Supplemental Material

Yifan Wang[1,2] Federico Perazzi[2] Brian McWilliams[2]
Alexander Sorkine-Hornung[2] Olga Sorkine-Hornung[1] Christopher Schroers[2]

[1]ETH Zurich [2]Disney Research

module	layers	# out
$v_{\{0,1,2\}}$	CONV(3,3)	64
d_2	CONV(3,3)-RELU CONV(3,3)-RELU-AVGPOOL	64
d_1	CONV(3,3)-RELU CONV(3,3)-RELU-AVGPOOL	64
d_0	CONV(3,3)-RELU CONV(3,3)-RELU-AVGPOOL	128
	CONV(3,3)-RELU CONV(3,3)-RELU-AVGPOOL	256
	CONV(3,3)-RELU-AVGPOOL CONV(3,3)-RELU-AVGPOOL CONV(3,3)	512

Table 1: Detailed architecture specification for discriminator.

	PSNR	S14	B100	U100	DIV2K
$2\times$	ProSR$_\ell$	33.93	32.32	32.81	36.42
	ProGanSR	32.49	30.94	31.36	34.78
$4\times$	ProSR$_\ell$	28.90	27.77	26.77	30.79
	ProGanSR	26.82	25.71	25.13	28.61
$8\times$	ProSR$_\ell$	25.23	24.97	22.94	27.17
	ProGanSR	23.56	23.10	20.71	24.75

Table 2: PSNR values of ProSR and ProGanSR.

1. Extended Evaluation

Comprehensive Quantitative Comparison. In addition to the PSNR comparison, we provide the SSIM values of the proposed ProSRwith other state-of-the-art approaches in Table 4. When evaluating ProGanSR in terms of PSNR, we observe a drop of up to 2dB, as shown in 2, which aligns with the measures reported in other GAN-extended SISR methods [6, 8].

Visual Comparison. We present more results in Figure 2-4. Figure 2 and Figure 1 show more ProGanSR$_\ell$ results and the hallucinated fine details; Figure 4 and Figure 3 show more examples of ProSRin comparison with other approaches.

2. Implementation Details

2.1. Network Specification

In Table 3 we list the detailed architecture of the final proposed ProSR$_\ell$ and ProSR$_s$. In the very deep model ProSR$_\ell$, local and pyramidal residual links are adopted to facilitate gradient flow, hence compression units of each DCU always compress to the features to a fixed number e.g. 160 in order to enable element-wise addition in residual links; whereas in ProSR$_s$, we use simple sequential connection without residual links, and set the compression rate of each DCU to 0.4. The growth-rate used in ProSR$_\ell$ and ProSR$_s$ is 40 and 12 respectively. The number parameters in ProSR$_\ell$ for upsampling ratio $2\times \sim 8\times$ are 9.5M, 13.4M and 15.5M, while in ProSR$_s$ these are 1.1M, 2.1M and 3.1M respectively.

ProGanSRuses ProSR$_\ell$ as the generator; the architecture for discriminator is specified in Table 1. Depending on the upsampling ratio, the input is downsampled by a factor of 64, 32 and 16. The final output of the discriminator is a 512-channel feature.

2.2. Training Details

We train our network using DIV2K dataset [1], which consists of 800 high resolution training image (2K). The training patch size is 48×48, 40×40, 32×32 for upsample ratio $2\times$, $4\times$ and $8\times$ respectively. Random cropping, flipping and transpose is applied as data augmentation. Additionally we subtract the mean from DIV2K dataset and rescale the image to range $[-127.5, -127.5]$ as in [7]. Adam optimizer [3] is used with initial learning rate 0.0001 and the $1st$ and $2nd$ momentum is set to 0.9 and 0.999 respectively. The learning rate is halved after when the evaluation result hasn't improved for 40 epochs.

	ProSR$_\ell$			ProSR$_s$	
module	layers	# out	module	layers	# out
$v_{\{0,1,2\}}$	CONV(3,3)	160	$v_{\{0,1,2\}}$	CONV(3,3)	24
u_0	9 DCUs CONV(3,3) SP-CONV(3,3)	160 160 160	u_0	4 DCUs CONV(3,3) SP-CONV(3,3)	138 138 138
u_1	3 DCUs CONV(3,3) SP-CONV(3,3)	160 160 160	u_1	2 DCUs CONV(3,3) SP-CONV(3,3)	142 142 142
u_2	1 DCUs CONV(3,3) SP-CONV(3,3)	160 160 160	u_2	1 DCUs CONV(3,3) SP-CONV(3,3)	143 143 143
$r_{\{0,1,2\}}$	CONV(3,3)	3	$r_{\{0,1,2\}}$	CONV(3,3)	3

Table 3: Detailed architecture specification for ProSR$_\ell$ and ProSR$_s$. SP-CONV denotes sub-pixel convolution layer.

SSIM	2×				4×				8×			
	S14	B100	U100	DIV2K	S14	B100	U100	DIV2K	S14	B100	U100	DIV2K
VDSR	0.913	0.896	0.914	0.939	0.768	0.726	0.754	0.822	0.614	0.583	0.571	0.699
DRCN	0.913	0.894	0.913	-	0.768	0.724	0.752	-	0.614	0.582	0.571	0.694
DRRN	0.914	0.897	0.919	0.941	0.772	0.728	0.764	0.827	0.622	0.587	0.583	0.704
LapSRN	0.913	0.895	0.959	0.942	0.772	0.727	0.756	0.825	0.620	0.586	0.581	0.704
MsLapSRN	0.915	0.898	0.919	0.942	0.774	0.731	0.768	0.829	0.629	0.592	0.598	0.711
EnhanceNet	-	-	-	-	0.778	0.734	0.771	-	-	-	-	-
SRDenseNet	-	-	-	-	0.778	0.734	0.782	-	-	-	-	-
ProSR$_s$ (Ours)	**0.916**	**0.898**	**0.921**	**0.943**	**0.782**	**0.736**	**0.783**	**0.836**	**0.641**	**0.598**	**0.616**	**0.721**
EDSR	0.920	0.901	0.935	**0.949**	0.788	0.742	0.803	0.845	0.645	0.601	0.621	0.724
MDSR	0.920	0.901	0.935	0.948	0.786	0.742	0.804	0.845	-	-	-	-
ProSR$_\ell$ (Ours)	**0.921**	**0.902**	0.935	0.948	**0.790**	**0.743**	**0.809**	**0.846**	**0.652**	**0.606**	**0.645**	**0.731**

Table 4: SSIM evaluation compared with the state-of-the-art approaches.

References

[1] E. Agustsson and R. Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017. 1

[2] J. Kim, J. Kwon Lee, and K. Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1646–1654, 2016. 5, 6

[3] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014. 1

[4] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang. Deep laplacian pyramid networks for fast and accurate super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2017. 5, 6

[5] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang. Fast and accurate image super-resolution with deep laplacian pyramid networks. *arXiv preprint arXiv:1710.01992*, 2017. 5, 6

[6] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. *arXiv preprint arXiv:1609.04802*, 2016. 1, 4

[7] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee. Enhanced deep residual networks for single image super-resolution. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017. 1, 5, 6

[8] M. S. M. Sajjadi, B. Schölkopf, and M. Hirsch. Enhancenet: Single image super-resolution through automated texture synthesis. *CoRR*, abs/1612.07919, 2016. 1, 4

[9] Y. Tai, J. Yang, and X. Liu. Image super-resolution via deep recursive residual network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017. 5, 6

Figure 1: $8\times$ ProGanSR results compared to ProSR.

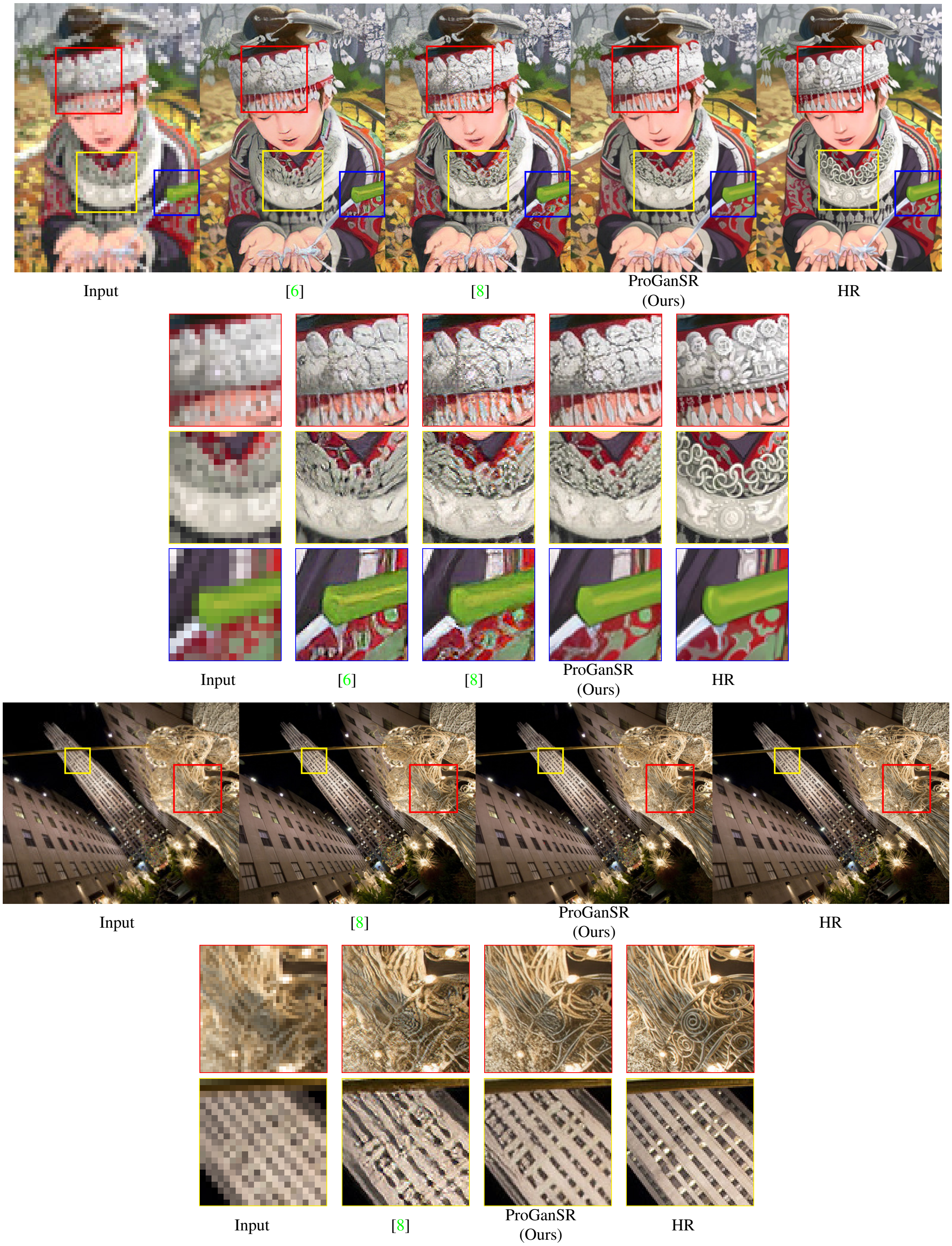

Figure 2: Comparison of $4\times$ GAN results.

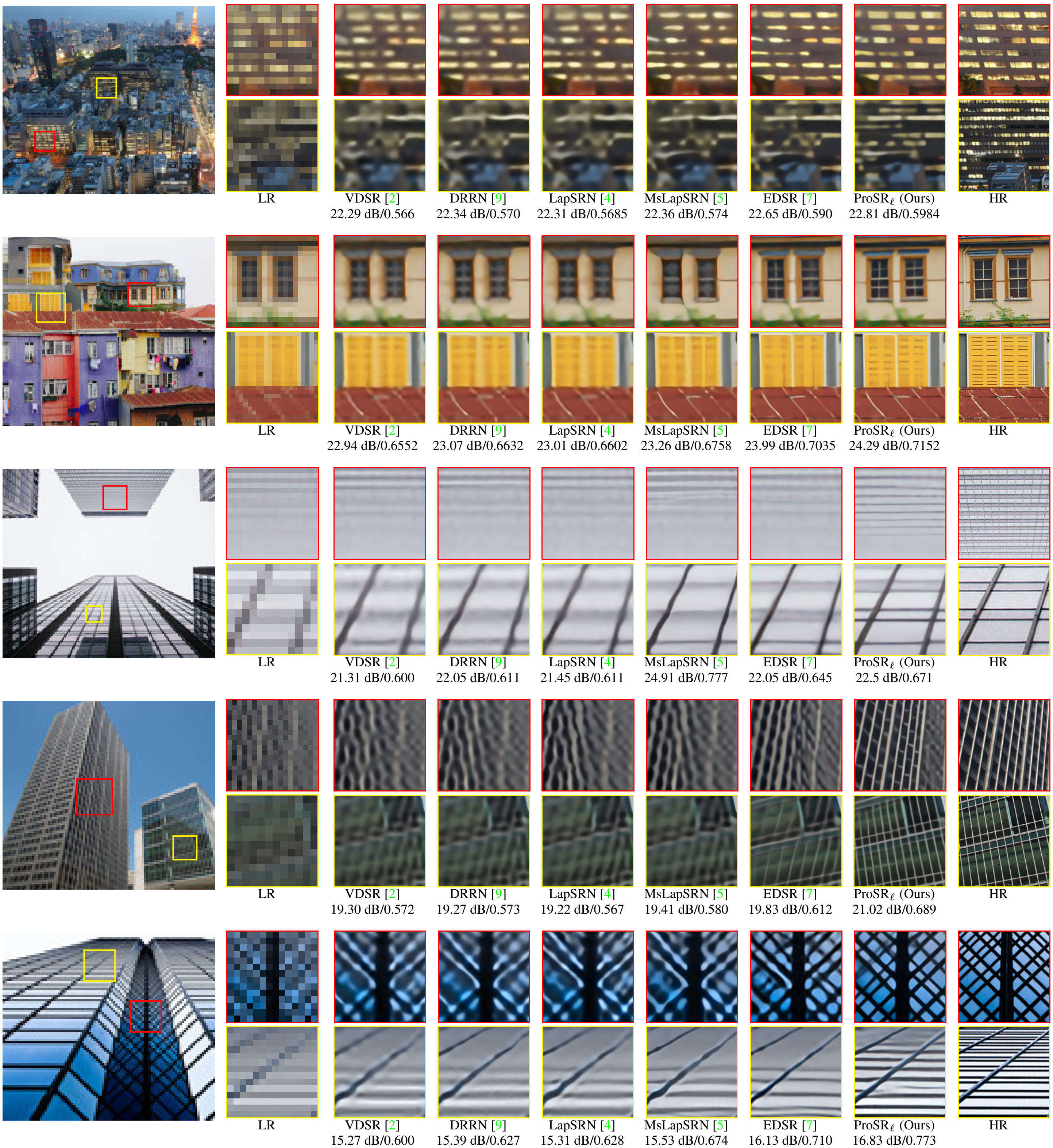

Figure 3: Comparison of 8× results between ProSR$_\ell$ (Ours) with other existing PSNR-driven models.

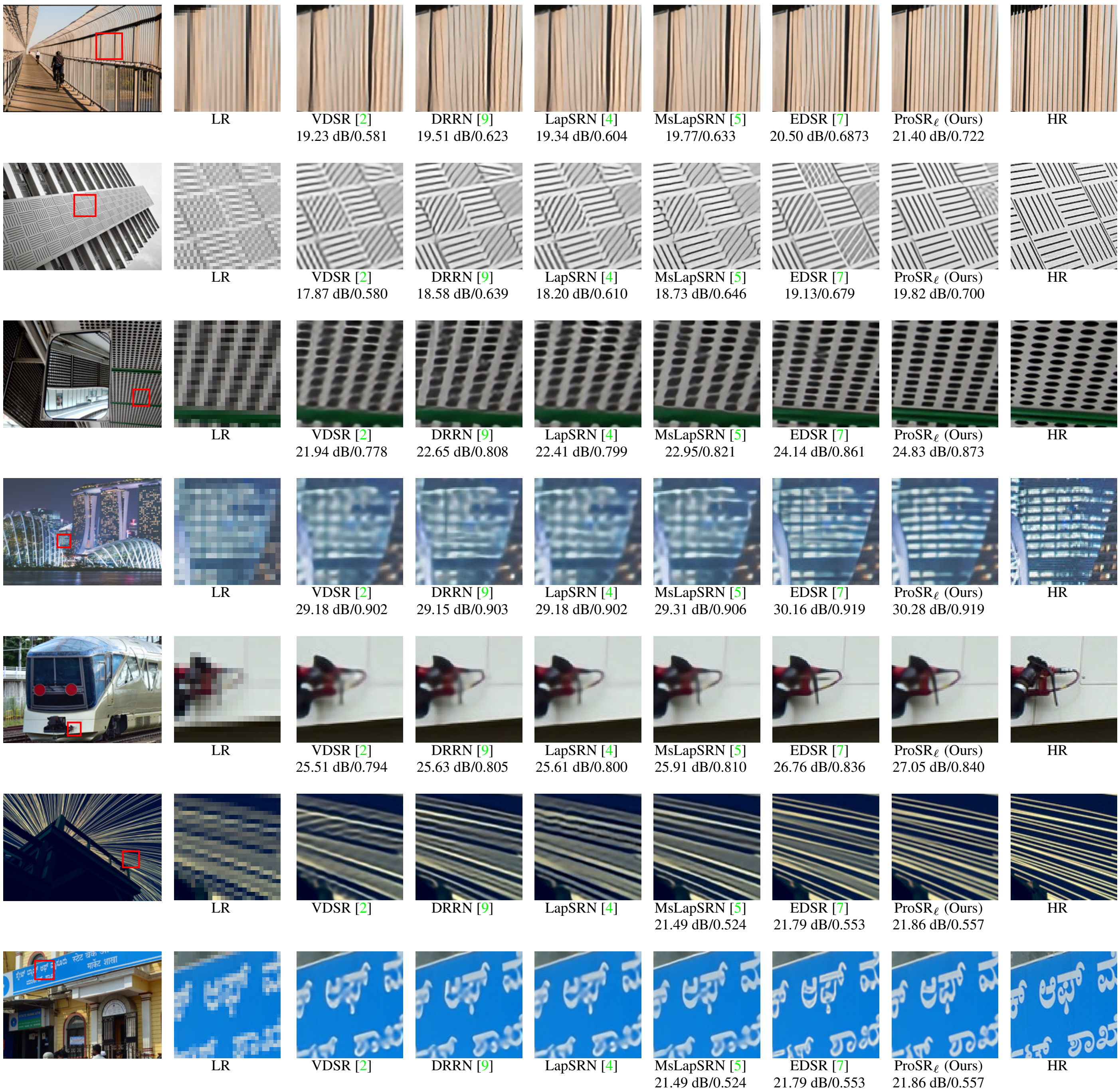

Figure 4: Comparison of $4\times$ results between ProSR$_\ell$ (Ours) with other existing PSNR-driven models.