SPIR: SYMMETRIC-PROBE IMAGE RESTORATION WITH FLOW-MATCHING PRIORS

Anonymous authors
Paper under double-blind review

ABSTRACT

Flow-matching priors provide clean-image estimates from degraded inputs, but a limited evaluation budget leaves a choice between a single central prediction and aggregation over a noisy neighborhood. We introduce Symmetric-Probe Image Restoration (SPIR), which averages predictions from opposite Gaussian probes and applies a data-consistency correction. The symmetric construction excludes odd perturbation responses while retaining nonlinear neighborhood information. Reusing the probes keeps the sampled prior estimator fixed within each time step. Our analysis identifies conditions under which symmetric probes improve estimation and characterizes the stability of data correction. Experiments on four datasets and five restoration tasks demonstrate competitive fidelity and perceptual quality, with task-dependent trade-offs.

1 INTRODUCTION

Recovering an unknown image $\mathbf{x} \in \mathbb{R}^d$ from degraded observations is a fundamental problem in computational imaging. We consider the linear measurement model

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

where $\mathbf{x} \in \mathbb{R}^d$ is the unknown image, $\mathbf{y} \in \mathbb{R}^{d_y}$ is the observation, and $\mathbf{H} : \mathbb{R}^d \to \mathbb{R}^{d_y}$ is a potentially ill-conditioned or rank-deficient measurement operator. The noise $\mathbf{n}$ has standard deviation σ_y, and $\mathbf{I}$ is the identity on measurement space. Image priors help resolve ambiguity and stabilize restoration. Maximum a posteriori (MAP) estimation takes the form

$$\widehat{\mathbf{x}} \in \arg\min_{\mathbf{x}} \underbrace{\frac{1}{2\sigma_y^2} \|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2}_{\ell_{\mathbf{y}}(\mathbf{x})} - \log p_{\mathrm{data}}(\mathbf{x}), \tag{2}$$

where $\widehat{\mathbf{x}}$ is a MAP estimate, $\ell_{\mathbf{y}}$ is the Gaussian data-fidelity term, and p_{data} is the clean-image density. Classical priors encode image structure through total variation (Rudin et al., 1992), wavelets (Mallat, 1999), or sparsity (Donoho, 2006). Learning-based restoration methods such as SRCNN (Dong et al., 2016) and DnCNN (Zhang et al., 2017) replace hand-crafted models with data-driven mappings, but are susceptible to perturbations (Antun et al., 2020). The central challenge is to exploit an expressive learned prior while preserving measurement consistency.

Plug-and-Play methods (Venkatakrishnan et al., 2013; Romano et al., 2017) address this challenge by inserting a pretrained denoiser into an iterative optimization procedure. Diffusion and score-based models learn expressive image distributions (Ho et al., 2020; Song et al., 2021; Dhariwal & Nichol, 2021) and support effective zero-shot inverse solvers (Chung et al., 2023; Song et al., 2023), although their iterative sampling requires repeated network evaluations. Flow matching (FM) (Lipman et al., 2023), rectified-flow (Liu et al., 2023) and stochastic-interpolant formulations (Albergo & Vanden-Eijnden, 2023) instead learn a velocity field that transports a simple source distribution to the data distribution. In addition to defining an ODE sampler, a flow trained with straight-line interpolation provides a single-evaluation estimate of the data at each interpolation time. A pretrained flow can thus define a time-indexed family of restoration operators without full ODE integration.

Recent flow-based inverse solvers use destination predictions as time-dependent denoisers, trajectory guides, or targets for measurement refinement (Pokle et al., 2024; Zhang et al., 2024; Martin

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.

et al., 2025; Pourya et al., 2026). Under a limited model-evaluation budget, a central question is how to construct the prior estimate. A center prediction requires only one evaluation but does not capture the predictor's responses to Gaussian perturbations. Averaging predictions from independently perturbed inputs estimates the expected response under Gaussian noise. With only a few evaluations, however, this estimate is sensitive to the sampled noise. Rather than simply increasing the number of model evaluations, we investigate whether structured perturbations improve estimation at a fixed number of prior evaluations and benefit restoration after data-consistency correction.

We introduce Symmetric-Probe Image Restoration (SPIR), which constructs a prior estimate from symmetric Gaussian perturbations and refines it through data-consistency correction. We refer to the Gaussian noise vectors used to perturb the flow inputs as probes, and pair each probe with its negative following classical antithetic sampling (Hammersley & Morton, 1956; Ren et al., 2019). This construction eliminates local first-order perturbation effects while retaining nonlinear information. The probes are reused within each time step. SPIR uses a frozen flow-matching prior without ODE integration or network backpropagation. Our contributions are summarized as follows:

1. **A symmetric-probe restoration procedure.** SPIR forms prior estimates by aggregating symmetric-probe destination predictions, reuses the probes within each time step, and corrects the estimates using the observation under a limited model-evaluation budget.
2. **A finite-sample mechanism analysis.** We characterize the center-to-expectation discrepancy and sampling variance under Gaussian perturbations, identify when symmetric probes improve estimation with the same number of prior evaluations, and quantify the stability of exact and inexact data correction.
3. **Systematic experimental evaluation.** We combine multi-dataset restoration benchmarks with a controlled prior-estimation ablation and broader parameter-sensitivity analyses to evaluate SPIR and characterize its practical behavior.

The rest of the paper is organized as follows: Section 2 introduces flow matching and destination estimation. Section 3 develops the motivation, algorithm, and finite-sample analysis. Section 4 reviews related inverse solvers. Section 5 presents benchmarks, a prior-estimation ablation, and sensitivity analysis. Section 6 concludes the paper.

2 Background and Problem Setup

2.1 Conditional flow matching

Flow matching (Lipman et al., 2023) learns a time-dependent velocity field $\mathbf{v}^\theta : \mathbb{R}^d \times [0,1] \to \mathbb{R}^d$ whose ODE transports a source distribution $p_{\mathbf{X}_0}$ to a target distribution $p_{\mathbf{X}_1}$. Conditional flow matching (Tong et al., 2024) avoids direct evaluation of the generally intractable marginal velocity. Given a coupling $\pi(\mathbf{x}_0, \mathbf{x}_1)$ between the marginals, consider the straight interpolation

$$\mathbf{X}_t = t\mathbf{X}_1 + (1-t)\mathbf{X}_0, \qquad t \in (0,1), \tag{3}$$

where $\mathbf{X}_0$ and $\mathbf{X}_1$ are source and target random variables coupled by π, and $\mathbf{X}_t$ is the interpolated state. The corresponding conditional velocity is $\mathbf{X}_1 - \mathbf{X}_0$, yielding the regression objective

$$\mathcal{L}_{\mathrm{CFM}}(\theta) = \mathbb{E}_{t,(\mathbf{X}_0,\mathbf{X}_1)\sim\pi}\left[\left\|\mathbf{v}^\theta(\mathbf{X}_t,t) - (\mathbf{X}_1 - \mathbf{X}_0)\right\|_2^2\right], \tag{4}$$

where $\mathcal{L}_{\mathrm{CFM}}$ is the mean-squared velocity loss, θ denotes the velocity-network parameters, and the expectation is over the coupling π and the training-time sampling law for t. The ideal minimizer of the expected loss in Equation 4 satisfies

$$\mathbf{v}^\star(\mathbf{x}_t, t) = \mathbb{E}[\mathbf{X}_1 - \mathbf{X}_0 \mid \mathbf{X}_t = \mathbf{x}_t], \tag{5}$$

where $\mathbf{v}^\star$ is the optimal velocity field for this expected loss. The coupling π determines the regression problem. Independent coupling, $\pi = p_{\mathbf{X}_0} \otimes p_{\mathbf{X}_1}$, admits a Gaussian-denoising interpretation when $p_{\mathbf{X}_0} = \mathcal{N}(\mathbf{0}, \mathbf{I})$. Optimal transport provides a principled coupling of the marginals (Peyré & Cuturi, 2019), whereas minibatch couplings can reduce regression variance and straighten the resulting flows (Pooladian et al., 2023; Tong et al., 2024). The conditional-mean identity below holds for this ideal velocity under any straight-line coupling.

Straight conditional interpolation paths do not imply globally straight trajectories of the learned marginal ODE. Rectification (Liu, 2022), curvature minimization (Lee et al., 2023), velocity-consistency training (Yang et al., 2024), and optimal flow matching (Kornilov et al., 2024) address generative-flow geometry.

2.2 Flow destinations

A velocity field trained on the interpolation in Equation 3 defines the single-step destination estimator used in flow-based restoration (Martin et al., 2025; Pourya et al., 2026):

$$\mathcal{D}_t^\theta(\mathbf{x}_t) := \mathbf{x}_t + (1-t)\mathbf{v}^\theta(\mathbf{x}_t, t), \tag{6}$$

where $\mathcal{D}_t^\theta$ maps an intermediate state to its estimated clean destination. Substituting Equation 5 into Equation 6 and applying the straight interpolation yields

$$\mathcal{D}_t^\star(\mathbf{x}_t) = \mathbb{E}[\mathbf{X}_1 \mid \mathbf{X}_t = \mathbf{x}_t], \tag{7}$$

where $\mathcal{D}_t^\star$ denotes the destination estimator induced by the ideal velocity $\mathbf{v}^\star$. The ideal estimator therefore returns a conditional mean. Equation 7 does not imply the pointwise identity $\mathcal{D}_t^\star(t\mathbf{x}_1 + (1-t)\boldsymbol{\epsilon}) = \mathbf{x}_1$ for every $\mathbf{x}_1$ and perturbation $\boldsymbol{\epsilon}$, which would require a degenerate conditional distribution. Aggregating a finite collection of destination predictions consequently yields a probe-dependent prior estimate, not an exact image or an observation-conditioned solution.

3 Method

3.1 Prior estimation under a model-evaluation budget

Under the observation model in Equation 1, flow-based restoration uses a prior estimate together with information from the observation. Given the current estimate, the prior estimate can be obtained from a single prediction at the zero-probe center or by averaging predictions under Gaussian perturbations. The predictor's nonlinearity generally makes these estimates different. We define the expected response to Gaussian perturbations as

$$\mathcal{Q}_t^\theta(\mathbf{x}) := \mathbb{E}_{\boldsymbol{\epsilon}\sim\mathcal{N}(\mathbf{0},\mathbf{I})}\left[\mathcal{D}_t^\theta(t\mathbf{x} + (1-t)\boldsymbol{\epsilon})\right]. \tag{8}$$

where $\mathcal{Q}_t^\theta$ is the Gaussian-averaged destination at the current estimate $\mathbf{x}$ and time t. This inference-time averaging target is defined by Gaussian perturbations, distinct from the OT training coupling. Section 5.3 evaluates both its estimation accuracy and the restoration quality obtained after data correction.

A center prediction $\mathcal{D}_t^\theta(t\mathbf{x})$ eliminates sampling variance using a single model evaluation, but generally differs from $\mathcal{Q}_t^\theta(\mathbf{x})$. This difference reflects nonlinear responses to perturbed inputs. Averaging predictions from independent probes estimates this expected response, but introduces sampling variation at a finite evaluation budget. We therefore investigate whether structured perturbations can retain this information while reducing sampling variation with the same number of prior evaluations.

SPIR evaluates opposite Gaussian perturbations. Their average retains the shared nonlinear response, while sign-changing components sum to zero. For smooth predictors, this removes the local first-order perturbation. With the same number of prior evaluations, the estimation benefit depends on the balance between sign-changing and retained variation (Section 3.3). Two opposite probes form the simplest configuration.

To incorporate the observation, SPIR applies a data-proximal correction to the symmetric prediction average. The probes are reused within each time step, and the corrected estimate initializes the next step. This yields the finite-budget restoration procedure in Section 3.2.

3.2 Symmetric-probe image restoration

As shown in Figure 1, SPIR maintains a clean-image estimate and alternates a symmetric flow-prior estimate with an observation-based correction. The OT-trained flow prior remains fixed during restoration. A prescribed time schedule controls input perturbations and proximal weighting, with at most M updates per time step and optional inner-loop early stopping. Algorithm 1 summarizes the procedure using single-step predictions and frozen symmetric probes.

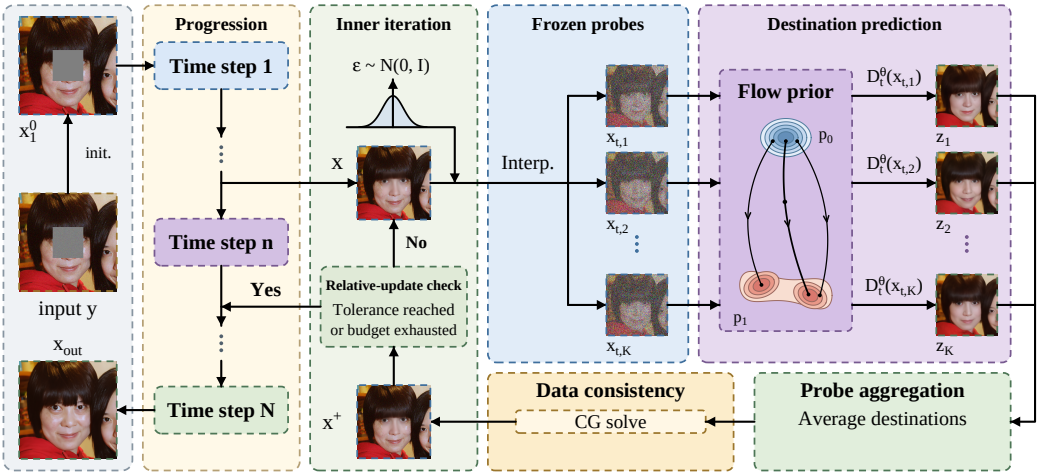

Figure 1: **SPIR workflow.** Symmetric destination estimates are aggregated and corrected using the observation. The relative-update test or update budget triggers time advancement. $\mathcal{D}_t^\theta$ is the destination estimator in Equation 6; Algorithm 1 specifies the complete SPIR procedure.

Initialization and time schedule. We initialize $\mathbf{x}_1^0 = \mathbf{H}^\top \mathbf{y}$, where the adjoint $\mathbf{H}^\top$ maps observations into image space. For prescribed $0 < t_1 := t_{\min} < \cdots < t_N := t_{\max} < 1$, we adopt a time-annealing schedule (Pourya et al., 2026),

$$w_{t_n}^2 = \frac{(1-t_n)^2}{t_n^2 + (1-t_n)^2}, \tag{9}$$

where $w_{t_n}^2 > 0$ controls the data-proximal correction. Increasing t_n reduces input noise and strengthens attachment to the prior estimate. Excluding the endpoints $t = 0$ and $t = 1$ preserves dependence on the current estimate and keeps the proximal precision finite. For $N \geq 2$, we space the proximal precision uniformly on a logarithmic scale:

$$\log \frac{1}{w_{t_n}^2} = \log \frac{1}{w_{t_{\min}}^2} + \frac{n-1}{N-1} \log \frac{w_{t_{\min}}^2}{w_{t_{\max}}^2}, \qquad n = 1, \ldots, N, \tag{10}$$

where $1/w_{t_n}^2$ is the proximal precision at time step n. Equation 10 preserves the prescribed endpoints and gives constant successive precision ratios. We recover the time steps as $\tau_n = \sqrt{w_{t_n}^2/(1-w_{t_n}^2)}$ and $t_n = 1/(1+\tau_n)$, where $\tau_n = (1-t_n)/t_n$ is the noise-to-image interpolation ratio. At time step $n = 1, \ldots, N$, $\mathbf{x}_n^0$ is the initial estimate and $\mathbf{x}_n^m$ the estimate after update m. Each step holds $t = t_n$ fixed for at most M updates, indexed by $m = 1, \ldots, M$.

Frozen symmetric probes. At each time step, SPIR draws $K/2$ independent vectors $\boldsymbol{\epsilon}_{2k-1} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and sets $\boldsymbol{\epsilon}_{2k} = -\boldsymbol{\epsilon}_{2k-1}$ for $k = 1, \ldots, K/2$. At a fixed input, their predictions estimate the average in Equation 8 with zero contribution from sign-changing responses (Section 3.3). A single pair ($K = 2$) provides a practical balance between fidelity and cost in Section 5.4. The set $\mathcal{E} = \{\boldsymbol{\epsilon}_k\}_{k=1}^K$ is fixed within each time step and defines

$$\mathbf{x}_{t,k}(\mathbf{x}) = t\mathbf{x} + (1-t)\boldsymbol{\epsilon}_k, \qquad k = 1, \ldots, K, \tag{11}$$

where $\mathbf{x}_{t,k}(\mathbf{x})$ is the kth interpolated flow input. Each update reuses these noise vectors and evaluates the network K times at the updated inputs in Equation 11. The $K = 1$ variant in Section 5.4 uses a single unpaired probe.

Averaged prediction and data consistency. For each perturbed input, the shared flow prior predicts the single-step destination $\mathcal{D}_t^\theta$ in Equation 6. Averaging the K predictions gives the prior estimate

$$\overline{\mathcal{D}}_{t,\mathcal{E}}^\theta(\mathbf{x}) = \frac{1}{K} \sum_{k=1}^K \mathcal{D}_t^\theta(t\mathbf{x} + (1-t)\boldsymbol{\epsilon}_k), \tag{12}$$

where $\overline{\mathcal{D}}^{\theta}_{t,\mathcal{E}}$ denotes the mean destination prediction over the fixed probe set $\mathcal{E}$. To incorporate the measurement constraint, we apply a data-proximal operator to the averaged estimate (Pourya et al., 2026). Writing the data-fidelity term in Equation 2 as $f_y := \ell_{\mathbf{y}}$, we set $\mathcal{C}_t := \mathrm{prox}_{w_t^2 f_y}$. For a predicted destination $\mathbf{z}$, this operator is

$$\mathcal{C}_t(\mathbf{z}) := \mathrm{prox}_{w_t^2 f_y}(\mathbf{z}) := \arg\min_{\mathbf{x}} \left\{ \frac{1}{2\sigma_y^2} \|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2 + \frac{1}{2w_t^2} \|\mathbf{x} - \mathbf{z}\|_2^2 \right\}, \tag{13}$$

where $\mathcal{C}_t$ is the exact data-proximal reference operator and $\mathbf{z}$ is its input estimate. The first term enforces consistency with the observation, whereas the second keeps the corrected image close to the prior estimate; decreasing w_t increases the relative weight of the latter. The positive coefficient $1/w_t^2$ makes the objective strictly convex even when $\mathbf{H}$ is rank deficient. Its closed-form solution is

$$\mathcal{C}_t(\mathbf{z}) = \left(\frac{\mathbf{H}^\top \mathbf{H}}{\sigma_y^2} + \frac{1}{w_t^2} \mathbf{I} \right)^{-1} \left(\frac{\mathbf{H}^\top \mathbf{y}}{\sigma_y^2} + \frac{1}{w_t^2} \mathbf{z} \right). \tag{14}$$

Our implementation approximates Equation 14 using conjugate gradients (CG), initialized at the current image estimate, with $\mathbf{z} = \overline{\mathcal{D}}^{\theta}_{t,\mathcal{E}}(\mathbf{x})$. With exact data correction, the reference update is

$$\mathbf{x}^m = \mathcal{T}_{t,\mathcal{E}}(\mathbf{x}^{m-1}), \qquad \mathcal{T}_{t,\mathcal{E}} := \mathcal{C}_t \circ \overline{\mathcal{D}}^{\theta}_{t,\mathcal{E}}, \tag{15}$$

where $\mathcal{T}_{t,\mathcal{E}}$ is the composition of prediction and correction at time t and $\mathbf{x}^m$ denotes the updated estimate. The finite CG solve approximates Equation 15; Section 3.3 analyzes data-correction stability and the resulting solver error.

Inner-loop stopping and time advancement. After retaining the corrected estimate at $t = t_n$, SPIR tests

$$r_t^m := \frac{\|\mathbf{x}^m - \mathbf{x}^{m-1}\|_2}{\max(\|\mathbf{x}^{m-1}\|_2, \varepsilon_{\mathrm{den}})} < \varepsilon_{\mathrm{upd}}, \tag{16}$$

where r_t^m is the relative update, $\varepsilon_{\mathrm{den}} > 0$ is a norm floor preventing division by zero, and $\varepsilon_{\mathrm{upd}} \geq 0$ is the dimensionless stopping threshold. Setting $\varepsilon_{\mathrm{upd}} = 0$ disables inner-loop early stopping. If the test fails and $m < M$, SPIR repeats with the same time and probes; otherwise, it sets $\mathbf{x}_n^{\mathrm{out}} = \mathbf{x}^m$ and $\mathbf{x}_{n+1}^0 = \mathbf{x}_n^{\mathrm{out}}$ for $n < N$. The output is $\widehat{\mathbf{x}} = \mathbf{x}_N^{\mathrm{out}}$. Equation 16 allocates updates according to the relative change in the estimate. Satisfying the test ends only the current inner loop; the prescribed time schedule continues until t_N.

3.3 Properties of symmetric prediction and data correction

At a fixed estimate $\mathbf{x}$ and time t, averaging predictions from opposite perturbations gives zero contribution from the response that changes sign with the perturbation. For a smooth predictor, this includes the first-order perturbation, while nonlinear responses shared by both signs remain. A single pair ($K = 2$) is sufficient for this property. Under local smoothness assumptions, the difference from the center prediction $\mathcal{D}_t^{\theta}(t\mathbf{x})$ is second order in the perturbation magnitude, characterizing the local structure of symmetric prediction (Appendix C.2).

For a fixed state independent of the probes, the comparison at matched prior-evaluation count involves K independent predictions or $K/2$ independent symmetric pairs. Symmetric averaging has lower mean-squared error relative to $\mathcal{Q}_t^{\theta}(\mathbf{x})$ when the variance of the sign-changing response exceeds that of the retained response. Appendix C.1 derives this condition and illustrates the resulting trade-off. In SPIR, this comparison applies to the first update at each time step with newly sampled probes. Subsequent updates reuse these probes; the algebraic symmetry and local remainder bound continue to apply under their stated assumptions. Appendix E.1 reports a fixed-state diagnostic of this variance condition using independently sampled probes.

The exact data correction is nonexpansive in its input estimate and preserves differences in the measurement null space. With finite CG, the error relative to the exact correction depends on the attained linear-system residual (Appendix C.3). Section 5.3 complements these stepwise properties by examining prediction error and final restoration quality.

4 Related Work

Gradient-step methods. Plug-and-Play (PnP) replaces a proximal operator in an iterative solver with a denoiser (Venkatakrishnan et al., 2013), whereas regularization by denoising (RED) constructs a regularizer from the denoising residual (Romano et al., 2017). PnP-GS (Hurault et al., 2022) parameterizes the denoiser as a gradient step on a learned scalar potential, enabling convergence to stationary points of an explicit objective under suitable assumptions. Broader operator-based perspectives include consensus equilibrium (Buzzard et al., 2018), which balances data and prior agents without requiring an explicit objective, and FiRe (Terris et al., 2025), which constructs implicit priors from fixed points of compositions of degradation and restoration. These frameworks are not restricted to gradient-step denoisers.

Diffusion-based methods. Diffusion and score-based models learn multiscale image priors through denoising or score estimation (Ho et al., 2020; Song et al., 2021). ILVR (Choi et al., 2021) conditions generation on low-frequency reference information, while DPS (Chung et al., 2023) guides denoised predictions with the measurement likelihood. DDRM (Kawar et al., 2022) and ΠGDM (Song et al., 2023) exploit the measurement operator, and DDS (Chung et al., 2024) incorporates CG data correction. Iteratively preconditioned guidance (Garber & Tirer, 2024) interpolates between back-projection and least-squares guidance. DiffPIR (also denoted PnP-Diff) (Zhu et al., 2023) alternates diffusion denoising with data-proximal updates, whereas DAPS (Zhang et al., 2025) decouples noise annealing from measurement-conditioned restoration for posterior sampling. BIRD (Chihaoui et al., 2024) uses diffusion inversion to jointly optimize input noise and unknown degradation parameters for blind restoration.

Flow-matching-based methods. Flow matching transports a source distribution to data using a learned velocity field (Lipman et al., 2023). OT-ODE (Pokle et al., 2024) constructs measurement-guided velocities for Gaussian probability paths, while Flow-Priors (Zhang et al., 2024) iteratively matches corrupted trajectories. D-Flow (Ben-Hamu et al., 2024) optimizes source noise through ODE differentiation. PnP-Flow (Martin et al., 2025) alternates data-fidelity descent, flow-path reprojection, and time-dependent denoising, and also considers averaging predictions from multiple random perturbations. Its sequence-convergence result requires boundedness and a summability condition on its time-dependent steps, without identifying the limit as a stationary point of a restoration objective. FLOWER (Pourya et al., 2026) combines destination estimation, measurement refinement, and time progression for approximate posterior sampling; its conditional transition analysis concerns sampling consistency rather than convergence of a repeated restoration update. FlowDPS (Kim et al., 2025) guides clean-image estimates by likelihood gradients and injects stochasticity into noise estimates. FlowChef (Patel et al., 2025) exploits rectified-flow structure for guidance without network backpropagation. Conditional Wasserstein approaches (Chemseddine et al., 2025) learn conditional OT flow-matching models for Bayesian inverse problems.

5 Numerical Results

5.1 Evaluation protocol

Datasets. The evaluation uses four datasets: AFHQ-Cat (Choi et al., 2020) and FFHQ (Karras et al., 2019), resized to 256×256, and CelebA (Liu et al., 2015) and Imagenette2, an ImageNet subset (Deng et al., 2009), resized to 128×128. The main benchmark uses 100 validation images per dataset, shared across all evaluated methods.

Tasks and metrics. The five tasks are Gaussian denoising, Gaussian deblurring, super-resolution by decimation, centered box inpainting, and random pixel inpainting. The measurement-noise standard deviations are 0.2, 0.05, 0.05, 0.05, and 0.01, respectively. Deblurring uses a 61×61 Gaussian kernel with blur standard deviation 3.0 at 256×256 and 1.0 at 128×128. Super-resolution uses factors 4 and 2, and box masks use half-sizes 40 and 20, respectively; random inpainting removes 70% of pixels in both cases. We report PSNR (dB, higher is better), SSIM (Wang et al., 2004) (higher is better), and LPIPS (Zhang et al., 2018) (lower is better).

Table 1: Mean restoration performance on four datasets, each evaluated on 100 validation images under the protocols in Section 5.1. Methods are grouped as gradient-step, diffusion, and flow-based. Best and second-best means are bold and underlined, respectively; ties are retained. N/A denotes unavailable results. Table 2 in Appendix D reports the corresponding ± standard deviations.

Dataset	Method	Denoising			Gaussian deblurring			Super-resolution			Box inpainting			Random inpainting		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
AFHQ Cat 256×256	PnP-GS	22.94	0.604	0.465	27.16	0.734	0.370	24.46	0.638	0.362	22.86	0.847	0.139	14.07	0.246	0.739
	DDRM	32.46	0.884	0.140	29.20	0.783	0.292	22.62	0.599	0.454	21.92	0.898	0.114	32.19	0.893	0.069
	DPS	27.99	0.736	0.177	23.37	0.619	0.378	25.84	0.699	0.302	22.19	0.746	0.147	22.82	0.497	0.388
	PnP-Diff	21.06	0.344	0.470	N/A	N/A	N/A	24.05	0.622	0.402	23.53	0.749	0.136	27.91	0.771	0.178
	D-Flow	19.90	0.299	0.521	28.15	0.742	**0.145**	24.79	0.600	0.193	25.47	0.843	0.068	32.28	0.889	0.045
	Flow-Priors	24.22	0.476	0.311	27.00	0.708	0.191	18.51	0.398	0.484	25.88	0.856	0.129	29.60	0.823	0.122
	OT-ODE	30.20	0.817	0.094	N/A	N/A	N/A	26.25	0.711	**0.112**	24.48	0.870	0.096	29.08	0.823	0.097
	FlowDPS	28.38	0.782	0.200	25.94	0.675	0.260	15.91	0.456	0.492	23.07	0.815	0.172	24.71	0.694	0.256
	FlowChef	22.36	0.644	0.217	19.38	0.427	0.338	18.66	0.377	0.418	19.42	0.598	0.267	13.93	0.331	0.514
	PnP-Flow	31.56	0.860	0.181	28.63	0.761	0.345	23.01	0.644	0.349	27.04	0.900	0.129	34.12	0.930	0.047
	FLOWER	32.45	0.887	0.131	29.52	0.790	0.284	25.35	0.708	0.278	27.59	**0.924**	0.067	33.71	0.921	0.050
	SPIR	**32.94**	**0.900**	**0.071**	**29.58**	**0.793**	0.237	**27.26**	**0.755**	0.168	**28.05**	0.917	**0.062**	**34.25**	**0.933**	**0.038**
CelebA 128×128	PnP-GS	20.79	0.553	0.331	33.46	0.931	0.044	25.70	0.852	0.072	22.40	0.844	0.172	15.40	0.434	0.497
	DDRM	32.72	0.922	0.041	34.52	0.947	0.024	30.38	0.900	0.054	21.22	0.902	0.113	31.13	0.926	0.033
	DPS	21.74	0.440	0.323	19.43	0.422	0.418	28.20	0.778	0.061	22.83	0.739	0.129	18.95	0.390	0.409
	PnP-Diff	21.26	0.406	0.322	N/A	N/A	N/A	25.90	0.739	0.127	21.58	0.755	0.110	26.22	0.803	0.109
	D-Flow	19.88	0.356	0.367	22.71	0.468	0.211	27.91	0.733	0.062	28.97	0.863	0.036	33.09	0.941	0.017
	Flow-Priors	24.37	0.532	0.213	31.34	0.860	0.059	27.23	0.761	0.095	29.20	0.908	0.058	28.69	0.837	0.068
	OT-ODE	30.32	0.870	0.038	N/A	N/A	N/A	31.18	0.907	**0.020**	26.67	0.906	0.048	27.14	0.835	0.078
	FlowDPS	30.75	0.887	0.039	30.20	0.883	0.077	22.96	0.716	0.148	26.20	0.856	0.085	24.38	0.756	0.121
	FlowChef	26.62	0.647	0.116	28.10	0.832	0.060	26.79	0.773	0.090	25.42	0.876	0.049	19.88	0.553	0.273
	PnP-Flow	32.07	0.912	0.059	34.46	0.941	0.051	31.15	0.906	0.057	29.75	0.936	0.048	33.61	0.953	0.023
	FLOWER	32.93	0.927	0.040	**35.35**	**0.954**	0.034	32.71	0.931	0.041	**30.59**	**0.949**	0.028	33.61	0.953	0.021
	SPIR	**33.08**	**0.930**	**0.020**	35.26	0.953	**0.023**	**33.01**	**0.938**	0.027	30.24	0.947	**0.027**	**34.02**	**0.959**	**0.014**
FFHQ 256×256	PnP-GS	21.63	0.638	0.428	25.21	0.754	0.295	23.22	0.657	0.319	22.11	0.833	0.164	15.17	0.412	0.624
	DDRM	33.29	0.906	0.086	29.07	**0.831**	0.135	23.39	0.708	0.233	21.14	0.899	0.117	32.29	0.917	0.044
	DPS	28.75	0.768	0.115	25.21	0.730	0.153	25.73	0.738	0.165	24.02	0.739	0.131	23.39	0.552	0.300
	PnP-Diff	21.18	0.326	0.428	N/A	N/A	N/A	24.93	0.727	0.186	29.15	0.775	0.070	28.89	0.831	0.082
	D-Flow	19.93	0.282	0.479	28.31	0.778	**0.100**	24.16	0.622	0.196	29.99	0.864	**0.030**	32.76	0.910	0.039
	Flow-Priors	24.30	0.451	0.274	26.96	0.747	0.175	16.87	0.412	0.550	29.37	0.895	0.062	28.61	0.790	0.109
	OT-ODE	30.62	0.831	0.066	N/A	N/A	N/A	25.95	0.740	**0.107**	25.25	0.878	0.090	28.52	0.822	0.109
	FlowDPS	28.85	0.826	0.143	25.57	0.736	0.243	15.16	0.461	0.473	25.26	0.847	0.127	24.21	0.729	0.214
	FlowChef	22.33	0.674	0.202	19.32	0.498	0.321	17.18	0.370	0.478	20.24	0.647	0.232	13.45	0.361	0.529
	PnP-Flow	30.00	0.828	0.171	25.46	0.715	0.305	22.37	0.690	0.289	27.27	0.805	0.192	24.84	0.717	0.287
	FLOWER	33.08	0.895	0.094	29.38	0.825	0.201	24.42	0.731	0.221	**30.79**	**0.936**	0.034	33.95	0.933	0.036
	SPIR	**33.65**	**0.908**	**0.050**	**29.52**	0.827	0.165	**27.00**	**0.785**	0.136	30.61	**0.936**	0.036	**34.71**	**0.945**	**0.028**
Imagenette2 128×128	PnP-GS	18.98	0.505	0.472	29.19	0.872	0.111	18.51	0.585	0.328	21.41	0.844	0.153	15.31	0.407	0.555
	DDRM	31.32	0.901	0.065	31.99	0.922	**0.036**	27.00	0.829	0.099	21.99	0.903	0.111	27.98	0.878	0.055
	DPS	22.89	0.554	0.247	20.03	0.507	0.365	25.18	0.717	0.101	22.67	0.768	0.118	18.08	0.367	0.414
	PnP-Diff	21.17	0.461	0.266	N/A	N/A	N/A	23.99	0.690	0.142	22.10	0.786	0.104	24.14	0.747	0.129
	D-Flow	19.96	0.416	0.295	17.72	0.334	0.372	19.73	0.451	0.375	23.67	0.835	0.088	28.01	0.862	0.042
	Flow-Priors	24.22	0.574	0.161	29.16	0.838	0.067	24.50	0.721	0.137	23.74	0.886	0.127	25.68	0.791	0.091
	OT-ODE	28.18	0.808	0.072	N/A	N/A	N/A	26.07	0.798	0.069	22.29	0.861	0.113	24.13	0.745	0.147
	FlowDPS	28.76	0.842	0.067	26.90	0.792	0.130	21.35	0.597	0.277	21.28	0.753	0.177	21.89	0.620	0.240
	FlowChef	25.85	0.657	0.103	24.68	0.705	0.117	23.81	0.685	0.118	20.90	0.821	0.116	18.46	0.465	0.362
	PnP-Flow	29.94	0.862	0.107	31.13	0.903	0.095	27.77	0.836	0.103	24.99	0.897	0.114	29.33	0.904	0.047
	FLOWER	31.07	0.898	0.062	32.24	0.923	0.058	28.86	0.872	0.079	25.38	**0.916**	**0.086**	29.35	0.904	0.045
	SPIR	**31.41**	**0.905**	**0.030**	**32.40**	**0.926**	0.038	**29.26**	**0.889**	**0.060**	**25.42**	0.914	0.093	**29.87**	**0.919**	**0.039**

Baselines and priors. All methods use their respective pretrained priors. For FLOWER and PnP-Flow, we average five restored images before computing the evaluation metrics. We compare SPIR with PnP-GS (Hurault et al., 2022), DDRM (Kawar et al., 2022), DPS (Chung et al., 2023), PnP-Diff (DiffPIR) (Zhu et al., 2023), D-Flow (Ben-Hamu et al., 2024), Flow-Priors (Zhang et al., 2024), OT-ODE (Pokle et al., 2024), PnP-Flow (Martin et al., 2025), FlowDPS (Kim et al., 2025), FlowChef (Patel et al., 2025), and FLOWER (Pourya et al., 2026). Appendices B and G give the SPIR algorithm and computational costs, respectively.

5.2 Main results

We evaluate SPIR on four datasets and five restoration tasks. Table 1 shows leading image fidelity in most evaluated settings and competitive perceptual quality. For denoising, SPIR achieves the best performance throughout the evaluation, improving upon competitive baselines such as DDRM and

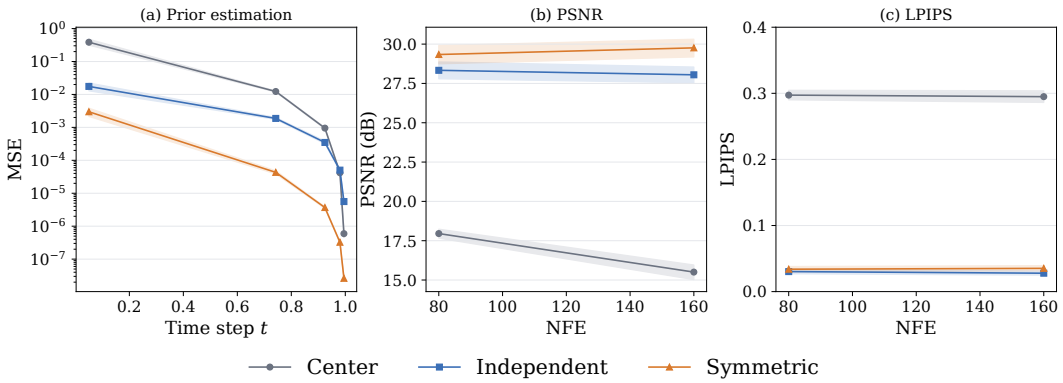

Figure 2: Prior-estimation ablation on CelebA box inpainting. (a) Per-pixel prediction MSE to an independent Monte Carlo neighborhood mean at fixed states from 16 images. Center uses one prediction; Independent and Symmetric use two. (b,c) Final PSNR and LPIPS at matched NFE on 100 images. Shading denotes 95% image-cluster bootstrap intervals after averaging probe repeats within each image.

FLOWER. For Gaussian deblurring, SPIR attains leading or near-leading image fidelity, although FLOWER retains an advantage in some settings. Perceptual performance is more varied, with D-Flow and DDRM achieving lower LPIPS in several comparisons. For super-resolution, SPIR consistently leads in image fidelity, whereas OT-ODE often obtains lower LPIPS, reflecting a fidelity-perception trade-off. For box inpainting, SPIR combines competitive image fidelity with strong LPIPS performance. FLOWER remains a close competitor and outperforms SPIR in fidelity in some settings, while D-Flow also achieves favorable perceptual results. For random inpainting, SPIR consistently achieves the best results across all metrics, outperforming otherwise competitive methods such as FLOWER, PnP-Flow, and D-Flow. Overall, the results demonstrate broad restoration effectiveness while retaining task-dependent trade-offs between fidelity and perception. Appendix A provides qualitative comparisons. Appendix F lists dataset-specific solver settings in Tables 4, 5, 6, and 7, including SPIR's probe count, inner-update cap, and number of time steps.

5.3 Ablation Studies

We compare prior-estimation strategies using fixed-state prediction diagnostics and restoration runs with matched flow-network evaluation counts. Center makes a single destination prediction from the unperturbed input $t\mathbf{x}$; Independent averages predictions from independently sampled Gaussian probes; Symmetric averages predictions from a Gaussian probe and its negative. Independent and Symmetric use the same numbers of probes and updates, isolating the effect of pairing. To match the number of flow-network evaluations (NFE), Center performs additional inner updates and corresponding data-correction steps. The matched budget counts network evaluations, with data-correction costs varying across strategies. The fixed-state diagnostic uses an independent Monte Carlo estimate of the expected prediction as its reference; the end-to-end runs evaluate restoration quality against clean images. Appendix E.1 details the configurations, sampling, and statistical protocol.

Figure 2(a) shows that Symmetric has lower error to the Monte Carlo reference than Independent and Center at the tested states, indicating more accurate estimation of the expected prediction. In Figure 2(b), Symmetric achieves the highest PSNR at both budgets, while Independent also outperforms Center. Figure 2(c) shows a different ordering: Independent has the lowest LPIPS, Symmetric has slightly higher LPIPS, and both outperform Center. Together, these results support a fidelity benefit from pairing in this setting, with a perceptual trade-off relative to independent sampling. Figure 8 in Appendix E.1 complements these comparisons with a diagnostic of the normalized variance advantage. Appendix E.2 presents a controlled two-factor ablation of probe pairing and within-time reuse, with the time grids and prior-term coefficients specified for each setting.

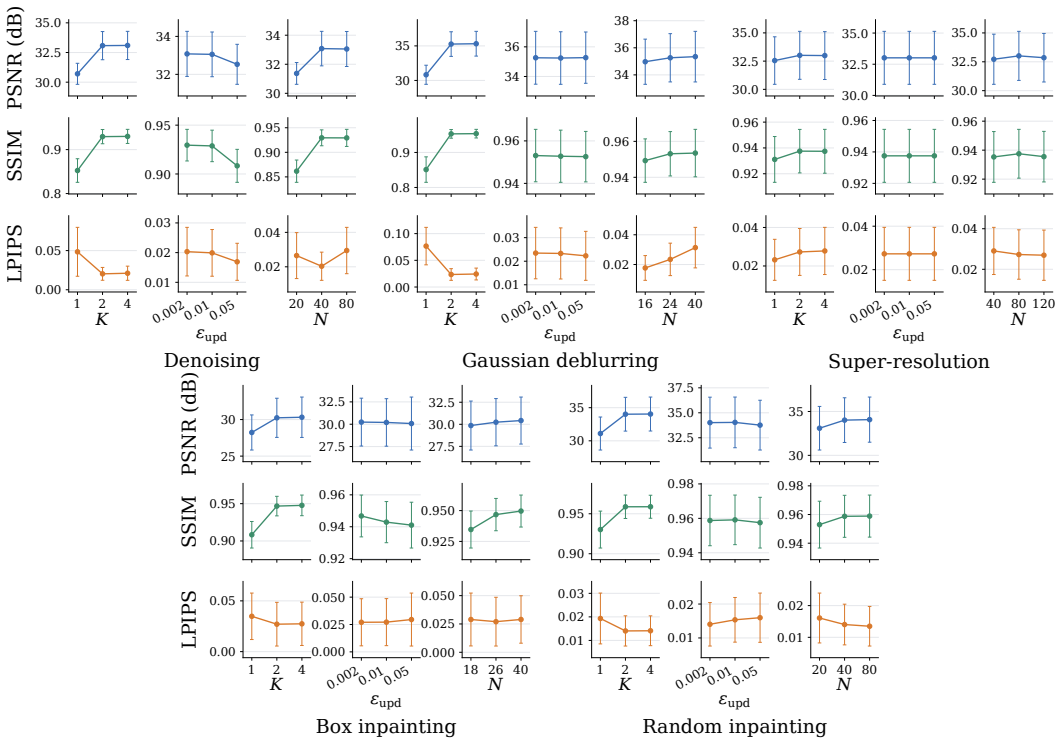

Figure 3: Sensitivity to probe count K, relative-update tolerance ε_{upd}, and number of time steps N across five restoration tasks. Rows show PSNR, SSIM, and LPIPS.

5.4 Parameter sensitivity

Figure 3 evaluates probe count K, update tolerance ε_{upd}, and number of time steps N. Using $K = 2$ captures most of the fidelity improvement over $K = 1$; $K = 4$ provides limited additional benefit at twice the per-update model cost, making $K = 2$ a practical choice. Small tolerances, approximately 2×10^{-3} to 10^{-2} in the tested settings, retain similar restoration quality. Larger tolerances can reduce fidelity, although their effect on LPIPS varies by task. The tolerance controls inner-loop early stopping. Increasing N initially improves fidelity but yields diminishing returns, favoring moderate, task-dependent choices rather than the largest grid. Excessive time steps increase the evaluation budget and can slightly reduce super-resolution fidelity or worsen deblurring LPIPS. These sensitivity analyses inform parameter selection, whereas Section 5.3 isolates the estimation strategy; protocol details are in Appendix E.

6 Conclusion

We presented SPIR, a novel image restoration method based on symmetric-probe estimation with flow-matching priors. Symmetric averaging incorporates predictions from Gaussian-perturbed inputs, with the probes reused within each time step. Our analysis characterizes the conditions for improved estimation and the stability of data correction. Systematic performance evaluations demonstrate strong restoration fidelity and competitive perceptual quality. Fixed-state diagnostics support the variance condition in the two evaluated settings, while controlled ablations show fidelity gains from pairing under both sampling strategies and no consistent additional benefit from within-time reuse. The perceptual effect of pairing varies by task. Parameter sensitivity analysis identifies settings with favorable restoration quality under the tested evaluation budgets. Together, these results support symmetric estimation with data-consistency correction as an effective approach under the evaluated restoration tasks and model-evaluation budgets.

AI USE STATEMENT

In this work, we used OpenAI Codex for manuscript polishing. We have reviewed all AI-assisted work, including the revised manuscript text. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

Appendix B specifies the SPIR algorithm, and Appendix C provides the proofs and their assumptions. Section 5.1 defines the benchmark protocol. Appendices D and E describe image-wise variability, ablation protocols, and parameter-sensitivity settings. Appendix F lists baseline solver settings and SPIR's K, M, and N. Appendix G reports measured runtime and memory; Appendix G.1 defines the theoretical network-evaluation count and its upper bound.

REFERENCES

Michael S. Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. In International Conference on Learning Representations (ICLR), 2023.

Vegard Antun, Francesco Renna, Clarice Poon, Ben Adcock, and Anders C Hansen. On instabilities of deep learning in image reconstruction and the potential costs of AI. Proceedings of the National Academy of Sciences, 117(48):30088–30095, 2020.

Heli Ben-Hamu, Omri Puny, Itai Gat, Brian Karrer, Uriel Singer, and Yaron Lipman. D-Flow: Differentiating through flows for controlled generation. In International Conference on Machine Learning (ICML), 2024.

Gregery T. Buzzard, Stanley H. Chan, Suhas Sreehari, and Charles A. Bouman. Plug-and-play unplugged: Optimization-free reconstruction using consensus equilibrium. SIAM Journal on Imaging Sciences, 11(3): 2001–2020, 2018. doi: 10.1137/17M1122451.

Jannis Chemseddine, Paul Hagemann, Gabriele Steidl, and Christian Wald. Conditional wasserstein distances with applications in bayesian OT flow matching. Journal of Machine Learning Research (JMLR), 26:141:1–141:47, 2025.

Hamadi Chihaoui, Abdelhak Lemkhenter, and Paolo Favaro. Blind image restoration via fast diffusion inversion. In Advances in Neural Information Processing Systems (NeurIPS), 2024.

Jooyoung Choi, Sungwon Kim, Yonghyun Jeong, Youngjune Gwon, and Sungroh Yoon. ILVR: conditioning method for denoising diffusion probabilistic models. In IEEE/CVF International Conference on Computer Vision (ICCV), 2021.

Yunjey Choi, Youngjung Uh, Jaejun Yoo, and Jung-Woo Ha. StarGAN v2: Diverse image synthesis for multiple domains. In IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.

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

Hyungjin Chung, Suhyeon Lee, and Jong Chul Ye. Decomposed diffusion sampler for accelerating large-scale inverse problems. In International Conference on Learning Representations (ICLR), 2024.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 248–255, 2009. doi: 10.1109/CVPR.2009.5206848.

Prafulla Dhariwal and Alexander Quinn Nichol. Diffusion models beat GANs on image synthesis. In Advances in Neural Information Processing Systems (NeurIPS), 2021.

Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(2):295–307, 2016.

David L Donoho. Compressed sensing. IEEE Transactions on information theory, 52(4):1289–1306, 2006.

Tomer Garber and Tom Tirer. Image restoration by denoising diffusion models with iteratively preconditioned guidance. In IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024.

J. M. Hammersley and K. W. Morton. A new Monte Carlo technique: Antithetic variates. Mathematical Proceedings of the Cambridge Philosophical Society, 52(3):449–475, 1956. doi: 10.1017/S0305004100031455.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In Advances in Neural Information Processing Systems (NeurIPS), 2020.

Samuel Hurault, Arthur Leclaire, and Nicolas Papadakis. Gradient step denoiser for convergent plug-and-play. In International Conference on Learning Representations (ICLR), 2022.

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), 2019.

Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. In Advances in Neural Information Processing Systems (NeurIPS), 2022.

Jeongsol Kim, Bryan Sangwoo Kim, and Jong Chul Ye. FlowDPS: Flow-driven posterior sampling for inverse problems. In IEEE/CVF International Conference on Computer Vision (ICCV), pp. 12328–12337, 2025.

Nikita Kornilov, Petr Mokrov, Alexander V. Gasnikov, and Alexander Korotin. Optimal flow matching: Learning straight trajectories in just one step. In Advances in Neural Information Processing Systems (NeurIPS), 2024.

Sangyun Lee, Beomsu Kim, and Jong Chul Ye. Minimizing trajectory curvature of ode-based generative models. In International Conference on Machine Learning (ICML), 2023.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In International Conference on Learning Representations (ICLR), 2023.

Qiang Liu. Rectified flow: A marginal preserving approach to optimal transport. arXiv preprint arXiv:2209.14577, 2022.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In International Conference on Learning Representations (ICLR), 2023.

Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In IEEE International Conference on Computer Vision (ICCV), pp. 3730–3738, 2015. doi: 10.1109/ICCV.2015.425.

Stéphane Mallat. A wavelet tour of signal processing. Academic Press, 2 edition, 1999.

Ségolène Tiffany Martin, Anne Gagneux, Paul Hagemann, and Gabriele Steidl. PnP-Flow: Plug-and-play image restoration with flow matching. In International Conference on Learning Representations (ICLR), 2025.

Maitreya Patel, Song Wen, Dimitris N. Metaxas, and Yezhou Yang. FlowChef: Steering of rectified flow models for controlled generations. In IEEE/CVF International Conference on Computer Vision (ICCV), pp. 15308–15318, 2025.

Gabriel Peyré and Marco Cuturi. Computational optimal transport: With applications to data science. Now Foundations and Trends, 2019.

Ashwini Pokle, Matthew J. Muckley, Ricky T. Q. Chen, and Brian Karrer. Training-free linear image inverses via flows. Transactions on Machine Learning Research, 2024.

Aram-Alexandre Pooladian, Heli Ben-Hamu, Carles Domingo-Enrich, Brandon Amos, Yaron Lipman, and Ricky T. Q. Chen. Multisample flow matching: Straightening flows with minibatch couplings. In International Conference on Machine Learning (ICML), 2023.

Mehrsa Pourya, Bassam El Rawas, and Michael Unser. Flower: A flow-matching solver for inverse problems. In International Conference on Learning Representations (ICLR), 2026.

Hongyu Ren, Shengjia Zhao, and Stefano Ermon. Adaptive antithetic sampling for variance reduction. In Proceedings of the 36th International Conference on Machine Learning, volume 97 of Proceedings of Machine Learning Research, pp. 5420–5428, 2019. URL https://proceedings.mlr.press/v97/ren19b.html.

Yaniv Romano, Michael Elad, and Peyman Milanfar. The little engine that could: Regularization by denoising (RED). SIAM Journal on Imaging Sciences, 10(4):1804–1844, 2017.

Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. Physica D: Nonlinear Phenomena, 60(1-4):259–268, 1992.

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.

Matthieu Terris, Ulugbek S. Kamilov, and Thomas Moreau. FiRe: Fixed-points of restoration priors for solving inverse problems. In IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 23185–23194, 2025.

Alexander Tong, Kilian Fatras, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. Transactions on Machine Learning Research, 2024, 2024.

Singanallur V. Venkatakrishnan, Charles A. Bouman, and Brendt Wohlberg. Plug-and-play priors for model based reconstruction. In IEEE Global Conference on Signal and Information Processing (GlobalSIP), 2013.

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. doi: 10.1109/TIP.2003.819861.

Ling Yang, Zixiang Zhang, Zhilong Zhang, Xingchao Liu, Minkai Xu, Wentao Zhang, Chenlin Meng, Stefano Ermon, and Bin Cui. Consistency flow matching: Defining straight flows with velocity consistency. arXiv preprint arXiv:2407.02398, 2024.

Bingliang Zhang, Wenda Chu, Julius Berner, Chenlin Meng, Anima Anandkumar, and Yang Song. Improving diffusion inverse problem solving with decoupled noise annealing. In IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 20895–20905, 2025.

Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. IEEE Transactions on Image Processing, 26(7):3142–3155, 2017.

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 586–595, 2018. doi: 10.1109/CVPR.2018.00068.

Yasi Zhang, Peiyu Yu, Yaxuan Zhu, Yingshan Chang, Feng Gao, Ying Nian Wu, and Oscar Leong. Flow priors for linear inverse problems via iterative corrupted trajectory matching. In Advances in Neural Information Processing Systems (NeurIPS), 2024.

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 (CVPR) Workshops, pp. 1219–1229, 2023.

A Qualitative restoration results

Figures 4, 5, 6, and 7 present two qualitative examples per task and dataset across two consecutive panels. Rows correspond to denoising, deblurring, super-resolution, box inpainting, and random inpainting; columns compare ground truth, degraded inputs, and restored outputs. N/A denotes unavailable results.

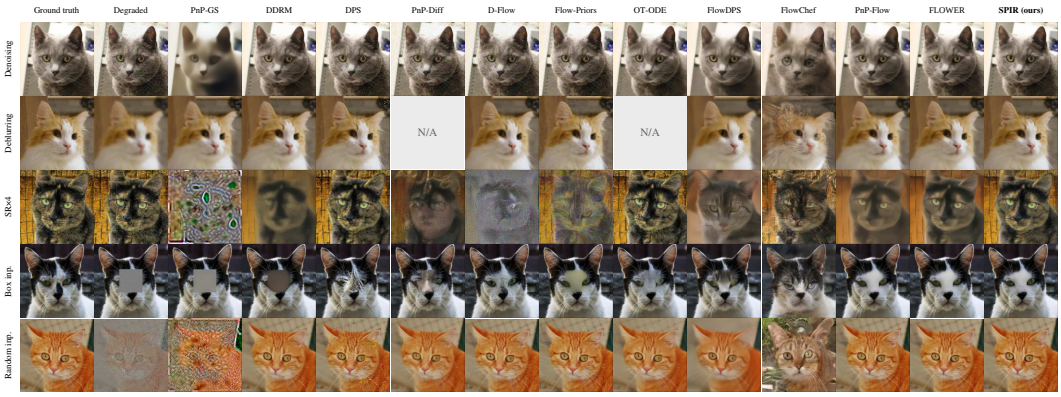

Figure 4: AFHQ-Cat qualitative results (a).

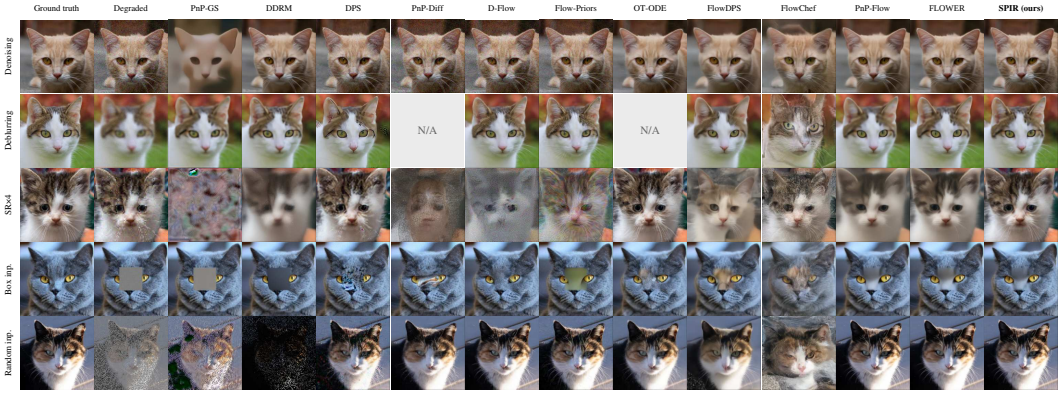

Figure 4: AFHQ-Cat qualitative results (b), continued.



Figure 5: CelebA qualitative results (a).

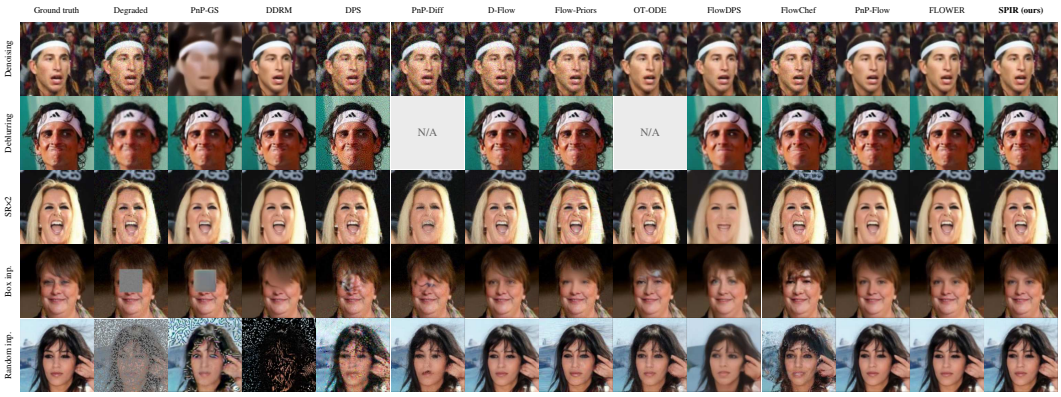

Figure 5: CelebA qualitative results (b), continued.

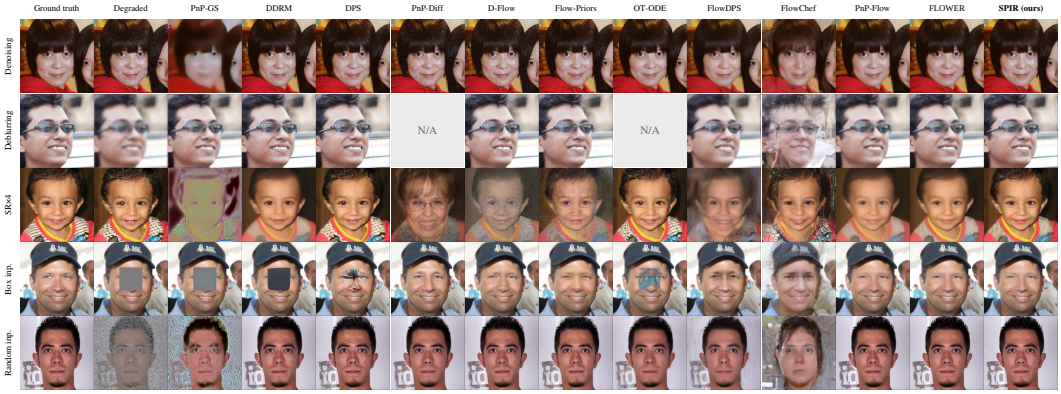

Figure 6: FFHQ qualitative results (a).

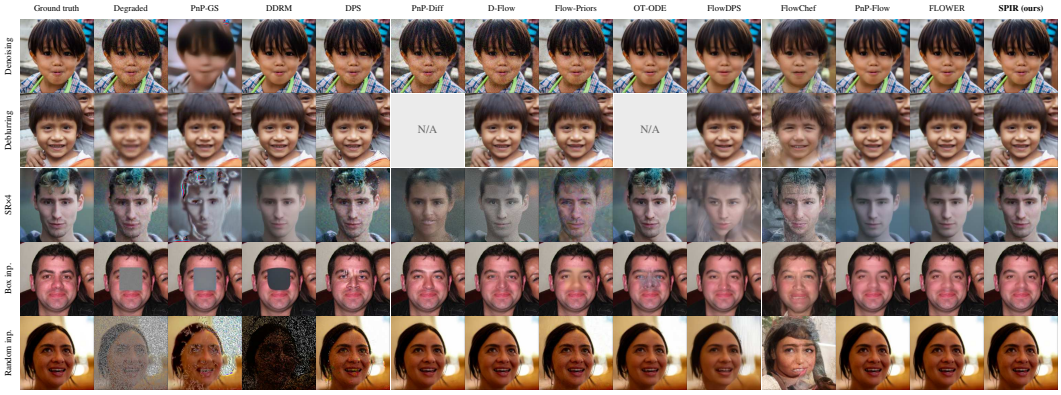

Figure 6: FFHQ qualitative results (b), continued.

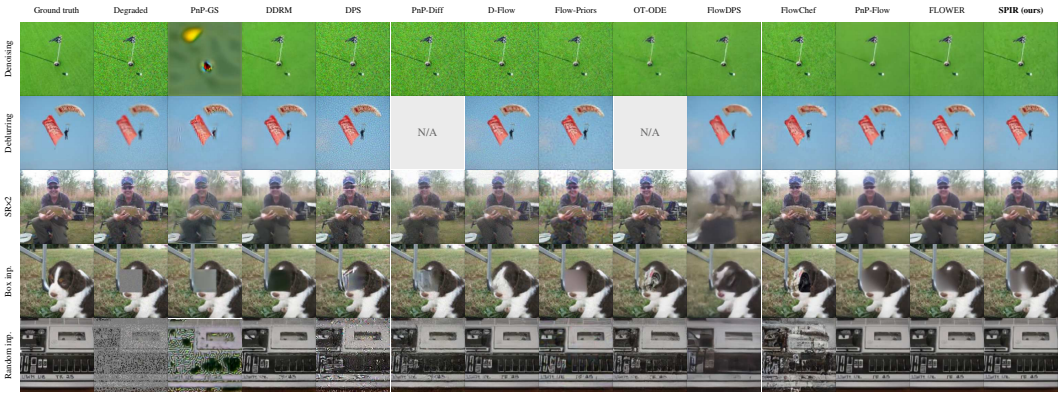

Figure 7: Imagenette2 qualitative results (a).

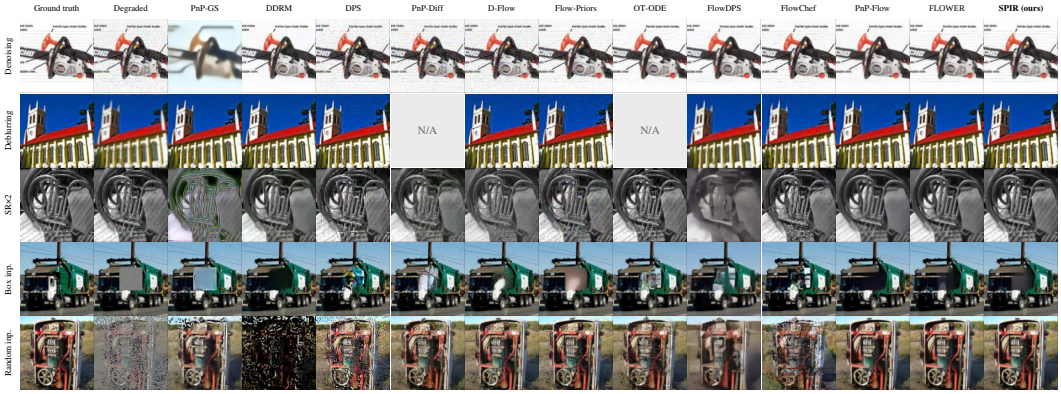

Figure 7: Imagenette2 qualitative results (b), continued.

B ALGORITHM

Algorithm 1 Symmetric-Probe Image Restoration (SPIR)

Require: Measurement $\mathbf{y}$; operators $\mathbf{H}, \mathbf{H}^\top$; destination model $\mathcal{D}_t^\theta$; time steps N; total probes K (even); endpoints $t_{\min}, t_{\max}$; inner-update cap M; relative-update tolerance $\varepsilon_{\text{upd}} \geq 0$; denominator safeguard $\varepsilon_{\text{den}} > 0$; noise level σ_y

$\mathbf{x} \leftarrow \mathbf{H}^\top \mathbf{y}$ ▷ Initialize
$(t_n, w_{t_n}^2)_{n=1}^N \leftarrow$ Equations 9, 10
for $n = 1$ **to** N **do**
  $\boldsymbol{\epsilon}_{n,2k-1} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}),\ k = 1, \ldots, K/2$
  $\boldsymbol{\epsilon}_{n,2k} = -\boldsymbol{\epsilon}_{n,2k-1},\ k = 1, \ldots, K/2$ ▷ Freeze symmetric probes
  $\mathbf{A}_n \leftarrow \mathbf{H}^\top \mathbf{H}/\sigma_y^2 + \mathbf{I}/w_{t_n}^2$
  for $m = 1$ **to** M **do**
    $\mathbf{z} \leftarrow \frac{1}{K} \sum_{k=1}^K \mathcal{D}_{t_n}^\theta(t_n \mathbf{x} + (1 - t_n)\boldsymbol{\epsilon}_{n,k})$ ▷ Aggregate predictions
    $\mathbf{b} \leftarrow \mathbf{H}^\top \mathbf{y}/\sigma_y^2 + \mathbf{z}/w_{t_n}^2$
    $\widetilde{\mathbf{x}} \leftarrow \mathrm{CG}(\mathbf{A}_n, \mathbf{b}; \mathbf{x})$ ▷ Data consistency
    $r \leftarrow \frac{\|\widetilde{\mathbf{x}} - \mathbf{x}\|_2}{\max(\|\mathbf{x}\|_2, \varepsilon_{\text{den}})}$
    $\mathbf{x} \leftarrow \widetilde{\mathbf{x}}$
    if $r < \varepsilon_{\text{upd}}$ **then** ▷ Inner-loop early stopping
      break
    end if
  end for
end for
return $\widehat{\mathbf{x}} = \mathbf{x}$

C FINITE-SAMPLE ANALYSIS AND DATA-CORRECTION ERROR

This appendix proves the finite-sample statements in Section 3.3 for symmetric prediction and data correction.

C.1 MATCHED PRIOR-EVALUATION COMPARISON AND THE CENTER ALTERNATIVE

Fix the current estimate $\mathbf{x}$ and time t, and define $g(\boldsymbol{\epsilon}) := \mathcal{D}_t^\theta(t\mathbf{x} + (1-t)\boldsymbol{\epsilon})$, with $\mathbf{x}$ and t held fixed. Its even and odd responses are $g_{\mathrm{e}}(\boldsymbol{\epsilon}) := [g(\boldsymbol{\epsilon}) + g(-\boldsymbol{\epsilon})]/2$ and $g_{\mathrm{o}}(\boldsymbol{\epsilon}) := [g(\boldsymbol{\epsilon}) - g(-\boldsymbol{\epsilon})]/2$. A prediction averaged over symmetric probes keeps g_{e} and removes g_{o} exactly, regardless of predictor smoothness or training coupling.

Proposition 1 (Estimation at matched prior-evaluation count)**.** For a state $\mathbf{x}$ fixed independently of the probes, suppose $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and $\mathbb{E}\|g(\boldsymbol{\epsilon})\|_2^2 < \infty$. Set $\mathbf{q} = \mathbb{E}g(\boldsymbol{\epsilon}) = \mathcal{Q}_t^\theta(\mathbf{x})$, $V_{\mathrm{e}} = \mathbb{E}\|g_{\mathrm{e}} - \mathbf{q}\|_2^2$, and $V_{\mathrm{o}} = \mathbb{E}\|g_{\mathrm{o}}\|_2^2$. For even K, let $\widehat{\mathbf{q}}_{\mathrm{iid}}$ average K independent predictions and $\widehat{\mathbf{q}}_{\mathrm{pair}}$ average $K/2$ independent opposite pairs. Both estimators have mean $\mathbf{q}$, and

$$\mathbb{E}\|\widehat{\mathbf{q}}_{\mathrm{iid}} - \mathbf{q}\|_2^2 = \frac{V_{\mathrm{e}} + V_{\mathrm{o}}}{K}, \qquad \mathbb{E}\|\widehat{\mathbf{q}}_{\mathrm{pair}} - \mathbf{q}\|_2^2 = \frac{2V_{\mathrm{e}}}{K}. \tag{17}$$

where V_{e} and V_{o} quantify the even- and odd-response variation, and both estimators require K prior evaluations.

Equation 17 states the trade-off precisely: pairing removes odd variation but uses only $K/2$ independent even responses. It reduces mean-squared estimation error exactly when $V_{\mathrm{o}} > V_{\mathrm{e}}$, with equality when the two are equal. The deterministic center has error $\|g(\mathbf{0}) - \mathbf{q}\|_2^2$ relative to this design target.

Proof of Proposition 1. The Gaussian probe distribution is invariant under sign reversal. Since g_{e} is even and g_{o} is odd, $\mathbb{E}g_{\mathrm{e}} = \mathbf{q}$, $\mathbb{E}g_{\mathrm{o}} = \mathbf{0}$, and $\mathbb{E}\langle g_{\mathrm{e}} - \mathbf{q}, g_{\mathrm{o}}\rangle = 0$. Here $\langle\cdot,\cdot\rangle$ is the Euclidean inner product. Consequently, $\mathbb{E}\|g - \mathbf{q}\|_2^2 = V_{\mathrm{e}} + V_{\mathrm{o}}$. The independent estimator averages K independent copies of g, so its mean-squared error is $(V_{\mathrm{e}} + V_{\mathrm{o}})/K$. Each opposite pair equals one copy of g_{e}. Averaging $K/2$ independent pairs therefore gives error $2V_{\mathrm{e}}/K$. This proves Equation 17. □

Two examples illustrate this classical antithetic variance comparison (Hammersley & Morton, 1956; Ren et al., 2019). For an affine response $g(\boldsymbol{\epsilon}) = \mathbf{a} + \mathbf{R}\boldsymbol{\epsilon}$, where $\mathbf{a}$ is a fixed vector and $\mathbf{R}$ a fixed matrix, the symmetric estimator and center prediction both equal $\mathbf{q} = \mathbf{a}$. In contrast, for the scalar response $g(\epsilon) = \epsilon^2$ with $\epsilon \sim \mathcal{N}(0,1)$, $\mathbf{q} = 1$, $V_{\mathrm{e}} = 2$, and $V_{\mathrm{o}} = 0$. The symmetric and independent errors are $4/K$ and $2/K$; the center's squared error is one.

Conditioned on the state entering a time step, the variance comparison applies to its first update with newly sampled probes. Later states depend on the reused probes; Appendix C.2 characterizes their local symmetric structure through a pathwise second-order remainder bound.

C.2 LOCAL DEVIATION FROM THE CENTER

Proposition 2 (Second-order remainder)**.** Fix $t \in (0,1)$ and a symmetric probe set $\mathcal{E}$. Suppose $\mathcal{D}_t^\theta$ is continuously differentiable with an L_t-Lipschitz Jacobian along the input segments from $t\mathbf{x}$ to $t\mathbf{x} \pm (1-t)\boldsymbol{\epsilon}_k$. Then

$$\left\|\overline{\mathcal{D}}_{t,\mathcal{E}}^\theta(\mathbf{x}) - \mathcal{D}_t^\theta(t\mathbf{x})\right\|_2 \le \frac{L_t(1-t)^2}{2K}\sum_{k=1}^{K}\|\boldsymbol{\epsilon}_k\|_2^2. \tag{18}$$

where $L_t \ge 0$ bounds Jacobian variation in operator norm.

Proof. Write $F = \mathcal{D}_t^\theta$, $\mathbf{u} = t\mathbf{x}$, and $\mathbf{h} = (1-t)\boldsymbol{\epsilon}$ for one pair. Taylor's formula gives

$$F(\mathbf{u} \pm \mathbf{h}) = F(\mathbf{u}) \pm \nabla F(\mathbf{u})\mathbf{h} + \mathbf{r}^{\pm}, \qquad \|\mathbf{r}^{\pm}\|_2 \le L_t\|\mathbf{h}\|_2^2/2,$$

where ∇F is the Jacobian and $\mathbf{r}^{\pm}$ are the remainders. The linear terms sum to zero. Summing the remainder bounds over all $K/2$ pairs proves Equation 18. □

Proposition 2 also permits a state depending on the probes, subject to the stated smoothness condition. The reference is the center prediction. For Gaussian probes and a uniform deterministic L_t, the expected bound is $L_t(1-t)^2 d/2$, scaling linearly with dimension d and independently of K. This is a local structural bound relative to the center, rather than a restoration-error guarantee. The K-dependent sampling error relative to $\mathcal{Q}_t^\theta(\mathbf{x})$ is instead described by Proposition 1.

C.3 Exact correction and attained CG error

For the quadratic fidelity in Equation 13, perturbations of the prior estimate pass through the exact data step according to

$$\mathcal{C}_t(\mathbf{z}+\boldsymbol{\eta}) - \mathcal{C}_t(\mathbf{z}) = \mathbf{B}_t\boldsymbol{\eta}, \qquad \mathbf{B}_t = \beta(\mathbf{A}+\beta\mathbf{I})^{-1}, \quad \|\mathbf{B}_t\|_2 \le 1. \tag{19}$$

where $\boldsymbol{\eta}$ is a perturbation of the prior estimate, $\mathbf{A} = \mathbf{H}^\top\mathbf{H}/\sigma_y^2$ is the measurement Gram matrix and $\beta = 1/w_t^2 > 0$ is the proximal precision.

Let $\boldsymbol{\Gamma}_t = \mathbf{A} + \beta\mathbf{I}$ and $\mathbf{b} = \mathbf{H}^\top\mathbf{y}/\sigma_y^2$. Then $\mathcal{C}_t(\mathbf{z}) = \boldsymbol{\Gamma}_t^{-1}(\mathbf{b}+\beta\mathbf{z})$. Subtracting the systems for two estimates proves Equation 19. The eigenvalues of $\mathbf{B}_t = \beta\boldsymbol{\Gamma}_t^{-1}$ are $\beta/(\lambda_i(\mathbf{A})+\beta) \in (0,1]$, where $\lambda_i(\mathbf{A})$ are the nonnegative eigenvalues of $\mathbf{A}$. They equal one on the measurement null space, where estimation errors are unchanged.

For any finite-CG output $\widetilde{\mathbf{x}}$ and input estimate $\mathbf{z}$, define its attained algebraic residual by $\boldsymbol{\rho} = \boldsymbol{\Gamma}_t\widetilde{\mathbf{x}} - \mathbf{b} - \beta\mathbf{z}$. Direct subtraction yields

$$\widetilde{\mathbf{x}} - \mathcal{C}_t(\mathbf{z}) = \boldsymbol{\Gamma}_t^{-1}\boldsymbol{\rho}, \qquad \|\widetilde{\mathbf{x}} - \mathcal{C}_t(\mathbf{z})\|_2 \le \frac{\|\boldsymbol{\rho}\|_2}{\beta}. \tag{20}$$

where $\boldsymbol{\rho}$ measures the error in satisfying the linear system; Equation 16 instead measures successive image changes. The bound follows from $\boldsymbol{\Gamma}_t \succeq \beta\mathbf{I}$ and holds even at the solver cap.

Combining Equations 19 and 20, for any reference estimate $\mathbf{z}_0$,

$$\|\widetilde{\mathbf{x}} - \mathcal{C}_t(\mathbf{z}_0)\|_2 \le \|\mathbf{B}_t(\mathbf{z}-\mathbf{z}_0)\|_2 + \|\boldsymbol{\rho}\|_2/\beta.$$

This separates estimation error from numerical error. Setting $\mathbf{z}_0 = \mathbf{q}$ uses the expected prediction as the reference. For either fixed-state estimator $\widehat{\mathbf{q}}$, its corrected mean-squared error relative to $\mathcal{C}_t(\mathbf{q})$ follows directly by applying $\mathbf{B}_t$ to the prediction error. Equation 17 then holds with V_e and V_o replaced by $\mathbb{E}\|\mathbf{B}_t(g_\mathrm{e}-\mathbf{q})\|_2^2$ and $\mathbb{E}\|\mathbf{B}_t g_\mathrm{o}\|_2^2$, respectively. Thus the observation operator weights the variance comparison, and the attained CG residual quantifies the additional numerical error. Section 5.3 examines prediction error and restoration quality separately.

D Main-benchmark standard deviations

Table 2 reports standard deviations for the same 100 validation images per dataset as the mean-only main table. Values are ± standard deviations and describe variation across images.

E Supplementary ablation and sensitivity protocols

E.1 Prior-estimation ablation protocol

The CelebA box-inpainting ablation uses a shared OT prior with optional extensions disabled. Its update allocation differs from the main-benchmark configuration in Appendix F.

For each of 16 diagnostic images, Figure 2(a) compares estimators at common states from a deterministic Center trajectory with $N = 20$, $M = 2$, at time-step indices 1, 5, 10, 15, and 20; the horizontal axis shows the corresponding flow time t, rather than the step index n. The independent Monte Carlo reference for Equation 8 pools two disjoint streams of 1,024 predictions. Reference probes do not overlap the tested probes. Prediction MSE is the mean squared difference from this 2,048-probe reference, not from the clean image. Independent and Symmetric use 64 probe repeats per state; Center is evaluated once. Repeats are averaged within each image before computing means and 95% image-cluster bootstrap intervals with 10,000 resamples, using images as the statistical sampling units.

Table 2: Image-wise sample standard deviations corresponding to Table 1, using the same 100 validation images per dataset. Dataset, method, and task order match the main table; N/A denotes unavailable results.

Dataset	Method	Denoising			Gaussian deblurring			Super-resolution			Box inpainting			Random inpainting		
		PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
AFHQ Cat 256×256	PnP-GS	±1.93	±0.108	±0.083	±1.90	±0.068	±0.060	±4.04	±0.119	±0.109	±2.27	±0.015	±0.022	±1.93	±0.099	±0.083
	DDRM	±1.62	±0.025	±0.038	±2.42	±0.064	±0.055	±1.78	±0.104	±0.077	±4.24	±0.015	±0.032	±2.55	±0.032	±0.015
	DPS	±0.96	±0.042	±0.047	±3.50	±0.152	±0.135	±2.07	±0.055	±0.064	±2.09	±0.042	±0.035	±0.87	±0.072	±0.086
	PnP-Diff	±0.27	±0.075	±0.101	N/A	N/A	N/A	±2.00	±0.091	±0.075	±2.22	±0.037	±0.030	±2.38	±0.048	±0.041
	D-Flow	±0.02	±0.074	±0.100	±2.40	±0.070	±0.035	±2.16	±0.069	±0.047	±2.43	±0.015	±0.013	±2.85	±0.040	±0.014
	Flow-Priors	±0.15	±0.079	±0.082	±2.57	±0.086	±0.040	±1.29	±0.054	±0.067	±2.81	±0.019	±0.028	±1.99	±0.022	±0.025
	OT-ODE	±1.70	±0.038	±0.022	N/A	N/A	N/A	±2.53	±0.080	±0.027	±2.83	±0.018	±0.023	±2.33	±0.051	±0.028
	FlowDPS	±1.86	±0.059	±0.041	±2.33	±0.098	±0.056	±1.89	±0.086	±0.056	±2.72	±0.032	±0.037	±1.42	±0.066	±0.046
	FlowChef	±0.96	±0.054	±0.036	±1.03	±0.085	±0.045	±1.10	±0.045	±0.048	±1.45	±0.056	±0.044	±1.24	±0.052	±0.039
	PnP-Flow	±1.68	±0.035	±0.041	±2.57	±0.076	±0.070	±1.54	±0.099	±0.060	±3.66	±0.018	±0.029	±2.85	±0.024	±0.012
	FLOWER	±1.62	±0.026	±0.034	±2.46	±0.064	±0.059	±1.75	±0.084	±0.049	±3.81	±0.013	±0.019	±2.71	±0.026	±0.011
	SPIR	±1.61	±0.020	±0.022	±2.40	±0.061	±0.050	±2.35	±0.067	±0.038	±3.31	±0.012	±0.018	±3.00	±0.024	±0.012
CelebA 128×128	PnP-GS	±1.53	±0.111	±0.069	±1.60	±0.012	±0.017	±3.17	±0.040	±0.041	±1.50	±0.016	±0.030	±0.72	±0.069	±0.080
	DDRM	±1.22	±0.019	±0.015	±1.73	±0.014	±0.009	±2.53	±0.025	±0.016	±3.35	±0.012	±0.070	±2.37	±0.020	±0.012
	DPS	±1.50	±0.105	±0.100	±2.14	±0.111	±0.124	±1.32	±0.033	±0.030	±1.77	±0.044	±0.045	±0.48	±0.063	±0.094
	PnP-Diff	±0.25	±0.067	±0.098	N/A	N/A	N/A	±1.54	±0.026	±0.030	±2.40	±0.040	±0.051	±1.89	±0.037	±0.026
	D-Flow	±0.03	±0.064	±0.103	±3.53	±0.160	±0.171	±1.27	±0.045	±0.028	±2.25	±0.043	±0.024	±2.69	±0.023	±0.009
	Flow-Priors	±0.15	±0.068	±0.084	±1.17	±0.020	±0.033	±1.37	±0.022	±0.040	±2.77	±0.021	±0.038	±1.89	±0.024	±0.030
	OT-ODE	±1.29	±0.025	±0.016	N/A	N/A	N/A	±2.21	±0.024	±0.008	±2.91	±0.019	±0.024	±1.83	±0.035	±0.027
	FlowDPS	±1.24	±0.021	±0.014	±1.84	±0.028	±0.030	±1.87	±0.068	±0.060	±2.69	±0.036	±0.038	±2.08	±0.064	±0.051
	FlowChef	±0.41	±0.052	±0.051	±1.68	±0.036	±0.027	±1.66	±0.029	±0.027	±2.84	±0.021	±0.026	±1.18	±0.054	±0.058
	PnP-Flow	±1.26	±0.022	±0.022	±1.90	±0.012	±0.023	±1.71	±0.023	±0.019	±2.46	±0.016	±0.026	±2.38	±0.015	±0.008
	FLOWER	±1.27	±0.018	±0.016	±1.88	±0.013	±0.014	±2.00	±0.018	±0.015	±2.62	±0.013	±0.023	±2.37	±0.015	±0.008
	SPIR	±1.19	±0.016	±0.008	±1.78	±0.012	±0.011	±2.12	±0.017	±0.012	±2.68	±0.013	±0.022	±2.55	±0.015	±0.006
FFHQ 256×256	PnP-GS	±1.15	±0.078	±0.060	±1.89	±0.053	±0.061	±1.32	±0.054	±0.047	±2.07	±0.014	±0.029	±1.02	±0.081	±0.082
	DDRM	±1.20	±0.034	±0.038	±2.04	±0.047	±0.048	±2.26	±0.063	±0.050	±5.47	±0.026	±0.053	±2.37	±0.028	±0.013
	DPS	±0.90	±0.045	±0.028	±3.20	±0.124	±0.103	±1.70	±0.046	±0.050	±2.88	±0.031	±0.043	±1.05	±0.095	±0.087
	PnP-Diff	±0.19	±0.047	±0.092	N/A	N/A	N/A	±2.16	±0.054	±0.046	±1.11	±0.027	±0.025	±2.24	±0.034	±0.017
	D-Flow	±0.02	±0.045	±0.093	±2.05	±0.053	±0.029	±2.71	±0.068	±0.074	±1.84	±0.007	±0.007	±2.42	±0.028	±0.014
	Flow-Priors	±0.11	±0.041	±0.065	±1.76	±0.047	±0.046	±1.06	±0.069	±0.053	±1.90	±0.008	±0.015	±1.76	±0.019	±0.022
	OT-ODE	±1.26	±0.023	±0.016	N/A	N/A	N/A	±2.68	±0.073	±0.036	±3.17	±0.020	±0.041	±2.22	±0.045	±0.025
	FlowDPS	±1.48	±0.042	±0.041	±2.24	±0.065	±0.057	±2.23	±0.075	±0.094	±3.25	±0.031	±0.043	±1.86	±0.047	±0.058
	FlowChef	±1.12	±0.050	±0.045	±1.28	±0.071	±0.055	±1.33	±0.046	±0.060	±1.66	±0.055	±0.057	±1.46	±0.045	±0.073
	PnP-Flow	±1.43	±0.036	±0.038	±1.89	±0.063	±0.069	±1.82	±0.062	±0.061	±1.66	±0.036	±0.042	±1.57	±0.063	±0.058
	FLOWER	±1.25	±0.019	±0.030	±2.03	±0.047	±0.049	±2.18	±0.063	±0.038	±2.32	±0.006	±0.010	±2.20	±0.018	±0.012
	SPIR	±1.19	±0.014	±0.018	±1.93	±0.042	±0.040	±2.38	±0.049	±0.028	±2.44	±0.006	±0.010	±2.35	±0.016	±0.011
Imagenette2 128×128	PnP-GS	±1.82	±0.114	±0.104	±2.28	±0.063	±0.064	±2.58	±0.120	±0.136	±2.42	±0.032	±0.065	±1.62	±0.108	±0.113
	DDRM	±1.77	±0.043	±0.043	±2.46	±0.036	±0.018	±3.10	±0.065	±0.039	±4.06	±0.016	±0.046	±2.83	±0.053	±0.022
	DPS	±1.91	±0.131	±0.097	±4.72	±0.178	±0.180	±2.77	±0.091	±0.051	±2.60	±0.062	±0.059	±0.66	±0.087	±0.124
	PnP-Diff	±0.38	±0.115	±0.104	N/A	N/A	N/A	±2.53	±0.058	±0.039	±3.34	±0.057	±0.056	±2.64	±0.073	±0.043
	D-Flow	±0.46	±0.117	±0.108	±4.48	±0.163	±0.221	±7.41	±0.237	±0.386	±3.21	±0.043	±0.040	±3.29	±0.086	±0.027
	Flow-Priors	±0.22	±0.111	±0.075	±1.88	±0.037	±0.033	±2.55	±0.047	±0.045	±2.77	±0.022	±0.068	±2.26	±0.049	±0.032
	OT-ODE	±1.55	±0.064	±0.031	N/A	N/A	N/A	±4.55	±0.140	±0.116	±2.45	±0.025	±0.040	±2.50	±0.073	±0.055
	FlowDPS	±1.68	±0.051	±0.030	±2.62	±0.085	±0.055	±1.94	±0.116	±0.095	±2.44	±0.078	±0.055	±2.28	±0.112	±0.086
	FlowChef	±0.71	±0.088	±0.049	±2.48	±0.087	±0.045	±2.44	±0.058	±0.038	±2.45	±0.032	±0.037	±1.79	±0.084	±0.082
	PnP-Flow	±1.65	±0.065	±0.057	±2.54	±0.036	±0.051	±2.51	±0.066	±0.041	±3.22	±0.025	±0.056	±2.77	±0.047	±0.020
	FLOWER	±1.61	±0.041	±0.036	±2.36	±0.034	±0.034	±2.89	±0.052	±0.033	±3.50	±0.019	±0.051	±2.74	±0.047	±0.018
	SPIR	±1.60	±0.034	±0.016	±2.25	±0.030	±0.024	±3.06	±0.042	±0.032	±3.47	±0.018	±0.049	±2.92	±0.042	±0.020

Figure 2(b) and (c) use 100 images and five probe repeats per image for random methods; Center is evaluated once. Fixed update allocations use $N = 20$ without inner-loop early stopping: at 80 and 160 NFE, Center uses $K = 1$, $M = 4, 8$, respectively, while Independent and Symmetric use $K = 2$, $M = 2, 4$. After averaging repeats within images, 95% image-cluster bootstrap intervals are computed separately for each method mean.

These results concern the evaluated CelebA box-inpainting setting. The variance condition in Appendix C.1 concerns centered response variation, which differs from prediction error to the Monte Carlo reference.

Empirical assessment of the variance condition. Figure 8 examines the condition in Proposition 1 at fixed restoration states. For CelebA box inpainting and AFHQ-Cat super-resolution, the diagnostic uses 16 validation images selected with a fixed seed, with each setting's prior, degradation, and restoration configuration held unchanged. States are recorded before the first inner update at five approximately equally spaced time-step indices, including the first and last (1, 11, 21, 30,

and 40 when $N = 40$). At each fixed state, a separate random stream supplies $R = 64$ independent Gaussian probes and their negatives. These diagnostic evaluations do not enter the restoration trajectory. The horizontal axis uses the actual flow times t_n, rather than normalized step indices.

Using the response definitions in Appendix C.1, the centered per-pixel variances are computed separately for each image:

$$\widehat{V}_{\mathrm{e}} = \frac{1}{(R-1)d} \sum_{r=1}^{R} \|g_{\mathrm{e}}(\boldsymbol{\epsilon}_r) - \bar{g}_{\mathrm{e}}\|_2^2, \quad \widehat{V}_{\mathrm{o}} = \frac{1}{(R-1)d} \sum_{r=1}^{R} \|g_{\mathrm{o}}(\boldsymbol{\epsilon}_r) - \bar{g}_{\mathrm{o}}\|_2^2,$$
$$A(t) = 100 \, \frac{\overline{V}_{\mathrm{o}}(t) - \overline{V}_{\mathrm{e}}(t)}{\overline{V}_{\mathrm{o}}(t) + \overline{V}_{\mathrm{e}}(t)}. \tag{21}$$

where R is the number of independent diagnostic pairs, $d = CHW$ is the number of image entries for C channels, height H, and width W, $\bar{g}_{\mathrm{e}}$ and $\bar{g}_{\mathrm{o}}$ are the respective response means over the R probes, and $\overline{V}_{\mathrm{e}}(t)$ and $\overline{V}_{\mathrm{o}}(t)$ average the per-image variance estimates at time t. The normalized variance advantage $A(t)$, defined as the normalized difference between the two variances, estimates the percentage reduction in estimation MSE implied by Equation 17 at matched prior-evaluation count; positive values support the variance condition. It concerns estimation of $\mathcal{Q}_t^\theta(\mathbf{x})$, not error relative to the clean image. Pointwise 95% confidence intervals use 10,000 image-level bootstrap resamples, retaining all time points of each sampled image and recomputing the ratio after averaging variances. Points with both variances at numerical precision are omitted as undefined.

Figure 8 shows a positive normalized variance advantage at all displayed times in both settings. For CelebA box inpainting, the advantage increases toward later flow times; for AFHQ-Cat super-resolution, it remains high throughout the sampled range and increases more modestly. These observations support the variance-reduction condition for symmetric prediction at the evaluated states, complementing the restoration-quality comparisons in Section 5.3. This reduction concerns sampling error around the expected prediction at a fixed state. Final restoration error also depends on the expected prediction's bias relative to the clean image and the subsequent restoration trajectory. As shown in Appendix C.3, exact data correction weights prediction perturbations by $\mathbf{B}_t$, attenuating measured directions while leaving the measurement null space unchanged. The normalized variance advantage therefore does not imply a proportional reduction in final image error.

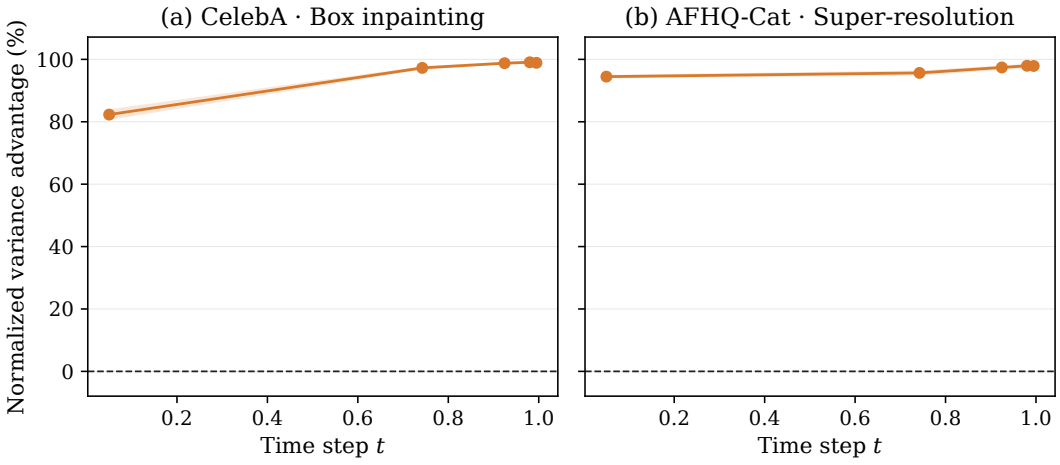

Figure 8: Empirical variance-condition diagnostic on (a) CelebA box inpainting and (b) AFHQ-Cat super-resolution. The horizontal axis is actual flow time t, and the vertical axis is the normalized variance advantage $A(t)$ (%) in Equation 21. Positive values indicate the lower-estimation-error condition for symmetric averaging at matched prior-evaluation count; the dashed line marks zero.

E.2 Probe Pairing and Within-Time Reuse

We evaluate probe pairing and within-time reuse through the two-factor ablation in Table 3. Independent uses two independent Gaussian probes; Symmetric uses a Gaussian probe and its negative.

Table 3: Probe pairing and within-time reuse at 160 NFE under the supplementary configurations in Appendix E.2. Cells report means with 95% image-cluster bootstrap intervals in brackets.

		CelebA box inpainting		AFHQ-Cat super-resolution	
Probes	Sampling strategy	PSNR ↑	LPIPS ↓	PSNR ↑	LPIPS ↓
Independent	Resampled	30.091 [29.625, 30.554]	0.0325 [0.0294, 0.0363]	26.263 [25.782, 26.749]	0.1682 [0.1602, 0.1762]
Independent	Frozen	29.725 [29.262, 30.178]	0.0292 [0.0263, 0.0326]	25.466 [25.012, 25.909]	0.1763 [0.1684, 0.1846]
Symmetric	Resampled	30.461 [29.980, 30.949]	0.0352 [0.0317, 0.0394]	26.708 [26.206, 27.209]	0.1667 [0.1590, 0.1744]
Symmetric	Frozen	30.448 [29.965, 30.931]	0.0351 [0.0315, 0.0392]	26.682 [26.183, 27.190]	0.1672 [0.1596, 0.1750]

Frozen reuses the probes throughout each time step and redraws them at the next time step, whereas Resampled draws a fresh set at every inner update. Symmetric with Frozen corresponds to SPIR's probe strategy. In this supplementary ablation, CelebA box inpainting uses a linear-time grid, and AFHQ-Cat super-resolution uses a log-precision grid. The coefficient of the quadratic prior term in Equation 13 is $2.5/(2w_t^2)$ and $0.003/(2w_t^2)$, respectively. Each setting holds its grid and coefficient fixed across all four variants, with $K = 2$, $M = 2$, $N = 40$, and early stopping disabled, giving 160 NFE regardless of reuse. The clean images, observations, masks, measurement noise, and random seeds are matched across variants. Metrics are computed separately for each restoration and averaged over five random trials per image, then over 100 images per setting. Confidence intervals use image-cluster bootstrap resampling, and paired comparisons use image-level metric differences.

Symmetric probes improve PSNR over independent probes under both sampling strategies in both settings, supported by paired image-level comparisons. Their LPIPS effect is task-dependent: pairing increases LPIPS for CelebA box inpainting but reduces it for AFHQ-Cat super-resolution. Probe reuse does not provide a consistent additional gain. With symmetric probes, the PSNR change on CelebA is small and its paired confidence interval includes zero, whereas reuse slightly reduces PSNR and increases LPIPS on AFHQ-Cat. With independent probes, reuse lowers PSNR in both settings, while its LPIPS effect varies by task. In the SPIR update, reuse keeps the sampled predictor fixed within each time step, allowing the relative-update test to compare successive states under the same probe set.

E.3 Parameter sensitivity and inner-loop early stopping

Figure 3 summarizes sensitivity across five restoration tasks. The CelebA denoising and box-inpainting tolerance and time-step sweeps use 32 test images and seed 0; their reported standard deviations describe image-wise variation under this fixed seed. Increasing N changes the computation budget, whereas the tolerance can change realized NFE through inner-loop early stopping. Under the strict test in Algorithm 1, zero tolerance disables inner-loop early stopping. Section 5.3 provides the controlled prior-estimation comparisons.

The inner-loop early-stopping comparison protocol varies $\varepsilon_{\mathrm{upd}}$ at fixed K, N, M, grid, and observations, including zero tolerance as the fixed-update reference. Quality-cost comparisons use PSNR/LPIPS, actual NFE, latency, and threshold- and budget-exit fractions, with fixed update allocations as the reference.

F Main-benchmark configurations

Tables 4, 5, 6, and 7 list the configurations for AFHQ-Cat, CelebA, FFHQ, and Imagenette2, respectively, used in Table 1. Each dataset has a separate table covering all five tasks. In the task order denoising, Gaussian deblurring, super-resolution, box inpainting, and random inpainting, the measurement-noise standard deviations are 0.2, 0.05, 0.05, 0.05, and 0.01, respectively. The operators are the identity, FFT-based Gaussian convolution, decimation, a centered box mask, and a random mask. All datasets use a 61×61 blur kernel and a random missing-pixel ratio of 0.7.

Table 4: Selected solver configurations for AFHQ-Cat (256×256). The Gaussian blur standard deviation is 3.0, the super-resolution factor is 4, and the box-mask half-size is 40. N/A denotes not applicable or unsupported.

Method	Parameter	Gaussian denoising	Gaussian deblurring	Super-resolution	Box inpainting	Random inpainting
PnP-GS	max_iter	30	30	50	30	30
	lr_pnp	1.0	1.0	2.0	1.0	1.0
	α	0.5	0.5	1.0	0.5	0.5
	algo	pgd	pgd	pgd	pgd	pgd
	σ_factor	1.0	1.0	5.0	1.0	1.0
DDRM	max_iter	100	100	100	100	100
	η	0.85	0.85	0.85	0.85	0.85
	η_b	1.0	1.0	1.0	1.0	1.0
DPS	max_iter	1000	1000	1000	1000	1000
	η	1.0	1.0	1.0	1.0	1.0
PnP-Diff	λ	7.0	N/A	100.0	7.0	7.0
	ζ	0.3	N/A	1.0	0.3	0.3
	σ	0.05	N/A	0.05	0.05	0.05
	max_iter	100	N/A	100	100	100
D-Flow	steps_euler	6	6	6	6	6
	λ	0.01	0.01	0.001	0.01	0.01
	α	0.1	0.1	0.1	0.1	0.1
	max_iter	20	20	20	20	20
	LBFGS_iter	20	20	20	20	20
	start_time	0.0	0.0	0.0	0.0	0.0
Flow-Priors	start_time	0.0	0.0	0.0	0.0	0.0
	K	1	1	1	1	1
	N	100	100	100	100	100
	λ	1000	1000	1000	1000	1000
	η	0.01	0.01	0.01	0.01	0.01
OT-ODE	steps_ode	100	N/A	100	100	100
	start_time	0.2	N/A	0.2	0.2	0.2
	γ	constant	N/A	constant	constant	constant
FlowDPS	steps	28	28	28	28	28
	dc_steps	3	3	3	3	3
	dc_lr	15.0	15.0	15.0	15.0	15.0
	stochastic	True	True	True	True	True
	num_samples	1	1	1	1	1
FlowChef	steps	100	100	100	100	100
	gradient_scale	500.0	500.0	500.0	500.0	500.0
	num_samples	1	1	1	1	1
PnP-Flow	steps_pnp	100	100	100	100	200
	lr_pnp	1.0	1.0	1.0	1.0	1.0
	γ_style	$(1-t)^\alpha$	$(1-t)^\alpha$	$(1-t)^\alpha$	$(1-t)^\alpha$	$(1-t)^\alpha$
	num_samples	5	5	5	5	5
	α	0.8	0.01	0.3	0.5	0.01
FLOWER	steps	100	100	100	100	100
	num_samples	5	5	5	5	5
SPIR	K	2	2	2	2	2
	M	2	2	1	2	2
	N	40	20	60	20	40

Table 5: Selected solver configurations for CelebA (128×128). The Gaussian blur standard deviation is 1.0, the super-resolution factor is 2, and the box-mask half-size is 20. N/A denotes not applicable or unsupported.

Method	Parameter	Gaussian denoising	Gaussian deblurring	Super-resolution	Box inpainting	Random inpainting
PnP-GS	max_iter	30	30	30	30	30
	lr_pnp	1.0	1.0	1.0	1.0	1.0
	α	0.5	0.5	0.5	0.5	0.5
	algo	pgd	pgd	pgd	pgd	pgd
	σ_factor	1.0	1.0	1.0	1.0	1.0
DDRM	max_iter	100	100	100	100	100
	η	0.85	0.85	0.85	0.85	0.85
	η_b	1.0	1.0	1.0	1.0	1.0
DPS	max_iter	1000	1000	1000	1000	1000
	η	1.0	1.0	1.0	1.0	1.0
PnP-Diff	λ	7.0	N/A	7.0	7.0	7.0
	ζ	0.3	N/A	0.3	0.3	0.3
	σ	0.05	N/A	0.05	0.05	0.05
	max_iter	100	N/A	100	100	100
D-Flow	steps_euler	6	6	6	6	6
	λ	0.01	0.01	0.01	0.01	0.01
	α	0.1	0.1	0.1	0.1	0.1
	max_iter	20	20	20	20	20
	LBFGS_iter	20	20	20	20	20
	start_time	0.0	0.0	0.0	0.0	0.0
Flow-Priors	start_time	0.0	0.0	0.0	0.0	0.0
	K	1	1	1	1	1
	N	100	100	100	100	100
	λ	1000	1000	1000	1000	1000
	η	0.01	0.01	0.01	0.01	0.01
OT-ODE	steps_ode	100	N/A	100	100	100
	start_time	0.2	N/A	0.2	0.2	0.2
	γ	constant	N/A	constant	constant	constant
FlowDPS	steps	28	28	28	28	28
	dc_steps	3	3	3	3	3
	dc_lr	15.0	15.0	15.0	15.0	15.0
	stochastic	True	True	True	True	True
	num_samples	1	1	1	1	1
FlowChef	steps	100	100	100	100	100
	gradient_scale	500.0	500.0	500.0	500.0	500.0
	num_samples	1	1	1	1	1
PnP-Flow	steps_pnp	100	100	100	100	100
	lr_pnp	1.0	1.0	1.0	1.0	1.0
	γ_style	$(1-t)^\alpha$	$(1-t)^\alpha$	$(1-t)^\alpha$	$(1-t)^\alpha$	$(1-t)^\alpha$
	num_samples	5	5	5	5	5
	α	0.8	0.01	0.3	0.5	0.01
FLOWER	steps	100	100	100	100	100
	num_samples	5	5	5	5	5
SPIR	K	2	2	2	2	2
	M	2	2	1	2	2
	N	40	20	80	30	40

Table 6: Selected solver configurations for FFHQ (256×256). The Gaussian blur standard deviation is 3.0, the super-resolution factor is 4, and the box-mask half-size is 40. N/A denotes not applicable or unsupported.

Method	Parameter	Gaussian denoising	Gaussian deblurring	Super-resolution	Box inpainting	Random inpainting
PnP-GS	max_iter	30	30	50	30	30
	lr_pnp	1.0	1.0	2.0	1.0	1.0
	α	0.5	0.5	1.0	0.5	0.5
	algo	pgd	pgd	pgd	pgd	pgd
	σ_factor	1.0	1.0	5.0	1.0	1.0
DDRM	max_iter	100	100	100	100	100
	η	0.85	0.85	0.85	0.85	0.85
	η_b	1.0	1.0	1.0	1.0	1.0
DPS	max_iter	1000	1000	1000	1000	1000
	η	1.0	1.0	1.0	1.0	1.0
PnP-Diff	λ	7.0	N/A	100.0	7.0	7.0
	ζ	0.3	N/A	1.0	0.3	0.3
	σ	0.05	N/A	0.05	0.05	0.05
	max_iter	100	N/A	100	100	100
D-Flow	steps_euler	6	6	6	6	6
	λ	0.01	0.01	0.001	0.01	0.01
	α	0.1	0.1	0.1	0.1	0.1
	max_iter	20	20	20	20	20
	LBFGS_iter	20	20	20	20	20
	start_time	0.0	0.0	0.0	0.0	0.0
Flow-Priors	start_time	0.0	0.0	0.0	0.0	0.0
	K	1	1	1	1	1
	N	100	100	100	100	100
	λ	1000	1000	1000	1000	1000
	η	0.01	0.01	0.01	0.01	0.01
OT-ODE	steps_ode	100	N/A	100	100	100
	start_time	0.2	N/A	0.2	0.2	0.2
	γ	constant	N/A	constant	constant	constant
FlowDPS	steps	28	28	28	28	28
	dc_steps	3	3	3	3	3
	dc_lr	15.0	15.0	15.0	15.0	15.0
	stochastic	True	True	True	True	True
	num_samples	1	1	1	1	1
FlowChef	steps	100	100	100	100	100
	gradient_scale	500.0	500.0	500.0	500.0	500.0
	num_samples	1	1	1	1	1
PnP-Flow	steps_pnp	100	100	100	100	100
	lr_pnp	1.0	1.0	1.0	1.0	1.0
	γ_style	$(1-t)^\alpha$	$(1-t)^\alpha$	$(1-t)^\alpha$	$(1-t)^\alpha$	$(1-t)^\alpha$
	num_samples	5	5	5	5	5
	α	0.8	0.01	0.3	0.5	0.01
FLOWER	steps	100	100	100	100	100
	num_samples	5	5	5	5	5
SPIR	K	2	2	2	2	2
	M	2	2	1	2	2
	N	40	20	80	30	40

Table 7: Selected solver configurations for Imagenette2 (128×128). The Gaussian blur standard deviation is 1.0, the super-resolution factor is 2, and the box-mask half-size is 20. N/A denotes not applicable or unsupported.

Method	Parameter	Gaussian denoising	Gaussian deblurring	Super-resolution	Box inpainting	Random inpainting
PnP-GS	max_iter	30	30	30	30	30
	lr_pnp	1.0	1.0	1.0	1.0	1.0
	α	0.5	0.5	0.5	0.5	0.5
	algo	pgd	pgd	pgd	pgd	pgd
	σ_factor	1.0	1.0	1.0	1.0	1.0
DDRM	max_iter	100	100	100	100	100
	η	0.85	0.85	0.85	0.85	0.85
	η_b	1.0	1.0	1.0	1.0	1.0
DPS	max_iter	1000	1000	1000	1000	1000
	η	1.0	1.0	1.0	1.0	1.0
PnP-Diff	λ	7.0	N/A	7.0	7.0	7.0
	ζ	0.3	N/A	0.3	0.3	0.3
	σ	0.05	N/A	0.05	0.05	0.05
	max_iter	100	N/A	100	100	100
D-Flow	steps_euler	6	6	6	6	6
	λ	0.01	0.01	0.01	0.01	0.01
	α	0.1	0.1	0.1	0.1	0.1
	max_iter	20	20	20	20	20
	LBFGS_iter	20	20	20	20	20
	start_time	0.0	0.0	0.0	0.0	0.0
Flow-Priors	start_time	0.0	0.0	0.0	0.0	0.0
	K	1	1	1	1	1
	N	100	100	100	100	100
	λ	1000	1000	1000	1000	1000
	η	0.01	0.01	0.01	0.01	0.01
OT-ODE	steps_ode	100	N/A	100	100	100
	start_time	0.2	N/A	0.2	0.2	0.2
	γ	constant	N/A	constant	constant	constant
FlowDPS	steps	28	28	28	28	28
	dc_steps	3	3	3	3	3
	dc_lr	15.0	15.0	15.0	15.0	15.0
	stochastic	True	True	True	True	True
	num_samples	1	1	1	1	1
FlowChef	steps	100	100	100	100	100
	gradient_scale	500.0	500.0	500.0	500.0	500.0
	num_samples	1	1	1	1	1
PnP-Flow	steps_pnp	100	100	100	100	100
	lr_pnp	1.0	1.0	1.0	1.0	1.0
	γ_style	$(1-t)^\alpha$	$(1-t)^\alpha$	$(1-t)^\alpha$	$(1-t)^\alpha$	$(1-t)^\alpha$
	num_samples	5	5	5	5	5
	α	0.8	0.01	0.3	0.5	0.01
FLOWER	steps	100	100	100	100	100
	num_samples	5	5	5	5	5
SPIR	K	2	2	2	2	2
	M	2	2	1	2	2
	N	40	40	80	20	40

G Detailed inference resource costs

G.1 Computational cost

For the schedule in Section 3.2, if time step n uses $M_n \leq M$ inner updates, a complete SPIR run using single-step destination predictions requires

$$\mathrm{NFE} = K \sum_{n=1}^{N} M_n \leq KNM, \tag{22}$$

where NFE counts probe-wise flow-network evaluations and M_n is the realized number of inner updates at time step n. The bound in Equation 22 is attained when every time step uses all M updates; K already includes both signs of each pair. The relative-update test requires no additional prior evaluation, and inner-loop early stopping can reduce NFE. For predictors using $s_{n,m}$ evaluations per probe at update m of time step n, the count becomes $K \sum_{n=1}^{N} \sum_{m=1}^{M_n} s_{n,m}$, before optional derivative-based corrections, which require separate accounting. CG incurs additional applications of $\mathbf{H}$ and $\mathbf{H}^\top$. Appendix G.2 reports measured runtime and memory.

The implementation specifies a denominator safeguard of 10^{-12}, a batch-wise maximum for the relative-update test, and a warm start for CG. The CG solver is configured to terminate when the maximum Euclidean norm of its algebraic residual over the batch is below 10^{-5} or its iteration cap is reached. Deterministic network inference is also required for the frozen-probe interpretation.

G.2 Measured latency and memory

Figure 9 reports a separate FFHQ evaluation on 32 test images using an NVIDIA A100 and batch size 1. Task averages exclude unsupported entries; the task sets therefore differ across methods. Across the five tasks, SPIR uses 0.290 to 0.300 GiB and 2.811 to 4.684 s/image, with lower latency than FLOWER in these runs. Table 8 provides all methods' resource measurements and their definitions; Appendix F lists the main-benchmark configurations.

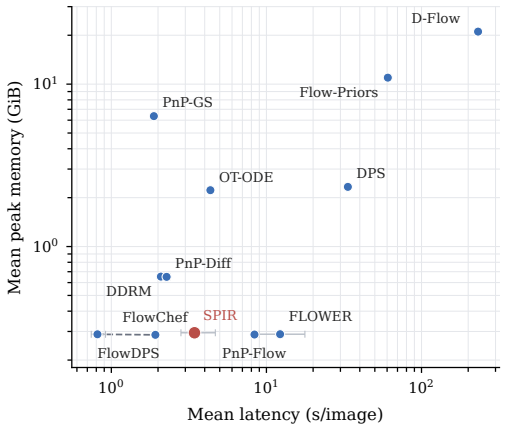

Figure 9: FFHQ inference resource costs on an NVIDIA A100, averaged over each method's supported tasks. Detailed values are reported in Table 8.

G.3 Per-task measurements

Table 8 gives the detailed values behind Figure 9, using the separate 32-image resource-evaluation protocol.

Table 8: FFHQ inference resource costs. The memory and latency panels report peak allocated memory (GiB) and latency (s/image), respectively; lower is better for both. DN, DB, SR, BI, and RI denote denoising, deblurring, super-resolution, box inpainting, and random inpainting. The lowest and second-lowest distinct values among supported methods are bold and underlined, respectively; ties are retained.

Method	DN	DB	SR	BI	RI
Peak allocated memory (GiB)					
PnP-GS	6.350	6.355	6.354	6.352	6.352
DDRM	0.652	0.657	0.652	0.652	0.652
DPS	2.330	2.333	2.330	2.330	2.330
PnP-Diff	0.651	N/A	0.650	0.651	0.651
D-Flow	21.052	21.057	21.053	21.052	21.052
Flow-Priors	10.960	10.968	10.960	10.960	10.960
OT-ODE	2.223	N/A	2.222	2.224	2.224
FlowDPS	<u>0.288</u>	<u>0.290</u>	<u>0.287</u>	<u>0.288</u>	<u>0.288</u>
FlowChef	**0.286**	**0.288**	**0.285**	**0.286**	**0.286**
PnP-Flow	**0.286**	<u>0.290</u>	<u>0.287</u>	**0.286**	**0.286**
FLOWER	<u>0.288</u>	0.291	0.288	<u>0.288</u>	<u>0.288</u>
SPIR	0.292	0.300	0.290	0.297	0.293
Latency (s/image)					
PnP-GS	<u>1.869</u>	<u>1.851</u>	<u>1.869</u>	1.886	<u>1.928</u>
DDRM	2.087	2.113	2.061	2.073	2.098
DPS	33.157	33.492	33.208	33.448	34.248
PnP-Diff	2.203	N/A	2.323	2.175	2.388
D-Flow	229.658	229.221	238.477	229.530	231.026
Flow-Priors	60.509	60.191	60.314	60.274	61.895
OT-ODE	4.268	N/A	4.410	4.267	4.463
FlowDPS	**0.774**	**0.797**	**0.837**	**0.742**	**0.917**
FlowChef	1.909	1.920	1.900	<u>1.885</u>	2.001
PnP-Flow	8.298	8.326	8.384	8.280	8.590
FLOWER	8.944	17.707	10.907	9.071	14.634
SPIR	2.811	4.684	3.572	2.963	3.156

FFHQ test split, 32 batches of one image. Memory is the mean of per-batch CUDA peaks; synchronized latency excludes model loading and metric computation. N/A: unsupported or consistently failed. Hardware: NVIDIA A100.