BEYOND BEST-OF-k: PRIOR-CONSISTENT DAPS FOR AMBIGUOUS NONLINEAR INVERSE PROBLEMS

Anonymous authors
Paper under double-blind review

ABSTRACT

Diffusion models have become strong priors for solving inverse problems. However, nonlinear inverse problems such as phase retrieval remain challenging when multiple measurement-consistent solutions are distinguishable only by the prior; we call such measurements ambiguous. Existing diffusion-based samplers rely on local approximations during reconstruction. For an ambiguous measurement, these approximations can allow measurement guidance to favor a measurement-consistent mode that is close to a prior-supported mode but is itself not supported by the prior. Evaluation often masks this type of failure by reporting the best-of-k runs, where the best is selected using ground truth. We introduce Prior-Consistent DAPS (PC-DAPS), a variant of Decoupled Annealing Posterior Sampling (DAPS) designed to address this failure case by maintaining multiple candidate solutions within a single sampling run. As annealing noise decreases, PC-DAPS progressively prunes and replaces weaker candidates according to their prior consistency, without any ground-truth access. PC-DAPS reduces the incidence of prior-inconsistent samples and improves nonlinear reconstruction quality while preserving sample quality on linear inverse problems.

1 INTRODUCTION

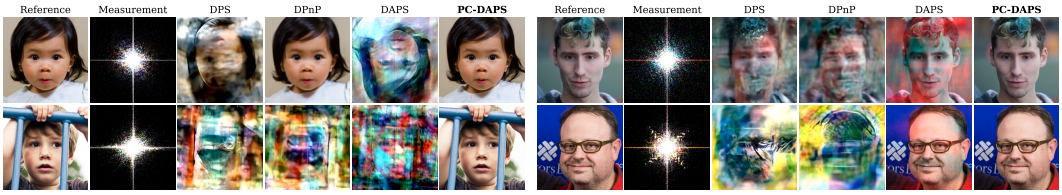


Figure 1: **Measurement consistency is not sufficient for reconstruction.** A single Fourier magnitude measurement is shared by rotations, translations, and color-channel inversions of an image. Existing solvers may return these prior-inconsistent modes, while PC-DAPS selects reconstructions supported by the learned image prior. Each displayed reconstruction is randomly selected from four independent runs for other methods and from the four particles of a single PC-DAPS run.

Inverse problems arise across science and engineering whenever an unknown signal must be recovered from indirect measurements, including image restoration (Bertero et al., 2021), medical imaging (Song et al., 2022), and astronomical imaging (Event Horizon Telescope Collaboration, 2019). Given an observed measurement y generated through a known but potentially noisy measurement process, the goal of the inverse problem is to recover the original signal x. For underdetermined problems, in which many different signals x may give rise to the same measurement y, Bayesian imaging defines the posterior distribution as $p(x \mid y) \propto p(y \mid x)p(x)$, where the likelihood $p(y \mid x)$ captures measurement consistency of a candidate image x and the prior $p(x)$ captures which candidate reconstructions are plausible natural signals. The goal of posterior sampling is thus to generate candidate reconstructions that are simultaneously consistent with the measurement and the prior.

In Fourier phase retrieval, for example, multiple images can produce identical Fourier magnitude measurements (Fienup, 1982). These measurement-equivalent solutions therefore have the same

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

likelihood, but can have very different posterior probabilities depending on their consistency with the image prior. In particular, we observe that existing phase-retrieval solvers applied to face images sometimes return a rotated or color-inverted face that is perfectly consistent with the measurement but implausible under a natural-image prior concentrated on upright faces. The challenge is therefore not simply to find measurement-consistent reconstructions, but to sample the measurement-consistent solutions supported by the prior.

Diffusion models provide strong image priors for posterior sampling (Ho et al., 2020; Song et al., 2021). At each noise level, a diffusion model implicitly defines a denoiser that maps a noisy sample x_t to an estimate of the corresponding clean image; this diffusion denoiser can be used to incorporate the prior $p(x)$ during posterior sampling. Diffusion-based inverse solvers apply this prior information along a sampling *trajectory*, i.e., the sequence of intermediate states that evolves from a noisy sample toward a reconstruction. Existing methods primarily either tilt the unconditional reverse diffusion process using an approximate likelihood gradient (Zhang et al., 2025; Chung et al., 2023; Song et al., 2023; Boys et al., 2024), or alternate prior and measurement updates across noise levels (Wu et al., 2024; Xu & Chi, 2024; Chen et al., 2026; Ahmed & Makin, 2026). In these methods, the prior information available at each noise level is represented through a local denoising approximation centered on a single clean-image estimate, rather than explicitly maintaining multiple competing prior modes at intermediate timesteps.

When the measurement is ambiguous, as in the case of phase retrieval and other underdetermined inverse problems, following a sampling trajectory can lead to *premature mode selection*. At high noise levels, multiple prior modes may remain plausible, but the trajectory can gradually move toward a mode that is not prior-supported but measurement-consistent. Once a trajectory selects a particular prior mode, subsequent refinement primarily improves measurement consistency in the local neighborhood of that mode rather than considering more distant alternatives. This becomes problematic when (i) the selected prior mode is not consistent with the measurement, and (ii) its nearest measurement-consistent mode is not supported by the prior. Trajectory-based samplers can prematurely commit to this pair of modes, ultimately resulting in hallucinatory sampled reconstructions that are not supported by the true posterior or even the true prior. Best-of-k evaluation partially masks this issue by generating multiple independent reconstructions and selecting the one closest to the ground truth (Chung et al., 2023; Zhang et al., 2025). However, this selection uses information unavailable at test time and does not ensure that the produced samples are prior-consistent. Particle-based approaches maintain multiple candidates (Cardoso et al., 2024; Dou & Song, 2024; Wu et al., 2023), but have primarily focused on linear inverse problems and particle diversity rather than prior-consistent sampling.

To address this issue, we introduce Prior-Consistent DAPS (PC-DAPS), which moves sample selection inside the sampler and removes the need for ground-truth-based best-of-k evaluation. PC-DAPS maintains a bag of K particles annealed under the same measurement, and models the denoiser posterior $p(x_0 \mid x_t)$ with a K-component Gaussian mixture rather than a unimodal estimate. Early in sampling, multiple candidates are kept while the noise level is high and the prior is still weakly informative. As the noise decreases during annealing, particles with low prior consistency are pruned and replaced by surviving candidates that are more prior-consistent. PC-DAPS thus delays mode selection until the prior becomes more informative, performing within a single sampling run (and without ground-truth access) what best-of-k does only after k independent sampling runs.

We evaluate PC-DAPS on FFHQ (Karras et al., 2019) and ImageNet (Deng et al., 2009) across three nonlinear and seven linear tasks, including the eight tasks reported in DAPS (Zhang et al., 2025) as well as high-pass filtering (Thaker et al., 2025) and 16× super-resolution. We report PSNR, SSIM (Wang et al., 2004), and LPIPS (Zhang et al., 2018) under both single-sample and best-of-k evaluation. We additionally compare the rate of prior-inconsistent hallucinations against independent DAPS runs. PC-DAPS reduces the incidence of prior-inconsistent reconstructions and improves performance on ambiguous nonlinear problems while maintaining image quality on linear inverse problems. Our contributions are as follows:

- We propose PC-DAPS, which maintains a bag of particles, models the denoiser posterior with a Gaussian mixture, and performs prior-based pruning without access to ground truth.
- We evaluate PC-DAPS across seven linear and three nonlinear inverse problems on FFHQ and ImageNet, showing fewer prior-inconsistent reconstructions and improved nonlinear reconstruction quality while preserving performance on linear tasks.

2 Related Work

Likelihood-guided diffusion sampling. Given a diffusion prior that has learned the unconditional score $\nabla_{x_t} \log p(x_t)$, we can decompose the conditional score (after observing measurement y) as $\nabla_{x_t} \log p(x_t \mid y) = \nabla_{x_t} \log p(x_t) + \nabla_{x_t} \log p(y \mid x_t)$. The likelihood term $\nabla_{x_t} \log p(y \mid x_t)$ is intractable, so a family of methods approximates it to steer the reverse diffusion process. In particular, DPS (Chung et al., 2023) evaluates the likelihood at the one-step posterior mean $p(y \mid x_t) \approx p(y \mid \mathbb{E}[x_0 \mid x_t])$ and backpropagates to compute the score. For linear measurement operators, ΠGDM (Song et al., 2023) and TMPD (Boys et al., 2024) instead propagate Gaussian approximations of $p(x_0 \mid x_t)$, while DDRM (Kawar et al., 2022) utilizes a diagonalization approach. MCG (Chung et al., 2022) adds a manifold-correction term to keep the guided iterate close to the data manifold. Jin et al. (2026) target hallucination inside this family of approaches by stabilizing the prior update, but remain committed to a single trajectory.

Split and annealed posterior sampling. In contrast, variable-splitting methods alternate between prior-only and measurement-only steps. Among methods using a diffusion model for the prior step, PnP-DM (Wu et al., 2024) uses a split Gibbs sampler for the likelihood, while DPnP (Xu & Chi, 2024) uses Metropolis-Hastings, and DiffPIR (Zhu et al., 2023) utilizes a proximal step. Finally, DAPS (Zhang et al., 2025) utilizes a repeated denoising-renoising scheme (Section 3.2). RePS (Ahmed & Makin, 2026) adds restarts to that schedule, and DAPS++ (Chen et al., 2026) reports that the DAPS reconstruction is largely driven by measurement consistency. We therefore design PC-DAPS as a modification of DAPS that retains its variable-splitting structure but introduces prior-based selection to discriminate between measurement-consistent reconstructions.

Particle-based posterior sampling. Another group of approaches maintains many particles throughout sampling, and uses sequential Monte Carlo to propagate, reweight, and resample particles to progressively improve measurement consistency (Cardoso et al., 2024; Janati et al., 2025). FPS-SMC (Dou & Song, 2024) casts the reverse sampling process as a state-space model and applies a particle filter. TDS (Wu et al., 2023) uses twisting functions to obtain asymptotic exactness for general measurement models. While particle-based methods are typically concerned with the fidelity of the ensemble in the limit of many particles, we are concerned instead with the probability that a single returned reconstruction is prior-inconsistent, and use particles for the purpose of exploring the posterior landscape.

Diagnosing hallucination and prior misfit. Hallucination has been studied as a property of reconstruction algorithms in inverse problems since before diffusion priors were adopted (Bhadra et al., 2021), and refers to any reconstruction that is inconsistent with either the measurement or the prior. Recent work has sought to quantify hallucination specifically for diffusion-based solvers (Jin et al., 2026). At the level of the model rather than the individual sample, DiME (Wang & Bouman, 2026) estimates the evidence of a diffusion prior by integrating over the time-marginals of a posterior sampler, and can be run alongside DAPS to detect prior misfit. Closest to our framing, Burns & Fridovich-Keil (2026) adopt a finite-sample perspective on reverse diffusion dynamics and show that widely used guidance approximations misestimate posterior spread at intermediate noise levels. This misestimation produces both prior modes absent from the true posterior and likelihood (measurement-consistent) modes unsupported by the prior, which is the specific failure that our PC-DAPS method aims to address. Benchmarks whose posteriors are analytically tractable provide a complementary check on the uncertainty a solver reports (Crafts & Villa, 2025; Zach et al., 2026; Zheng et al., 2025). All of these existing hallucination diagnostics are applied after sampling; PC-DAPS utilizes prior support during sampling to decide which reconstructions survive, thereby reducing the incidence of prior-inconsistent hallucinations.

3 Preliminaries

3.1 Bayesian inverse problems

We consider Bayesian inverse problems of the form:

$$y = \mathcal{A}(x) + n, \qquad n \sim \mathcal{N}(0, \beta_y^2 I), \tag{1}$$

where $x \in \mathbb{R}^d$ is the unknown signal of interest, $\mathcal{A} : \mathbb{R}^d \to \mathbb{R}^m$ is a known measurement operator that is deterministic but may be linear or nonlinear, $n \in \mathbb{R}^m$ is unknown additive Gaussian noise,

and $y \in \mathbb{R}^m$ is the observed measurement. Given a prior distribution $p(x)$, Bayes' rule gives the posterior distribution over likely signals x consistent with a provided measurement y:

$$p(x \mid y) \propto p(y \mid x)\, p(x), \qquad p(y \mid x) = \mathcal{N}\big(y; \mathcal{A}(x), \beta_y^2 I\big). \tag{2}$$

We are particularly interested in forward models where the set of measurement-consistent signals

$$\mathcal{M}(y) = \big\{x \in \mathbb{R}^d : \|\mathcal{A}(x) - y\| \leq \delta\big\} \tag{3}$$

contains multiple candidate reconstructions which are well-separated in signal space. In this context, two distinct signals in $\mathcal{M}(y)$ can be assigned equal likelihoods $p(y \mid x)$ but have drastically different posterior probabilities $p(x \mid y)$, as one image might be unrealistic (i.e., significantly lower prior probability $p(x)$). A common failure case of existing reconstruction methods is to sample both signals with roughly equal frequency, thereby misrepresenting their posterior probabilities as relatively equal.

For example, in Fourier phase retrieval, two images which are identical up to rotation, translation, and/or color channel sign appear visually quite distinct but produce identical Fourier magnitude measurements (Fienup, 1982). On structured face images, a solver can return a $180°$-rotated or color-inverted (e.g., blue) face that fits the observed Fourier measurement despite being inconsistent with a prior distribution supported only on the set of upright faces. We refer to such reconstructions as *prior-inconsistent*. Prior-inconsistent reconstructions commonly appear across classical phase-retrieval methods (Fienup, 1982), deep-learning approaches (Metzler et al., 2018; Işil & Oktem, 2025), and diffusion-based inverse-problem solvers (Xu & Chi, 2024; Zhang et al., 2025). Our goal is to reduce the frequency of prior-inconsistent solutions while retaining diversity of measurement- and prior-consistent samples.

3.2 Decoupled Annealing Posterior Sampling (DAPS)

To develop a sampler which produces fewer prior-inconsistent samples, we build on existing approaches by (i) identifying the root causes that produce the inconsistent samples and (ii) making adjustments to reduce the incidence of these samples. Specifically, we build on DAPS (Zhang et al., 2025), a diffusion-based variable-splitting approach which repeatedly (i) denoises to a clean image via the diffusion reverse process, (ii) improves the clean image's likelihood consistency via MCMC (e.g., Langevin dynamics (Roberts & Tweedie, 1996)), and (iii) renoises the MCMC output via the diffusion forward process. By adding less noise back than was originally present, DAPS aims to iteratively denoise towards a measurement-consistent clean sample.

Mathematically, DAPS utilizes the fact that given an intermediate image $x_t \sim p(x_t \mid y)$, drawing:

$$x_0 \sim p(x_0 \mid x_t, y), \qquad x_{t-1} \sim \mathcal{N}\big(x_0, \sigma_{t-1}^2 I\big)\,, \tag{4}$$

produces $x_{t-1} \sim p(x_{t-1} \mid y)$. To sample from the guided denoising distribution $p(x_0 \mid x_t, y)$, observe that y and x_t are independent conditioned on x_0; thus, by Bayes' rule we can factorize:

$$p(x_0 \mid x_t, y) \propto p(y \mid x_0) p(x_0 \mid x_t). \tag{5}$$

While $p(y \mid x_0)$ is induced by Equation (1), the denoising distribution $p(x_0 \mid x_t)$ is unknown and intractable to estimate via sampling. DAPS utilizes a Gaussian approximation for this distribution:

$$p(x_0 \mid x_t) \approx \mathcal{N}\big(x_0; \hat{x}_0(x_t), r_t^2 I\big), \tag{6}$$

where $\hat{x}_0(x_t)$ is an estimate of $\mathbb{E}[x_0 \mid x_t]$ and r_t is set heuristically. Observe that the (Stein) score w.r.t. x_0 of the *approximated* right-hand side of Equation (5) is available in closed form, allowing for sampling via the unadjusted Langevin algorithm with step size η_t from initialization $\hat{x}_0(x_t)$:

$$\hat{x}_0^{(j+1)} \leftarrow \hat{x}_0^{(j)} + \eta_t\big(\nabla_{\hat{x}_0} \log p(\hat{x}_0^{(j)} \mid x_t) + \nabla_{\hat{x}_0} \log p(y \mid \hat{x}_0^{(j)})\big) + \sqrt{2\eta_t}\epsilon^{(j)}, \quad \epsilon^{(j)} \sim \mathcal{N}(0, I). \tag{7}$$

Once the Langevin algorithm terminates (after a user-prescribed number of iterations N), the result is renoised using the EDM parametrization (Karras et al., 2022) of the diffusion forward process $x_{t-1} = \hat{x}_0^{(N)} + \sigma_{t-1}\epsilon$ with $\epsilon \sim \mathcal{N}(0, I)$. Renoising in DAPS is introduced to counteract the locality of Langevin refinement: because consecutive Langevin updates are local, the samples may fail to explore alternative prior modes distant from $\hat{x}_0(x_t)$. The renoising-denoising update intends to allow particles to make larger moves per iteration and to investigate alternative prior modes.

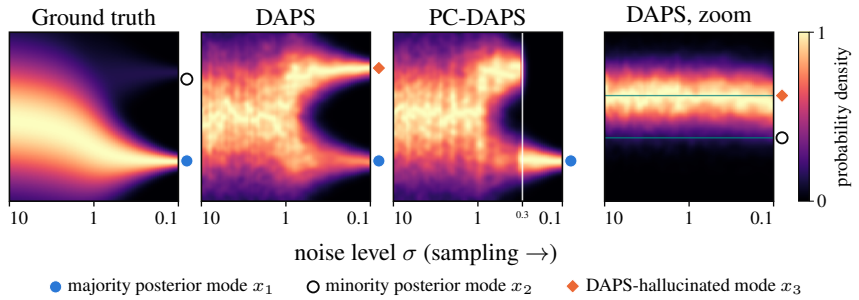


Figure 2: **Premature mode selection.** We consider a nonlinear inverse problem with quadratic $\mathcal{A}$ and two-atom prior; details are provided in Appendix A.2. We select y such that one atom is likelihood-consistent while the other is likelihood-inconsistent but neighboring a likelihood-consistent prior-inconsistent mode. The exact posterior, DAPS, and PC-DAPS are shown as distributions across noise levels σ. The exact posterior assigns positive weight to both atoms, with majority weight on the likelihood-consistent atom. However, DAPS assigns majority mass to the prior-inconsistent likelihood mode, as highlighted in the last panel, which is not present in the true posterior. PC-DAPS's selection algorithm prunes this mode, sampling only from the posterior-consistent atom. While neither DAPS nor PC-DAPS samples from the likelihood-inconsistent atom with adequate probability, PC-DAPS eliminates the hallucinated prior-inconsistent mode produced by DAPS.

However, the DAPS update rule still produces the prior-inconsistent samples described in Section 3.1, as shown in Figure 1. Observe that the DAPS procedure for approximately sampling from $p(x_0 \mid x_t, y)$ in effect produces a prior-consistent sample near $\mathbb{E}[x_0 \mid x_t]$ and samples from likelihood-consistent regions of its neighborhood. In contexts where x_t selects a prior-consistent sample that is likelihood-inconsistent, yet there exists a nearby region of signal space consistent with y, Equation (7) may sample from that region *even if it is prior-inconsistent*, as the Gaussian approximation to $p(x_0 \mid x_t)$ merely discourages the samples from deviating far in signal space from $\hat{x}_0(x_t)$. We observe that x_0 renoises to an x_t that subsequently denoises to the same region with high probability, causing the particle to become trapped in a prior-inconsistent but measurement-consistent mode. We refer to this phenomenon as *premature mode selection* because the sample x_t prematurely commits to only sampling x_0 from a particular prior mode, without being able to adequately decommit even if that prior mode is not present in the posterior.

Figure 2 illustrates this phenomenon. Here, the prior places mass on two candidate solutions, x_1 and x_2, while the measurement y is explained by x_1 and a third point x_3 unsupported by the prior. In analogy to Fourier phase retrieval, one can think of x_1 as the desired reconstruction, e.g., an upright, non-blue face that is both consistent with the measurement and supported by the image prior, x_2 as a realistic face inconsistent with the particular measurement, and x_3 as a structurally implausible transformation of the desired image x_1. The exact posterior should favor the x_1 mode while assigning zero mass to the prior-inconsistent x_3. However, DAPS samples from x_3's neighborhood by (i) sampling x_2 from the prior and (ii) sampling x_3 from x_2's local likelihood. Once a particle enters this locally attractive likelihood-consistent region, subsequent renoising and Langevin updates become trapped, resulting in premature mode selection.

This toy example motivates the design of PC-DAPS, which maintains multiple candidate measurement-consistent modes to identify and correct trapped prior-inconsistent samples. As shown in Figure 2, PC-DAPS places its mass only on the prior-supported and measurement-consistent mode at x_1 and avoids sampling from the prior-inconsistent, measurement-consistent region near x_3. We describe how PC-DAPS identifies and removes such prior-inconsistent particles in the next section.

4 Method

To address the presence of prior-inconsistent samples, a first approach might be to produce multiple samples via DAPS and discard any which are prior-inconsistent. There are two main shortcomings of this approach: (i) the effective sample size of this sampling method is suboptimal as substantial

computational resources are used to produce samples that are then thrown out, and (ii) it is unclear how to select which samples to discard without having access to the ground truth signal or posterior.

To address these issues, we introduce PC-DAPS, a procedure for correcting prematurely committed particles *during the sampling process* without requiring access to the ground truth image or to any pre-trained prior-inconsistency classifier. PC-DAPS maintains a bag of K particles, each of which is denoised using the same annealing schedule and pretrained denoiser, but uses different initializations and random noise in its Langevin and renoising updates. If no additional modifications are made to DAPS, the K particles will denoise independently, and can be partitioned into (nonempty) sets of "bad" prematurely committed particles and "good" particles that are posterior-consistent. The core idea of PC-DAPS is to improve DAPS's existing renoising strategy by identifying prematurely committed particles *during sampling* and moving them to more promising modes discovered by other particles in the bag. Additionally, PC-DAPS utilizes the particles to improve additional design decisions made by DAPS, such as by replacing the unimodal Gaussian approximation to the denoising distribution with an adaptive Gaussian mixture model (GMM) approximation. A toy example illustrating the benefit of Gaussian mixture approximation is shown in Appendix A.3.

What remains is to describe how PC-DAPS (i) identifies when particles commit prematurely, (ii) identifies which particles are problematic, (iii) corrects the identified particles, and (iv) constructs the multimodal denoising distribution approximation from the bag.

4.1 Pruning scheme

PC-DAPS begins by initializing a bag of K particles $\mathcal{B} = \{x_{t_{\max}}^{(k)}\}_{k=1}^{K}$ from the initial distribution of the reverse diffusion process, and denoises the particles as prescribed by DAPS. On tasks where $\mathcal{M}(y)$ (Equation (3)) contains multiple clearly separated candidates (e.g., for a nonlinear $\mathcal{A}$), the particles will begin to denoise to genuinely different candidate reconstructions. At the annealing time of the DAPS denoising process where the prior first becomes discriminative, PC-DAPS invokes a one-shot consensus check to identify low-prior particles.

To determine when the check should occur, we test the diversity of the bag at each iteration. Define:

$$\bar{r} = \frac{1}{K}\sum_{k=1}^{K}\left\|\mathcal{A}(x^{(k)}) - y\right\|^2, \qquad r^\star = \left\|\mathcal{A}(\bar{x}) - y\right\|^2, \qquad \bar{x} = \frac{1}{K}\sum_{k=1}^{K} x^{(k)}. \tag{8}$$

Observe that, for linear $\mathcal{A}$, Jensen's inequality forces $\bar{r} \geq r^\star$ by convexity of the squared norm. Using this convexity intuition, for nonlinear $\mathcal{A}$ the condition that $\bar{r} < r^\star$ indicates that the bag occupies multiple wells of the likelihood potential, indicating that mode-level pruning is meaningful.

To identify which particles require correcting, we partition the bag into clusters $\mathcal{C}_m$ and score each by its average prior consistency (obtained via Tweedie's formula) over several noise scales $\mathcal{S}$:

$$s(\mathcal{C}_m) = -\frac{1}{|\mathcal{C}_m||\mathcal{S}|}\sum_{x\in\mathcal{C}_m}\sum_{\sigma\in\mathcal{S}}\left\|x - D_\theta(x+\sigma\epsilon;\,\sigma)\right\|^2, \qquad \epsilon\sim\mathcal{N}(0,I). \tag{9}$$

The premise of this criterion is that a sample lying near the (prior) data manifold denoises with small score at every noise scale. If a cluster's score $s(\mathcal{C}_m)$ exceeds a threshold, its particles are unaltered. Otherwise, its particles are replaced randomly by copies of particles from clusters scoring above the threshold. Implementation choices including the clustering method, choice of noise scales $\mathcal{S}$, and determination of threshold are discussed in Appendix A.4.

4.2 Gaussian mixture approximation of the denoiser posterior

An additional benefit of maintaining a bag of particles is that we can improve upon DAPS's unimodal approximation in Equation (6) by selecting a multimodal distribution informed by the bag's diversity. We propose replacing the DAPS Gaussian approximation with a Gaussian mixture. Denote the clean-signal estimate of the k-th particle by $\tilde{x}_0^{(k)} = D_\theta\big(x_t^{(k)};\,\sigma_t\big)$. Before pruning described in Section 4.1 activates, we use a uniformly-weighted mixture of Gaussians centered at the clean estimates:

$$q(x_0 \mid x_t) = \frac{1}{K}\sum_{k=1}^{K}\mathcal{N}\Big(x_0;\tilde{x}_0^{(k)},\,\mathrm{diag}\left(\mathrm{Var}_{\mathcal{B}}\big(\tilde{x}_0^{(k)}\big)\right)\Big). \tag{10}$$

After pruning, we maintain a mode for each unpruned *cluster*, where the corresponding component parameters are given by the cluster's empirical weight, mean, and diagonal covariance:

$$q(x_0 \mid x_t) = \sum_{m=1}^{M} \pi_m \mathcal{N}\big(x_0; \mu_m, \Sigma_m\big), \tag{11}$$

$$\pi_m = \frac{|\mathcal{C}_m|}{K}, \qquad \mu_m = \frac{1}{|\mathcal{C}_m|} \sum_{x^{(k)} \in \mathcal{C}_m} \tilde{x}_0^{(k)}, \qquad \Sigma_m = \operatorname{diag}\Big(\operatorname{Var}_{\mathcal{C}_m}\big(\tilde{x}_0^{(k)}\big)\Big). \tag{12}$$

When a cluster contains only a single particle, its component uses the full bag's diagonal covariance.

Unlike DAPS, which centers the denoising approximation at the point estimate $\hat{x}_0(x_t)$, PC-DAPS shares the bag-mean center across all particles in a mode, giving a variance-reduced prior estimate that we find improves reconstruction quality on linear reconstruction tasks. Furthermore, when the full bag lies in a single cluster, the mixture reduces to a single Gaussian centered on the bag's average Tweedie estimate with data-dependent covariance. This contrasts with the heuristic, geometry-independent covariance utilized by DAPS.

Observe that the score of the GMM denoising approximation $q(x_0 \mid x_t)$ is accessible in closed form:

$$\nabla_{x_0} \log q(x_0 \mid x_t) = \sum_{m=1}^{M} q_m(x_0) \big(\Sigma_m\big)^{-1} \big(\mu_m - x_0\big), \tag{13}$$

with weights $q_m(x_0) = \pi_m \mathcal{N}(x_0; \mu_m, \Sigma_m) / \sum_{m'} \pi_{m'} \mathcal{N}(x_0; \mu_{m'}, \Sigma_{m'})$, and thus can be utilized in the DAPS Langevin update introduced in Equation (7). Observe that when the weights q_m are approximately one-hot, this update approximately reduces to a data-dependent Gaussian. Recalling that the DAPS Gaussian approximation disincentivizes the Langevin dynamics from deviating far from the point estimate $\hat{x}_0(x_t)$, the PC-DAPS Gaussian approximation discourages the Langevin dynamics from deviating far from the *prior mode* identified by a collective cluster of the bag.

The full algorithm (Algorithm 1) can be found in Appendix A.1. PC-DAPS makes two core changes to DAPS utilizing K particles annealed in parallel: a pruning technique to correct particles trapped in prior-inconsistent likelihood modes in tasks with likelihood ambiguity (Section 4.1), and a multimodal denoising distribution approximation estimated from the diversity of the particles (Section 4.2). Additional implementation details are provided in Appendix A.4, such as omitting the final Langevin refinement and adaptive ODE stepping from high to low noise to improve reconstruction quality while reducing computational cost.

5 Results

Main Results. Table 1 shows that while PC-DAPS is not uniformly superior to all baselines, it performs competitively across all experiments and often achieves the highest sample quality, with the most substantial improvements in phase retrieval. In particular, the phase-retrieval results for PC-DAPS are competitive across metrics despite its handicap in evaluation protocol; for PC-DAPS we average metrics across the four samples it generates in a single run, whereas all baselines are evaluated on only the best sample (chosen using ground-truth knowledge) generated across four independent runs.

Two implementation choices described in Appendix A.4 produce additional improvements: skipping Langevin refinement in the last DAPS iteration, and using adaptive ODE steps at low noise. Although PC-DAPS uses multiple particles at each timestep, its computational cost remains comparable to that of DAPS due to the adaptive ODE steps, as shown in Appendix A.5.

Prior-Inconsistent Hallucinations. Table 1 reports standard metrics to facilitate baseline comparisons. However, the evaluation procedure masks the failure mode PC-DAPS aims to address (measurement-consistent reconstructions that the prior does not support), because the best-of-k evaluation protocol used by baseline methods discards $k - 1$ of the reconstructed images. On phase retrieval, we instead classify every sample as *faithful*, *upside-down / color-inverted* (a measurement-consistent likelihood mode the natural-image prior would down-weight), or *failed*. Labels are assigned automatically using per-channel PSNR to the ground truth and its rotated version, maximizing over ± 1-pixel circular shifts to account for translation ambiguity. A channel is assigned to

Table 1: **Quantitative evaluation on FFHQ and ImageNet.** Comparison of methods on FFHQ (left) and ImageNet (right) across seven linear and three nonlinear inverse problems using PSNR, SSIM, and LPIPS, evaluated over 100 test images. For the nonlinear problems, mean $\pm$ standard deviation over 100 images is reported to highlight variation across samples. PC-DAPS produces 4 samples for each input, so the metrics reported for PC-DAPS are averaged over these 4 samples. For phase retrieval, for all methods other than PC-DAPS, we report oracle results from best-of-4 runs. We indicate the best and second-best results for each experiment with **bold** and underline annotations. We use a noisy measurement with noise level $\beta_y = 0.05$ in all tasks.

Task	Method	FFHQ PSNR (↑)	FFHQ SSIM (↑)	FFHQ LPIPS (↓)	ImageNet PSNR (↑)	ImageNet SSIM (↑)	ImageNet LPIPS (↓)
Super-resolution 4×	PC-DAPS	**30.25**	**0.841**	**0.166**	**26.44**	**0.714**	**0.245**
	DAPS	29.36	0.782	0.173	25.51	0.639	0.281
	DDNM	28.01	0.797	0.196	23.92	0.606	0.472
	DPS	25.86	0.753	0.269	21.13	0.489	0.361
	FPS-SMC	28.45	0.812	0.206	24.79	0.705	0.311
Super-resolution 16×	PC-DAPS	**22.21**	**0.623**	**0.368**	**20.90**	**0.454**	**0.535**
	DAPS	22.14	0.597	0.373	20.35	0.450	0.545
	DPS	21.09	0.586	0.570	18.41	0.384	0.592
	FPS-SMC	22.09	0.604	0.496	18.88	0.423	0.588
High-pass filter	PC-DAPS	**18.84**	**0.815**	**0.155**	**17.33**	**0.778**	**0.204**
	DAPS	16.85	0.702	0.204	15.51	0.732	0.217
	DPS	4.92	0.437	0.687	12.84	0.606	0.473
	FGPS	18.04	0.761	0.184	14.77	0.708	0.229
Inpaint (box)	PC-DAPS	24.73	0.825	0.152	21.56	0.738	0.221
	DAPS	24.69	0.746	0.166	21.12	0.714	0.229
	DDNM	24.44	**0.839**	0.233	21.67	**0.746**	0.321
	DPS	22.51	0.792	0.209	18.94	0.722	0.257
	FPS-SMC	**24.86**	0.823	**0.146**	**22.16**	0.726	**0.208**
Inpaint (random)	PC-DAPS	**31.83**	**0.875**	**0.121**	28.18	0.781	**0.159**
	DAPS	30.73	0.801	0.129	27.32	0.725	0.189
	DDNM	29.87	0.816	0.123	**31.14**	**0.839**	0.193
	DPS	25.46	0.823	0.203	23.52	0.745	0.297
	FPS-SMC	28.17	0.824	0.258	24.55	0.699	0.318
Gaussian deblurring	PC-DAPS	**30.35**	**0.839**	**0.164**	26.37	**0.703**	**0.250**
	DAPS	29.60	0.787	0.179	25.94	0.659	0.267
	DDNM	28.18	0.805	0.214	**28.11**	0.701	0.281
	DPS	25.87	0.764	0.219	20.31	0.598	0.397
	FPS-SMC	26.47	0.771	0.251	23.86	0.603	0.390
Motion deblurring	PC-DAPS	**32.13**	**0.872**	**0.119**	**29.06**	**0.800**	**0.162**
	DAPS	31.65	0.837	0.137	28.66	0.770	0.175
	DPS	24.52	0.801	0.246	18.96	0.629	0.423
	FPS-SMC	27.31	0.825	0.229	24.60	0.645	0.329
Phase retrieval	PC-DAPS	**30.09**$_{\pm 1.77}$	**0.834**$_{\pm 0.089}$	**0.152**$_{\pm 0.088}$	**21.75**$_{\pm 1.10}$	0.469$_{\pm 0.132}$	**0.373**$_{\pm 0.128}$
	DAPS	29.66$_{\pm 4.17}$	0.770$_{\pm 0.100}$	0.177$_{\pm 0.090}$	20.98$_{\pm 6.72}$	0.461$_{\pm 0.226}$	0.387$_{\pm 0.163}$
	DPS	17.64$_{\pm 2.97}$	0.441$_{\pm 0.129}$	0.410$_{\pm 0.090}$	16.81$_{\pm 3.61}$	0.427$_{\pm 0.143}$	0.447$_{\pm 0.099}$
	PnP-DM	21.70$_{\pm 5.50}$	0.542$_{\pm 0.122}$	0.398$_{\pm 0.101}$	17.68$_{\pm 6.70}$	0.458$_{\pm 0.191}$	0.499$_{\pm 0.101}$
	DPnP	20.39$_{\pm 6.15}$	0.592$_{\pm 0.139}$	0.392$_{\pm 0.173}$	16.74$_{\pm 3.21}$	**0.487**$_{\pm 0.196}$	0.535$_{\pm 0.107}$
	RED-Diff	15.72$_{\pm 4.41}$	0.401$_{\pm 0.191}$	0.588$_{\pm 0.089}$	15.11$_{\pm 3.68}$	0.389$_{\pm 0.058}$	0.529$_{\pm 0.127}$
	HIO	13.53$_{\pm 2.50}$	0.359$_{\pm 0.093}$	0.726$_{\pm 0.068}$	-	-	-
Nonlinear deblur	PC-DAPS	29.04$_{\pm 0.82}$	**0.804**$_{\pm 0.017}$	0.180$_{\pm 0.023}$	27.63$_{\pm 2.58}$	0.739$_{\pm 0.053}$	**0.211**$_{\pm 0.032}$
	DAPS	28.49$_{\pm 1.52}$	0.762$_{\pm 0.031}$	0.191$_{\pm 0.031}$	27.25$_{\pm 3.18}$	0.717$_{\pm 0.089}$	0.213$_{\pm 0.055}$
	DPS	23.39$_{\pm 2.01}$	0.623$_{\pm 0.082}$	0.278$_{\pm 0.060}$	22.49$_{\pm 3.20}$	0.591$_{\pm 0.101}$	0.306$_{\pm 0.081}$
	RED-Diff	**30.72**$_{\pm 0.56}$	0.793$_{\pm 0.029}$	**0.162**$_{\pm 0.032}$	**29.91**$_{\pm 1.36}$	**0.751**$_{\pm 0.024}$	0.214$_{\pm 0.080}$
High dynamic range	PC-DAPS	**28.25**$_{\pm 1.17}$	**0.839**$_{\pm 0.020}$	**0.163**$_{\pm 0.039}$	**26.01**$_{\pm 2.77}$	**0.822**$_{\pm 0.099}$	**0.173**$_{\pm 0.094}$
	DAPS	27.67$_{\pm 3.57}$	**0.839**$_{\pm 0.078}$	0.165$_{\pm 0.069}$	25.80$_{\pm 4.37}$	0.819$_{\pm 0.113}$	0.177$_{\pm 0.106}$
	DPS	22.73$_{\pm 6.07}$	0.591$_{\pm 0.141}$	0.264$_{\pm 0.156}$	19.23$_{\pm 2.52}$	0.582$_{\pm 0.082}$	0.503$_{\pm 0.106}$
	RED-Diff	21.94$_{\pm 3.52}$	0.508$_{\pm 0.081}$	0.261$_{\pm 0.091}$	21.88$_{\pm 5.77}$	0.598$_{\pm 0.097}$	0.281$_{\pm 0.191}$

an orientation only when it is more than 1 dB closer to that orientation; samples with all channels matching the ground truth are *faithful*, those matching the rotation or disagreeing across channels are *upside-down / color-inverted*, and the remainder are *failed*. Table 2 reports the resulting frequencies, with examples in Figure 3.

On FFHQ, PC-DAPS lowers the rate of upside-down or color-inverted reconstructions from 14.5% to 1.5% and raises the rate of faithful reconstructions from 79.5% to 97.5%. This addresses the main failure mode PC-DAPS is meant to fix. On ImageNet, the improvements are smaller: the proportion of faithful reconstructions increases from 38.5% to 45.0%, while the proportion of unrecoverable runs/bags (runs/bags where there exists no faithful sample) drops from 25% to 15%, and the rate of upside-down or color-inverted reconstructions decreases from 42.0% to 30.5%. PC-DAPS relies on prior consistency to disambiguate the color and orientation ambiguities in phase-retrieval measure-

Table 2: **Phase-retrieval outcome frequencies for DAPS and PC-DAPS on FFHQ and ImageNet.** Individual run outcomes classified as Faithful (correctly reconstructed), Upside-down/color-inverted (converged but with rotation or color ambiguities), or Failed (diverged or noisy output). Unrecoverable denotes instances where all 4 runs or particles within a bag are unfaithful, yielding no reliable sample. Overall, PC-DAPS substantially reduces upside-down/color-inverted samples while increasing faithful reconstructions. The performance gap between datasets highlights the inherent difficulty of phase retrieval on highly diverse, unaligned images (ImageNet) compared to structured prior domains (FFHQ).

Dataset	Outcome	DAPS (4 runs)	PC-DAPS (bag of 4)
FFHQ	Faithful ↑	79.5%	**97.5%**
	Upside-down / color-inverted ↓	14.5%	**1.5%**
	Failed ↓	6.0%	**1.0%**
	Unrecoverable runs/bags ↓	4%	**1%**
ImageNet	Faithful ↑	38.5%	**45.0%**
	Upside-down / color-inverted ↓	42.0%	**30.5%**
	Failed ↓	**19.5%**	24.5%
	Unrecoverable runs/bags ↓	25%	**15%**



Figure 3: **PC-DAPS resolves the upside-down orientation inconsistencies seen in standard DAPS best-of-k runs.** The diversity rows show that, although PC-DAPS is globally stable, it preserves subtle variation among the four particles generated within a single run in comparison to the four independent DAPS runs.

ments, but the multi-class ImageNet prior provides weaker disambiguating signals at varying noise scales when compared to the FFHQ prior. When the local geometry of the prior support is more complex, data-manifold-based prior-consistency pruning appears to be less effective.

6 Discussion

Motivation, limitations, and future work. Results in the diffusion literature for ambiguous inverse problems are usually reported as best-of-k, with selection using the ground truth. This evaluation approach became common because existing solvers are inconsistent across runs, and a single-sample result would have large variations in quality. We see this instability as a symptom of premature mode selection in the context of measurement ambiguity. We introduce PC-DAPS to utilize the learned prior information to reduce this unwanted variance during sampling without the need for a ground-truth reference. Even in high-dimensional datasets, including FFHQ, our method demonstrates significant improvements due to the local structure of the prior. The method obtains a smaller improvement on more structurally diverse datasets, namely ImageNet, because the prior landscape is more complex, exposing a remaining limitation and open problem to be addressed. We encourage future work to further explore our proposed paradigm of leveraging the prior score to assess candidate reconstructions at inference time, and work towards eliminating prior-inconsistent reconstructions even under less-structured priors.

Posterior sampling metrics. Standard metrics like PSNR, SSIM, and LPIPS measure how similar a generated image is to a single reference image. Observe that these evaluation metrics penalize methods that find several plausible solutions that are equally valid under the measurement and prior if the solutions are dissimilar from the fixed reference. Furthermore, an image with (for example) high PSNR when compared to the ground truth may be inconsistent with the provided measurement. Because sample diversity and likelihood consistency are explicit goals of posterior sampling, including in applications such as uncertainty quantification, these sample-quality metrics are not adequate replacements for quality of the sampled posterior. In our work, we report a separate axis of reconstruction quality: prior consistency. We believe this or similar multifaceted metrics should be considered alongside standard fidelity metrics.

AI USE STATEMENT

In this work, we used generative AI tools for: designing research experiments and implementing methods (editing pre-existing implementations of baselines from publicly available repositories).

We have not used generative AI tools for: help with developing theoretical models or conceptual frameworks, formulating mathematical claims, providing critical ingredients for proving mathematical claims, assisting in the writing of proofs, proposing or refining hypotheses, providing feedback on research methodology, supporting qualitative and thematic data analysis, or interpreting results.

Generating synthetic data sets, assisting with translation, and cleaning and reformatting datasets are not applicable to this work.

Additionally, we used generative AI tools for creating or modifying scientific figures or images, creating or editing software code, and editing a research paper to improve readability.

We have reviewed all AI-assisted work by validating the performance of the baseline implementations with (i) the original repositories and (ii) the original papers cited in this work. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

We follow ethical practices in our research, especially in how we use data. All datasets and checkpoints in this study are publicly available and widely recognized in the computer vision community. We use this data and reproduce the methods in accordance with their licenses and for academic purposes.

REPRODUCIBILITY STATEMENT

We are committed to ensuring that others can reproduce the research in this paper. We include detailed information about our method and experiments in the Supplementary Material, covering all procedures and hyperparameters. Upon acceptance, we will also share all source code publicly.

REFERENCES

Bilal Ahmed and Joseph G Makin. Solving diffusion inverse problems with restart posterior sampling. In *European Conference on Computer Vision (ECCV)*, pp. 79–96. Springer, 2026.

Mario Bertero, Patrizia Boccacci, and Christine De Mol. *Introduction to inverse problems in imaging*. CRC press, 2021.

Sayantan Bhadra, Varun A Kelkar, Frank J Brooks, and Mark A Anastasio. On hallucinations in tomographic image reconstruction. *IEEE Transactions on Medical Imaging*, 40(11):3249–3260, 2021.

Benjamin Boys, Mark Girolami, Jakiw Pidstrigach, Sebastian Reich, Alan Mosca, and Omer Deniz Akyildiz. Tweedie moment projected diffusions for inverse problems. *Transactions on Machine Learning Research (TMLR)*, 2024. ISSN 2835-8856.

Benjamin A Burns and Sara Fridovich-Keil. When, why, and how do diffusion posterior samplers fail? A finite-sample lens. *arXiv preprint arXiv:2605.30330*, 2026.

Gabriel Cardoso, Yazid Janati El Idrissi, Sylvain Le Corff, and Eric Moulines. Monte Carlo guided denoising diffusion models for Bayesian linear inverse problems. In *International Conference on Learning Representations (ICLR)*, volume 2024, pp. 44001–44037, 2024.

Hao Chen, Renzheng Zhang, and Scott S Howard. DAPS++: Rethinking diffusion inverse problems with decoupled posterior annealing. In *European Conference on Computer Vision (ECCV)*, pp. 223–240. Springer, 2026.

Hyungjin Chung, Byeongsu Sim, Dohoon Ryu, and Jong Chul Ye. Improving diffusion models for inverse problems using manifold constraints. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 35, pp. 25683–25696, 2022.

Hyungjin Chung, Jeongsol Kim, Michael Thompson Mccann, Marc Louis Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *The Eleventh International Conference on Learning Representations (ICLR)*, 2023.

Evan Scope Crafts and Umberto Villa. Benchmarking diffusion annealing-based bayesian inverse problem solvers. *IEEE Open Journal of Signal Processing*, 2025.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 248–255. IEEE, 2009.

Zehao Dou and Yang Song. Diffusion posterior sampling for linear inverse problem solving: A filtering perspective. In *International Conference on Learning Representations (ICLR)*, volume 2024, pp. 12069–12095, 2024.

Event Horizon Telescope Collaboration. First M87 Event Horizon Telescope Results. IV. Imaging the Central Supermassive Black Hole. *The Astrophysical Journal Letters*, 875(1):L4, apr 2019.

James R Fienup. Phase retrieval algorithms: A comparison. *Applied Optics*, 21(15):2758–2769, 1982.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pp. 6840–6851. Curran Associates, Inc., 2020.

Çağatay Işil and Figen S Oktem. Deep plug-and-play HIO approach for phase retrieval. *Applied Optics*, 64(5):A84–A94, 2025.

Yazid Janati, Badr Moufad, Mehdi Abou El Qassime, Alain Oliviero Durmus, Eric Moulines, and Jimmy Olsson. A mixture-based framework for guiding diffusion models. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *International Conference on Machine Learning (ICML)*, volume 267 of *Proceedings of Machine Learning Research*, pp. 26830–26876. PMLR, 13–19 Jul 2025.

Pengfei Jin, Yiqi Tian, Kailong Fan, Bingjie Qi, and Quanzheng Li. Hallucination-aware diffusion sampling for inverse problems via robust prior updates. *arXiv preprint arXiv:2606.02331*, 2026.

Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4396–4405. IEEE, 2019.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems (NeurIPS)*, volume 35, pp. 26565–26577. Curran Associates, Inc., 2022. doi: 10.52202/068431-1926.

Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 35, pp. 23593–23606, 2022.

Morteza Mardani, Jiaming Song, Jan Kautz, and Arash Vahdat. A variational perspective on solving inverse problems with diffusion models. In *International Conference on Learning Representations (ICLR)*, volume 2024, pp. 28027–28053, 2024.

Christopher Metzler, Phillip Schniter, Ashok Veeraraghavan, and Richard Baraniuk. prDeep: Robust phase retrieval with a flexible deep network. In *International Conference on Machine Learning (ICML)*, pp. 3501–3510. PMLR, 2018.

Gareth O. Roberts and Richard L. Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341–363, 1996. doi: 10.2307/3318418.

Gideon Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978. doi: 10.1214/aos/1176344136.

Jiaming Song, Arash Vahdat, Morteza Mardani, and Jan Kautz. Pseudoinverse-guided diffusion models for inverse problems. In *International conference on learning representations (ICLR)*, 2023.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations (ICLR)*, 2021.

Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Solving inverse problems in medical imaging with score-based generative models. In *International Conference on Learning Representations (ICLR)*, 2022.

Darshan Thaker, Abhishek Goyal, and René Vidal. Frequency-guided posterior sampling for diffusion-based image restoration. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 12873–12882. IEEE, 2025.

Frederic Wang and Katherine Bouman. Sample-efficient evidence estimation of score based priors for model selection. In *International Conference on Learning Representations (ICLR)*, volume 2026, pp. 74960–74987, 2026.

Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-shot image restoration using denoising diffusion null-space model. In *The Eleventh International Conference on Learning Representations (ICLR)*, 2023.

Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

Luhuan Wu, Brian Trippe, Christian Naesseth, David Blei, and John Cunningham. Practical and asymptotically exact conditional sampling in diffusion models. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 36, pp. 31372–31403. Curran Associates, Inc., 2023. doi: 10.52202/075280-1363.

Zihui Wu, Yu Sun, Yifan Chen, Bingliang Zhang, Yisong Yue, and Katherine Bouman. Principled probabilistic imaging using diffusion models as plug-and-play priors. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2024.

Xingyu Xu and Yuejie Chi. Provably robust score-based diffusion posterior sampling for plug-and-play image reconstruction. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems (NeurIPS)*, volume 37, pp. 36148–36184. Curran Associates, Inc., 2024. doi: 10.52202/079017-1140.

Martin Zach, Youssef Haouchat, and Michael Unser. A statistical benchmark for diffusion-posterior-sampling algorithms. In *International Conference on Learning Representations (ICLR)*, volume 2026, pp. 108984–109021, 2026.

Bingliang Zhang, Wenda Chu, Julius Berner, Chenlin Meng, Anima Anandkumar, and Yang Song. Improving Diffusion Inverse Problem Solving with Decoupled Noise Annealing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20895–20905, 2025.

Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 586–595. IEEE, 2018.

Hongkai Zheng, Wenda Chu, Bingliang Zhang, Zihui Wu, Austin Wang, Berthy T. Feng, Caifeng Zou, Yu Sun, Nikola Kovachki, Zachary E. Ross, Katherine L. Bouman, and Yisong Yue. InverseBench: Benchmarking plug-and-play diffusion priors for inverse problems in physical sciences. In *International Conference on Learning Representations (ICLR)*, volume 2025, pp. 90912–90940, 2025.

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 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1219–1229. IEEE, 2023.

A Supplementary Material

A.1 Full Algorithm

Algorithm 1 Prior-Consistent DAPS (PC-DAPS)

Require: measurement y, pretrained denoiser D_θ, measurement operator $\mathcal{A}$, measurement noise β_y
Require: noise schedule $\sigma_{t_{N_A}} > \cdots > \sigma_{t_0} = 0$, bag size K, Langevin steps N, threshold γ

$x_{t_{N_A}}^{(k)} \sim \mathcal{N}(0, \sigma_{t_{N_A}}^2 I)$ for $k = 1, \ldots, K$
`pruned` $\leftarrow$ `false`
for $i = N_A, N_A - 1, \ldots, 1$ **do**
  Clean-signal estimates and denoising distribution
    $\tilde{x}_0^{(k)} \leftarrow D_\theta\Big(x_{t_i}^{(k)}; \sigma_{t_i}\Big)$
    if `pruned` $=$ `false` **then**
      $q(x_0 \mid x_t) \leftarrow \frac{1}{K} \sum_{k=1}^{K} \mathcal{N}\Big(x_0; \tilde{x}_0^{(k)}, \mathrm{diag}(\mathrm{Var}_{\mathcal{B}}(\tilde{x}_0^{(k)}))\Big)$
    else
      cluster $\{\tilde{x}_0^{(k)}\}_{k=1}^K$ into $\{\mathcal{C}_m\}_{m=1}^M$
      construct $q(x_0 \mid x_t)$ using Equations (11) and (12)
    end if
  Langevin refinement
    if $i > 1$ **then**
      $\hat{x}_0^{(k,0)} \leftarrow \tilde{x}_0^{(k)}$
      for $j = 0, \ldots, N-1$ **do**
        $\hat{x}_0^{(k,j+1)} \leftarrow$ Langevin step (Equation (7)) using the score in Equation (13)
      end for
  One-shot consensus pruning
      if $\neg$`pruned` and $\bar{r} < r^\star$ and clustering returns $M > 1$ clusters **then**
        compute $s(\mathcal{C}_m)$ for each cluster using Equation (9)
        $\mathcal{G} \leftarrow \{\mathcal{C}_m : s(\mathcal{C}_m) \geq (1+\gamma) \max s(\mathcal{C}_m)\}$
        randomly replace particles in clusters $\mathcal{C}_m \notin \mathcal{G}$ by copies from clusters in $\mathcal{G}$
        `pruned` $\leftarrow$ `true`
      end if
  Re-noising
      $x_{t_{i-1}}^{(k)} \leftarrow \hat{x}_0^{(k,N)} + \sigma_{t_{i-1}} \epsilon^{(k)}, \quad \epsilon^{(k)} \sim \mathcal{N}(0, I)$
    else
      $\hat{x}_0^{(k)} \leftarrow$ ODE-Solve$(x_{t_1}^{(k)}, \sigma_{t_1} \to 0)$
    end if
end for
return $\{\hat{x}_0^{(k)}\}_{k=1}^K$

A.2 Toy Experiment Details

Prior. The signals x are elements of $\mathbb{R}^2$. We define a reference vector u, in this case the first standard basis vector $u = e_1$. The prior has two atoms at $\pm au$ for scalar $a > 0$:

$$p(x_0) = \tfrac{1}{2}\,\delta_{-au} + \tfrac{1}{2}\,\delta_{au}, \quad a = 1. \tag{14}$$

Observe that the prior denoiser is known exactly for EDM-parametrized forward model $x_\sigma = x_0 + \sigma\epsilon$ with $\epsilon \sim \mathcal{N}(0, I)$:

$$D(x, \sigma) = \mathbb{E}[x_0 \mid x_\sigma = x] = a \tanh\big(a\langle x, u\rangle/\sigma^2\big)\, u. \tag{15}$$

DAPS and PC-DAPS use this closed-form denoiser in place of a trained network, both for the probability-flow ODE and for all Tweedie estimates.

Measurement and likelihood. The measurement operator is a quadratic function in the inner product:

$$\mathcal{A}(x) = f(\langle x, u\rangle), \quad f(z) = \frac{c}{2ab}\,(z-a)(z+a-b), \tag{16}$$

chosen to interpolate $f(-a) = c$, $f(a+b) = c$ and $f(a) = 0$. Measurements follow $y = \mathcal{A}(x_0) + n$ with $n \sim \mathcal{N}(0, \beta_y^2)$, so the likelihood is $p(y \mid x_0) = \mathcal{N}(y; \mathcal{A}(x_0), \beta_y^2)$. The observation is set to $y = c = \kappa\beta_y$. The precise values used in Figure 2 are $\beta_y = 0.05$, $b = 0.1$, $\kappa = 2$ (thus $c = 0.1$).

Exact posterior. By Bayes' rule, the ground truth posterior $p(x_0 \mid y)$ is accessible in closed form:

$$p(x_0 \mid y) = (1 - \pi)\,\delta_{-au} + \pi\,\delta_{au}, \quad \pi = \frac{p(y \mid au)}{p(y \mid -au) + p(y \mid au)}. \tag{17}$$

Observe that although $\mathcal{A}((a+b)u) = \mathcal{A}(-au)$, $(a+b)u$ has no prior mass, thus samples from (the neighborhood of) $(a+b)u$ are prior-inconsistent hallucinations.

A.3 More Details on Gaussian Mixture Approximation

To illustrate the role of the Gaussian mixture $q(x_0 \mid x_t)$ introduced in Section 4.2, used by PC-DAPS to approximate the denoising distribution $p(x_0 \mid x_t)$, we show a side-by-side comparison of the Langevin dynamics refinement when using (i) the DAPS Gaussian approximation $\mathcal{N}(x_0; \hat{x}_0(x_t), r_t^2 I)$ and (ii) the PC-DAPS GMM approximation $q(x_0 \mid x_t)$ on a toy example.

In the illustrated example depicted in Figure 4, we consider a multimodal prior distribution on the plane $\mathbb{R}^2$, and construct $\mathcal{A}$ to simply select the second component x_2 of the signal vector x. Four panes are shown depicting the Langevin algorithm procedure of a single (i) DAPS iteration, (ii) PC-DAPS iteration using the DAPS Gaussian approximation, (iii) PC-DAPS iteration using the PC-DAPS GMM approximation, and (iv) ground truth (denoiser and Langevin) iteration. The high-probability level sets of both the (approximated) denoiser and the induced Langevin target distribution are depicted in each pane.

Observe that, in both panes using the DAPS Gaussian approximation, the Langevin algorithm effectively searches for points (i) consistent with the provided measurement $y = x_2$ and (ii) close to the initialization point in a Euclidean sense. As a result, the algorithm terminates by returning a likelihood-consistent yet posterior-inconsistent point. In contrast, the Gaussian mixture approximation is attracted to the region of the measurement's preimage that is consistent with the prior thanks to other particles in the bag which explored and identified prior-consistent regions distant from the Langevin initialization. Additionally, observe that although the DAPS Gaussian is isotropic and thus places high probability outside the high-probability level set of the prior, the PC-DAPS approximation's component covariances better adapt to the geometry of the prior's support.

A.4 Additional implementation choices

There are two implementation choices that are orthogonal to the core PC-DAPS algorithm but improve reconstruction quality or computational efficiency in practice: skipping the final Langevin refinement and using adaptive ODE stepping. Their effects are summarized in Table 3 and described in detail below.

Skipping the last Langevin refinement. At the final annealing time ($t_0 = 0$), the Langevin update is not performed, and the ODE endpoint $\hat{x}_0(x_{t_0})$ is used as the output. When the noise is nearly zero, the Langevin step size becomes very small, and the added noise can slightly reduce sample quality.

Adaptive ODE steps. To compute the DAPS clean estimate $\hat{x}_0(x_{t_i})$, the probability flow ODE is integrated using Euler steps. One Euler step is used at high noise levels, where the clean estimate is less sensitive to the precise ODE integration, and this is increased to five steps at lower noise levels, where the prior becomes more discriminative and finer integration reduces discretization error near the data manifold. Since PC-DAPS utilizes a bag of multiple particles, reducing unnecessary ODE evaluations is important for controlling the additional computational cost. Adaptive ODE stepping is therefore used to improve the time efficiency of the particle-based sampler.

Descriptions of the implementation details of the particle clustering and pruning procedure used by PC-DAPS are given below.

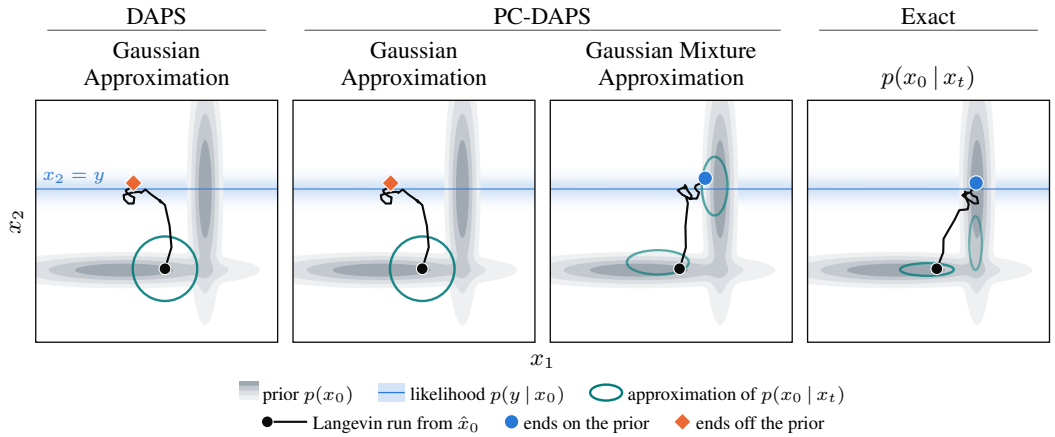


Figure 4: **Gaussian mixture approximations preserve prior-supported modes in Langevin refinement.** Each pane illustrates N iterations of Langevin refinement from the same starting point $\hat{x}_0$ (black dot; path smoothed for readability). The same prior $p(x_0)$ (gray) and likelihood $p(y \mid x_0)$ (blue) are shown in each pane, while the approximation $q(x_0 \mid x_t)$ (teal) differs. When utilizing the Gaussian approximation, both DAPS (left) and PC-DAPS (middle-left) produce Langevin targets which are measurement-consistent yet prior-inconsistent, and the refinement ends outside the prior (orange diamond). In contrast, the PC-DAPS Gaussian mixture (middle-right) represents multiple prior-supported regions explored by particles in the bag, allowing Langevin refinement to move toward the prior-supported mode associated with the measurement (blue dot). The exact denoising distribution (right) exhibits the same multimodal structure.

Clustering technique. To assign particles to components, the corresponding clean samples $\{\tilde{x}_0^{(k)}\}_{k=1}^K$ described in Section 4.2 are downsampled to a coarse resolution (e.g., 32×32), centered, and projected onto the leading principal components that explain 90% of the variance. On this subspace, a Gaussian mixture model is fit with M selected via Bayesian information criterion (Schwarz, 1978). Each particle is then hard-assigned to its most likely component, giving a partition $\{\mathcal{C}_m\}_{m=1}^M$ of the bag.

Pruning threshold. Pruning is applied to the post-Langevin samples only when the test in Equation (8) fires ($r^\star > \bar{r}$) and the clustering returns more than one cluster. Each cluster is scored using the non-positive function Equation (9) with $\mathcal{S} = \{0.5, 1, 2\}$. Denoting $s^\star = \max_m s(\mathcal{C}_m)$, all clusters $\mathcal{C}_m$ satisfying $s(\mathcal{C}_m) < (1+\gamma)s^\star$ are pruned, while clusters satisfying $(1+\gamma)s^\star \le s(\mathcal{C}_m) \le s^\star$ are preserved. Each pruned particle is replaced by a copy of a surviving particle, cycling through the survivors. Observe that the cluster producing score $s^\star$ always survives. γ is calibrated from within-cluster score variation: following Zhang et al. (2025), 10 calibration images per dataset are denoised via PC-DAPS (including the clustering procedure) without pruning. The relative average score difference for all particle pairs within each cluster is computed, and γ is set to the 95th percentile of these differences.

Ablations of the main computational parameters and their comparisons are presented in the following paragraphs.

Ablation of ODE steps and annealing noise scheduling steps. PC-DAPS benefits from the same implementation improvements identified in DAPS. The effects of the number of ODE steps used to compute the clean estimate and the number of annealing noise-scheduling steps are ablated, and trends consistent with those reported for DAPS are observed: increasing the number of ODE steps generally improves perceptual quality, while denser annealing schedules are particularly beneficial for nonlinear inverse problems. The corresponding ablations are shown in Figure 5 and Figure 6.

Number of particles in a bag. Using more particles provides the selection procedure with more candidate trajectories, which improves its ability to distinguish distinct modes and reject prior-inconsistent but measurement-consistent components. On phase retrieval, increasing the bag size

Table 3: **Comparison of the ODE schedule and final Langevin refinement**. Ablated over 10 randomly selected FFHQ images using the PC-DAPS-1K configuration in phase retrieval task.

ODE steps	Final Langevin	Mean reconstruction			Oracle selection		
		PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
5	Included	29.50	0.787	0.168	29.63	0.789	0.167
	Dropped	30.01	0.831	0.144	30.14	0.833	0.143
1	Included	25.02	0.692	0.274	28.43	0.780	0.188
	Dropped	25.41	0.739	0.254	28.93	0.831	0.164
1→5	Included	28.11	0.774	0.181	28.34	0.778	0.179
	Dropped	29.83	0.818	0.158	29.99	0.822	0.149

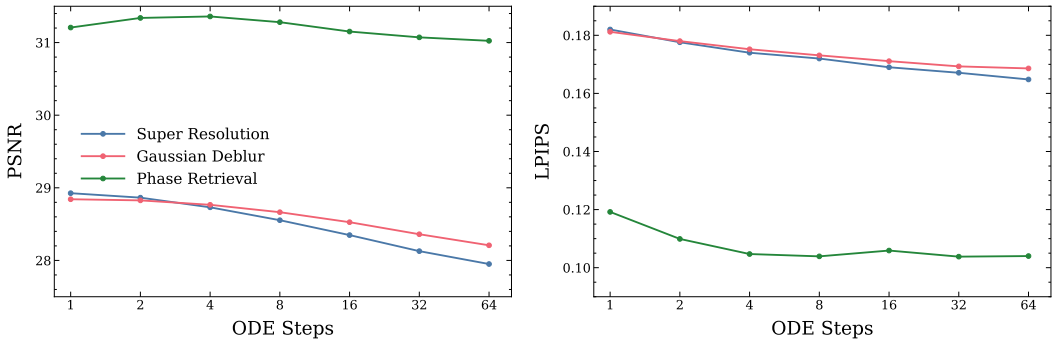


Figure 5: **Effect of the number of ODE steps on PC-DAPS.** Increasing the number of ODE steps generally improves perceptual quality (LPIPS) but can reduce PSNR, with the effect depending on the inverse problem.

substantially improves both PSNR and LPIPS, with the gains becoming smaller beyond $K = 8$–16. This improvement comes at an approximately linear increase in computational cost, since all K particles are propagated through the annealing schedule. Compared with DAPS best-of-k, PC-DAPS achieves higher reconstruction quality while using less sampling time for larger bag sizes, as shown in Figure 7.

A fair comparison on phase retrieval. Table 1 reports the best-of-k results used by the baseline methods, while PC-DAPS is evaluated using the mean over its K samples. For a direct comparison under the same evaluation protocol, Table 4 reports mean-of-k results for all methods. The baseline methods show lower mean performance because their k reconstructions can include measurement-consistent but prior-inconsistent solutions, such as color inversions, rotations, or translations, whereas best-of-k selects only the reconstruction closest to the ground truth.

A.5 Time Cost

PC-DAPS keeps K particles in one sampling trajectory, while DAPS needs K separate runs to get K candidate reconstructions. If implementation overhead is ignored, both methods need about the same number of denoiser evaluations, which increases roughly linearly with K. PC-DAPS does have extra steps, including clustering, estimating mixture parameters, and a one-shot pruning step. However, these add only a small overhead compared to the cost of the ODE and Langevin updates.

The implementation choices in Appendix A.4, such as skipping the final Langevin refinement and using adaptive ODE steps at low noise, help lower the sampling cost of PC-DAPS. Figure 8 shows the measured wall-clock time and reconstruction quality. For super-resolution, PC-DAPS takes longer than DAPS because it uses multiple particles. For phase retrieval, though, the DAPS runtime includes four separate runs because it uses a best-of-4 evaluation protocol. PC-DAPS, on the other hand, creates the particle set in one run and selects internally. These results suggest that using

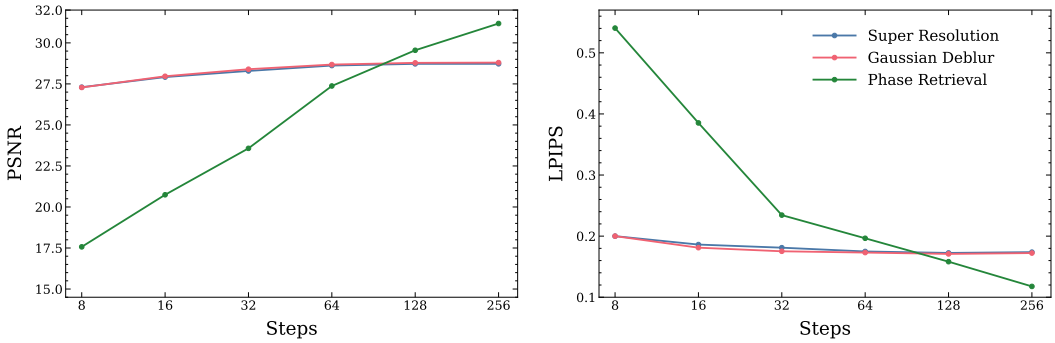


Figure 6: **Effect of the number of annealing noise scheduling steps on PC-DAPS.** Increasing the number of annealing steps has little effect on super-resolution and Gaussian deblurring, while phase retrieval shows substantial improvement with a finer annealing schedule.

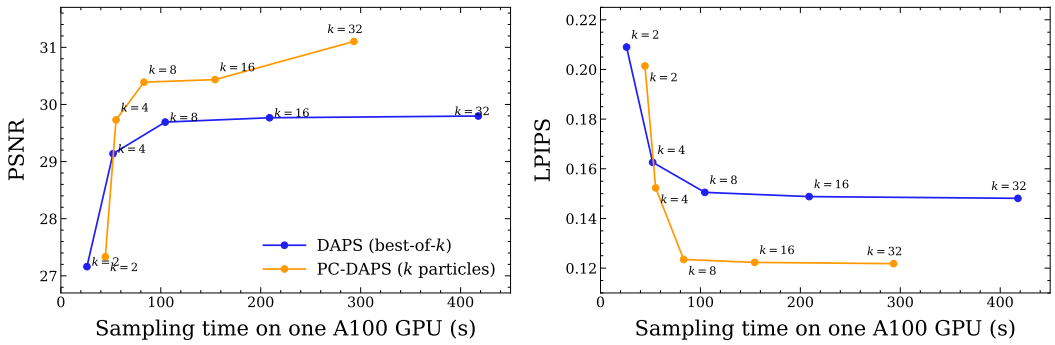


Figure 7: **Effect of the number of particles in a bag on phase retrieval.** Increasing the bag size improves PC-DAPS reconstruction quality, with both PSNR and LPIPS improving rapidly up to $K = 8$–16 and then beginning to saturate. PC-DAPS is competitive with DAPS best-of-k at comparable values of k, while requiring less sampling time for larger k.

multiple particles does not always mean a much higher wall-clock cost to achieve the same quality metric.

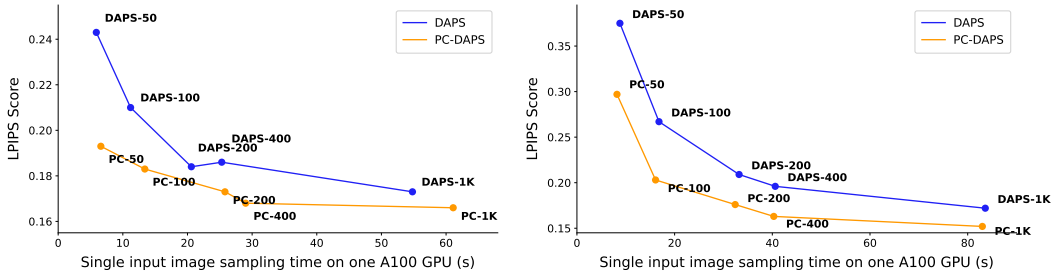
A.6 Inverse Problem Setup

In Table 1, following the inverse-problem settings of DAPS (Zhang et al., 2025), the benchmark includes the eight tasks from DAPS: $4\times$ super-resolution, box inpainting, random inpainting, Gaussian deblurring, motion deblurring, nonlinear deblurring, phase retrieval, and high dynamic range (HDR) reconstruction. For these tasks, the same measurement operators, measurement noise level, and evaluation protocol as DAPS are used. Unless otherwise stated, measurements are corrupted with additive Gaussian noise with standard deviation $\beta_y = 0.05$. Methods are evaluated on 100 test images from each dataset, FFHQ and ImageNet, at a resolution of 256×256. Additionally, $16\times$ super-resolution and high-pass filtering from Thaker et al. (2025) are included in the table and described in detail below.

Super-resolution $16\times$**.** Bicubic downsampling is applied from 256×256 to 16×16, followed by additive Gaussian noise with standard deviation $\beta_y = 0.05$. This uses the same linear bicubic downsampling operator as the $4\times$ super-resolution task in DAPS, but with a larger downsampling factor. The resulting operator has a larger null space, leaving more image content unconstrained by the measurement and providing a more challenging setting. All other sampler settings follow the $4\times$ super-resolution configuration.

Table 4: **Phase retrieval under a mean-of-k evaluation.** All methods are evaluated using the mean performance across k reconstructions, rather than selecting the reconstruction closest to the ground truth. This exposes failures that best-of-k evaluation can mask, since individual reconstructions may be measurement-consistent but differ from the ground truth due to color inversion, rotation, or translation. PC-DAPS achieves higher reconstruction quality compared to other baselines.

Task	Method	FFHQ			ImageNet		
		PSNR (↑)	SSIM (↑)	LPIPS (↓)	PSNR (↑)	SSIM (↑)	LPIPS (↓)
Phase retrieval	PC-DAPS	**30.09**$_{\pm 1.77}$	**0.834**$_{\pm 0.089}$	**0.152**$_{\pm 0.088}$	**21.75**$_{\pm 1.10}$	**0.469**$_{\pm 0.132}$	**0.373**$_{\pm 0.128}$
	DAPS	25.11$_{\pm 5.73}$	0.666$_{\pm 0.141}$	0.283$_{\pm 0.138}$	15.70$_{\pm 3.90}$	0.317$_{\pm 0.153}$	0.516$_{\pm 0.101}$
	DPS	14.13$_{\pm 2.72}$	0.398$_{\pm 0.106}$	0.508$_{\pm 0.085}$	13.38$_{\pm 4.03}$	0.311$_{\pm 0.208}$	0.560$_{\pm 0.123}$
	PnP-DM	17.68$_{\pm 5.50}$	0.458$_{\pm 0.191}$	0.523$_{\pm 0.149}$	14.80$_{\pm 5.60}$	0.382$_{\pm 0.170}$	0.570$_{\pm 0.110}$
	DPnP	16.57$_{\pm 6.76}$	0.479$_{\pm 0.213}$	0.499$_{\pm 0.193}$	14.69$_{\pm 3.61}$	0.427$_{\pm 0.180}$	0.584$_{\pm 0.101}$
	RED-Diff	14.30$_{\pm 3.90}$	0.347$_{\pm 0.180}$	0.630$_{\pm 0.095}$	13.90$_{\pm 3.30}$	0.354$_{\pm 0.070}$	0.558$_{\pm 0.105}$



(a) 4 × super-resolution (b) Phase retrieval

Figure 8: **Comparison of single input sampling time and reconstruction quality between DAPS and PC-DAPS.** The x-axis shows the sampling time for a single input sample measured on an NVIDIA A100-SXM4-80GB GPU, while the y-axis reports LPIPS. Evaluation is performed on 100 FFHQ images. For phase retrieval, DAPS requires four independent runs for an input image; therefore, the reported DAPS sampling time corresponds to the total time of all four runs.

High-pass filtering. Following Thaker et al. (2025), a 61×61 convolutional kernel is used, constructed as a Dirac delta minus a Gaussian blur kernel with standard deviation 5.0. The resulting kernel has zero sum and therefore removes the DC component. Gaussian noise with standard deviation $\beta_y = 0.05$ is added to the filtered image.

A.6.1 Baseline Details

For DAPS (Zhang et al., 2025), DPS (Chung et al., 2023), DDNM (Wang et al., 2023), FPS-SMC (Dou & Song, 2024), RED-Diff (Mardani et al., 2024), FGPS (Thaker et al., 2025), PnP-DM (Wu et al., 2024), and DPnP (Xu & Chi, 2024), the authors' official public implementations and recommended task-specific hyperparameters are used throughout the benchmark. DAPS uses the DAPS-4K configuration for nonlinear tasks and DAPS-1K for linear tasks. For consistency in computational budget, PC-DAPS uses the DAPS-1K configuration for both linear and nonlinear tasks, including its adaptive variant across all tasks. When a baseline specifies task-specific step sizes or noise-schedule parameters, the reported values are used without modification. HIO, a classical phase retrieval method, is included as an additional reference baseline, with results taken from Zhang et al. (2025).

For the two additional tasks not included in the DAPS benchmark, $16\times$ super-resolution and high-pass filtering, the closest corresponding configuration available for each baseline is used: the $4\times$ super-resolution configuration for $16\times$ super-resolution and the Gaussian deblurring configuration for high-pass filtering. Only the measurement operator and the associated measurement generation are modified to match the task definitions described above.

Unless otherwise specified, PC-DAPS uses the 1K setting with 200 annealing steps and adaptive ODE stepping, and generates $K = 4$ reconstructions for each measurement. Reported PC-DAPS metrics are averaged over the four reconstructions. For phase retrieval, baselines other than PC-

Table 5: **Sampling configurations.** ODE steps, annealing steps and number of function evaluations (NFE) for the DAPS and PC-DAPS configurations used in the ablation studies. PC-DAPS uses a bag of $K = 4$ particles; $1{\to}k$ means one Euler step for the first 65% of the annealing steps and k steps afterwards. NFE is the number of denoiser evaluations per sample (per run for DAPS, per particle for PC-DAPS), counted during sampling. PC-DAPS configurations are named after the DAPS configuration with the same annealing steps and final ODE steps; they use fewer ODE steps early on but one extra denoiser evaluation per annealing step to fit the mixture, so their NFE differs from the name.

Configuration	**DAPS-50**	**DAPS-100**	**DAPS-200**	**DAPS-400**	**DAPS-1K**	**DAPS-2K**	**DAPS-4K**
ODE Steps	2	2	2	4	5	8	10
Annealing Steps	25	50	100	100	200	250	400
NFE	50	100	200	400	1000	2000	4000
Configuration	**PC-DAPS-50**	**PC-DAPS-100**	**PC-DAPS-200**	**PC-DAPS-400**	**PC-DAPS-1K**		
ODE Steps	$1{\to}2$	$1{\to}2$	$1{\to}2$	$1{\to}4$	$1{\to}5$		
Annealing Steps	25	50	100	100	200		
NFE	66	131	261	331	732		

DAPS are evaluated using the best of four independent runs, selecting the reconstruction closest to the ground-truth image, following prior work (Chung et al., 2023; Zhang et al., 2025). This protocol gives the baselines access to ground-truth information that is unavailable to PC-DAPS at test time. PSNR, SSIM (Wang et al., 2004), and LPIPS (Zhang et al., 2018) are reported in Table 1 for all methods. For ablation studies, the PC-DAPS variants follow the corresponding DAPS settings reported in Table 5.

A.7 More Qualitative Samples

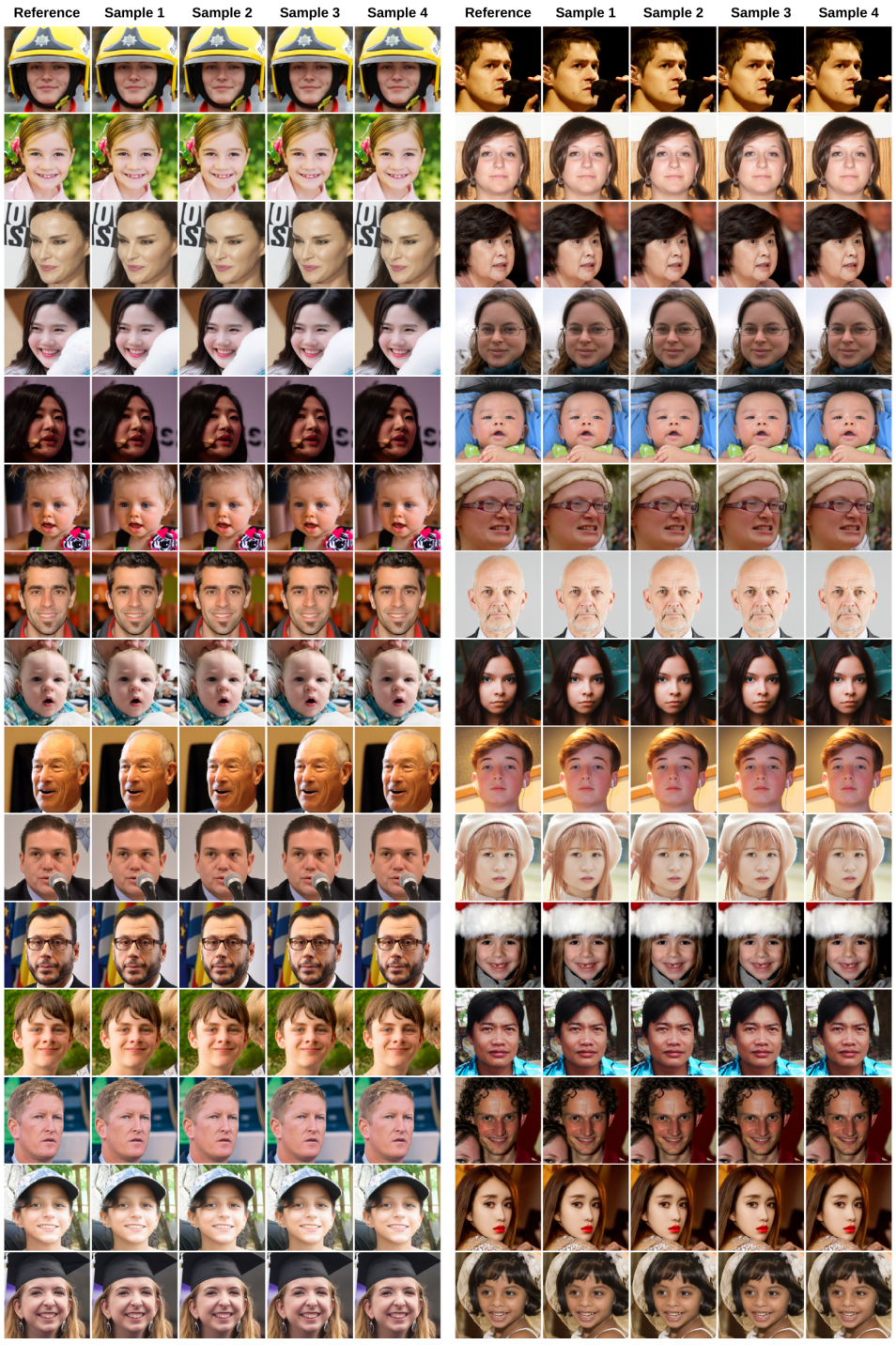


Figure 9: **PC-DAPS samples for phase retrieval on FFHQ.** Samples are shown without manual selection to better visualize success rate and sample quality.

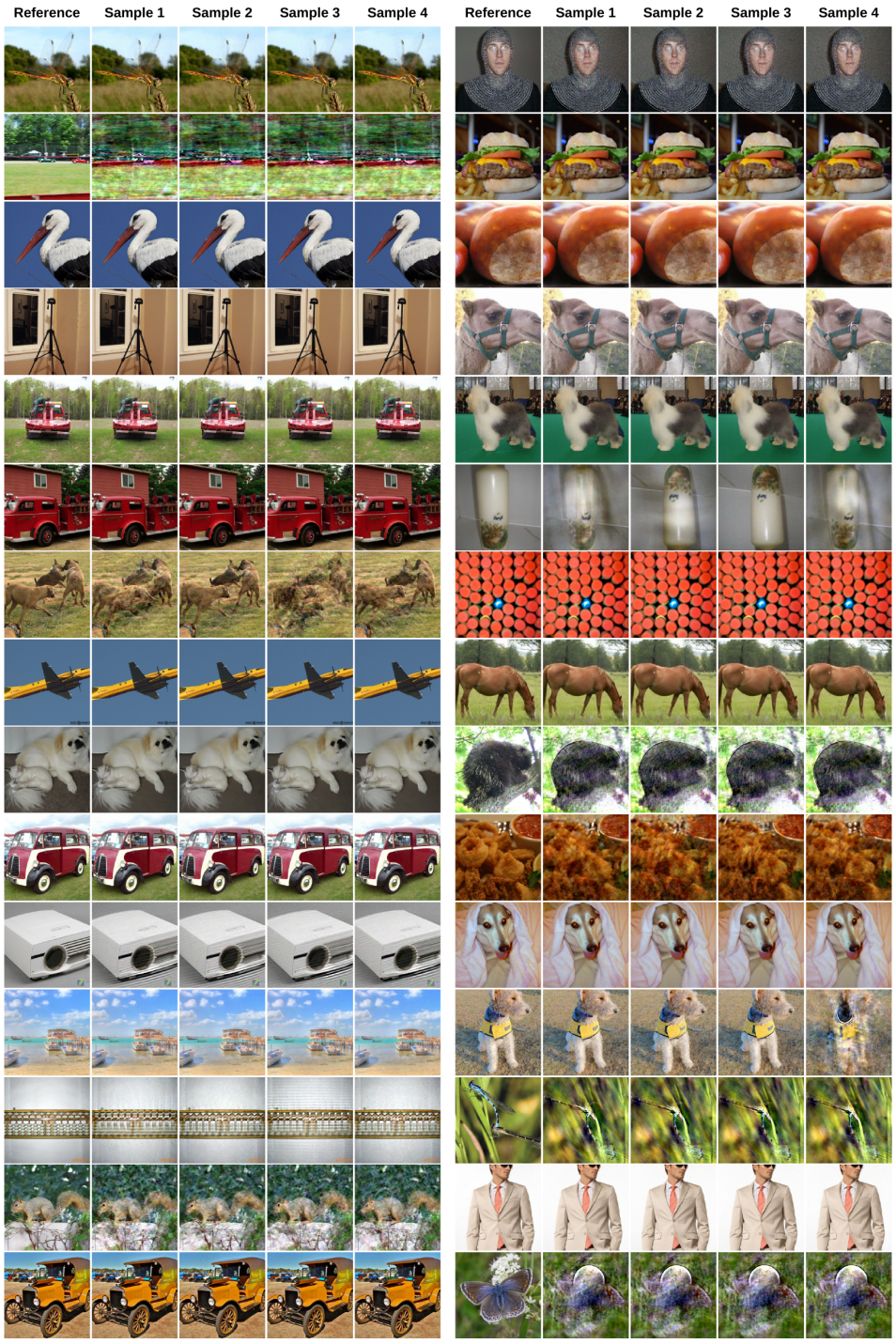


Figure 10: **PC-DAPS samples for phase retrieval on ImageNet.** Samples are shown without manual selection to better visualize the success rate and sample quality. As discussed in Section 5, ImageNet samples in a bag still contain failed (diverged/noisy) particles and underperform compared to FFHQ.

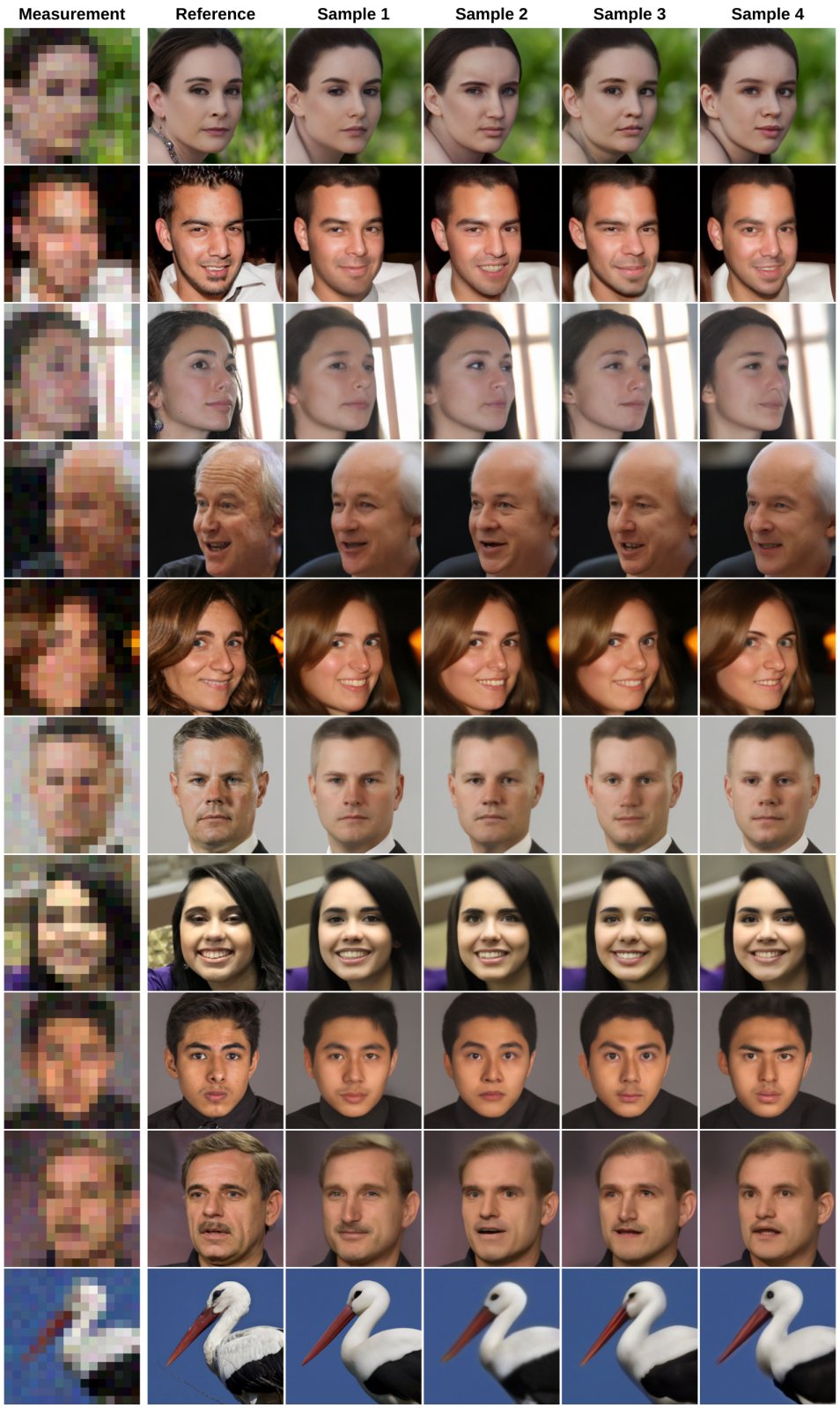


Figure 11: **PC-DAPS samples for** $16\times$ **super-resolution.** PC-DAPS is able to keep the diversity in the multimodal posterior distribution.

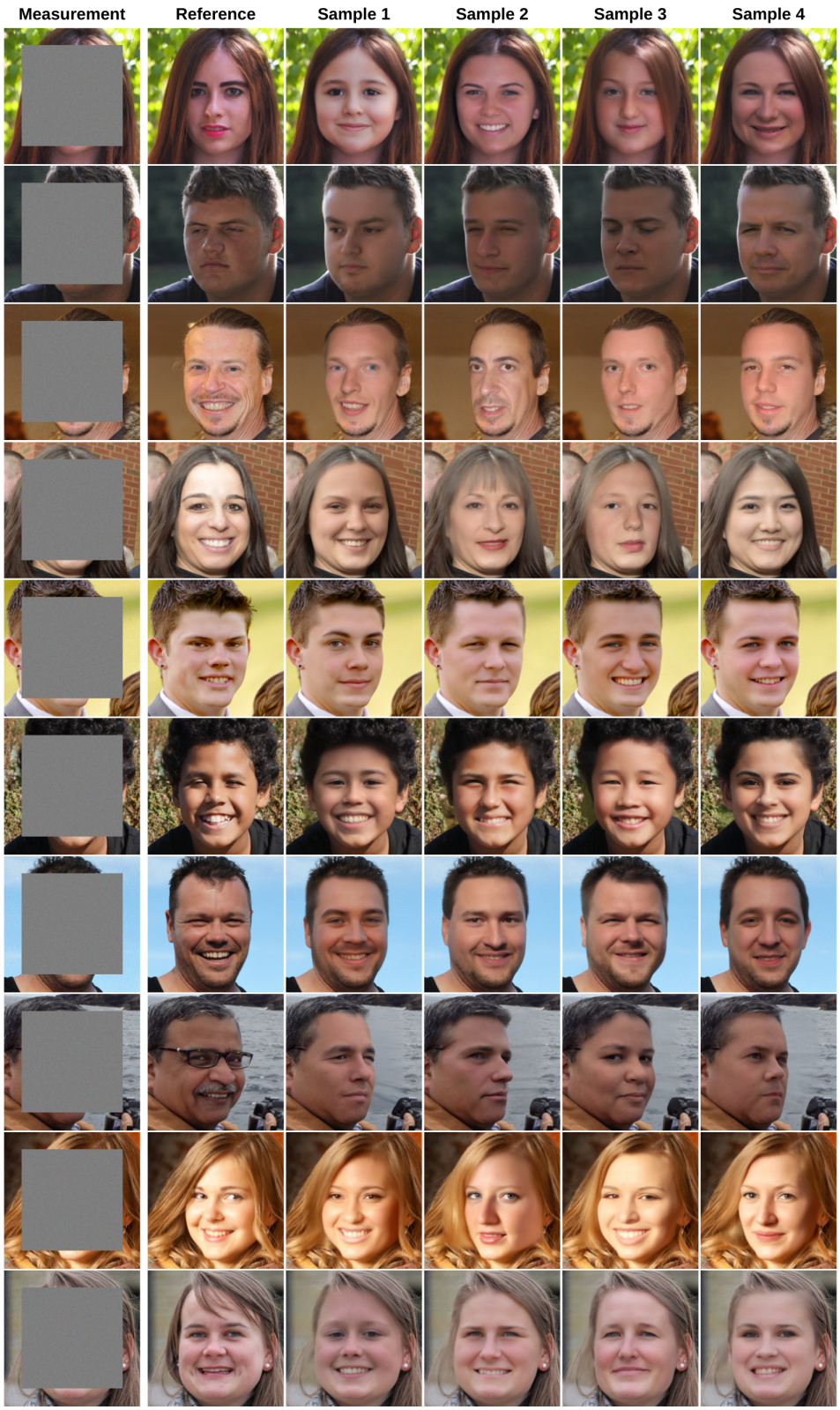
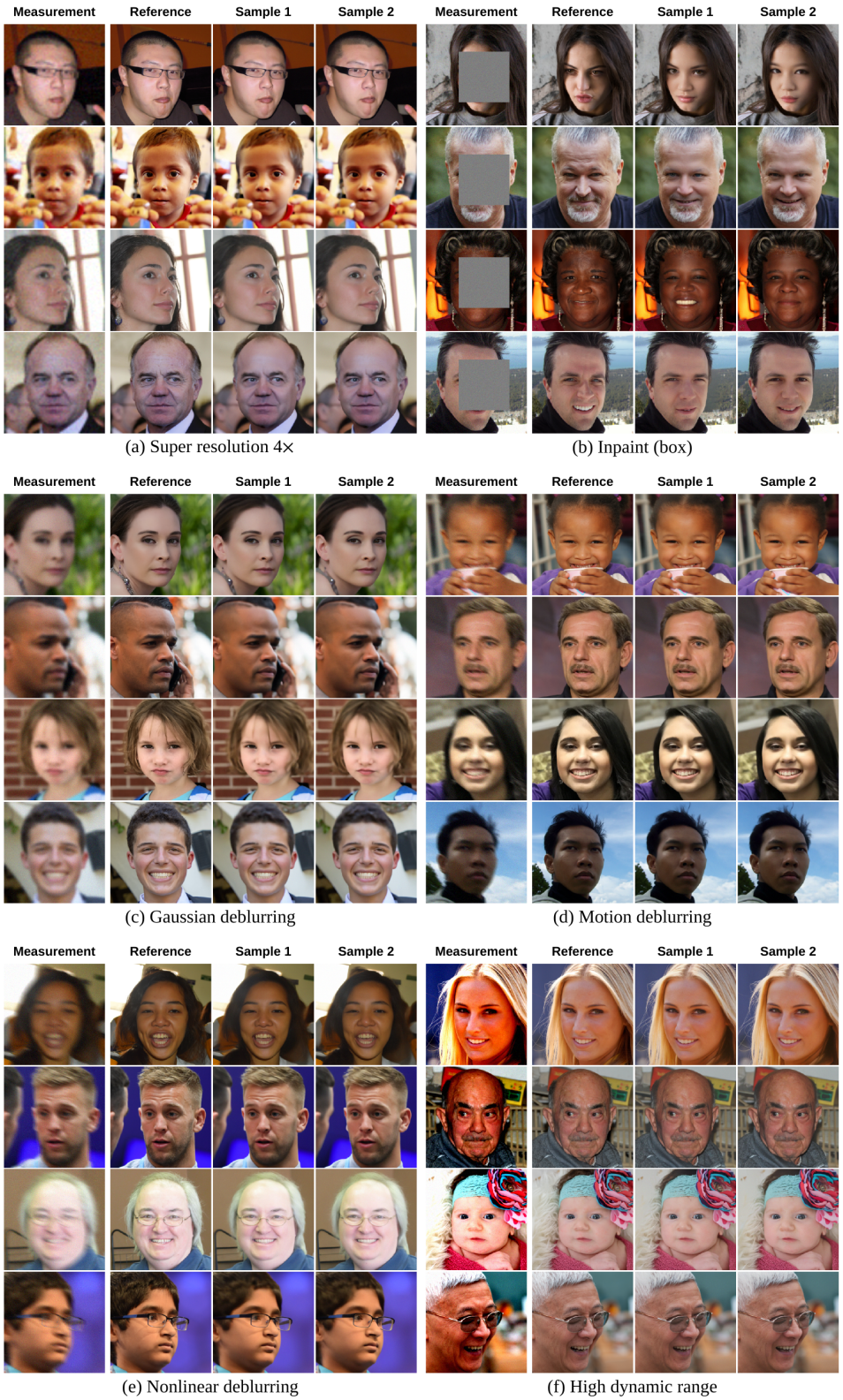


Figure 12: **PC-DAPS samples for** 190×190 **box inpainting.** PC-DAPS is able to keep the diversity in the multimodal posterior distribution.

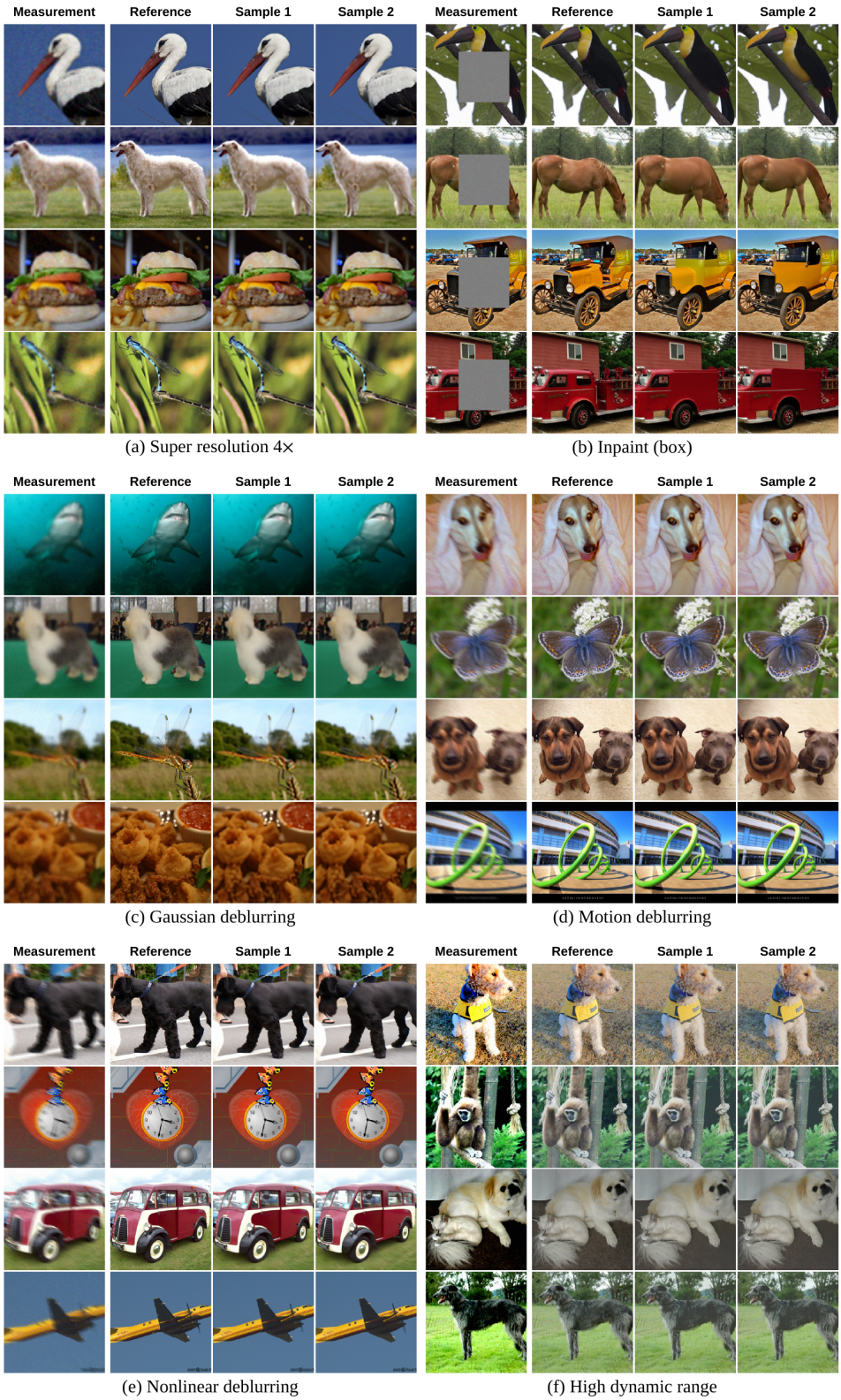


(a) Super resolution 4× (b) Inpaint (box)

(c) Gaussian deblurring (d) Motion deblurring

(e) Nonlinear deblurring (f) High dynamic range

Figure 13: **PC-DAPS samples for various tasks on FFHQ.** PC-DAPS obtains competitive reconstructions for linear and nonlinear tasks.



(a) Super resolution 4× (b) Inpaint (box)

(c) Gaussian deblurring (d) Motion deblurring

(e) Nonlinear deblurring (f) High dynamic range

Figure 14: **PC-DAPS samples for various tasks on ImageNet.** PC-DAPS obtains competitive reconstructions for linear and nonlinear tasks.