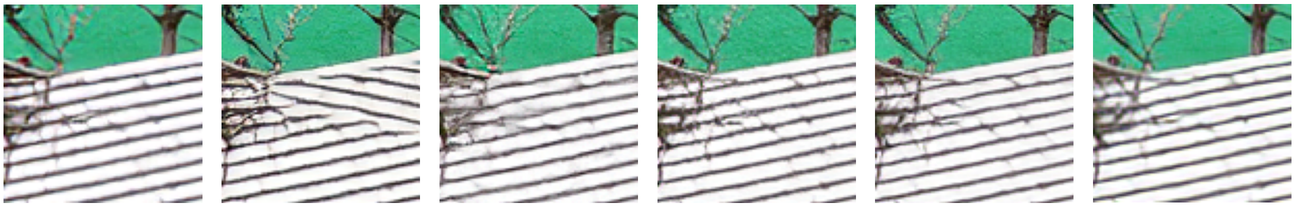


(b) Urban100 [4]

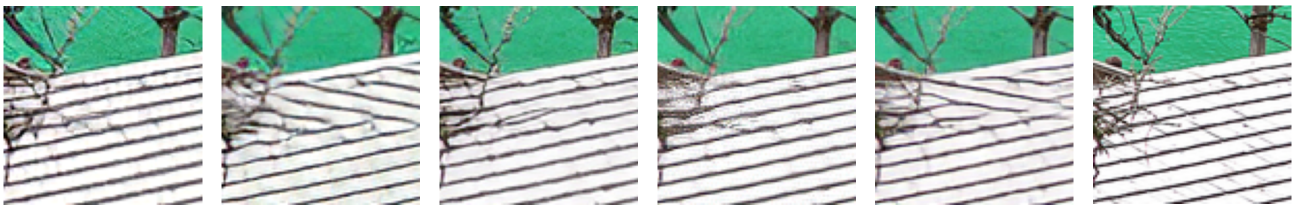


(c) DIV2K [1]

Figure 1. Perception-distortion trade-off performance of our model WGSR compared to other state-of-the-art methods on the PSNR-NRQM plane.

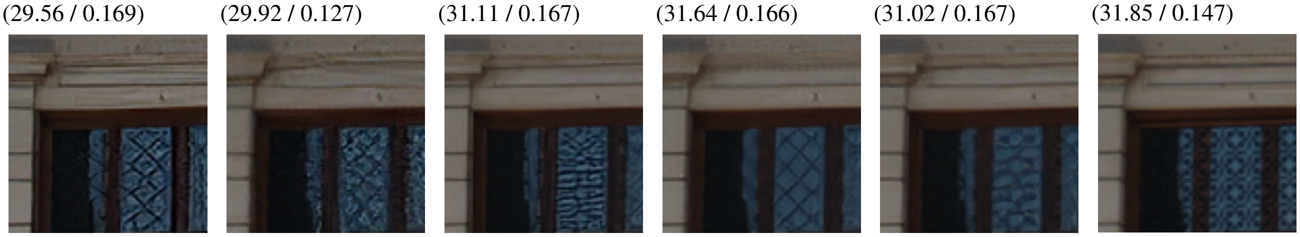


ESRGAN-FS [3]	SPSR [10]	SRFlow-DA [6]	FxSR [11]	SROOE [12]	WGSR (Ours)
(18.20 / 0.216)	(14.64 / 0.179)	(18.24 / 0.193)	(18.98 / 0.188)	(19.26 / 0.156)	(19.64 / 0.176)

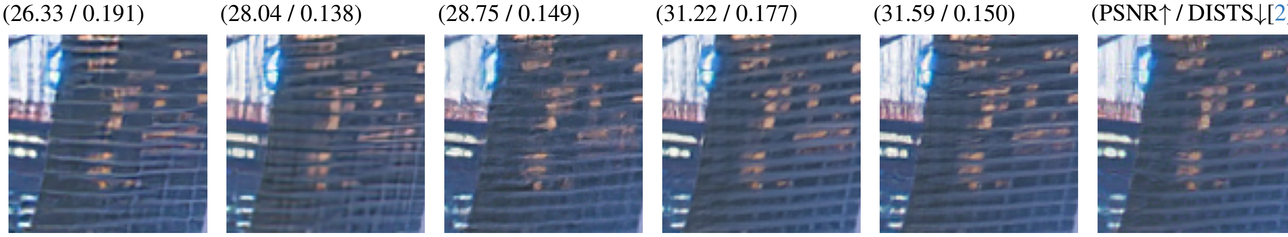


ESRGAN+ [13]	RankSRGAN [15]	LDL [7]	PDASR [16]	DualFormer [8]	HR (img-807)
(17.83 / 0.187)	(14.99 / 0.230)	(18.02 / 0.167)	(19.40 / 0.200)	(16.58 / 0.172)	(PSNR↑ / DISTS↓[2])

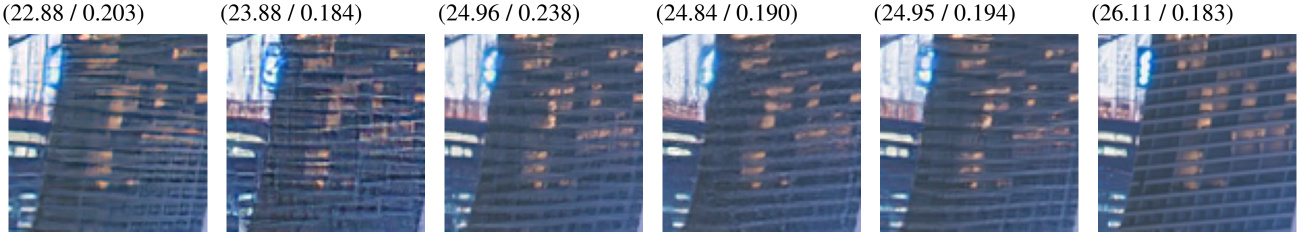
ESRGAN-FS [3]	SPSR [10]	SRFlow-DA [6]	FxSR [11]	SROOE [12]	WGSR (Ours)
(29.56 / 0.169)	(29.92 / 0.127)	(31.11 / 0.167)	(31.64 / 0.166)	(31.02 / 0.167)	(31.85 / 0.147)



ESRGAN+ [13]	RankSRGAN [15]	LDL [7]	PDASR [16]	DualFormer [8]	HR (img-812)
(26.33 / 0.191)	(28.04 / 0.138)	(28.75 / 0.149)	(31.22 / 0.177)	(31.59 / 0.150)	(PSNR↑ / DISTS↓[2])

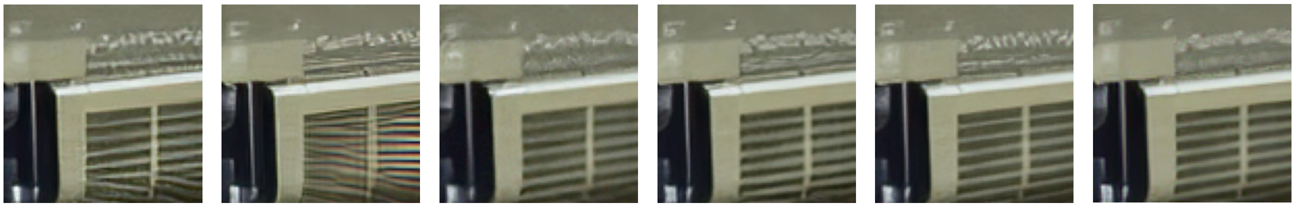


ESRGAN-FS [3]	SPSR [10]	SRFlow-DA [6]	FxSR [11]	SROOE [12]	WGSR (Ours)
(22.88 / 0.203)	(23.88 / 0.184)	(24.96 / 0.238)	(24.84 / 0.190)	(24.95 / 0.194)	(26.11 / 0.183)

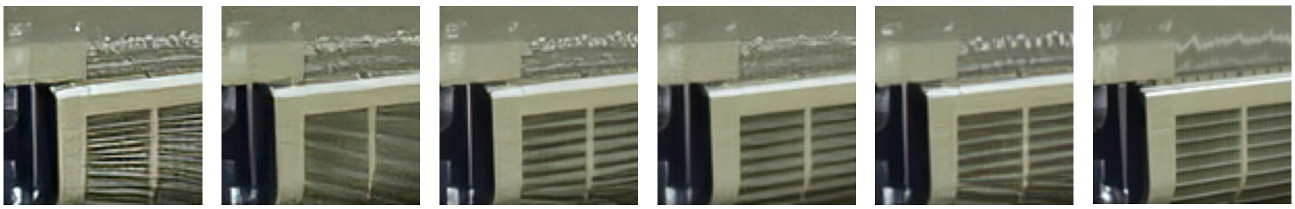


ESRGAN+ [13]	RankSRGAN [15]	LDL [7]	PDASR [16]	DualFormer [8]	HR (img-814)
(22.49 / 0.218)	(21.09 / 0.214)	(24.41 / 0.179)	(25.72 / 0.222)	(24.73 / 0.206)	(PSNR↑ / DISTS↓[2])

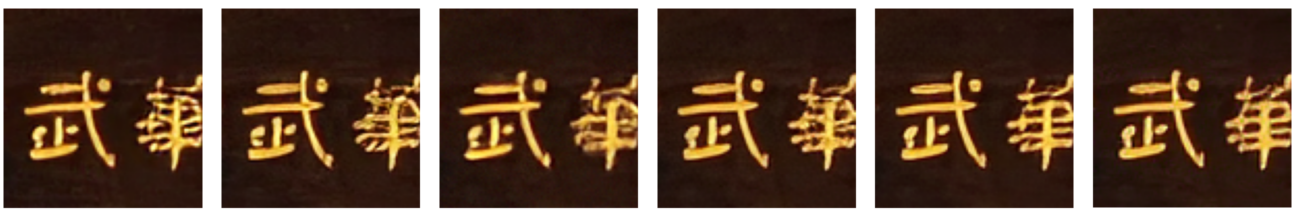
Figure 2. Visual comparison of the proposed wavelet-guided perceptual optimization method with the state-of-the-art for $\times 4$ SR on natural images from DIV2K validation set [1]. The proposed WGSR has clear advantages in reconstructing realistic high-frequency details while inhibiting artifacts.



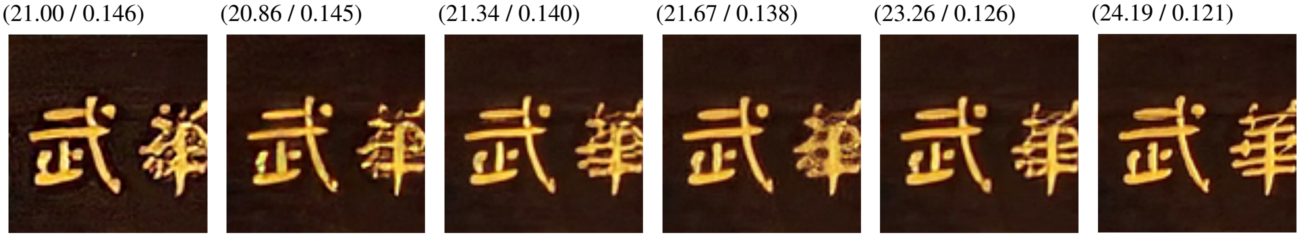
ESRGAN-FS [3]	SPSR [10]	SRFlow-DA [6]	FxSR [11]	SROOE [12]	WGSR (Ours)
(24.76 / 0.130)	(22.28 / 0.175)	(27.88 / 0.126)	(26.28 / 0.140)	(27.39 / 0.120)	(27.99 / 0.115)



ESRGAN+ [13]	RankSRGAN [15]	LDL [7]	PDASR [16]	DualFormer [8]	HR (img-820)
(20.79 / 0.179)	(24.66 / 0.175)	(24.91 / 0.133)	(27.77 / 0.120)	(26.08 / 0.118)	(PSNR↑ / DISTS↓[2])



ESRGAN-FS [3]	SPSR [10]	SRFlow-DA [6]	FxSR [11]	SROOE [12]	WGSR (Ours)
(21.00 / 0.146)	(20.86 / 0.145)	(21.34 / 0.140)	(21.67 / 0.138)	(23.26 / 0.126)	(24.19 / 0.121)



ESRGAN+ [13]	RankSRGAN [15]	LDL [7]	PDASR [16]	DualFormer [8]	HR (img-826)
(19.53 / 0.168)	(20.29 / 0.148)	(21.66 / 0.136)	(22.66 / 0.154)	(23.24 / 0.117)	(PSNR↑ / DISTS↓[2])

ESRGAN-FS [3]	SPSR [10]	SRFlow-DA [6]	FxSR [11]	SROOE [12]	WGSR (Ours)
(26.38 / 0.129)	(27.25 / 0.123)	(28.32 / 0.136)	(28.70 / 0.134)	(29.13 / 0.127)	(29.70 / 0.120)

ESRGAN+ [13]	RankSRGAN [15]	LDL [7]	PDASR [16]	DualFormer [8]	HR (img-832)
(24.51 / 0.173)	(25.07 / 0.170)	(27.81 / 0.124)	(29.99 / 0.136)	(28.77 / 0.122)	(PSNR↑ / DISTS↓[2])

Figure 3. Visual comparison of the proposed wavelet-guided perceptual optimization method with the state-of-the-art for $\times 4$ SR on natural images from DIV2K validation set [1].

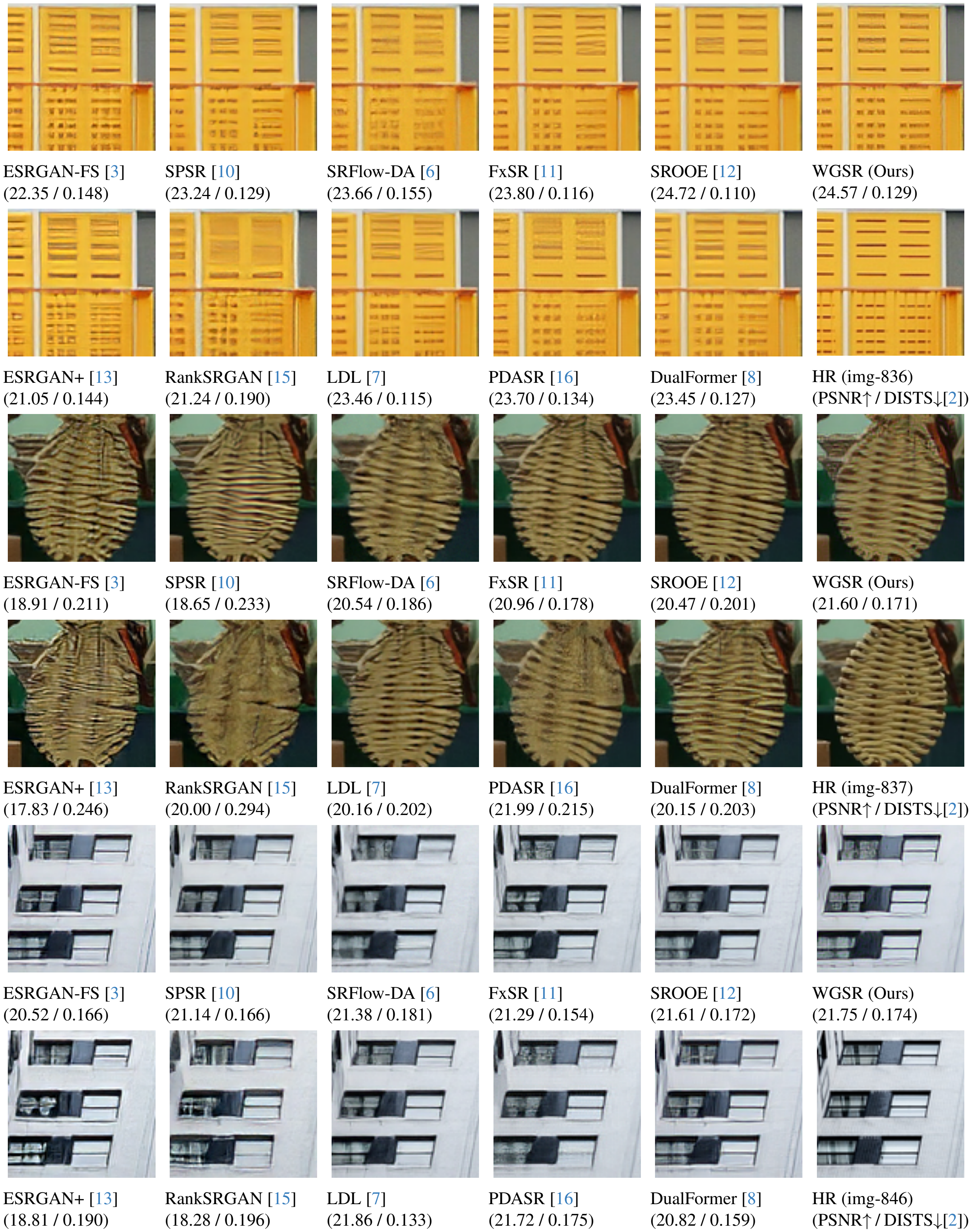
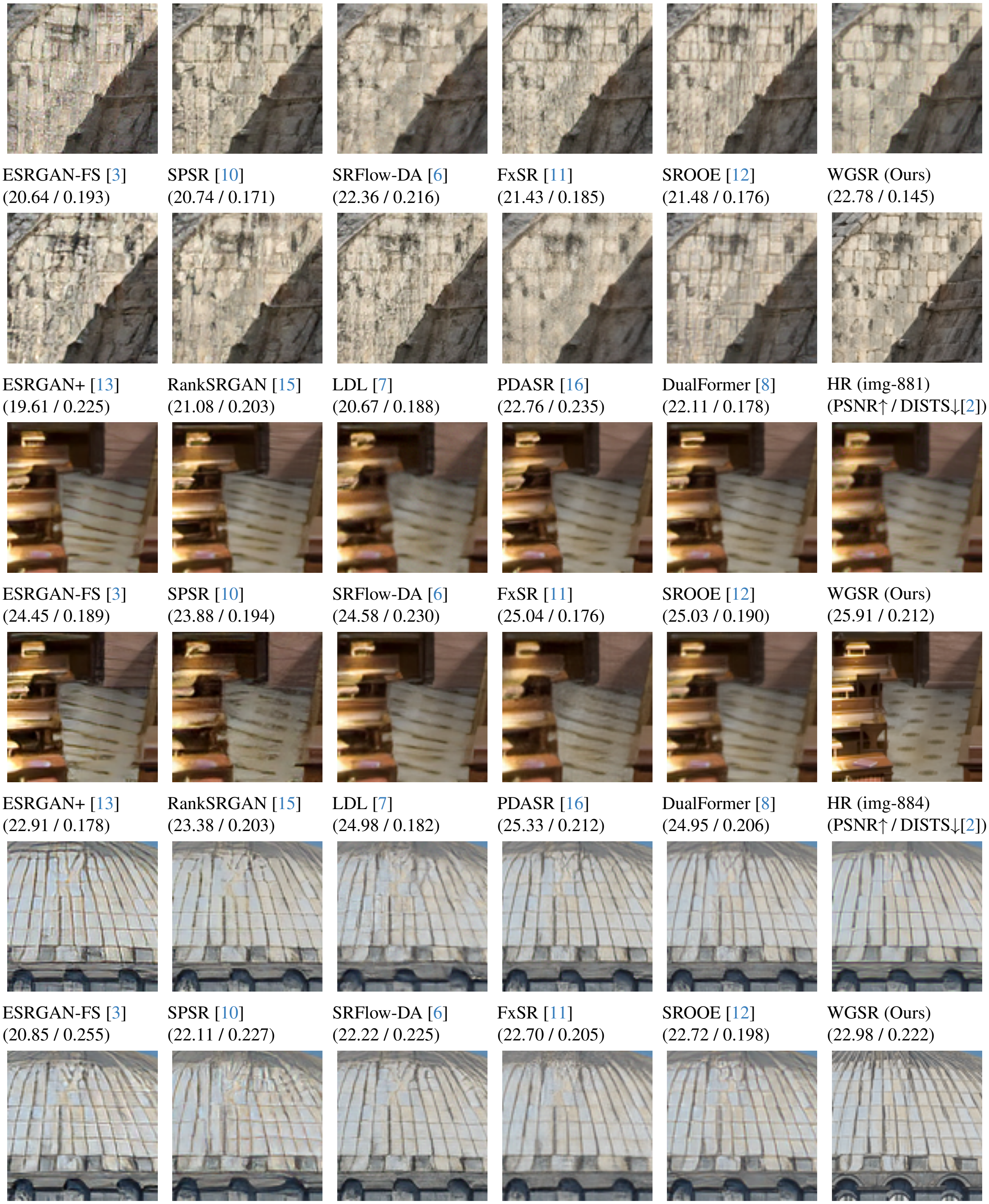


Figure 4. Visual comparison of the proposed wavelet-guided perceptual optimization method with the state-of-the-art for $\times 4$ SR on natural images from DIV2K validation set [1].



ESRGAN-FS [3] (20.64 / 0.193) | SPSR [10] (20.74 / 0.171) | SRFlow-DA [6] (22.36 / 0.216) | FxSR [11] (21.43 / 0.185) | SROOE [12] (21.48 / 0.176) | WGSR (Ours) (22.78 / 0.145)

ESRGAN+ [13] (19.61 / 0.225) | RankSRGAN [15] (21.08 / 0.203) | LDL [7] (20.67 / 0.188) | PDASR [16] (22.76 / 0.235) | DualFormer [8] (22.11 / 0.178) | HR (img-881) (PSNR↑ / DISTS↓[2])

ESRGAN-FS [3] (24.45 / 0.189) | SPSR [10] (23.88 / 0.194) | SRFlow-DA [6] (24.58 / 0.230) | FxSR [11] (25.04 / 0.176) | SROOE [12] (25.03 / 0.190) | WGSR (Ours) (25.91 / 0.212)

ESRGAN+ [13] (22.91 / 0.178) | RankSRGAN [15] (23.38 / 0.203) | LDL [7] (24.98 / 0.182) | PDASR [16] (25.33 / 0.212) | DualFormer [8] (24.95 / 0.206) | HR (img-884) (PSNR↑ / DISTS↓[2])

ESRGAN-FS [3] (20.85 / 0.255) | SPSR [10] (22.11 / 0.227) | SRFlow-DA [6] (22.22 / 0.225) | FxSR [11] (22.70 / 0.205) | SROOE [12] (22.72 / 0.198) | WGSR (Ours) (22.98 / 0.222)

ESRGAN+ [13] (21.07 / 0.212) | RankSRGAN [15] (21.58 / 0.234) | LDL [7] (22.58 / 0.229) | PDASR [16] (23.23 / 0.234) | DualFormer [8] (22.70 / 0.210) | HR (img-890) (PSNR↑ / DISTS↓[2])

Figure 5. Visual comparison of the proposed wavelet-guided perceptual optimization method with the state-of-the-art for $\times 4$ SR on natural images from DIV2K validation set [1].

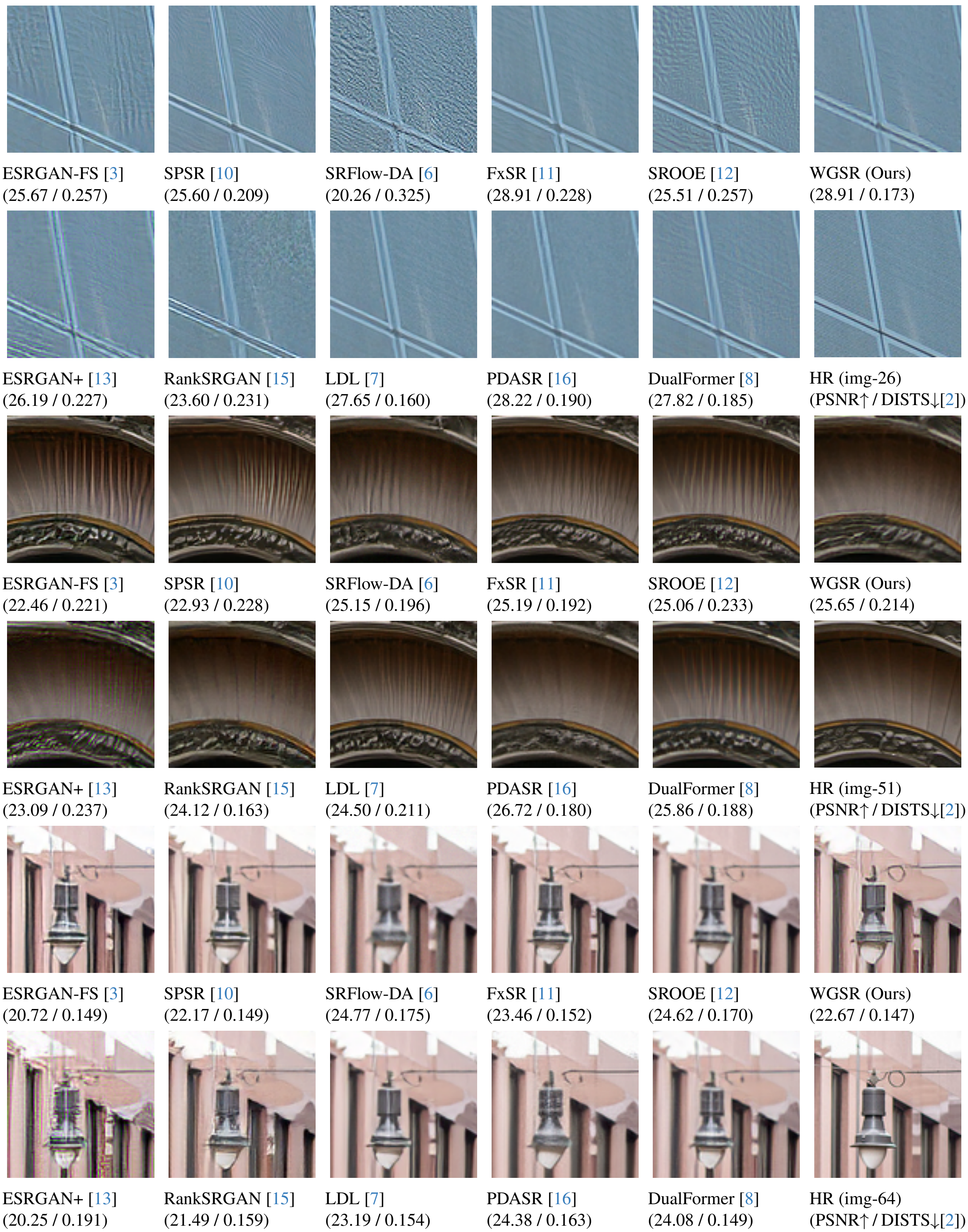
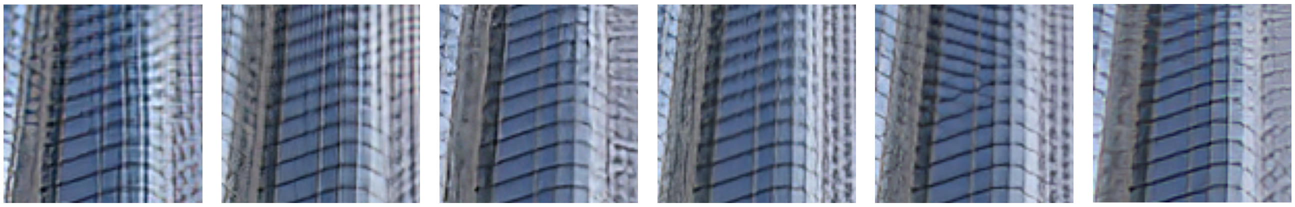
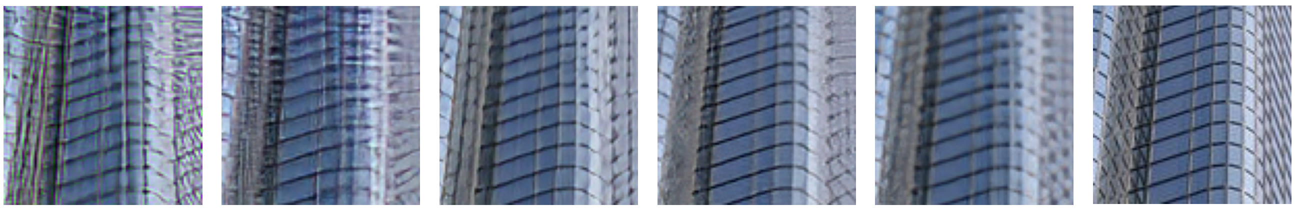


Figure 6. Visual comparison of the proposed wavelet-guided perceptual optimization method with the state-of-the-art for $\times 4$ SR on natural images from Urban100 validation set [4].

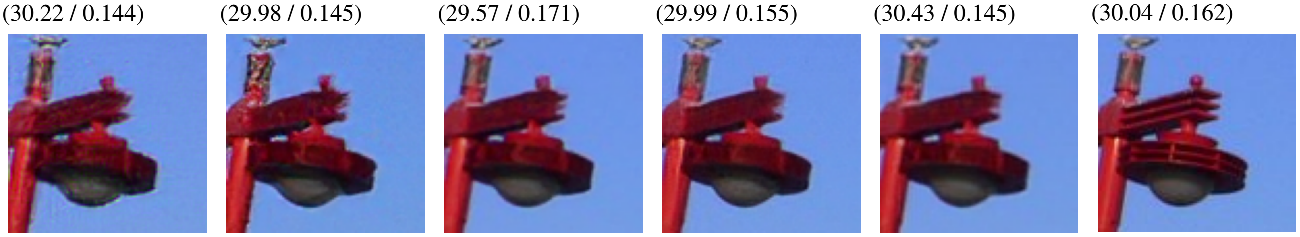


ESRGAN-FS [3] (18.36 / 0.283)	SPSR [10] (19.31 / 0.278)	SRFlow-DA [6] (19.92 / 0.258)	FxSR [11] (19.44 / 0.251)	SROOE [12] (19.94 / 0.256)	WGSR (Ours) (20.98 / 0.252)

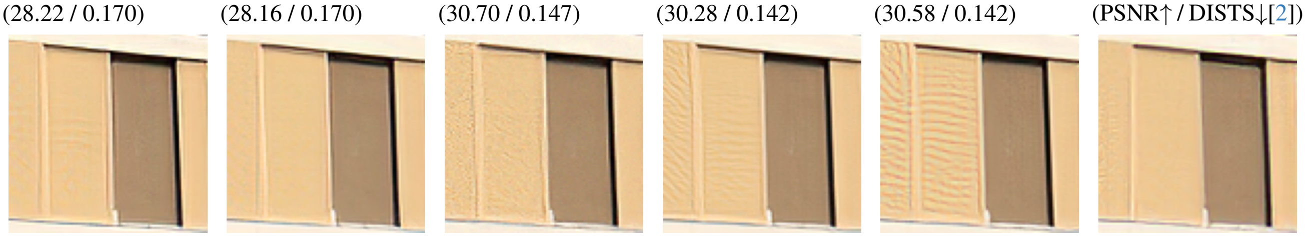


ESRGAN+ [13] (17.85 / 0.258)	RankSRGAN [15] (18.81 / 0.294)	LDL [7] (20.06 / 0.245)	PDASR [16] (21.08 / 0.219)	DualFormer [8] (20.66 / 0.251)	HR (img-74) (PSNR↑ / DISTS↓[2])

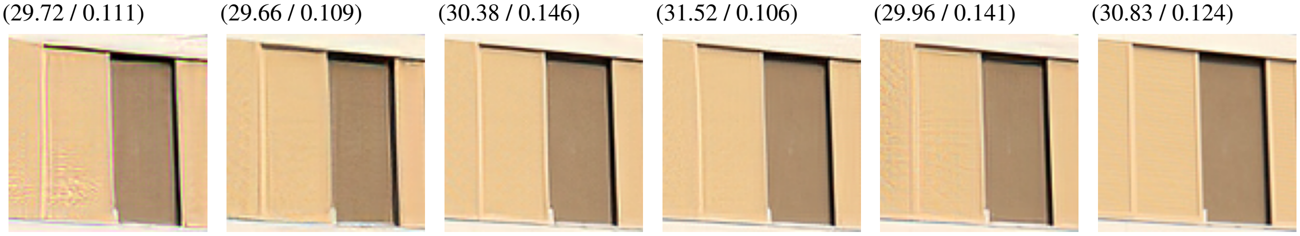
ESRGAN-FS [3] (30.22 / 0.144)	SPSR [10] (29.98 / 0.145)	SRFlow-DA [6] (29.57 / 0.171)	FxSR [11] (29.99 / 0.155)	SROOE [12] (30.43 / 0.145)	WGSR (Ours) (30.04 / 0.162)



ESRGAN+ [13] (28.22 / 0.170)	RankSRGAN [15] (28.16 / 0.170)	LDL [7] (30.70 / 0.147)	PDASR [16] (30.28 / 0.142)	DualFormer [8] (30.58 / 0.142)	HR (img-86) (PSNR↑ / DISTS↓[2])



ESRGAN-FS [3] (29.72 / 0.111)	SPSR [10] (29.66 / 0.109)	SRFlow-DA [6] (30.38 / 0.146)	FxSR [11] (31.52 / 0.106)	SROOE [12] (29.96 / 0.141)	WGSR (Ours) (30.83 / 0.124)



ESRGAN+ [13] (26.59 / 0.161)	RankSRGAN [15] (27.75 / 0.127)	LDL [7] (30.61 / 0.109)	PDASR [16] (31.10 / 0.094)	DualFormer [8] (30.26 / 0.122)	HR (img-87) (PSNR↑ / DISTS↓[2])

Figure 7. Visual comparison of the proposed wavelet-guided perceptual optimization method with the state-of-the-art for $\times 4$ SR on natural images from Urban100 validation set [4].

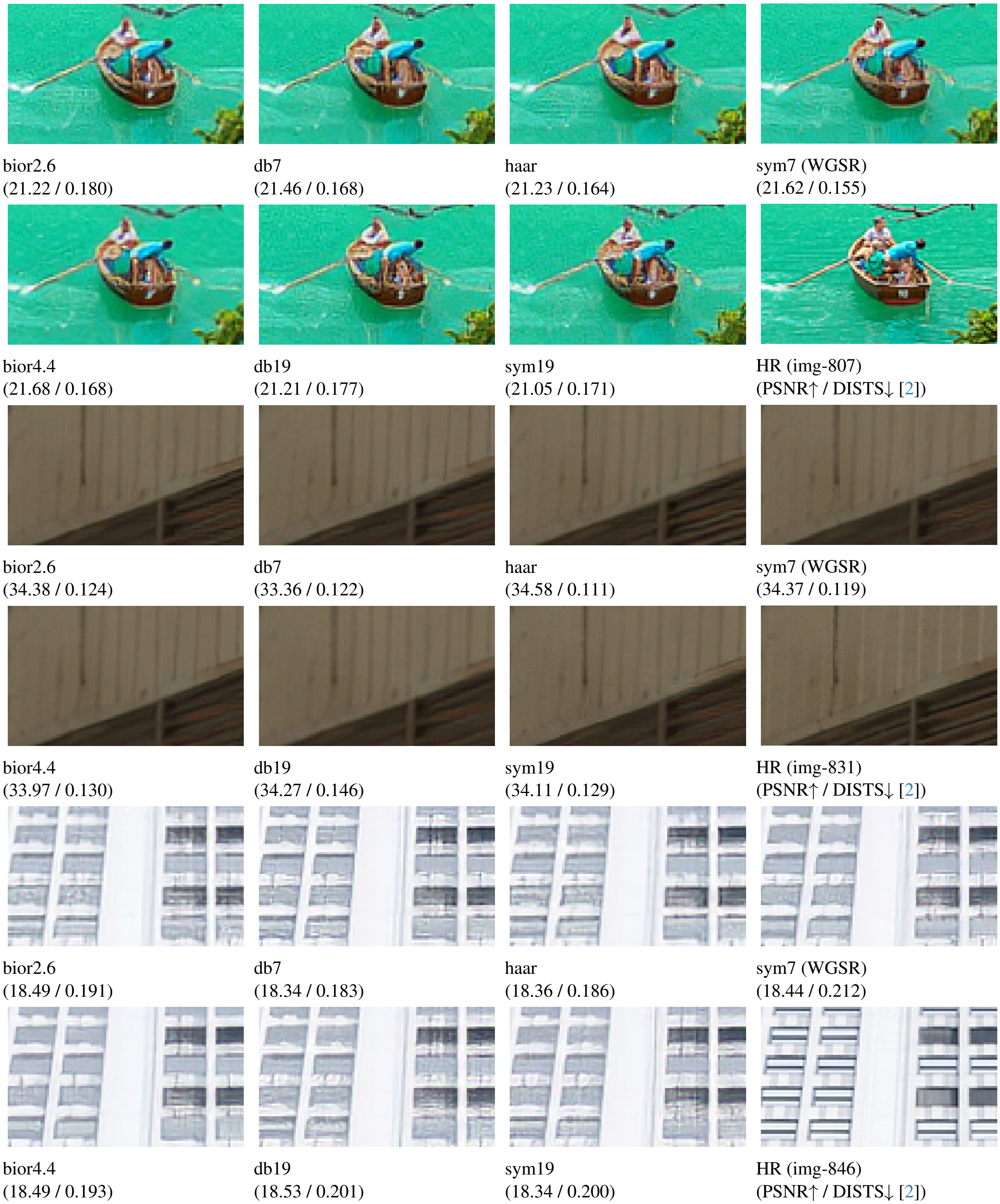


Figure 8. Visual comparison of WGSR method with different wavelet families for 4× SR on DIV2K [1].

References

[1] E Agustsson and R. Timofte. NTIRE 2017 Challenge on single image super-resolution: Dataset and study. In *IEEE/CVF Conf. on Comp. Vision and Patt. Recog. (CVPR) Workshops*, 2017. 1, 2, 3, 4, 5, 8

[2] K. Ding, K. Ma, S. Wang, and E. P. Simoncelli. Image quality assessment: Unifying structure and texture similarity. *IEEE Trans. on Patt Anal. and Mach. Intel.*, 44:2567–2581, 2020. 2, 3, 4, 5, 6, 7, 8

[3] Manuel Fritsche, Shuhang Gu, and Radu Timofte. Frequency separation for real-world super-resolution. In *IEEE/CVF Int. Conf. on Computer Vision Workshop (ICCVW)*, pages 3599–3608, 2019. 1, 2, 3, 4, 5, 6, 7

[4] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *IEEE Conf. on Comp. Vision and Patt. Recog. (CVPR)*, pages 5197–5206, 2015. 1, 6, 7

[5] Bjorn Jawerth and Wim Sweldens. An overview of wavelet based multiresolution analyses. *SIAM Review*, 36(3):377–412, 1994. 1

[6] Younghyun Jo, Sejong Yang, and Seon Joo Kim. Srflow-da: Super-resolution using normalizing flow with deep convolutional block. In *IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2021. 1, 2, 3, 4, 5, 6, 7

[7] Jie Liang, Hui Zeng, and Lei Zhang. Details or artifacts: A locally discriminative learning approach to realistic image super-resolution. In *IEEE/CVF Conf. on Comp. Vision and Patt. Recog. (CVPR)*, pages 5657–5666, 2022. 1, 2, 3, 4, 5, 6, 7

[8] Xin Luo, Yunan Zhu, Shunxin Xu, and Dong Liu. On the effectiveness of spectral discriminators for perceptual quality improvement. In *ICCV*, 2023. 1, 2, 3, 4, 5, 6, 7

[9] Chao Ma, Chih-Yuan Yang, Xiaokang Yang, and Ming-Hsuan Yang. Learning a no-reference quality metric for single-image super-resolution. *Comput. Vis. Image Underst.*, 158:1–16, 2017. 1

[10] Cheng Ma, Yongming Rao, Yean Cheng, Ce Chen, Jiwen Lu, and Jie Zhou. Structure-preserving super resolution with gradient guidance. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2020. 1, 2, 3, 4, 5, 6, 7

[11] Seung Ho Park, Young Su Moon, and Nam Ik Cho. Flexible style image super-resolution using conditional objective. *IEEE Access*, 10:9774–9792, 2022. 1, 2, 3, 4, 5, 6, 7

[12] Seung Ho Park, Young Su Moon, and Nam Ik Cho. Perception-oriented single image super-resolution using optimal objective estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1725–1735, 2023. 1, 2, 3, 4, 5, 6, 7

[13] N. C. Rakotonirina and A. Rasoanaivo. Esrgan+: Further improving enhanced super-resolution generative adversarial network. In *IEEE Int. Conf. on Acoust., Speech and Signal Processing (ICASSP)*, pages 3637–3641, 2020. 1, 2, 3, 4, 5, 6, 7

[14] Roman Zeyde, Michael Elad, and Matan Protter. On single image scale-up using sparse-representations. In *Curves and Surfaces*, pages 711–730, Berlin, Heidelberg, 2012. Springer. 1

[15] Wenlong Zhang, Yihao Liu, Chao Dong, and Yu Qiao. Ranksrgan: Super resolution generative adversarial networks with learning to rank. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 44(10):7149–7166, 2021. 1, 2, 3, 4, 5, 6, 7

[16] Yuehan Zhang, Bo Ji, Jia Hao, and Angela Yao. Perception-distortion balanced admm optimization for single-image super-resolution. In *European Conf. on Comp. Vision (ECCV)*, 2022. 1, 2, 3, 4, 5, 6, 7