ISP2HRNet: Learning to Reconstruct High Resolution Image from Irregularly Sampled Pixels via Hierarchical Gradient Learning

Yuanlin Wang[1], Ruiqin Xiong[1*], Rui Zhao[1], Jin Wang[2], Xiaopeng Fan[3], Tiejun Huang[1]
[1]State Key Laboratory of Multimedia Information Processing,
School of Computer Science, Peking University
[2]College of Computer Science, Beijing University of Technology
[3]School of Computer Science and Technology, Harbin Institute of Technology
{wangyuanlin, ruizhao}@stu.pku.edu.cn {rqxiong, tjhuang}@pku.edu.cn
ijinwang@bjut.edu.cn fxp@hit.edu.cn

6. Model Parameters

We present the parameter of each comparison model for the task of reconstruction from incomplete image with random missing pixels (Sec. 4.2) in Tab. 5. We also present model parameters for the task of super resolution (SR) from incomplete image with random missing pixels (Sec. 4.3) in Tab. 6. Our model achieves favorable performance with fewer parameters.

Method	Params(M)	DIV2K Val [1]	
		PSNR↑	SSIM↑
Partial Conv [7]	51.55	27.04	0.7998
RFR [5]	31.22	27.77	0.8343
MISF [6]	65.87	21.83	0.6733
DPDNN [2]	1.37	29.15	0.8583
DPIR [9]	32.64	29.51	0.8781
RRSNet [3]	3.77	29.15	0.8637
DiffPIR [11]	93.56	28.90	0.8482
CODE [10]	12.23	27.79	0.8383
Ours	**1.71**	**30.67**	**0.8932**

Table 5. Model parameters and quantitative results of reconstruction from incomplete image with 80% random missing pixels on DIV2K validation set [1].

7. Representation of Irregularly Sampled Pixels

The irregularly sampled pixels are unordered and unstructured. If the coordinates of the sampled pixels happen to be the integer coordinates of a regular grid, these sampled pixels can be represented by a regular image array, where the unknown pixels are set to 0. A binary mask can be used to indicate the sampled pixels and unknown pixels. However, this scenario is rare. In most cases, the coordinates of the randomly sampled pixels are fractional and cannot be easily represented by a finite-resolution regular grid. Therefore, following the strategy in point cloud processing [8], we organize irregular pixels into a list as input for ISP2HRNet.

*Corresponding author.

Method	Params(M)	DIV2K Val [1]				
		×1.5	×2	×2.5	×3*	×4*
DPIR+Bicubic	32.64	30.80	28.76	27.43	26.52	25.20
DPIR+EDSR^b	34.01	-	29.50	-	27.13	25.72
DPIR+EDSR	73.37	-	29.52	-	27.14	25.73
DPIR+SwinIR	44.39	-	29.52	-	27.14	25.74
DPIR+LIIF	34.21	31.45	29.43	28.02	27.08	25.69
Ours	**1.71**	**31.82**	**29.58**	**28.17**	**27.16**	**25.76**

Table 6. Model parameters and quantitative results of super resolution from incomplete image with 50% random missing pixels on DIV2K validation set [1]. (EDSR^b: EDSR-baseline, the encoder in LIIF is the EDSR-baseline, * indicates the equivalent sampling ratio is out of our model's training distribution.)

8. Experimental Details

8.1. Relationship Between M and N

During training, we set the number of input irregularly sampled pixels N ***equal*** to the number of regular grid points $\mathrm{M} = \mathrm{h} \times \mathrm{w} = 48 \times 48 = 2304$ in IRC module (Sec. 4.1). This setting ensures compact information transfer and reduces potential information loss.

8.2. Inference Details

The memory required for searching nearest neighbors is proportional to the square of the number of irregular pixels (N^2). As N increases, the memory demand rises sharply.

Due to memory constraints, we reconstruct a high resolution (HR) image by stitching multiple reconstructed patches and averaging the overlapping areas.

Specifically, we first obtain a ratio r' between the number of sampled pixels and the resolution of an image that we expect to reconstruct. Then, we divide the target resolution plane into multiple patches from left to right and top to bottom. The size of each patch is $round(48/\sqrt{r'}) \times round(48/\sqrt{r'})$, which ensures that the number of irregular pixels falling into each patch is approximately 2304. The stride for dividing patches is $\frac{1}{2} \cdot round(48/\sqrt{r'})$, i.e., there is an overlapping area of size $\left(\frac{1}{2} \cdot round(48/\sqrt{r'})\right) \times round\left(48/\sqrt{r'}\right)$ between any two adjacent patches. For high resolution image reconstruction of a patch, we choose 2304 irregular pixels falling within the patch as the network input. If the number of irregular pixels falling in the patch is less than 2304, we randomly choose repeated pixels until the number reaches 2304. If the number of irregular pixels falling in the patch exceeds 2304, we randomly select 2304 pixels from them. The proposed ISP2HRNet generates 48×48 regular grid features in IRC module, and reconstructs a high resolution patch using the implicit neural representation. After generating HR reconstruction results of all patches, we obtain the target image by stitching all reconstructed patches and averaging the overlapping areas.

8.3. Training and Testing Details in Sec. 4.2

For a fair comparison, all comparison methods (except DPIR and DiffPIR) are retrained on the DIV2K training dataset. For DPIR and DiffPIR, we retrain the denoiser for DPIR and the diffusion model for DiffPIR on the DIV2K training dataset.

These comparison methods are retrained and tested using three-channel 2D image arrays as input, with randomly missing pixels set to zero, as these methods cannot handle input formats other than image arrays. Although the coordinates of the irregular pixels in this task are integers (i.e., they can be represented by a 2D image array with unknown pixels set to zero), we still organize these pixels into a list as the input to ISP2HRNet. Following the strategy in image inpainting, the sampled pixels of input are fused to the output image to generate the final reconstruction result. For ISP2HRNet, we trained a single model that is used to address all tasks in Sec. 4.

Additionally, we note that RRSNet [3] was originally trained and evaluated on grayscale (Y channel) images, as reflected by its source code. In our experiments, we retrain and test RRSNet on color images, which largely accounts for the difference between our reported results (Tab. 1) and those in the original paper. Another factor contributing to this discrepancy lies in different training strategies.

8.4. Comparison Methods in Sec. 4.3

Our work focuses on high resolution image reconstruction from irregular pixels, which has been less explored in the community recently. Therefore, we employ a straightforward idea to combine existing image restoration and SR methods, which could achieve similar functions for comparison. We choose DPIR [9] as the image restoration method as it outperforms other methods in Tab. 1. Specifically, DPIR first reconstructs an LR image from the degraded LR input. Then, image SR methods take the reconstructed LR image as input and output HR image.

8.5. Difference between Tasks in Sec. 4.2 and Sec. 4.4

Tasks in Sec. 4.2 and Sec. 4.4 both use HR images from the test datasets as the source of sampled pixels and the ground truth. In Sec. 4.2, the coordinates of the irregularly sampled pixels are integers, and the pixel values at these coordinates can be directly obtained. The coordinates of sampled pixels in Sec. 4.4 are decimal, and the pixel values at these coordinates are obtained by bicubic interpolation. In Sec. 4.2, the sampled pixels are fused to the output image to form the final results for evaluation. As the sampling ratio increases, the final results in Sec. 4.2 contain more known pixel values, resulting in an increasing performance gap between Tab. 1 and Tab. 3. For example, on the DIV2K validation set, when the sampling ratio is 0.2, the PSNR is 30.67 dB in Tab. 1 and 30.55 dB in Tab. 3, resulting in a performance difference of 0.12 dB. In contrast, when the sampling ratio increases to 0.8, the PSNR is 42.19 dB in Tab. 1 and 38.07 dB in Tab. 3, yielding a much larger gap of 4.12 dB.

8.6. Ablation Study on K Nearest Neighbors

We explore the choice of K, i.e., the number of nearest neighbors used for extracting image gradient structure at irregular positions, as presented in Tab. 7. For a trade-off between performance and computation, we consider $\mathrm{K} = 8$ a reasonable choice.

Ratio	$\mathrm{K}=4$		$\mathrm{K}=6$		$\mathrm{K}=8$		$\mathrm{K}=12$	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
0.2	26.44	0.8826	27.35	0.8854	**27.43**	**0.8864**	27.45	0.8867
0.4	31.20	0.9447	31.38	0.9453	**31.40**	**0.9455**	31.47	0.9460

Table 7. Ablation study on K, i.e., the number of nearest neighbors used for extracting image gradient structure at irregular positions.

9. More Visual Results

9.1. Reconstruction from Incomplete Image with Random Missing Pixels

We provide more visual results of reconstruction from incomplete image with random missing pixels in Fig. 7. Most comparison methods produce reconstruction results with severe artifacts. Although DPIR [9] and DiffPIR [11] are able to restore clear texture structures in certain cases (*e.g.* img027, img044 from Urban100 [4]), our method achieves better performance in reconstructing fine details.

9.2. Super Resolution from Incomplete Image with Random Missing Pixels

We provide more visual results of super resolution from incomplete image with random missing pixels in Figs. 8 and 9. Note that for large super resolution scales, both our method and comparion methods struggle to reconstruct complete details. However, our method preserves neat edges, while the comparison methods suffer from visually annoying artifacts and rough edges.

9.3. Reconstruction from Irregularly Sampled Pixels

We present the visual results of reconstruction from irregularly sampled pixels at different sampling ratios in Fig. 10. Fig. 10 includes the irregularly sampled pixels that serve as the input to ISP2HRNet, the initial regular grid pixel value reconstruction results which are generated by the IRC module, and the reconstruction results output by ISP2HRNet. The resolution of initial reconstruction in IRC module is related to the number of irregularly sampled pixels N, because $\mathrm{M} = \mathrm{N}$ (M is the number of regular grid coordinates in IRC module). The more sampled pixels, the higher the resolution of initial reconstruction. We indicate the resolution of initial reconstruction and final reconstruction results in Fig. 10. The proposed ISP2HRNet can handle reconstruction from varying numbers of irregularly sampled pixels. As the number of sampled pixels increases, the reconstructed texture becomes closer to ground truth.

9.4. Super Resolution from a Fixed Number of Irregularly Sampled Pixels

The proposed ISP2HRNet can handle arbitrary-size high resolution image reconstruction from a fixed number of irregularly sampled pixels, as shown in Fig. 11. The input is irregularly sampled from the 100×100 origin image (with decimal coordinates), and we perform super resolution at multiple upsampling scales. We present HR reconstruction results of varying resolutions. As the upsampling scale increases, the reconstruction results maintain clear textures. It is noted that a single model is used to address all tasks in Secs. 9.1 to 9.4.

References

[1] Eirikur Agustsson and Radu Timofte. 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)*, pages 126–135, 2017. 1, 8

[2] Weisheng Dong, Peiyao Wang, Wotao Yin, Guangming Shi, Fangfang Wu, and Xiaotong Lu. Denoising prior driven deep neural network for image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(10):2305–2318, 2018. 1

[3] Yanchen Dong, Rui Zhao, Ruiqin Xiong, Shuyuan Zhu, Xiaopeng Fan, and Tiejun Huang. Optimization-inspired deep network for image restoration from partial random samples. In *IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–5, 2023. 1, 2

[4] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5197–5206, 2015. 3

[5] Jingyuan Li, Ning Wang, Lefei Zhang, Bo Du, and Dacheng Tao. Recurrent feature reasoning for image inpainting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 1

[6] Xiaoguang Li, Qing Guo, Di Lin, Ping Li, Wei Feng, and Song Wnag. MISF: Multi-level interactive siamese filtering for high-fidelity image inpainting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 1

[7] Guilin Liu, Fitsum A Reda, Kevin J Shih, Ting-Chun Wang, Andrew Tao, and Bryan Catanzaro. Image inpainting for irregular holes using partial convolutions. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 85–100, 2018. 1

[8] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. PointNet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 652–660, 2017. 1

[9] Kai Zhang, Yawei Li, Wangmeng Zuo, Lei Zhang, Luc Van Gool, and Radu Timofte. Plug-and-play image restoration with deep denoiser prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):6360–6376, 2021. 1, 2, 3

[10] Haiyu Zhao, Yuanbiao Gou, Boyun Li, Dezhong Peng, Jiancheng Lv, and Xi Peng. Comprehensive and delicate: An efficient transformer for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14122–14132, 2023. 1

[11] Yuanzhi Zhu, Kai Zhang, Jingyun Liang, Jiezhang Cao, Bihan Wen, Radu Timofte, and Luc Van Gool. Denoising diffusion models for plug-and-play image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2023. 1, 3

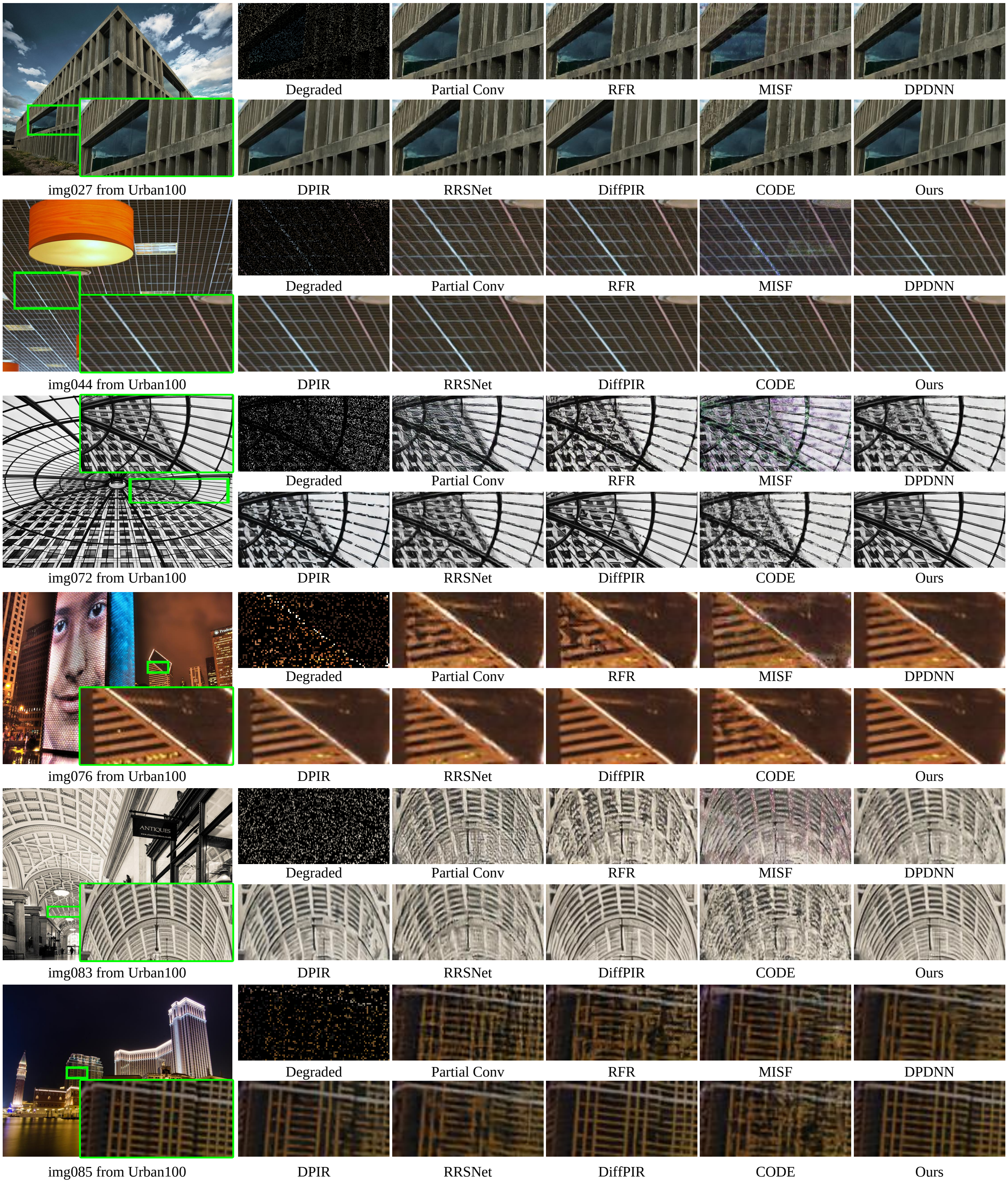

Figure 7. Visual results of reconstruction from incomplete image with random missing pixels. The sampling ratio is 0.2. Please enlarge the figure for better comparison.

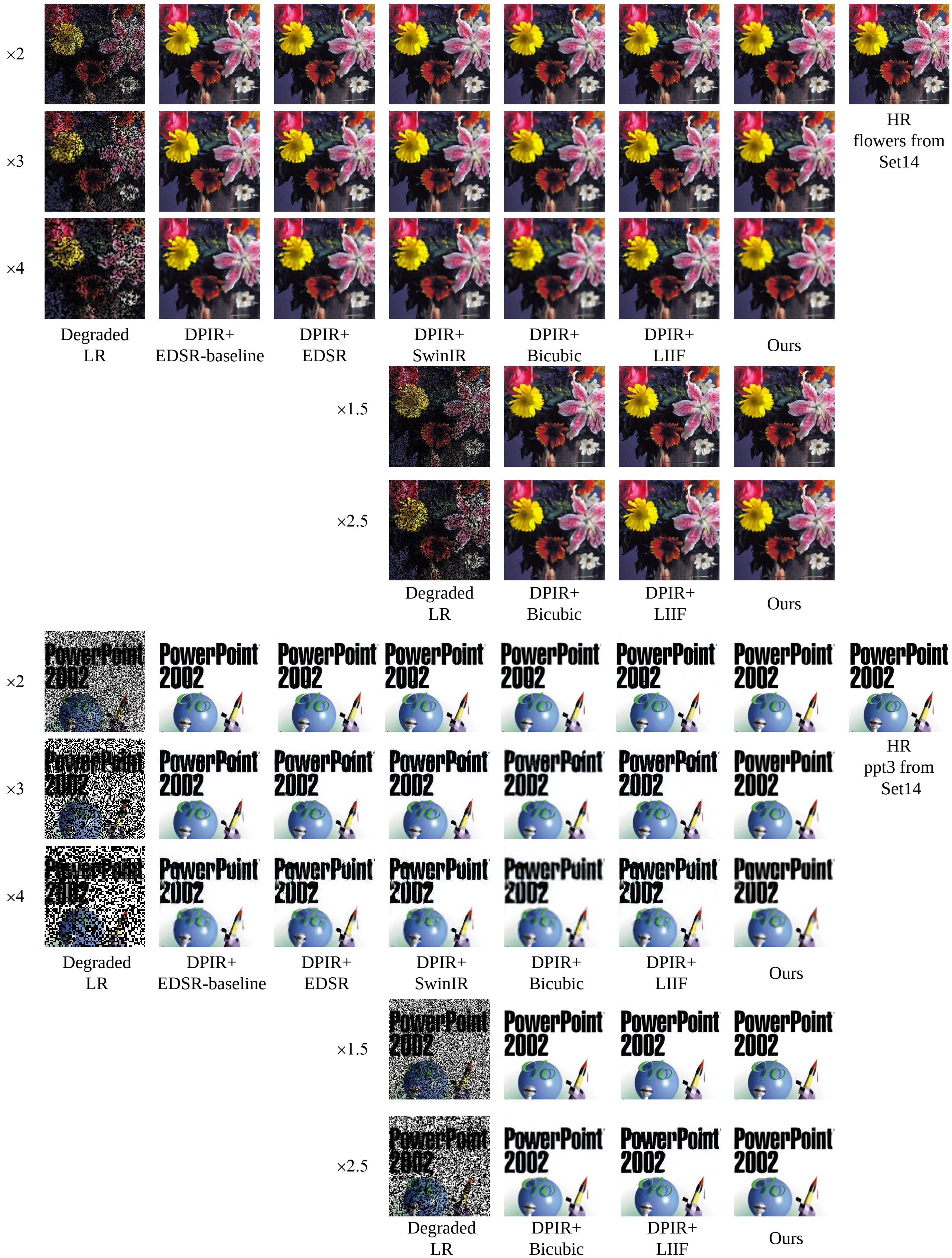

Figure 8. Visual results of super resolution from incomplete image with random missing pixels. The sampling ratio is 0.5 in low resolution image. Please enlarge the figure for better comparison.

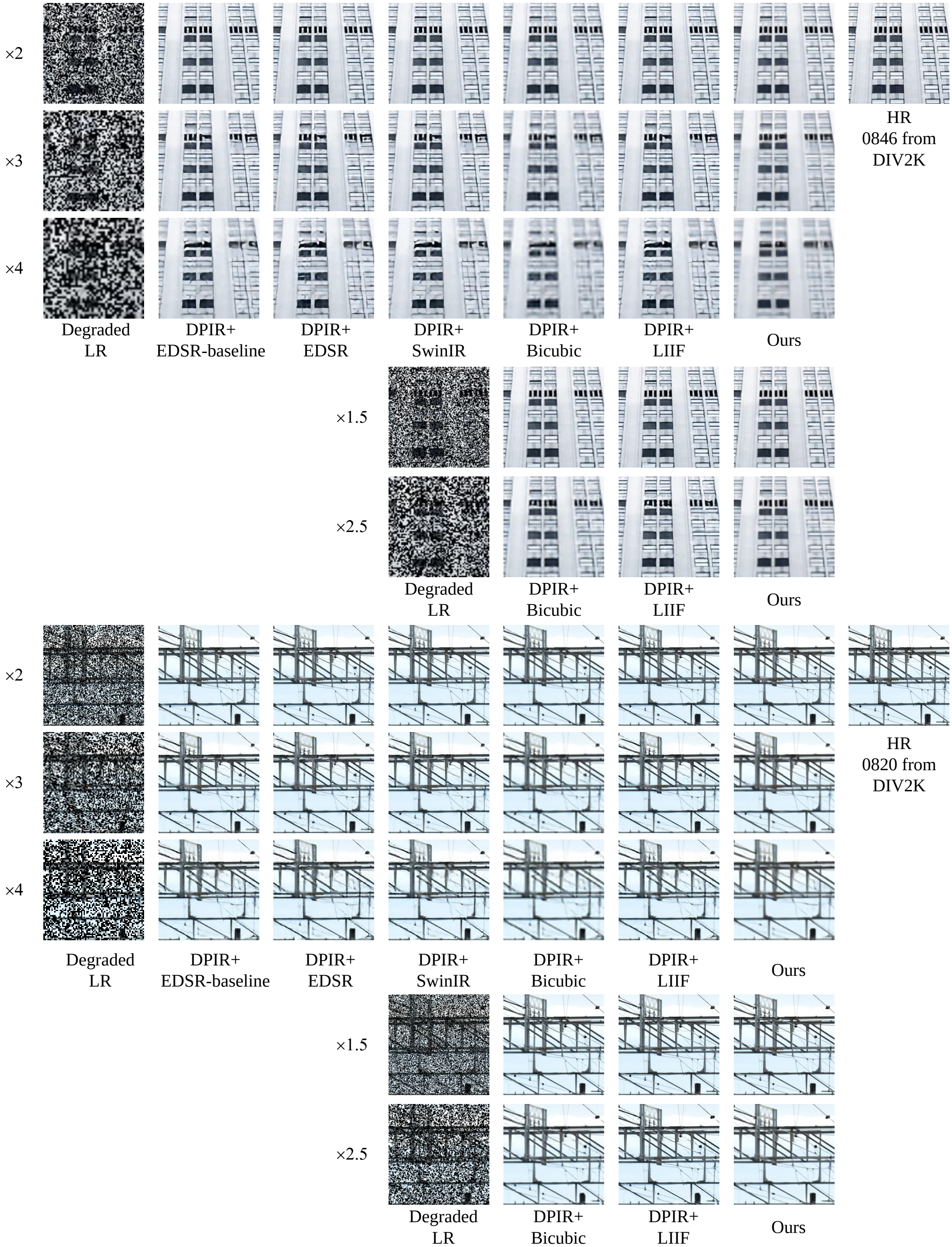

Figure 9. Visual results of super resolution from incomplete image with random missing pixels. The sampling ratio is 0.5 in low resolution image. Please enlarge the figure for better comparison.

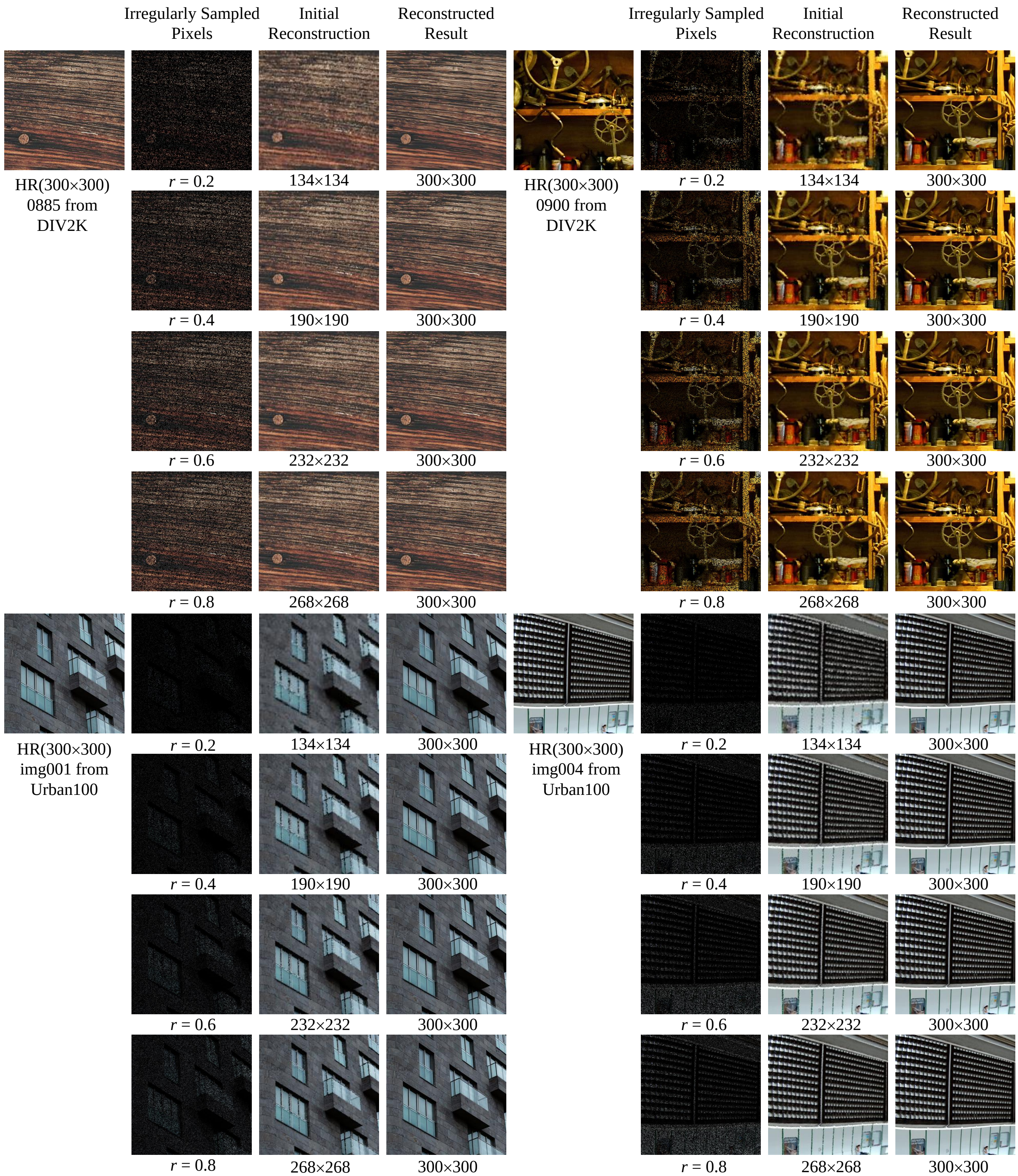

Figure 10. Visual results of reconstruction from irregularly sampled pixels. r is the sampling ratio.

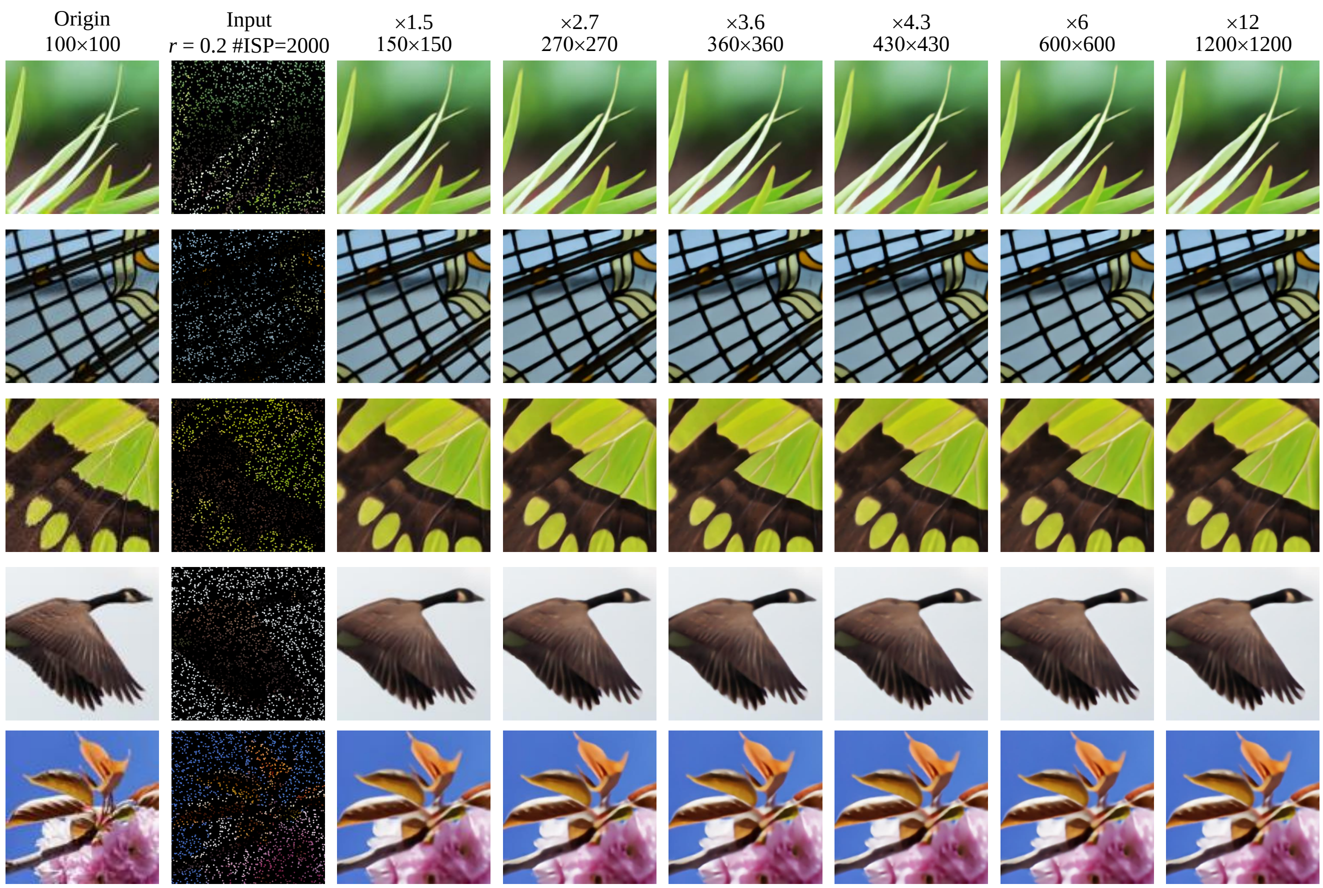

Figure 11. Visual results of super resolution from a fixed number of irregularly sampled pixels. The sampling ratio is 0.2 in the original image. #ISP is the number of irregularly sampled pixels. From top to bottom: 0803, 0821, 0829, 0896, and 0898 from the DIV2K dataset [1].