
Latent Refinement via Flow Matching for Training-free Linear Inverse Problem Solving

Hossein Askari¹ Yadan Luo¹ Hongfu Sun¹ Fred Roosta^{1,2}

¹The University of Queensland

²ARC Training Centre for Information Resilience (CIRES)

{h.askari, yadan.luo, hongfu.sun, fred.roosta}@uq.edu.au

Abstract

Recent advances in *inverse problem* solving have increasingly adopted flow *priors* over diffusion models due to their ability to construct straight probability paths from noise to data, thereby enhancing efficiency in both training and inference. However, current flow-based inverse solvers face two primary limitations: (i) they operate directly in pixel space, which demands heavy computational resources for training and restricts scalability to high-resolution images, and (ii) they employ guidance strategies with *prior*-agnostic posterior covariances, which can weaken alignment with the generative trajectory and degrade posterior coverage. In this paper, we propose **LFlow** (Latent Refinement via **F**lows), a *training-free* framework for solving linear inverse problems via pretrained latent flow priors. LFlow leverages the efficiency of flow matching to perform ODE sampling in latent space along an optimal path. This latent formulation further allows us to introduce a theoretically grounded posterior covariance, derived from the optimal vector field, enabling effective flow guidance. Experimental results demonstrate that our proposed method outperforms state-of-the-art latent diffusion solvers in reconstruction quality across most tasks. The code will be publicly available at [GitHub](#).

1 Introduction

Linear *inverse problems* are fundamental to a variety of significant image processing tasks, such as super-resolution [1], inpainting [2], deblurring [3], and denoising [4]. Solving such problems involves inferring an unknown image $\mathbf{x}_0 \in \mathbb{R}^n$, which is assumed to follow an unknown prior distribution $q(\mathbf{x}_0)$, from incomplete and noisy observations $\mathbf{y} \in \mathbb{R}^m$, commonly modeled as:

$$\mathbf{y} = \mathcal{A}\mathbf{x}_0 + \mathbf{n}, \quad \mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma_y^2 \mathbf{I}), \quad (1)$$

where $\mathcal{A} \in \mathbb{R}^{m \times n}$ represents a known linear operator and $\mathbf{n}$ denotes additive i.i.d. Gaussian noise. When the operator $\mathcal{A}$ is singular (e.g., if $m < n$), the inverse problem becomes ill-posed [5], hindering the unique or stable recovery of $\mathbf{x}_0$ from $\mathbf{y}$. Consequently, accurate and plausible inferences demand strong *priors* that effectively integrate domain-specific knowledge to constrain the solution space.

Deep generative models that perform progressive *refinement* via stochastic differential equations (SDEs), particularly diffusion models [6–8], have solidified their role as powerful *priors* for solving a broad spectrum of inverse problems [9, 10]. Specifically, these models have proven effective for *zero-shot* inference of images from partially acquired and noisy measurements, with extensive research focusing on the design of *guidance* mechanisms to inject data consistency into the generative process [11–32]. Building on these advancements, diffusion-based inverse solvers have been further extended to operate in latent spaces [33–39] rather than raw pixel spaces, aiming to reduce the computational cost of training and improve generalization [40]. However, these approaches often neglect posterior variability by assuming zero covariance in likelihood-based guidance, which can lead to unstable sampling and reduced coverage of the posterior distribution [21, 30].

Recently, *flow matching* [41, 42] has gained prominence as a compelling alternative for generative modeling. By parameterizing transformation dynamics with ordinary differential equations (ODEs), these models can generate arbitrary probability paths, including those grounded in optimal transport (OT) principles [41]. This flexibility enables the design of straight-line generative trajectories, leading to more efficient training and sampling compared to diffusion-based approaches [43, 44]. Motivated by these capabilities, several recent works have explored the use of flow-based priors for inverse problems, achieving faster and higher-quality solutions across diverse tasks [45–50]. Nevertheless, existing methods still suffer from two key drawbacks: (1) they operate in pixel space, which restricts scalability to high-dimensional data and limits generalizability across different types of inverse problems; and (2) they adopt guidance techniques originally developed for diffusion models, which estimate posterior covariances independently of the learned prior. This disconnect may steer the sampling trajectory away from high-probability regions, leading to degraded sample quality and reduced fidelity, with slower convergence often observed when *adaptive* ODE solvers are employed.

To address these limitations, we propose **LFlow** (Latent Refinement via **Fl**ows), a framework that utilizes latent flow matching to solve linear inverse problems without additional training. By applying flow matching in latent space, LFlow achieves enhanced computational efficiency and enables more scalable and effective inverse solutions in reduced-dimensional domains. Additionally, we introduce a well-founded, time-dependent variance for the latent identity posterior covariance, formulated using Tweedie’s covariance formula and the optimal vector field under the assumption of a *Gaussian latent representation*. This posterior covariance is explicitly informed by the pretrained optimal vector field, ensuring that guidance remains consistent with the generative dynamics. Our empirical evaluations demonstrate that images inferred via latent ODE sampling along conditional OT paths exhibit superior perceptual quality compared to those generated through latent diffusion-based probability paths.

Our primary contributions are as follows:

- **Methodological:** We propose a training-free framework based on latent flow matching and posterior-guided ODE sampling for solving linear inverse problems, significantly outperforming latent diffusion-based approaches in both efficiency and reconstruction quality.
- **Analytical:** We derive a principled correction to the pretrained latent flow using the measurement likelihood gradient and introduce an analytically justified, time-dependent posterior covariance to improve sampling accuracy and convergence speed.
- **Empirical:** We validate the performance of LFlow through extensive experiments on image reconstruction tasks, including deblurring, super-resolution, and inpainting, achieving state-of-the-art results *without* requiring substantial problem-specific hyperparameter tuning.

2 Overview of Related Work

Training-free inverse problem solvers that exploit diffusion or flow priors can be generally categorized into four methodological classes: (1) *Variable splitting* methods decompose inference into two alternating steps: one enforces data fidelity and the other imposes regularization [51, 36, 20]; (2) *Variational Bayesian* methods introduce a parameterized surrogate posterior distribution, typically Gaussian, whose parameters are optimized using a variational objective [27, 52, 53]; (3) *Asymptotically exact* methods combine generative priors with classical samplers—such as MCMC, SMC, or Gibbs sampling—to approximate the true posterior with convergence guarantees as the sample size grows [54–57]; and (4) *Guidance-based* methods correct the generative trajectory using an approximate likelihood gradient to steer samples toward the posterior [12, 21–23]. Our work centers on the fourth category and further elaborates on related methods built on various types of *priors*.

Diffusion Guidance Approximation refers to estimating the likelihood score $\nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\mathbf{x}_t)$ during the reverse-time diffusion process governed by an SDE of the form:

$$d\mathbf{x}_t \approx [\mathbf{f}(\mathbf{x}_t, t) - \mathbf{g}(t)^2 (\mathbf{s}_\theta(\mathbf{x}_t, t) + \nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\mathbf{x}_t))] dt + \mathbf{g}(t) d\mathbf{w}_t, \quad (2)$$

where $\mathbf{f}(\cdot, \cdot)$ is the drift, $\mathbf{g}(\cdot)$ the diffusion coefficient, $\mathbf{w}_t$ standard Brownian motion, and $\mathbf{s}_\theta(\cdot, \cdot)$ a pretrained score network. Notable methods such as DPS [21], IIGDM [22], TMPD [23], OPC [29], and MMPS [30] define the likelihood score as $\nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\mathbf{x}_t) = \nabla_{\mathbf{x}_t} \log \int p(\mathbf{y}|\mathbf{x}_0) p(\mathbf{x}_0|\mathbf{x}_t) d\mathbf{x}_0$. The main challenge lies in computing the expectation over all possible denoised states $\mathbf{x}_0$ given $\mathbf{x}_t$, which requires sampling from $p(\mathbf{x}_0|\mathbf{x}_t)$ at every reverse step—posing significant computational demands. A

common remedy is a local Gaussian approximation $p(\mathbf{x}_0|\mathbf{x}_t) \approx \mathcal{N}(\mathbb{E}[\mathbf{x}_0|\mathbf{x}_t], \text{Cov}[\mathbf{x}_0|\mathbf{x}_t])$, reducing the problem to estimating posterior moments. These approaches differ mainly in how they specify $\text{Cov}[\mathbf{x}_0|\mathbf{x}_t]$. For instance, DPS sets it to zero, and IIGDM obtains a *prior-agnostic*, time-scaled identity matrix derived from the *forward process* under a *Gaussian data assumption*. OPC performs a *post hoc* constant variance optimization for each step. TMPD approximates the covariance by replacing the Jacobian in the Tweedie relation $\text{Cov}[\mathbf{x}_0|\mathbf{x}_t] = \frac{\sigma(t)^2}{\alpha(t)} \nabla_{\mathbf{x}_t}^\top \mathbb{E}[\mathbf{x}_0|\mathbf{x}_t]$ with a diagonal row-sum surrogate, whereas MMPS evaluates the full Jacobian via automatic differentiation and solves the induced linear system with Conjugate Gradients [58]. Estimating the full covariance, however, remains computationally expensive in high dimensions, with prohibitive memory and runtime costs.

Inference in Latent Space has become feasible with latent diffusion models (LDMs) [59], allowing inverse solvers to reduce training costs and improve scalability [33, 36–38]. As an initial attempt, PS�D [33] augments DPS with a “gluing” objective to enforce posterior mean consistency under the autoencoder mapping, aiming to mitigate encoder–decoder nonlinearity. However, this constraint remains empirically ineffective under noisy measurements, and reconstruction artifacts persist [34]. Other approaches avoid explicit nonlinearity correction: In particular, Resample [36] constructs a Gaussian posterior by fusing a supposedly Gaussian prior on the unconditional *reverse* sample with a Gaussian pseudo-likelihood centered at a *forward-projected* measurement-consistent posterior mean. Similarly, DAPS [37] enhances posterior sampling by decoupling consecutive diffusion steps through a two-step procedure: first drawing $\mathbf{z}_0 \sim p(\mathbf{z}_0|\mathbf{z}_t, \mathbf{y})$, then re-noising $\mathbf{z}_{t-\Delta t} \sim \mathcal{N}(\mathbf{z}_0, \sigma_{t-\Delta t}^2 \mathbf{I})$, which provides *global* corrections—particularly effective in non-linear tasks, though at the expense of weaker local guidance in low-noise or linear settings. SITCOM [60] enforces three consistency conditions—data, forward, and backward diffusion—at each step, thereby enabling sampling with fewer steps. Nevertheless, these methods largely assume zero covariances, limiting posterior coverage.

Flow Matching in Inverse Problems has recently proven effective for achieving fast and high-quality solutions across various tasks [45–50]. A prime example is OT-ODE [45], which adopts IIGDM [22] gradient correction within the flow regime and employs an ODE solver scheme based on the conditional OT path. C-IIGFM [48] introduces a plug-and-play framework that projects conditional flow dynamics into a more amenable space, accelerating inference. PnP-Flow [61] solves imaging inverse problems by alternating a data-fidelity gradient step, a re-projection onto the flow path via latent-noise interpolation, and a time-dependent FM denoiser, all without backpropagating through the ODE. However, each of these methods typically (1) borrows guidance strategies from diffusion models that assume either zero or prior-agnostic posterior covariances, and (2) operates in pixel space, which limits their scalability and applicability to high-dimensional problems.

3 Preliminaries

Continuous Normalizing Flow (CNF) [62] constructs a smooth probability path $\{p_t(\mathbf{x}_t)\}_{t \in [0,1]}$ that transports samples from a data distribution $q(\mathbf{x}_0)$, with $\mathbf{x}_0 \in \mathbb{R}^d$, to a standard Gaussian $\mathcal{N}(\mathbf{0}, \mathbf{I})$ at $t = 1$. This evolution follows a time-varying vector field $\mathbf{v} : \mathbb{R}^d \times [0, 1] \rightarrow \mathbb{R}^d$, governed by the ODE $d\mathbf{x}_t = \mathbf{v}(\mathbf{x}_t, t) dt$. In practice, $\mathbf{v}(\cdot, t)$ is approximated by a learnable field $\mathbf{v}_\theta$, trained via maximum likelihood, which requires expensive ODE simulations.

Flow Matching (FM) [41] avoids inefficient likelihood-based training for CNFs by directly regressing a learnable vector field toward an analytically defined target field. In particular, *Conditional Flow Matching* (CFM) [41] introduces a time-dependent *conditional vector field* $\mathbf{v}(\mathbf{x}_t | \mathbf{x}_0)$ that governs the evolution of samples $\mathbf{x}_t$ conditioned on an initial point $\mathbf{x}_0$. This field induces a *conditional probability path* $p_t(\mathbf{x}_t | \mathbf{x}_0)$ satisfying the boundary conditions $p_0 = \delta(\mathbf{x}_0)$ and $p_1 = \mathcal{N}(\mathbf{0}, \mathbf{I})$. The training objective then becomes:

$$\mathcal{L}_{\text{CFM}}(\theta) = \mathbb{E}_{t, \mathbf{x}_0 \sim q, \mathbf{x}_t \sim p_t(\mathbf{x}_t | \mathbf{x}_0)} \|\mathbf{v}_\theta(\mathbf{x}_t, t) - \mathbf{v}(\mathbf{x}_t | \mathbf{x}_0)\|^2. \quad (3)$$

Using Gaussian paths, we define $p_t(\mathbf{x}_t | \mathbf{x}_0) = \mathcal{N}(\alpha(t)\mathbf{x}_0, \sigma(t)^2 \mathbf{I})$ with interpolation $\mathbf{x}_t = \alpha(t)\mathbf{x}_0 + \sigma(t)\mathbf{x}_1$, where $\mathbf{x}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Differentiating $\mathbf{x}_t$ with respect to t gives $\dot{\alpha}(t)\mathbf{x}_0 + \dot{\sigma}(t)\mathbf{x}_1$. Substituting the inversion $\mathbf{x}_0 = [\mathbf{x}_t - \sigma(t)\mathbf{x}_1]/\alpha(t)$ yields the true vector field:

$$\mathbf{v}(\mathbf{x}_t | \mathbf{x}_0) = \frac{\dot{\alpha}(t)}{\alpha(t)} \mathbf{x}_t + \sigma(t) \left(\frac{\dot{\sigma}(t)}{\sigma(t)} - \frac{\dot{\alpha}(t)}{\alpha(t)} \right) \mathbf{x}_1, \quad (4)$$

where $\dot{\alpha}(t)$ and $\dot{\sigma}(t)$ denote time derivatives. In particular, choosing $\alpha(t) = 1 - t$ and $\sigma(t) = t$ recovers the OT path, which induces straight trajectories and improves efficiency.

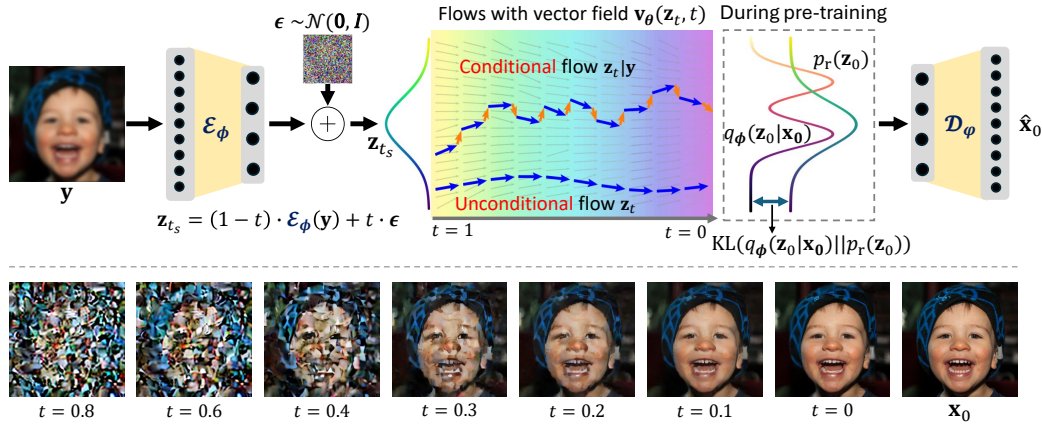


Figure 1: **(Top)** LFlow pipeline: A VAE encoder maps the observation y to a latent z , from which a noisy start z_{t_s} is formed. During VAE pre-training, the KL term encourages the encoder’s approximate posterior $q_\phi(z_0 | x_0)$ to approach the Gaussian prior $p_r(z_0) = \mathcal{N}(\mathbf{0}, \mathbf{I})$. A pre-trained flow field $\mathbf{v}_\theta(z_t, t)$ defines the unconditional trajectory z_t . At inference, orange arrows ($\rightarrow$) denote likelihood-based guidance that corrects the prior field, yielding the conditional latent path $z_t | y$ toward $p(z_0 | y)$. Decoding with $\mathcal{D}_\phi$ produces $\hat{x}_0$. **(Bottom)** Gaussian deblurring snapshots along the conditional path as t decreases from 0.8 to 0.

4 Method

To address the ill-posedness of linear inverse problems, we adopt a Bayesian view and target the posterior $p(x_0 | y) \propto p(y | x_0) p(x_0)$, where $p(y | x_0)$ is the likelihood and $p(x_0)$ is a learned prior induced by a pretrained latent flow (via the decoder). Our goal is to generate samples from this posterior without retraining a task-specific model.

4.1 Latent Refinement via Flows (LFlow)

We represent the prior distribution $p(x_0)$ implicitly via a latent prior $p(z_0)$ and the decoder push-forward $p(x_0) = \mathcal{D}_\phi \# p(z_0)$. Specifically, let $\mathcal{E}_\phi : \mathbb{R}^d \rightarrow \mathbb{R}^k$ and $\mathcal{D}_\phi : \mathbb{R}^k \rightarrow \mathbb{R}^d$ be a pretrained autoencoder. The latent $z_0 = \mathcal{E}_\phi(x_0)$ follows $p(z_0)$, modeled by a flow-matching velocity $\mathbf{v}_\theta(z_t, t)$ defining the ODE

$$dz_t = \mathbf{v}_\theta(z_t, t) dt, \quad t \in [0, 1]. \quad (5)$$

This flow transports latent samples from $z_0 \sim p(z_0)$ to a noise distribution $z_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. At inference, we integrate *backward* from z_1 to $\hat{z}_0$, adding measurement-driven guidance to follow the conditional trajectory, then decode $\hat{x}_0 = \mathcal{D}_\phi(\hat{z}_0)$. Figure 1 illustrates the overall inference pipeline and provides a visual example of decoded reconstructions along the conditional trajectory, offering intuition for how the flow progresses in practice.

Posterior Sampling via Conditional Flows in Latent Space We aim to sample from the posterior $p(z_0 | y)$ by simulating a reverse-time ODE characterized by a conditional velocity field. The following proposition provides the theoretical foundation for this procedure, ensuring that if the conditional density evolves according to a continuity equation, then the reverse-time flow recovers samples from the posterior. A formal proof is deferred to Appendix A.1.

Proposition 4.1 (Posterior Sampling via Reverse-Time Conditional Flows). *Let $p_t(z_t | y)$ denote the conditional distribution of latent variables $z_t \in \mathbb{R}^k$ at time $t \in [0, 1]$, with terminal condition $p_1(z_1) = \mathcal{N}(\mathbf{0}, \mathbf{I})$. Suppose this distribution evolves over time according to the reverse-time continuity equation:*

$$\partial_t p_t(z_t | y) = \nabla_{z_t} \cdot [p_t(z_t | y) \mathbf{v}_t(z_t | y)], \quad (6)$$

for some conditional velocity field $\mathbf{v}_t(z_t | y)$. Then, the solution to the reverse-time ODE

$$dz_t = -\mathbf{v}_t(z_t | y) dt, \quad z_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \quad (7)$$

yields samples from the posterior $p(z_0 | y)$ as $t \rightarrow 0$.

Conditional Vector Field Estimation Having established that posterior samples can be obtained by integrating the reverse-time ODE in Eq. (7), it remains to estimate the corresponding conditional

vector field $\mathbf{v}_t(\mathbf{z}_t|\mathbf{y})$. Under Gaussian latent dynamics, this field takes the form:

$$\mathbf{v}_t(\mathbf{z}_t | \mathbf{y}) = \frac{\dot{\alpha}(t)}{\alpha(t)} \mathbf{z}_t + \lambda(t) \nabla_{\mathbf{z}_t} \log p_t(\mathbf{z}_t | \mathbf{y}), \quad \lambda(t) = \frac{d}{dt} \left(\frac{\sigma(t)}{\alpha(t)} \right). \quad (8)$$

By applying Bayes' rule, we obtain a principled decomposition of the velocity field, which leads to the following practical approximation via the pretrained flow and a likelihood correction.

$$\mathbf{v}_t(\mathbf{z}_t | \mathbf{y}) \approx \mathbf{v}_\theta(\mathbf{z}_t, t) - \frac{t}{1-t} \nabla_{\mathbf{z}_t} \log p_t(\mathbf{y} | \mathbf{z}_t), \quad (9)$$

where the additional term $\nabla_{\mathbf{z}_t} \log p_t(\mathbf{y} | \mathbf{z}_t)$, commonly referred to as *guidance*, steers the flow toward consistency with the measurements $\mathbf{y}$. A complete proof of the results in Eq. (8) and Eq. (9) in the latent domain is provided in Appendix A.2.

Likelihood Approximation A key challenge in Eq. (9) arises from approximating the gradient of the noise-conditional distribution, which we define as

$$\nabla_{\mathbf{z}_t} \log p(\mathbf{y} | \mathbf{z}_t) = \nabla_{\mathbf{z}_t} \log \int p(\mathbf{y} | \mathbf{x}_0, \mathbf{z}_t) p(\mathbf{x}_0 | \mathbf{z}_t) d\mathbf{x}_0 = \nabla_{\mathbf{z}_t} \log \mathbb{E}_{\mathbf{x}_0 \sim p(\mathbf{x}_0 | \mathbf{z}_t)} [p(\mathbf{y} | \mathbf{x}_0)]. \quad (10)$$

Here, the likelihood term can be expressed as $p(\mathbf{y} | \mathbf{x}_0) = \mathcal{N}(\mathbf{y}; \mathcal{A}\mathcal{D}_\varphi(\mathbf{z}_0), \sigma_y^2 \mathbf{I})$, which is a Gaussian distribution with a nonlinear mean induced by the decoder $\mathcal{D}_\varphi$. Due to the nonlinearity of $\mathcal{D}_\varphi$, the marginal likelihood $p(\mathbf{y} | \mathbf{z}_t)$ deviates from Gaussianity. To facilitate the computation of its gradient, we approximate $\mathcal{D}_\varphi(\mathbf{z}_0)$ via a first-order Taylor expansion around $\bar{\mathbf{z}}_0 := \mathbb{E}[\mathbf{z}_0 | \mathbf{z}_t]$ as:

$$\mathcal{D}_\varphi(\mathbf{z}_0) \approx \mathcal{D}_\varphi(\bar{\mathbf{z}}_0) + J_{\mathcal{D}}(\mathbf{z}_0 - \bar{\mathbf{z}}_0), \quad (11)$$

where $J_{\mathcal{D}} := J_{\mathcal{D}}(\bar{\mathbf{z}}_0)$ is the Jacobian of $\mathcal{D}_\varphi$ at $\bar{\mathbf{z}}_0$. Assuming the posterior $p(\mathbf{z}_0 | \mathbf{z}_t) \sim \mathcal{N}(\bar{\mathbf{z}}_0, \Sigma_{\mathbf{z}})$, the distribution of the image $\mathbf{x}_0 = \mathcal{D}_\varphi(\mathbf{z}_0)$ becomes approximately Gaussian as well:

$$\mathbf{x}_0 \sim \mathcal{N}(\mathbb{E}[\mathbf{x}_0 | \mathbf{z}_t], \text{Cov}[\mathbf{x}_0 | \mathbf{z}_t]), \text{ where } \mathbb{E}[\mathbf{x}_0 | \mathbf{z}_t] \approx \mathcal{D}_\varphi(\bar{\mathbf{z}}_0), \quad \text{Cov}[\mathbf{x}_0 | \mathbf{z}_t] \approx J_{\mathcal{D}} \Sigma_{\mathbf{z}} J_{\mathcal{D}}^\top. \quad (12)$$

Using this approximation, we estimate the gradient of the log-likelihood in latent space based on the decoded mean and the propagated covariance (see Appendix A.4):

$$\nabla_{\mathbf{z}_t} \log p(\mathbf{y} | \mathbf{z}_t) \approx \underbrace{(\nabla_{\mathbf{z}_t} \bar{\mathbf{z}}_0)^\top J_{\mathcal{D}}^\top}_{\mathbf{J}} \underbrace{\mathcal{A}^\top (\sigma_y^2 \mathbf{I} + \mathcal{A} J_{\mathcal{D}} \Sigma_{\mathbf{z}} J_{\mathcal{D}}^\top \mathcal{A}^\top)^{-1} (\mathbf{y} - \mathcal{A} \mathcal{D}_\varphi(\bar{\mathbf{z}}_0))}_{\mathbf{v}}, \quad (13)$$

which involves a Jacobian-vector product ($\mathbf{J} \cdot \mathbf{v}$), and can be efficiently computed using automatic differentiation (Appendix B). Furthermore, the Gaussian posterior assumption enables the application of Tweedie's formula [63], which connects the pretrained vector field $\mathbf{v}_\theta(\mathbf{z}_t, t)$ to the posterior moments $\mathbb{E}[\mathbf{z}_0 | \mathbf{z}_t]$ and $\text{Cov}[\mathbf{z}_0 | \mathbf{z}_t]$. Specifically, we have (refer to Appendix A.3):

$$\mathbb{E}[\mathbf{z}_0 | \mathbf{z}_t] = \mathbf{z}_t - t \mathbf{v}_\theta(\mathbf{z}_t, t), \quad (14)$$

$$\text{Cov}[\mathbf{z}_0 | \mathbf{z}_t] = \frac{t^2}{1-t} (\mathbf{I} - t \nabla_{\mathbf{z}_t} \mathbf{v}_\theta(\mathbf{z}_t, t)). \quad (15)$$

Latent Posterior Covariance Computing the posterior covariance $\text{Cov}[\mathbf{z}_0 | \mathbf{z}_t]$ in low-dimensional latent spaces is more tractable and can be efficiently estimated via automatic differentiation. However, inverting the matrix $(\sigma_y^2 \mathbf{I} + \mathcal{A} J_{\mathcal{D}} \Sigma_{\mathbf{z}} J_{\mathcal{D}}^\top \mathcal{A}^\top)$ in Eq. (13) remains a significant computational bottleneck. To avoid explicit inversion, one may solve the corresponding linear system using the Conjugate Gradient (CG) method. Yet, CG's convergence crucially depends on the symmetry and positive definiteness of the system matrix—conditions that may be violated in practice due to imperfections in the pretrained velocity field $\mathbf{v}_\theta$, potentially leading to instability or divergence. To address this, we further analyze the structure of the posterior covariance and the Jacobian $\nabla_{\mathbf{z}_t} \mathbf{v}_\theta(\mathbf{z}_t, t)$, under certain regularity assumptions on the latent prior $p(\mathbf{z}_0)$.

Assumption 4.2 (Strong Log-Concavity of the Latent Prior [64]). *Let $p(\mathbf{z}_0) = \exp(-\Phi(\mathbf{z}_0))$ be the latent prior over $\mathbb{R}^d$, where Φ is a twice continuously differentiable potential function. We assume that $p(\mathbf{z}_0)$ is γ -strongly log-concave for some $\gamma > 0$; that is, $\nabla^2 \Phi(\mathbf{z}_0) \succeq \gamma \mathbf{I}_d$, for all $\mathbf{z}_0 \in \mathbb{R}^d$.*

(1) Log-concave Prior: This assumption imposes a uniform lower bound on the Hessian of the prior's potential function, which translates to a lower bound on the Jacobian of the vector field.

Proposition 4.3 (Bound on the Jacobian of the Vector Field). Let $\mathbf{v}_\theta(\mathbf{z}_t, t)$ denote the velocity field of the interpolant $\mathbf{z}_t = \alpha(t)\mathbf{z}_0 + \sigma(t)\mathbf{z}_1$ between a standard Gaussian prior and a target distribution $p(\mathbf{z}_0)$, defined via coefficients $\alpha(t), \sigma(t) \in \mathbb{R}$. Under Assumption 4.2, the Jacobian of the vector field satisfies the following bound:

$$\frac{d}{dt} \left(\frac{1}{2} \log (\alpha(t)^2 + \gamma \sigma(t)^2) \right) \cdot \mathbf{I}_{d_z} \preceq \nabla_{\mathbf{z}_t} \mathbf{v}_\theta(\mathbf{z}_t, t) \prec \frac{d}{dt} \log \sigma(t) \cdot \mathbf{I}_{d_z} \quad \forall t \in (0, 1], \mathbf{z} \in \mathbb{R}^d. \quad (16)$$

The proof is provided in Appendix A.5. The resulting sandwich bound ensures the Jacobian remains well-behaved—the upper bound guarantees valid and stable posterior covariance, while the lower bound prevents overestimated uncertainty during posterior-guided inference.

(2) Gaussian Special Case: For practical inference and efficient computation, we now consider a tractable special case where the latent prior is Gaussian, $\mathbf{z}_0 \sim \mathcal{N}(0, \sigma_{\text{latr}}^2 \mathbf{I})$, a choice that is both theoretically justified and widely used in practice—for instance, in variational autoencoders (VAEs), where the latent prior is regularized toward a standard Gaussian via KL divergence [65]. This enables us to derive the optimal velocity field and its Jacobian in closed form, which is characterized in the next proposition, whose proof can be found in Appendix A.6.

Proposition 4.4 (Optimal Vector Field). Let $\mathbf{z}_0 \sim \mathcal{N}(0, \sigma_{\text{latr}}^2 \mathbf{I}_{d_z})$ and $\mathbf{z}_1 \sim \mathcal{N}(0, \mathbf{I}_{d_z})$ be independent random variables. Define $\mathbf{z}_t = (1 - t)\mathbf{z}_0 + t\mathbf{z}_1$ for $t \in [0, 1]$. The optimal vector field $\mathbf{v}^*(\mathbf{z}_t, t)$ that minimizes the expected squared error $\arg \min_{\mathbf{v}} \mathbb{E} [\|\mathbf{v}(\mathbf{z}_t, t) - (\mathbf{z}_0 - \mathbf{z}_1)\|^2]$ is given by

$$\mathbf{v}^*(\mathbf{z}_t, t) = \frac{(1 - t)\sigma_{\text{latr}}^2 - t}{(1 - t)^2\sigma_{\text{latr}}^2 + t^2} \mathbf{z}_t. \quad (17)$$

Remark 4.5 (Tightness of Bound). When $\mathbf{z}_0 \sim \mathcal{N}(0, \sigma_{\text{latr}}^2 \mathbf{I})$, the optimal vector field $\mathbf{v}^*(\mathbf{z}_t, t)$ achieves the Jacobian lower bound in Proposition 4.3. This follows from the potential function $\Phi(\mathbf{z}_0) = \frac{1}{2\sigma_{\text{latr}}^2} \|\mathbf{z}_0\|^2$, yielding $\gamma = \sigma_{\text{latr}}^{-2}$, and interpolation coefficients $\alpha(t) = 1 - t$, $\sigma(t) = t$. Substituting into the bound confirms it matches the exact Jacobian of $\mathbf{v}^*$, making the bound tight.

Plugging the Jacobian of the optimal vector field into Eq. (15) yields:

$$\text{Cov}[\mathbf{z}_0 | \mathbf{z}_t] = r^2(t) \cdot \mathbf{I}_{d_z}, \quad \text{with} \quad r^2(t) = \frac{t^2 [(1 - t)(1 - 2t) + 2t^2]}{(1 - t) [(1 - t)^2 + t^2]}, \quad (18)$$

where we assumed $\sigma_{\text{latr}} = 1$. As a result, the propagated covariance can also be simplified as

$$\text{Cov}[\mathbf{x}_0 | \mathbf{z}_t] \approx J_{\mathcal{D}} \cdot \text{Cov}[\mathbf{z}_0 | \mathbf{z}_t] \cdot J_{\mathcal{D}}^\top \approx r^2(t) \cdot J_{\mathcal{D}} J_{\mathcal{D}}^\top. \quad (19)$$

Computing the full Jacobian $J_{\mathcal{D}} \in \mathbb{R}^{d_x \times d_z}$ is often infeasible in practice. To simplify, we assume that the decoder acts approximately as a local isometry near $\bar{\mathbf{z}}_0$ [66], such that $J_{\mathcal{D}} J_{\mathcal{D}}^\top \approx \cdot P$, where P is the orthogonal projector onto the image of $J_{\mathcal{D}}$. For computational convenience, we approximate this behavior as isotropic in the full space, resulting in:

$$\text{Cov}[\mathbf{x}_0 | \mathbf{z}_t] \approx r^2(t) \cdot \mathbf{I}_{d_x}. \quad (20)$$

Comparison with IIGDM and OT-ODE The posterior covariance in IIGDM is derived solely from the *forward* process under the strong assumption of *Gaussian data space*, resulting in a variance for the identity covariance as $r^2(t) = \frac{\sigma(t)^2}{\alpha(t)^2 + \sigma(t)^2} \mathbf{I}$. Please see Appendix A.7 for details.

Additionally, IIGDM leverages the inverse of this posterior covariance as Fisher information for natural gradient [67] updates on samples during the guidance step, thereby enhancing performance. OT-ODE can be regarded as a flow-based extension of IIGDM, where it adopts the same covariance expression but replaces the diffusion forward process with a flow-specific noise scheduler, specifically setting $\alpha(t) = 1 - t$ and $\sigma(t) = t$. Therefore, both methods rely on an identity posterior covariance that is *independent* of the learned score or vector field. In contrast, our proposed covariance explicitly incorporates information from the pre-trained vector field. This in turn allows for more effective guidance during flow-based ODE sampling, ultimately improving reconstruction quality. Figure 2 visually illustrates how our time-dependent variance differs in magnitude and effect from the simpler identity-based covariance used in OT-ODE.

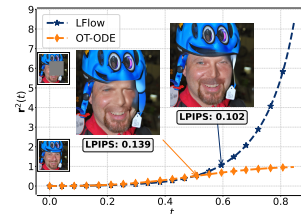


Figure 2: **Posterior covariance values** across $t \in [0, 0.9]$ for our method and OT-ODE, along with reconstructed images for a single FFHQ sample in the super-resolution task.

Initiating the Flow Sampling Process Inspired by prior works in pixel spaces [68, 45], we propose initializing the reverse ODE—across all tasks—from a *partially corrupted* version of the *encoding* of $\mathbf{y}$ in the latent space. Specifically, rather than starting from pure noise $\mathbf{z}_1 \sim \mathcal{N}(0, \mathbf{I})$ at $t = 1$, we initialize at $t_s < 1$:

$$\mathbf{z}_{t_s} = (1 - t_s) \mathcal{E}_\phi(\mathbf{y}) + t_s \mathbf{z}_1, \quad \mathbf{z}_1 \sim \mathcal{N}(0, \mathbf{I}). \quad (21)$$

This initialization ensures that $\mathbf{z}_{t_s}$ is closer to the posterior mode $\mathbf{z}_0 \mid \mathbf{y}$, making the subsequent backward integration more stable and likely to remain on a plausible manifold.

We summarize the complete sampling algorithm via latent ODE flows in Algorithm 1 (Appendix C).

5 Experiments

Datasets and Tasks We evaluate our method on three datasets: FFHQ [69], ImageNet [70], and CelebA-HQ [71], each containing images with a resolution of $256 \times 256 \times 3$ pixels. We use 200 randomly selected validation samples per dataset. All images are normalized to the range $[-1, 1]$. We present our findings on several linear inverse problem tasks, including Gaussian deblurring, motion deblurring, super-resolution, and box inpainting. The measurement operators are configured as follows: (i) Gaussian deblurring convolves images with a Gaussian blurring kernel of size 61×61 and a standard deviation of 3.0 [29]; (ii) Motion Deblurring uses motion blur kernels that are randomly generated with a size of 61×61 and an intensity value of 0.5 [29]; (iii) Super-Resolution (SR) involves bicubic downsampling by a factor of 4; and (iv) Box Inpainting simulates missing data by masking a 128×128 pixel box, which is randomly positioned around the center with a margin of $[16, 16]$ pixels, following the methodology of [21]. All measurements in these experiments are corrupted by Gaussian noise with a standard deviation of $\sigma_y = 0.01$, ensuring that the reconstruction methods are evaluated under realistic noisy conditions.

Baselines and Metrics Since the primary goal of this study is to improve solving inverse problems in the latent space, we focus on comparing our proposed method against state-of-the-art latent diffusion solvers—specifically, **PSLD** [33], **MPGD** [38], **Resample** [36], **DMplug** [51], **DAPS** [37], and **SITCOM** [60] through both quantitative and qualitative analyses. Additionally, we include results from the pixel-based **IIGDM** [22], **OT-ODE** [45], and **C-IIGFM** [48], when reported, for a more comprehensive comparison. Following the evaluation protocols of prior works, we report quantitative metrics including Learned Perceptual Image Patch Similarity (LPIPS) [72], Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Fréchet Inception Distance (FID) [73] (Appendix D.3).

Architecture, Training, and Sampling In contrast to previous latent diffusion inverse solvers, which rely on the LDM-VQ-4 or Stable Diffusion (SD) v1.5 models [59] built on U-Net [74], we adopt the LFM-VAE framework [75] based on DiT transformer architectures [76] for all datasets. Following the methodology outlined in [75], we leverage their pre-trained model. It is important to note that, despite our architectural differences, the quality of our prior does not exceed that of SDs or LDMs. This limitation may partly arise from using a lower-dimensional latent space of 32×32 . After evaluating various numerical ODE integration methods, we selected `adaptive_heun` as the default solver, utilizing its reliable implementation from the open-source `torchdiffeq` library [62]. All experiments were conducted on a single NVIDIA 3090 GPU with a batch size of 1. Further details on baselines, model configurations, and other hyper-parameters are provided in Appendix C.

5.1 Results

For clarity, the quantitative results for the four tasks on both FFHQ and ImageNet are presented in Tables 1 and 2, while qualitative comparisons are shown in Figures 3 and 4. The following subsections provide a detailed discussion of each task. Due to space constraints, results for CelebA-HQ are included in Table 5 and Figure 7 in Appendix D. Additional results are also presented in this section.

Gaussian Deblurring LFlow achieves the best perceptual quality across both datasets, attaining the lowest LPIPS and highest SSIM—surpassing the second-best method by approximately 8.8% in LPIPS on FFHQ. While PSLD slightly leads in PSNR, LFlow closes the gap within 1.18 dB and avoids the over-smoothing and loss of fine detail often observed in PSLD outputs (e.g., blurred backgrounds and softened skin textures). On ImageNet, LFlow similarly excels in perceptual quality, restoring structures like the dog’s snout, fur, and facial wrinkles more faithfully. The reconstructions

Table 1: Quantitative evaluation of inverse problem solving on **FFHQ** samples of the validation dataset. **Bold** and underline indicates the best and second-best respectively. The methods shaded in gray are in pixel space.

Method	Deblurring (Gaussian)			Deblurring (Motion)			SR ($\times 4$)			Inpainting (Box)		
	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$
LFlow (ours)	29.10	0.837	0.166	30.04	0.849	0.168	<u>29.12</u>	0.841	0.176	23.85	0.867	0.132
SITCOM [60]	30.42	0.828	0.237	28.78	0.828	<u>0.183</u>	29.26	0.833	0.191	24.12	0.839	0.198
DAPS [37]	28.52	0.789	0.231	29.00	<u>0.831</u>	0.252	29.38	0.826	0.197	<u>24.83</u>	0.819	0.191
DMplug [51]	27.43	0.784	0.240	27.95	0.817	0.243	29.45	0.838	<u>0.183</u>	22.55	0.807	0.220
Resample [36]	28.73	0.801	0.201	29.19	0.828	0.184	28.90	0.804	<u>0.189</u>	20.40	<u>0.825</u>	0.243
MPGD [38]	<u>29.34</u>	0.815	0.308	27.98	0.803	0.324	27.49	0.788	0.295	20.58	0.806	0.324
PSLD [33]	30.28	<u>0.836</u>	0.281	29.21	0.812	0.303	29.07	0.834	0.270	24.21	<u>0.847</u>	0.169
OT-ODE [45]	29.73	0.819	<u>0.198</u>	28.15	0.792	0.238	28.56	0.823	0.198	25.77	0.751	0.225
IIGDM [22]	28.62	0.809	<u>0.182</u>	27.18	0.773	0.223	27.78	0.815	0.201	26.82	0.767	0.214

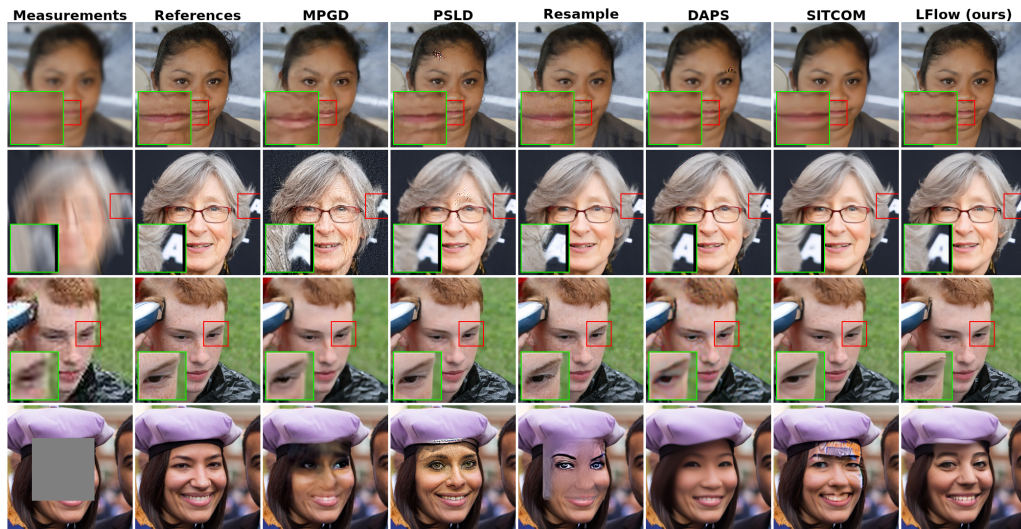


Figure 3: Qualitative results on **FFHQ** test set. Row 1: Deblur (gaussian), Row 2: Deblur (motion), Row 3: SR $\times 4$, Row 4: Inpainting. Our approach better preserves fine image details than *latent-based* diffusion methods.

remain sharp and natural, free from the spurious high-frequency artifacts or texture distortions that affect other approaches.

Motion Deblurring LFlow reconstructs motion-blurred scenes with strong structural consistency and minimal artifacts, effectively preserving sharp transitions along motion boundaries. Fine details—such as hair contours, facial edges, and accessories—are recovered with smooth gradients and natural appearances. Without introducing ringing or over-enhancement, LFlow balances fidelity and realism across both FFHQ and ImageNet, achieving state-of-the-art performance on all metrics. Notably, LFlow improves LPIPS by a clear margin of 0.015 over the second-best method, SITCOM, on FFHQ, reflecting its ability to retain high-frequency content while maintaining spatial coherence—qualities that are visually evident in the restored hair and facial structures.

Super-Resolution High-frequency structures such as hair strands, floral patterns, and facial details are reconstructed with remarkable clarity by our method, without the over-sharpening or softness often observed in competing approaches. On FFHQ, it achieves a perceptual gain of 0.021 LPIPS over the next-best method, while also maintaining competitive PSNR. On ImageNet, it establishes a clear lead across all metrics, with an LPIPS margin of 0.018. These improvements translate to visually cleaner textures and more coherent spatial gradients. Unlike MPGD, which may produce overly synthetic details, or PSLD and Resample, which occasionally yield hazy regions, our approach preserves the natural appearance of fine structures while avoiding haloing and texture overshoot.

Inpainting When completing missing regions, LFlow excels in producing semantically coherent and perceptually consistent content that blends naturally with surrounding areas. On FFHQ, it surpasses the second-best method in LPIPS by a margin of 0.037 while also achieving the highest SSIM, reflecting both perceptual sharpness and structural accuracy. On ImageNet, it continues to lead in LPIPS and SSIM, recovering fine textures and edges with minimal boundary artifacts.

Table 2: Quantitative results of inverse problem solving on **ImageNet** samples of the validation dataset. **Bold** and underline indicates the best and second-best respectively. The methods shaded in gray are in pixel space.

Method	Deblurring (Gaussian)			Deblurring (Motion)			SR ($\times 4$)			Inpainting (Box)		
	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$
LFlow (ours)	25.55	<u>0.697</u>	0.328	26.10	0.711	0.344	25.29	0.696	0.338	<u>21.92</u>	0.772	0.227
SITCOM [60]	<u>25.38</u>	0.672	0.388	24.78	0.686	0.382	25.62	0.687	0.374	20.34	0.698	0.291
DAPS [37]	24.12	0.681	0.413	<u>25.97</u>	<u>0.706</u>	0.362	25.18	0.667	<u>0.356</u>	21.13	<u>0.701</u>	0.286
Resample [36]	25.04	0.665	0.408	24.32	0.623	0.390	24.81	0.683	0.404	19.42	0.663	0.305
MPGD [38]	24.27	0.695	0.397	24.81	0.662	0.404	<u>25.50</u>	0.648	0.398	17.05	0.672	0.324
PSLD [33]	26.79	0.721	0.372	25.45	0.692	0.351	26.16	<u>0.692</u>	0.363	20.58	0.687	<u>0.274</u>
IIGDM [22]	25.27	0.636	<u>0.332</u>	23.03	0.617	<u>0.347</u>	24.73	0.629	0.359	22.13	0.589	0.361

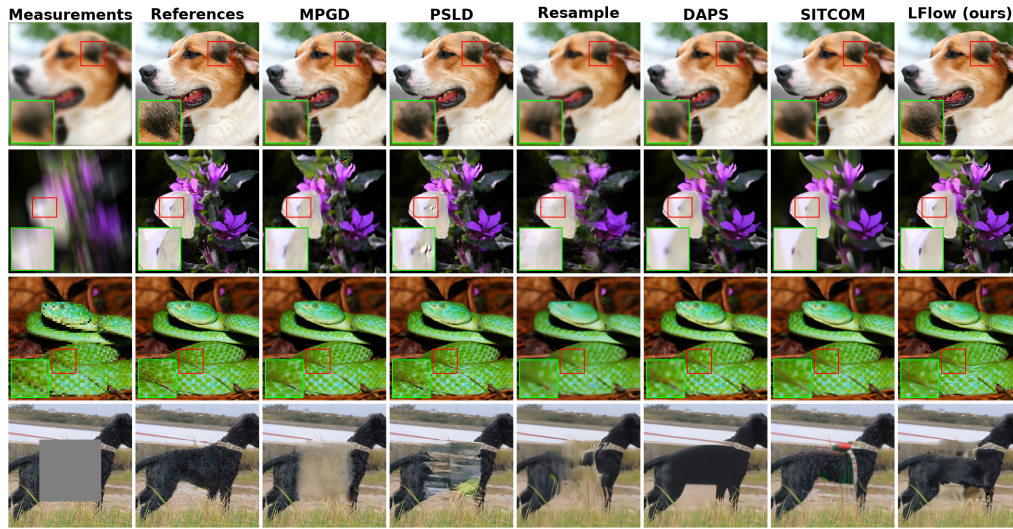


Figure 4: Qualitative results on **ImageNet** test set. Row 1: Deblur (gaussian), Row 2: Deblur (motion), Row 3: SR $\times 4$, Row 4: Inpainting. Our method reconstructs fine image details more faithfully than the baselines.

In contrast to methods that may introduce visible seams, blotchy patterns, or inconsistent colors, LFlow reconstructs faces, objects, and natural scenes with smoother transitions and well-aligned local details—resulting in reconstructions that appear complete and visually seamless.

Perceptual-Fidelity Trade-off While LFlow occasionally reports slightly lower PSNR than pixel-fidelity-oriented baselines such as PSLD, this reflects the well-known trade-off between distortion and perception. Our approach consistently achieves substantially lower LPIPS and sharper, more natural reconstructions, indicating that it prioritizes perceptual realism and fine-detail preservation over pixel-wise averaging effects that often inflate PSNR.

5.2 Ablation Study

Posterior Covariance To examine how the time-dependent posterior covariance influences overall performance, we conducted a series of ablation studies. As shown in Table 3, adapting the posterior covariance (labeled Cov_LFlow) provides systematic gains over the baseline “Cov_IIGDM”. Across both FFHQ and ImageNet, Cov_LFlow yields higher PSNR and lower LPIPS for deblurring and $\times 4$ SR, indicating that LFlow helps reduce perceptual artifacts while preserving fine details. Figure 6 presents qualitative comparisons between Cov_LFlow and Cov_IIGDM on FFHQ. Across both motion deblurring and inpainting, Cov_LFlow yields sharper facial structures, cleaner textures, and fewer artifacts. For motion deblurring, it better restores contours and eye details without ringing, while inpainting results show improved shading consistency and reduced boundary errors. These visual observations align with the quantitative improvements reported in Table 3.

Starting Time t_s We also investigated the impact of the starting time t_s on the results. Figure 5 illustrates the impact of varying the start time t_s for the flow process (SR task). We see that overly large t_s values tend to slightly degrade perceptual quality, whereas overly small t_s values can cause excessive smoothing or artifacts. The plot indicates that an intermediate choice of t_s (around 0.7–0.8) strikes a favorable balance, leading to consistently lower LPIPS on both FFHQ and ImageNet.

Table 3: Ablations on the effect of $\text{Cov}[\mathbf{z}_0|\mathbf{z}_t]$. **Bold** indicates the best.

Dataset	FFHQ				ImageNet			
	Deblur (G)		Inpainting		Deblur (M)		SR ($\times 4$)	
	PSNR $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	LPIPS $\downarrow$
Cov_LFlow	29.10	0.166	23.85	0.132	26.10	0.344	25.29	0.338
Cov_IIIGDM	29.04	0.179	22.69	0.151	25.22	0.363	24.80	0.351

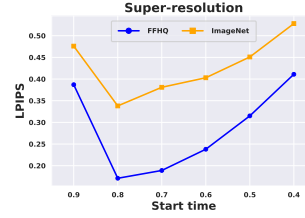


Figure 5: Ablation study on the start time t_s .

Table 4: **Average inference time** (seconds per image) for latent-based inverse solvers on ImageNet samples. Timings are measured on an NVIDIA 3090 GPU.

Method	Gaussian Deblur	SR $\times 4$
Resample [36]	550.26s	410.68s
MPGD [38]	566.58s	548.96s
PSLD [33]	705.31s	675.85s
SITCOM [60]	345.60s	328.12s
LFlow	267.76s	227.92s
LFlow (Cov_IIIGDM)	477.95s	406.02s

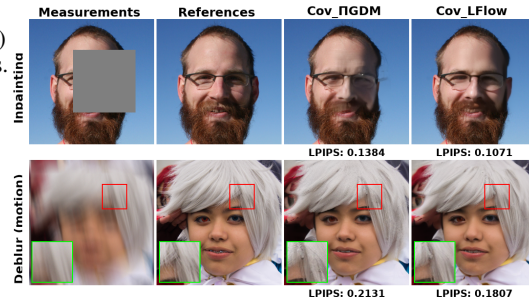


Figure 6: Visual results on the effect of $\text{Cov}[\mathbf{z}_0|\mathbf{z}_t]$.

Inference Time To highlight the practical advantages of posterior covariance modeling beyond standard evaluation metrics, we report the average inference time per image in Table 4. Despite using an adaptive ODE solver with potentially higher NFE, LFlow achieves significantly faster inference than prior latent-based solvers. This efficiency stems from faster convergence and more accurate trajectory estimation. The variant Cov_IIIGDM, which employs a less accurate covariance, converges more slowly—demonstrating the importance of proper covariance modeling. Notably, PSLD runs for 1000 iterations and Resample for 500, yet both are slower than LFlow.

6 Discussions

In this paper, we present LFlow, which efficiently addresses linear inverse problems by utilizing flow matching in the latent space of pre-trained autoencoders without additional training. Based on a justified latent *Gaussian representation assumption*, our approach introduces a theoretically sound, time-dependent latent posterior covariance that enhances gradient-based inference. Experimental results across such tasks as deblurring, super-resolution, and inpainting demonstrate that LFlow outperforms current latent diffusion models in reconstruction quality.

Limitation. One limitation of LFlow lies in its runtime: the current implementation can require approximately 3:30–10 minutes to solve inverse problems on an NVIDIA RTX 3090 GPU, with Gaussian deblurring and super-resolution averaging 2:30–6 minutes, and motion deblurring and inpainting taking around 3:30–10 minutes. While this is more efficient than existing latent diffusion solvers, as demonstrated in our ablation studies, it may still pose challenges in time-sensitive applications. Nonetheless, the improved reconstruction quality and the efficiency gains achieved during training help mitigate this drawback.

Future work. Future work will focus on optimizing the solver and investigating alternative numerical integration schemes to further reduce inference time. In addition, we plan to extend LFlow to improve its robustness under distributional shifts, enabling broader applicability to real-world scenarios and downstream tasks.

Acknowledgement

This research was partially supported by the Australian Research Council through an Industrial Transformation Training Centre for Information Resilience (IC200100022).

References

- [1] Brian B Moser, Arundhati S Shanbhag, Federico Raue, Stanislav Frolov, Sebastian Palacio, and Andreas Dengel. Diffusion models, image super-resolution, and everything: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, 2024.
- [2] Andreas Lugmayr, Andres Danelljan, Fisher Yu, Radu Timofte, and Luc Van Gool. Repaint: Inpainting using denoising diffusion probabilistic models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11461–11471, 2022.
- [3] Kaihao Zhang, Wenqi Ren, Wenhan Luo, Wei-Sheng Lai, Björn Stenger, Ming-Hsuan Yang, and Hongdong Li. Deep image deblurring: A survey. *International Journal of Computer Vision*, 130(9):2103–2130, 2022.
- [4] Peyman Milanfar and Mauricio Delbracio. Denoising: A powerful building-block for imaging, inverse problems, and machine learning. *arXiv preprint arXiv:2409.06219*, 2024.
- [5] Finbarr O’Sullivan. A statistical perspective on ill-posed inverse problems. *Statistical science*, pages 502–518, 1986.
- [6] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, pages 2256–2265. PMLR, 2015.
- [7] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.
- [8] 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*, 2021.
- [9] Giannis Daras, Hyungjin Chung, Chieh-Hsin Lai, Yuki Mitsufuji, Jong Chul Ye, Peyman Milanfar, Alexandros G Dimakis, and Mauricio Delbracio. A survey on diffusion models for inverse problems. *arXiv preprint arXiv:2410.00083*, 2024.
- [10] Hongkai Zheng, Wenda Chu, Bingliang Zhang, Zihui Wu, Austin Wang, Berthy Feng, Caifeng Zou, Yu Sun, Nikola Borislavov Kovachki, Zachary E Ross, Katherine Bouman, and Yisong Yue. Inversebench: Benchmarking plug-and-play diffusion priors for inverse problems in physical sciences. In *The Thirteenth International Conference on Learning Representations*, 2025.
- [11] Jooyoung Choi, Sungwon Kim, Yonghyun Jeong, Youngjune Gwon, and Sungroh Yoon. Ilvr: Conditioning method for denoising diffusion probabilistic models. *arXiv preprint arXiv:2108.02938*, 2021.
- [12] Ajil Jalal, Marius Arvinte, Giannis Daras, Eric Price, Alexandros G Dimakis, and Jon Tamir. Robust compressed sensing mri with deep generative priors. *Advances in Neural Information Processing Systems*, 34:14938–14954, 2021.
- [13] Bahjat Kawar, Gregory Vaksman, and Michael Elad. Snips: Solving noisy inverse problems stochastically. *Advances in Neural Information Processing Systems*, 34:21757–21769, 2021.
- [14] Yang Song, Liyue Shen, Lei Xing, and Stefano Ermon. Solving inverse problems in medical imaging with score-based generative models. In *NeurIPS 2021 Workshop on Deep Learning and Inverse Problems*, 2021.
- [15] Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. *Advances in Neural Information Processing Systems*, 35:23593–23606, 2022.
- [16] Yang Song, Liyue Shen, Lei Xing, and Stefano Ermon. Solving inverse problems in medical imaging with score-based generative models. In *International Conference on Learning Representations*, 2022.
- [17] Hyungjin Chung and Jong Chul Ye. Score-based diffusion models for accelerated mri. *Medical image analysis*, 80:102479, 2022.

- [18] Marc Anton Finzi, Anudhyan Boral, Andrew Gordon Wilson, Fei Sha, and Leonardo Zepeda-Núñez. User-defined event sampling and uncertainty quantification in diffusion models for physical dynamical systems. In *International Conference on Machine Learning*, pages 10136–10152. PMLR, 2023.
- [19] Arpit Bansal, Hong-Min Chu, Avi Schwarzschild, Soumyadip Sengupta, Micah Goldblum, Jonas Geiping, and Tom Goldstein. Universal guidance for diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 843–852, 2023.
- [20] Yuanzhi Zhu, Kai Zhang, Jingyun Liang, Bihan Cao, Radu Timofte, and Luc Van Gool. Denoising diffusion models for plug-and-play image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1219–1229, 2023.
- [21] Hyungjin Chung, Jeongsol Kim, Michael Thompson Mccann, Marc Louis Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *The Eleventh International Conference on Learning Representations*, 2023.
- [22] Jiaming Song, Arash Vahdat, Morteza Mardani, and Jan Kautz. Pseudoinverse-guided diffusion models for inverse problems. In *International Conference on Learning Representations*, 2023.
- [23] 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. Featured Certification.
- [24] Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-shot image restoration using denoising diffusion null-space model. In *The Eleventh International Conference on Learning Representations*, 2023.
- [25] Jiaming Song, Qinsheng Zhang, Hongxu Yin, Morteza Mardani, Ming-Yu Liu, Jan Kautz, Yongxin Chen, and Arash Vahdat. Loss-guided diffusion models for plug-and-play controllable generation. In *International Conference on Machine Learning*, pages 32483–32498. PMLR, 2023.
- [26] Jiwen Yu, Yinhuai Wang, Chen Zhao, Bernard Ghanem, and Jian Zhang. Freedom: Training-free energy-guided conditional diffusion model. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 23117–23127. IEEE Computer Society, 2023.
- [27] Berthy T. Feng, Jamie Smith, Michael Rubinstein, Huiwen Chang, Katherine L. Bouman, and William T. Freeman. Score-based diffusion models as principled priors for inverse imaging. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 10520–10531, October 2023.
- [28] Hyungjin Chung, Suhyeon Lee, and Jong Chul Ye. Decomposed diffusion sampler for accelerating large-scale inverse problems. In *The Twelfth International Conference on Learning Representations*, 2024.
- [29] Xinyu Peng, Ziyang Zheng, Wenrui Dai, Nuoqian Xiao, Chenglin Li, Junni Zou, and Hongkai Xiong. Improving diffusion models for inverse problems using optimal posterior covariance. In *Forty-first International Conference on Machine Learning*, 2024.
- [30] François Rozet, G r me Andry, Francois Lanusse, and Gilles Louppe. Learning diffusion priors from observations by expectation maximization. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.
- [31] Hossein Askari, Fred Roosta, and Hongfu Sun. Training-free medical image inverses via bi-level guided diffusion models. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 75–84, 2025.
- [32] Severi Rissanen, Markus Heinonen, and Arno Solin. Free hunch: Denoiser covariance estimation for diffusion models without extra costs. In *The Thirteenth International Conference on Learning Representations*, 2025.

- [33] Litu Rout, Negin Raoof, Giannis Daras, Constantine Caramanis, Alex Dimakis, and Sanjay Shakkottai. Solving linear inverse problems provably via posterior sampling with latent diffusion models. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.
- [34] Hyungjin Chung, Jong Chul Ye, Peyman Milanfar, and Mauricio Delbracio. Prompt-tuning latent diffusion models for inverse problems. In *Forty-first International Conference on Machine Learning*, 2024.
- [35] Litu Rout, Yujia Chen, Abhishek Kumar, Constantine Caramanis, Sanjay Shakkottai, and Wen-Sheng Chu. Beyond first-order tweedie: Solving inverse problems using latent diffusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9472–9481, 2024.
- [36] Bowen Song, Soo Min Kwon, Zecheng Zhang, Xinyu Hu, Qing Qu, and Liyue Shen. Solving inverse problems with latent diffusion models via hard data consistency. In *The Twelfth International Conference on Learning Representations*, 2024.
- [37] 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 Computer Vision and Pattern Recognition Conference*, pages 20895–20905, 2025.
- [38] Yutong He, Naoki Murata, Chieh-Hsin Lai, Yuhta Takida, Toshimitsu Uesaka, Dongjun Kim, Wei-Hsiang Liao, Yuki Mitsufuji, J Zico Kolter, Ruslan Salakhutdinov, and Stefano Ermon. Manifold preserving guided diffusion. In *The Twelfth International Conference on Learning Representations*, 2024.
- [39] Rayhan Zirvi, Bahareh Tolooshams, and Anima Anandkumar. Diffusion state-guided projected gradient for inverse problems. In *The Thirteenth International Conference on Learning Representations*, 2025.
- [40] Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Yael Pritch, Michael Rubinstein, and Kfir Aberman. Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 22500–22510, 2023.
- [41] Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations*, 2023.
- [42] Xingchao Liu, Chengyue Gong, and qiang liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *The Eleventh International Conference on Learning Representations*, 2023.
- [43] Neta Shaul, Ricky TQ Chen, Maximilian Nickel, Matthew Le, and Yaron Lipman. On kinetic optimal probability paths for generative models. In *International Conference on Machine Learning*, pages 30883–30907. PMLR, 2023.
- [44] Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Müller, Harry Saini, Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, et al. Scaling rectified flow transformers for high-resolution image synthesis. In *Forty-first International Conference on Machine Learning*, 2024.
- [45] Ashwini Pople, Matthew J. Muckley, Ricky T. Q. Chen, and Brian Karrer. Training-free linear image inversion via flows, 2024.
- [46] Heli Ben-Hamu, Omri Puny, Itai Gat, Brian Karrer, Uriel Singer, and Yaron Lipman. D-flow: Differentiating through flows for controlled generation. In *Forty-first International Conference on Machine Learning*, 2024.
- [47] 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 *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.

- [48] Kushagra Pandey, Ruihan Yang, and Stephan Mandt. Fast samplers for inverse problems in iterative refinement models. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.
- [49] Yici Yan, Yichi Zhang, Xiangming Meng, and Zhizhen Zhao. FIG: Flow with interpolant guidance for linear inverse problems. In *The Thirteenth International Conference on Learning Representations*, 2025.
- [50] Ségolène Tiffany Martin, Anne Gagneux, Paul Hagemann, and Gabriele Steidl. Pnp-flow: Plug-and-play image restoration with flow matching. In *The Thirteenth International Conference on Learning Representations*, 2025.
- [51] 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 *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.
- [52] Berthy Feng and Katherine Bouman. Efficient bayesian computational imaging with a surrogate score-based prior. In *NeurIPS 2023 Workshop on Deep Learning and Inverse Problems*, 2023.
- [53] Morteza Mardani, Jiaming Song, Jan Kautz, and Arash Vahdat. A variational perspective on solving inverse problems with diffusion models. In *The Twelfth International Conference on Learning Representations*, 2024.
- [54] Luhuan Wu, Brian L. Trippe, Christian A Naesseth, John Patrick Cunningham, and David Blei. Practical and asymptotically exact conditional sampling in diffusion models. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.
- [55] Brian L. Trippe, Jason Yim, Doug Tischer, David Baker, Tamara Broderick, Regina Barzilay, and Tommi S. Jaakkola. Diffusion probabilistic modeling of protein backbones in 3d for the motif-scaffolding problem. In *The Eleventh International Conference on Learning Representations*, 2023.
- [56] Gabriel Cardoso, Yazid Janati el idrissi, Sylvain Le Corff, and Eric Moulines. Monte carlo guided denoising diffusion models for bayesian linear inverse problems. In *The Twelfth International Conference on Learning Representations*, 2024.
- [57] Zehao Dou and Yang Song. Diffusion posterior sampling for linear inverse problem solving: A filtering perspective. In *The Twelfth International Conference on Learning Representations*, 2024.
- [58] Åke Björck. *Numerical methods in matrix computations*, volume 59. Springer, 2015.
- [59] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022.
- [60] 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 *International Conference on Machine Learning (ICML)*, 2025.
- [61] Ségolène Martin, Anne Gagneux, Paul Hagemann, and Gabriele Steidl. Pnp-flow: Plug-and-play image restoration with flow matching. *arXiv preprint arXiv:2410.02423*, 2024.
- [62] Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. *Advances in neural information processing systems*, 31, 2018.
- [63] Bradley Efron. Tweedie’s formula and selection bias. *Journal of the American Statistical Association*, 106(496):1602–1614, 2011.
- [64] Patrick Cattiaux and Arnaud Guillin. Semi log-concave markov diffusions. *Séminaire de probabilités XLVI*, pages 231–292, 2014.
- [65] Arash Vahdat, Karsten Kreis, and Jan Kautz. Score-based generative modeling in latent space. In A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, 2021.

- [66] Bin Dai, Ziyu Wang, and David Wipf. The usual suspects? reassessing blame for vae posterior collapse. In *International conference on machine learning*, pages 2313–2322. PMLR, 2020.
- [67] Shun-Ichi Amari. Natural gradient works efficiently in learning. *Neural computation*, 10(2):251–276, 1998.
- [68] Hyungjin Chung, Byeongsu Sim, and Jong Chul Ye. Come-closer-diffuse-faster: Accelerating conditional diffusion models for inverse problems through stochastic contraction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12413–12422, 2022.
- [69] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4401–4410, 2019.
- [70] 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*, pages 248–255. Ieee, 2009.
- [71] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proceedings of the IEEE international conference on computer vision*, pages 3730–3738, 2015.
- [72] 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*, pages 586–595, 2018.
- [73] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017.
- [74] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015.
- [75] Quan Dao, Hao Phung, Binh Nguyen, and Anh Tran. Flow matching in latent space. *arXiv preprint arXiv:2307.08698*, 2023.
- [76] William Peebles and Saining Xie. Scalable diffusion models with transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4195–4205, 2023.
- [77] Sinho Chewi and Aram-Alexandre Pooladian. An entropic generalization of caffarelli’s contraction theorem via covariance inequalities. *Comptes Rendus. Mathématique*, 361(G9):1471–1482, 2023.
- [78] Nanye Ma, Mark Goldstein, Michael S Albergo, Nicholas M Boffi, Eric Vanden-Eijnden, and Saining Xie. Sit: Exploring flow and diffusion-based generative models with scalable interpolant transformers. In *European Conference on Computer Vision*, pages 23–40. Springer, 2024.
- [79] Kai Zhang, Luc Van Gool, and Radu Timofte. Deep unfolding network for image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3217–3226, 2020.
- [80] Su Wang, Chitwan Saharia, Ceslee Montgomery, Jordi Pont-Tuset, Shai Noy, Stefano Pellegrini, Yasumasa Onoe, Sarah Laszlo, David J Fleet, Radu Soricut, et al. Imagen editor and editbench: Advancing and evaluating text-guided image inpainting. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 18359–18369, 2023.
- [81] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4713–4726, 2022.
- [82] Noam Elata, Hyungjin Chung, Jong Chul Ye, Tomer Michaeli, and Michael Elad. Inversion: Bridging supervised and zero-shot diffusion for inverse problems. *arXiv preprint arXiv:2504.01689*, 2025.

A Proofs

Lemma A.1 (Tweedie's Mean Formula). Suppose the joint distribution of $\mathbf{z}_0$ and $\mathbf{z}_t$ factors as

$$p_t(\mathbf{z}_0, \mathbf{z}_t) = p(\mathbf{z}_0) p_t(\mathbf{z}_t | \mathbf{z}_0)$$

with

$$p_t(\mathbf{z}_t | \mathbf{z}_0) = \mathcal{N}(\mathbf{z}_t | \alpha(t) \mathbf{z}_0, \sigma(t)^2 \mathbf{I}).$$

Then

$$\mathbb{E}[\mathbf{z}_0 | \mathbf{z}_t] = \frac{1}{\alpha(t)} \left(\mathbf{z}_t + \sigma(t)^2 \nabla_{\mathbf{z}_t} \log p_t(\mathbf{z}_t) \right). \quad (22)$$

Proof. Starting from the definition of the score,

$$\begin{aligned} \nabla_{\mathbf{z}_t} \log p_t(\mathbf{z}_t) &= \frac{\nabla_{\mathbf{z}_t} p_t(\mathbf{z}_t)}{p_t(\mathbf{z}_t)} = \frac{1}{p_t(\mathbf{z}_t)} \nabla_{\mathbf{z}_t} \int p_t(\mathbf{z}_0, \mathbf{z}_t) d\mathbf{z}_0 \\ &= \frac{1}{p_t(\mathbf{z}_t)} \int \nabla_{\mathbf{z}_t} (p_t(\mathbf{z}_0, \mathbf{z}_t)) d\mathbf{z}_0 = \frac{1}{p_t(\mathbf{z}_t)} \int p_t(\mathbf{z}_0, \mathbf{z}_t) \nabla_{\mathbf{z}_t} \log p_t(\mathbf{z}_0, \mathbf{z}_t) d\mathbf{z}_0 \\ &= \int p_t(\mathbf{z}_0 | \mathbf{z}_t) \nabla_{\mathbf{z}_t} \log p_t(\mathbf{z}_t | \mathbf{z}_0) d\mathbf{z}_0 = \int p_t(\mathbf{z}_0 | \mathbf{z}_t) \frac{1}{\sigma(t)^2} (\alpha(t) \mathbf{z}_0 - \mathbf{z}_t) d\mathbf{z}_0 \\ &= \frac{1}{\sigma(t)^2} (\alpha(t) \mathbb{E}[\mathbf{z}_0 | \mathbf{z}_t] - \mathbf{z}_t). \end{aligned}$$

Rearranging completes the proof. $\square$

Lemma A.2 (Tweedie's Covariance Formula). For any distribution $p(\mathbf{z}_0)$ and

$$p_t(\mathbf{z}_t | \mathbf{z}_0) = \mathcal{N}(\mathbf{z}_t | \alpha(t) \mathbf{z}_0, \sigma(t)^2 \mathbf{I}),$$

the posterior $p_t(\mathbf{z}_0 | \mathbf{z}_t)$ is also Gaussian with mean $\mathbb{E}[\mathbf{z}_0 | \mathbf{z}_t]$ and covariance $\mathbb{Cov}[\mathbf{z}_0 | \mathbf{z}_t]$. These are connected to the score function $\nabla_{\mathbf{z}_t} \log p_t(\mathbf{z}_t)$ via

$$\mathbb{Cov}[\mathbf{z}_0 | \mathbf{z}_t] = \frac{\sigma(t)^2}{\alpha(t)^2} \left(\mathbf{I} + \sigma(t)^2 \nabla_{\mathbf{z}_t}^2 \log p_t(\mathbf{z}_t) \right). \quad (23)$$

Proof. We start with the Hessian of $\log p_t(\mathbf{z}_t)$:

$$\begin{aligned} \nabla_{\mathbf{z}_t}^2 \log p_t(\mathbf{z}_t) &= \nabla_{\mathbf{z}_t} (\nabla_{\mathbf{z}_t}^\top \log p_t(\mathbf{z}_t)) = \frac{\partial}{\partial \mathbf{z}_{tj}} \left(\frac{\partial \log p_t(\mathbf{z}_t)}{\partial \mathbf{z}_{ti}} \right) \\ &= \nabla_{\mathbf{z}_t} \left(\frac{\alpha(t) \mathbb{E}[\mathbf{z}_0 | \mathbf{z}_t] - \mathbf{z}_t}{\sigma(t)^2} \right)^\top = \frac{1}{\sigma(t)^2} \nabla_{\mathbf{z}_t} (\alpha(t) \mathbb{E}[\mathbf{z}_0 | \mathbf{z}_t] - \mathbf{z}_t)^\top \\ &= \frac{\alpha(t)}{\sigma(t)^2} \nabla_{\mathbf{z}_t} (\mathbb{E}[\mathbf{z}_0 | \mathbf{z}_t]^\top) - \frac{1}{\sigma(t)^2} \mathbf{I} \\ &= \frac{\alpha(t)}{\sigma(t)^2} \int p_t(\mathbf{z}_0 | \mathbf{z}_t) \nabla_{\mathbf{z}_t} \log \left(\frac{p_t(\mathbf{z}_t | \mathbf{z}_0)}{p_t(\mathbf{z}_t)} \right) \mathbf{z}_0^\top d\mathbf{z}_0 - \frac{1}{\sigma(t)^2} \mathbf{I} \\ &= \left(\frac{\alpha(t)}{\sigma(t)^2} \right)^2 \int p_t(\mathbf{z}_0 | \mathbf{z}_t) (\mathbf{z}_0 - \mathbb{E}[\mathbf{z}_0 | \mathbf{z}_t]) \mathbf{z}_0^\top d\mathbf{z}_0 - \frac{1}{\sigma(t)^2} \mathbf{I} \\ &= \left(\frac{\alpha(t)}{\sigma(t)^2} \right)^2 \underbrace{\left(\mathbb{E}[\mathbf{z}_0 \mathbf{z}_0^\top | \mathbf{z}_t] - \mathbb{E}[\mathbf{z}_0 | \mathbf{z}_t] \mathbb{E}[\mathbf{z}_0 | \mathbf{z}_t]^\top \right)}_{\mathbb{Cov}[\mathbf{z}_0 | \mathbf{z}_t]} - \frac{1}{\sigma(t)^2} \mathbf{I}. \end{aligned}$$

Solving for $\mathbb{Cov}[\mathbf{z}_0 | \mathbf{z}_t]$ yields the stated formula. $\square$

Lemma A.3 (Connection between Posterior Mean and Vector Field). *Let $\mathbf{z}_t = \alpha(t) \mathbf{z}_0 + \sigma(t) \mathbf{z}_1$ be a one-sided interpolant, where $\mathbf{z}_0 \sim p_{\text{data}}$ and $\mathbf{z}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Then the vector field $\mathbf{v}(\mathbf{z}_t, t)$ satisfies*

$$\mathbf{v}(\mathbf{z}_t, t) = \frac{\dot{\sigma}(t)}{\sigma(t)} \mathbf{z}_t + \left(\dot{\alpha}(t) - \frac{\dot{\sigma}(t) \alpha(t)}{\sigma(t)} \right) \mathbb{E}[\mathbf{z}_0 | \mathbf{z}_t]. \quad (24)$$

Proof. Starting from the definition,

$$\mathbf{v}(\mathbf{z}_t, t) = \mathbb{E}[\dot{\alpha}(t) \mathbf{z}_0 + \dot{\sigma}(t) \mathbf{z}_1 | \mathbf{z}_t] = \dot{\alpha}(t) \mathbb{E}[\mathbf{z}_0 | \mathbf{z}_t] + \dot{\sigma}(t) \mathbb{E}[\mathbf{z}_1 | \mathbf{z}_t].$$

where $\dot{\sigma}(t)$, $\dot{\alpha}(t)$ represent the first-order time derivatives of $\sigma(t)$ and $\alpha(t)$, respectively. Noting that

$$\mathbb{E}[\mathbf{z}_1 | \mathbf{z}_t] = \frac{\mathbf{z}_t - \alpha(t) \mathbb{E}[\mathbf{z}_0 | \mathbf{z}_t]}{\sigma(t)},$$

we obtain

$$\mathbf{v}(\mathbf{z}_t, t) = \dot{\alpha}(t) \mathbb{E}[\mathbf{z}_0 | \mathbf{z}_t] + \dot{\sigma}(t) \frac{\mathbf{z}_t - \alpha(t) \mathbb{E}[\mathbf{z}_0 | \mathbf{z}_t]}{\sigma(t)} = \frac{\dot{\sigma}(t)}{\sigma(t)} \mathbf{z}_t + \left(\dot{\alpha}(t) - \frac{\dot{\sigma}(t) \alpha(t)}{\sigma(t)} \right) \mathbb{E}[\mathbf{z}_0 | \mathbf{z}_t].$$

□

Lemma A.4 (Cramér–Rao inequality). *Let $\mu(d\mathbf{x}) = \exp(-\Phi(\mathbf{x}))d\mathbf{x}$ be a probability measure on $\mathbb{R}^d$, where $\Phi : \mathbb{R}^d \rightarrow \mathbb{R}$ is twice continuously differentiable¹. Then for any $f \in C^1(\mathbb{R}^d)$,*

$$\text{Var}_\mu(f) \geq \left\langle \mathbb{E}_\mu[\nabla f], (\mathbb{E}_\mu[\nabla^2 \Phi])^{-1} \mathbb{E}_\mu[\nabla f] \right\rangle. \quad (25)$$

Proof. For comprehensive proofs of the Cramér–Rao inequality, we refer readers to [77], and the references cited therein. □

Lemma A.5 (Posterior Covariance Lower Bound via Cramér–Rao). *Let $\mathbf{z}_t = \alpha(t) \mathbf{z}_0 + \sigma(t) \mathbf{z}_1$, where $\mathbf{z}_0 \sim p(\mathbf{z}_0) \propto \exp(-\Phi(\mathbf{z}_0))$ with $\Phi \in C^2(\mathbb{R}^d)$, and $\mathbf{z}_1 \sim \mathcal{N}(0, \mathbf{I}_d)$ are independent. Suppose:*

- (i) *The likelihood is Gaussian: $p(\mathbf{z}_t | \mathbf{z}_0) = \mathcal{N}(\alpha(t) \mathbf{z}_0, \sigma(t)^2 \mathbf{I}_d)$.*
- (ii) *Φ is γ -strongly convex, i.e., $\nabla^2 \Phi(\mathbf{z}_0) \succeq \gamma \mathbf{I}_d$.*

Then the posterior satisfies:

$$\text{Cov}(\mathbf{z}_0 | \mathbf{z}_t) \succeq \left(\gamma + \frac{\alpha(t)^2}{\sigma(t)^2} \right)^{-1} \mathbf{I}_d. \quad (26)$$

Proof. Consider assumption (i). By applying Bayes' rule, the negative log-posterior is:

$$-\log p(\mathbf{z}_0 | \mathbf{z}_t) = -\log p(\mathbf{z}_t | \mathbf{z}_0) - \log p(\mathbf{z}_0) + \text{const}.$$

Taking second derivatives yields:

$$\nabla_{\mathbf{z}_0}^2 [-\log p(\mathbf{z}_0 | \mathbf{z}_t)] = \frac{\alpha(t)^2}{\sigma(t)^2} \mathbf{I}_d + \nabla^2 \Phi(\mathbf{z}_0).$$

Using assumption (ii), $\nabla^2 \Phi(\mathbf{z}_0) \succeq \gamma \mathbf{I}_d$, we obtain:

$$\nabla_{\mathbf{z}_0}^2 [-\log p(\mathbf{z}_0 | \mathbf{z}_t)] \succeq \left(\gamma + \frac{\alpha(t)^2}{\sigma(t)^2} \right) \mathbf{I}_d.$$

Finally, applying Lemma A.4, the Cramér–Rao bound gives:

$$\text{Cov}[\mathbf{z}_0 | \mathbf{z}_t] \succeq (\mathbb{E}_{\mathbf{z}_0 | \mathbf{z}_t} [\nabla_{\mathbf{z}_0}^2 (-\log p(\mathbf{z}_0 | \mathbf{z}_t))])^{-1} \succeq \left(\gamma + \frac{\alpha(t)^2}{\sigma(t)^2} \right)^{-1} \mathbf{I}_d.$$

□

¹We write $\Phi \in C^2(\mathbb{R}^d)$ to denote that Φ has continuous first and second derivatives on $\mathbb{R}^d$. Likewise, $f \in C^1(\mathbb{R}^d)$ means that f has a continuous gradient.

A.1 Proof of Proposition 4.1

Proof. The proof follows from Appendix C of [78]. We assume that the time-dependent latent variable $\mathbf{z}_t \in \mathbb{R}^d$ is generated by a known forward process of the form:

$$\mathbf{z}_t = \alpha(t) \mathbf{z}_0 + \sigma(t) \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}),$$

where $\mathbf{z}_0$ is the initial (clean) latent sample and $\alpha(t), \sigma(t)$ are scalar functions of time.

This process induces a time-indexed family of conditional distributions $\bar{p}_t(\mathbf{z} \mid \mathbf{y})$, where the expectation is taken over the joint distribution of $\mathbf{z}_0 \mid \mathbf{y}$ and the noise $\boldsymbol{\varepsilon}$.

We define the characteristic function of the marginal at time t as:

$$\hat{p}_t(\mathbf{k}) := \int e^{i\mathbf{k}^\top \mathbf{z}} \bar{p}_t(\mathbf{z} \mid \mathbf{y}) d\mathbf{z} = \tilde{\mathbb{E}} \left[e^{i\mathbf{k}^\top \mathbf{z}_t} \right], \quad (27)$$

where $\tilde{\mathbb{E}}$ denotes expectation over the conditional joint distribution of $\mathbf{z}_0 \mid \mathbf{y}$ and $\boldsymbol{\varepsilon}$.

We now differentiate this characteristic function with respect to time:

$$\partial_t \hat{p}_t(\mathbf{k}) = \partial_t \tilde{\mathbb{E}} \left[e^{i\mathbf{k}^\top \mathbf{z}_t} \right] = \tilde{\mathbb{E}} \left[\underbrace{\frac{d}{dt} e^{i\mathbf{k}^\top \mathbf{z}_t}}_{=i\mathbf{k}^\top \dot{\mathbf{z}}_t \cdot e^{i\mathbf{k}^\top \mathbf{z}_t}} \right] = \tilde{\mathbb{E}} \left[i\mathbf{k}^\top \dot{\mathbf{z}}_t \cdot e^{i\mathbf{k}^\top \mathbf{z}_t} \right].$$

Since $\mathbf{z}_t$ is a deterministic function of $\mathbf{z}_0$ and $\boldsymbol{\varepsilon}$, we may condition on $\mathbf{z}_t$ and write:

$$\begin{aligned} \partial_t \hat{p}_t(\mathbf{k}) &= \mathbb{E}_{\bar{p}_t(\mathbf{z})} \left[\tilde{\mathbb{E}} \left[i\mathbf{k}^\top \dot{\mathbf{z}}_t \mid \mathbf{z}_t = \mathbf{z} \right] \cdot e^{i\mathbf{k}^\top \mathbf{z}} \right] \\ &= i\mathbf{k}^\top \int \underbrace{\tilde{\mathbb{E}} [\dot{\mathbf{z}}_t \mid \mathbf{z}_t = \mathbf{z}]}_{:=\tilde{\mathbf{v}}_t(\mathbf{z})} \cdot e^{i\mathbf{k}^\top \mathbf{z}} \bar{p}_t(\mathbf{z} \mid \mathbf{y}) d\mathbf{z}. \end{aligned}$$

On the other hand, by the definition of the characteristic function:

$$\begin{aligned} \partial_t \hat{p}_t(\mathbf{k}) &= \partial_t \int e^{i\mathbf{k}^\top \mathbf{z}} \bar{p}_t(\mathbf{z} \mid \mathbf{y}) d\mathbf{z} \\ &= \int e^{i\mathbf{k}^\top \mathbf{z}} \partial_t \bar{p}_t(\mathbf{z} \mid \mathbf{y}) d\mathbf{z}. \end{aligned}$$

Equating the two expressions for $\partial_t \hat{p}_t(\mathbf{k})$, we obtain:

$$\int e^{i\mathbf{k}^\top \mathbf{z}} \partial_t \bar{p}_t(\mathbf{z} \mid \mathbf{y}) d\mathbf{z} = i\mathbf{k}^\top \int \tilde{\mathbf{v}}_t(\mathbf{z}) e^{i\mathbf{k}^\top \mathbf{z}} \bar{p}_t(\mathbf{z} \mid \mathbf{y}) d\mathbf{z}.$$

We now use the identity $\nabla_{\mathbf{z}} e^{i\mathbf{k}^\top \mathbf{z}} = i\mathbf{k} \cdot e^{i\mathbf{k}^\top \mathbf{z}}$ and integration by parts to move the gradient from $e^{i\mathbf{k}^\top \mathbf{z}}$ to the density term:

$$\int e^{i\mathbf{k}^\top \mathbf{z}} \partial_t \bar{p}_t(\mathbf{z} \mid \mathbf{y}) d\mathbf{z} = - \int e^{i\mathbf{k}^\top \mathbf{z}} \nabla_{\mathbf{z}} \cdot (\tilde{\mathbf{v}}_t(\mathbf{z}) \bar{p}_t(\mathbf{z} \mid \mathbf{y})) d\mathbf{z}.$$

Finally, since both expressions match for all $\mathbf{k}$, their inverse Fourier transforms must be equal. Hence, we conclude:

$$\partial_t \bar{p}_t(\mathbf{z} \mid \mathbf{y}) = -\nabla_{\mathbf{z}} \cdot (\tilde{\mathbf{v}}_t(\mathbf{z}) \bar{p}_t(\mathbf{z} \mid \mathbf{y})).$$

This is the continuity (Fokker–Planck) equation with drift field $\tilde{\mathbf{v}}_t$. Therefore, integrating the ODE:

$$\frac{d\mathbf{z}_t}{dt} = -\tilde{\mathbf{v}}_t(\mathbf{z}_t),$$

backward in time from $\mathbf{z}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, yields samples $\mathbf{z}_0 \sim \bar{p}_0(\mathbf{z} \mid \mathbf{y})$, as desired. $\square$

A.2 Conditional Vector Field

Proof. From Lemma A.3, the unconditional vector field is

$$\mathbf{v}(\mathbf{z}_t, t) = \frac{\dot{\sigma}(t)}{\sigma(t)} \mathbf{z}_t + \left(\dot{\alpha}(t) - \frac{\dot{\sigma}(t) \alpha(t)}{\sigma(t)} \right) \mathbb{E}[\mathbf{z}_0 | \mathbf{z}_t].$$

By Lemma A.1, we have

$$\mathbb{E}[\mathbf{z}_0 | \mathbf{z}_t] = \frac{1}{\alpha(t)} \left(\mathbf{z}_t + \sigma(t)^2 \nabla_{\mathbf{z}_t} \log p_t(\mathbf{z}_t) \right).$$

Substituting this in,

$$\mathbf{v}(\mathbf{z}_t, t) = \frac{\dot{\alpha}(t)}{\alpha(t)} \mathbf{z}_t + \frac{\sigma(t)}{\alpha(t)} \left(\dot{\alpha}(t) \sigma(t) - \dot{\sigma}(t) \alpha(t) \right) \nabla_{\mathbf{z}_t} \log p_t(\mathbf{z}_t).$$

Conditioning on $\mathbf{y}$, the score becomes

$$\nabla_{\mathbf{z}_t} \log p_t(\mathbf{z}_t | \mathbf{y}) = \nabla_{\mathbf{z}_t} \log p_t(\mathbf{z}_t) + \nabla_{\mathbf{z}_t} \log p_t(\mathbf{y} | \mathbf{z}_t).$$

Hence, the conditional vector field is

$$\mathbf{v}(\mathbf{z}_t, \mathbf{y}, t) = \mathbf{v}(\mathbf{z}_t, t) + \frac{\sigma(t)}{\alpha(t)} \left(\dot{\alpha}(t) \sigma(t) - \dot{\sigma}(t) \alpha(t) \right) \nabla_{\mathbf{z}_t} \log p_t(\mathbf{y} | \mathbf{z}_t).$$

Finally, choosing the linear schedule $\alpha(t) = 1 - t$ and $\sigma(t) = t$ simplifies this to

$$\mathbf{v}(\mathbf{z}_t, \mathbf{y}, t) = \mathbf{v}(\mathbf{z}_t, t) - \frac{t}{1-t} \nabla_{\mathbf{z}_t} \log p_t(\mathbf{y} | \mathbf{z}_t).$$

Approximating the unconditional velocity $\mathbf{v}(\mathbf{z}_t, t)$ by a parametric estimator $\mathbf{v}_\theta(\mathbf{z}_t, t)$ gives

$$\mathbf{v}(\mathbf{z}_t, \mathbf{y}, t) \approx \mathbf{v}_\theta(\mathbf{z}_t, t) - \frac{t}{1-t} \nabla_{\mathbf{z}_t} \log p_t(\mathbf{y} | \mathbf{z}_t).$$

□

A.3 Derivation of Posterior Mean (Eq. (14)) and Posterior Covariance (Eq. (15))

Proof. From Lemma A.3, we know

$$\mathbf{v}(\mathbf{z}_t, t) = \frac{\dot{\sigma}(t)}{\sigma(t)} \mathbf{z}_t + \left(\dot{\alpha}(t) - \frac{\dot{\sigma}(t) \alpha(t)}{\sigma(t)} \right) \mathbb{E}[\mathbf{z}_0 | \mathbf{z}_t].$$

With $\alpha(t) = 1 - t$ and $\sigma(t) = t$, one obtains

$$\mathbb{E}[\mathbf{z}_0 | \mathbf{z}_t] = \mathbf{z}_t - t \mathbf{v}(\mathbf{z}_t, t).$$

Meanwhile, from Lemmas A.1 and A.2, one can show

$$\mathbb{Cov}[\mathbf{z}_0 | \mathbf{z}_t] = \frac{\sigma(t)^2}{\alpha(t)} \nabla_{\mathbf{z}_t}^T \mathbb{E}[\mathbf{z}_0 | \mathbf{z}_t].$$

Thus,

$$\mathbb{Cov}[\mathbf{z}_0 | \mathbf{z}_t] = \frac{t^2}{1-t} \left(\mathbf{I} - t \nabla_{\mathbf{z}_t} \mathbf{v}_\theta(\mathbf{z}_t, t) \right).$$

□

A.4 Proof of Noise Conditional Score Approximation in Latent Space

Proof. We assume the likelihood and posterior over $\mathbf{x}_0$ are Gaussian:

$$p(\mathbf{y} | \mathbf{x}_0) = \mathcal{N}(\mathcal{A}\mathbf{x}_0, \sigma_y^2 \mathbf{I}_m), \quad p(\mathbf{x}_0 | \mathbf{x}_t) = \mathcal{N}(\mathbf{m}, \Sigma)$$

with:

$$\mathbf{x}_0, \mathbf{x}_t \in \mathbb{R}^n, \quad \mathbf{y} \in \mathbb{R}^m, \quad \mathcal{A} \in \mathbb{R}^{m \times n}, \quad \mathbf{m} := \mathbb{E}[\mathbf{x}_0 | \mathbf{x}_t], \quad \Sigma := \text{Cov}[\mathbf{x}_0 | \mathbf{x}_t]$$

Define:

$$\mathbf{S} := \sigma_y^2 \mathbf{I}_m + \mathcal{A} \Sigma \mathcal{A}^\top \in \mathbb{R}^{m \times m}, \quad \mathbf{r} := \mathbf{y} - \mathcal{A} \mathbf{m} \in \mathbb{R}^m$$

Then:

$$p(\mathbf{y} \mid \mathbf{x}_t) = \int p(\mathbf{y} \mid \mathbf{x}_0) p(\mathbf{x}_0 \mid \mathbf{x}_t) d\mathbf{x}_0 = \mathcal{N}(\mathcal{A} \mathbf{m}, \mathbf{S})$$

The log-likelihood becomes:

$$\log p(\mathbf{y} \mid \mathbf{x}_t) = -\frac{m}{2} \log(2\pi) - \frac{1}{2} \log \det \mathbf{S} - \frac{1}{2} \mathbf{r}^\top \mathbf{S}^{-1} \mathbf{r}$$

We compute its gradient w.r.t. $\mathbf{x}_t \in \mathbb{R}^n$ in two parts:

(1) Log-determinant term:

$$\nabla_{\mathbf{x}_t} \log \det \mathbf{S} = \text{Tr}(\mathbf{S}^{-1} \nabla_{\mathbf{x}_t} \mathbf{S}), \quad \frac{\partial \mathbf{S}}{\partial x_{t,i}} = \mathcal{A} \frac{\partial \Sigma}{\partial x_{t,i}} \mathcal{A}^\top$$

So,

$$\nabla_{\mathbf{x}_t} \log \det \mathbf{S} = \begin{bmatrix} \text{Tr}(\mathcal{A}^\top \mathbf{S}^{-1} \mathcal{A} \cdot \frac{\partial \Sigma}{\partial x_{t,1}}) \\ \vdots \\ \text{Tr}(\mathcal{A}^\top \mathbf{S}^{-1} \mathcal{A} \cdot \frac{\partial \Sigma}{\partial x_{t,n}}) \end{bmatrix} \in \mathbb{R}^n$$

(2) Quadratic term:

$$\nabla_{\mathbf{x}_t} (\mathbf{r}^\top \mathbf{S}^{-1} \mathbf{r}) = -2(\nabla_{\mathbf{x}_t} \mathbf{m})^\top \mathcal{A}^\top \mathbf{S}^{-1} \mathbf{r} - \begin{bmatrix} \text{Tr}(\mathbf{S}^{-1} \mathbf{r} \mathbf{r}^\top \mathbf{S}^{-1} \cdot \frac{\partial \mathbf{S}}{\partial x_{t,1}}) \\ \vdots \\ \text{Tr}(\mathbf{S}^{-1} \mathbf{r} \mathbf{r}^\top \mathbf{S}^{-1} \cdot \frac{\partial \mathbf{S}}{\partial x_{t,n}}) \end{bmatrix} \in \mathbb{R}^n$$

Putting both together:

$$\nabla_{\mathbf{x}_t} \log p(\mathbf{y} \mid \mathbf{x}_t) = -(\nabla_{\mathbf{x}_t} \mathbf{m})^\top \mathcal{A}^\top \mathbf{S}^{-1} \mathbf{r} - \frac{1}{2} \begin{bmatrix} \text{Tr}(\mathcal{A}^\top \mathbf{S}^{-1} \mathcal{A} \cdot \frac{\partial \Sigma}{\partial x_{t,1}}) + \text{Tr}(\mathbf{S}^{-1} \mathbf{r} \mathbf{r}^\top \mathbf{S}^{-1} \cdot \frac{\partial \mathbf{S}}{\partial x_{t,1}}) \\ \vdots \\ \text{Tr}(\mathcal{A}^\top \mathbf{S}^{-1} \mathcal{A} \cdot \frac{\partial \Sigma}{\partial x_{t,n}}) + \text{Tr}(\mathbf{S}^{-1} \mathbf{r} \mathbf{r}^\top \mathbf{S}^{-1} \cdot \frac{\partial \mathbf{S}}{\partial x_{t,n}}) \end{bmatrix}$$

Assuming $\text{Cov}[\mathbf{x}_0 \mid \mathbf{x}_t]$ is slowly varying w.r.t. $\mathbf{x}_t$, we neglect its gradient, yielding the approximation:

$$\begin{aligned} \nabla_{\mathbf{x}_t} \log p(\mathbf{y} \mid \mathbf{x}_t) &\approx (\nabla_{\mathbf{x}_t} \mathbf{m})^\top \mathcal{A}^\top \mathbf{S}^{-1} \mathbf{r} \\ &= (\nabla_{\mathbf{x}_t} \mathbb{E}[\mathbf{x}_0 \mid \mathbf{x}_t])^\top \mathcal{A}^\top (\sigma_y^2 \mathbf{I}_m + \mathcal{A} \Sigma \mathcal{A}^\top)^{-1} (\mathbf{y} - \mathcal{A} \mathbb{E}[\mathbf{x}_0 \mid \mathbf{x}_t]) \end{aligned}$$

Now, we aim to derive a principled extension of this formula to the *latent space*. Let $\mathbf{z}_0$ and $\mathbf{z}_t$ be latent variables, and $\mathcal{D}_\varphi$ be a (generally nonlinear) decoder such that $\mathbf{x}_0 = \mathcal{D}_\varphi(\mathbf{z}_0)$. Then the likelihood can be written as:

$$p(\mathbf{y} \mid \mathbf{z}_0) = \mathcal{N}(\mathbf{y}; \mathcal{A} \mathcal{D}_\varphi(\mathbf{z}_0), \sigma_y^2 \mathbf{I})$$

which is a Gaussian with **nonlinear mean**. We further assume:

$$p(\mathbf{z}_0 \mid \mathbf{z}_t) = \mathcal{N}(\bar{\mathbf{z}}_0, \Sigma_z), \quad \text{where } \bar{\mathbf{z}}_0 := \mathbb{E}[\mathbf{z}_0 \mid \mathbf{z}_t], \quad \Sigma_z := \text{Cov}[\mathbf{x}_0 \mid \mathbf{z}_t]$$

Since $\mathcal{D}_\varphi$ is nonlinear, $p(\mathbf{y} \mid \mathbf{z}_t)$ is not Gaussian. However, we approximate $\mathcal{D}_\varphi(\mathbf{z}_0)$ via a first-order Taylor expansion around $\bar{\mathbf{z}}_0$:

$$\mathcal{D}_\varphi(\mathbf{z}_0) \approx \mathcal{D}_\varphi(\bar{\mathbf{z}}_0) + J_{\mathcal{D}}(\bar{\mathbf{z}}_0)(\mathbf{z}_0 - \bar{\mathbf{z}}_0)$$

where $J_{\mathcal{D}}(\bar{\mathbf{z}}_0)$ is the Jacobian of $\mathcal{D}_\varphi$ at $\bar{\mathbf{z}}_0$.

Using this approximation, the distribution over $\mathbf{x}_0$ becomes approximately Gaussian:

$$\mathbb{E}[\mathbf{x}_0 | \mathbf{z}_t] \approx \mathcal{D}_\varphi(\bar{\mathbf{z}}_0), \quad \text{Cov}[\mathbf{x}_0 | \mathbf{z}_t] \approx J_{\mathcal{D}}(\bar{\mathbf{z}}_0) \boldsymbol{\Sigma}_z J_{\mathcal{D}}(\bar{\mathbf{z}}_0)^\top$$

We now apply the same posterior score approximation used in pixel space, but using the decoded mean and propagated covariance:

$$\nabla_{\mathbf{z}_t} \log p(\mathbf{y} | \mathbf{z}_t) \approx (\nabla_{\mathbf{z}_t} \bar{\mathbf{z}}_0)^\top J_{\mathcal{D}}(\bar{\mathbf{z}}_0)^\top \mathcal{A}^\top (\sigma_y^2 \mathbf{I} + \mathcal{A} J_{\mathcal{D}}(\bar{\mathbf{z}}_0) \boldsymbol{\Sigma}_z J_{\mathcal{D}}(\bar{\mathbf{z}}_0)^\top \mathcal{A}^\top)^{-1} (\mathbf{y} - \mathcal{A} \mathcal{D}_\varphi(\bar{\mathbf{z}}_0))$$

This expression is now a well-defined and justified approximation to the posterior score in latent space, based on first-order decoder linearization and Gaussian propagation. $\square$

A.5 Proof of Proposition 4.3

Proof. We aim to compute the Jacobian $\nabla_{\mathbf{z}_t} \mathbf{v}(\mathbf{z}_t, t)$. From Lemma A.3, we have:

$$\mathbf{v}(\mathbf{z}_t, t) = \frac{\dot{\sigma}(t)}{\sigma(t)} \mathbf{z}_t + \left(\dot{\alpha}(t) - \frac{\dot{\sigma}(t)\alpha(t)}{\sigma(t)} \right) \mathbb{E}[\mathbf{z}_0 | \mathbf{z}_t].$$

Taking the gradient with respect to $\mathbf{z}_t$, we obtain:

$$\nabla_{\mathbf{z}_t} \mathbf{v}(\mathbf{z}_t, t) = \frac{\dot{\sigma}(t)}{\sigma(t)} \mathbf{I} + \left(\dot{\alpha}(t) - \frac{\dot{\sigma}(t)\alpha(t)}{\sigma(t)} \right) \nabla_{\mathbf{z}_t} \mathbb{E}[\mathbf{z}_0 | \mathbf{z}_t].$$

From Lemma A.1, the posterior mean is given by:

$$\mathbb{E}[\mathbf{z}_0 | \mathbf{z}_t] = \frac{1}{\alpha(t)} (\mathbf{z}_t + \sigma(t)^2 \nabla_{\mathbf{z}_t} \log p_t(\mathbf{z}_t)),$$

and therefore its gradient is:

$$\nabla_{\mathbf{z}_t} \mathbb{E}[\mathbf{z}_0 | \mathbf{z}_t] = \frac{1}{\alpha(t)} (\mathbf{I} + \sigma(t)^2 \nabla_{\mathbf{z}_t}^2 \log p_t(\mathbf{z}_t)).$$

Substituting this into the expression for the Jacobian, we get:

$$\nabla_{\mathbf{z}_t} \mathbf{v}(\mathbf{z}_t, t) = \frac{\dot{\sigma}(t)}{\sigma(t)} \mathbf{I} + \left(\dot{\alpha}(t) - \frac{\dot{\sigma}(t)\alpha(t)}{\sigma(t)} \right) \cdot \frac{1}{\alpha(t)} (\mathbf{I} + \sigma(t)^2 \nabla_{\mathbf{z}_t}^2 \log p_t(\mathbf{z}_t)).$$

Applying Lemma A.2, the Hessian of the log-density can be expressed as:

$$\sigma(t)^2 \nabla_{\mathbf{z}_t}^2 \log p_t(\mathbf{z}_t) = \frac{\alpha(t)^2}{\sigma(t)^2} \text{Cov}[\mathbf{z}_0 | \mathbf{z}_t] - \mathbf{I},$$

which implies:

$$\mathbf{I} + \sigma(t)^2 \nabla_{\mathbf{z}_t}^2 \log p_t(\mathbf{z}_t) = \frac{\alpha(t)^2}{\sigma(t)^2} \text{Cov}[\mathbf{z}_0 | \mathbf{z}_t].$$

Therefore, the Jacobian becomes:

$$\nabla_{\mathbf{z}_t} \mathbf{v}(\mathbf{z}_t, t) = \frac{\dot{\sigma}(t)}{\sigma(t)} \mathbf{I} + \left(\dot{\alpha}(t) - \frac{\dot{\sigma}(t)\alpha(t)}{\sigma(t)} \right) \cdot \frac{\alpha(t)}{\sigma(t)^2} \text{Cov}[\mathbf{z}_0 | \mathbf{z}_t].$$

By Lemma A.4, if $p(\mathbf{z}_0)$ is γ -semi-log-convex, then the Cramér–Rao inequality yields:

$$\text{Cov}[\mathbf{z}_0 | \mathbf{z}_t] \succeq \left(\gamma + \frac{\alpha(t)^2}{\sigma(t)^2} \right)^{-1} \mathbf{I}_d.$$

Substituting this bound into the Jacobian expression gives:

$$\nabla_{\mathbf{z}_t} \mathbf{v}(\mathbf{z}_t, t) \succeq \left(\frac{\dot{\sigma}(t)}{\sigma(t)} + \left(\dot{\alpha}(t) - \frac{\dot{\sigma}(t)\alpha(t)}{\sigma(t)} \right) \cdot \frac{\alpha(t)}{\sigma(t)^2} \cdot \left(\gamma + \frac{\alpha(t)^2}{\sigma(t)^2} \right)^{-1} \right) \mathbf{I}_d. \quad (28)$$

This expression simplifies to:

$$\nabla_{\mathbf{z}_t} \mathbf{v}(\mathbf{z}_t, t) \succeq \left(\frac{\gamma \sigma(t) \dot{\sigma}(t) + \alpha(t) \dot{\alpha}(t)}{\gamma \sigma(t)^2 + \alpha(t)^2} \right) \mathbf{I}_d. \quad (29)$$

Finally, observe that the right-hand side can be written as:

$$\eta(t) := \frac{d}{dt} \left(\frac{1}{2} \log (\alpha(t)^2 + \gamma \sigma(t)^2) \right) = \frac{\gamma \sigma(t) \dot{\sigma}(t) + \alpha(t) \dot{\alpha}(t)}{\gamma \sigma(t)^2 + \alpha(t)^2},$$

$$\nabla_{\mathbf{z}_t} \mathbf{v}(\mathbf{z}_t, t) \succeq \eta(t) \mathbf{I}_d.$$

Now, assume the posterior covariance is positive semidefinite:

$$\mathbb{Cov}[\mathbf{z}_0 \mid \mathbf{z}_t] \succeq 0.$$

Define the scalar coefficient

$$c(t) := \left(\dot{\alpha}(t) - \frac{\dot{\sigma}(t) \alpha(t)}{\sigma(t)} \right) \cdot \frac{\alpha(t)}{\sigma(t)^2}.$$

Then the Jacobian simplifies to

$$\nabla_{\mathbf{z}_t} \mathbf{v}(\mathbf{z}_t, t) = \frac{\dot{\sigma}(t)}{\sigma(t)} \cdot \mathbf{I} + c(t) \cdot \mathbb{Cov}[\mathbf{z}_0 \mid \mathbf{z}_t].$$

Since $\mathbb{Cov}[\mathbf{z}_0 \mid \mathbf{z}_t] \succeq 0$, the sign of $c(t)$ determines the direction of the inequality:

- If $c(t) \geq 0$, then $c(t) \cdot \mathbb{Cov}[\mathbf{z}_0 \mid \mathbf{z}_t] \succeq 0$, and

$$\nabla_{\mathbf{z}_t} \mathbf{v}(\mathbf{z}_t, t) \succeq \frac{\dot{\sigma}(t)}{\sigma(t)} \cdot \mathbf{I}.$$

- If $c(t) < 0$, then $c(t) \cdot \mathbb{Cov}[\mathbf{z}_0 \mid \mathbf{z}_t] \preceq 0$, and

$$\nabla_{\mathbf{z}_t} \mathbf{v}(\mathbf{z}_t, t) \preceq \frac{\dot{\sigma}(t)}{\sigma(t)} \cdot \mathbf{I}.$$

In particular, for the common case where $\alpha(t) = 1 - t$ and $\sigma(t) = t$, we have $c(t) < 0$ for all $t \in (0, 1]$, and thus the Jacobian is upper bounded. This completes the proof. $\square$

A.6 Proof of Proposition 4.4

Proposition A.6 (Optimal Vector Field). *Let $\mathbf{z}_0 \sim \mathcal{N}(\mathbf{0}, \sigma_{latr}^2 \mathbf{I})$ and $\mathbf{z}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ be independent. Define $\mathbf{z}_t = (1 - t) \mathbf{z}_0 + t \mathbf{z}_1$ for $t \in [0, 1]$. The optimal vector field $\mathbf{v}^*(\mathbf{z}_t, t)$ that minimizes*

$$\arg \min_{\mathbf{v}} \mathbb{E} [\|\mathbf{v}(\mathbf{z}_t, t) - (\mathbf{z}_0 - \mathbf{z}_1)\|^2]$$

is

$$\mathbf{v}^*(\mathbf{z}_t, t) = \frac{(1 - t) \sigma_{latr}^2 - t}{(1 - t)^2 \sigma_{latr}^2 + t^2} \mathbf{z}_t.$$

Proof. Since the loss function is quadratic and $\mathbf{v}^*(\mathbf{z}_t, t)$ depends only on $\mathbf{z}_t$ and t , the optimal vector field is the conditional expectation:

$$\mathbf{v}^*(\mathbf{z}_t, t) = \mathbb{E} [\mathbf{z}_0 - \mathbf{z}_1 \mid \mathbf{z}_t].$$

Given that $\mathbf{z}_0$ and $\mathbf{z}_1$ are independent Gaussian random variables, and $\mathbf{z}_t$ is a linear combination of $\mathbf{z}_0$ and $\mathbf{z}_1$, the joint distribution of $\mathbf{z}_0$, $\mathbf{z}_1$, and $\mathbf{z}_t$ is multivariate Gaussian. We will compute $\mathbb{E} [\mathbf{z}_0 \mid \mathbf{z}_t]$ and $\mathbb{E} [\mathbf{z}_1 \mid \mathbf{z}_t]$ using the properties of multivariate normal distributions.

First, we identify the covariance matrices:

$$\Sigma_{\mathbf{z}_0 \mathbf{z}_0} = \sigma_{latr}^2 \mathbf{I}, \quad \Sigma_{\mathbf{z}_1 \mathbf{z}_1} = \mathbf{I}, \quad \Sigma_{\mathbf{z}_0 \mathbf{z}_1} = \mathbf{0},$$

since $\mathbf{z}_0$ and $\mathbf{z}_1$ are independent.

Next, compute the covariance between $\mathbf{z}_0$ and $\mathbf{z}_t$:

$$\Sigma_{\mathbf{z}_0 \mathbf{z}_t} = \mathbb{E} [\mathbf{z}_0 \mathbf{z}_t^\top] = (1-t) \mathbb{E} [\mathbf{z}_0 \mathbf{z}_0^\top] + t \mathbb{E} [\mathbf{z}_0 \mathbf{z}_1^\top] = (1-t) \sigma_{latr}^2 \mathbf{I}.$$

Similarly, the covariance between $\mathbf{z}_1$ and $\mathbf{z}_t$:

$$\Sigma_{\mathbf{z}_1 \mathbf{z}_t} = \mathbb{E} [\mathbf{z}_1 \mathbf{z}_t^\top] = (1-t) \mathbb{E} [\mathbf{z}_1 \mathbf{z}_0^\top] + t \mathbb{E} [\mathbf{z}_1 \mathbf{z}_1^\top] = t \mathbf{I}.$$

The variance of $\mathbf{z}_t$ is:

$$\Sigma_{\mathbf{z}_t \mathbf{z}_t} = (1-t)^2 \Sigma_{\mathbf{z}_0 \mathbf{z}_0} + t^2 \Sigma_{\mathbf{z}_1 \mathbf{z}_1} = (1-t)^2 \sigma_{latr}^2 \mathbf{I} + t^2 \mathbf{I} = ((1-t)^2 \sigma_{latr}^2 + t^2) \mathbf{I} = s^2 \mathbf{I},$$

where $s^2 = (1-t)^2 \sigma_{latr}^2 + t^2$.

The joint covariance matrix of $\mathbf{z}_0$, $\mathbf{z}_1$, and $\mathbf{z}_t$ is:

$$\Sigma_{\mathbf{w}} = \begin{bmatrix} \Sigma_{\mathbf{z}_0 \mathbf{z}_0} & \Sigma_{\mathbf{z}_0 \mathbf{z}_1} & \Sigma_{\mathbf{z}_0 \mathbf{z}_t} \\ \Sigma_{\mathbf{z}_1 \mathbf{z}_0} & \Sigma_{\mathbf{z}_1 \mathbf{z}_1} & \Sigma_{\mathbf{z}_1 \mathbf{z}_t} \\ \Sigma_{\mathbf{z}_t \mathbf{z}_0} & \Sigma_{\mathbf{z}_t \mathbf{z}_1} & \Sigma_{\mathbf{z}_t \mathbf{z}_t} \end{bmatrix} = \begin{bmatrix} \sigma_{latr}^2 \mathbf{I} & \mathbf{0} & (1-t) \sigma_{latr}^2 \mathbf{I} \\ \mathbf{0} & \mathbf{I} & t \mathbf{I} \\ (1-t) \sigma_{latr}^2 \mathbf{I} & t \mathbf{I} & s^2 \mathbf{I} \end{bmatrix}.$$

Using the properties of conditional expectations for multivariate normals, we compute the conditional expectations:

$$\begin{aligned} \mathbb{E} [\mathbf{z}_0 | \mathbf{z}_t] &= \Sigma_{\mathbf{z}_0 \mathbf{z}_t} (\Sigma_{\mathbf{z}_t \mathbf{z}_t})^{-1} \mathbf{z}_t = \frac{(1-t) \sigma_{latr}^2}{s^2} \mathbf{z}_t, \\ \mathbb{E} [\mathbf{z}_1 | \mathbf{z}_t] &= \Sigma_{\mathbf{z}_1 \mathbf{z}_t} (\Sigma_{\mathbf{z}_t \mathbf{z}_t})^{-1} \mathbf{z}_t = \frac{t}{s^2} \mathbf{z}_t. \end{aligned}$$

Therefore, the optimal vector field is:

$$\mathbf{v}^*(\mathbf{z}_t, t) = \mathbb{E} [\mathbf{z}_0 - \mathbf{z}_1 | \mathbf{z}_t] = \mathbb{E} [\mathbf{z}_0 | \mathbf{z}_t] - \mathbb{E} [\mathbf{z}_1 | \mathbf{z}_t] = \left(\frac{(1-t) \sigma_{latr}^2}{s^2} - \frac{t}{s^2} \right) \mathbf{z}_t = \frac{(1-t) \sigma_{latr}^2 - t}{(1-t)^2 \sigma_{latr}^2 + t^2} \mathbf{z}_t.$$

□

A.7 Proof of Covariance in Π GDM

(Π GDM [22]). Let $\mathbf{x}_0 \sim \mathcal{N}(\mathbf{0}, \sigma_{data}^2 \mathbf{I})$, and consider the forward process

$$\mathbf{x}_t = \alpha(t) \mathbf{x}_0 + \sigma(t) \mathbf{x}_1, \quad \mathbf{x}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{I}).$$

Then the conditional covariance $\text{Cov}[\mathbf{x}_0 | \mathbf{x}_t]$ is

$$\text{Cov}[\mathbf{x}_0 | \mathbf{x}_t] = \frac{\sigma_{data}^2 \sigma(t)^2}{\alpha(t)^2 \sigma_{data}^2 + \sigma(t)^2} \mathbf{I}. \quad (30)$$

Proof. Again $\mathbf{x}_0$ and $\mathbf{x}_1$ are independent Gaussians, and $\mathbf{x}_t$ is a linear combination. Thus $\mathbf{x}_0, \mathbf{x}_t$ are jointly Gaussian with

$$\text{Cov}[\mathbf{x}_t] = \alpha(t)^2 \sigma_{data}^2 \mathbf{I} + \sigma(t)^2 \mathbf{I},$$

and

$$\text{Cov}[\mathbf{x}_t, \mathbf{x}_0] = \mathbb{E} [(\mathbf{x}_t - \mathbb{E}[\mathbf{x}_t])(\mathbf{x}_0 - \mathbb{E}[\mathbf{x}_0])^\top] = \alpha(t) \text{Cov}[\mathbf{x}_0] = \alpha(t) \sigma_{data}^2 \mathbf{I}.$$

Using the standard formula for conditional covariances in a Gaussian,

$$\text{Cov}[\mathbf{x}_0 | \mathbf{x}_t] = \text{Cov}[\mathbf{x}_0] - \text{Cov}[\mathbf{x}_0, \mathbf{x}_t] \text{Cov}[\mathbf{x}_t]^{-1} \text{Cov}[\mathbf{x}_t, \mathbf{x}_0].$$

Since $\text{Cov}[\mathbf{x}_0] = \sigma_{data}^2 \mathbf{I}$, one obtains

$$\text{Cov}[\mathbf{x}_0 | \mathbf{x}_t] = \sigma_{data}^2 \mathbf{I} - (\alpha(t) \sigma_{data}^2 \mathbf{I}) (\alpha(t)^2 \sigma_{data}^2 + \sigma(t)^2)^{-1} (\alpha(t) \sigma_{data}^2 \mathbf{I}),$$

which simplifies to

$$\text{Cov}[\mathbf{x}_0 | \mathbf{x}_t] = \frac{\sigma_{data}^2 \sigma(t)^2}{\alpha(t)^2 \sigma_{data}^2 + \sigma(t)^2} \mathbf{I}.$$

□

B Closed-form Solutions for computing vector $\mathbf{v}$ in Eq. (13)

In this section, we derive efficient closed-form expressions for computing the vector $\mathbf{v}$ under the assumption of isotropic posterior covariance, i.e. $\mathbb{C}\text{ov}[\mathbf{z}_0 | \mathbf{z}_t] = r(t)^2 \mathbf{I}$. We begin by introducing essential notation.

Notations.

- Let $\mathbf{m} \in \{0, 1\}^{d \times 1}$ represent the sampling positions in an image or signal.
- The downsampling operator associated with $\mathbf{m}$ is $\mathbf{D}_{\mathbf{m}} \in \{0, 1\}^{\|\mathbf{m}\|_0 \times d}$. It selects only those rows (i.e. entries) of a vector or matrix corresponding to the non-zero entries of $\mathbf{m}$. For example, s -fold downsampling with evenly spaced ones is denoted $\mathbf{D}_{\downarrow s}$.
- $\mathbf{D}_{\downarrow s}$ represents a *distinct block* downsampler, which averages s blocks (each of size d/s) from a vector.
- $\mathcal{F}$ is the (unitary) Fourier transform matrix of dimension $d \times d$, and $\mathcal{F}_{\downarrow s}$ is the analogous transform matrix for signals of dimension d/s .
- $\hat{\mathbf{v}}$ denotes the Fourier transform of the vector $\mathbf{v}$, and $\bar{\mathbf{v}}$ denotes its complex conjugate.
- The notation $\odot$ refers to element-wise (Hadamard) multiplication. Divisions such as $'/'$ or $'\div'$ also apply element-wise when the vectors/matrices match in dimension.

Lemma B.1 (Downsampling Equivalence). *Standard s -fold downsampling in the spatial domain is equivalent to s -fold block downsampling in the frequency domain. Concretely,*

$$\mathbf{D}_{\downarrow s} = \mathcal{F}_{\downarrow s} \mathbf{D}_{\downarrow s} \mathcal{F}^{-1}.$$

Proof. Please see [79, 29] for details. □

B.1 Image Inpainting

The observation model for image inpainting can be written as

$$\mathbf{y} = \underbrace{\mathbf{D}_{\mathbf{m}}}_{=\mathcal{A}} \mathbf{x}_0 + \mathbf{n}, \quad (31)$$

where $\mathbf{n}$ is noise. A convenient *zero-filling* version of $\mathbf{y}$ can be defined as

$$\tilde{\mathbf{y}} = \mathbf{D}_{\mathbf{m}}^{\top} \mathbf{y} = \mathbf{m} \odot (\mathbf{x}_0 + \tilde{\mathbf{n}}), \quad \tilde{\mathbf{n}} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}).$$

The closed-form solution for $\mathbf{v}$ in image inpainting is then

$$\mathbf{v} = \frac{\tilde{\mathbf{y}} - (\mathbf{m} \odot \mathcal{D}_{\varphi}(\mathbb{E}[\mathbf{z}_0 | \mathbf{z}_t]))}{\sigma_{\mathbf{y}}^2 + r(t)^2}.$$

Proof. Starting with the more general form,

$$\mathbf{v} = \mathbf{D}_{\mathbf{m}}^{\top} (\sigma_{\mathbf{y}}^2 \mathbf{I} + r(t)^2 \mathbf{D}_{\mathbf{m}} \mathbf{D}_{\mathbf{m}}^{\top})^{-1} (\mathbf{y} - \mathbf{D}_{\mathbf{m}} \mathcal{D}_{\varphi}(\mathbb{E}[\mathbf{z}_0 | \mathbf{z}_t])).$$

Since $\mathbf{D}_{\mathbf{m}} \mathbf{D}_{\mathbf{m}}^{\top} = \mathbf{I}$ on the support of $\mathbf{y}$ and $(\sigma_{\mathbf{y}}^2 \mathbf{I} + r(t)^2 \mathbf{I})^{-1} = 1/(\sigma_{\mathbf{y}}^2 + r(t)^2)$, it simplifies to

$$\mathbf{v} = \frac{\mathbf{D}_{\mathbf{m}}^{\top} (\mathbf{y} - \mathbf{D}_{\mathbf{m}} \mathcal{D}_{\varphi}(\mathbb{E}[\mathbf{z}_0 | \mathbf{z}_t]))}{\sigma_{\mathbf{y}}^2 + r(t)^2} = \frac{\tilde{\mathbf{y}} - \mathbf{m} \odot \mathcal{D}_{\varphi}(\mathbb{E}[\mathbf{z}_0 | \mathbf{z}_t])}{\sigma_{\mathbf{y}}^2 + r(t)^2}.$$

Recalling that $\tilde{\mathbf{y}} = \mathbf{D}_{\mathbf{m}}^{\top} \mathbf{y} = \mathbf{m} \odot (\mathbf{x}_0 + \tilde{\mathbf{n}})$, we arrive at the stated closed-form solution. □

B.2 Image Deblurring

For image deblurring, the observation model is

$$\mathbf{y} = \mathbf{x}_0 * \mathbf{k} + \mathbf{n}, \quad (32)$$

where $\mathbf{k}$ is the blurring kernel and $*$ is the circular convolution operator. Using the Fourier transform, this can be written as

$$\mathbf{y} = \underbrace{\mathcal{F}^{-1} \text{diag}(\hat{\mathbf{k}}) \mathcal{F}}_{=\mathcal{A}} \mathbf{x}_0 + \mathbf{n},$$

where $\hat{\mathbf{k}}$ is the DFT of $\mathbf{k}$. Under isotropic-covariance assumption, the closed-form solution for $\mathbf{v}$ is

$$\mathbf{v} = \mathcal{F}^{-1} \left(\hat{\mathbf{k}} \odot \frac{\mathcal{F}[\mathbf{y} - \mathcal{A} \mathcal{D}_{\varphi}(\mathbb{E}[\mathbf{z}_0 | \mathbf{z}_t])]}{\sigma_{\mathbf{y}}^2 + r(t)^2 |\hat{\mathbf{k}}|^2} \right).$$

Proof. Because $\mathcal{A}$ is a real linear operator of convolution type, we have $\mathcal{A}^T = \mathcal{A}^H$. Thus,

$$\mathbf{v} = \mathcal{A}^T \left(\sigma_{\mathbf{y}}^2 \mathbf{I} + r(t)^2 \mathcal{A} \mathcal{A}^T \right)^{-1} [\mathbf{y} - \mathcal{A} \mathcal{D}_{\varphi}(\mathbb{E}[\mathbf{z}_0 | \mathbf{z}_t])].$$

Substituting $\mathcal{A} = \mathcal{F}^{-1} \text{diag}(\hat{\mathbf{k}}) \mathcal{F}$ and simplifying in the Fourier domain (using the diagonal structure in frequency space), one obtains

$$\mathbf{v} = \mathcal{F}^{-1} \left(\hat{\mathbf{k}} \odot \frac{\mathcal{F}[\mathbf{y} - \mathcal{A} \mathcal{D}_{\varphi}(\cdot)]}{\sigma_{\mathbf{y}}^2 + r(t)^2 |\hat{\mathbf{k}}|^2} \right).$$

□

B.3 Super-Resolution

Following [79], the super-resolution observation model is approximately

$$\mathbf{y} = (\mathbf{x}_0 * \mathbf{k})_{\downarrow s} + \mathbf{n}, \quad (33)$$

which, in “canonical form,” can be written as

$$\mathbf{y} = \mathbf{D}_{\downarrow s} \mathcal{F}^{-1} \text{diag}(\hat{\mathbf{k}}) \mathcal{F} \mathbf{x}_0 + \mathbf{n}.$$

Hence, $\mathcal{A} = \mathbf{D}_{\downarrow s} \mathcal{F}^{-1} \text{diag}(\hat{\mathbf{k}}) \mathcal{F}$. The closed-form solution under the isotropic assumption is

$$\mathbf{v} = \mathcal{F}^{-1} \left(\hat{\mathbf{k}} \odot_s \frac{\mathcal{F}_{\downarrow s} [\mathbf{y} - \mathcal{A} \mathcal{D}_{\varphi}(\mathbb{E}[\mathbf{z}_0 | \mathbf{z}_t])]}{\sigma_{\mathbf{y}}^2 + r(t)^2 (\hat{\mathbf{k}} \odot \hat{\mathbf{k}})_{\downarrow s}} \right),$$

where $\odot_s$ denotes block-wise Hadamard multiplication.

Proof. Since $\mathcal{A}^T = \mathcal{A}^H$ and

$$\mathcal{A} = \mathbf{D}_{\downarrow s} \mathcal{F}^{-1} \text{diag}(\hat{\mathbf{k}}) \mathcal{F},$$

we get $\mathcal{A}^T = \mathcal{F}^T \text{diag}(\hat{\mathbf{k}}) (\mathcal{F}^{-1})^T \mathbf{D}_{\downarrow s}^T$. Applying Lemma B.1, namely $\mathbf{D}_{\downarrow s} \mathcal{F}^{-1} = \mathcal{F}_{\downarrow s}^{-1} \mathbf{D}_{\downarrow s}$, and its conjugate-transpose version, reduces the inverse $(\sigma_{\mathbf{y}}^2 \mathbf{I} + r(t)^2 \mathcal{A} \mathcal{A}^T)^{-1}$ to diagonal form in the “downsampled” Fourier domain. One obtains

$$\mathbf{v} = \mathcal{F}^{-1} \left(\hat{\mathbf{k}} \odot_s \frac{\mathcal{F}_{\downarrow s} [\mathbf{y} - \mathcal{A} \mathcal{D}_{\varphi}(\cdot)]}{\sigma_{\mathbf{y}}^2 + r(t)^2 (\hat{\mathbf{k}} \odot \hat{\mathbf{k}})_{\downarrow s}} \right).$$

□

C Implementation details

Algorithm 1 LFlow Sampling: Posterior-Guided Latent ODE Inference for Linear Inverse Problems

```

1: Input: measurements  $\mathbf{y}$ , encoder  $\mathcal{E}_\phi$ , decoder  $\mathcal{D}_\varphi$ , forward operator  $\mathcal{A}$ , pre-trained vector field  $\mathbf{v}_\theta(\cdot, t)$  for  $t \in [t_s, 0]$ ,  $t_s = 0.8$ , and  $K = 2$ .
2: Initialize.  $\mathbf{z}_{t_s} \leftarrow (1 - t_s) \mathcal{E}_\phi(\mathbf{y}) + t_s \mathbf{z}_1$ ,  $\mathbf{z}_1 \sim \mathcal{N}(0, \mathbf{I})$ 
3: for  $t = t_s$  down to 0 do
4:    $\bar{\mathbf{z}}_0 \leftarrow \mathbf{z}_t - t \mathbf{v}_\theta(\mathbf{z}_t, t)$  ▷ Posterior mean prediction Eq. (14)
5:    $r^2(t) \leftarrow \frac{t^2}{1-t} (1 - t \nabla_{\mathbf{z}_t} \mathbf{v}^*(\mathbf{z}_t, t))$  ▷ Posterior covariance estimation Eq. (18)
6:    $\nabla_{\mathbf{z}_t} \log p(\mathbf{y} | \mathbf{z}_t) \leftarrow (\nabla_{\mathbf{z}_t} \mathcal{D}_\varphi(\bar{\mathbf{z}}_0))^\top \mathcal{A}^\top \frac{(\mathbf{y} - \mathcal{A} \mathcal{D}_\varphi(\bar{\mathbf{z}}_0))}{(\sigma_{\mathbf{y}}^2 \mathbf{I} + r^2(t) \mathcal{A} \mathcal{A}^\top)}$  ▷ Eq. (13)
7:   for  $k = 1$  to  $K$  do
8:      $\mathbf{v}_\theta(\mathbf{z}_t, \mathbf{y}, t) \leftarrow \mathbf{v}_\theta(\mathbf{z}_t, t) - \frac{t}{1-t} \nabla_{\mathbf{z}_t} \log p(\mathbf{y} | \mathbf{z}_t)$  ▷  $K$ -step update of vector field Eq. (9)
9:   end for
10:   $\mathbf{z}_{t-\Delta t} \leftarrow \text{ODESolverStep}(\mathbf{z}_t, \mathbf{v}_\theta(\mathbf{z}_t, \mathbf{y}, t))$ 
11: end for
12: Return:  $\mathbf{x}_0 \leftarrow \mathcal{D}_\varphi(\mathbf{z}_0)$ 

```

C.1 LFlow

- For the solver parameters, we set the absolute and relative tolerances (atol and rtol) to 10^{-3} for inpainting and motion deblurring tasks, and to 10^{-5} for Gaussian deblurring and super-resolution tasks.
- We set the hyperparameters to $K = 2$ and $t_s = 0.8$ for all tasks. These values were selected via ablation on validation performance and were found to balance guidance strength and reconstruction fidelity across tasks.
- To ensure dimensional compatibility, the measurement $\mathbf{y}$ is upsampled (e.g., via bicubic interpolation) to match the input size of the latent encoder. For consistency, we adopt the same super-resolution and deblurring operators as in [29] across both our method and the relevant baselines.
- For inpainting tasks, we incorporate the strategy proposed in the PSLD method [33] into our LFlow algorithm. This strategy reconstructs missing regions that align seamlessly with the known parts of the image, expressed as $\mathbf{x}_0 = \mathcal{A}^T \mathcal{A} \mathbf{x}_0 + (\mathbf{I} - \mathcal{A}^T \mathcal{A}) \mathcal{D}_\varphi(\mathbf{z}_0)$. Unlike the DPS sampler, which generates the entire image and may lead to inconsistencies with the observed data, this approach ensures that observations are directly applied to the corresponding parts of the generated image, leaving unmasked areas unchanged [80]. For other tasks, such as motion deblurring, Gaussian deblurring, and super-resolution, this extra step is unnecessary since no box inpainting is involved, i.e., $\mathbf{x}_0 = \mathcal{D}_\varphi(\mathbf{z}_0)$.

C.2 Comparison methods

PSLD [33] applies an orthogonal projection onto the subspace of $\mathcal{A}$ between decoding and encoding to enforce fidelity:

$$\begin{aligned}
 \mathbf{z}_{t-1} = & \text{DDIM}(\mathbf{z}_t) - \rho \nabla_{\mathbf{z}_t} \left(\|\mathbf{y} - \mathcal{A} \mathcal{D}_\varphi(\mathbb{E}[\mathbf{z}_0 | \mathbf{z}_t])\|_2^2 + \gamma \|\mathbb{E}[\mathbf{z}_0 | \mathbf{z}_t] - \mathcal{E}_\phi(\mathcal{D}_\varphi(\mathbb{E}[\mathbf{z}_0 | \mathbf{z}_t]))\|_2^2 \right. \\
 & \left. - \mathcal{E}_\phi(\mathcal{A}^\top \mathbf{y} + (\mathbf{I} - \mathcal{A}^\top \mathcal{A}) \mathcal{D}_\varphi(\mathbb{E}[\mathbf{z}_0 | \mathbf{z}_t])) \right). \quad (34)
 \end{aligned}$$

We use a fixed step size of ρ and select γ as recommended in [36]. For our experiments, we rely on the official PSLD implementation² with its default configurations. Specifically, we conduct ImageNet experiments using Stable Diffusion v1.5, which is generally considered more robust compared to the LDM-VQ4 models.

PSLD aims to ensure that latent variables remain close to the natural manifold by enforcing them to be fixed points after autoencoding. While this approach seems to be theoretically justified, it has proven empirically ineffective [34].

²<https://github.com/LituRout/PSLD>

Resample [36] estimates first a clean latent prediction $\mathbf{z}_0^{\text{est}}(\mathbf{z}_{t+1})$ from the previous sample $\mathbf{z}_{t+1}$ using Tweedie’s formula, as described in Eq. (22). This prediction is then used to update the latent state via DDIM:

$$\mathbf{z}'_t = \text{DDIM}(\mathbf{z}_0^{\text{est}}(\mathbf{z}_{t+1}), \mathbf{z}_{t+1}). \quad (35)$$

The updated sample $\mathbf{z}'_t$ is then projected back to a measurement-consistent latent variable $\mathbf{z}_t^{\text{proj}}$ via:

$$\mathcal{N}\left(\mathbf{z}_t^{\text{proj}}; \frac{\sigma_t^2 \sqrt{\bar{\alpha}_t} \mathbf{z}_0^{\text{cond}} + (1 - \bar{\alpha}_t) \mathbf{z}'_t}{\sigma_t^2 + (1 - \bar{\alpha}_t)}, \frac{\sigma_t^2 (1 - \bar{\alpha}_t)}{\sigma_t^2 + (1 - \bar{\alpha}_t)} \mathbf{I}_k\right), \quad (36)$$

where $\mathbf{z}_0^{\text{cond}}$ is a latent vector that satisfies the measurement constraint, obtained by solving the following optimization problem:

$$\mathbf{z}_0^{\text{cond}} \in \arg \min_{\mathbf{z}} \frac{1}{2} \|\mathbf{y} - \mathcal{A}(\mathcal{D}_{\varphi}(\mathbf{z}))\|_2^2, \quad \text{initialized at } \mathbf{z}_0^{\text{est}}(\mathbf{z}_{t+1}). \quad (37)$$

Here, σ_t^2 is a tunable hyperparameter controlling the trade-off between the prior and the data fidelity, and $\bar{\alpha}_t$ is a predefined DDIM noise schedule parameter. For our experiments, we adopt the publicly available implementation provided by the authors³, using the pre-trained LDM-VQ4 models on FFHQ and ImageNet [59], along with their default hyperparameters and a 500-step DDIM sampler.

Resample refines latent diffusion sampling by balancing the reverse-time prior from the unconditional model with a measurement-informed likelihood centered on a consistent latent—ensuring the sample aligns with both the data manifold and observed measurements.

MPGD [38] accelerates inference by computing gradients only with respect to the clean latent estimate instead of the noisy input, thus avoiding heavy chain-rule expansions. Their gradient update in latent space is as follows:

$$\mathbf{z}_{t-1} = \text{DDIM}(\mathbf{z}_t) - \eta \nabla_{\mathbb{E}[\mathbf{z}_0 | \mathbf{z}_t]} \|\mathbf{y} - \mathcal{A}(\mathcal{D}_{\varphi}(\mathbb{E}[\mathbf{z}_0 | \mathbf{z}_t]))\|_2. \quad (38)$$

Note that in MPGD, we leveraged Stable Diffusion v1.5 as for PS LD. For more information, please refer to the GitHub repository⁴.

DAPS [37]⁵ refines latent estimates via Langevin dynamics guided by a latent prior and a measurement likelihood. The initial estimate $\mathbf{z}_0^{(0)}$ is obtained by solving a probability flow ODE from $\mathbf{z}_t$ using the latent score model. At each inner iteration $j = 0, \dots, N-1$, the latent estimate $\mathbf{z}_0^{(j)} \in \mathbb{R}^d$ is updated as follows:

$$\mathbf{z}_0^{(j+1)} = \mathbf{z}_0^{(j)} + \eta_t \left(\nabla_{\mathbf{z}_0^{(j)}} \log p(\mathbf{z}_0^{(j)} | \mathbf{z}_t) + \nabla_{\mathbf{z}_0^{(j)}} \log p(\mathbf{y} | \mathbf{z}_0^{(j)}) \right) + \sqrt{2\eta_t} \epsilon_j, \quad \epsilon_j \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \quad (39)$$

Here, η_t denotes the Langevin step size. The first term reflects prior guidance via the latent score model, while the second enforces consistency with the measurement through decoding and evaluating the likelihood.

DAPS avoids the limitations of local Markovian updates in diffusion models by decoupling time steps and directly sampling each noisy state $\mathbf{z}_t$ from the marginal posterior $p(\mathbf{z}_t | \mathbf{y})$. It performs posterior sampling by alternating between posterior-guided denoising via MCMC and noise re-injection, enabling large global corrections and improved inference in nonlinear inverse problems.

SITCOM (Step-wise Triple-Consistent Sampling) [60] enforces three complementary consistency conditions—measurement, forward diffusion, and step-wise backward diffusion—allowing diffusion trajectories to remain measurement-consistent with fewer reverse steps. By optimizing the input of a pre-trained diffusion model at each step, SITCOM ensures triple consistency across the data manifold, measurement space, and diffusion process, leading to efficient inverse problem solving.

At each step t , SITCOM enforces three consistencies:

³<https://github.com/soominkwon/resample>

⁴https://github.com/KellyYutongHe/mpgd_pytorch/

⁵<https://github.com/zhangbingliang2019/DAPS>

(S1) Measurement-consistent optimization:

$$\hat{\mathbf{v}}_t := \arg \min_{\mathbf{v}'_t} \underbrace{\left\| \mathcal{A} \left(\underbrace{\frac{1}{\sqrt{\alpha_t}} [\mathbf{v}'_t - \sqrt{1 - \bar{\alpha}_t} \epsilon_\theta(\mathbf{v}'_t, t)]}_{\mathbf{C}_2} \right) - \mathbf{y} \right\|_2^2}_{\mathbf{C}_1} + \lambda \underbrace{\|\mathbf{z}_t - \mathbf{v}'_t\|_2^2}_{\mathbf{C}_3}. \quad (40)$$

(S2) Denoising:

$$\hat{\mathbf{z}}'_0 = \frac{1}{\sqrt{\alpha_t}} [\hat{\mathbf{v}}_t - \sqrt{1 - \bar{\alpha}_t} \epsilon_\theta(\hat{\mathbf{v}}_t, t)], \quad (41)$$

(S3) Sampling (Forward step):

$$\mathbf{z}_{t-1} = \sqrt{\bar{\alpha}_{t-1}} \hat{\mathbf{z}}'_0 + \sqrt{1 - \bar{\alpha}_{t-1}} \epsilon. \quad (42)$$

Together, these steps enforce **C1**: measurement, **C2**: backward trajectory, and **C3**: forward diffusion consistency. For implementation details and hyperparameters, we rely on the official GitHub repository⁶.

SITCOM nudges the input to the denoiser at each diffusion step so its denoised output matches the measurements while staying close to the current state. It then computes the clean estimate (Tweedie) and re-noises via the forward kernel to keep the next input in distribution.

DMPlug [51]⁷ views the entire reverse diffusion process $R(\cdot)$ as a deterministic function mapping seeds to objects, and solves the inverse problem by optimizing directly over the seed $\mathbf{z}$:

$$\mathbf{z}^* \in \arg \min_{\mathbf{z}} \ell(\mathbf{y}, \mathcal{A}(R(\mathbf{z}))) + \Omega(R(\mathbf{z})), \quad \mathbf{x}^* = R(\mathbf{z}^*). \quad (43)$$

Most existing DM-based methods for inverse problems interleave reverse diffusion with measurement projections in a *step-wise* manner, but this often breaks both *manifold feasibility* (staying on the data manifold $\mathcal{M}$) and *measurement feasibility* (satisfying $\{\mathbf{x} \mid \mathbf{y} = \mathcal{A}(\mathbf{x})\}$). In contrast, DMPlug is not step-wise: it preserves manifold feasibility by keeping the pretrained reverse process **intact** while promoting $\mathbf{y} \approx \mathcal{A}(\mathbf{x})$ via global optimization.

IIGDM [22] considers the following gradient update scheme

$$\mathbf{x}_{t-1} = \text{DDIM}(\mathbf{x}_t) - \eta \left((\mathbf{y} - \mathcal{A}(\mathbb{E}[\mathbf{x}_0 \mid \mathbf{x}_t]))^\top (r_t^2 \mathcal{A} \mathcal{A}^\top + \sigma_y^2 \mathbf{I})^{-1} \mathcal{A} \frac{\partial \mathbb{E}[\mathbf{x}_0 \mid \mathbf{x}_t]}{\partial \mathbf{x}_t} \right)^\top. \quad (44)$$

where η controls the step size, σ_y represents the noise level of the measurement, and r_t is the time-dependent scale for identity posterior covariance. For this method, we utilize the official, reliable code provided by [29].

OT-ODE [45] extends the gradient guidance of IIGDM [22] to ODE sampling via an optimal transport path, resulting in a variance for the identity covariance as $r^2(t) = \frac{\sigma(t)^2}{\alpha(t)^2 + \sigma(t)^2} \mathbf{I}$. Moreover, the conditional expectation $\mathbb{E}[\mathbf{x}_0 \mid \mathbf{x}_t]$ is computed from the velocity field $\mathbf{v}_\theta(\mathbf{x}_t, t)$, according to the relation in A.3. For a fair comparison, we used the same solver as LFlow, i.e., the adaptive Heun.

⁶<https://github.com/sjames40/SITCOM>

⁷<https://github.com/sun-umn/DMPlug>

D Additional Experiments and Ablations

D.1 CelebA-HQ ($256 \times 256 \times 3$).

Table 5: Quantitative results of linear inverse problem solving on **CelebA-HQ** samples of the validation dataset. **Bold** and underline indicate the best and second-best respectively. The method shaded in gray is in pixel space.

Method	Deblurring (Gauss)			Deblurring (Motion)			SR ($\times 4$)			Inpainting (Box)		
	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$
LFlow (ours)	29.06	<u>0.825</u>	0.164	30.14	0.845	0.167	28.92	0.830	0.170	24.82	0.876	0.123
SITCOM [60]	28.74	0.792	0.275	28.05	0.776	0.324	28.45	0.812	0.208	21.45	0.733	0.216
DAPS [37]	26.66	0.773	0.314	27.22	0.766	0.251	28.29	0.798	0.227	21.15	0.807	0.202
Resample [36]	28.07	0.742	<u>0.239</u>	28.37	0.804	<u>0.232</u>	<u>29.84</u>	0.806	<u>0.193</u>	19.49	0.797	0.237
PSLD [33]	<u>29.47</u>	0.833	0.310	<u>29.75</u>	<u>0.821</u>	0.313	31.65	<u>0.829</u>	<u>0.246</u>	<u>24.03</u>	<u>0.812</u>	<u>0.165</u>
MPGD [38]	29.85	0.821	0.302	29.09	0.792	0.348	29.01	0.760	0.280	23.80	0.773	0.198
OT-ODE [45]	27.83	0.789	<u>0.292</u>	26.15	0.758	0.326	28.95	0.784	0.251	22.37	0.790	0.225
C-ITGFM [48]	28.26	0.801	<u>0.280</u>	27.18	0.743	0.335	29.52	0.805	0.226	22.84	0.798	0.219

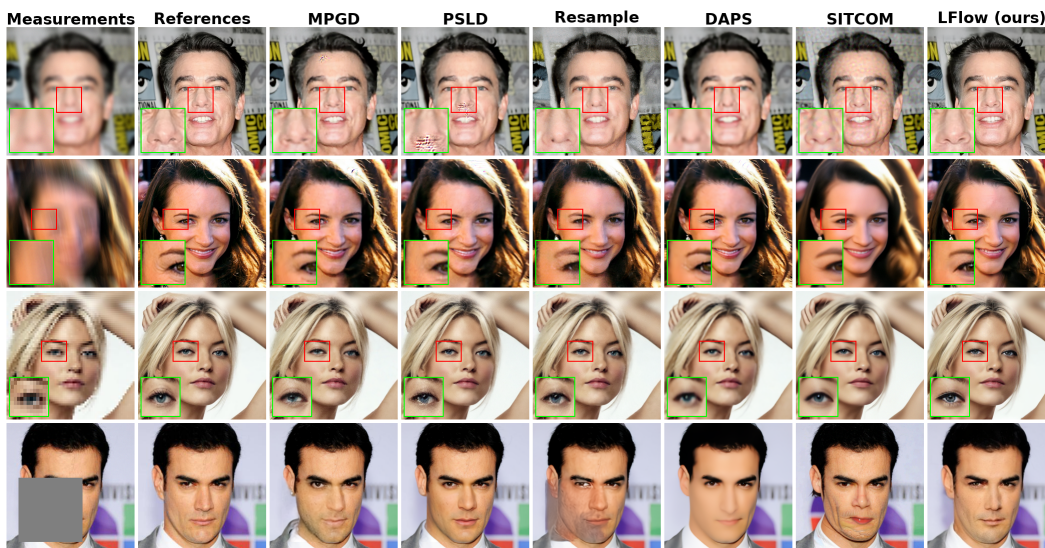


Figure 7: Qualitative results on **CelebA-HQ** test set. Row 1: Deblur (gauss), Row 2: Deblur (motion), Row 3: SR $\times 4$, Row 4: Inpainting.

CelebA-HQ LFlow attains the lowest LPIPS across all four tasks and the highest SSIM in three of four tasks. It reports LPIPS values of 0.164 (Gaussian deblurring), 0.167 (motion deblurring), 0.170 (super-resolution), and 0.123 (inpainting), surpassing the second-best method by 0.042 in inpainting. PSNR remains competitive throughout, ranking first in two tasks and remaining close elsewhere. These quantitative gains are reflected in the visual results: for both deblurring tasks, LFlow restores sharper eye contours, facial edges, and skin textures while avoiding ringing or oversharpening seen in MPGD and PSLD. In super-resolution, it preserves fine details such as eyelashes and lips with smooth transitions, maintaining natural gradients without introducing artifacts. In the inpainting task, LFlow offers semantically consistent completions with coherent tone and geometry—whereas other methods exhibit mismatched shading, seams, or patchy textures. These results highlight LFlow’s ability to recover fine structures while maintaining perceptual realism across diverse facial reconstructions.

D.2 CelebA-HQ ($512 \times 512 \times 3$).

We further tested and evaluated our method on the high-resolution CelebA-HQ dataset ($512 \times 512 \times 3$), demonstrating its robust capabilities in handling complex image processing tasks. In Figure 8, we showcase the effectiveness and versatility of our approach in enhancing image quality across various tasks.



Figure 8: Additional results on CelebA-HQ 512×512 dataset. Row 1: Deblur (gaussian), Row 2: Deblur (motion), Row 3: SRx4, Row 4: Inpainting.

D.3 FID Score Results

Table 6: FID ↓ scores across four inverse problems on **FFHQ**, **CelebA-HQ**, and **ImageNet**.

Method	Deblurring (Gaussian)			Deblurring (Motion)			SR ($\times 4$)			Inpainting (Box)		
	FFHQ	CelebA	ImageNet	FFHQ	CelebA	ImageNet	FFHQ	CelebA	ImageNet	FFHQ	CelebA	ImageNet
LFlow (ours)	52.48	47.79	88.76	<u>57.11</u>	48.53	82.89	58.49	51.07	92.28	34.40	30.78	117.45
SITCOM [60]	74.80	76.72	<u>72.95</u>	70.18	70.26	68.64	67.30	62.58	90.55	45.25	50.05	123.62
DAPS [37]	72.45	65.58	75.51	76.23	67.23	89.17	<u>65.78</u>	48.47	<u>83.42</u>	51.28	45.59	126.36
DMplug [51]	78.50	—	—	75.35	—	—	<u>80.86</u>	—	—	60.36	—	—
Resample [36]	59.64	52.47	63.35	69.74	63.12	85.90	78.62	59.47	105.25	55.60	68.31	138.84
MPGD [38]	64.20	61.37	102.58	70.32	88.7	146.58	90.55	84.43	119.12	84.53	53.15	154.28
PSLD [33]	62.49	57.92	87.39	68.94	75.21	124.73	66.22	70.66	80.58	43.89	40.18	<u>119.12</u>
OT-ODE [45]	<u>56.72</u>	62.23	—	53.55	58.12	—	60.71	<u>47.83</u>	—	<u>40.31</u>	37.92	—
C-ITGFM [48]	—	56.85	—	—	<u>50.54</u>	—	—	45.10	—	—	<u>34.96</u>	—

PSNR and **SSIM** are commonly used recovery metrics that quantify pixel-level fidelity, while **LPIPS** and **FID** are considered perceptual metrics that assess high-level semantic similarity or perceptual quality. In this paper, we focus on image reconstruction tasks from noisy measurements. In such settings, perceptual metrics like FID—although effective for evaluating generative models that prioritize visual realism—primarily measure distribution-level similarity and may overlook structural details, especially when fine-grained information is critical. Moreover, FID can be misleading when reconstructed images appear perceptually plausible but deviate significantly from the ground truth [49]. In contrast, PSNR and SSIM offer objective evaluations of noise suppression and content preservation, which are crucial in our experiments. That said, we report FID scores across all three datasets considered above for four different tasks. As shown in Table 6, our algorithm achieves a balanced trade-off between recovery and perceptual metrics. In the task of noisy image reconstruction, it not only delivers the best recovery metrics but also achieves strong perceptual scores.

D.4 Ablation on hyperparameter K

To evaluate the effect of hyperparameter K , we conducted an ablation study on the FFHQ dataset. As shown in Table 7, setting $K=2$ yields slightly better average performance across tasks compared to $K=1$. Although we also tested $K \geq 3$, the results showed negligible improvements while incurring higher computational costs. For this reason, we adopt $K=2$ as it provides consistent gains while maintaining reasonable inference time.

Table 7: Ablation of parameter K on FFHQ. Increasing K offers marginal gains with higher cost; $K = 2$ achieves the best trade-off between performance and efficiency.

K	GDB			MDB			SR			BIP		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
1	28.72	0.829	0.172	29.71	0.842	0.173	28.80	0.834	0.183	23.59	0.859	0.136
2	29.10	0.837	0.166	30.04	0.849	0.168	29.12	0.841	0.176	23.85	0.867	0.132

D.5 Comparing LFlow with supervised methods

We incorporated **supervised** results obtained using a conditional diffusion model. Our method offers several key advantages over **supervised** inverse approaches such as **SR3** [81] and **InvFussion** [82]:

- **LFlow** is a zero-shot method that generalizes across diverse tasks without retraining, whereas **supervised** methods require training a separate model for each specific task.
- **LFlow** is robust to varying degradation types, while **supervised** methods often exhibit poor generalization when faced with distribution shifts.
- **LFlow** achieves significantly better performance on certain datasets and resolutions—for example, FFHQ at 256×256 .

These advantages are clearly demonstrated in the results presented in Table 8.

Table 8: **Comparison of LFlow with supervised baselines** on two tasks: super-resolution and box inpainting. For *super-resolution*, we compare against **SR3**, trained on the *ImageNet* dataset. For *box inpainting*, we compare against **InvFusion**, trained on the FFHQ dataset.

Method	SR			Inpainting (Box)		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
SR3 [81]	24.65	0.708	0.347	—	—	—
InvFusion [82]	—	—	—	20.12	0.827	0.215
LFlow (ours)	25.29	0.696	0.338	23.85	0.867	0.132

D.6 Pretrained Models and Algorithm Conversion

Conversion of Pretrained Models. Conversion from discrete-time diffusion model to continuous-time flow model was first introduced in [45] by aligning the signal-to-noise ratio (SNR) of the two processes. In principle, this enables mapping discrete diffusion steps $\{\tau\}$ to continuous flow times t and rescaling the noise accordingly. However, this mapping is practically valid under restrictive assumptions—namely, that both trajectories follow the same distributional path (e.g., Gaussian) and employ a linear noise schedule. While such conversion is feasible for pretrained models with similar linear schedules, it becomes non-trivial for discrete-time latent diffusion models (e.g., LDMs, Stable Diffusion). Their custom nonlinear schedules break the clean SNR alignment, often requiring the solution of nonlinear or even cubic equations to infer consistent flow times.

Algorithmic Conversion. Beyond converting the pretrained model itself, one must also consider the sampling *algorithm*. Several baselines are inherently discrete-time: their update rules depend on the availability of a finite noise grid and stepwise re-noising kernels. Such designs do not always admit a continuous-time analogue, and therefore direct conversion is not universally possible. Nevertheless, we successfully extended two representative methods, **PSLD** and **MPGD**, to the continuous-time flow setting. By reformulating their projection–correction steps as infinitesimal updates within an ODE sampler, we obtain continuous-time counterparts that closely follow the spirit of the original algorithms while operating with a pretrained flow prior. In contrast, algorithms that fundamentally rely on discrete re-noising (e.g., DAPS) cannot be faithfully mapped without substantial redesign.

Table 9: Quantitative comparison of LFlow results on **FFHQ** dataset against continuous-time versions of CT-MPGD and CT-PSLD across four inverse problems. **Bold** and underline indicate the best and second-best respectively.

Method	Deblurring (Gaussian)			Deblurring (Motion)			SR ($\times 4$)			Inpainting (Box)		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
LFlow (ours)	29.10	0.837	0.166	30.04	0.849	0.168	<u>29.12</u>	0.841	0.176	23.85	0.867	0.132
PSLD [33]	30.28	0.836	0.281	<u>29.21</u>	0.812	0.303	29.07	<u>0.834</u>	0.270	24.21	0.847	<u>0.169</u>
CT-PSLD	27.11	0.783	0.332	25.98	0.770	0.356	26.53	0.772	0.331	20.60	0.790	0.252
MPGD [38]	29.34	0.815	0.308	27.98	0.803	0.324	27.49	0.788	0.295	20.58	0.806	0.324
CT-MPGD	24.96	0.765	0.343	24.30	0.751	0.375	24.69	0.740	0.338	20.01	0.780	0.353

Discussion. The results in Table 9 clearly show that *algorithmic conversion* is non-trivial. Although we successfully reformulated PSLD and MPGD into their continuous-time counterparts (CT-PSLD and CT-MPGD), both suffer a significant performance drop compared to their original discrete-time versions. Importantly, we kept the comparison fair by employing the same ODE solver (adaptive Heun) as used in LFlow. This indicates that simply replacing the diffusion prior with a flow prior, while retaining the algorithmic structure, does not guarantee competitive performance in continuous time. Instead, the gap highlights the necessity of designing sampling strategies that are intrinsically compatible with ODE-based flow formulations, as achieved in our proposed LFlow framework.

D.7 Additional Visual Results (Best Viewed When Zoomed in)

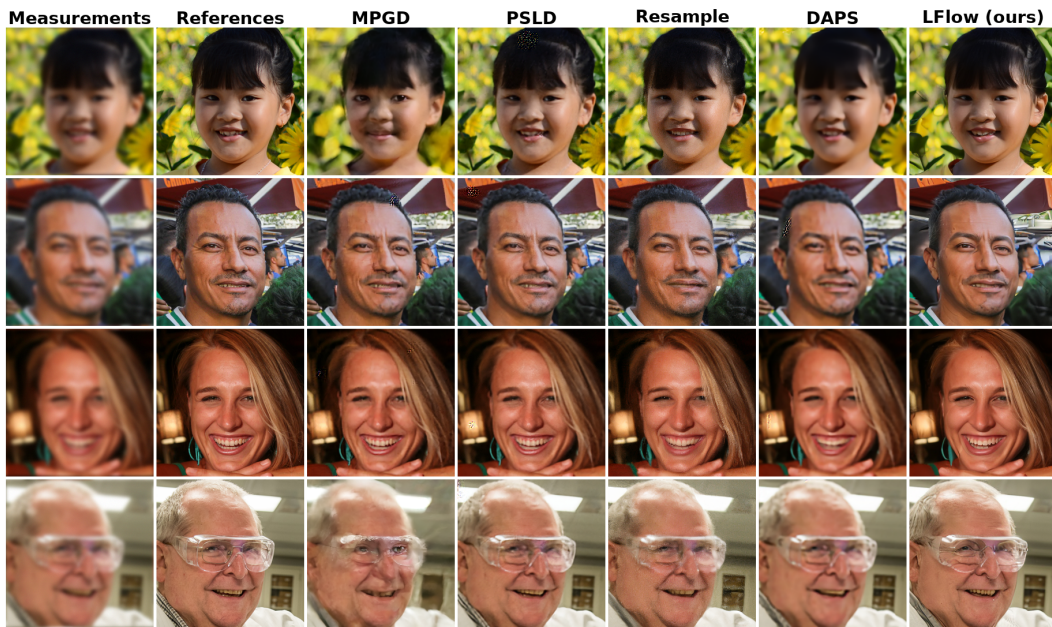


Figure 9: Additional **Gaussian deblurring** results on the **FFHQ** dataset.

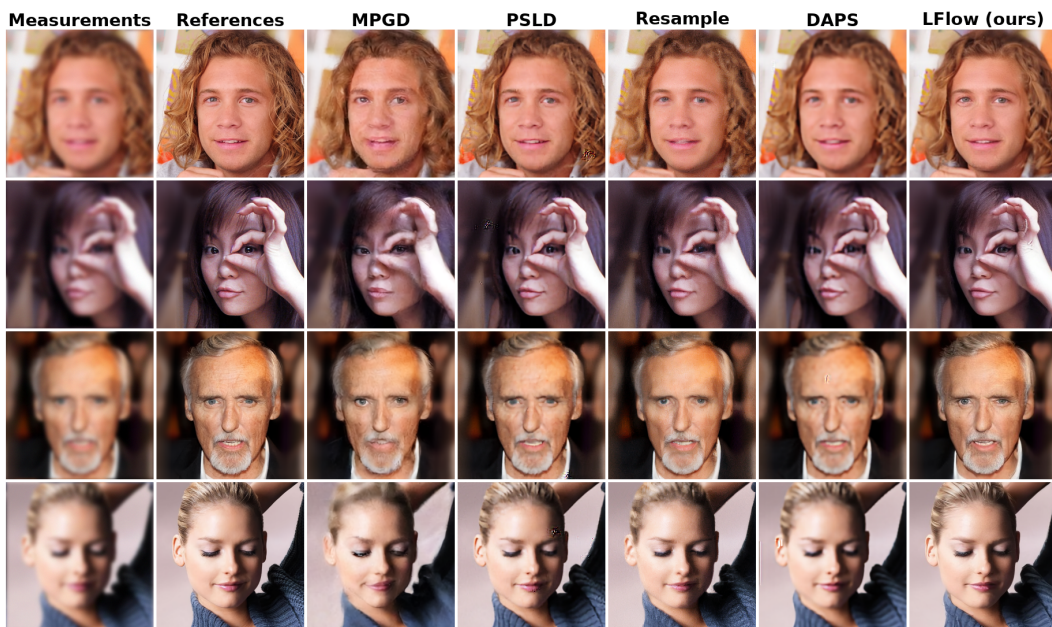


Figure 10: Additional **Gaussian deblurring** results on the **CelebA-HQ** dataset.

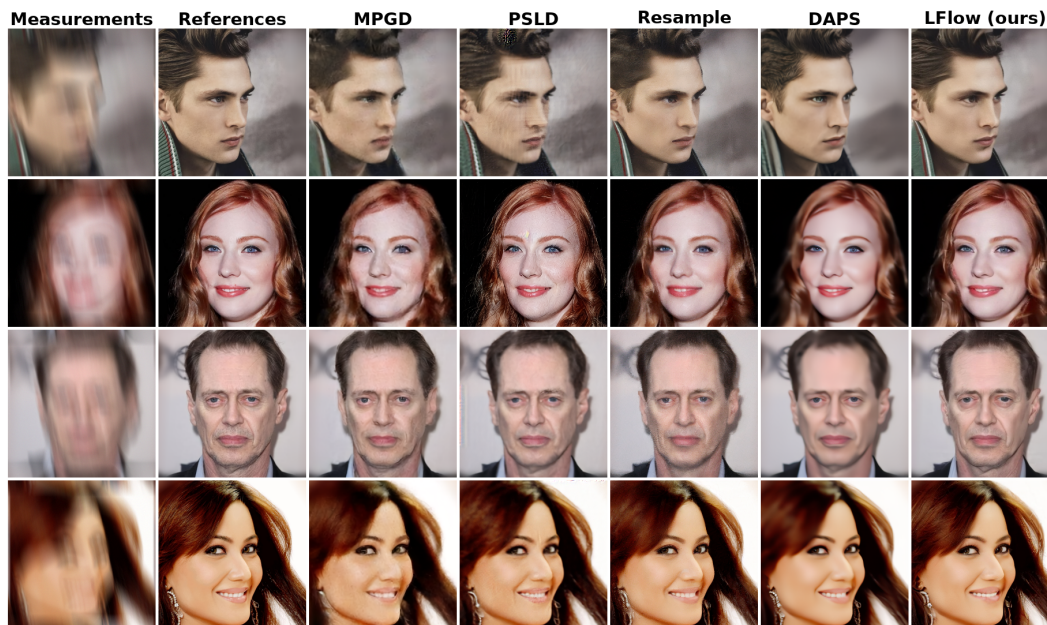


Figure 11: Additional **motion deblurring** results on the **CelebA-HQ** dataset.

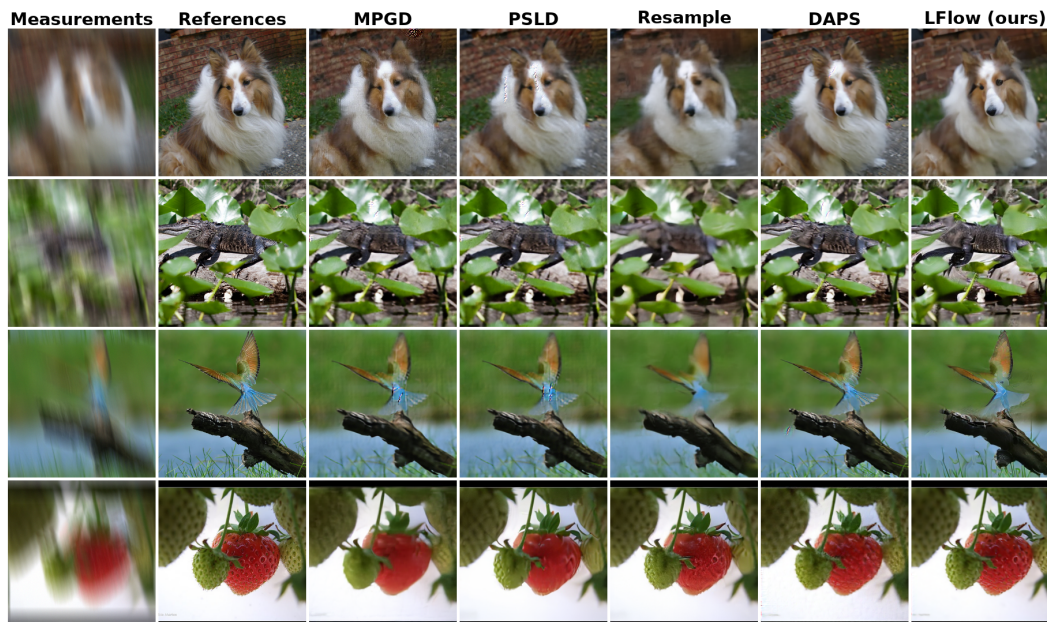


Figure 12: Additional **motion deblurring** results on the **ImageNet** dataset.

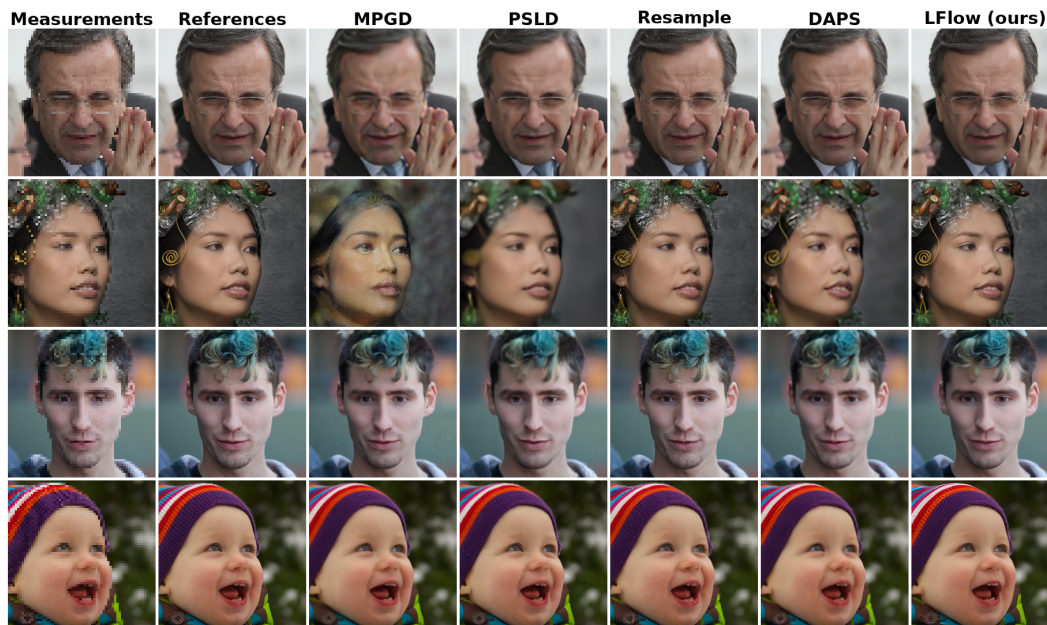


Figure 13: Additional **Super-resolution** results on the **FFHQ** dataset.

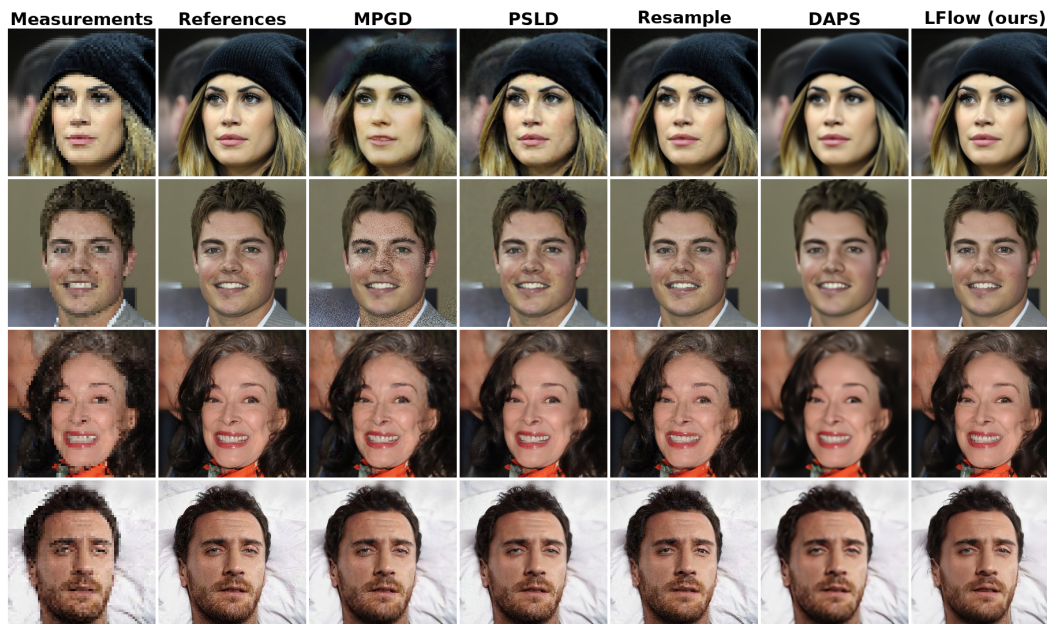


Figure 14: Additional **Super-resolution** results on the **CelebA-HQ** dataset.

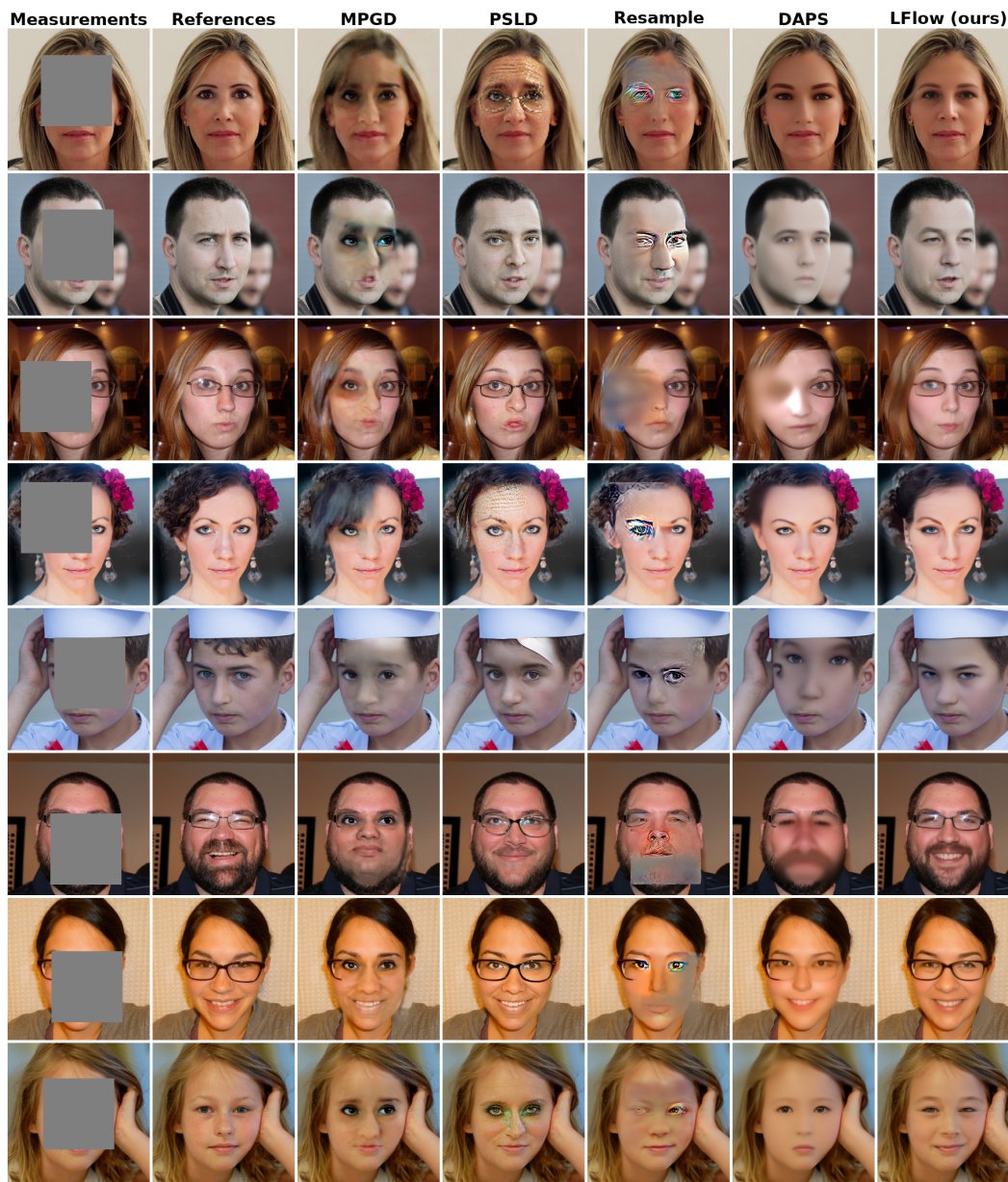


Figure 15: Additional **Inpainting** results on **FFHQ** dataset.

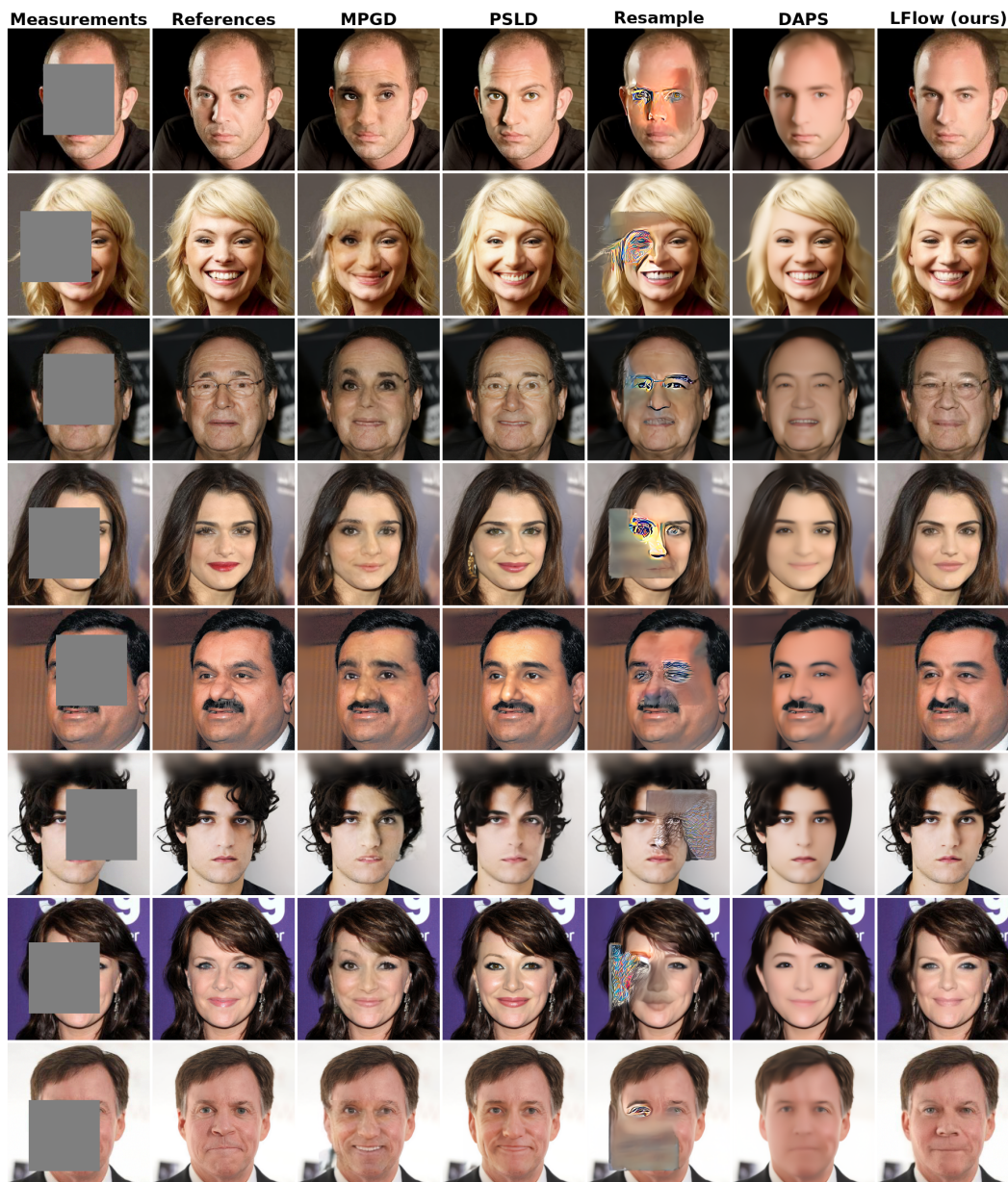


Figure 16: Additional **Inpainting** results on **CelebA-HQ** dataset.

NeurIPS Paper Checklist

The checklist is designed to encourage best practices for responsible machine learning research, addressing issues of reproducibility, transparency, research ethics, and societal impact. Do not remove the checklist: **The papers not including the checklist will be desk rejected.** The checklist should follow the references and follow the (optional) supplemental material. The checklist does NOT count towards the page limit.

Please read the checklist guidelines carefully for information on how to answer these questions. For each question in the checklist:

- You should answer [Yes], [No], or [NA].
- [NA] means either that the question is Not Applicable for that particular paper or the relevant information is Not Available.
- Please provide a short (1–2 sentence) justification right after your answer (even for NA).

The checklist answers are an integral part of your paper submission. They are visible to the reviewers, area chairs, senior area chairs, and ethics reviewers. You will be asked to also include it (after eventual revisions) with the final version of your paper, and its final version will be published with the paper.

The reviewers of your paper will be asked to use the checklist as one of the factors in their evaluation. While "[Yes]" is generally preferable to "[No]", it is perfectly acceptable to answer "[No]" provided a proper justification is given (e.g., "error bars are not reported because it would be too computationally expensive" or "we were unable to find the license for the dataset we used"). In general, answering "[No]" or "[NA]" is not grounds for rejection. While the questions are phrased in a binary way, we acknowledge that the true answer is often more nuanced, so please just use your best judgment and write a justification to elaborate. All supporting evidence can appear either in the main paper or the supplemental material, provided in appendix. If you answer [Yes] to a question, in the justification please point to the section(s) where related material for the question can be found.

IMPORTANT, please:

- **Delete this instruction block, but keep the section heading "NeurIPS Paper Checklist",**
- **Keep the checklist subsection headings, questions/answers and guidelines below.**
- **Do not modify the questions and only use the provided macros for your answers.**

1. Claims

Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

Answer: [Yes]

Justification: We only claim to present a new method, which we describe in Section 4. We compare our method against previous ones in Sections 5.1 and 5.2.

Guidelines:

- The answer NA means that the abstract and introduction do not include the claims made in the paper.
- The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
- The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. Limitations

Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [Yes]

Justification: We discuss the limitations in Section 2.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. Theory assumptions and proofs

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [NA]

Justification: For each theoretical result presented in the paper, we provide the full set of assumptions explicitly and include complete, step-by-step proofs. All derivations are checked for mathematical correctness, with no steps omitted. Where appropriate, we include formal statements of assumptions (e.g., regularity or convexity conditions), and provide full proofs either in the main text or in the appendix to ensure transparency and reproducibility.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. Experimental result reproducibility

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: The manuscript describes all methods and experiments. Algorithms are provided for the methods. The code will be available in a future update.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general, releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
 - (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
 - (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
 - (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
 - (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. Open access to data and code

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [No]

Justification: We plan to publicly release the full codebase and instructions to reproduce all experimental results upon acceptance. At submission time, we are unable to share the code due to anonymity requirements and ongoing cleanup of the repository. We confirm that all experiments are reproducible and will be documented with detailed scripts and setup instructions.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.

- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. Experimental setting/details

Question: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: Experiment details are provided in Appendix C.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. Experiment statistical significance

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [No]

Justification: Computing error bars for Table 1, 2, and 7 would require retraining every model several times for different datasets.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. Experiments compute resources

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [Yes]

Justification: All experiments were conducted on a local machine equipped with a single NVIDIA RTX 3090 GPU. Experiment details are provided in Appendix C.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. Code of ethics

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics <https://neurips.cc/public/EthicsGuidelines?>

Answer: [Yes]

Justification: We have reviewed and agree with the NeurIPS Code of Ethics.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. Broader impacts

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [NA]

Justification: There is no societal impact of the work performed.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. Safeguards

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: The paper poses no such risks.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. Licenses for existing assets

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: The FFHQ, ImageNet and CelebA datasets are used under their respective non-commercial research licenses and cited appropriately.

Any third-party code or pretrained models (e.g., for flow matching) are used under compatible open-source licenses (e.g., MIT) and are cited in the main paper.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, paperswithcode.com/datasets has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. New assets

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [NA]

Justification: : The paper does not release new assets.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.

- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. Crowdsourcing and research with human subjects

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: The paper does not involve crowdsourcing nor research with human subjects

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. Institutional review board (IRB) approvals or equivalent for research with human subjects

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: The paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. Declaration of LLM usage

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [NA]

Justification: This work does not involve large language models (LLMs) in any part of the core methodology. LLMs were not used for designing algorithms, generating model outputs, or conducting any experiments. They were only used for minor writing support and clarity improvement, which does not impact the scientific originality or rigor of the research.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (<https://neurips.cc/Conferences/2025/LLM>) for what should or should not be described.