PnP-CM: Consistency Models as Plug-and-Play Priors for Inverse Problems

Supplementary Material

A. Proof of Theorem 1

Proof. To analyze the effect of noise injection on the convergence behavior, we compare the combined residuals in the presence and absence of noise, decomposing the residual into its primal and dual components:

$$\Delta_k = \frac{1}{\sqrt{P}}\Big(\underbrace{||\mathbf{z}_{k+1}-\mathbf{z}_k||_2}_{\Delta_{\mathbf{z}_k}} + \underbrace{||\mathbf{x}_{k+1}-\mathbf{x}_k||_2}_{\Delta_{\mathbf{x}_k}} + \underbrace{||\mathbf{u}_{k+1}-\mathbf{u}_k||_2}_{\Delta_{\mathbf{u}_k}}\Big). \tag{14}$$

Since the primal variable $\mathbf{z}$ is updated prior to noise injection, it remains unaffected. We use the superscripts $\mathbf{0}$ and $\boldsymbol{\eta}$ to denote the cases without and with noise injection, respectively.

For the $\mathbf{x}$-update, we obtain

$$\begin{aligned}
\Delta^{\boldsymbol{\eta}}_{\mathbf{x}_k} &= ||\mathbf{x}_{k+1}-\mathbf{x}_k||_2 \\
&= ||f_\theta(\mathbf{z}_{k+1}+\mathbf{u}_k+\boldsymbol{\eta}_k)-\mathbf{x}_k||_2 \\
&= ||f_\theta(\mathbf{z}_{k+1}+\mathbf{u}_k+\boldsymbol{\eta}_k)-f_\theta(\mathbf{z}_{k+1}+\mathbf{u}_k) \\
&\quad + f_\theta(\mathbf{z}_{k+1}+\mathbf{u}_k)-\mathbf{x}_k||_2 \\
&\le ||f_\theta(\mathbf{z}_{k+1}+\mathbf{u}_k+\boldsymbol{\eta}_k)-f_\theta(\mathbf{z}_{k+1}+\mathbf{u}_k)||_2 \\
&\quad + ||f_\theta(\mathbf{z}_{k+1}+\mathbf{u}_k)-\mathbf{x}_k||_2 \\
&= ||f_\theta(\mathbf{z}_{k+1}+\mathbf{u}_k+\boldsymbol{\eta}_k)-f_\theta(\mathbf{z}_{k+1}+\mathbf{u}_k)||_2 + \Delta^{\mathbf{0}}_{\mathbf{x}_k} \\
&\le L||\boldsymbol{\eta}_k||_2 + \Delta^{\mathbf{0}}_{\mathbf{x}_k},
\end{aligned} \tag{15}$$

where L is the Lipschitz constant of f_θ.

Similarly, for the $\mathbf{u}$-update we have

$$\begin{aligned}
\Delta^{\boldsymbol{\eta}}_{\mathbf{u}_k} &= ||\mathbf{u}_{k+1}-\mathbf{u}_k||_2 \\
&= ||(\mathbf{u}_k+\mathbf{z}_{k+1}-\mathbf{x}_{k+1})-\mathbf{u}_k||_2 \\
&= ||\mathbf{z}_{k+1}-\mathbf{x}_{k+1}||_2 \\
&= ||\mathbf{z}_{k+1}-f_\theta(\mathbf{z}_{k+1}+\mathbf{u}_k+\boldsymbol{\eta}_k)||_2 \\
&= ||\mathbf{z}_{k+1}-f_\theta(\mathbf{z}_{k+1}+\mathbf{u}_k)+f_\theta(\mathbf{z}_{k+1}+\mathbf{u}_k) \\
&\quad - f_\theta(\mathbf{z}_{k+1}+\mathbf{u}_k+\boldsymbol{\eta}_k)||_2 \\
&\le ||\mathbf{z}_k-f_\theta(\mathbf{z}_{k+1}+\mathbf{u}_k)||_2 \\
&\quad + ||f_\theta(\mathbf{z}_{k+1}+\mathbf{u}_k)-f_\theta(\mathbf{z}_{k+1}+\mathbf{u}_k+\boldsymbol{\eta}_k)||_2 \\
&= \Delta^{\mathbf{0}}_{\mathbf{u}_k} + ||f_\theta(\mathbf{z}_{k+1}+\mathbf{u}_k)-f_\theta(\mathbf{z}_{k+1}+\mathbf{u}_k+\boldsymbol{\eta}_k)||_2 \\
&\le \Delta^{\mathbf{0}}_{\mathbf{u}_k} + L||\boldsymbol{\eta}_k||_2
\end{aligned} \tag{16}$$

Hence, the additional divergence caused by noise injection at iteration n is at most

$$\frac{2L}{\sqrt{P}}\|\boldsymbol{\eta}_k\|_2. \tag{17}$$

Formally, the total deviation introduced across all iterations is bounded by

$$\sum_k \Delta^{\boldsymbol{\eta}}_k - \sum_k \Delta^{\mathbf{0}}_k \;\le\; \frac{2L}{\sqrt{P}}\sum_k \|\boldsymbol{\eta}_k\|_2. \tag{18}$$

The convergence is preserved as long as the cumulative effect of noise injection remains finite, $\sum_{k=1}^{\infty}\|\boldsymbol{\eta}_k\|_2 < \infty$, and the noise injection is diminishing. This condition can be enforced by an appropriate noise scheduling strategy, e.g., by decaying the noise magnitude over iterations. Then the additional divergence introduced by noise injection remains bounded, ensuring that the overall convergence behavior of the algorithm is unaffected. □

B. Experimental Details

B.1. Implementation Details

We implemented our method following Alg. 1. For the JPEG restoration problem, we used the standard non-differentiable measurement, but employed a differentiable surrogate forward operator [47] to enable gradient computation. For the inpainting task, we initialized our algorithm with an additional median filter, consistent with [23]. For phase retrieval, we obtained the measurements via $\mathbf{y} = |\mathcal{F}(\mathcal{M} \odot \mathbf{x})|$, with $\mathcal{M}$ denoting the coded mask, $\mathcal{F}$ the Fourier transform, and $|\cdot|$ is entry-wise magnitude, similar to [62]. We note that phase retrieval suffers from ambiguity since magnitude-only measurements do not preserve phase information, resulting in multiple mathematically valid channel inversions. To identify the correct reconstruction, all eight inversions were evaluated, and the inversion minimizing LPIPS to a single pre-generated unconditional CM model output for the dataset was chosen. Note this does not increase the total NFE, since it is only performed at the output.

As described in Section 3.3, the data fidelity was implemented using an SVD-based approach for small-scale linear tasks (*e.g.*, super-resolution, Gaussian deblurring, inpainting). For MRI reconstruction, we employed conjugate gradient (CG), which is well suited for the large-scale linear systems arising in physics-driven reconstruction [3, 4, 17, 36, 63, 64]. For nonlinear inverse problems, first-order optimization methods were used, with iterations run until a convergence tolerance was met.

B.2. Hyperparameter Settings

Noise schedule. We adopt the noise scheduling strategy of CM4IR [23], which provides a compact parametrization requiring only a small number of hyperparameters while ensuring a stable, monotonic decrease in the injected noise across iterations. Following standard DDPM notation, the forward noising process uses a monotonically increasing variance schedule $\{\beta_i\}_{i=1}^{\mathrm{T}} \subset (0,1)$, with $\alpha_i = 1-\beta_i$,

Table 3. Complete set of hyperparameters used across all tasks in this study, including the noise schedule, ADMM penalty parameters, and momentum coefficients, for both natural image experiments and medical imaging applications.

Inverse Problems on Natural Images (CelebA-HQ [33])					
SR ×4	$i_{\mathrm{N}} = 150$	$\gamma = 0.2$	$\delta_n = (0.3, 0.05, 0.2, 0.2)$	$\rho = (-4.0, 4.0)$	$\mu_{\mathrm{N}} = 0.2$
Gaussian Deblur	$i_{\mathrm{N}} = 100$	$\gamma = 0.1$	$\delta_n = (0.3, 0.2, 0.1, 0.1)$	$\rho = (-4.0, 6.5)$	$\mu_{\mathrm{N}} = 0.1$
Inpainting	$i_{\mathrm{N}} = 100$	$\gamma = 0.1$	$\delta_n = (0.3, 0.2, 0.8, 0.8)$	$\rho = (-6.0, 12.0)$	$\mu_{\mathrm{N}} = 0.05$
JPEG Restoration	$i_{\mathrm{N}} = 250$	$\gamma = 0.3$	$\delta_n = (0.1, 0.1, 0.2, 4.0)$	$\rho = (-2.5, 2.5)$	$\mu_{\mathrm{N}} = 0.05$
Nonlinear Deblur	$i_{\mathrm{N}} = 250$	$\gamma = 0.15$	$\delta_n = (0.05, 0.05, 0.05, 0.1)$	$\rho = (-2.0, 0.0)$	$\mu_{\mathrm{N}} = 0.05$
Phase Retrieval	$i_{\mathrm{N}} = 250$	$\gamma = 0.03$	$\delta_n = (0.1, 0.1, 0.1, 0.2, 0.2, 0.2, 0.2, 0.2)$	$\rho = (-2.0, 0.0)$	$\mu_{\mathrm{N}} = 0.05$
Medical Imaging (MRI Reconstruction using fastMRI [37])					
Coranal PD – (R=4)	$i_{\mathrm{N}} = 50$	$\gamma = 0.1$	$\delta_n = (0.3, 2.0, 6.0, 2.5)$	$\rho = (-4.5, -1.0)$	$\mu_{\mathrm{N}} = 0.20$
Coranal PD – (R=8)	$i_{\mathrm{N}} = 50$	$\gamma = 0.1$	$\delta_n = (0.35, 3.5, 7.5, 3.5)$	$\rho = (-4.5, -1.5)$	$\mu_{\mathrm{N}} = 0.50$
Coranal PD-FS – (R=4)	$i_{\mathrm{N}} = 50$	$\gamma = 0.1$	$\delta_n = (0.2, 3.0, 4.0, 2.5)$	$\rho = (-2.5, -0.5)$	$\mu_{\mathrm{N}} = 0.05$
Coranal PD-FS – (R=8)	$i_{\mathrm{N}} = 50$	$\gamma = 0.1$	$\delta_n = (0.4, 9.5, 4.5, 1.5)$	$\rho = (-2.5, 0.5)$	$\mu_{\mathrm{N}} = 0.45$

$\bar{\alpha}_i = \prod_{j=1}^{i} \alpha_j$. In CM4IR, two hyperparameters are used to define the noising process through these DDPM parameters. First, an initial diffusion index $i_{\mathrm{N}} \in [1, 1000]$ selects the initial perturbation level as $t_{\mathrm{N}} = \sqrt{1 - \bar{\alpha}_{i_{\mathrm{N}}}}$. Then a decay rate γ is used to define the subsequent perturbation levels as $t_{\mathrm{N}-k} = \sqrt{1 - (1+\gamma)^k \bar{\alpha}_{i_{\mathrm{N}}}}$. Thus, a larger γ corresponds to a faster decay in the injected noise across sampling iterations, while a larger i_N increases the initial perturbation level from which the trajectory begins. We also follow the observation that CM noise level should slightly exceed the perturbation level to reflect the inherent measurement noise in the inverse problem, and use the same per-iteration offset vector $\{t_n\}_{n=1}^{\mathrm{N}} > 0$ as in CM4IR with CM noise level as $(1 + \delta_n)\, t_n$.

Penalty parameter. For ADMM, the penalty parameter must stay positive and typically increase across iterations to progressively tighten the consistency constraint between the primal and dual variables. To reduce the dimensionality of this hyperparameter space, we parametrize the full sequence using a softplus-transformed linear interpolation: $\rho_t = \mathrm{Softplus}\big(\mathrm{Linspace}(\rho_{\mathrm{start}}, \rho_{\mathrm{end}}, \mathrm{N})\big)$ This ensures positivity, provides smooth growth across iterations, and requires only two user-set endpoints $(\rho_{\mathrm{start}}, \rho_{\mathrm{end}})$.

Momentum coefficient. To improve stability and avoid oscillation, we use a simple decaying schedule for the momentum terms. The user specifies an initial coefficient μ_{N}, and subsequent coefficients are scaled down by the iteration index: $\mu_n = \mu_{\mathrm{N}}/(\mathrm{N} - n + 1)$. With this definition, the momentum gradually decreases toward the final iterations, reducing the risk of oscillatory behavior and preventing late-stage instability.

All hyperparameter values associated with our main experimental setup are listed in Tab. 3.

B.3. Comparison Methods

DPS. We followed the official implementation provided by Chung et al. [15]. Since the likelihood weight was originally tuned for ImageNet and FFHQ, we re-tuned it for LSUN Bedroom and FastMRI knee datasets, using $\eta = 1.0$ for natural images and $\eta = 0.85$ for medical imaging. For CelebA-HQ, we used the hyperparameters given for FFHQ.

ΠGDM. We relied on the official implementation reported by Song et al. [50]. For both CelebA-HQ and LSUN Bedroom, we tuned the likelihood weights for each nonlinear task to achieve the best performance, and used 100 NFEs with $\eta = 0.0$.

DiffPIR. For DiffPIR, we used the implementation provided by Zhu et al. [72]. For CelebA-HQ, we adopted the hyperparameters reported in Appendix B for the FFHQ dataset. Similarly, for LSUN Bedroom, we begin with the ImageNet hyperparameters from the same table and fine-tune them to achieve optimal performance on LSUN Bedroom.

DPnP. We followed the official implementation provided by Xu and Chi [62]. For the CelebA-HQ dataset, we initially adopted the hyperparameters used for the FFHQ dataset, as described in Appendix D, and subsequently adjusted them as necessary to achieve optimal performance.

For the LSUN Bedroom dataset, we began with the same hyperparameter settings used for CelebA-HQ and further refined them to obtain the best results.

PnP-DM. The official implementation as specified by Wu et al. [61] was employed. Hyperparameters for CelebA-HQ were kept as originally reported. For LSUN Bedroom, the noise level schedule parameter β_d was set to 14.9, which we found to achieve the best reconstruction performance. Although the codebase contains solvers for Gaussian deblurring and super-resolution, inpainting functionality was added through our own implementation, and its parameters were tuned to achieve optimal performance.

DDS. We used the official implementation of Chung et al. [16] with $\gamma = 1.0$ and $\eta = 0.85$. In contrast to clinical MRI reconstruction, where subsampled k-space measurements already contain inherent noise and no clean reference is available, the original method adds Gaussian noise to clean data. To better reflect the clinical setting, we applied DDS directly on the noisy subsampled k-space measurements. We tuned the CG steps depending on the data SNR.

CM. The official public repositories provided by Song et al. [53] were followed during the implementation. While the codebase includes iterative functions for inpainting and super-resolution, it does not provide one for Gaussian deblurring, which we implemented ourselves. We applied 40 steps for all inverse problems as advised by the authors.

CoSIGN. We used the official implementation provided by Zhao et al. [71]. Since ControlNet training checkpoints are not available for inpainting (70%) and Gaussian Deblurring tasks, we trained them from scratch. We used the provided checkpoint for super-resolution ($\times 4$).

CM4IR. For CM4IR, we relied on the official implementation and hyperparameter settings provided by Garber and Tirer [23]. Since these configurations were extensively tuned by the authors, we retained them without modification. We note that the public implementation of CM4IR contains a heuristic Tikhonov-type stabilization term in its pseudoinverse calculation for highly ill-conditioned problems, which is not discussed in [23]. Nonetheless, we incorporate this term with the suggested values.

C. Training Details of Diffusion and Consistency Models

We train the CMs on CelebA-HQ and fastMRI following the publicly available codes from [34, 53]. For both datasets, we first train an EDM-based DM [34] and then distill it into a robust CM. Notably, for fastMRI, we establish, for the first time, a CM specifically designed for MRI data.

We adopt the same U-Net hyperparameters as used for the LSUN Bedroom networks [53], except for the number of input and output channels for fastMRI. For CelebA-HQ, the DM is trained for 3.55M iterations with a batch size of 16, and subsequently distilled into a CM using LPIPS loss for 1.175M iterations with a batch size of 32. For fastMRI, the DM is trained for 1M iterations with a batch size of 4, and then distilled into a CM using ℓ_2 loss for 700k iterations with the same batch size.

Table 4. Effect of momentum across different NFEs on PD MRI (R=8). Momentum yields modest gains at higher NFEs, but provides noticeable improvements in the low-NFE regime.

Noise Inj.	Moment.	NFE=4	NFE=10	NFE=20
✓	✗	30.70 / 0.845	32.81 / 0.868	32.86 / 0.873
✓	✓	32.81 / 0.870	32.87 / 0.872	32.81 / 0.879

D. Further Experimental Results

D.1. Effect of Momentum Across NFEs

We study the effect of momentum across different numbers of NFEs. As shown in Table 4, the gains from momentum are modest at higher NFEs, which is consistent with standard ADMM behavior where convergence is already well-established. However, in the low-NFE regime considered in this work, momentum provides a noticeable improvement in reconstruction quality. This suggests that while momentum is not the primary driver of performance, it plays an important role in accelerating convergence when only a small number of iterations is used.

D.2. Hyperparameter Sensitivity Analysis

To assess the robustness of our chosen hyperparameters, we conducted a perturbation study on the Gaussian deblurring task using the CelebA-HQ dataset. Focusing on the case $\mathrm{N} = 4$, each of the nine hyperparameters given in Tab. 3 was independently perturbed by up to $\pm 25\%$, applying multiplicative factors drawn from a uniform distribution. For every image in the dataset, a new perturbed hyperparameter set was sampled and the full reconstruction process was rerun. We then compared the reconstruction quality against the baseline configuration.

Fig. 5 summarizes the effect of increasing perturbation magnitude on performance by reporting the *absolute* changes. As the perturbation level grows, both metrics degrade gradually, with PSNR decreasing and LPIPS increasing in a smooth and monotonic fashion. Importantly, the magnitude of these changes remains small: even at 25% hyperparameter perturbation, the performance drop is modest and the algorithm continues to produce stable and visually plausible reconstructions.

Table 5. Quantitative results on CelebA-HQ for different measurement noise levels (σ_y) and varying numbers of NFEs across super resolution ($\times 4$), Gaussian deblurring, and inpainting tasks.

NFE↓	σ_y	SR $\times$4	Gaussian Deblurring	Inpainting
		PSNR↑ / LPIPS↓	PSNR↑ / LPIPS↓	PSNR↑ / LPIPS↓
2	0.025	27.44 / 0.272	29.22 / 0.251	28.77 / 0.213
	0.05	26.99 / 0.281	28.78 / 0.256	28.91 / 0.209
4	0.025	27.78 / 0.265	29.57 / 0.238	29.14 / 0.206
	0.05	27.27 / 0.285	28.94 / 0.249	29.23 / 0.201
8	0.025	27.91 / 0.264	29.71 / 0.237	29.23 / 0.205
	0.05	27.37 / 0.286	29.05 / 0.250	29.32 / 0.199

D.3. Performance Across Different Noise Levels and Sampling Steps

Tab. 5 reports extended quantitative results for varying noise levels and sampling steps on CelebA-HQ. PnP-CM maintains consistently strong performance across different inverse tasks, showing robustness to both the number of function evaluations and measurement noise. Increasing the sampling steps generally improves quality as expected, while performance remains competitive even in the few-step regime.

D.4. Additional Qualitative Results for Natural Image Datasets

Alongside the results presented in the main paper, we include further qualitative comparisons on CelebA-HQ for two representative inverse problems: Gaussian deblurring and $4\times$ super-resolution. As shown in Figs. 6, 7, and 8, PnP-CM consistently reconstructs sharper facial details and more realistic high-frequency textures compared to all baseline methods.

For LSUN Bedroom, we also provide further qualitative results against all baseline methods discussed in the main text in Figs. 9, 10, and 11, illustrating that the visual trends observed there hold across a wider set of examples.

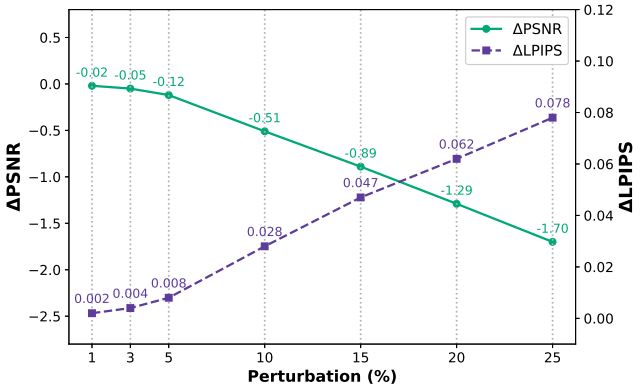

Figure 5. Absolute change in PSNR and LPIPS when perturbing the reported PnP-CM hyperparameters. Even with up to 25% variation, the deviations remain small, indicating strong robustness.

D.5. Additional Qualitative Results on MRI Reconstruction

We provide additional qualitative reconstruction results for coronal PD and PD-FS datasets at acceleration factors $R = 4$ and $R = 8$, with representative examples from each case shown in Fig. 12.

Figure 6. Representative Gaussian deblurring results on CelebA-HQ with $\sigma_y = 0.05$. Unlike competing approaches, PnP-CM better preserves fine-scale facial structures and recovers sharper, more realistic textures.

Figure 7. Demonstration of inpainting results on CelebA-HQ with $\sigma_y = 0.05$. PnP-CM reconstructs the missing regions with coherent facial geometry and realistic textures, producing completions that align more closely with the ground truth than competing methods.

Figure 8. Illustrative $4\times$ super-resolution results on CelebA-HQ with $\sigma_y = 0.05$. PnP-CM restores high-frequency details with improved clarity and denoising, yielding visually richer reconstructions.

Figure 9. Representative Gaussian deblurring results on LSUN Bedroom with $\sigma_y = 0.05$. Comparisons with all baseline methods show that PnP-CM restores textures more faithfully and avoids oversmoothing.

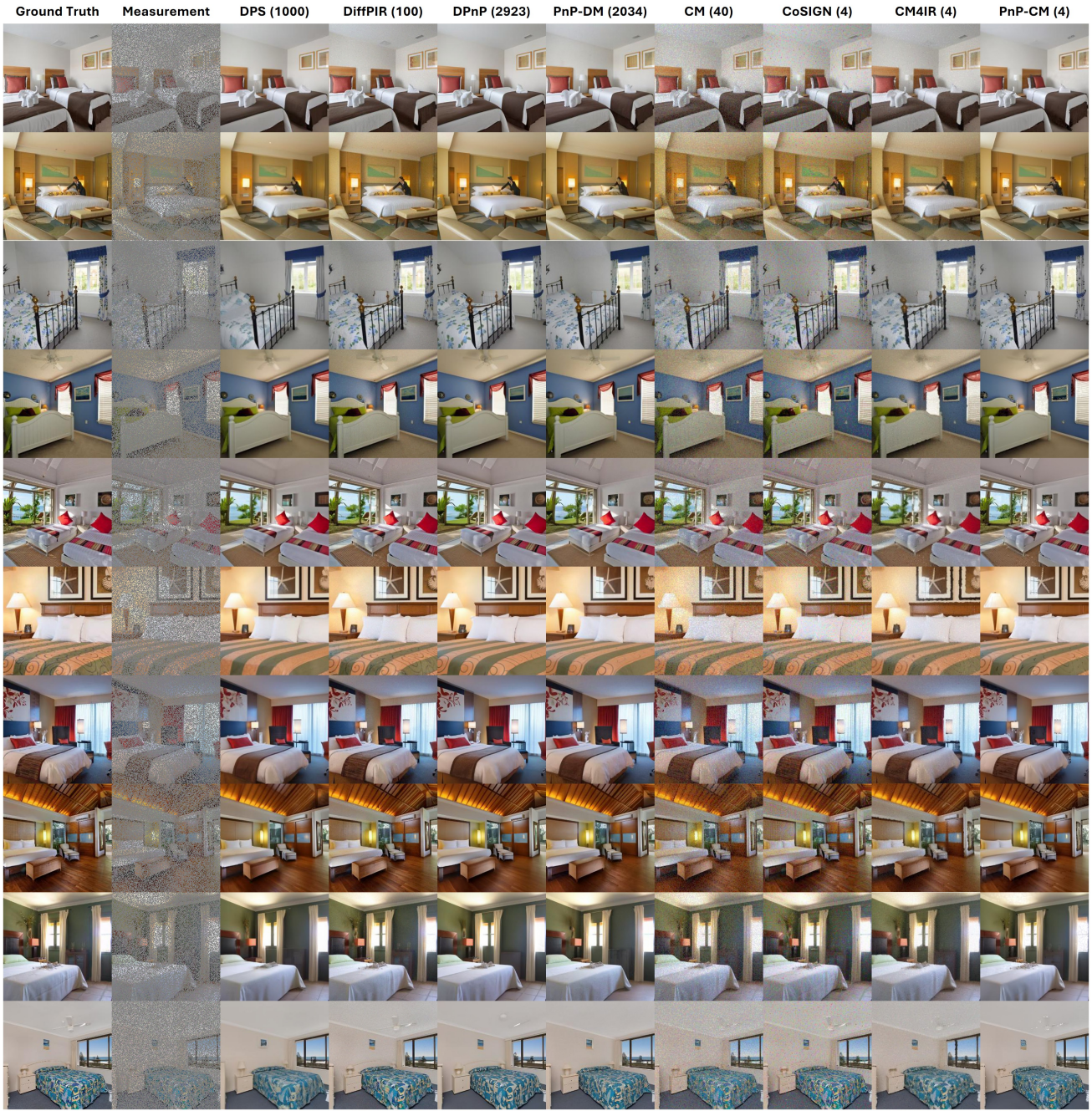

Figure 10. Illustrative inpainting results on LSUN Bedroom with $\sigma_y = 0.05$. Compared to other methods, PnP-CM fills missing regions with coherent structures and yields reconstructions closer to the ground truth.

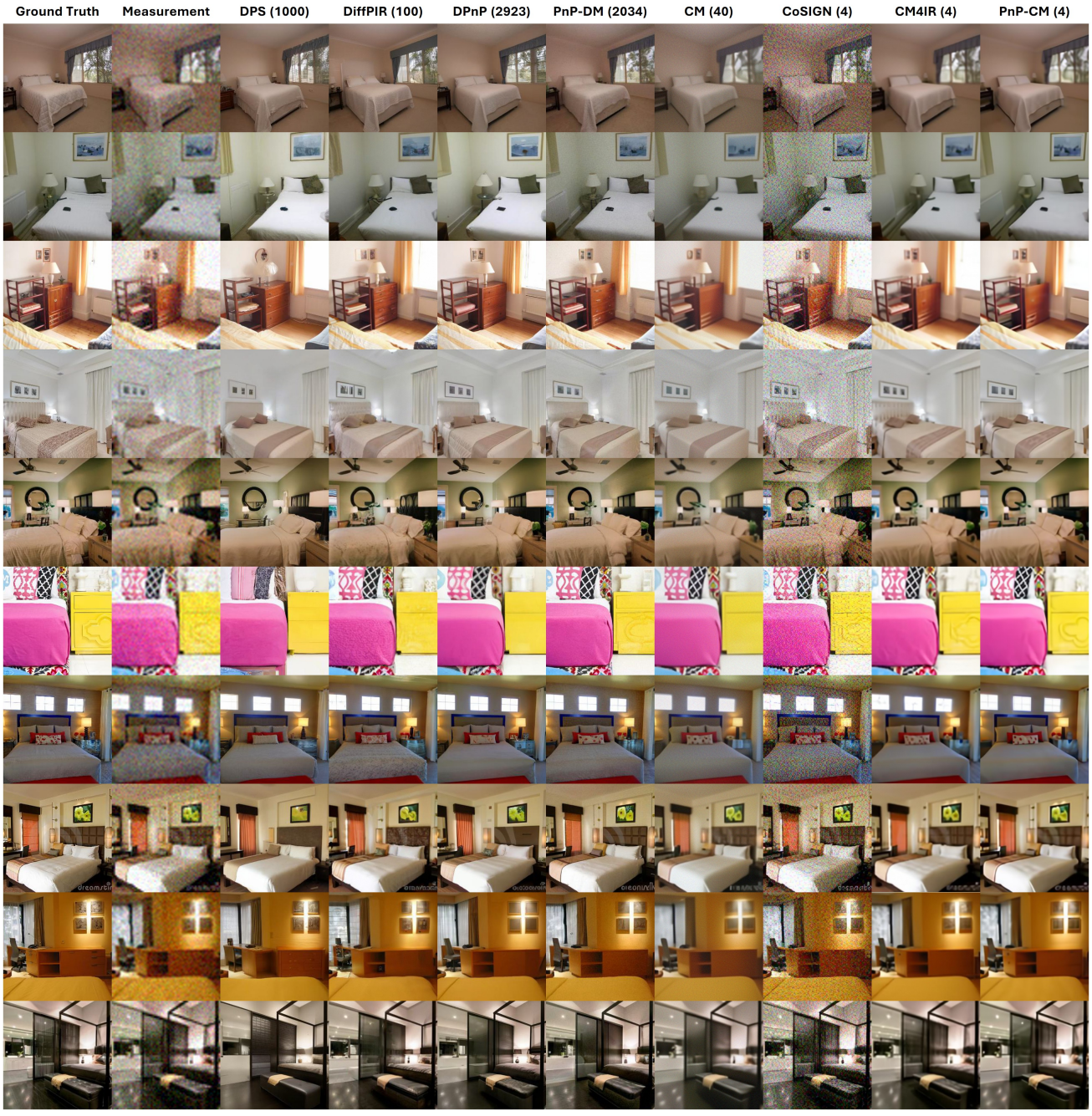

Figure 11. Demonstration of super-resolution ($\times 4$) results on LSUN Bedroom with $\sigma_y = 0.05$. Reconstructions are compared against all baseline methods, with PnP-CM producing sharper details and closer resemblance to the ground truth.

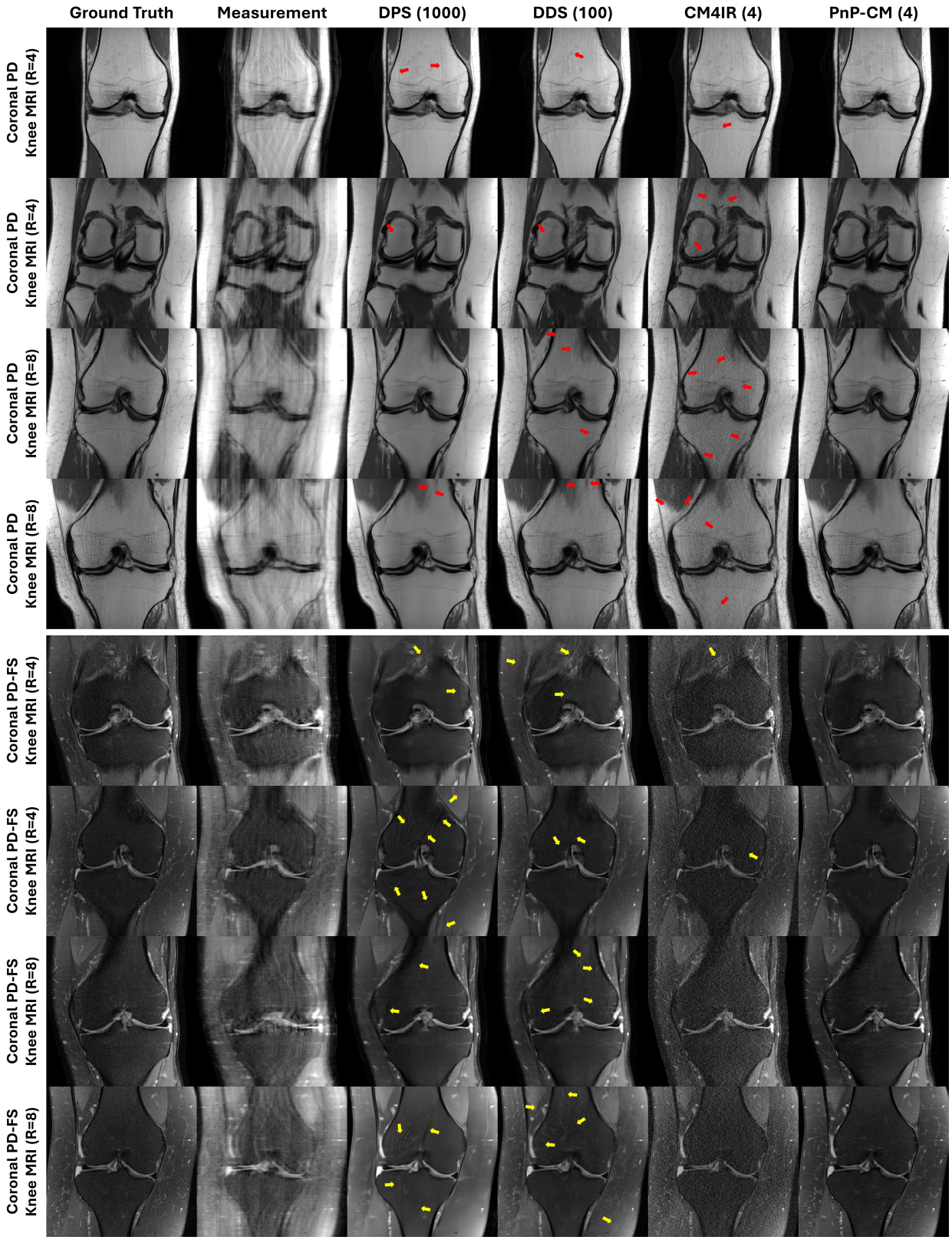

Figure 12. Qualitative comparisons for $R = 4$ and $R = 8$ on the Coronal PD and Coronal PD-FS datasets across different methods. The proposed method, PnP-CM, consistently demonstrates superior performance by effectively reducing artifacts.