Consistency Posterior Sampling for Diverse Image Synthesis

Vishal Purohit*, Matthew Repasky*, Jianfeng Lu, Qiang Qiu, Yao Xie, Xiuyuan Cheng

*Equal Contribution

Bayesian Inverse Problems

Bayesian inverse problems are concerned with noisy measurements:

$$\mathbf{y} = \mathcal{A}(\mathbf{x}) + \boldsymbol{\epsilon}$$

$\mathbf{x}$

$\mathbf{y} \sim p(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathcal{A}(\mathbf{x}), \sigma^2\mathbf{I})$

$\hat{\mathbf{x}} \sim p(\mathbf{x}|\mathbf{y})$

$\mathbf{y}$

$\mathcal{A}$: Forward Model
ϵ: Noise $\sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$

Solutions can be provided via posterior sampling.

Diffusion Posterior Sampling

Existing methods guide diffusion step using conditional information:

p_T $\rightarrow$ $$d\mathbf{x}_t = \left[\mathbf{f}_t(\mathbf{x}_t) - g_t^2 \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t|\mathbf{y})\right] dt + g_t d\bar{\mathbf{w}}_t$$ p_0 $\rightarrow$

Score model $\rightarrow$ $\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t) + \nabla_{\mathbf{x}_t} \log p_t(\mathbf{y}|\mathbf{x}_t)$ $\leftarrow$ **Intractable, must be approximated**

Other issues:

- Targets point estimate (MAP) solution $\mathbf{x}^\star$ rather than sampling posterior.
- Each sample requires simulation of full Langevin dynamics.

“Diffusion Posterior Sampling for General Noisy Inverse Problems,” Chung et al., ICLR 2023.

Sampling from Generative Models

Generally, samplers of generative models define a pushforward:

$$p_0 \approx \tilde{p}_0 = \Phi_{\#} p_1,$$

e.g., p_1 is typically a simple Gaussian $\mathcal{N}(\mathbf{0}, \eta^2 \mathbf{I})$.

Hence, $\mathbf{x}_1 \sim p_1$ can be denoted *noise space* samples corresponding to *data space* samples $\mathbf{x}_0 = \Phi(\mathbf{x}_1) \sim \tilde{p}_0$.

Model Posterior

We aim to sample from the **true posterior** of $\mathbf{x}_0$ given $\mathbf{y}$:

$$p_{0,\mathbf{y}}(\mathbf{x}_0) \coloneqq p(\mathbf{x}_0|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x}_0)p_0(\mathbf{x}_0).$$

Given $\mathbf{x}_0 = \Phi(\mathbf{x}_1)$, this corresponds to the **model posterior**:

$$p_{0,\mathbf{y}}(\mathbf{x}_0) \approx \tilde{p}_{0,\mathbf{y}}(\mathbf{x}_0) \propto p(\mathbf{y}|\mathbf{x}_0)\Phi_{\#}p_1(\mathbf{x}_0).$$

Finally, we arrive at the **noise space posterior**:

$$\tilde{p}_{1,\mathbf{y}}(\mathbf{x}_1) \propto p(\mathbf{y}|\Phi(\mathbf{x}_1))p_1(\mathbf{x}_1).$$

Method

Sample $\mathbf{x}_1 \sim \tilde{p}_{1,\mathbf{y}}$ and map to data space $\mathbf{x}_0 = \Phi(\mathbf{x}_1)$.

Langevin Dynamics

Langevin dynamics are defined by the following general SDE:

$$\mathrm{d}\mathbf{x} = \nabla_{\mathbf{x}} \log p_{\infty}(\mathbf{x})\mathrm{d}t + \sqrt{2}\mathrm{d}\mathbf{w}_t,$$

where $\mathbf{x} \sim p_{\infty}$ after long enough simulation.

To sample the noise-space posterior, we require its score:

$$\nabla_{\mathbf{x}_1} \log \tilde{p}_{1,\mathbf{y}}(\mathbf{x}_1) = \underbrace{\nabla_{\mathbf{x}_1} \log p\left(\mathbf{y}|\Phi(\mathbf{x}_1)\right)}_{\substack{= -\frac{1}{2\sigma^2}\nabla_{\mathbf{x}_1}\left\|\mathbf{y} - \mathcal{A}\left(\Phi(\mathbf{x}_1)\right)\right\|_2^2 \\ \text{for } p(\mathbf{y}|\mathbf{x}_0) = \mathcal{N}(\mathcal{A}(\mathbf{x}_0), \sigma^2\mathbf{I})}} + \underbrace{\nabla_{\mathbf{x}_1} \log p_1(\mathbf{x}_1)}_{\substack{= -\mathbf{x}_1 \\ \text{for } p_1 = \mathcal{N}(\mathbf{0}, \mathbf{I})}}$$

Langevin Dynamics for Posterior Sampling

Hence, the following SDE simulates $\tilde{p}_{1,\mathbf{y}}$:

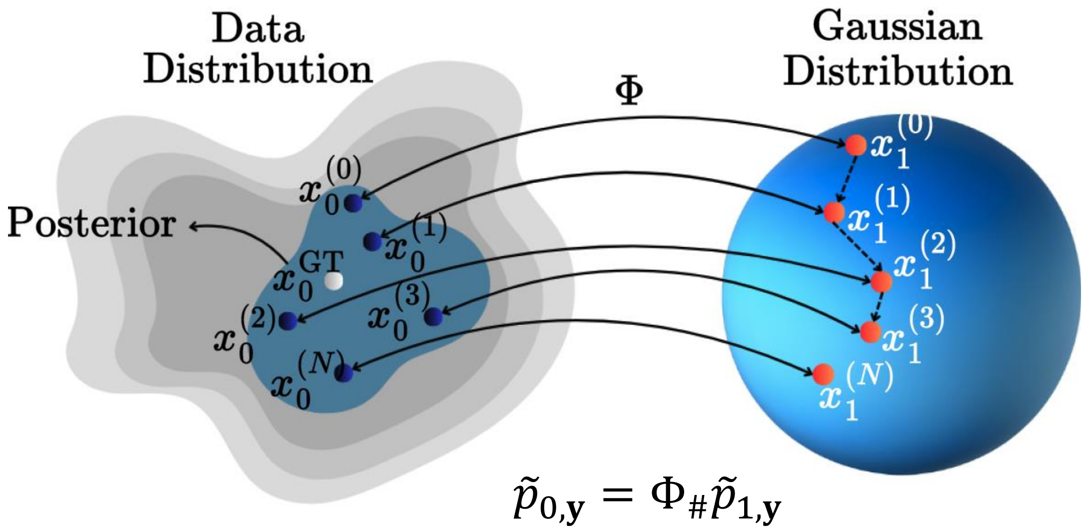
$$\mathrm{d}\mathbf{x}_1 = -\left(\mathbf{x}_1 + \frac{1}{2\sigma^2}\nabla_{\mathbf{x}_1}\left\|\mathbf{y} - \mathcal{A}\big(\Phi(\mathbf{x}_1)\big)\right\|_2^2\right)\mathrm{d}t + \sqrt{2}\,\mathrm{d}\mathbf{w}_t.$$

The dynamics can be simulated using Euler-Maruyama (EM):

$$\mathbf{x}_1^{i+1} = (1-\tau)\mathbf{x}_1^i - \tau\mathbf{g}^i + \sqrt{2\tau}\boldsymbol{\xi}^i.$$

$\mathbf{g}^i \coloneqq \frac{1}{2\sigma^2}\nabla_{\mathbf{x}_1^i}\left\|\mathbf{y} - \mathcal{A}\left(\Phi\left(\mathbf{x}_1^i\right)\right)\right\|_2^2$

$\boldsymbol{\xi}^i \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$



$$\tilde{p}_{0,\mathbf{y}} = \Phi_{\#}\tilde{p}_{1,\mathbf{y}}$$

Noise Space Sampling Algorithm

Each step requires the gradient of Φ.

- Few-step samplers like **consistency models** can alleviate this burden.

Convergence may require many steps.

- Conduct **optimization warm start** (Adam to minimize $\|\mathbf{y} - \mathcal{A}(\mathbf{x})\|_2^2$).

Produce N posterior samples via Langevin dynamics.

- Balance stability with sampling speed via step size τ.

<u>Posterior Sampling in Noise Space</u>

Inputs: $\mathbf{y}, \mathcal{A}, \sigma, \Phi, N, \tau, \text{WarmStart}(\cdot)$

$\mathbf{z_1} \leftarrow \text{WarmStart}(\cdot)$

$\mathbf{x}_1^{(0)} \leftarrow \mathbf{z}_1$

for $i = 0, \dots, N-1$ **do**

$\mathbf{x}_0^{(i)} \leftarrow \Phi\left(\mathbf{x}_1^{(i)}\right)$

$\mathbf{g}^{(i)} \leftarrow (2\sigma^2)^{-1} \nabla_{\mathbf{x}_1^{(i)}} \left\|\mathbf{y} - \mathcal{A}\left(\mathbf{x}_0^{(i)}\right)\right\|$

$\boldsymbol{\xi}^{(i)} \leftarrow \mathcal{N}(\mathbf{0}, \mathbf{I})$

$\mathbf{x}_1^{(i+1)} \leftarrow \mathbf{x}_1^{(i)} - \tau\left(\mathbf{x}_1^{(i)} + \mathbf{g}^{(i)}\right) + \sqrt{2\tau}\boldsymbol{\xi}^{(i)}$

end for

return $\mathbf{x}_0^1, \ \mathbf{x}_0^2, \dots, \mathbf{x}_0^N$

Image Inverse Problem Experiments

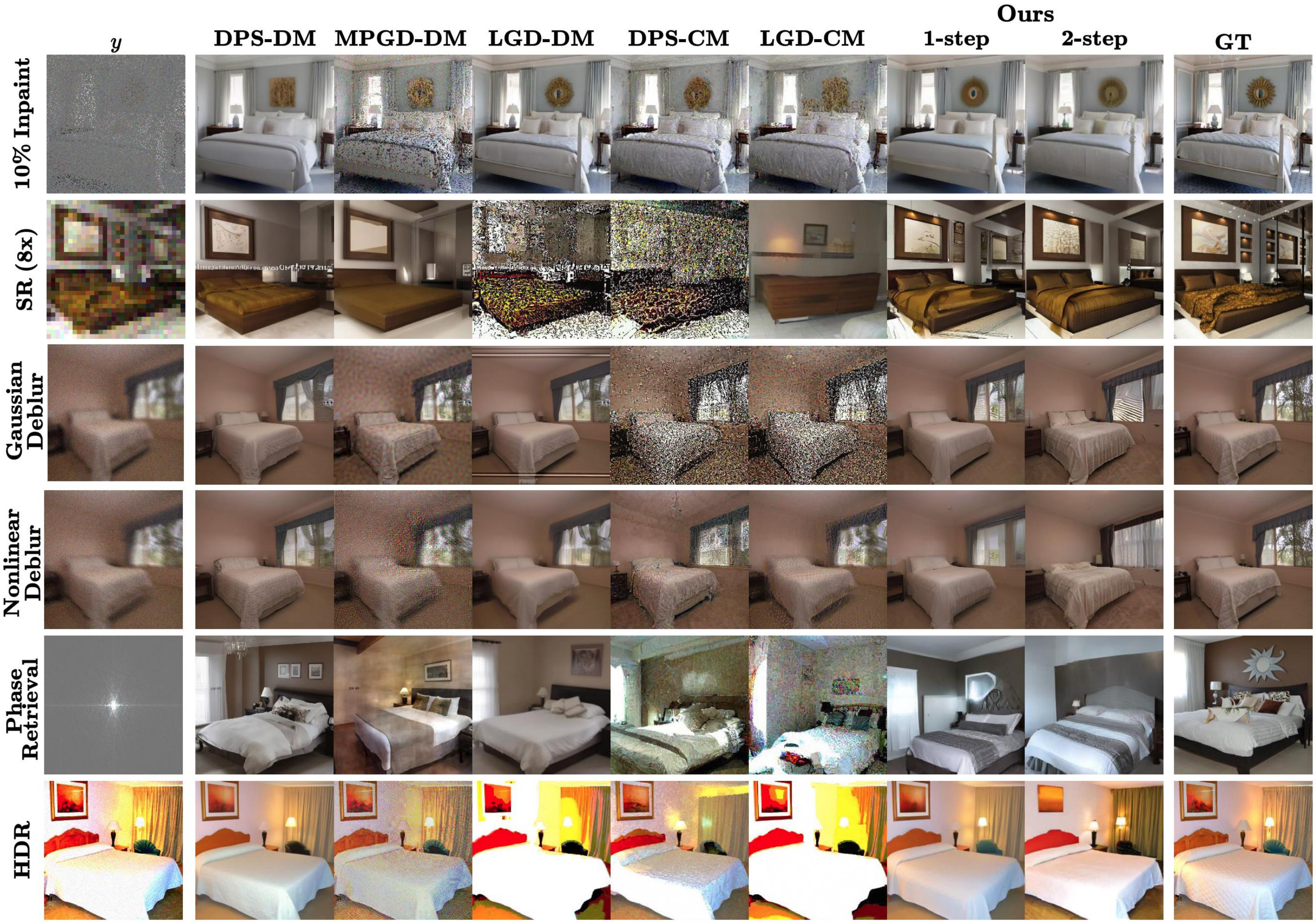
Using CMs from Song et al., which are U-nets

We consider noisy inverse problems in image data:

- **Linear problems**: super-resolution, inpainting, Gaussian blur
- **Nonlinear problems**: nonlinear blur, Fourier phase retrieval, high dynamic range (HDR) reconstruction

Fidelity: one goal can be to obtain highly accurate reconstructions of the ground truth.

Diversity: another goal is to explore and represent uncertain features within the posterior.

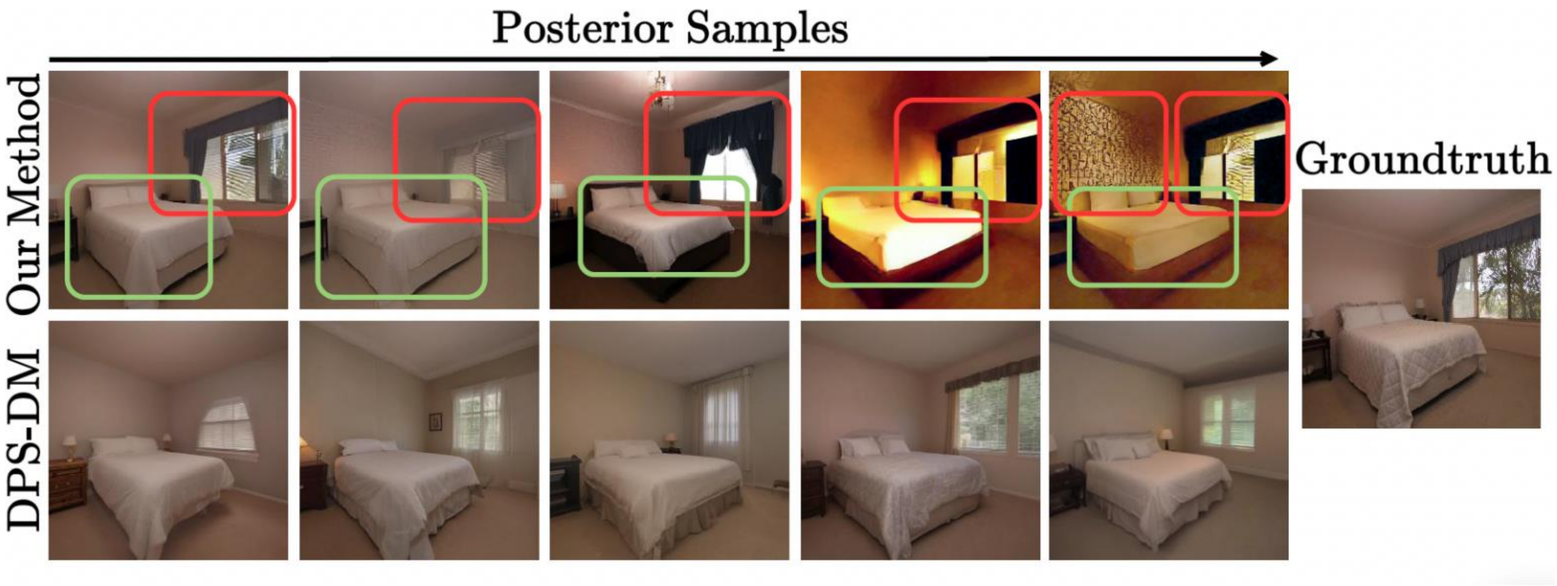

Ours
y
DPS-DM
MPGD-DM
LGD-DM
DPS-CM
LGD-CM
1-step
2-step
GT
10% Inpaint
SR (8x)
Gaussian Deblur
Nonlinear Deblur
Phase Retrieval
HDR

Uncertainty Quantification

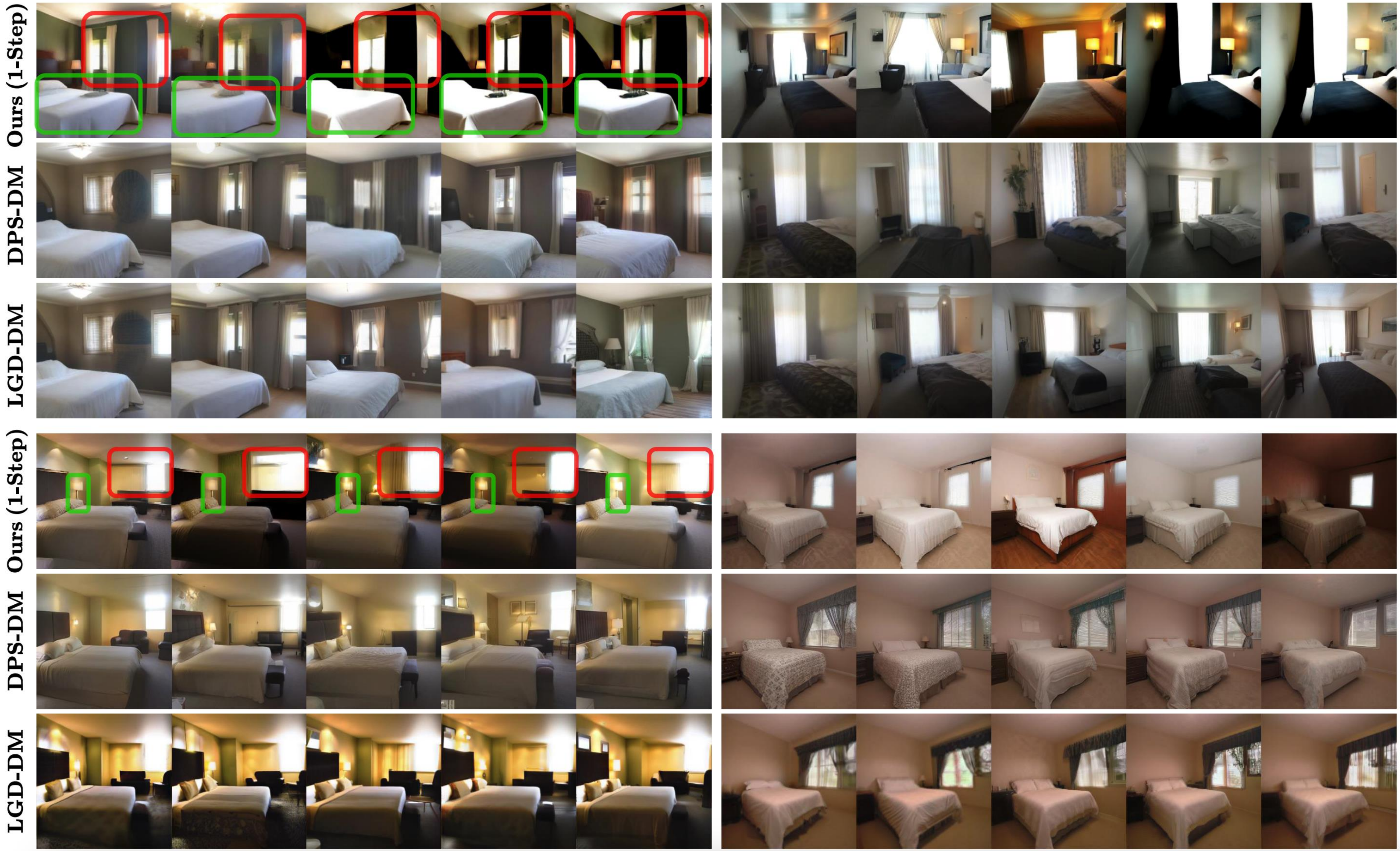
Our method also demonstrates superior diversity within the posterior, highlighting uncertain semantic features.

Red:
High Uncertainty

Green:
Low Uncertainty



Green boxes are persistent features, while red boxes are highly variable.


Ours (1-Step)
DPS-DM
LGD-DM
Ours (1-Step)
DPS-DM
LGD-DM

Conclusions

Sampling in the noise space enables **efficient accumulation of posterior samples**, leveraging the generative mapping Φ.

Few-step generative models, such as consistency models, can enable scalable generation.

Our approach enables **uncertainty quantification** via the generation of diverse samples, surpassing diffusion-based methods.