

Supplementary Materials: Perceptual Oriented Image Restoration is a Multi-Objective Optimization Problem

Qiwen Zhu

State Key Lab of MIPT, Huazhong
University of Science and Technology
Wuhan, China
zhuqiwen@hust.edu.cn

Yanjie Wang

School of AIA, Huazhong University
of Science and Technology
Wuhan, China
aiawyj@hust.edu.cn

Shilv Cai

School of AIA, Huazhong University
of Science and Technology
Wuhan, China
caishilv@hust.edu.cn

Liqun Chen

School of AIA, Huazhong University
of Science and Technology
Wuhan, China
chenliqun@hust.edu.cn

Jiahuan Zhou

Wangxuan Institute of Computer
Technology, Peking University
Beijing, China
jiahuanzhou@pku.edu.cn

Luxin Yan

State Key Lab of MIPT, Huazhong
University of Science and Technology
Wuhan, China
yanluxin@hust.edu.cn

Sheng Zhong

State Key Lab of MIPT, Huazhong
University of Science and Technology
Wuhan, China
zhongsheng@hust.edu.cn

Xu Zou*

State Key Lab of MIPT, Huazhong
University of Science and Technology
Wuhan, China
zoux@hust.edu.cn

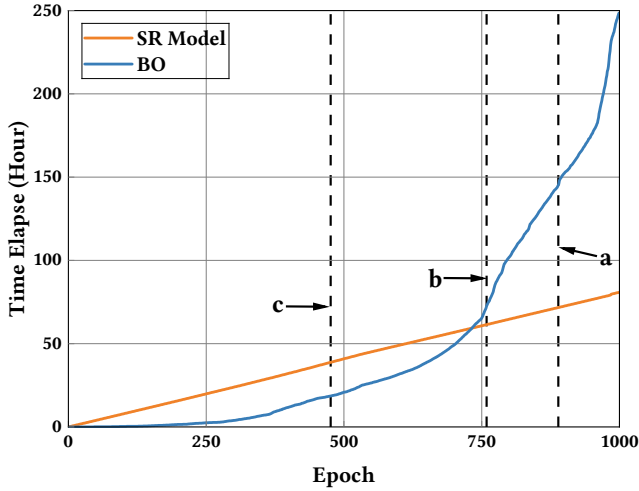


Figure 1: Comparative of the time consumed for training the SR model against the time required for BO during the training process of our MOBOSR. The temporal measurements were conducted on an server with NVIDIA RTX 3090 GPU and two Intel Xeon Gold 6226R CPUs. Epochs for Ours-[a,b,c] are labeled for clarity.

1 Time Analysis

Bayesian Optimization (BO) minimizes the number of evaluations by substituting the actual objective function with a surrogate function and heuristically determining the most promising points for improvement through an acquisition function for subsequent evaluation rounds. Consequently, this approach significantly conserves

*Corresponding author.

Table 1: Detailed comparison of time expended for sampling points Ours-[a,b,c] and at the optimization cessation.

Point	Epoch	SR Time Elapse	MOBO Time Elapse
Ours-c	476	38.8h	18.6h
Ours-b	759	61.4h	72.6h
Ours-a	889	71.7h	145.2h
End	1000	80.8h	248.4h

optimization iteration compared to evolutionary algorithms. However, the BO process necessitates the computation of the covariance matrix, resulting in a time complexity of $O(n^3)$. If the objective function be overly intricate or the problem dimensions too high, necessitating numerous optimization rounds, this would substantially increase the optimization time.

In the main text, the three sampling points selected from the perceptual-distortion Pareto front, labeled as Ours-[a,b,c], were obtained at epochs 476, 759 and 889, respectively. As shown in Table 1 and Figure 1, the time consumed to train the Super-Resolution (SR) model with Ours-c, obtained at epoch 476, was approximately half of that required for BO. At epoch 759, the training time for the SR model with Ours-b was nearly identical to that for BO. However, by epoch 889, the time expended on BO for Ours-a was roughly half of that for training the SR model. By the designated optimization halt at epoch 1000, the time consumption for BO had become threefold that of the SR model.

But, after optimization up to epoch 759 (Ours-c), there were no significant changes in the Pareto front (as shown in Figure 2). We believe that the substantial advantages brought by our Multi-Objective Bayesian Optimization Super-Resolution (MOBOSR), and without introducing any additional computational load during inference, justify the mere doubling of training duration.

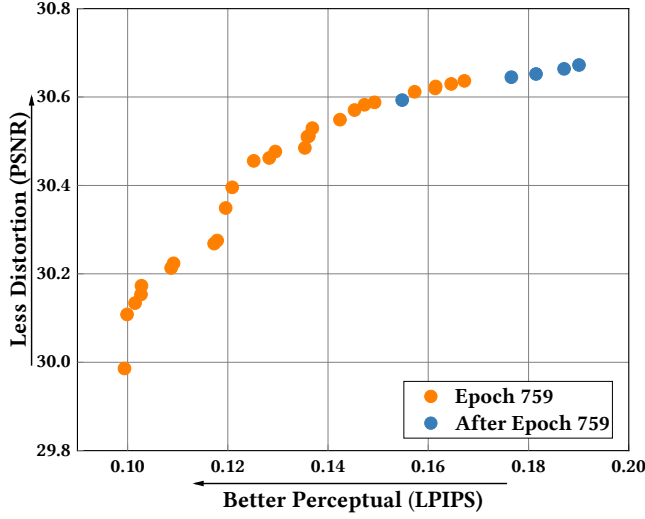


Figure 2: Changes in the perception-distortion Pareto frontier obtained by our MOBOSR after epoch 759.

Table 2: Analyses of multi-task learning on the DIV2K [1] validation set. The best and second-best results are highlighted in bold and underline, respectively. MTL methods show significant improvements over the manually weighted ESRGAN [12] but are not as effective as our MOBOSR.

Methods	PSNR↑	SSIM↑	LR-PSNR↑	LPIPS↓
ESRGAN [12]	27.6994	0.7610	41.2244	0.1193
CAGrad [5]	<u>28.4434</u>	<u>0.7803</u>	47.5291	<u>0.1161</u>
NashMTL [7]	28.1606	0.7717	43.7037	0.1168
MOBOSR (Ours)	28.5089	0.7834	<u>44.6847</u>	0.1145

2 Discussion on Multi-Task Learning

Focusing on a single task may overlook the information from related tasks that could improve the target task. By sharing parameters between different tasks (with different loss functions) to a certain extent may achieve better generalization for the original task. Hence, we discuss the impact of Multi-Task Learning (MTL) in balancing the distortion and perceptual quality for SISR models. We initially consider using MTL to optimize multiple loss functions to address the balance issue. We compare the results of our MOBOSR with 2 MTL methods (CAGrad [5] and NashMTL [7]), as well as with ESRGAN [12]. As shown in Table 2, MTL methods show significant improvements over the manually weighted ESRGAN [12] but are not as effective as our MOBOSR. We believe this is because MTL approaches just seek a single compromise between multi-task/multi-loss rather than searching for the entire Pareto frontier as MOBO does. But, the ability to achieve these outcomes indicates that MTL merits further investigation.

3 Metrics Recalculation Details

Due to the variations in datasets and metrics reported by the methods under comparison, as well as the differences in implementation

details during metric computation, we have adopted a uniform metric calculation method to re-evaluate the metrics of other methods. This ensures a fairer comparison. We generated SR results for all test sets using the model weights and inference code released by the authors, followed by calculations using our standardized metric computation program. Our codes for metric calculation are detailed in the GitHub repository: <https://github.com/ZhuKeven/MOBOSR>. Table 3 presents the metrics we recomputed alongside those reported by the authors, with most showing no significant differences and some even surpassing the reported results. Wang et al. only reported metrics for the RRDB-PSNR [12] model trained with L1 loss, without providing the metrics for the ESRGAN [12] model trained using GAN [2]. Consequently, we are unable to present the ESRGAN [12] metrics comparison in Table 3 in the same manner as for other methods.

4 More Quantitative Results

Due to the length constraints of the main text, we have included the complete results for Ours-[a,b,c] here, as well as the metrics for the RRDB-PSNR [12] model trained using L1 loss. Although RRDB-PSNR [12] exhibits superior performance in terms of PSNR and SSIM [13] metrics, the margin by which it surpasses Ours-a is considerably less than the extent to which Ours-a exceeds RRDB-PSNR [12] in perceptual (LPIPS) and consistency (LR-PSNR) metrics. Not to mention that the RRDB-PSNR [12] was trained on a significantly larger dataset, the DF2K-OST (13774 images), which comprises the DIV2K [1] training set (800 images), Flickr2K [10] (2650 images), and OST [11] (10,324 images), whereas our MOBOSR was trained solely on the DIV2K [1] training set (800 images).

5 More Visual Results

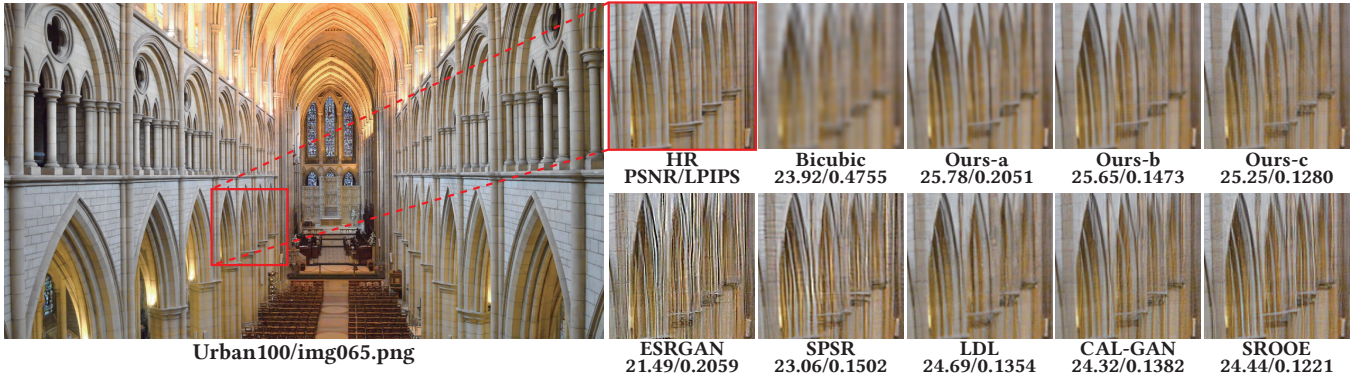
We have included more visual comparison results here, including the visualizations on the Urban100 [3] dataset as shown in Figure 3 and Figure 4, and the visualizations on the DIV2K [1] validation set as shown in Figure 5.

References

- [1] Eirikur Agustsson and Radu Timofte. 2017. NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 1122–1131.
- [2] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2020. Generative adversarial networks. *Commun. ACM* 63, 11 (oct 2020), 139–144.
- [3] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. 2015. Single Image Super-Resolution From Transformed Self-Exemplars. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 5197–5206.
- [4] Jie Liang, Hui Zeng, and Lei Zhang. 2022. 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 (CVPR)*. 5657–5666.
- [5] Bo Liu, Xingchao Liu, Xiaojie Jin, Peter Stone, and Qiang Liu. 2021. Conflict-Averse Gradient Descent for Multi-task Learning. In *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 34. 18878–18890.
- [6] Cheng Ma, Yongming Rao, Yean Cheng, Ce Chen, Jiwen Lu, and Jie Zhou. 2020. Structure-Preserving Super Resolution With Gradient Guidance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 7766–7775.
- [7] Aviv Navon, Aviv Shamsian, Idan Achituve, Haggai Maron, Kenji Kawaguchi, Gal Chechik, and Ethan Fetaya. 2022. Multi-Task Learning as a Bargaining Game. In *Proceedings of the International Conference on Machine Learning (ICML)*, Vol. 162. 16428–16446.
- [8] Joonkyu Park, Sanghyun Son, and Kyoung Mu Lee. 2023. Content-Aware Local GAN for Photo-Realistic Super-Resolution. In *Proceedings of the IEEE/CVF*

Table 3: Comparison of metrics calculated using our uniform method versus those reported by the authors, showing minimal difference, with some even surpassing the reported results. The higher results are highlighted in bold. The symbols \uparrow and \downarrow indicate that higher or lower values of the metric are preferable.

Metric	Dataset	RRDB-PSNR [12]		SPSR [6]		RRDB+LDL [4]		CAL-GAN [8]		SROOE [9]	
		Author	Recalculated	Author	Recalculated	Author	Recalculated	Author	Recalculated	Author	Recalculated
PSNR \uparrow	Set5	32.73	32.7010	30.400	30.3871	30.985	31.0007	31.177	31.0475	-	31.2455
	Set14	28.99	28.9831	26.640	26.6501	27.491	27.2064	-	27.3272	-	27.2561
	DIV2K	-	30.8888	-	28.1824	28.951	28.9510	28.863	28.9549	27.69	29.0990
	BSD100	27.85	27.8235	25.505	25.4949	-	26.0988	25.925	26.2581	24.87	26.1715
	Urban100	27.03	26.9859	24.799	24.8063	25.498	25.4781	25.290	25.2908	24.33	25.8452
	General100	-	31.9145	29.414	29.4794	30.232	30.1974	30.182	30.0742	28.74	30.4723
	Manga109	31.66	31.5637	-	28.6102	29.407	29.4111	-	29.1665	28.08	29.9017
SSIM \uparrow	Set5	0.9011	0.9010	0.8627	0.8432	0.8626	0.8610	0.863	0.8552	-	0.8651
	Set14	0.7917	0.7915	0.7930	0.7133	0.7476	0.7343	-	0.7353	-	0.7304
	DIV2K	-	0.8485	-	0.7720	0.7951	0.7952	0.790	0.7897	0.7932	0.7980
	BSD100	0.7455	0.7453	0.6576	0.6571	-	0.6811	0.676	0.6789	0.6869	0.6866
	Urban100	0.8153	0.8152	0.9481	0.7472	0.7673	0.7670	0.763	0.7623	0.7707	0.7764
	General100	-	0.8725	0.8537	0.8095	0.8277	0.8278	0.825	0.8262	0.8297	0.8332
	Manga109	0.9196	0.9195	-	0.8591	0.8746	0.8746	-	0.8676	0.8554	0.8786
LR-PSNR \uparrow	Set5	-	53.0951	-	46.3607	-	48.5067	-	42.4327	-	53.1781
	Set14	-	50.8098	-	43.6201	-	46.2893	-	41.5963	-	51.0679
	DIV2K	-	51.9030	-	44.8529	-	47.9757	-	42.8611	50.80	53.5488
	BSD100	-	50.3975	-	42.6756	-	45.1571	-	41.0666	49.19	51.2347
	Urban100	-	51.0009	-	42.6679	-	46.5827	-	41.6069	48.32	50.6700
	General100	-	52.6741	-	44.6786	-	48.0079	-	43.4227	50.11	52.9797
	Manga109	-	52.3321	-	44.3872	-	47.8923	-	42.8636	48.77	51.7820
LPIPS \downarrow	Set5	-	0.1691	0.0644	0.0616	0.0670	0.0637	0.061	0.0687	-	0.0603
	Set14	-	0.2718	0.1318	0.1313	0.1207	0.1309	-	0.1320	-	0.1131
	DIV2K	-	0.2526	-	0.1097	0.1011	0.1007	0.091	0.1072	0.0957	0.0956
	BSD100	-	0.3590	0.1611	0.1629	-	0.1635	0.151	0.1696	0.1500	0.1514
	Urban100	-	0.1956	0.1184	0.1186	0.1096	0.1097	0.108	0.1171	0.1065	0.1067
	General100	-	0.1668	0.0863	0.0866	0.0790	0.0794	0.077	0.0894	0.0753	0.0758
	Manga109	-	0.0977	-	0.0662	0.0553	0.0546	-	0.0688	0.0524	0.0511

**Figure 3: More visual comparisons on Urban100 [3].**

- International Conference on Computer Vision (ICCV). 10585–10594.
- [9] Seung Ho Park, Young Su Moon, and Nam Ik Cho. 2023. Perception-Oriented Single Image Super-Resolution using Optimal Objective Estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 1725–1735.
- [10] Radu Timofte, Eirikur Agustsson, Luc Van Gool, Ming-Hsuan Yang, Lei Zhang, Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, et al. 2017. NTIRE 2017 Challenge on Single Image Super-Resolution: Methods and Results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 1110–1121.
- [11] Xintao Wang, Ke Yu, Chao Dong, and Chen Change Loy. 2018. Recovering Realistic Texture in Image Super-Resolution by Deep Spatial Feature Transform. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 606–615.
- [12] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. 2019. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*. 63–79.

Table 4: Comparison of Ours-[a,b,c] with other artworks on 7 datasets. The best, second-best and third-best results are highlighted in bold, underline and *italic*, respectively. The symbols \uparrow and \downarrow indicate that higher or lower values of the metric are preferable.

Metric	Method	Train Datasets	Set5	Set14	DIV2K	BSD100	Urban100	General100	Manga109
PSNR \uparrow	RRDB-PSNR [12]	DF2K-OST	32.7010	28.9831	30.8888	27.8235	26.9859	31.9145	31.5637
	ESRGAN [12]	DF2K-OST	30.4618	26.2839	28.1778	25.2892	24.3617	29.4593	28.5041
	SPSR [6]	DIV2K	30.3871	26.6501	28.1824	25.4949	24.8063	29.4794	28.6102
	RRDB+LDL [4]	DIV2K	31.0007	27.2064	28.9510	26.0988	25.4781	30.1974	29.4111
	CAL-GAN [8]	DIV2K	31.0475	27.3272	28.9549	26.2581	25.2908	30.0742	29.1665
	SROOE [9]	DF2K	31.2455	27.2561	29.0990	26.1715	25.8452	30.4723	29.9017
	Ours-a	DIV2K	<u>32.3663</u>	<u>28.7621</u>	<u>30.6384</u>	<u>27.6546</u>	<u>26.5285</u>	<u>31.6047</u>	<u>30.9787</u>
	Ours-b	DIV2K	<u>32.2126</u>	<u>28.6426</u>	<u>30.4890</u>	<u>27.5176</u>	<u>26.4439</u>	<u>31.4769</u>	<u>30.8840</u>
	Ours-c	DIV2K	31.8272	28.1766	29.9858	27.0494	26.0764	31.1164	30.2763
SSIM \uparrow	RRDB-PSNR [12]	DF2K-OST	0.9010	0.7915	0.8485	0.7453	0.8152	0.8725	0.9195
	ESRGAN [12]	DF2K-OST	0.8518	0.6982	0.7761	0.6496	0.7341	0.8102	0.8604
	SPSR [6]	DIV2K	0.8432	0.7133	0.7720	0.6571	0.7472	0.8095	0.8591
	RRDB+LDL [4]	DIV2K	0.8610	0.7343	0.7952	0.6811	0.7670	0.8278	0.8746
	CAL-GAN [8]	DIV2K	0.8552	0.7353	0.7897	0.6789	0.7623	0.8262	0.8676
	SROOE [9]	DF2K	0.8651	0.7304	0.7980	0.6866	0.7764	0.8332	0.8786
	Ours-a	DIV2K	<u>0.8961</u>	<u>0.7847</u>	<u>0.8417</u>	<u>0.7379</u>	<u>0.7989</u>	<u>0.8665</u>	<u>0.9133</u>
	Ours-b	DIV2K	<i>0.8918</i>	<i>0.7800</i>	<i>0.8376</i>	<i>0.7323</i>	<i>0.7963</i>	<i>0.8629</i>	<i>0.9093</i>
	Ours-c	DIV2K	0.8804	0.7615	0.8203	0.7109	0.7812	0.8495	0.8938
LR-PSNR \uparrow	RRDB-PSNR [12]	DF2K-OST	53.0951	50.8098	51.9030	50.3975	51.0009	52.6741	52.3321
	ESRGAN [12]	DF2K-OST	46.7348	43.8433	45.9012	43.8190	42.9339	45.4220	43.9667
	SPSR [6]	DIV2K	46.3607	43.6201	44.8529	42.6756	42.6679	44.6786	44.3872
	RRDB+LDL [4]	DIV2K	48.5067	46.2893	47.9757	45.1571	46.5827	48.0079	47.8923
	CAL-GAN [8]	DIV2K	42.4327	41.5963	42.8611	41.0666	41.6069	43.4227	42.8636
	SROOE [9]	DF2K	53.1781	51.0679	53.5488	51.2347	50.6700	52.9797	51.7820
	Ours-a	DIV2K	<u>53.7806</u>	53.5768	<u>54.7418</u>	54.2712	53.7559	<u>54.1618</u>	<u>53.8049</u>
	Ours-b	DIV2K	<u>53.4204</u>	<i>53.1927</i>	<u>53.9850</u>	<i>53.3449</i>	<u>53.3426</u>	<u>53.9097</u>	53.9045
	Ours-c	DIV2K	54.3372	<u>53.3344</u>	55.2161	<u>53.3618</u>	<i>52.9401</i>	54.5283	<i>53.4195</i>
LPIPS \downarrow	RRDB-PSNR [12]	DF2K-OST	0.1691	0.2718	0.2526	0.3590	0.1956	0.1668	0.0977
	ESRGAN [12]	DF2K-OST	0.0750	0.1341	0.1155	<i>0.1617</i>	0.1228	0.0876	0.0647
	SPSR [6]	DIV2K	<i>0.0616</i>	0.1313	0.1097	0.1629	0.1186	0.0866	0.0662
	RRDB+LDL [4]	DIV2K	0.0637	<i>0.1309</i>	<i>0.1007</i>	0.1635	<u>0.1097</u>	<i>0.0794</i>	<u>0.0546</u>
	CAL-GAN [8]	DIV2K	0.0687	0.1320	0.1072	0.1696	0.1171	0.0894	0.0688
	SROOE [9]	DF2K	0.0603	0.1131	0.0956	<u>0.1514</u>	0.1067	0.0758	0.0511
	Ours-a	DIV2K	0.1293	0.2148	0.1887	0.2723	0.1811	0.1342	0.0794
	Ours-b	DIV2K	0.0849	0.1604	0.1365	0.2051	0.1432	0.0978	0.0598
	Ours-c	DIV2K	<u>0.0607</u>	<u>0.1240</u>	<u>0.0994</u>	0.1508	<i>0.1154</i>	<u>0.0776</u>	<i>0.0576</i>

[13] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. 2004. Image Quality Assessment: from Error Visibility to Structural Similarity. *IEEE Transactions on*

Image Processing 13, 4 (2004), 600–612.

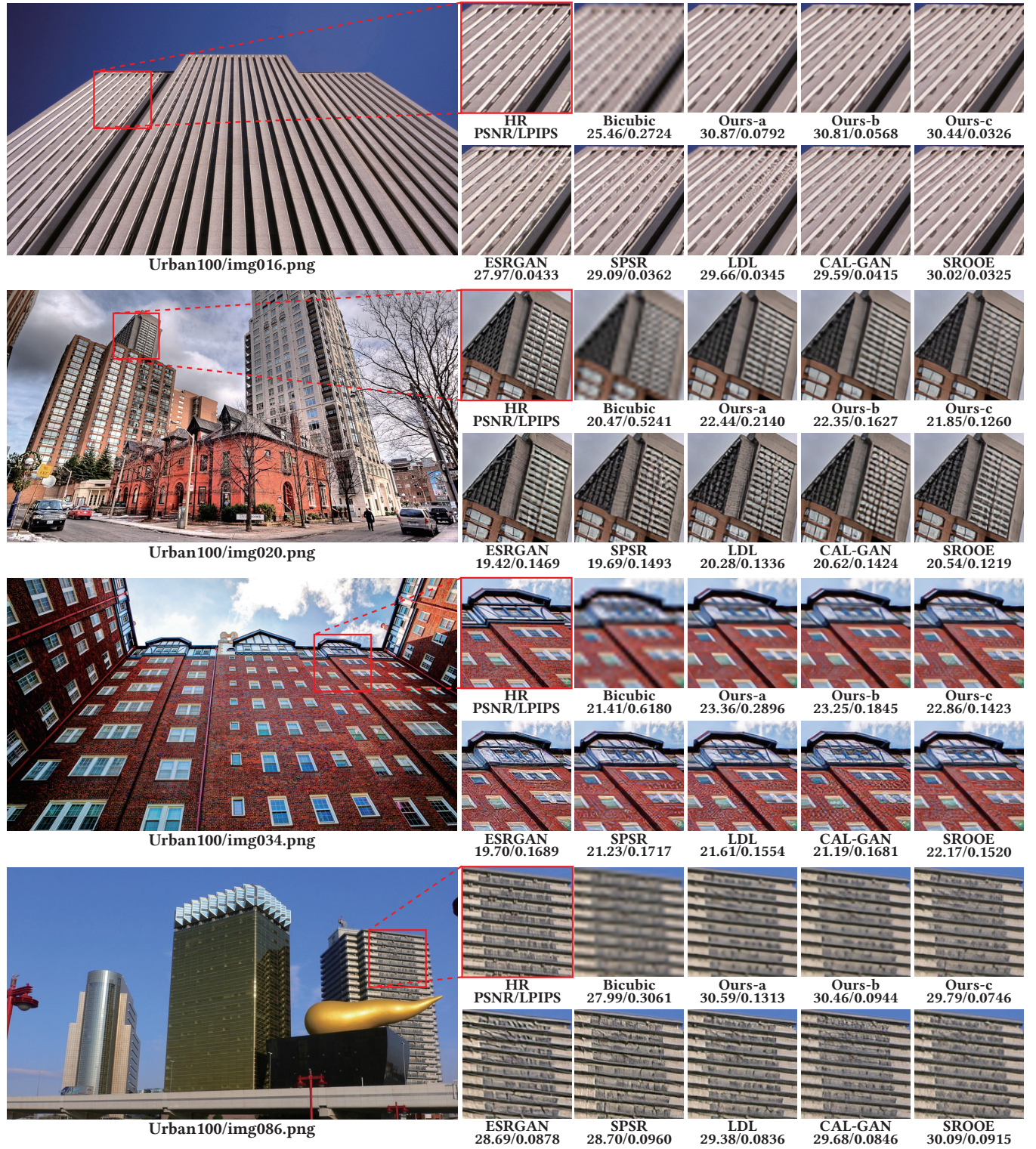


Figure 4: More visual comparisons on Urban100 [3].



Figure 5: More visual comparisons on DIV2K [1] validation set.