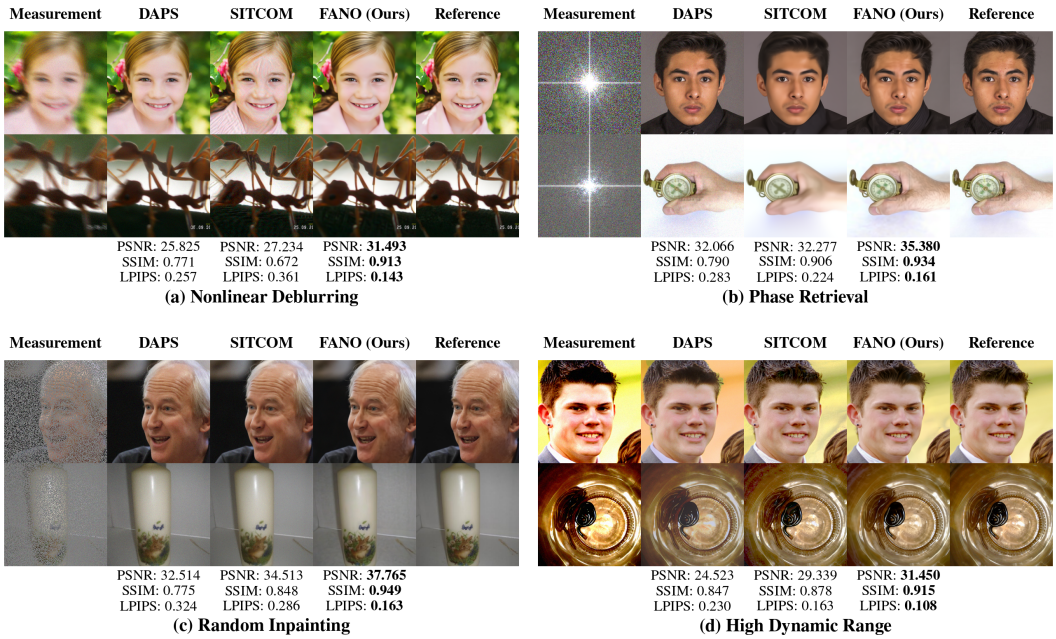


Figure 1: Visual comparison on FFHQ and ImageNet across four inverse problems (σ_y=0.05): (a) nonlinear deblurring, (b) phase retrieval, (c) random inpainting, (d) high dynamic range (HDR). Each panel shows the measurement, DAPS, SITCOM, FANO, and the reference. Metrics below each panel are averaged over the two examples, with the best in **bold**; FANO recovers sharper structures and finer textures across all tasks, with the advantage most pronounced on nonlinear problems.

FANO: Anchored Forward-Noise Optimization for Diffusion Inverse Problems

Anonymous authors
Paper under double-blind review

Abstract

Pre-trained diffusion models have become a widely adopted zero-shot prior for image inverse problems, among which annealed decoupling methods achieve leading reconstruction performance on nonlinear tasks. We find, however, that annealed methods that directly optimize the denoiser input $\boldsymbol{x}_t$ exhibit *radial drift* in later layers: within-layer optimization drives the noise-component norm of $\boldsymbol{x}_t$ away from its typical value, pushing the denoiser out of its high-performance regime and degrading reconstruction quality. To address this, we propose FANO (Anchored Forward-Noise Optimization). At each annealing layer, FANO fixes the previous denoiser estimate as an anchor and optimizes only the forward noise $\boldsymbol{e}$, whose distribution is known, instead of $\boldsymbol{x}_t$, thereby preserving the forward-diffusion structure of the denoiser input. The standard Gaussian prior on $\boldsymbol{e}$ concentrates its norm near the typical value, and deviations degrade denoiser performance. We accordingly decompose $\boldsymbol{e}$ in polar coordinates into a direction and a radial magnitude, and apply a soft potential to regulate the radius, keeping $\boldsymbol{x}_t$ in the high-density region of the forward distribution. On multiple linear and nonlinear inverse problems, FANO matches the strongest current baselines on linear tasks and achieves consistent improvements on nonlinear tasks.

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.

1 Introduction

From a degraded measurement $\boldsymbol{y} \in \mathbb{R}^m$, image inverse problems aim to recover an unknown image $\boldsymbol{x}^\star \in \mathbb{R}^d$, with the two related through the forward measurement model $\boldsymbol{y} = A(\boldsymbol{x}^\star) + \boldsymbol{\xi}$, where $A : \mathbb{R}^d \to \mathbb{R}^m$ is a linear or nonlinear measurement operator and $\boldsymbol{\xi} \in \mathbb{R}^m$ denotes measurement noise. Owing to information loss ($m < d$ or noninvertible A) and noise, inverse problems are typically severely ill-posed and require prior information to constrain the solution space.

Pretrained diffusion models have become the most widely adopted generative priors for zero-shot image inverse problems (Ho et al., 2020; Song et al., 2021b). These methods directly use the denoiser D_t provided by an unconditional diffusion model as an implicit image prior and incorporate measurement consistency into reverse sampling through the known A at inference time, without retraining the model for each degradation type.

Among diffusion-prior methods for inverse problems (Section 2), annealed decoupling methods (Zhang et al., 2025; Alkhouri et al., 2025; Kim et al., 2026) decompose inference into a sequence of within-layer subproblems along an annealing schedule, achieving the strongest current reconstruction performance on nonlinear inverse problems. However, when analyzing SITCOM (Alkhouri et al., 2025), an annealed method that directly optimizes the denoiser input, we find that the measurement residual continues to decrease in later annealing layers while reconstruction quality abruptly collapses (Figure 2(a,c)). Tracking the norm of the noise component in the denoiser input reveals a sharp increase in later layers (Figure 2(a)); controlled experiments show that this deviation is accompanied by degraded denoiser performance (Figure 2(b)). Moreover, the norm deviates more severely in the absence of measurement noise, ruling out the overfitting hypothesis. We refer to this phenomenon as *radial drift*. Section 3.1 presents the full analysis.

Based on this analysis, we propose *Anchored Forward-Noise Optimization* (FANO). Radial drift arises because the measurement-consistency objective does not constrain the norm of the noise component when directly optimizing $\boldsymbol{x}_t$. FANO replaces the denoiser input $\boldsymbol{x}_t$ with the forward noise $\boldsymbol{e}$ as the search variable: each layer fixes the previous layer's denoised estimate as an anchor, keeps the denoiser input within the conditional forward distribution $q_t(\cdot \mid \bar{\boldsymbol{x}})$, and progressively contracts the search region through annealing. High-dimensional Gaussian concentration indicates that $\|\boldsymbol{e}\|$ should concentrate near $\sqrt{d-1}$, and deviations degrade denoiser performance (Figure 2(b)). We therefore decompose $\boldsymbol{e}$ into a direction and a radial magnitude in polar coordinates: the measurement loss drives the directional search, while a concentration-inspired potential anchors the radial magnitude near its typical value, keeping the denoiser input in the high-density region of the forward distribution. Experiments on 5 linear and 3 nonlinear inverse problems on FFHQ and ImageNet show that FANO matches the strongest current baselines on linear tasks and achieves consistent improvements on nonlinear tasks (Figure 1, Table 1). Figure 3 provides an overview of the method.

Our contributions are:

- **Diagnosis of radial drift.** We identify *radial drift* in annealed methods that directly optimize $\boldsymbol{x}_t$: the equivalent noise norm deviates from its typical value in later layers, degrading denoiser performance. Controlled experiments confirm the association and show that radial correction can recover the quality lost at the final layer.
- **Forward-noise optimization with radial control.** We propose FANO, which uses the forward noise $\boldsymbol{e}$ as the within-layer search variable. This fixes the denoiser input to the forward-diffusion form and excludes the anchor from optimization. Exploiting the Gaussian prior, we decompose $\boldsymbol{e}$ into a direction and a radial magnitude, with a concentration-inspired potential to keep the denoiser input in the high-density region.
- **Comprehensive evaluation.** Experiments on 5 linear and 3 nonlinear inverse problems show that FANO matches the strongest baselines on linear tasks and achieves consistent improvements on nonlinear tasks. Ablation studies identify the $\boldsymbol{e}$-search framework as the primary source of gains.

2 Background and Related Work

For zero-shot inverse problem solving, the denoiser provided by a pretrained diffusion model is directly used as an implicit image prior, incorporating measurement consistency into reverse sampling

at inference time without retraining for each degradation type. Existing methods differ primarily in their search strategies and have broadly evolved along three directions: stepwise guidance locally corrects the diffusion state at each reverse step; end-to-end noise optimization encodes the reverse process as a deterministic mapping and performs a global search in the initial noise space; and annealed decoupling decomposes inference into a sequence of subproblems across noise levels, allowing a broader search within each layer.

2.1 Diffusion Models

Diffusion models transform the data distribution $p(\boldsymbol{x}_0)$ into an approximately Gaussian distribution through a forward process that progressively adds noise, and then learn a reverse denoising process to generate data from noise (Ho et al., 2020; Song et al., 2021b). Given noise-schedule parameters (a_t, σ_t), the conditional distribution of the forward process at timestep t is

$$q_t(\boldsymbol{x}_t \mid \boldsymbol{x}_0) = \mathcal{N}(a_t \boldsymbol{x}_0, \sigma_t^2 \boldsymbol{I}_d), \qquad \boldsymbol{x}_t = a_t \boldsymbol{x}_0 + \sigma_t \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_d), \tag{1}$$

Reverse generation is driven by the score function $\nabla_{\boldsymbol{x}_t} \log p_t(\boldsymbol{x}_t)$ (Anderson, 1982; Song et al., 2021a), approximated in practice by a learned network $\boldsymbol{s}_\theta(\boldsymbol{x}_t, t)$. Tweedie's formula (Boys et al., 2024) relates the score to $\mathbb{E}[\boldsymbol{x}_0 \mid \boldsymbol{x}_t]$; substituting $\boldsymbol{s}_\theta$ defines the denoiser

$$D_t(\boldsymbol{x}_t) := \frac{1}{a_t}\big(\boldsymbol{x}_t + \sigma_t^2 \boldsymbol{s}_\theta(\boldsymbol{x}_t, t)\big) \approx \mathbb{E}[\boldsymbol{x}_0 \mid \boldsymbol{x}_t], \tag{2}$$

which provides the minimum mean-square-error estimate of $\boldsymbol{x}_0$ given $\boldsymbol{x}_t$.

2.2 Stepwise Measurement Guidance

DPS (Chung et al., 2023) approximates $p_t(\boldsymbol{y} \mid \boldsymbol{x}_t)$ via the denoised estimate $D_t(\boldsymbol{x}_t)$, enabling guidance for arbitrary forward operators; subsequent work refines the approximation or exploits problem structure (Chung et al., 2022; Song et al., 2023; Meng & Kabashima, 2025). Spectral-domain methods such as DDRM (Kawar et al., 2022) and DDNM (Wang et al., 2023) achieve exact data consistency on decomposable linear operators but do not extend to nonlinear problems. DSG (Yang et al., 2024) and DiffRGD (Liao et al., 2026) exploit the spherical geometry of high-dimensional Gaussians to mitigate guidance-induced distribution shift, the latter performing Riemannian optimization on a sphere of a given radius. Nonetheless, the correction range of stepwise guidance remains restricted to a neighborhood of the current diffusion state, making it difficult to traverse multiple posterior modes in nonlinear problems(Zhang et al., 2025).

2.3 End-to-End Noise Optimization

End-to-end noise optimization treats the reverse process as a deterministic mapping $\mathcal{G}$ and directly solves $\min_{\boldsymbol{z}} \|A(\mathcal{G}(\boldsymbol{z})) - \boldsymbol{y}\|^2$ in the initial noise space (Wang et al., 2024). INO (Jia et al., 2026) constrains the search to the Gaussian typical shell $\|\boldsymbol{z}\| = \sqrt{d}$ via tangent-space optimization; N-HMC (Xia et al., 2026) instead samples the posterior $p(\boldsymbol{z} \mid \boldsymbol{y})$ with Hamiltonian Monte Carlo. This family avoids the $p_t(\boldsymbol{y} \mid \boldsymbol{x}_t)$ approximation but requires backpropagation through the entire generation chain, inherits the nonconvexity of the multistep composition, and loses the coarse-to-fine structure of diffusion.

2.4 Annealed Decoupling

Annealed decoupling methods decompose the reverse process into a sequence of within-layer subproblems along an annealing schedule, retaining diffusion's coarse-to-fine hierarchy(Rissanen et al., 2023; Dieleman, 2024) while allowing a broader search within each layer. Between layers, resampling—adding forward noise conditioned on a clean estimate (Lugmayr et al., 2022)—returns the state to the forward distribution.

DAPS (Zhang et al., 2025) uses $\boldsymbol{x}_0$ as the search variable and samples the within-layer posterior via Langevin dynamics; FAST-DIPS (Kim et al., 2026) replaces the Langevin chain with ADMM splitting for faster convergence (Li et al., 2026; Zhang et al., 2026). SITCOM (Alkhouri et al., 2025) shifts the search variable to the denoiser input $\boldsymbol{x}_t$, optimizing

$$\min_{\boldsymbol{v}_t} \|A(D_t(\boldsymbol{v}_t)) - \boldsymbol{y}\|^2 + \lambda \|\boldsymbol{x}_t - \boldsymbol{v}_t\|^2$$

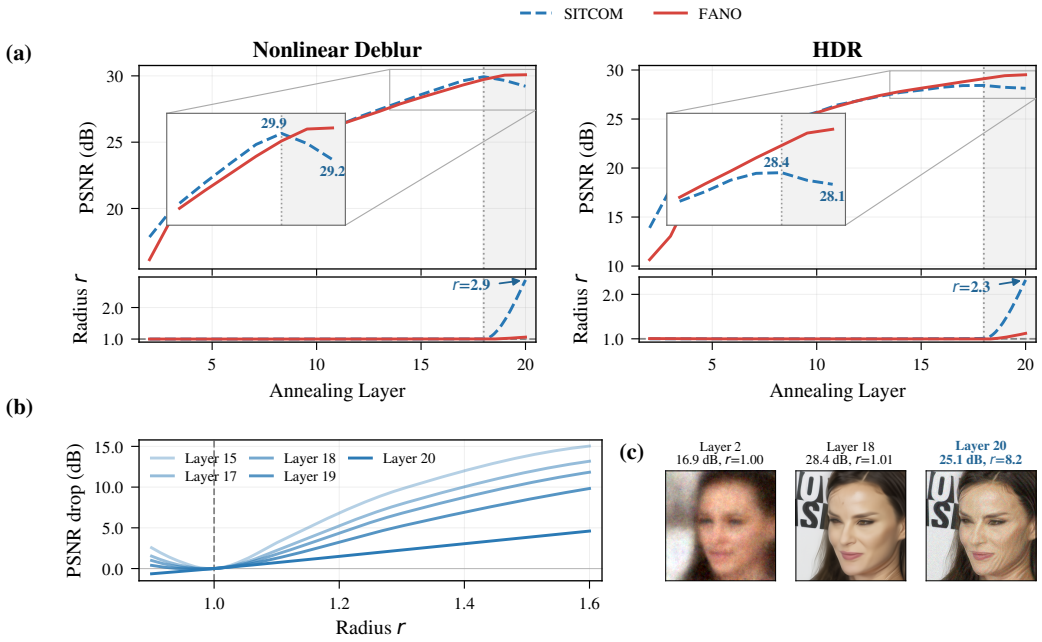


Figure 2: Radial drift on FFHQ (σ_y=0.05, N=20). **(a)** Mean PSNR and normalized radius r on FFHQ100 for SITCOM (Alkhouri et al., 2025) and FANO; insets magnify late-layer PSNR, and shading marks layers after SITCOM's peak. **(b)** Denoiser PSNR drop relative to r=1 when varying the radius along fixed random noise directions. **(c)** SITCOM reconstructions at layers 2, 18, and 20, illustrating late-layer degradation as r increases.

and leveraging the denoiser's implicit bias for regularization. However, its Euclidean regularizer does not directly constrain the distributional structure of $\boldsymbol{v}_t$; within-layer optimization can drive $\boldsymbol{v}_t$ away from the high-density region of the forward distribution (Section 3.1). FANO addresses this by using the forward noise $\boldsymbol{e}$ as the search variable and introducing a concentration-inspired radial potential (Section 3).

3 Method

This section first analyzes radial drift in existing annealed methods (Section 3.1), then introduces an optimization strategy using the forward noise $\boldsymbol{e}$ as the search variable together with a radial control mechanism (Section 3.2), and finally presents the complete algorithm, **FANO** (Anchored Forward-Noise Optimization, Section 3.3).

3.1 Radial Drift in Annealed Decoupling

Annealed decoupling methods share the following core recursion: given an annealing schedule $t_1 > \cdots > t_N$, each layer adds fresh noise to the previous layer's denoised result and searches for a new estimate consistent with the measurement at the current noise level. Taking SITCOM (Alkhouri et al., 2025) as a representative method, we track this recursion (Figure 2(a)) and observe a puzzling phenomenon: on two nonlinear inverse problems, the measurement residual RMSE decreases monotonically throughout, and reconstruction PSNR improves over the first 18 layers, yet abruptly reverses and decreases in the final two layers. The optimizer thus degrades reconstruction quality while successfully matching the measurement more closely. A natural alternative explanation is overfitting to measurement noise, but a control experiment with $\sigma_y = 0$ rules out this hypothesis: without noise, the final-layer PSNR loss increases from 0.7 dB to 3.4 dB for nonlinear deblurring and from 0.3 dB to 0.5 dB for HDR. This more severe degradation indicates that the issue arises from the optimization dynamics themselves rather than noise overfitting.

To trace the cause of this degradation, we examine the variable optimized within each SITCOM layer. SITCOM directly optimizes the denoiser input $\boldsymbol{x}_t$ at each annealing layer to make the denoised output match the measurement as closely as possible. Given the current layer's anchor $\bar{\boldsymbol{x}} := \hat{\boldsymbol{x}}^{k-1}$, the forward process $\boldsymbol{x}_t = a_t\bar{\boldsymbol{x}} + \sigma_t\boldsymbol{e}$ allows any denoiser input to be parameterized by an equivalent forward noise $\boldsymbol{e}$. Define the normalized equivalent radius as $r := \|\boldsymbol{e}\|_2/\sqrt{d-1}$; under $\boldsymbol{e} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_d)$, high-dimensional concentration ensures $r \approx 1$. Tracking r across layers (Figure 2(a), bottom row) reveals an anomalous signal coinciding with the degradation: over the first 18 layers, $r \approx 1.00$, but its mean rises to 2.9 for nonlinear deblurring and 2.3 for HDR in the final layer, with both statistics computed over the full FFHQ100 dataset. Panel (c) visualizes the layerwise reconstructions of a single nonlinear deblurring example: at layer 18, PSNR reaches 28.4 dB ($r = 1.01$); in the final layer, r increases to 8.2 and PSNR drops to 25.1 dB.

Figure 2(b) tests the association between radial deviation and performance degradation through a controlled experiment. Holding a randomly sampled noise direction $\boldsymbol{u} \sim \mathrm{Unif}(\mathbb{S}^{d-1}_{\sqrt{d-1}})$ fixed and scaling only the radius r, we find that denoiser performance is optimal near $r = 1$ at every annealing layer. This test uses random directions, whereas directions obtained through multiple measurement-driven optimization steps are no longer random samples. We therefore further examine the terminal state of the final layer in the same SITCOM run, fixing the anchor and optimized direction but reconstructing the denoiser input as $a_t\bar{\boldsymbol{x}} + \sigma_t(\boldsymbol{e}/r)$ to normalize the equivalent noise's radial magnitude to 1. This increases mean PSNR by 0.15 dB for HDR and 0.49 dB for nonlinear deblurring, recovering most of the final-layer degradation by correcting the radius alone. We refer to this late-stage degradation caused by the radial magnitude deviating from its typical value as *radial drift*. Note that r is measured relative to the estimated anchor rather than the ground truth, so anchor error can partially amplify the observed r. The red curves in Figure 2(a) show the layerwise behavior of our method, FANO (Sections 3.2–3.3), under the same conditions: mean PSNR continues to improve through the final layer, the normalized radius remains near 1, and radial drift disappears.

3.2 Forward-Noise Optimization and Radial Control

The preceding analysis directly links radial drift to the uncontrolled norm of the noise component during within-layer optimization. FANO therefore shifts the search variable from $\boldsymbol{x}_t$ to the forward noise $\boldsymbol{e}$: at each annealing layer, it fixes the previous layer's denoised estimate $\bar{\boldsymbol{x}} := \hat{\boldsymbol{x}}^{k-1}$ as an anchor and parameterizes the denoiser input as $\boldsymbol{x}_t = a_t\bar{\boldsymbol{x}} + \sigma_t\boldsymbol{e}$, using $\boldsymbol{e}$ as the sole within-layer search variable. This parameterization preserves the forward-diffusion structure $q_t(\cdot \mid \bar{\boldsymbol{x}})$ of the denoiser input. Searching over $\boldsymbol{e}$ treats the anchor $\bar{\boldsymbol{x}}$ as a fixed offset excluded from within-layer optimization, so within-layer updates modify only the noise component.

The prior $\mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_d)$ is uniform in direction and highly concentrated in radius (Figure 2(b)), motivating a polar decomposition of $\boldsymbol{e}$ into a radial magnitude r and a direction $\boldsymbol{u}$. The measurement loss drives the directional search, while a concentration-inspired soft constraint controls the radius.

We decompose the forward noise $\boldsymbol{e} \in \mathbb{R}^d$ into a normalized radial magnitude and a direction as $\boldsymbol{e} = r \cdot \boldsymbol{u}$, where $r := \|\boldsymbol{e}\|_2/\sqrt{d-1} > 0$. Let $\mathbb{S} := \mathbb{S}^{d-1}_{\sqrt{d-1}} := \{\boldsymbol{u} \in \mathbb{R}^d : \|\boldsymbol{u}\|_2 = \sqrt{d-1}\}$, so that $\boldsymbol{u} \in \mathbb{S}$. Under the polar change of variables, the joint density of the standard Gaussian prior $\mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_d)$ factorizes as

$$p(r, \boldsymbol{u}) \propto r^{d-1} \exp\left(-\frac{(d-1)\, r^2}{2}\right), \tag{3}$$

where the direction $\boldsymbol{u}$ is uniformly distributed on $\mathbb{S}$ and the radius r follows a normalized χ distribution(Laurent & Massart, 2000). The prior has no directional preference but strongly favors a particular radial range, leading to the following constraint strategy: *constrain only the radial dimension and let the measurement loss fully drive the angular search.*

The negative log of the radial prior yields the radial potential

$$\Phi_d(r) = \frac{d-1}{2}\big(r^2 - 1 - 2\log r\big), \tag{4}$$

which attains its minimum value of zero at $r = 1$ (i.e., $\|\boldsymbol{e}\|_2 = \sqrt{d-1}$; see Appendix A.1 for the derivation), providing soft radial control of the denoiser input inspired by Gaussian radial statistics. Let $\ell_y(\boldsymbol{x}) := \|A(\boldsymbol{x}) - \boldsymbol{y}\|_2^2$ denote the measurement loss. For a fixed annealing layer t and anchor $\bar{\boldsymbol{x}}$,

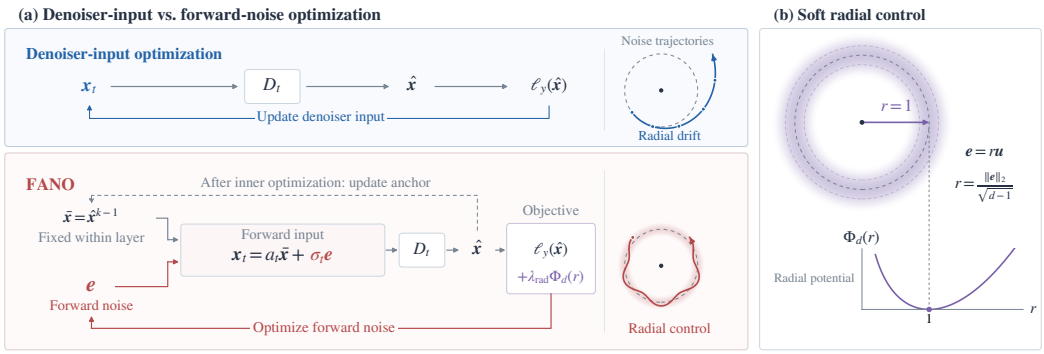


Figure 3: Overview of FANO. **(a)** Blue: direct optimization of $\boldsymbol{x}_t$; red: FANO fixes the anchor $\bar{\boldsymbol{x}}{=}\hat{\boldsymbol{x}}^{k-1}$ and optimizes only the forward noise $\boldsymbol{e}$. Circular diagrams contrast radial drift (blue) with controlled search (red). **(b)** The radial potential $\Phi_d(r)$ is minimized at $r{=}1$, softly penalizing deviations from the typical noise shell. FANO decomposes $\boldsymbol{e}{=}r\boldsymbol{u}$ and jointly optimizes direction $\boldsymbol{u}$ and magnitude r.

the measurement loss along direction $\boldsymbol{u}$ is $F_t(r;\boldsymbol{u}) := \ell_y(D_t(a_t\bar{\boldsymbol{x}}{+}\sigma_t r\boldsymbol{u}))$. Replacing the coefficient prescribed by the prior with a tunable hyperparameter $\lambda_{\text{rad}} \geq 0$ gives FANO's within-layer objective:

$$\min_{r,\boldsymbol{u}} \quad J_\lambda(r;\boldsymbol{u}) := F_t(r;\boldsymbol{u}) + \lambda_{\text{rad}}\,\Phi_d(r) \qquad \text{s.t.} \quad \boldsymbol{u} \in \mathbb{S}, \quad r > 0. \tag{5}$$

The following proposition characterizes the optimization landscape of the one-dimensional radial subproblem $\min_{r>0} J_\lambda(r;\boldsymbol{u})$.

Proposition 1 (Boundedness and benefits of radial relaxation)**.** *Assume $d \geq 2$, $\lambda_{\text{rad}} > 0$, and that $F_t(\cdot\,;\boldsymbol{u})$ is nonnegative and continuous on $(0,\infty)$. Then, $\forall\, \boldsymbol{u} \in \mathbb{S}$:*

(i) $J_\lambda(\cdot\,;\boldsymbol{u})$ attains a global minimum over $r > 0$ at $r_\lambda(\boldsymbol{u})$, and

$$|r_\lambda(\boldsymbol{u}) - 1| \;\leq\; \sqrt{\frac{2\,F_t(1;\boldsymbol{u})}{\lambda_{\text{rad}}(d-1)}}\,. \tag{6}$$

(ii) If $F_t(\cdot\,;\boldsymbol{u})$ is differentiable at $r{=}1$ and $\partial_r F_t(1;\boldsymbol{u}) \neq 0$, then $J_\lambda(r_\lambda(\boldsymbol{u});\boldsymbol{u}) < J_\lambda(1;\boldsymbol{u}) = F_t(1;\boldsymbol{u})$.

Part (i) guarantees that the optimal radius does not diverge: the radial potential anchors r_λ near 1, with the deviation shrinking as λ_{rad} increases. Part (ii) shows that when the measurement loss has a nonzero radial gradient, hard spherical search with $r \equiv 1$ cannot exploit this descent direction, whereas the soft constraint permits a bounded deviation and strictly reduces the objective. This result is consistent with the ablation in Table 2, where FANO outperforms FANO-Sphere on HDR by approximately 0.8 dB. The proof is provided in Appendix A.2, and the strong-regularization asymptotic expansion in Corollary 1.

3.3 Algorithm

FANO proceeds recursively along an annealing schedule $t_1 > t_2 > \cdots > t_N$. Layer k uses the denoised estimate from layer $k-1$, $\hat{\boldsymbol{x}}^{k-1}$, as its anchor. Each layer performs three steps: *(i) Initialization*: sample a random direction $\boldsymbol{u}$ on the sphere and set the radius to $r = 1$; *(ii) Forward-noise optimization*: solve Equation (5) for $r^\star, \boldsymbol{u}^\star$, using Riemannian Adam for the direction $\boldsymbol{u}$ (Absil et al., 2008) and standard Adam for the radius r. Each inner step requires only one forward and one backward pass through the denoiser to obtain gradients for both variables, requiring, over N layers with at most K steps per layer, at most NK NFEs; *(iii) Anchor update*: pass the optimized denoised estimate to the next layer:

$$\hat{\boldsymbol{x}}^k := D_{t_k}\big(a_{t_k}\hat{\boldsymbol{x}}^{k-1} + \sigma_{t_k} r^\star \boldsymbol{u}^\star\big). \tag{7}$$

The initial anchor $\hat{\boldsymbol{x}}^0$ is obtained by denoising pure noise with D_{t_1}.

Within-layer iterations use the measurement residual for early stopping. The current layer terminates when

$$\left\|A\big(D_{t_k}(\boldsymbol{z}_j)\big) - \boldsymbol{y}\right\|_2 \;\le\; \tau_{\text{stop}} := \sigma_y\sqrt{m} \tag{8}$$

where σ_y is the standard deviation of the measurement noise and m is the measurement dimension. The quantity $\sigma_y\sqrt{m}$ is the typical norm of additive Gaussian noise $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \sigma_y^2\boldsymbol{I}_m)$; a residual below this level indicates that optimization is beginning to fit the noise itself. In the noiseless case, $\tau_{\text{stop}} = 0$ and the criterion reduces to a fixed iteration budget. The full pseudocode is provided in Appendix B.1, and optimizer implementation details in Appendix B.2.

4 Experiments

We evaluate FANO on 5 linear tasks (super-resolution, Gaussian deblurring, motion deblurring, box inpainting, and random inpainting) and 3 nonlinear tasks (phase retrieval, high dynamic range (HDR) reconstruction, and nonlinear deblurring).

4.1 Experimental Setup

Datasets, tasks, and metrics Experiments are conducted on FFHQ 256×256 (Karras et al., 2021) and ImageNet 256×256 (Deng et al., 2009), using 100 test images from each dataset. We use publicly available pretrained VP SDE DDPMs: the FFHQ model is from Choi et al. (2021), and the ImageNet model is from Dhariwal & Nichol (2021), without task-specific fine-tuning. Degradation operators for linear tasks follow the standard settings of DAPS (Zhang et al., 2025). For nonlinear tasks, phase retrieval uses an oversampling ratio of 1.5 (a resolution ratio of $384/256$, corresponding to a 384×384 FFT; DAPS (Zhang et al., 2025) denotes this as OS $= 2$), HDR reconstruction uses a clipping factor of 2, and the spatially varying blur kernels for nonlinear deblurring are from BKSE (Tran et al., 2021). The measurement noise level is $\sigma_y = 0.05$ in our experiments; task-specific degradation parameters are detailed in Appendix B.4. We report PSNR↑, SSIM↑ (Wang et al., 2004), and LPIPS↓ (Zhang et al., 2018), with LPIPS consistently computed using a VGG16 feature network.

Methods and configurations We compare against DAPS (Zhang et al., 2025), SITCOM (Alkhouri et al., 2025), FAST-DIPS (Kim et al., 2026), DiffRGD (Liao et al., 2026), and INO (Jia et al., 2026). All methods are rerun using their public code and recommended hyperparameters with the same checkpoints, test images, degradation operators, and noise levels. Table 1 reports each baseline on the tasks supported by its public code: DiffRGD covers linear tasks except box inpainting, and INO does not cover motion deblurring, phase retrieval, or HDR. FANO uses the same algorithmic framework across all tasks (Appendix Algorithm 1), with learning rates $\eta_u = \eta_r = 0.02$ and the noise-ball early-stopping criterion in Equation (8); detailed configurations are listed in Appendix B.5. Following the convention in prior work (Chung et al., 2023; Zhang et al., 2025; Alkhouri et al., 2025), phase retrieval uses 4 independent runs per image and reports the best result; all baselines use the same protocol.

4.2 Quantitative Results

Table 1 summarizes the quantitative results across all tasks. On linear inverse problems, FANO achieves reconstruction quality comparable to or slightly better than the strongest current baselines: it outperforms the second-best method in PSNR by 0.48–1.10 dB on random inpainting, 0.63–0.76 dB on Gaussian deblurring, and 0.50–0.67 dB on motion deblurring, with box inpainting remaining a relative weakness. The performance gap is more pronounced on nonlinear tasks: on FFHQ, FANO exceeds the second-best method in PSNR by 1.89 dB for phase retrieval, 1.52 dB for nonlinear deblurring, and 1.75 dB for HDR reconstruction. Its advantage on nonlinear tasks is particularly pronounced relative to SITCOM, which shares the annealing framework but searches in Euclidean space. This is consistent with the analysis in Section 3.1: radial drift occurs more frequently with nonlinear measurement operators, and soft radial control effectively keeps the denoiser input within its operating regime.

Appendix D compares the computational costs and runtimes of the main methods. FANO's reconstruction quality degrades smoothly as measurement noise increases, with detailed noise-robustness

Table 1: Quantitative results at $\sigma_y = 0.05$, reported as the mean $\pm$ standard deviation over 100 test images. The best results are in **bold**, and the second-best results are underlined.

		FFHQ			ImageNet		
Task	**Method**	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
SR $\times 4$	DAPS	29.35±1.47	0.782±0.028	0.268±0.035	25.47±2.86	0.638±0.087	0.376±0.061
	SITCOM	30.55±1.87	0.868±0.033	0.223±0.034	**27.09±3.84**	**0.757±0.116**	**0.307±0.077**
	FAST-DIPS	29.59±1.75	0.843±0.041	0.241±0.042	26.55±3.65	0.725±0.125	0.325±0.072
	DiffRGD	29.24±1.93	0.838±0.042	**0.214±0.037**	25.96±3.70	0.712±0.136	0.322±0.084
	INO	29.24±1.80	0.809±0.037	0.260±0.038	25.65±3.50	0.694±0.131	0.310±0.090
	FANO	**30.68±1.88**	**0.872±0.032**	0.232±0.036	27.03±3.87	0.755±0.118	0.321±0.082
Gauss. deblur	DAPS	29.61±1.63	0.786±0.031	0.259±0.032	25.88±3.23	0.657±0.105	0.356±0.062
	SITCOM	29.89±1.95	0.849±0.038	0.242±0.038	26.38±3.60	0.726±0.128	0.340±0.083
	FAST-DIPS	29.20±1.92	0.830±0.045	0.259±0.045	25.92±3.74	0.694±0.142	0.358±0.082
	DiffRGD	27.05±5.12	0.780±0.162	0.248±0.106	22.87±3.36	0.564±0.171	0.413±0.091
	INO	28.72±1.88	0.789±0.048	0.275±0.042	24.76±3.46	0.653±0.145	0.334±0.087
	FANO	**30.65±2.04**	**0.868±0.036**	**0.225±0.038**	**27.01±3.95**	**0.746±0.128**	**0.317±0.084**
Motion deblur	DAPS	31.65±1.34	0.836±0.022	0.208±0.029	28.65±3.11	0.769±0.066	0.244±0.055
	SITCOM	32.37±1.70	0.897±0.024	0.188±0.031	29.38±3.68	0.832±0.080	0.235±0.062
	FAST-DIPS	31.38±1.59	0.873±0.035	0.200±0.036	28.79±3.92	0.796±0.106	0.261±0.078
	DiffRGD	29.24±2.07	0.835±0.044	0.213±0.035	23.40±3.44	0.591±0.169	0.421±0.100
	FANO	**33.04±1.82**	**0.909±0.024**	**0.167±0.029**	**29.88±4.19**	**0.840±0.085**	**0.225±0.070**
Inpaint (box)	DAPS	24.32±2.67	0.747±0.031	0.236±0.034	21.05±3.44	0.713±0.036	0.271±0.057
	SITCOM	24.47±2.52	0.737±0.033	0.245±0.030	19.62±2.91	0.699±0.038	0.255±0.060
	FAST-DIPS	25.18±2.45	0.857±0.028	0.182±0.030	**21.53±4.17**	**0.797±0.058**	0.257±0.044
	INO	22.79±2.91	0.827±0.039	0.222±0.041	18.69±3.02	0.739±0.105	0.282±0.070
	FANO	**25.33±2.94**	**0.866±0.028**	**0.156±0.030**	20.40±3.30	0.786±0.040	**0.206±0.040**
Inpaint (rand.)	DAPS	30.72±1.35	0.801±0.020	0.230±0.026	27.35±3.20	0.725±0.066	0.248±0.050
	SITCOM	32.40±1.44	0.863±0.018	0.182±0.030	28.90±3.57	0.814±0.054	0.180±0.053
	FAST-DIPS	31.78±1.57	0.892±0.030	0.185±0.034	29.46±4.11	0.828±0.093	0.214±0.058
	DiffRGD	32.73±1.91	0.912±0.023	0.167±0.030	28.90±4.15	0.825±0.094	0.239±0.070
	INO	32.97±2.14	0.915±0.031	0.166±0.035	28.33±4.55	0.810±0.124	0.218±0.082
	FANO	**34.07±1.97**	**0.931±0.017**	**0.120±0.027**	**29.94±4.61**	**0.864±0.074**	**0.157±0.050**
Phase retrieval	DAPS	30.09±2.96	0.781±0.057	0.239±0.058	22.94±6.22	0.505±0.216	0.411±0.148
	SITCOM	29.21±4.99	0.817±0.178	0.264±0.122	20.49±6.38	0.467±0.274	0.465±0.163
	FAST-DIPS	28.37±3.71	0.849±0.094	0.221±0.085	21.38±6.72	0.548±0.274	0.440±0.191
	FANO	**31.98±4.14**	**0.877±0.118**	**0.187±0.091**	**23.83±7.99**	**0.567±0.289**	**0.388±0.187**
HDR	DAPS	27.48±3.64	0.837±0.082	0.210±0.073	26.07±4.08	0.828±0.096	0.204±0.110
	SITCOM	28.13±3.12	0.810±0.107	0.217±0.086	25.75±3.66	0.771±0.147	0.232±0.131
	FAST-DIPS	25.90±3.25	0.837±0.077	0.227±0.071	24.58±4.04	0.809±0.113	0.237±0.116
	FANO	**29.88±3.42**	**0.878±0.082**	**0.157±0.076**	**27.32±3.68**	**0.832±0.116**	**0.196±0.113**
NL deblur	DAPS	28.53±1.53	0.763±0.032	0.259±0.034	27.25±3.13	0.720±0.080	0.284±0.064
	SITCOM	29.31±2.17	0.801±0.085	0.240±0.073	28.47±3.96	0.802±0.094	**0.204±0.064**
	FAST-DIPS	27.55±1.63	0.803±0.044	0.282±0.048	25.50±3.59	0.696±0.140	0.383±0.111
	INO	27.53±2.41	0.797±0.056	0.252±0.041	25.39±3.74	0.688±0.137	0.318±0.081
	FANO	**30.83±1.77**	**0.873±0.031**	**0.183±0.033**	**29.09±4.25**	**0.819±0.100**	0.222±0.072

comparisons provided in Appendix E.1. Figure 1 presents visual reconstruction comparisons on representative tasks, and qualitative results for additional tasks are provided in Appendix F. Appendix E.2 further explores the ability of generative priors to compensate for missing oversampling information in phase retrieval without oversampling (OS $= 1$).

4.3 Ablation Studies

Our ablation isolates the respective contributions of the polar parameterization and radial freedom (Table 2). All three variants share the $\boldsymbol{e}$-search framework (N, K, residual-based early stopping, and hyperparameters), differing only in parameterization and radial freedom:

Table 2: Ablation of the polar parameterization and radial freedom (FFHQ100, $\sigma_y = 0.05$). The three variants share the $\boldsymbol{e}$-search framework (N, K, and residual-based early stopping) and differ in two design choices: *Polar* (the polar decomposition $\boldsymbol{e} = r\boldsymbol{u}$) and *Radial freedom* (allowing r to vary freely; otherwise, $r \equiv 1$ is fixed). The best results are in **bold**.

	Component		HDR			NL deblur		
Variant	Polar	Radial freedom	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
FANO	✓	✓	**29.88**	**0.878**	**0.157**	**30.83**	**0.873**	0.183
FANO-Sphere	✓	×	29.06	0.858	0.167	30.72	0.870	0.183
FANO-Euc	×	✓	29.18	0.859	0.171	30.80	0.871	**0.182**

- **FANO** (default): uses the polar decomposition $\boldsymbol{e} = r\boldsymbol{u}$ and jointly optimizes $(r, \boldsymbol{u})$;
- **FANO-Sphere**: uses the polar decomposition, fixes $r \equiv 1$, and optimizes only the direction $\boldsymbol{u}$;
- **FANO-Euc**: directly optimizes $\boldsymbol{e}$ with standard Adam, retaining the radial potential $\lambda_{\text{rad}}\Phi_d(\|\boldsymbol{e}\|_2/\sqrt{d-1})$ but without the polar decomposition.

All three variants outperform every comparison method in Table 1 in PSNR, SSIM, and LPIPS on HDR and nonlinear deblurring, indicating that the $\boldsymbol{e}$-search framework itself provides the primary gains. In particular, FANO-Sphere fixes $r \equiv 1$ and requires no tuning of λ_{rad}, yet its PSNR still exceeds the strongest comparison method by 0.93 dB and 1.41 dB on the two tasks, respectively, making it a ready-to-use, tuning-free default. The complete comparison across 8 tasks is given in Appendix Table 5, and sensitivity to λ_{rad} is analyzed in Appendix C.2.

Differences within the framework are task-dependent. On nonlinear deblurring, the PSNR gap among the three variants is at most 0.11 dB, with no notable separation in SSIM or LPIPS either. This indicates that the $\boldsymbol{e}$-search framework already realizes its advantages on this task, leaving limited additional benefit from the polar decomposition or radial freedom. HDR behaves differently: allowing radial deviations improves FANO over FANO-Sphere by 0.82 dB, while the polar decomposition improves FANO over FANO-Euc by 0.70 dB. Both gains also hold for SSIM and LPIPS, indicating independent contributions from the two mechanisms.

5 Conclusion

We analyze *radial drift* in annealed diffusion methods for inverse problems: within-layer optimization drives the equivalent noise norm of the denoiser input away from the typical value of the Gaussian prior, degrading reconstruction quality in later layers. To address this limitation, we propose FANO, which uses the forward noise $\boldsymbol{e}$ instead of the denoiser input $\boldsymbol{x}_t$ as the within-layer search variable in an annealing framework. The anchor is excluded from within-layer updates, the denoiser input retains the forward-diffusion structure, and the search region contracts progressively through annealing. The Gaussian prior on $\boldsymbol{e}$ further provides a statistical basis for radial control, with a concentration-inspired radial potential keeping the denoiser input in the high-density region of the forward distribution. Experiments on 8 linear and nonlinear tasks validate the effectiveness of FANO. Ablation studies identify the $\boldsymbol{e}$-search framework as the primary source of improvement, and the gains on nonlinear tasks are consistent with the observations from our radial drift analysis.

The current framework relies on deterministic optimization and restricts the search space to a neighborhood of a single optimization trajectory, preventing exploration of the multimodal structure of the inverse problem posterior. Incorporating the geometric structure of noise into a stochastic sampling framework, for example by constructing Langevin dynamics on the sphere, offers a possible path from single-point reconstruction to sampling multiple solutions and quantifying uncertainty.

Reproducibility statement

Algorithm 1 provides the complete pseudocode for FANO. All hyperparameters (annealing layers, inner steps, learning rates, and radial regularization coefficients) are listed in Appendix Table 4; task-specific degradation parameters are given in Appendix Table 3; and the Riemannian Adam optimizer

and positivity safeguards are detailed in Appendix B.2. All experiments use publicly available pre-trained diffusion models from Choi et al. (2021) and Dhariwal & Nichol (2021) without task-specific fine-tuning. Code to reproduce all experiments will be released upon acceptance.

Ethics statement

This work proposes a general-purpose optimization method for image inverse problems using pre-trained diffusion models. All experiments are conducted on publicly available face (FFHQ) and object (ImageNet) datasets. The method does not involve human subjects, collection of new data, or deployment in sensitive applications. As with other image reconstruction techniques, the underlying diffusion models were trained on large-scale image datasets whose curation may carry biases; we do not address dataset bias in this work.

AI use statement

In this work, we have *not* used generative AI tools for any of the following required-disclosure tasks: generating synthetic data, developing theoretical models or conceptual frameworks, formulating mathematical claims, assisting in the writing of proofs, proposing or refining hypotheses, designing research methodology or experiments, implementing methods, interpreting results, assisting with translation, cleaning or reformatting datasets, or supporting qualitative and thematic data analysis. All research ideas, mathematical derivations, algorithm design, and experimental conclusions are solely the work of the authors.

Additionally, we used generative AI tools (Claude via Cursor IDE) to edit and polish draft text for improved readability and English fluency. All AI-edited text was reviewed and revised by the authors to ensure accuracy. We take responsibility for the final content of this work, including all text, claims, and artifacts.

References

P.-A. Absil, Robert Mahony, and Rodolphe Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2008. ISBN 978-0-691-13298-3. URL https://sites.uclouvain.be/absil/amsbook/.

Ismail Alkhouri, Shijun Liang, Cheng-Han Huang, Jimmy Dai, Qing Qu, Saiprasad Ravishankar, and Rongrong Wang. SITCOM: Step-wise triple-consistent diffusion sampling for inverse problems. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 1128–1158. PMLR, 2025. URL https://proceedings.mlr.press/v267/alkhouri25b.html.

Brian D. O. Anderson. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3):313–326, 1982. doi: 10.1016/0304-4149(82)90051-5. URL https://doi.org/10.1016/0304-4149(82)90051-5.

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*, 2024. URL https://openreview.net/forum?id=4unJi0qrTE.

Jooyoung Choi, Sungwon Kim, Yonghyun Jeong, Youngjune Gwon, and Sungroh Yoon. ILVR: Conditioning method for denoising diffusion probabilistic models. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 14347–14356. IEEE, 2021. doi: 10.1109/ICCV48922.2021.01410. URL https://doi.org/10.1109/ICCV48922.2021.01410.

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*, volume 35, pp. 25683–25696. Curran Associates, Inc., 2022. doi: 10.52202/068431-1862. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/a48e5877c7bf86a513950ab23b360498-Abstract-Conference.html.

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*, 2023. URL https://openreview.net/forum?id=OnD9zGAGT0k.

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

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat GANs on image synthesis. In *Advances in Neural Information Processing Systems*, volume 34, pp. 8780–8794. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/hash/49ad23d1ec9fa4bd8d77d02681df5cfa-Abstract.html.

Sander Dieleman. Diffusion is spectral autoregression. Sander Dieleman's blog, September 2024. URL https://sander.ai/2024/09/02/spectral-autoregression.html.

James R. Fienup. Phase retrieval algorithms: A comparison. *Applied Optics*, 21(15):2758–2769, 1982. doi: 10.1364/AO.21.002758. URL https://doi.org/10.1364/AO.21.002758.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/hash/4c5bcfec8584af0d967f1ab10179ca4b-Abstract.html.

Jing Jia, Wei Yuan, Sifan Liu, Liyue Shen, and Guanyang Wang. Weak diffusion priors can still achieve strong inverse-problem performance. In *International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=fdkSA4F0lN.

Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(12):4217–4228, 2021. doi: 10.1109/TPAMI.2020.2970919. URL https://doi.org/10.1109/TPAMI.2020.2970919.

Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. In *Advances in Neural Information Processing Systems*, volume 35, pp. 23593–23606. Curran Associates, Inc., 2022. doi: 10.52202/068431-1714. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/95504595b6169131b6ed6cd72eb05616-Abstract-Conference.html.

Minwoo Kim, Seunghyeok Shin, and Hongki Lim. FAST-DIPS: Adjoint-free analytic steps and hard-constrained likelihood correction for diffusion-prior inverse problems. In *International Conference on Learning Representations*, volume 2026, pp. 104696–104730, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/ab378fb084f313a204432f0a1e697bae-Abstract-Conference.html.

Beatrice Laurent and Pascal Massart. Adaptive estimation of a quadratic functional by model selection. *The Annals of Statistics*, 28(5):1302–1338, 2000. doi: 10.1214/aos/1015957395. URL https://doi.org/10.1214/aos/1015957395.

Xiang Li, Soo Min Kwon, Shijun Liang, Ismail R. Alkhouri, Saiprasad Ravishankar, and Qing Qu. Decoupled data consistency with diffusion purification for image restoration. In *International Conference on Computational Photography*, 2026. URL https://arxiv.org/abs/2403.06054.

Jia-Wei Liao, Li-Xuan Peng, Mei-Heng Yueh, Min Sun, Cheng-Fu Chou, and Jun-Cheng Chen. DiffRGD: An inference-time diffusion guidance through Riemannian gradient descent. In *European Conference on Computer Vision*, 2026. URL https://media.eventhosts.cc/Conferences/ECCV2026/pdfs/9638.pdf.

Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. RePaint: Inpainting using denoising diffusion probabilistic models. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11451–11461. IEEE, 2022. doi: 10.1109/CVPR52688.2022.01117. URL https://doi.org/10.1109/CVPR52688.2022.01117.

Xiangming Meng and Yoshiyuki Kabashima. Diffusion model based posterior sampling for noisy linear inverse problems. In *Proceedings of the 16th Asian Conference on Machine Learning*, volume 260 of *Proceedings of Machine Learning Research*, pp. 623–638. PMLR, 2025. URL https://proceedings.mlr.press/v260/meng25a.html.

Severi Rissanen, Markus Heinonen, and Arno Solin. Generative modelling with inverse heat dissipation. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=4PJUBT9f2Ol.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2021a. URL https://openreview.net/forum?id=St1giarCHLP.

Jiaming Song, Arash Vahdat, Morteza Mardani, and Jan Kautz. Pseudoinverse-guided diffusion models for inverse problems. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=9_gsMA8MRKQ.

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*, 2021b. URL https://openreview.net/forum?id=PxTIG12RRHS.

Phong Tran, Anh Tuan Tran, Quynh Phung, and Minh Hoai. Explore image deblurring via encoded blur kernel space. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11951–11960. IEEE, 2021. doi: 10.1109/CVPR46437.2021.01178. URL https://doi.org/10.1109/CVPR46437.2021.01178.

Hengkang Wang, Xu Zhang, Taihui Li, Yuxiang Wan, Tiancong Chen, and Ju Sun. DMPlug: A plug-in method for solving inverse problems with diffusion models. In *Advances in Neural Information Processing Systems*, volume 37, pp. 117881–117916. Curran Associates, Inc., 2024. doi: 10.52202/079017-3744. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/d5c04aa72b92c53bda5b525b60958295-Abstract-Conference.html.

Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-shot image restoration using denoising diffusion null-space model. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=mRieQgMtNTQ.

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. URL https://doi.org/10.1109/TIP.2003.819861.

Yingzhi Xia, Setthakorn Tanomkiattikun, Liangli Zhen, and Zaiwang Gu. Noise-adaptive diffusion sampling for inverse problems without task-specific tuning. In *International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/1e45dc45f1649f880130169b2271129e-Abstract-Conference.html.

Lingxiao Yang, Shutong Ding, Yifan Cai, Jingyi Yu, Jingya Wang, and Ye Shi. Guidance with spherical Gaussian constraint for conditional diffusion. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 56071–56095. PMLR, 2024. URL https://proceedings.mlr.press/v235/yang24h.html.

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*, pp. 20895–20905, 2025. doi: 10.1109/CVPR52734.2025.01946. URL https://openaccess.thecvf.com/content/CVPR2025/html/Zhang_Improving_Diffusion_Inverse_Problem_Solving_with_Decoupled_Noise_Annealing_CVPR_2025_paper.html.

Jiawei Zhang, Ziyuan Liu, Leon Yan, Zhenyu Xiao, and Yuantao Gu. Stage-wise distortion–perception traversal in zero-shot inverse problems with diffusion models. In *International Conference on Machine Learning*, 2026. URL https://arxiv.org/abs/2605.28711.

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 586–595, 2018. URL https://openaccess.thecvf.com/content_cvpr_2018/html/Zhang_The_Unreasonable_Effectiveness_CVPR_2018_paper.html.

A Theoretical Derivations and Proofs

A.1 Derivation of the Radial Potential

A standard Gaussian vector $\boldsymbol{e} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_d)$ has density $p(\boldsymbol{e}) \propto \exp(-\|\boldsymbol{e}\|_2^2/2)$. We decompose $\boldsymbol{e}$ in polar coordinates as $\boldsymbol{e} = r \cdot \boldsymbol{u}$, where the normalized radial magnitude is $r := \|\boldsymbol{e}\|_2/\sqrt{d-1} > 0$ and the direction is $\boldsymbol{u} \in \mathbb{S}^{d-1}_{\sqrt{d-1}} := \{\boldsymbol{u} \in \mathbb{R}^d : \|\boldsymbol{u}\|_2 = \sqrt{d-1}\}$. Since $\|\boldsymbol{e}\|_2 = r\sqrt{d-1}$, the Jacobian of the polar change of variables (the spherical surface element in d dimensions multiplied by the radial factor $\|\boldsymbol{e}\|_2^{d-1}$) yields the joint density

$$p(r, \boldsymbol{u}) \propto (r\sqrt{d-1})^{d-1} \exp\left(-\frac{(d-1)\, r^2}{2}\right), \tag{A.1}$$

where the direction $\boldsymbol{u}$ is uniformly distributed on $\mathbb{S}^{d-1}_{\sqrt{d-1}}$. Integrating over $\boldsymbol{u}$ gives the radial marginal density

$$p(r) \propto r^{d-1} \exp\left(-\frac{(d-1)\, r^2}{2}\right), \quad r > 0. \tag{A.2}$$

Taking the negative log and subtracting a constant so that the minimum is zero gives the radial potential

$$\Phi_d(r) := -\log p(r) + \text{const} = \frac{d-1}{2}\big(r^2 - 1 - 2\log r\big), \tag{A.3}$$

which attains its minimum value of zero at $r = 1$ (i.e., $\|\boldsymbol{e}\|_2 = \sqrt{d-1}$). Its curvature at $r = 1$ is $\Phi_d''(1) = 2(d-1)$, reflecting the strength of high-dimensional concentration: larger d imposes a greater cost for deviations from the typical value.

A.2 Proof of Proposition 1

Following the notation in the main text, fix the annealing layer t, anchor $\bar{\boldsymbol{x}}$, and direction $\boldsymbol{u} \in \mathbb{S}^{d-1}_{\sqrt{d-1}}$, and write $F(r) := F_t(r; \boldsymbol{u})$, $J(r) := J_\lambda(r; \boldsymbol{u}) = F(r) + \lambda_{\text{rad}}\Phi_d(r)$, and $r_\lambda := r_\lambda(\boldsymbol{u})$.

Proof of Proposition 1. **(i): Existence and bounded deviation.** Since $\log r \le r - 1$, we have $r^2 - 1 - 2\log r = (r-1)^2 + 2(r - 1 - \log r) \ge (r-1)^2$, and therefore

$$\Phi_d(r) \ge \frac{d-1}{2}(r-1)^2. \tag{A.4}$$

Moreover, $\Phi_d(r) \to +\infty$ as $r \to 0^+$ or $r \to +\infty$. Together with $F(r) \ge 0$, this implies $J(r) \to +\infty$. Every sublevel set at a finite level is closed and bounded, so continuity ensures that J admits a global minimizer $r_\lambda \in (0, \infty)$. By the global optimality of r_λ and $\Phi_d(1) = 0$,

$$\lambda_{\text{rad}}\Phi_d(r_\lambda) \le J(r_\lambda) \le J(1) = F(1).$$

Combining this with Equation (A.4) gives $\frac{\lambda_{\text{rad}}(d-1)}{2}(r_\lambda - 1)^2 \le F(1)$, or equivalently $|r_\lambda - 1| \le \sqrt{2F(1)/[\lambda_{\text{rad}}(d-1)]}$. Define $\boldsymbol{z}_t(r, \boldsymbol{u}) := a_t\bar{\boldsymbol{x}} + \sigma_t r\boldsymbol{u}$; then $\|\boldsymbol{z}_t(r, \boldsymbol{u}) - \boldsymbol{z}_t(1, \boldsymbol{u})\|_2 = \sigma_t|r - 1|\sqrt{d-1}$. Substituting the radius bound above yields the following bound on the denoiser input shift:

$$\|\boldsymbol{z}_t(r_\lambda, \boldsymbol{u}) - \boldsymbol{z}_t(1, \boldsymbol{u})\|_2 \le \sigma_t\sqrt{\frac{2F(1)}{\lambda_{\text{rad}}}}\,. \tag{A.5}$$

(ii): Strict relaxation gain. Since $\Phi_d'(1) = 0$, we have $J'(1) = F'(1) =: s_t(\boldsymbol{u})$. If $s_t(\boldsymbol{u}) \ne 0$, then $r{=}1$ is not a stationary point of J. Set $\delta = -\varepsilon\,\text{sign}(s_t(\boldsymbol{u}))$ for sufficiently small $\varepsilon > 0$. Then

$$J(1+\delta) = J(1) + s_t(\boldsymbol{u})\delta + o(|\delta|) = J(1) - \varepsilon|s_t(\boldsymbol{u})| + o(\varepsilon) < J(1).$$

By the global optimality of r_λ, $J(r_\lambda) \le J(1+\delta) < J(1) = F(1)$. □

A.3 Strong-Regularization Asymptotic Expansion

Corollary 1 (Strong-regularization asymptotics)**.** *Under the assumptions of Proposition 1, further assume that F is twice continuously differentiable in a neighborhood of $r{=}1$, and let $c_\lambda := \lambda_{\text{rad}}(d-1)$ and $s_t(\boldsymbol{u}) := F'(1)$. Fix d, t, $\bar{\boldsymbol{x}}$, and $\boldsymbol{u}$, and let $\lambda_{\text{rad}} \to \infty$ (equivalently, $c_\lambda \to \infty$). Then*

$$r_\lambda(\boldsymbol{u}) = 1 - \frac{s_t(\boldsymbol{u})}{2c_\lambda} + O(c_\lambda^{-2}), \tag{A.6}$$

$$J_\lambda(r_\lambda(\boldsymbol{u}); \boldsymbol{u}) = F_t(1; \boldsymbol{u}) - \frac{s_t(\boldsymbol{u})^2}{4c_\lambda} + O(c_\lambda^{-2}). \tag{A.7}$$

Proof. By part (i), $r_\lambda \to 1$ as $c_\lambda \to \infty$. Let $\delta_\lambda := r_\lambda - 1$. The first-order condition at an interior minimizer is

$$0 = F'(r_\lambda) + c_\lambda \left[(1+\delta_\lambda) - \frac{1}{1+\delta_\lambda} \right].$$

Part (i) gives $\delta_\lambda = O(c_\lambda^{-1/2})$, but a tighter order is needed below. Since $r_\lambda \to 1$ and F' is continuous near 1, there exists a c_λ-independent constant $M > 0$ such that $|F'(r_\lambda)| \le M$. Using $r_\lambda - 1/r_\lambda = \delta_\lambda(r_\lambda + 1)/r_\lambda$, the first-order condition gives $|\delta_\lambda| = |F'(r_\lambda)| \cdot r_\lambda/[c_\lambda(r_\lambda+1)] \le M/c_\lambda$, hence $\delta_\lambda = O(c_\lambda^{-1})$.

Expanding the bracketed term as $2\delta_\lambda + O(\delta_\lambda^2)$ and using $F'(1+\delta_\lambda) = s_t(\boldsymbol{u}) + O(\delta_\lambda)$ gives

$$0 = s_t(\boldsymbol{u}) + 2c_\lambda\delta_\lambda + O(c_\lambda^{-1}),$$

which yields Equation (A.6). A Taylor expansion of the objective gives

$$F(1+\delta_\lambda) = F(1) + s_t(\boldsymbol{u})\delta_\lambda + O(\delta_\lambda^2),$$
$$\lambda_{\text{rad}}\Phi_d(1+\delta_\lambda) = c_\lambda\delta_\lambda^2 + O(c_\lambda\delta_\lambda^3).$$

Substituting $\delta_\lambda = -s_t(\boldsymbol{u})/(2c_\lambda) + O(c_\lambda^{-2})$, the measurement loss contributes $-s_t(\boldsymbol{u})^2/(2c_\lambda)$ and the potential contributes $s_t(\boldsymbol{u})^2/(4c_\lambda)$. Combining the two yields Equation (A.7). □

Equations (A.6)–(A.7) show that, under strong regularization, the optimal radius has an $O(c_\lambda^{-1})$ deviation from 1; when $s_t(\boldsymbol{u}) \neq 0$, this corresponds to an objective decrease of order $s_t(\boldsymbol{u})^2/(4c_\lambda)$. Radial regularization thus limits the magnitude of radial deviations without entirely eliminating measurement-gradient-driven radial adjustment.

B Implementation Details and Experimental Setup

B.1 Algorithm Pseudocode

Algorithm 1. FANO: Anchored Forward-Noise Optimization

Input Measurement $\boldsymbol{y}$, forward operator A, denoiser D_t.
Schedule $\{t_k, a_{t_k}, \sigma_{t_k}\}_{k=1}^{N}$; at most K inner steps per layer.
Learning rates η_u, η_r; radial penalty λ_{rad}; threshold τ_{stop}.
Output Restored image $\hat{\boldsymbol{x}}^N$.

Sample $\boldsymbol{\epsilon}_0 \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_d)$; set $\hat{\boldsymbol{x}}^0 \leftarrow D_{t_1}(\sigma_{t_1}\boldsymbol{\epsilon}_0)$
Outer loop: annealing
for $k = 1, \dots, N$ **do**
  $\bar{\boldsymbol{x}} \leftarrow \hat{\boldsymbol{x}}^{k-1}$; sample $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_d)$
  $\boldsymbol{u} \leftarrow \sqrt{d-1}\,\boldsymbol{\epsilon}/\|\boldsymbol{\epsilon}\|_2$; $r \leftarrow 1$
  Inner loop: polar optimization
  for $j = 1, \dots, K$ **do**
    $\boldsymbol{z} \leftarrow a_{t_k}\bar{\boldsymbol{x}} + \sigma_{t_k} r \boldsymbol{u}$; $\boldsymbol{x} \leftarrow D_{t_k}(\boldsymbol{z})$
    $\rho \leftarrow \|A(\boldsymbol{x}) - \boldsymbol{y}\|_2$
    if $\rho \le \tau_{\text{stop}}$ **then**
      break
    end if
    $\boldsymbol{g}_z \leftarrow \nabla_{\boldsymbol{z}} \|A(D_{t_k}(\boldsymbol{z})) - \boldsymbol{y}\|_2^2$
    $\tilde{\boldsymbol{g}}_u \leftarrow P_{T_{\boldsymbol{u}}\mathbb{S}}(\sigma_{t_k} r \boldsymbol{g}_z)$
    $g_r \leftarrow \sigma_{t_k}\langle \boldsymbol{g}_z, \boldsymbol{u}\rangle + \lambda_{\text{rad}}\Phi_d'(r)$
    $\boldsymbol{u} \leftarrow \text{RiemannianAdam}(\boldsymbol{u}, \tilde{\boldsymbol{g}}_u, \eta_u)$; $r \leftarrow \text{Adam}(r, g_r, \eta_r)$
  end for
  Update the anchor
  $\hat{\boldsymbol{x}}^k \leftarrow D_{t_k}(a_{t_k}\bar{\boldsymbol{x}} + \sigma_{t_k} r \boldsymbol{u})$
end for
return $\hat{\boldsymbol{x}}^N$

B.2 Polar Gradients and Optimizers

Separation of polar gradients The within-layer objective in Equation (5) is

$$J_\lambda(r; \boldsymbol{u}) = \ell_y(D_t(a_t\bar{\boldsymbol{x}} + \sigma_t r \boldsymbol{u})) + \lambda_{\text{rad}}\,\Phi_d(r).$$

Since $\Phi_d(r)$ is independent of $\boldsymbol{u}$, the gradient with respect to $\boldsymbol{u}$ comes entirely from the measurement loss:

$$\nabla_{\boldsymbol{u}} J_\lambda(r; \boldsymbol{u}) = \sigma_t r\, \nabla_{\boldsymbol{z}} \ell_y(D_t(\boldsymbol{z}))\Big|_{\boldsymbol{z}=a_t\bar{\boldsymbol{x}}+\sigma_t r\boldsymbol{u}}. \tag{B.1}$$

The gradient with respect to r contains contributions from both the measurement loss and the radial potential:

$$\frac{\partial J_\lambda(r; \boldsymbol{u})}{\partial r} = \sigma_t \left\langle \nabla_{\boldsymbol{z}} \ell_y(D_t(\boldsymbol{z})), \boldsymbol{u}\right\rangle\Big|_{\boldsymbol{z}=a_t\bar{\boldsymbol{x}}+\sigma_t r\boldsymbol{u}} + \lambda_{\text{rad}}\,\Phi_d'(r), \tag{B.2}$$

where $\Phi_d'(r) = (d-1)(r - 1/r)$. Because the angular Riemannian gradient uses the tangent-space projection (B.3) to further remove the radial component along $\boldsymbol{u}$, the updates to the two variables are geometrically separated: the angular update moves $\boldsymbol{u}$ along a tangent direction of the sphere, while the radial update adjusts the scalar r. The radial regularization term $\lambda_{\text{rad}}\Phi_d'(r)$ appears only in the latter.

Riemannian Adam for the direction variable Fix the outer layer k, let $t = t_k$, and write $\mathbb{S} := \mathbb{S}^{d-1}_{\sqrt{d-1}}$. The angular variable $\boldsymbol{u}$ is optimized on $\mathbb{S}$. Denote the variables at step $j \ge 1$ by $(r_j, \boldsymbol{u}_j)$. The denoiser input and its input-space gradient are, respectively,

$$\boldsymbol{z}_j := a_t\bar{\boldsymbol{x}} + \sigma_t r_j \boldsymbol{u}_j, \qquad \boldsymbol{g}_{z,j} := \nabla_{\boldsymbol{z}} \ell_y(D_t(\boldsymbol{z}))\big|_{\boldsymbol{z}=\boldsymbol{z}_j}.$$

The corresponding Euclidean angular gradient is $\boldsymbol{g}_{u,j} := \sigma_t r_j \boldsymbol{g}_{z,j} = \nabla_{\boldsymbol{u}} J_\lambda(r_j; \boldsymbol{u})\big|_{\boldsymbol{u}=\boldsymbol{u}_j}$. The update rules are as follows:

(i) *Tangent-space projection.* Project the Euclidean gradient onto the tangent space at $\boldsymbol{u}_j$:

$$\tilde{\boldsymbol{g}}_{u,j} = \boldsymbol{g}_{u,j} - \frac{\langle \boldsymbol{g}_{u,j}, \boldsymbol{u}_j \rangle}{\|\boldsymbol{u}_j\|_2^2} \boldsymbol{u}_j. \tag{B.3}$$

(ii) *Adam moment updates.* Maintain first- and second-moment estimates in the tangent space:

$$\begin{aligned} \boldsymbol{m}_j &= \beta_1 \boldsymbol{m}_{j-1} + (1-\beta_1)\tilde{\boldsymbol{g}}_{u,j}, \\ \boldsymbol{v}_j &= \beta_2 \boldsymbol{v}_{j-1} + (1-\beta_2)\tilde{\boldsymbol{g}}_{u,j} \odot \tilde{\boldsymbol{g}}_{u,j}. \end{aligned} \tag{B.4}$$

Apply bias correction $\hat{\boldsymbol{m}}_j = \boldsymbol{m}_j/(1-\beta_1^j)$ and $\hat{\boldsymbol{v}}_j = \boldsymbol{v}_j/(1-\beta_2^j)$. After computing the preconditioned update direction, project it onto the tangent space once more:

$$\boldsymbol{\Delta}_j = P_{T_{\boldsymbol{u}_j}\mathbb{S}}\left(\frac{\hat{\boldsymbol{m}}_j}{\sqrt{\hat{\boldsymbol{v}}_j} + \epsilon}\right). \tag{B.5}$$

(iii) *Retraction.* Retract the tangent update onto the sphere:

$$\boldsymbol{u}_{j+1} = \sqrt{d-1}\,\frac{\boldsymbol{u}_j - \eta_u \boldsymbol{\Delta}_j}{\|\boldsymbol{u}_j - \eta_u \boldsymbol{\Delta}_j\|_2}. \tag{B.6}$$

(iv) *Momentum transport.* After the update, reproject the first moment onto the tangent space at the new point:

$$\boldsymbol{m}_j \leftarrow P_{T_{\boldsymbol{u}_{j+1}}\mathbb{S}}(\boldsymbol{m}_j).$$

This keeps the momentum on the manifold. The default hyperparameters are $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$.

Adam for the radius variable and positivity safeguards The radial scalar $r > 0$ is updated with standard Adam. If the updated value satisfies $r' \le r_{\min}$ (with the default $r_{\min} = 10^{-6}$), a backtracking search repeatedly halves the step size until $r' > r_{\min}$ or the maximum number of backtracking steps is reached. This mechanism keeps r positive throughout optimization, ensuring that the potential $\Phi_d(r)$ has a well-defined $\log r$ term. Backtracking was never triggered on any task in our experiments, indicating that r does not approach zero along normal optimization trajectories.

B.3 Statistical Motivation for Early Stopping

For a fixed outer layer k, within-layer optimization uses the measurement residual for early stopping, terminating when $\|A(D_{t_k}(\boldsymbol{z}_j)) - \boldsymbol{y}\|_2 \le \tau_{\text{stop}}$. The threshold is set to

$$\tau_{\text{stop}} := \sigma_y \sqrt{m}, \tag{B.7}$$

where σ_y is the standard deviation of the measurement noise and m is the measurement dimension. This choice is motivated by the concentration of additive Gaussian noise: if $\boldsymbol{\xi} \sim \mathcal{N}(\boldsymbol{0}, \sigma_y^2 \boldsymbol{I}_m)$, then $\mathbb{E}[\|\boldsymbol{\xi}\|_2^2] = m\sigma_y^2$, and concentration inequalities for the χ^2 distribution imply $\|\boldsymbol{\xi}\|_2/(\sigma_y\sqrt{m}) \approx 1$, with deviations of order $O(1/\sqrt{m})$. Thus, $\sigma_y\sqrt{m}$ is the typical norm of the measurement noise: a residual below this level indicates that the optimizer is beginning to fit the noise itself, so further optimization harms generalization. In the noiseless case $\sigma_y = 0$, we have $\tau_{\text{stop}} = 0$, and the criterion reduces to a fixed iteration budget.

B.4 Task-Specific Degradation Parameters

Table 3 lists the degradation-operator parameters for each task.

Table 3: Degradation-operator parameters for each task.

Task	Type	Parameters
SR $\times 4$	Linear	Bicubic, factor 4
Gauss. deblur	Linear	61×61 Gaussian, σ_{blur}=3.0
Motion deblur	Linear	61×61 motion, intensity 0.5
Inpaint (box)	Linear	128×128 random mask
Inpaint (rand.)	Linear	Pixel mask prob. 0.7
Phase retrieval	Nonlinear	Oversampling ratio 1.5
HDR	Nonlinear	Clipping factor 2
NL deblur	Nonlinear	BKSE varying kernel

Table 4: Annealing configurations and radial regularization coefficients used in the main FANO experiments on FFHQ and ImageNet ($\sigma_y = 0.05$).

Task	Annealing layers N	Inner steps K	λ_{rad}
SR $\times 4$	50	30	0
Gauss. deblur	50	30	0.0015
Motion deblur	50	30	0.01
Inpaint (box)	20	20	0
Inpaint (rand.)	20	30	1
Phase retrieval	50	30	0.01
HDR	50	30	0.0015
NL deblur	50	30	0.0015

B.5 Task-Specific Hyperparameters for FANO

Table 4 lists the number of annealing layers N, the number of inner steps K, and the radial regularization coefficient λ_{rad} used for each task in the main FANO experiments. The same configurations are used for FFHQ and ImageNet. The direction is optimized with Riemannian Adam and the radius with Adam, both using learning rates $\eta_u = \eta_r = 0.02$, $\beta_1 = 0.9$, and $\beta_2 = 0.999$. The annealing schedule $\{t_k\}_{k=1}^{N}$ is sampled at uniform intervals from a 1000-step VP diffusion process. The forward coefficients at each layer are $a_{t_k} = \sqrt{\bar{\alpha}_{t_k}}$ and $\sigma_{t_k} = \sqrt{1-\bar{\alpha}_{t_k}}$, where $\bar{\alpha}_t$ denotes the cumulative signal-to-noise ratio schedule of VP diffusion. All tasks use the same scheduling procedure.

C Ablation and Parameter Sensitivity Analysis

C.1 A Tuning-Free Baseline: FANO-Sphere

When task-specific tuning of λ_{rad} is undesirable, FANO-Sphere with $r \equiv 1$ provides a tuning-free default. Table 5 compares its PSNR with the two most closely related annealed baselines in Table 1, SITCOM and DAPS. Among the 16 dataset–task combinations, FANO-Sphere outperforms both SITCOM and DAPS in 15; the sole exception is ImageNet box inpainting, where DAPS leads by 0.88 dB. Its advantage over SITCOM is particularly pronounced on nonlinear tasks: 2.85 dB on FFHQ and 3.37 dB on ImageNet for phase retrieval, and 0.93–1.05 dB for HDR. The default FANO configuration in the main text, which jointly optimizes $r, \boldsymbol{u}$, provides a further improvement of approximately 0.8 dB on HDR.

C.2 Sensitivity to the Hyperparameter λ_{rad}

We assess the sensitivity of reconstruction quality to the radial regularization coefficient λ_{rad} through a parameter sweep. For box inpainting, random inpainting, and HDR, we use the first 30 FFHQ images to sweep $\lambda_{\text{rad}} \in \{0, 0.0005, 0.0015, 0.01, 0.1, 1, 2, 5, 10\}$. Each task uses the annealing budget in Table 4, with the remaining hyperparameters fixed at η_u=η_r=0.02, σ_y=0.05, and seed=42. Within each task, all coefficients share the same images, measurements, and initializations (Figure C.1).

Table 5: PSNR comparison of FANO-Sphere ($r \equiv 1$, without tuning λ_{rad}) against annealed baselines ($\sigma_y = 0.05$). The best result for each combination is in **bold**.

	FFHQ			ImageNet		
Task	SITCOM	DAPS	**Sphere**	SITCOM	DAPS	**Sphere**
SR $\times 4$	30.55	29.35	**30.57**	27.09	25.47	**27.14**
Gauss. deblur	29.89	29.61	**30.66**	26.38	25.88	**27.01**
Motion deblur	32.37	31.65	**33.03**	29.38	28.65	**29.89**
Inpaint (box)	24.47	24.32	**25.14**	19.62	**21.05**	20.17
Inpaint (rand.)	32.40	30.72	**34.06**	28.90	27.35	**29.90**
Phase retrieval	29.21	30.09	**32.06**	20.49	22.94	**23.86**
HDR	28.13	27.48	**29.06**	25.75	26.07	**26.80**
NL deblur	29.31	28.53	**30.72**	28.47	27.25	**29.16**

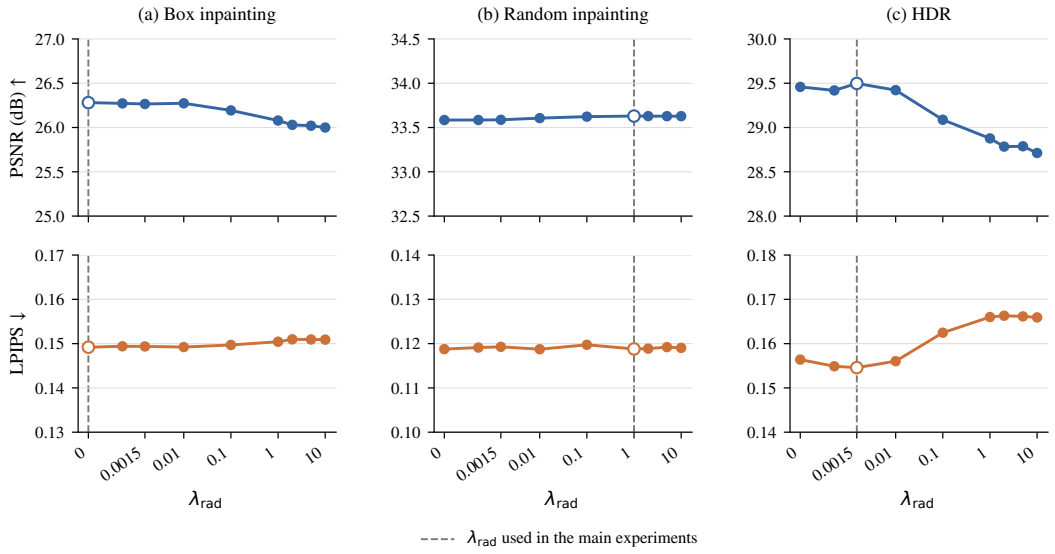


Figure C.1: Sensitivity to λ_{rad} on three representative tasks (the first 30 FFHQ images, σ_y=0.05). The top row shows PSNR and the bottom row shows LPIPS (VGG16), both averaged over images. Dashed lines and open circles mark the default value used for each task in the main text.

The curves for box inpainting and random inpainting are relatively flat. Across the full sweep, box inpainting varies by 0.28 dB in PSNR and 0.0018 in LPIPS, with the default λ_{rad}=0 achieving both the highest PSNR and the lowest LPIPS. Random inpainting varies by at most 0.05 dB in PSNR and 0.0010 in LPIPS; the default λ_{rad}=1 achieves the highest PSNR, and its LPIPS differs from the lowest value in the sweep by less than 0.0001.

HDR is more sensitive to radial regularization: the default λ_{rad}=0.0015 achieves both the highest PSNR and the lowest LPIPS on the sweep grid. Increasing the coefficient to 0.1 or above reduces PSNR by at least 0.41 dB and increases LPIPS by at least 0.0079, with stronger regularization further limiting reconstruction quality. This is consistent with the role of radial freedom in nonlinear tasks (see Section 3.1): when the measurement operator is strongly nonlinear, the optimizer needs moderate radial flexibility to match the measurement, and excessive radial constraints limit this ability. These results use the FFHQ30 subset and a single random seed to illustrate sensitivity trends; the configurations used for the full-dataset experiments are listed in Table 4.

D Computational Cost and Efficiency Analysis

Do FANO's reconstruction gains arise from the method itself rather than greater computational cost? We address this question through three comparisons: (1) a direct comparison with the most closely

Table 6: Matched-budget comparison of SITCOM-TB and FANO at $\sigma_y = 0.05$. For each task, both methods use the same number of annealing layers N, inner steps K, residual-based early-stopping criterion, and number of candidate runs. The best results are in **bold**.

		FFHQ			ImageNet		
Task	**Method**	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
SR $\times 4$	SITCOM-TB	30.61	0.869	**0.226**	**27.11**	**0.759**	**0.311**
	FANO	**30.68**	**0.872**	0.232	27.03	0.755	0.321
Gauss. deblur	SITCOM-TB	30.09	0.854	0.244	26.56	0.732	0.337
	FANO	**30.65**	**0.868**	**0.225**	**27.01**	**0.746**	**0.317**
Motion deblur	SITCOM-TB	32.62	0.902	0.187	29.68	0.838	0.231
	FANO	**33.04**	**0.909**	**0.167**	**29.88**	**0.840**	**0.225**
Inpaint (box)	SITCOM-TB	24.47	0.737	0.245	19.62	0.699	0.256
	FANO	**25.33**	**0.866**	**0.156**	**20.40**	**0.786**	**0.206**
Inpaint (rand.)	SITCOM-TB	32.38	0.863	0.182	28.89	0.814	0.181
	FANO	**34.07**	**0.931**	**0.120**	**29.94**	**0.864**	**0.157**
Phase retrieval	SITCOM-TB	30.47	0.868	0.232	23.36	**0.577**	0.402
	FANO	**31.98**	**0.877**	**0.187**	**23.83**	0.567	**0.388**
HDR	SITCOM-TB	28.03	0.809	0.216	25.81	0.765	0.237
	FANO	**29.88**	**0.878**	**0.157**	**27.32**	**0.832**	**0.196**
NL deblur	SITCOM-TB	29.68	0.815	0.237	29.00	**0.820**	**0.197**
	FANO	**30.83**	**0.873**	**0.183**	**29.09**	0.819	0.222

related method, SITCOM, under matched NFE budgets to isolate the method's contribution; (2) a comparison of NFE and GPU memory efficiency with INO, which also performs spherical noise optimization; and (3) a comparison of actual runtime with DAPS.

D.1 Matched-Budget Comparison with SITCOM

SITCOM and FANO share the same algorithmic structure of annealing layers with within-layer optimization, making SITCOM the most directly comparable baseline in the main text. Table 1 uses the $N = 20$ configuration recommended in the SITCOM paper, whereas FANO defaults to $N = 50, K = 30$, resulting in different NFE budgets. To remove this confounding factor, Table 6 adjusts both the number of SITCOM annealing layers and its inner steps to match FANO (denoted SITCOM-TB), giving the two methods the same maximum NFE budget. Both methods use the residual-based early-stopping criterion in Equation (8).

Under the same annealing budget, FANO outperforms SITCOM-TB in FFHQ PSNR on all eight tasks, with particularly pronounced gains on inpainting and nonlinear tasks: 0.86 dB for box inpainting and 1.51 dB for phase retrieval on FFHQ, consistent with the analysis in Section 3.1. On ImageNet, its PSNR for SR $\times 4$ is lower by 0.08 dB and its LPIPS for nonlinear deblurring is higher by 0.025, with trade-offs on individual metrics more apparent for linear degradation tasks. Overall, FANO's gains on nonlinear tasks arise from using $\boldsymbol{e}$ as the search variable rather than from an additional NFE budget.

D.2 Computational Cost Comparison with INO

Both FANO and INO (Jia et al., 2026) optimize noise variables on a sphere, but use different inference frameworks: FANO progressively refines the estimate across annealing layers, requiring gradients through only a single denoiser call per layer, whereas INO treats the diffusion model as a black box and optimizes the initial noise of a three-step DDIM chain end to end. FANO uses a total NFE of $N \times K = 50 \times 30 = 1500$, whereas each of INO's 1000 AdamSphere steps requires 3 DDIM forward steps, for a total NFE of 3000, twice that of FANO. In terms of GPU memory, INO must backpropagate through the full DDIM chain and reaches approximately 5.34 GiB of peak allocated GPU memory, compared with only 2.30 GiB for FANO, less than half as much. Despite its sub-

Table 7: Average core reconstruction time (seconds) of FANO and DAPS on FFHQ, executed sequentially on a single A100.

Task	DAPS	FANO
SR $\times 4$	83	74
Gauss. deblur	68	88
Motion deblur	68	88
Inpaint (box)	46	86
Inpaint (rand.)	46	34
Phase retrieval	107	87
HDR	57	90
NL deblur	606	101

stantially lower NFE and memory usage, across the 30 metrics reported by both methods, FANO achieves better mean performance on 29 (Table 1). Under these experimental conditions, spherical search within a layerwise annealing framework outperforms end-to-end black-box optimization in both reconstruction quality and computational cost.

D.3 Runtime Comparison with DAPS

Table 7 reports the average core reconstruction time of FANO and DAPS (Zhang et al., 2025) on FFHQ, with all configurations executed sequentially on the same NVIDIA A100-PCIE-40GB GPU. DAPS uses its recommended 1k configuration, while FANO uses the task-specific annealing budgets, with $N = 50, K = 30$ for most tasks. For nonlinear deblurring, DAPS uses 200 Langevin updates per layer, compared with 100 for the other tasks. The nonlinear blur operator also requires calls to the BKSE network, making this task substantially more time-consuming than linear tasks.

D.4 Total Budget and Budget Allocation

This section presents two sets of budget experiments. Figure D.1 sweeps FANO's annealing budgets on the FFHQ10 subset, and Table 8 compares two annealing configurations on the full datasets.

Figure D.1 shows FANO's performance on one linear task (motion deblurring) and one nonlinear task (nonlinear deblurring) for different numbers of annealing layers N and inner steps K, as a function of total NFE ($= N \times K$).

Monotonic improvement for fixed N For each fixed N, increasing K monotonically improves both PSNR and LPIPS, with diminishing marginal gains as NFE increases.

Convergence of curves for different N The three curves corresponding to different N tend to converge in the high-NFE regime (at NFE$\approx$2000, LPIPS differences are $<$0.005). This indicates that FANO's final reconstruction quality depends primarily on the total computational budget and is relatively insensitive to its allocation between N and K. Notably, curves for larger N retain greater curvature at high NFE, suggesting that further increasing K may allow larger-N configurations to match or exceed the asymptotic performance of smaller-N configurations.

Trade-off between annealing layers and inner steps at a fixed budget Table 8 uses the same maximum NFE budget of 1500 to compare $N = 50, K = 30$ (more layers with fewer steps per layer) and $N = 20, K = 75$ (fewer layers with more steps per layer). Linear tasks are insensitive to this allocation, with PSNR differences of at most 0.14 dB between the two configurations. The three nonlinear tasks respond differently: $N = 50$ improves PSNR by 0.82–1.06 dB for phase retrieval and 0.11–0.27 dB for HDR, whereas $N = 20$ improves PSNR by 0.18–0.21 dB for nonlinear deblurring and consistently achieves better LPIPS. The gains on phase retrieval substantially outweigh the losses on nonlinear deblurring, so Table 1 adopts $N = 50, K = 30$ as the default configuration. Table 8 omits box inpainting and random inpainting; the budgets used for these two tasks in the main experiments are listed in Table 4.

Table 8: Results for two FANO configurations on FFHQ100 and ImageNet100 ($\sigma_y = 0.05$): $N = 50, K = 30$ and $N = 20, K = 75$, both with a maximum NFE of 1500. The best metric for each task and dataset is in **bold**.

		FFHQ			ImageNet		
Task	**Config**	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
SR ×4	$N=50, K=30$	**30.68**	**0.872**	0.232	27.03	0.755	0.321
	$N=20, K=75$	30.64	0.869	**0.226**	**27.10**	**0.756**	**0.315**
Gauss. deblur	$N=50, K=30$	**30.65**	**0.868**	0.225	27.01	0.746	0.317
	$N=20, K=75$	30.55	0.866	**0.217**	**27.08**	**0.750**	**0.306**
Motion deblur	$N=50, K=30$	**33.04**	**0.909**	0.167	29.88	0.840	0.225
	$N=20, K=75$	33.01	0.908	**0.161**	**30.02**	**0.845**	**0.208**
Phase retrieval	$N=50, K=30$	**31.98**	**0.877**	**0.187**	**23.83**	**0.567**	**0.388**
	$N=20, K=75$	31.16	0.852	0.199	22.77	0.526	0.410
HDR	$N=50, K=30$	**29.88**	**0.878**	**0.157**	**27.32**	**0.832**	0.196
	$N=20, K=75$	29.61	0.868	0.162	27.21	0.827	**0.182**
NL deblur	$N=50, K=30$	30.83	0.873	0.183	29.09	0.819	0.222
	$N=20, K=75$	**31.01**	**0.874**	**0.173**	**29.30**	**0.828**	**0.186**

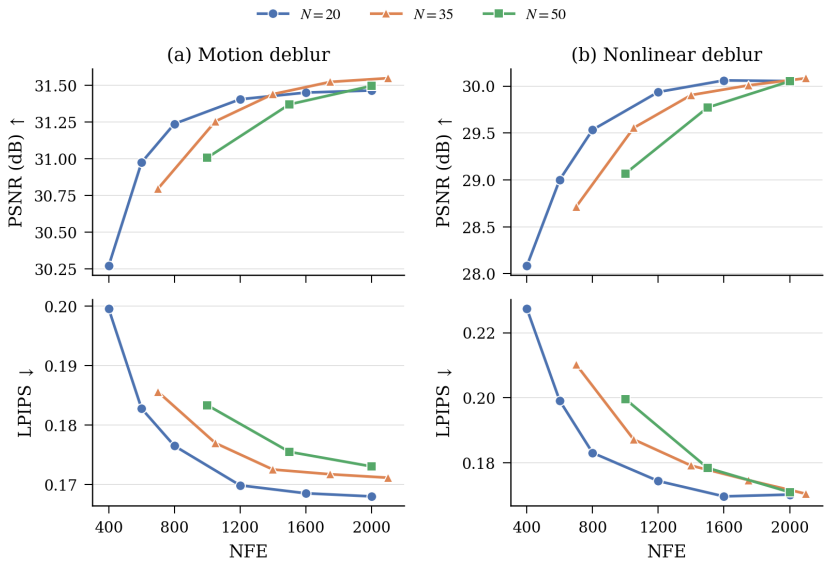


Figure D.1: FANO reconstruction performance under different NFE budgets. The left column shows motion deblurring and the right column nonlinear deblurring; the top row shows PSNR (↑) and the bottom row LPIPS (↓). Each curve corresponds to a number of annealing layers N, with total NFE ($= N \times K$) on the horizontal axis.

E Extended Experiments

E.1 Noise Robustness

Figure E.1 compares FANO and DAPS on motion deblurring and nonlinear deblurring as the measurement noise σ_y increases from 0.01 to 0.30, using 20 FFHQ images. Both metrics for FANO degrade smoothly and monotonically: PSNR decreases from 34.21 dB to 27.89 dB for motion deblurring and from 31.34 dB to 26.27 dB for nonlinear deblurring, without abrupt changes or collapse. By contrast, DAPS performs relatively close to FANO for $\sigma_y \leq 0.10$ but degrades sharply for

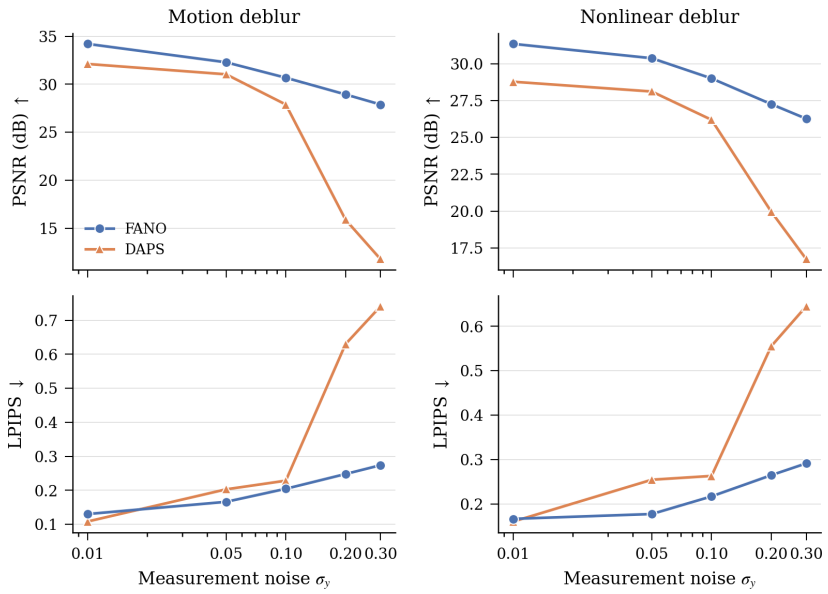
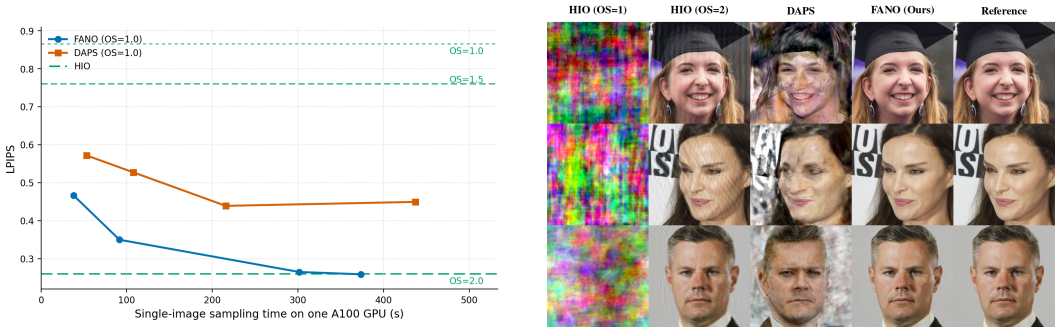


Figure E.1: Reconstruction performance of FANO and DAPS at different measurement noise levels σ_y. The left column shows motion deblurring and the right column nonlinear deblurring; the top row shows PSNR (↑) and the bottom row LPIPS (↓). FANO uses $N = 50, K = 30$ and DAPS uses its recommended 1k configuration. All results are averaged over 20 FFHQ images.



(a) Time–LPIPS trade-off. **(b)** Visual comparison at OS $= 1$.

Figure E.2: Phase retrieval without oversampling (OS $= 1$, noiseless) on 100 FFHQ images. (a) LPIPS as a function of per-image wall-clock time on a single A100: solid curves show FANO and DAPS under varying compute budgets; dashed lines mark HIO at three oversampling rates. (b) Three representative examples comparing HIO (OS $= 1$), HIO (OS $= 2$), DAPS, FANO, and the reference. At roughly 300 s per image, FANO reaches perceptual quality comparable to HIO at OS $=$ 2, recovering clean structures without any oversampling.

$\sigma_y \geq 0.20$: its motion deblurring PSNR drops from 27.88 dB at $\sigma_y = 0.10$ to 15.88 dB at $\sigma_y = 0.20$, while LPIPS increases from 0.228 to 0.630.

E.2 Phase Retrieval without Oversampling: Compensation by Generative Priors

Classical phase retrieval methods, such as HIO (Fienup, 1982), already provide reliable recovery in the noiseless setting when the oversampling ratio satisfies OS ≥ 2. Oversampling reduces ill-conditioning by increasing the measurement dimension, but lower oversampling ratios impose fewer requirements on optical systems and reduce acquisition costs in practice. A potential advantage of

diffusion generative priors is their ability to compensate for missing oversampling information and enable high-quality recovery even at OS $= 1$, without oversampling.

Figure E.2 compares FANO, DAPS, and classical HIO on 100 FFHQ test images. The experiment follows the phase retrieval configuration in Section 4.1, changing only OS to 1 and removing measurement noise. At its largest budget of 437 seconds per image, DAPS reduces LPIPS to only approximately 0.45, whereas FANO reaches 0.26 at approximately 300 seconds, comparable to HIO at OS $= 2$. Representative examples in Figure E.2(b) further corroborate this gap: at OS $= 1$, FANO recovers structures and textures close to the reference, whereas DAPS exhibits substantial distortions and HIO fails completely at OS $= 1$.

HIO is implemented using Fourier-magnitude projections and a known image-support constraint, processing the three RGB channels separately and clipping the final output to $[0, 1]$. For all OS $\in \{1, 1.5, 2\}$, the feedback parameter is fixed at $\beta = 0.9$, and each candidate undergoes 500 updates.

F Additional Qualitative Results

We present representative reconstructions for each task below. Each figure contains two tasks (left and right); for each task, the upper half shows FFHQ and the lower half ImageNet, with three examples per dataset. The columns, from left to right, show the degraded measurement (Measurement), DAPS, SITCOM, FANO (ours), and the ground truth (Reference).

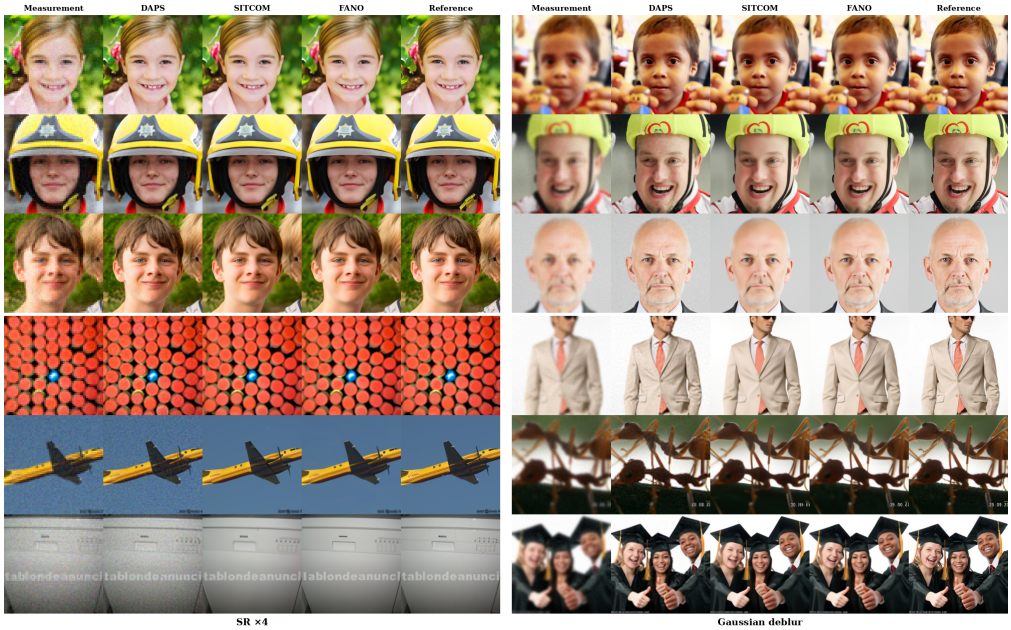


Figure F.1: **Super resolution** (*left*) and **Gaussian deblurring** (*right*) results. First (resp. last) three rows are for the FFHQ (resp. ImageNet) dataset.

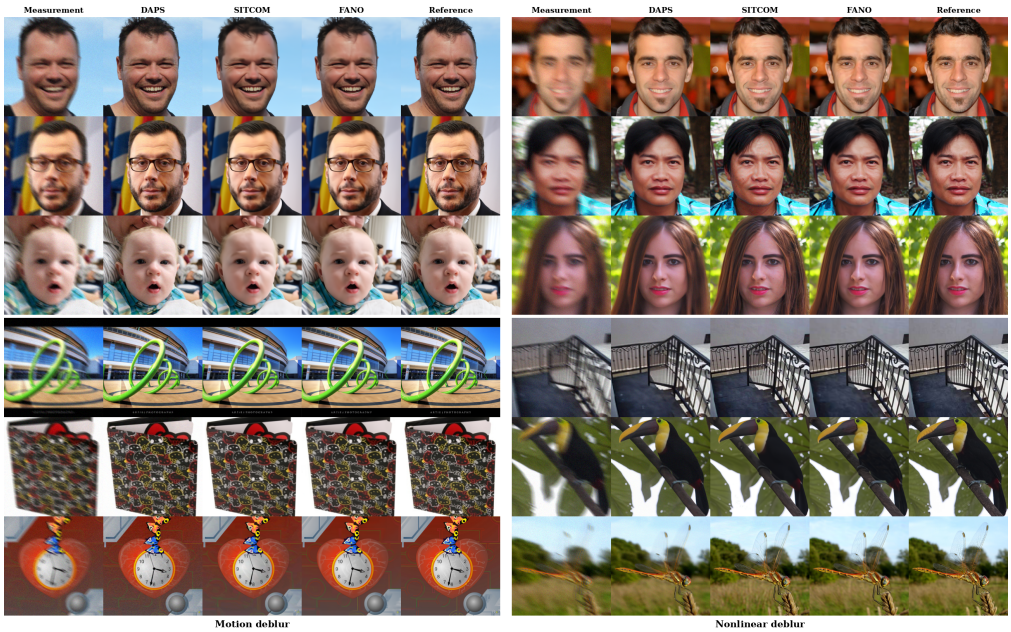


Figure F.2: **Motion deblurring** (*left*) and **nonlinear deblurring** (*right*) results. First (resp. last) three rows are for the FFHQ (resp. ImageNet) dataset.

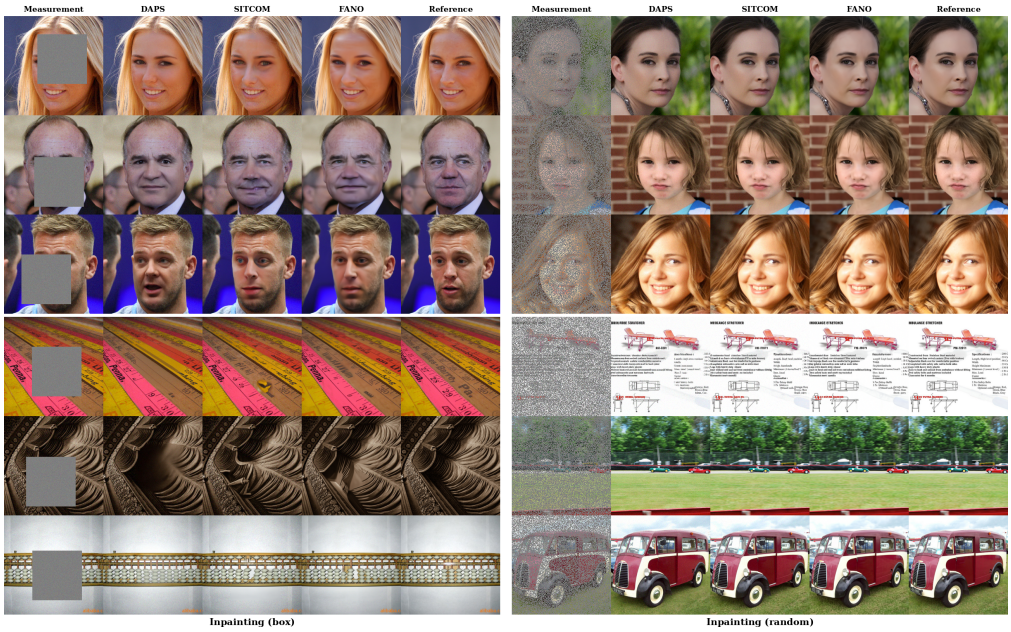


Figure F.3: **Box inpainting** (*left*) and **random inpainting** (*right*) results. First (resp. last) three rows are for the FFHQ (resp. ImageNet) dataset.



Figure F.4: **Phase retrieval** (*left*) and **high dynamic range** (*right*) results. First (resp. last) three rows are for the FFHQ (resp. ImageNet) dataset.