Blockwise Divide-and-Aggregate for Image Restoration using Diffusion Priors - Appendix

Table A. **Improved fidelity metrics across degradations**. The improved metrics are shown in green compared to our method in the main paper.

Method	Sampler	NFEs	Agg	Degradation	PSNR↑	CMMD↓	LPIPS↓	FID↓
DDNM+ (from main paper)	DDIM	100	–	SR×4	24.91	0.785	0.405	99.62
				Deblurring	25.05	1.091	0.281	88.43
				Inpainting	26.58	0.237	0.243	58.72
DDNM+	EDM	500	–	SR×4	25.57	0.706	0.346	96.37
				Deblurring	25.73	0.782	0.424	125.61
				Inpainting	25.91	0.702	0.488	74.21
DAPS-1K (from main paper)	EDM	1000	–	SR×4	25.89	0.662	0.276	83.57
				Deblurring	26.15	0.721	0.253	75.68
				Inpainting	28.44	0.423	0.135	67.25
Naive baseline (MMSE)	EDM	500	–	SR×4	27.64	0.923	0.374	138.15
				Deblurring	27.70	0.864	0.331	112.21
				Inpainting	31.03	0.916	0.205	74.19
Ours (from main paper)	DDIM	100	MMSE-style	SR×4	26.64	0.913	0.358	133.26
				Deblurring	26.70	0.824	0.339	111.72
				Inpainting	30.03	0.904	0.211	62.55
Ours	EDM	500	Patchwise	SR×4	24.87	**0.635**	**0.289**	**89.34**
				Deblurring	25.79	**0.841**	**0.320**	**106.16**
				Inpainting	27.41	**0.389**	**0.102**	**41.72**

A. Additional Qualitative Results on ImageNet Dataset

In Figure A and Figure B, we present additional qualitative results for ImageNet experiments shown in Table 1.

B. Quantitative Results on LSUN-Bedroom Dataset

The quantitative results for the LSUN-Bedroom dataset are presented in Table 2.

C. Additional Details of Our Method

Patch-wise Aggregation. MMSE-style aggregation across independently sampled trajectories can lead to over-smoothing. Importantly, blur is not intrinsic to the proposed divide-and-aggregate framework but depends on the aggregation scheme. Instead of globally averaging trajectories, we aggregate locally in overlapping patches of size 16×16. For each patch, we select the trajectory with the highest local gradient energy, computed from finite-difference spatial gradients, and stitch patches using a smooth Tukey window. This avoids averaging misaligned high-frequency structures-the primary cause of blur-while periodic re-aggregation enforces global consistency. Visual comparisons in Fig. C and Fig. D further demonstrate the improved sharpness of our method.

Handling Overlap : In the patch-wise aggregation scheme, Tukey window smoothing mitigates artifacts arising from overlapping patches.

MMSE & Naive Baseline: We also report a naive MMSE baseline, obtained via Monte Carlo averaging of multiple denoised trajectories, in Table A. While this yields high PSNR, it performs significantly worse in CMMD, FID, and LPIPS.

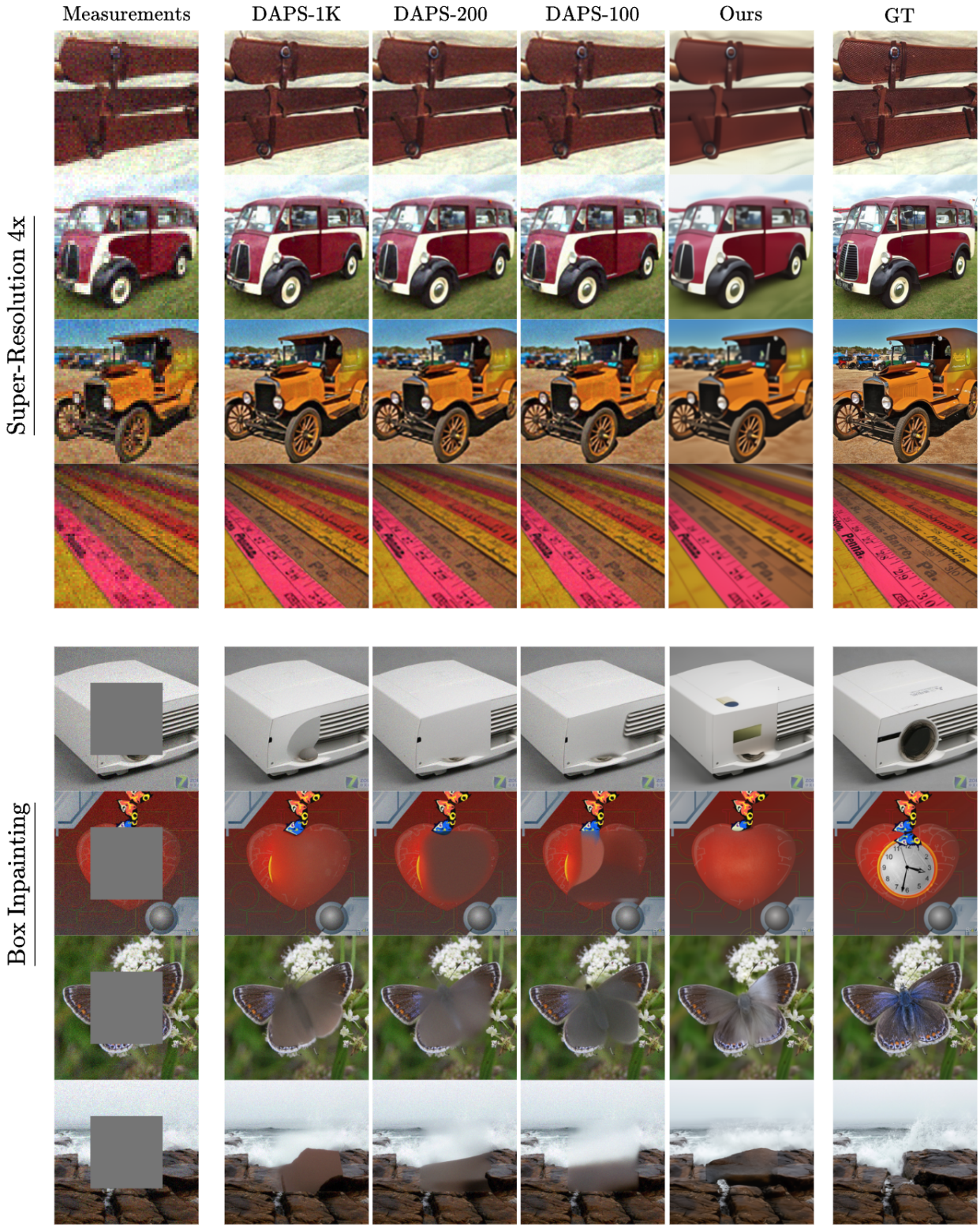

Figure A. Additional qualitative results for ImageNet experiments on SR 4x and Box inpainting.

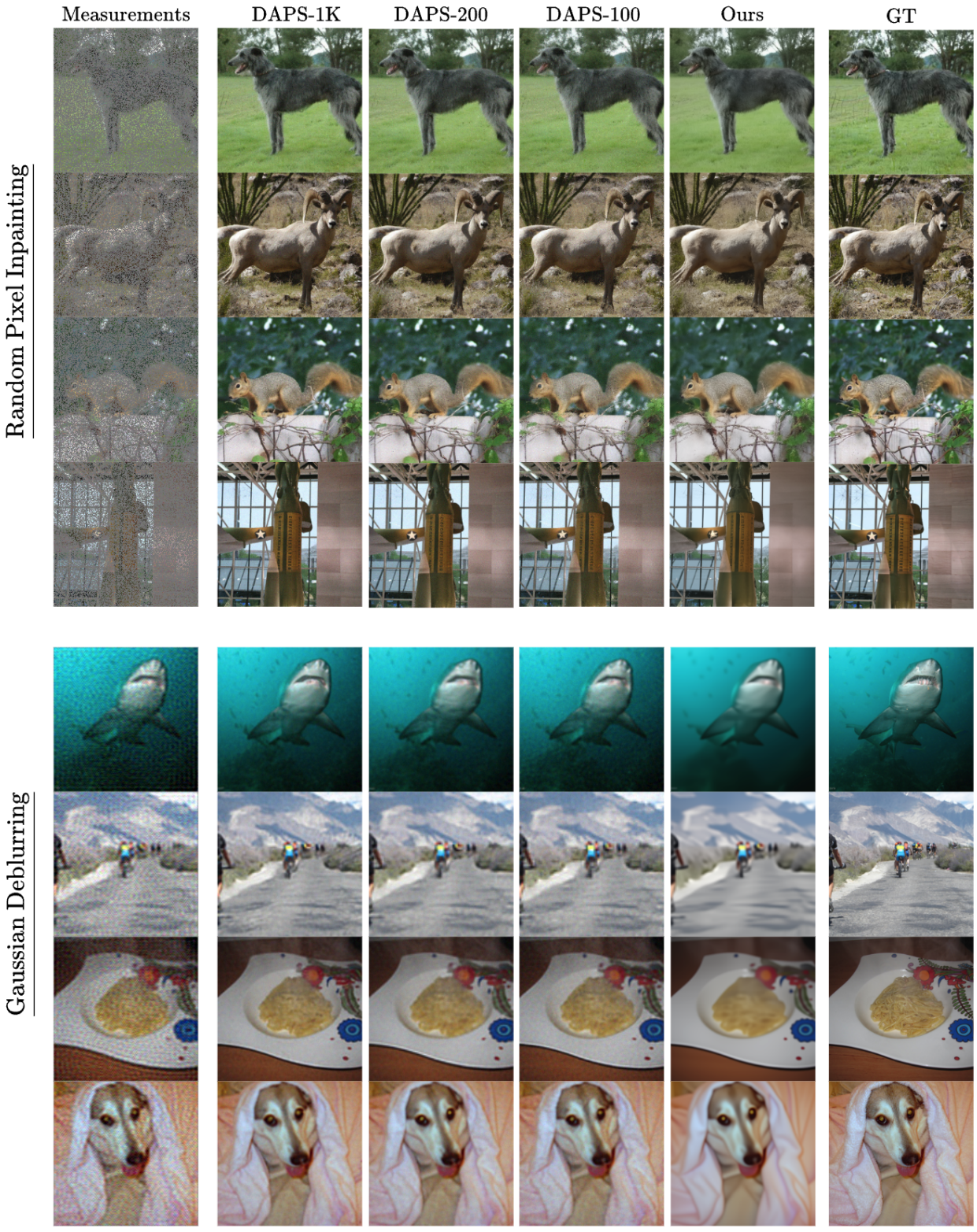

Figure B. Additional qualitative results for ImageNet experiments on random pixel inpainting and Gaussian deblurring.

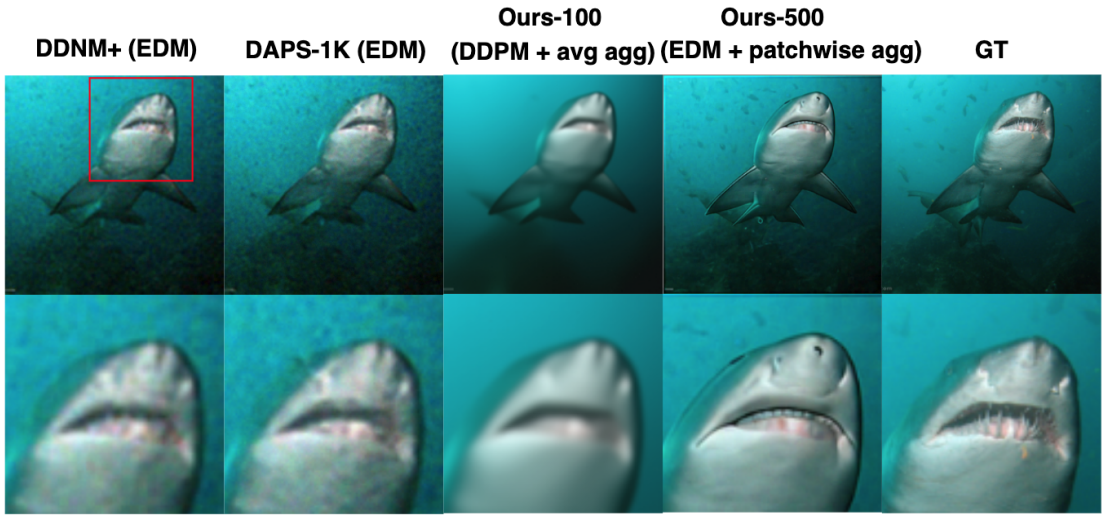

Figure C. Reconstruction sharpness improvements using patchwise aggregation shown for SR4×. Ours-500 with EDM sampler and patchwise aggregation produces sharper reconstructions.

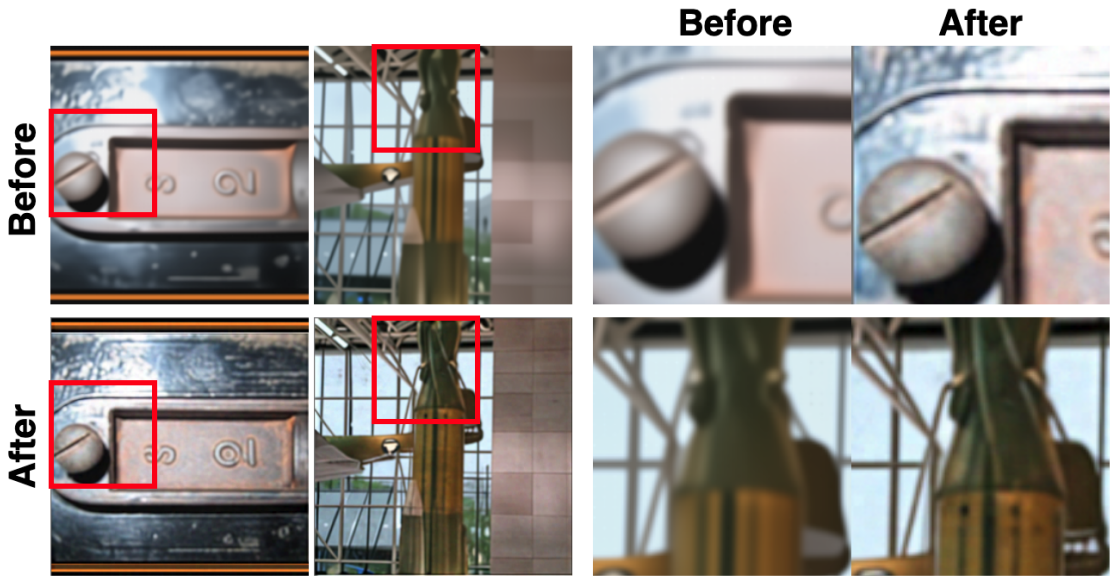

Figure D. A close comparison of results from the main paper (Before) and improved results (After).

D. Experiment Design Details - Hyper-parameters and Degradation Operators

This section provides additional details about the hyperparameters used by baseline methods and our method for each experiment table provided in the main paper. Additionally, we provide the details about the forward operators used in our experiments.

D.1. Hyper-parameters

ImageNet Experiments. For ImageNet experiments we follow the same settings as DAPS [46]. We use the same set of 100 images from the official repo to benchmark our method. All the numbers for DDRM, DPS, FPS-SMC, and DAPS-1k in Table 1 are borrowed from DAPS [46]. We run the experiments for DAPS-200, DAPS-100, and our method. We use $N = M = 4$, where M and N are the number of trajectories and blocks/patches, respectively. This provides a good tradeoff between speed and metrics for our method. Note that we use batch size M, i.e., we parallelly denoise all the trajectories. We use deterministically paired updates followed by MMSE-style aggregation scheme.

LSUN-Bedroom Experiments. η - Hyper-parameter for the DDIM sampler; T - Number of denoising steps; $|\mathcal{B}|$ - Batch size; $\sigma_{\mathbf{y}}$ - measurement noise

- **DDRM :** η_b - DDRM hyper-parmeter. We use the default value of 1 as suggested by authors of the paper in their github repo.
- **DPS :** ζ - controls the strength of gradient update. We find best value for each degradation.
- **IDPG and DDPG:** Scale LS (SLS), Gamma (g), Step-size-mode (smode), ζ, inject-noise (in-noise), Eta-Tilde $\tilde{\eta}$
- **Our method :** N - number of blocks; M - number of trajectories.

Table B. Hyper-parameters for LSUN-Bedroom (256 × 256) experiments in the main paper.

Method	8× Super-Resolution	Deblur	20% Pixel Inpainting	Compressed Sensing
DDRM	$\eta = 0.85, \eta_b = 1.0, T = 20$	$\eta = 0.85, \eta_b = 1.0, T = 20$	$\eta = 0.85, \eta_b = 1.0, T = 20$	$\eta = 0.85, \eta_b = 1.0, T = 20$
DPS	$\zeta = 55.0, \eta = 0.5, T = 100$	$\zeta = 25.0, \eta = 0.85, T = 100$	$\zeta = 25.0, \eta = 0.85, T = 100$	$\zeta = 25.0, \eta = 0.85, T = 100$
DDNM+	$\eta = 0.5, \sigma_{\mathbf{y}} = 0.1, T = 100$	$\eta = 0.85, \sigma_{\mathbf{y}} = 0.1, T = 100$	$\eta = 0.85, \sigma_{\mathbf{y}} = 0.1, T = 100$	$\eta = 0.85, \sigma_{\mathbf{y}} = 0.1, T = 100$
IDPG	$\eta = 0.5, \sigma_{\mathbf{y}} = 0.1, T = 100$ SLS = 1.0, $g = 300$, smode=1 in-noise = 0, $\zeta = 0.5$, $\tilde{\eta}$ = 1.0	$\eta = 0.85, \sigma_{\mathbf{y}} = 0.1, T = 100$ SLS = 1.0, $g = 11$, smode=1 in-noise = 0, $\zeta = 0.7$, $\tilde{\eta}$ = 1.0	$\eta = 0.5, \sigma_{\mathbf{y}} = 0.1, T = 100$ SLS = 1.0, $g = 300$, smode=1 in-noise = 0, $\zeta = 0.5$, $\tilde{\eta}$ = 1.0	$\eta = 0.5, \sigma_{\mathbf{y}} = 0.1, T = 100$ SLS = 1.0, $g = 11$, smode=1 in-noise = 0, $\zeta = 0.7$, $\tilde{\eta}$ = 1.0
DDPG	$\eta = 0.5, \sigma_{\mathbf{y}} = 0.1, T = 100$ SLS = 1.0, $g = 300$, smode=1 in-noise = 1, $\zeta = 0.5$, $\tilde{\eta}$ = 1.0	$\eta = 0.85, \sigma_{\mathbf{y}} = 0.1, T = 100$ SLS = 1.0, $g = 11$, smode=1 in-noise = 1, $\zeta = 0.7$, $\tilde{\eta}$ = 1.0	$\eta = 0.85, \sigma_{\mathbf{y}} = 0.1, T = 100$ SLS = 1.0, $g = 300$, smode=1 in-noise = 1, $\zeta = 0.5$, $\tilde{\eta}$ = 1.0	$\eta = 0.85, \sigma_{\mathbf{y}} = 0.1, T = 100$ SLS = 1.0, $g = 11$, smode=1 in-noise = 1, $\zeta = 0.7$, $\tilde{\eta}$ = 1.0
Ours	$\eta = 0.5, \sigma_{\mathbf{y}} = 0.1, T = 100$ $N = M = 4$	$\eta = 0.5, \sigma_{\mathbf{y}} = 0.1, T = 100$ $N = M = 4$	$\eta = 0.85, \sigma_{\mathbf{y}} = 0.1, T = 100$ $N = M = 4$	$\eta = 0.85, \sigma_{\mathbf{y}} = 0.1, T = 100$ $N = M = 4$

D.2. Details of Degradation Operators for LSUN-Bedroom Dataset

- **Super-Resolution:** For super-resolution experiments, we utilize an average pooling downsampler with 8x and 4x downsampling for LSUN-Bedroom and ImageNet datasets, respectively. The pseudo-inverse of this operator is computed following the approach described in DDNM [42].
- **Deblurring:** For deblurring experiments, we consider a Gaussian kernel with a kernel size of 5 and a standard deviation of 20.0.
- **Pixel Inpainting:** For random pixel inpainting experiments, we employ a predefined mask with 20% visible pixels (for Table 1) or 10% visible pixels (for Table 2). The same mask is used across all experiments to ensure consistency.
- **Mask Inpainting:** For mask inpainting experiments, we utilize a predefined random mask that remains consistent across all experiments.
- **Compressed Sensing (CS):** For compressed sensing, we adopt a Walsh-Hadamard sampling matrix with sampling ratios of 0.20 (20%) and 0.15 (15%).
- **Bicubic Downsampling:** For additional qualitative evaluations, we use a bicubic downsampler with a degradation strength of 4x.
- **Colorization:** For colorization, we define the measurement operator as $\mathbf{H} = [\frac{1}{3}, \frac{1}{3}, \frac{1}{3}]$ for each pixel.
- **Gaussian Denoising:** For Gaussian denoising experiments, noise is sampled from a Gaussian distribution with zero mean and a standard deviation of $\sigma_{\mathbf{y}} = 0.10$.
- **Anisotropic Deblurring:** For anisotropic deblurring, we adopt settings similar to those described in DDNM [42].

E. Comparison with Other Degradation Operators

This section presents sample results for our method and DDNM+ across various degradations. Figure E illustrates results for the anisotropic deblurring and Gaussian denoising tasks $\sigma_{\mathbf{y}} = 0.25$. Figure F showcases colorization results, while Figure G provides results for mask inpainting. Additionally, Figure H demonstrates results for the bi-cubic downsampling degradation.

It is important to note that the results presented in this section serve as evidence that our method is applicable to a wide range of degradations. However, we do not claim superior performance over DDNM+ for these specific tasks.

F. Additional Qualitative Results

We present additional qualitative results from Table 1 in the main paper in this section.

- **LSUN-Bedroom (256 x 256) :** 20% inpainting in Figure. I, Walsh-Hadamard CS for 20% compression ration in Figure. J, Gaussian deblurring in Figure. K, SR-8x in Figure. L.

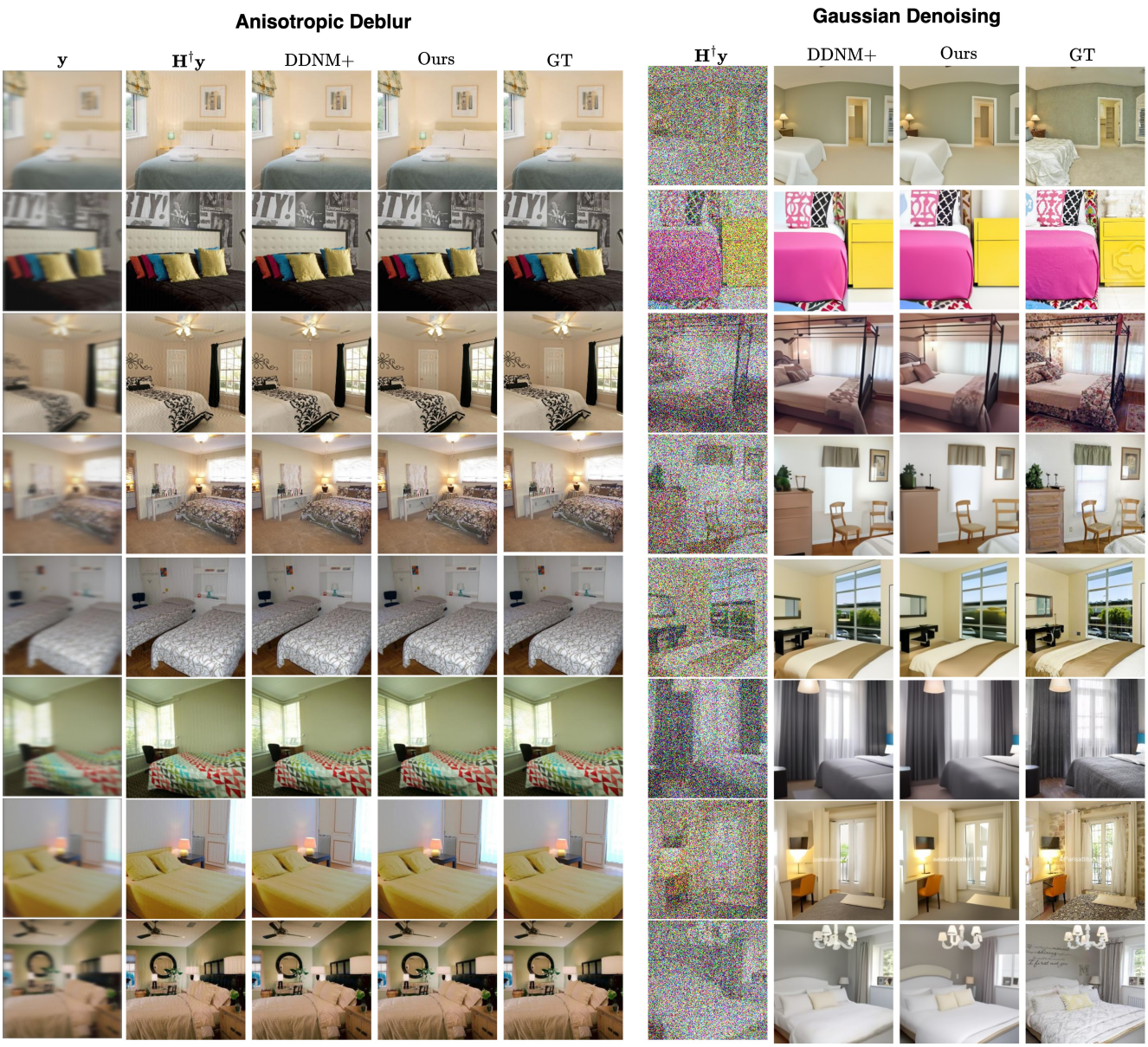

Figure E. **Visual comparison of reconstruction results for Anisotropic Deblurring and Gaussian Denoising tasks. Anisotropic Deblur (Left Panel):** Shows the degraded input image ($\mathbf{y}$), $\mathbf{H}^\dagger\mathbf{y}$, reconstructed outputs from DDNM+, the proposed method (*Ours*), and the ground truth (GT). The proposed method provides reasonable reconstructions relative to DDNM+. **Gaussian Denoising (Right Panel):** Displays similar comparisons for Gaussian noise removal. The reconstructions from the proposed method (*Ours*) yield visually cleaner outputs with fewer artifacts than DDNM+, demonstrating superior denoising capability. (Zoom in to see spot the differences)

- **ImageNet (256 x 256) :** 20% inpainting in Figure. M, Walsh-Hadamard CS for 20% compression ration in Figure. N, Gaussian deblurring in Figure. O, SR-4x in Figure. P.

G. Limitations

Applicability. Since our method builds upon DDNM [42], it is specifically designed for linear inverse problems. However, the substantial performance improvements demonstrated by our approach highlight its potential for solving a wide range of linear problems.

Quality of Reconstructions. While our method may produce slightly blurrier reconstructions for tasks like super-resolution

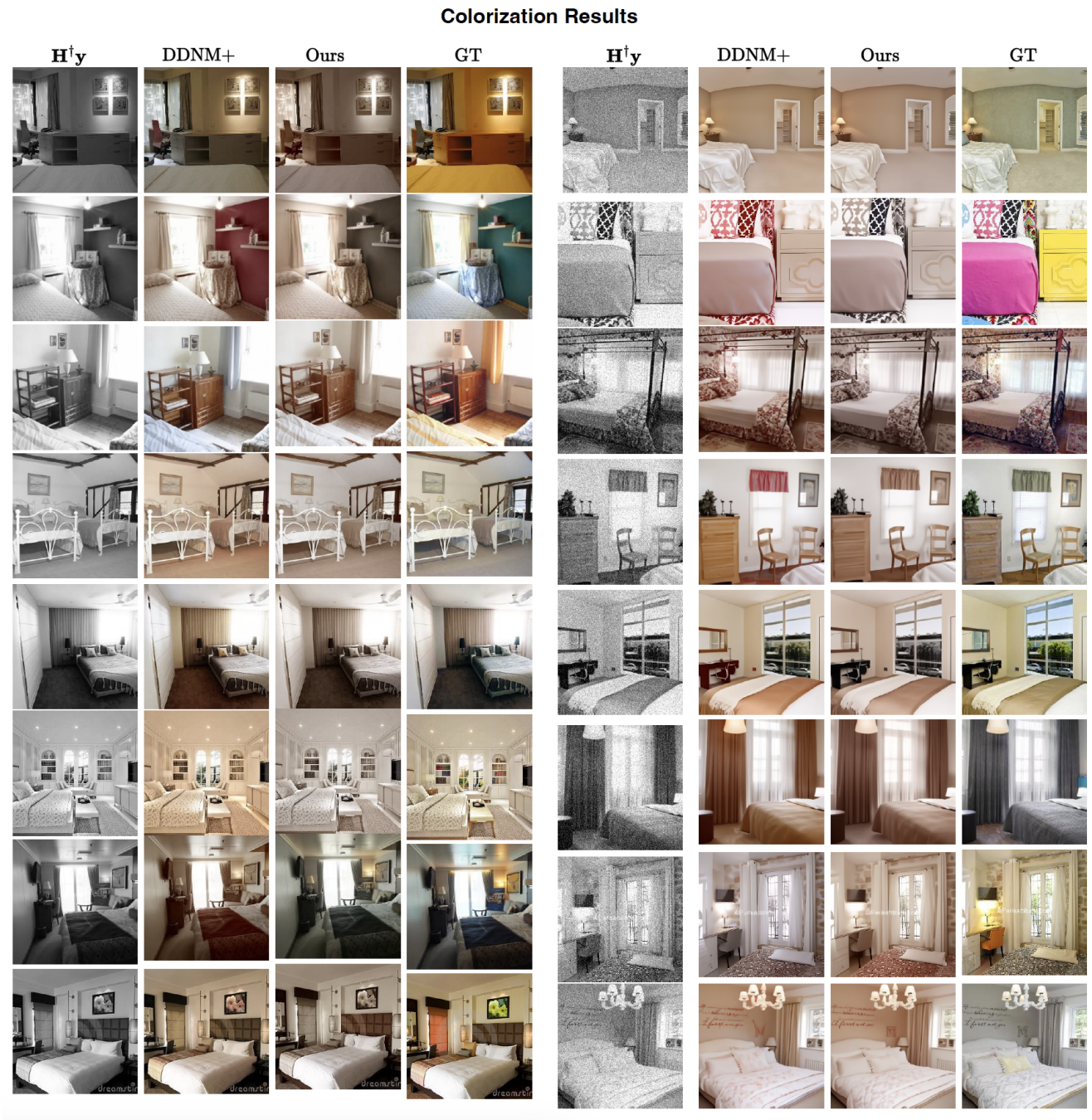

Figure F. **Visual comparison of reconstruction results for the Colorization task.** The figure demonstrates the performance of DDNM+ and the proposed method (*Ours*) compared to the ground truth (GT) across various scenes. **Colorization no noise (Left Panel)** : Colorization results under non noisy case. In some cases, DDNM+ results are more similar but in other examples, our method has reconstruction similar to GT. **Colorization with noise** $\sigma_{\mathbf{y}} = 0.1$ **(Right Panel)** : Under noisy case DDNM+ and our method provides reasonable reconstructions.

(see Figure A), it excels in inverse problems where measurements are spatially disjoint, such as compressed sensing. We attribute this blurriness to the aggregation step in our approach. Despite this, our method represents a significant step toward addressing noisy linear inverse problems by leveraging diffusion priors more effectively.

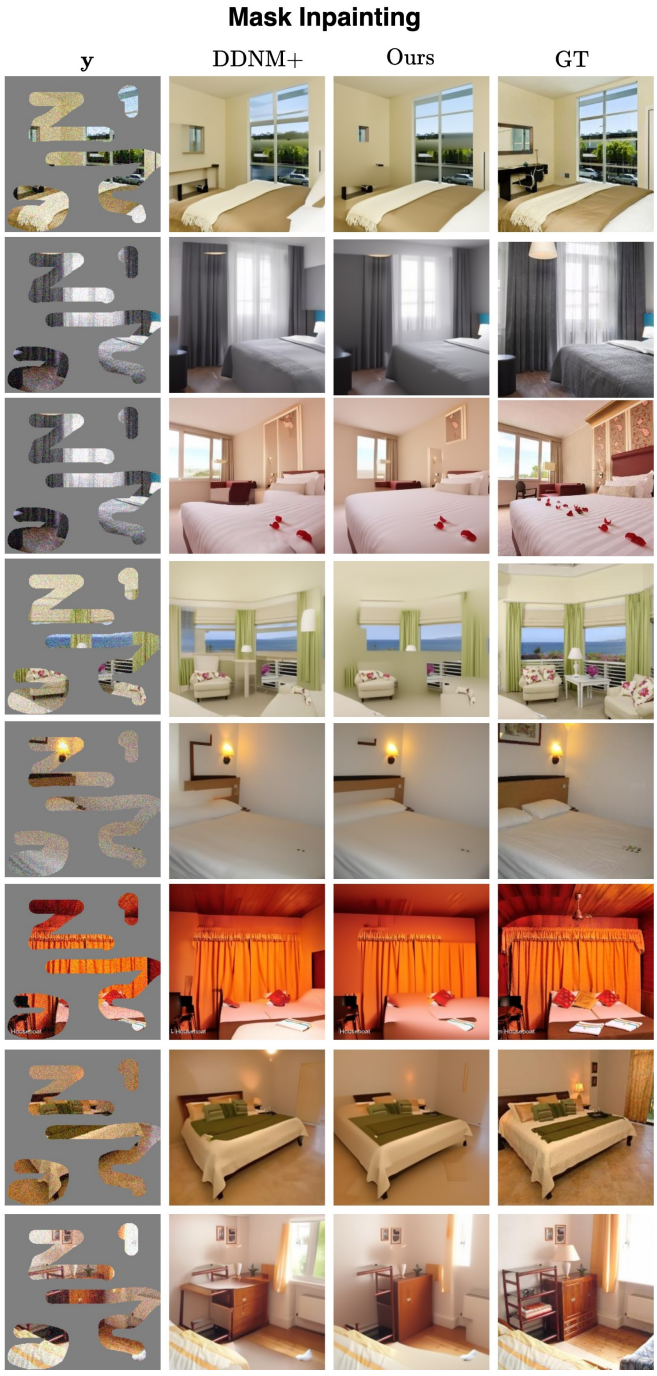

Figure G. **Visual comparison of reconstruction results for the noisy mask inpainting task with $\sigma_{\mathbf{y}} = 0.1$.** We observe that our method underperforms when challenged with extreme cases. The reconstruction may have missing features compared to DDNM+.

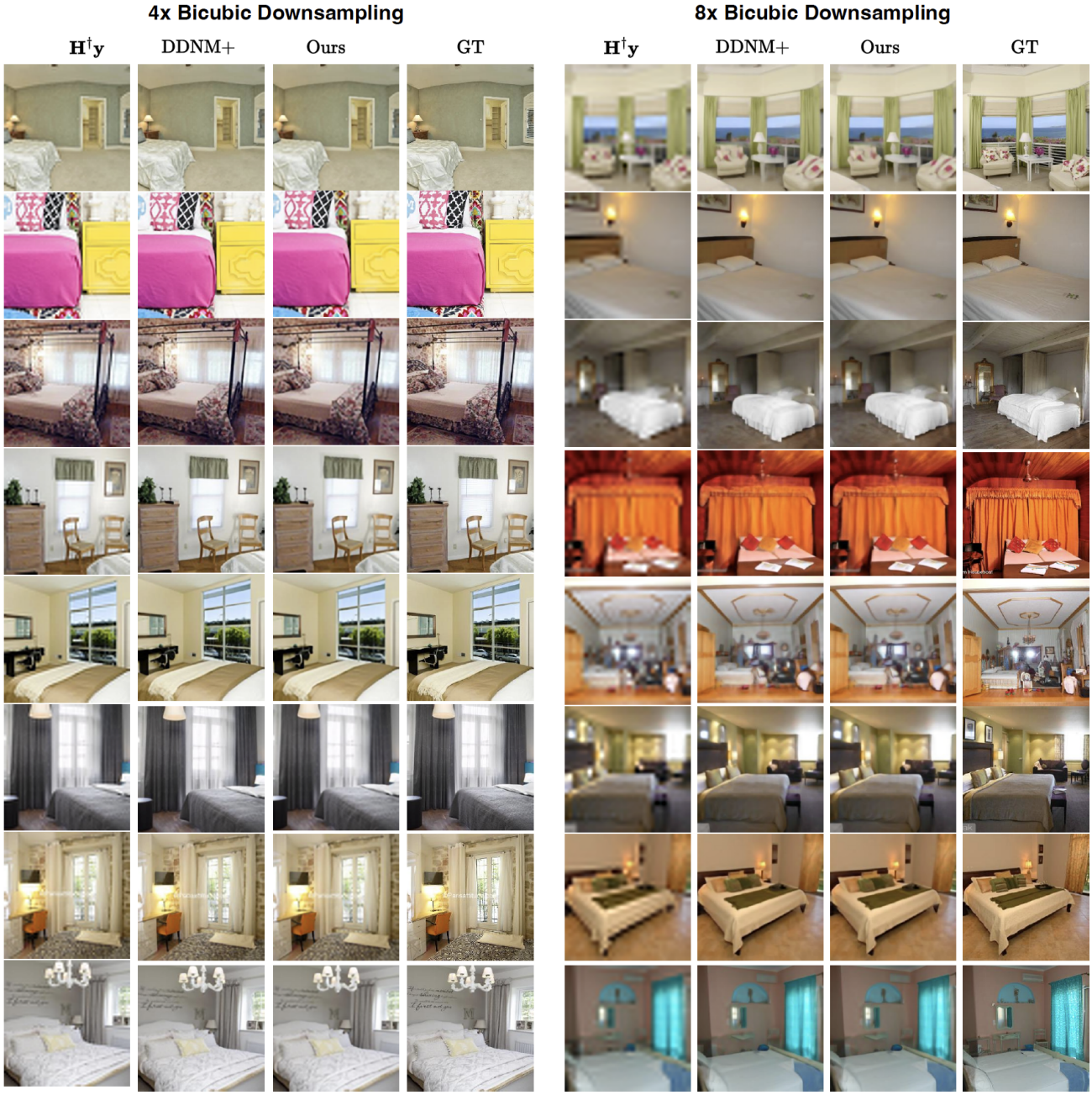

Figure H. **Visual comparison of reconstruction results for the bi-cubic downsampling (4x left panel) and (8x right panel) task.** We found the reconstruction performance for our method to be reasonable across images compared to DDNM+. However, we note that the images for both methods can be slightly blurry.

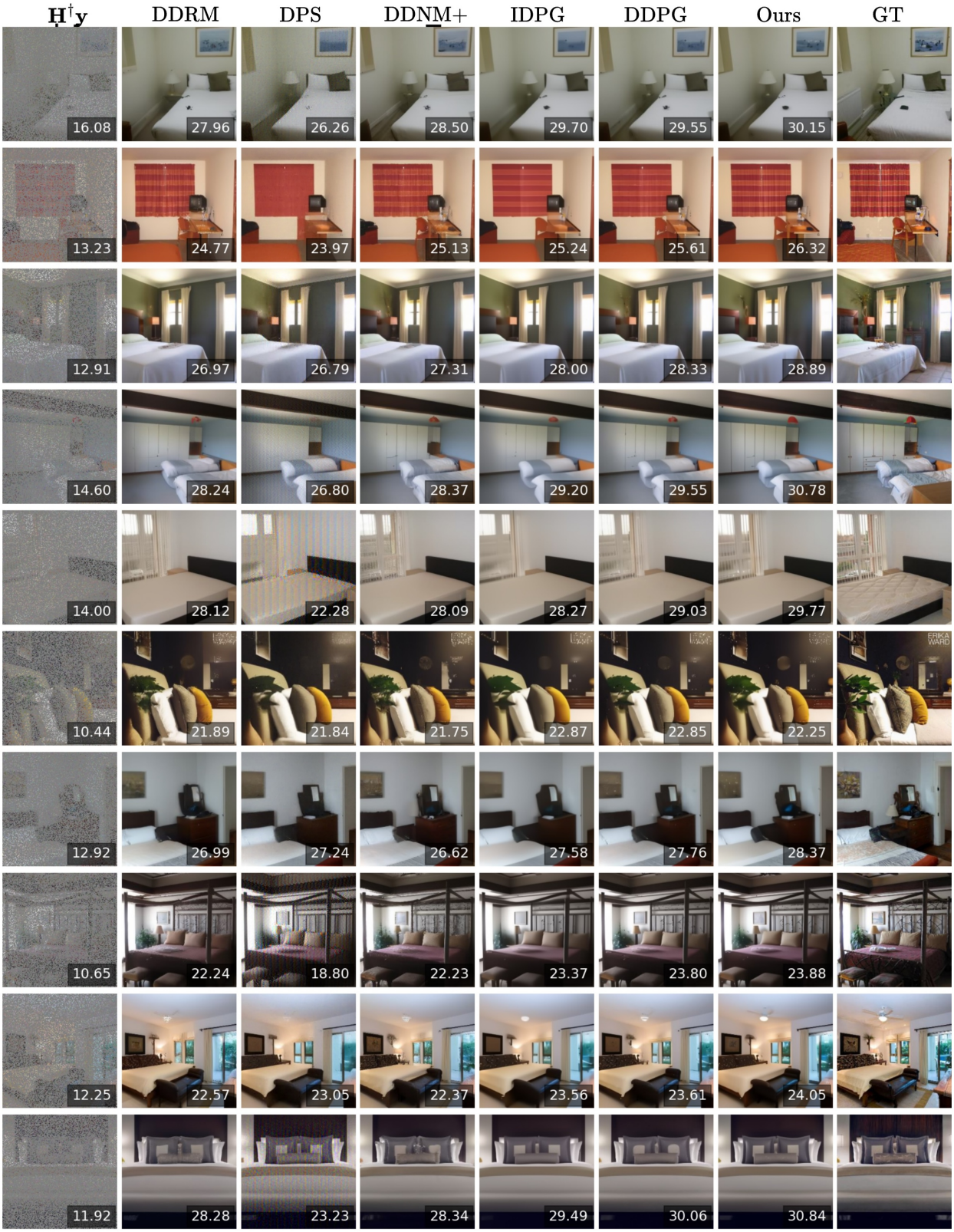

Figure I. Reconstruction results for random pixel inpainting 20% with $\sigma_{\mathbf{y}} = 0.1$ on LSUN-bedroom dataset. The PSNR values are annotated. We outperform other methods by a reasonable margin.

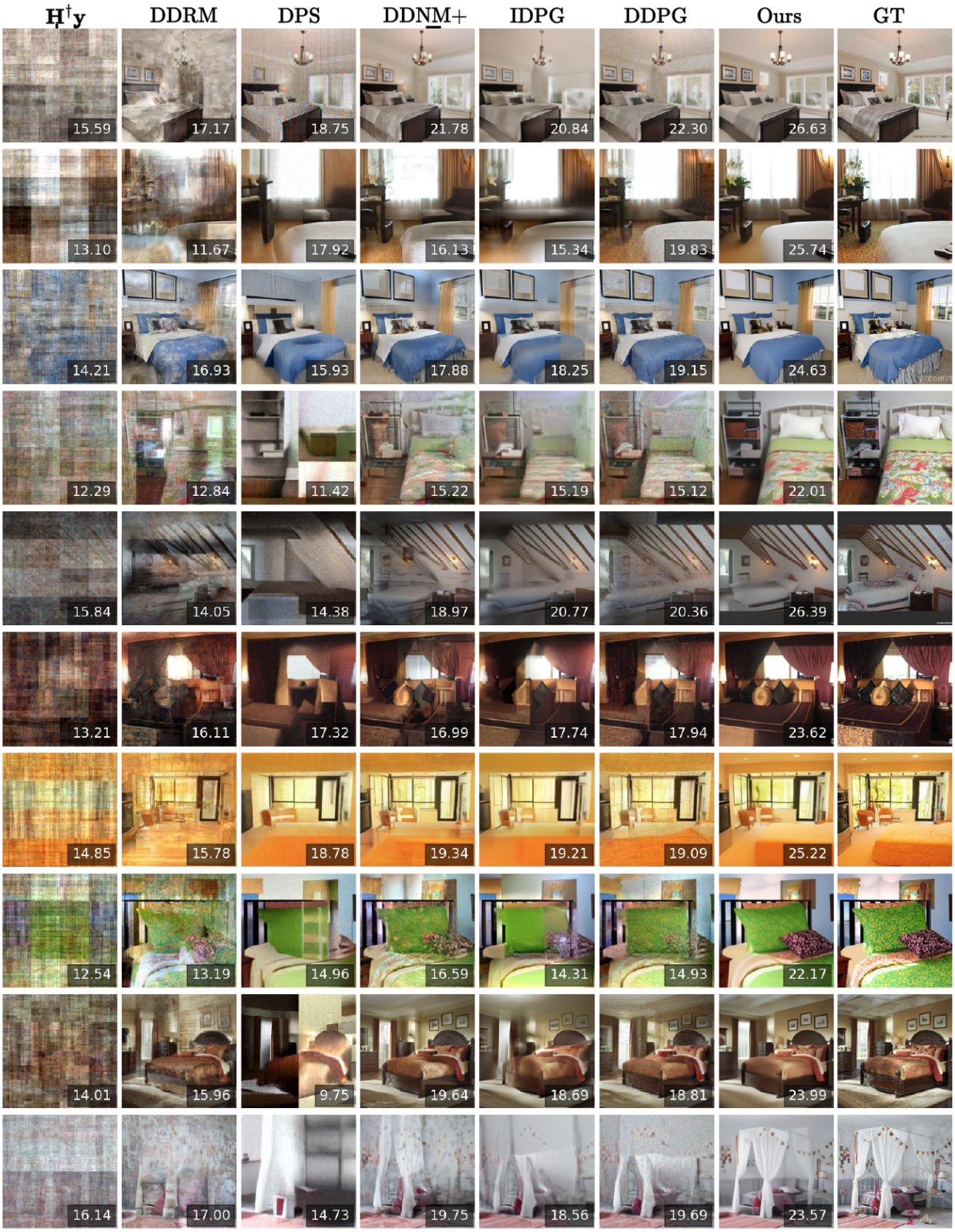

Figure J. Reconstruction results for Walsh-Hadamard compressed sensing task with a compression ratio of 20% or 0.20 with $\sigma_{\mathbf{y}} = 0.1$ on LSUN-bedroom dataset. The PSNR values are annotated. Our method comprehensively outperforms other methods.

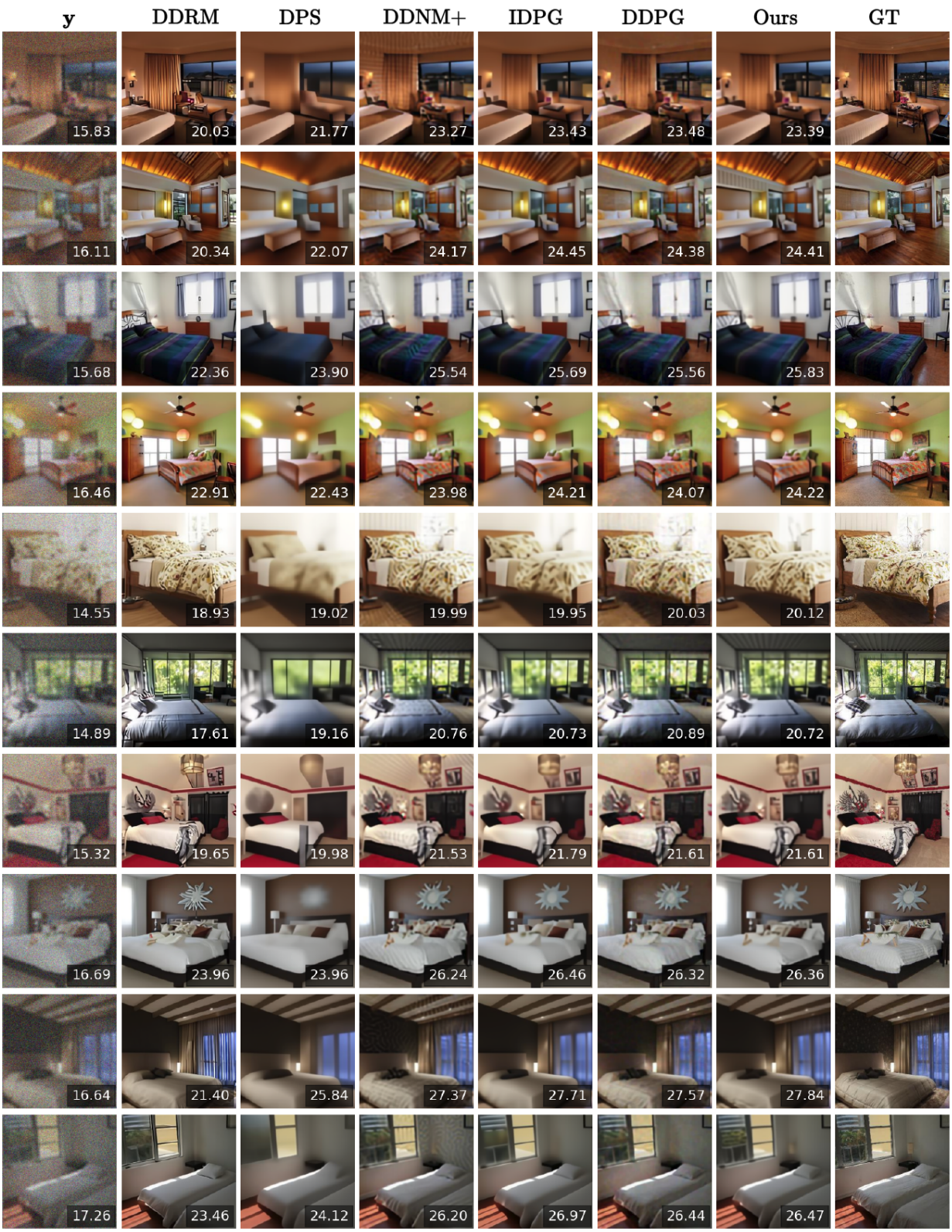

Figure K. Reconstruction results for Gaussian deblurring task with a compression $\sigma_{\mathbf{y}} = 0.1$ on LSUN-bedroom dataset. The PSNR values are annotated. Our method performs slightly better than baselines.

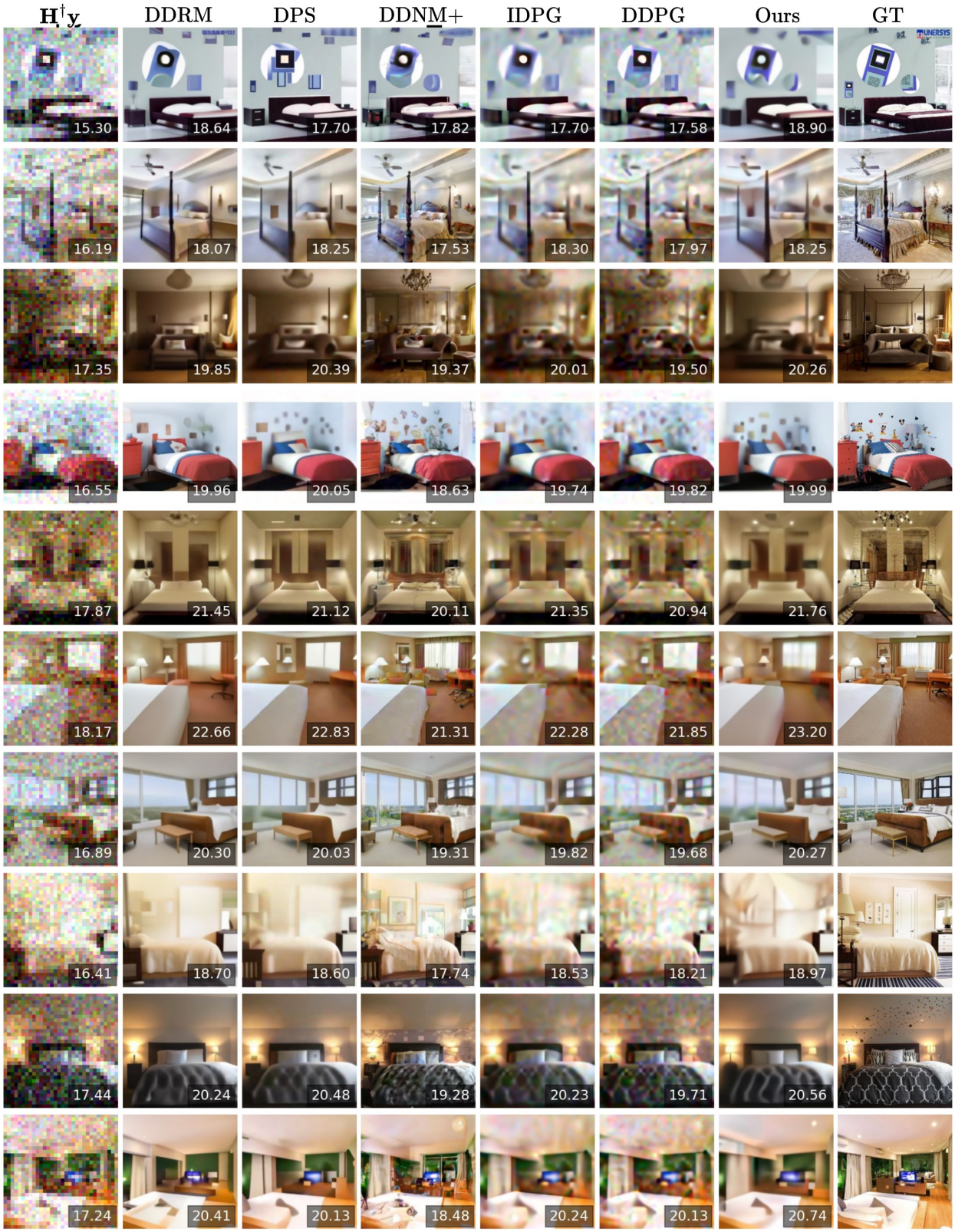

Figure L. Reconstruction results for super-resolution with 8x average pooling kernel and $\sigma_{\mathbf{y}} = 0.1$ on LSUN-bedroom dataset. The PSNR values are annotated. Our method performs slightly better than baselines.

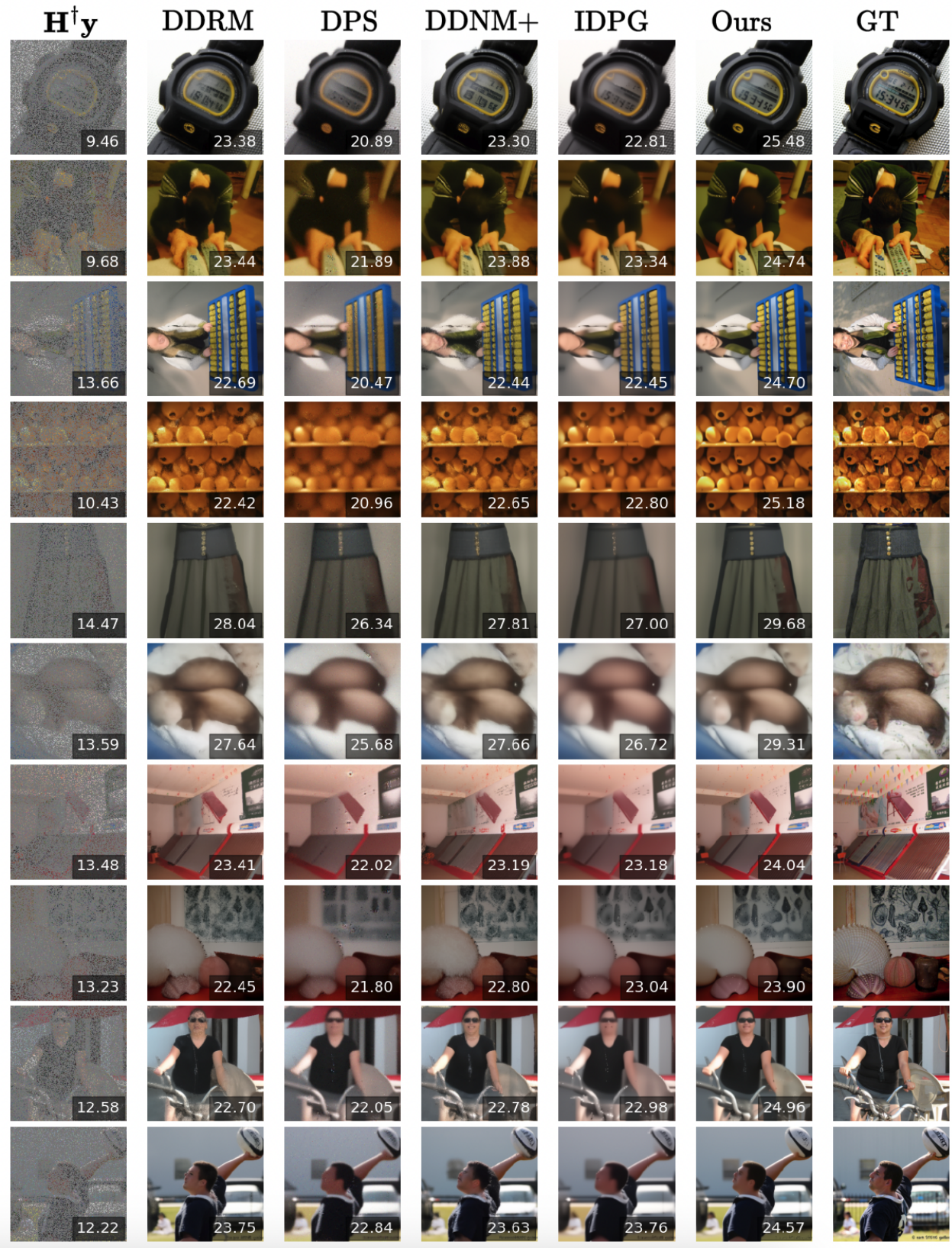

Figure M. Reconstruction results for random pixel inpainting 20% with $\sigma_{\mathbf{y}} = 0.1$ on ImageNet dataset. The PSNR values are annotated. We outperform other methods by a reasonable margin.

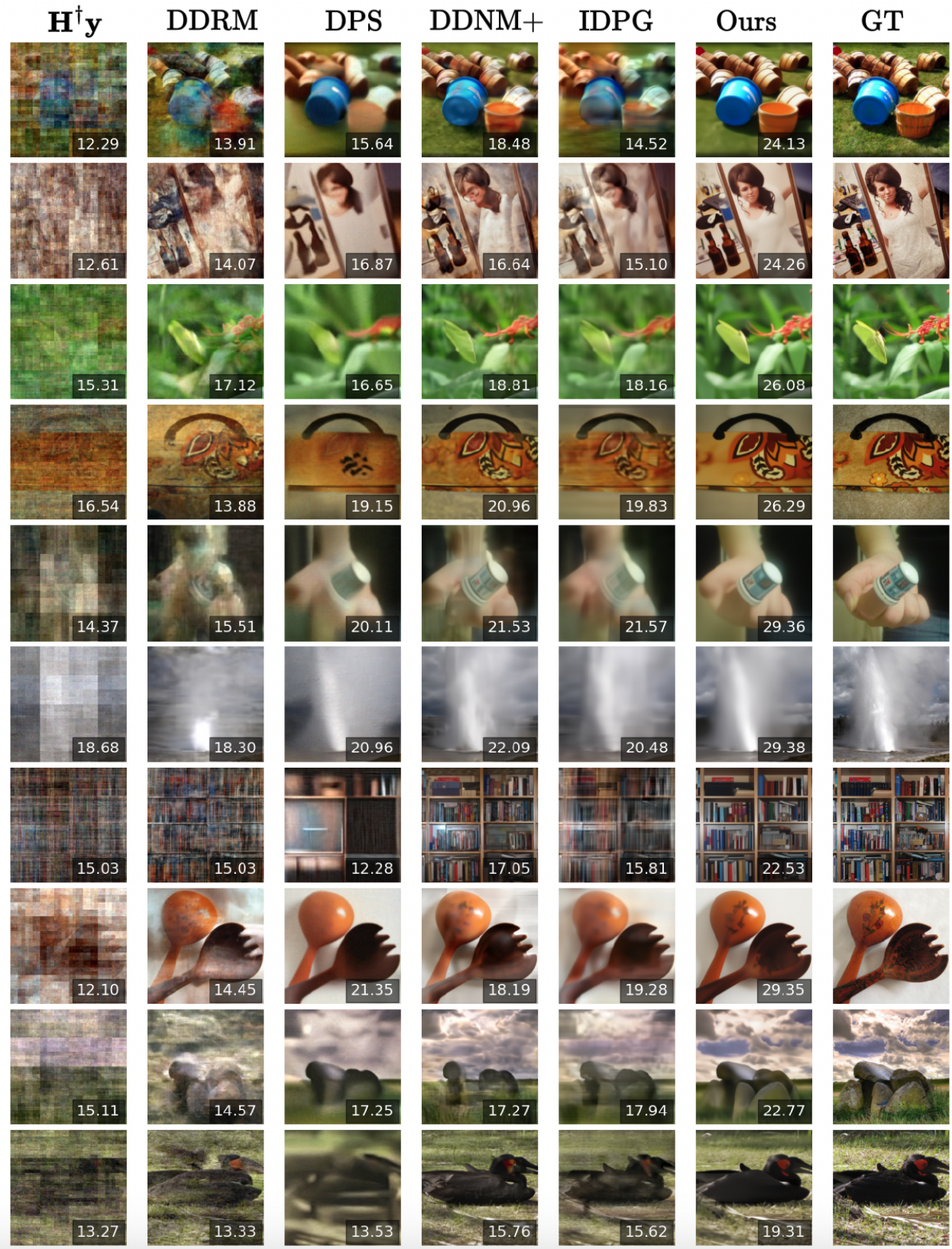

Figure N. Reconstruction results for Walsh-Hadamard compressed sensing task with a compression ratio of 20% or 0.20 with $\sigma_{\mathbf{y}} = 0.1$ on ImageNet dataset. The PSNR values are annotated. Our method comprehensively outperforms other methods.

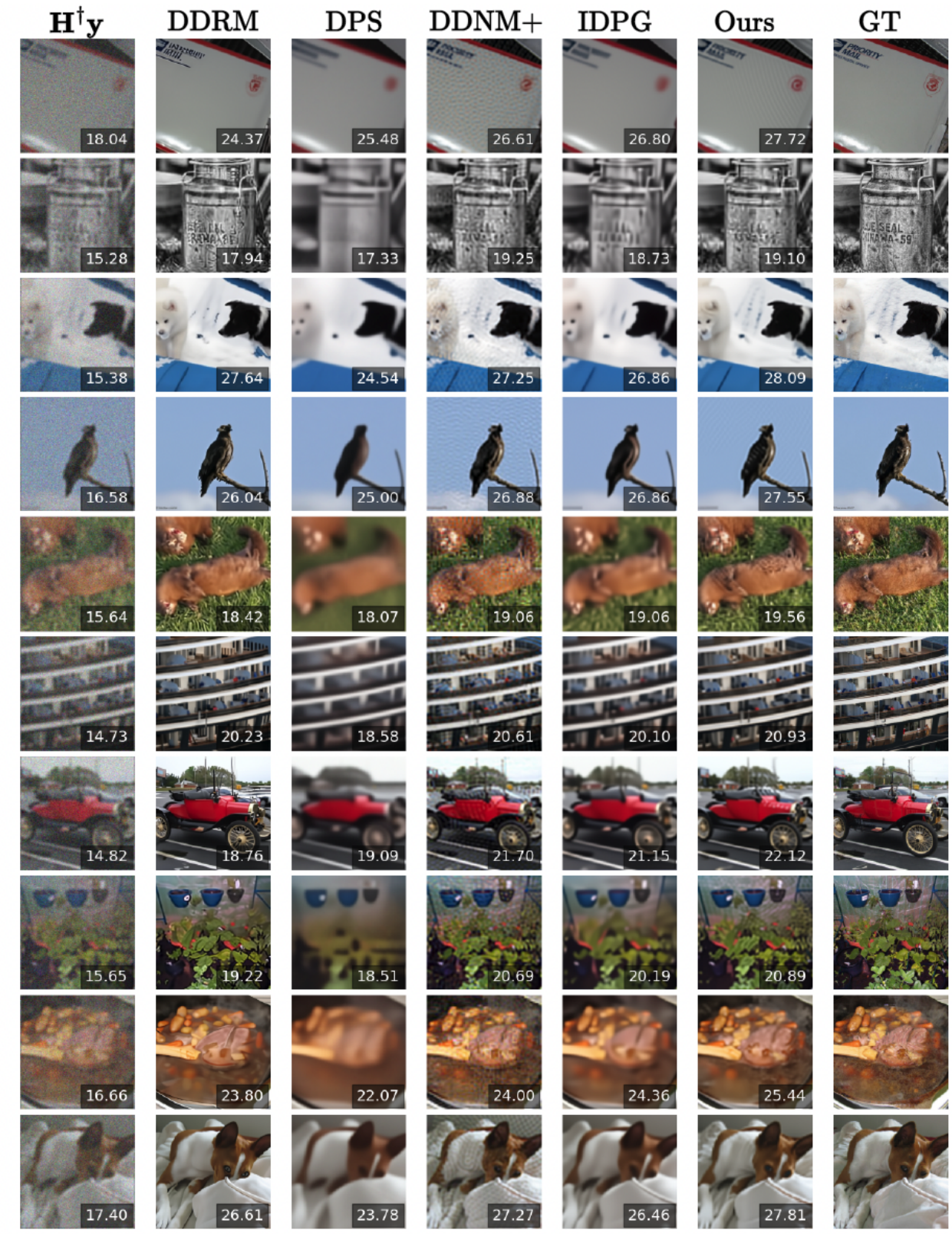

Figure O. Reconstruction results for Gaussian deblurring task with a compression $\sigma_{\mathbf{y}} = 0.1$ on ImageNet dataset. The PSNR values are annotated. Our method performs slightly better than baselines.

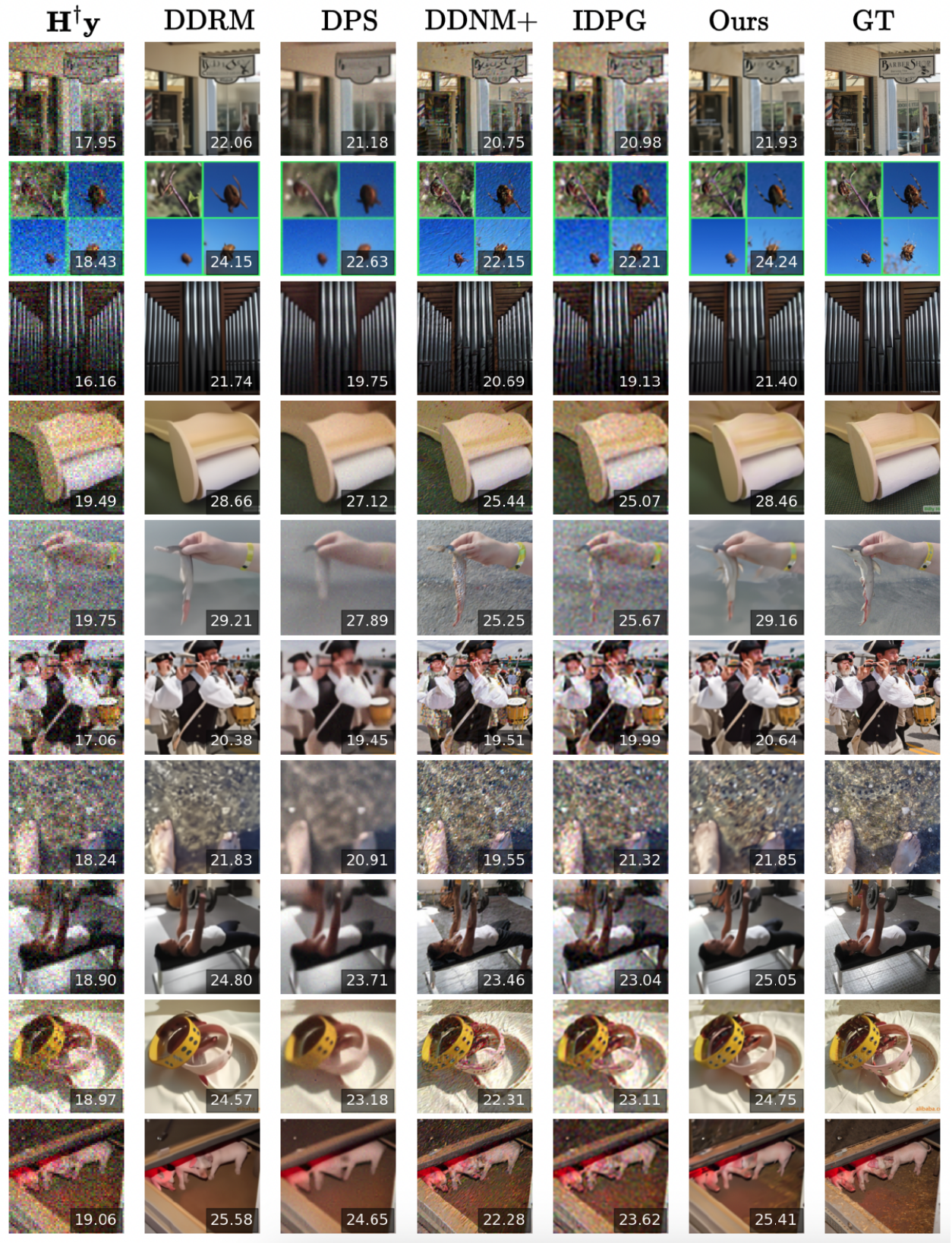

Figure P. Reconstruction results for super-resolution with 4x average pooling kernel and $\sigma_{\mathbf{y}} = 0.1$ on ImageNet dataset. The PSNR values are annotated. Our method performs slightly better than baselines.