Whitened Score Diffusion: A Structured Prior for Imaging Inverse Problems

Jeffrey Alido[1]   **Tongyu Li**[1]   **Yu Sun**[2]   **Lei Tian**[1]

[1]Department of Electrical and Computer Engineering,
Boston University, Boston, MA 02215, USA
[2]Department of Electrical and Computer Engineering,
Johns Hopkins University, Baltimore, MD 21205, USA
{jalido, tongyuli, leitian}@bu.edu, ysun214@jh.edu

Abstract

Conventional score-based diffusion models (DMs) may struggle with anisotropic Gaussian diffusion processes due to the required inversion of covariance matrices in the denoising score matching training objective [58]. We propose Whitened Score (WS) diffusion models, a novel framework based on stochastic differential equations that learns the Whitened Score function instead of the standard score. This approach circumvents covariance inversion, extending score-based DMs by enabling stable training of DMs on arbitrary Gaussian forward noising processes. WS DMs establish equivalence with flow matching for arbitrary Gaussian noise, allow for tailored spectral inductive biases, and provide strong Bayesian priors for imaging inverse problems with structured noise. We experiment with a variety of computational imaging tasks using the CIFAR, CelebA (64×64), and CelebA-HQ (256×256) datasets and demonstrate that WS diffusion priors trained on anisotropic Gaussian noising processes consistently outperform conventional diffusion priors based on isotropic Gaussian noise. Our code is open-sourced at github.com/jeffreyalido/wsdiffusion.

1   Introduction

Diffusion models (DMs) are a powerful class of generative models that implicitly learn a complex data distribution by modeling the (Stein) score function [48, 49, 18, 13, 26, 25]. The score function is then plugged into a reverse denoising process described by an ordinary differential equation (ODE) or a stochastic differential equation (SDE) to generate novel samples from noise. Typically, the forward noising process is defined by adding different levels of isotropic Gaussian noise to a clean data sample, which enables a simple and tractable denoising score matching (DSM) objective [58]. However, the DSM objective exhibits instability when the forward diffusion noise covariance is ill-conditioned or singular, as its computation requires inverting the covariance matrix.

Flow matching (FM) [33, 35, 1, 65] is an alternative generative modeling paradigm that reshapes an arbitrary known noise distribution into a complex data distribution according to an implicit probability path constructed by the flow. For the isotropic Gaussian case, [34, 52] established that FM and DMs are equivalent up to a rescaling of the noise parameters that define the SDE and probability paths. However, for anisotropic Gaussian noise, there exists a gap between score-based DMs and FM, where score-based DMs cannot be as easily trained for arbitrary Gaussian forward noising processes due to the necessary inversion of the covariance matrix in the conditional score [58].

A denoising DM capable of denoising structured, correlated noise is desirable in many scientific inverse problems, especially in imaging, as it may serve as a rich Bayesian prior [49, 16, 63, 28]. Imaging through fog, turbulence and scattering [3, 64, 31], wide-field microscopy [39], diffraction tomography [32, 30], optical coherence tomography (OCT) [21], interferometry [56] and many other

imaging modalities have an image formation process corrupted by structured, spatially correlated noise [62, 6], in contrast to the widely assumed additive isotropic (white) Gaussian noise. Conventional DMs are trained on isotropic Gaussian noise, which may render them practically insufficient Bayesian priors for realistic use cases with correlated noise.

Motivated by FM's ability to model arbitrary probability paths and the expressiveness of diffusion priors for inverse problems, we propose **Whitened Score Diffusion**, a framework for learning DMs based on arbitrary Gaussian noising processes. Instead of learning the (time-dependent) score function, $\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)$, we learn $\mathbf{G}_t \mathbf{G}_t^\top \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)$, with $\mathbf{G}_t$ the diffusion matrix in the forward diffusion process (Fig. 1). We term our framework Whitened Score (WS) DMs, after the whitening transformation that transforms the score vector field into an isotropic vector field. This extends the current SDE framework for score-based DMs as it avoids the computation of the inverse covariance for any anisotropic Gaussian noise in DSM objective, enabling an arbitrary choice of Gaussian probability paths, similar to FM. We elaborate on the equivalence of our framework to FM and draw a connection to the reverse-time diffusion process derivation by [8], where $\mathbf{G}_t \mathbf{G}_t^\top \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)$ is a *predictable process* of the stochastic term in a reverse-time SDE.

This work presents an extension of score-based DMs to arbitrary Gaussian forward processes, bridging a gap between DMs and FM. Our framework enables a principled construction of denoising generative priors that incorporate spectral bias aligned with correlated measurement noise, leading to improved performance in inverse problems with structured noise. Empirical results on CIFAR-10, CelebA (64×64), and CelebA-HQ (256×256) across several imaging tasks show consistently higher peak signal-to-noise ratio (PSNR) reconstructions compared to conventional models trained with isotropic noise. Our contributions are: (i) a framework for training DMs that supports arbitrary Gaussian probability paths, (ii) theoretical insights and a connection to FM, and (iii) a demonstration of effective priors for imaging inverse problems under structured noise.
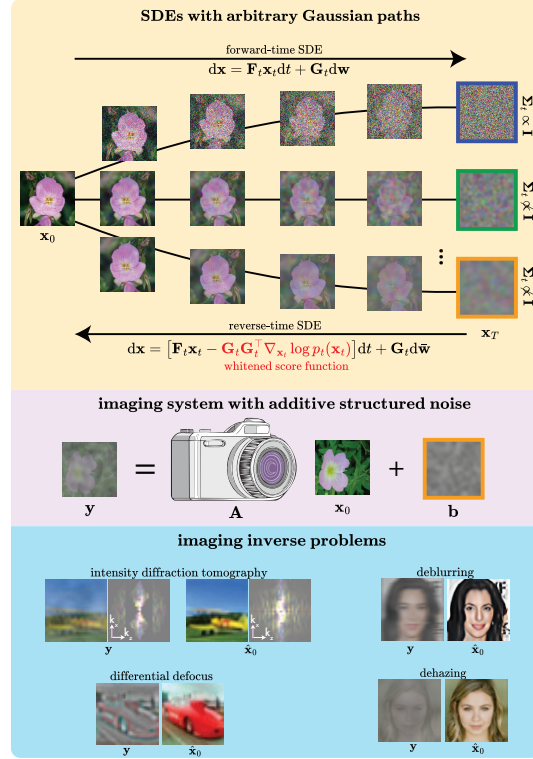

Figure 1: Our framework enables arbitrary Gaussian diffusion processes, allowing us to train a denoising DM on a diverse set of structured noise. The WS DM applies to a variety of imaging inverse problems corrupted with correlated, structured noise.

2 Background

2.1 Score-based diffusion models

Score-based DMs are a class of generative models that estimate a probability density function $p(\mathbf{x}_0)$ by reversing a time-dependent noising process. In continuous time, the forward noising process is described by an Itô SDE in the form of,

$$\mathrm{d}\mathbf{x}_t = \mathbf{F}_t \mathbf{x}_t \mathrm{d}t + \mathbf{G}_t \mathrm{d}\mathbf{w}. \tag{1}$$

$\mathbf{F}_t \in \mathbb{R}^{m \times m}$ is the drift coefficient, $\mathbf{w} \in \mathbb{R}^m$ is the standard Wiener process (Brownian motion), and $\mathbf{G}_t \in \mathbb{R}^{m \times m}$ is the diffusion matrix that controls the structure of the noise. The noise level is indexed by time $t \in [0, T]$ such that $\mathbf{x}_0 \sim p(\mathbf{x}_0)$, $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and $\mathbf{x}_t \sim p(\mathbf{x}_t \mid \mathbf{x}_0)$, a probability transition Gaussian kernel defined by Eq. 1.

The corresponding reverse-time SDE for Eq. 1 is:

$$d\mathbf{x}_t = \left[\mathbf{F}_t\mathbf{x}_t - \mathbf{G}_t\mathbf{G}_t^\top \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)\right] dt + \mathbf{G}_t \, d\bar{\mathbf{w}}_t, \tag{2a}$$

and the deterministic ODE, also known as probability flow, with the same time-marginals is:

$$d\mathbf{x}_t = \left[\mathbf{F}_t\mathbf{x}_t - \frac{1}{2}\mathbf{G}_t\mathbf{G}_t^\top \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)\right] dt, \tag{2b}$$

where $\bar{\mathbf{w}}$ is the reverse-time standard Wiener process and $\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)$ is the Stein score function.

Sampling from $p(\mathbf{x}_0)$ requires solving Eq. 2 and thereby knowing the score function, $\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)$. [48, 49] approximated the score function with a neural network $\mathbf{s}_\theta(\mathbf{x}_t, t)$ by optimizing the DSM objective [58]:

$$\hat{\theta} = \arg\min_\theta \mathbb{E}_{t\sim U(0,1], \mathbf{x}_t\sim p(\mathbf{x}_t|\mathbf{x}_0), \mathbf{x}_0\sim p(\mathbf{x})} \left\{\|\mathbf{s}_\theta(\mathbf{x}_t, t) - \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t \mid \mathbf{x}_0)\|_2^2\right\}, \tag{3}$$

where the conditional score function has a closed form expression given by

$$\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t \mid \mathbf{x}_0) = \mathbf{\Sigma}_t^{-1}(\boldsymbol{\mu}_t - \mathbf{x}_t), \tag{4}$$

with $\boldsymbol{\mu}_t$ and $\mathbf{\Sigma}_t$ the mean and covariance of the Gaussian transition kernel $p(\mathbf{x}_t \mid \mathbf{x}_0)$. $\boldsymbol{\mu}_t$ and $\mathbf{\Sigma}_t$ are functions of the drift coefficient and diffusion matrix in Eq. 1 attained by solving the ODEs in Eqs. 5.50 and 5.51 in [54], creating a linearly proportional relationship as $\boldsymbol{\mu}_t \propto \mathbf{x}_0$ and $\mathbf{\Sigma}_t \propto \mathbf{G}_t\mathbf{G}_t^\top$. This leads the transformed score function term, $\mathbf{G}_t\mathbf{G}_t^\top \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)$ in Eq. 2b to always be isotropic, as the covariance will multiply with its inverse, regardless of the diffusion coefficient, $\mathbf{G}_t$.

2.2 Structured Forward Processes in Diffusion Models

Conventional score-based diffusion models (DMs) typically employ uncorrelated white Gaussian noise, corresponding to a diagonal diffusion matrix $\mathbf{G}_t$ [49]. However, this formulation constrains the learned score function $\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t)$ to isotropic noise settings. Extending beyond diagonal $\mathbf{G}_t$ poses numerical challenges, as the inversion of the covariance matrix in Eq. 4 can become unstable for ill-conditioned or singular cases.

Recent studies have demonstrated that controlling the spatial frequency content of the forward noise can influence the model's inductive spectral bias and enhance generative flexibility [23]. Yet, a unified framework for training diffusion models under arbitrary noising processes remains lacking. Our work addresses this gap through the lens of score-based DMs in the SDE formulation, establishing a principled foundation for frequency-controlled and structured forward processes that can be adapted to diverse DM tasks.

Several prior approaches have introduced structured forward processes to improve generative expressivity. CLD [15] and PSLD [40] extend the state space by incorporating velocity variables, injecting noise in phase space to simplify score estimation, albeit at the cost of auxiliary dynamics. MDMs [45] and Blurring Diffusion [19] employ anisotropic or spatially correlated noise but require inversion of dense covariance matrices. Flexible Diffusion [17] parameterizes the forward SDE to allow adaptive noise scheduling, increasing model complexity and training cost.

These advances collectively underscore the importance of moving beyond isotropic Gaussian noise while revealing practical limitations related to stability and computational overhead. Motivated by this, our proposed WS model enables arbitrary Gaussian forward processes without covariance inversion, offering a simple, stable, and general mechanism for structured generative modeling.

2.3 Flow matching

Flow matching (FM) [33, 35, 1] is another paradigm in generative modeling that connects a noise distribution and a data distribution with an ODE

$$\frac{d\phi_t(\mathbf{x}_t)}{dt} = \mathbf{u}_t(\phi_t(\mathbf{x}_t)), \tag{5}$$

for FM vector field $\mathbf{u}_t(\mathbf{x}_t)$ and initial condition $\phi_0(\mathbf{x}_0) = \mathbf{x}_0$. Noise samples are transformed along time into a sample from the data distribution using a neural network that models the conditional FM vector field

$$\mathbf{u}_t(\mathbf{x}_t \mid \mathbf{x}_0) = \mathbf{\Sigma}_t'(\mathbf{x}_0)\mathbf{\Sigma}_t^{-1}(\mathbf{x}_0)(\mathbf{x}_t - \boldsymbol{\mu}_t(\mathbf{x}_0)) + \boldsymbol{\mu}_t'(\mathbf{x}_0), \tag{6}$$

3

where $\boldsymbol{\mu}_t(\mathbf{x}_0)$ and $\boldsymbol{\Sigma}_t(\mathbf{x}_0)$ are the mean and covariance of the probability path p_t, and f' denotes the time derivative of f. Because $\boldsymbol{\Sigma}'_t$ is proportional to $\boldsymbol{\Sigma}_t$ up to a scalar coefficient, multiplying by the inverse to yield identity [54], the functional form of $\mathbf{u}_t(\mathbf{x}_t \mid \mathbf{x}_0)$ allows simple and stable training of FM models with arbitrary Gaussian probability paths. This diagonal matrix-yielding multiplication currently lacks in score-based models due to the necessary inversion of the covariance matrix in Eq. 4.

We note that WS aligns with FM in the sense that both frameworks aim to enable arbitrary probability paths. In Section 3.2, we present a formal connection between WS and FM. Nevertheless, FM may require new approaches to incorporate the measurement likelihood for solving inverse problems [65], whereas our WS framework can be readily combined with existing techniques for enforcing measurement consistency (see Section 2.4 for a review).

2.4 Imaging inverse problems with diffusion model priors

Reconstructing an unknown signal $\mathbf{x}_0 \in \mathbb{R}^m$ from a measurement $\mathbf{y} \in \mathbb{R}^n$ given a known forward model $\mathbf{y} \sim \mathcal{N}(\mathbf{A}\mathbf{x}_0, \boldsymbol{\Sigma}_\mathbf{y})$—with $\boldsymbol{\Sigma}_\mathbf{y} \in \mathbb{R}^{n \times n}$ the covariance of the additive Gaussian noise and $\mathbf{A} \in \mathbb{R}^{m \times n}$ the measurement forward model—is a central challenge in computational imaging and scientific problems. Recent advances employ DMs as flexible priors [14], using plug-and-play schemes [66, 67, 60], likelihood-guided sampling via posterior score approximations [22, 49, 9, 11, 47, 27], Markov Chain Monte Carlo (MCMC) techniques [38, 7, 61, 63, 53, 55], variational methods [16, 37], and latent DM frameworks [43, 12, 46].

Here, we adopt methods that approximate the posterior. The posterior score can be factored into the prior score and the likelihood score using Bayes's rule to arrive at a modification of Eq. 2 for the stochastic reverse diffusion

$$\mathrm{d}\mathbf{x}_t = \left[\mathbf{F}_t \mathbf{x}_t - \mathbf{G}_t \mathbf{G}_t^\top \left(\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t) + \nabla_{\mathbf{x}_t} \log p_t(\mathbf{y} \mid \mathbf{x}_t) \right) \right] \mathrm{d}t + \mathbf{G}_t \, \mathrm{d}\bar{\mathbf{w}}_t, \qquad (7a)$$

and the deterministic reverse diffusion

$$\mathrm{d}\mathbf{x}_t = \left[\mathbf{F}_t \mathbf{x}_t - \frac{1}{2} \mathbf{G}_t \mathbf{G}_t^\top \left(\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t) + \nabla_{\mathbf{x}_t} \log p_t(\mathbf{y} \mid \mathbf{x}_t) \right) \right] \mathrm{d}t. \qquad (7b)$$

The prior score is approximated by the denoising DM. However, the measurement likelihood score is intractable due to the time-dependence. Methods in [11, 5, 47, 50] make simplifying assumptions about the prior distribution, while those in [22, 10, 28, 43] treat the likelihood score approximation as an empirically designed update using the measurement as a guiding signal. All these likelihood score approximations can thus be plugged into Eq. 7 to solve the inverse problem.

A major gap in current research on imaging inverse problems is the consideration of additive *structured* noise. Most research on DM priors for imaging inverse problems has largely focused on scenarios with isotropic Gaussian noise, employing corresponding isotropic Gaussian denoising DMs. Recent work by [20] explored structured priors for imaging inverse problems using stochastic restoration priors achieving superior performance over conventional denoising DMs trained on isotropic Gaussian noising processes in cases involving both correlated and uncorrelated noise. However, a formal treatment of structured noise in diffusion-based frameworks lacks, which we seek to address.

3 Whitened Score Diffusion

We define our forward-time SDE with non-diagonal diffusion matrix as,

$$\mathrm{d}\mathbf{x}_t = \underbrace{-\frac{1}{2}\beta_t \, \mathbf{x}_t \mathrm{d}t}_{:=\mathbf{F}_t} + \underbrace{\sqrt{\beta_t}\mathbf{K}}_{:=\mathbf{G}_t} \mathrm{d}\mathbf{w}, \qquad (8)$$

adopting from the variance-preserving (VP) SDE [49]. In our experiments, we constrain $\mathbf{K}$ to be in the class of circulant convolution matrices due to their ability to be implemented with the fast Fourier transform (FFT). However, our method generalizes to any $\mathbf{K}$ that is positive semidefinite. When $\mathbf{K} = \mathbf{I}$, we recover exactly the VP-SDE. The corresponding probability transition kernel of Eq. 8 is[1]

$$p(\mathbf{x}_t \mid \mathbf{x}_0) = \mathcal{N}\left(\mathbf{x}_t \mid \alpha_t \mathbf{x}_0, (1 - \alpha_t^2)\mathbf{K}\mathbf{K}^\top \right), \qquad (9)$$

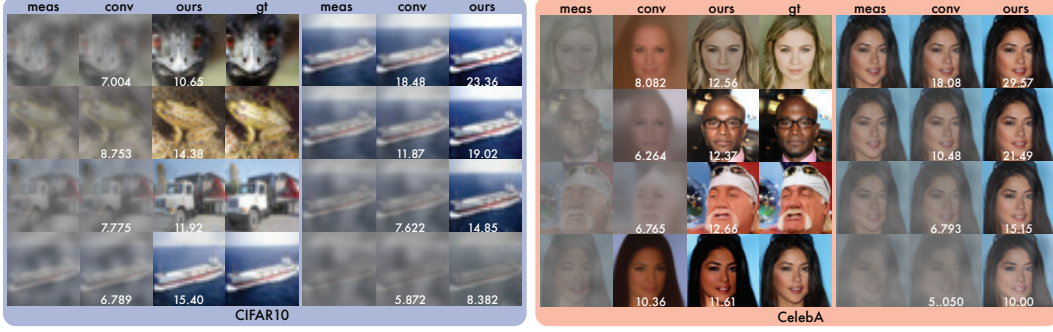[1]The mean and covariance of the transition kernel are solved in Eqs. 5.50 and 5.51 in [54].

Figure 2: Denoising correlated noise on CIFAR10 and CelebA (64×64). We benchmark our WS DM trained on anisotropic Gaussian noise with the conventional DM (conv) trained on isotropic Gaussian noise. Left: results with a fixed SNR of 0.26; Right: measurements $\mathbf{y}$ with decreasing SNR from 1.4 to 0.12 using additive grayscale noise filtered by a Gaussian kernel of std 2.5 and 5 pixels for CIFAR10 and CelebA, respectively. The PSNR is labeled in white.

where $\alpha_t = e^{-\frac{1}{2} \int_0^t \beta_s \mathrm{d}s}$. In general, α_t is defined as the integral of the drift coefficient from 0 to t, $\alpha_t = \int_0^t \mathbf{F}_s \mathrm{d}s$. By leveraging the parameterization trick for Gaussian distributions, we may rewrite Eq. 9 as the following continuous time system:

$$\mathbf{x}_t = \alpha_t \mathbf{x}_0 + \sqrt{1 - \alpha_t^2} \mathbf{K} \mathbf{z}, \quad \mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \tag{10}$$

Note that we may use other drift and diffusion matrices, such as the variance-exploding (VE) SDE, ending up with scalar multiples of $\mathbf{x}_0$ and $\mathbf{G}_t \mathbf{G}_t^\top$ for the mean and covariance, respectively, given the initial conditions of $\boldsymbol{\mu}_0 = \mathbf{x}_0$ and $\boldsymbol{\Sigma}_0 = \mathbf{0}$. Specific to our SDE in Eq. 8, we define the signal-to-noise ratio (SNR) to be the ratio $\alpha_t / \sqrt{1 - \alpha_t^2}$.

3.1 Whitened Score matching objective

From Eq. 9, the conditional score to solve Eq. 3 is

$$\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t \mid \mathbf{x}_0) = \left((1 - \alpha_t^2) \mathbf{K} \mathbf{K}^\top \right)^{-1} (\alpha_t \mathbf{x}_0 - \mathbf{x}_t), \tag{11}$$

and inverting the matrix may often lead to instability in the score computation. For example, the condition number of a Gaussian convolution matrix grows as the Gaussian kernel $\mathbf{K}$ widens, amplifying high spatial frequency features, leading to poor model training for the DSM objective in Eq. 3.

To mitigate these numerical instabilities in the score computation during training, we apply a *whitening* transformation to the score by naturally multiplying it with $\mathbf{G}_t \mathbf{G}_t^\top$, where $\mathbf{G}_t \mathbf{G}_t^\top \propto \boldsymbol{\Sigma}_t$, the forward diffusion process covariance. Similar to DSM, for our SDE in Eq. 8, we approximate $\mathbf{G}_t \mathbf{G}_t^\top \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)$ as $\mathbf{G}_t \mathbf{G}_t^\top \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t \mid \mathbf{x}_0)$ which has the following closed-form expression after canceling $\boldsymbol{\Sigma}_t$ with $\mathbf{G}_t \mathbf{G}_t^\top$:

$$\mathbf{G}_t \mathbf{G}_t^\top \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t \mid \mathbf{x}_0) = \beta_t \frac{\alpha_t \mathbf{x}_0 - \mathbf{x}_t}{1 - \alpha_t^2}. \tag{12}$$

We train a model $\mathbf{n}_\theta(\mathbf{x}_t, t)$ using the following denoising WS matching loss:

$$L = \mathbb{E}_{t \sim U(0,1), \mathbf{x}_t \sim p(\mathbf{x}_t | \mathbf{x}_0), \mathbf{x}_0 \sim p(\mathbf{x})} \left\{ \| \mathbf{n}_\theta(\mathbf{x}_t, t) - \mathbf{G}_t \mathbf{G}_t^\top \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t \mid \mathbf{x}_0) \|_2^2 \right\}, \tag{13}$$

with proof in Appendix A. This objective accounts for varying levels of spatial correlation in the noise to enable our model to denoise arbitrary Gaussian noise.

This objective defines a new learning target within the broader landscape of diffusion model losses. Our approach can be seen as a generalization of noise prediction [18] to the setting of correlated Gaussian noise, where the preconditioning term $\mathbf{G}_t \mathbf{G}_t^\top$ captures the noise structure. Unlike conventional noise prediction, which assumes isotropic noise, our formulation enables stable training under
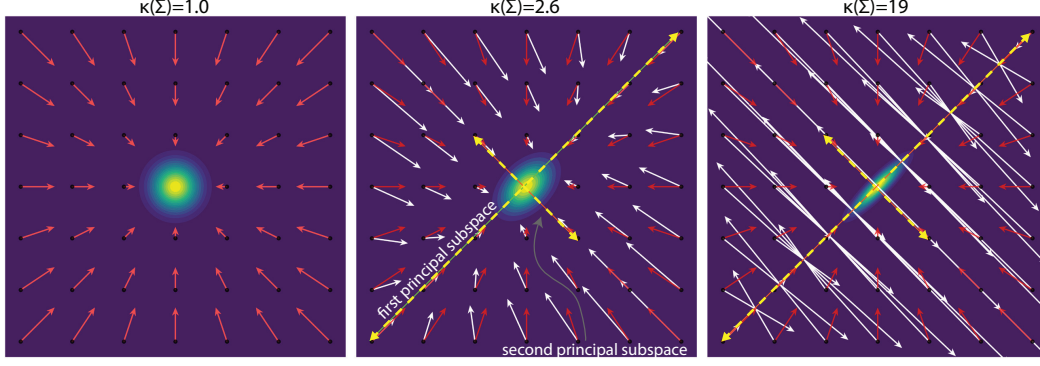
5

Figure 3: $\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t \mid \mathbf{x}_0)$ vector field (white) and $\mathbf{G}_t\mathbf{G}_t^\top \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t \mid \mathbf{x}_0)$ vector field (red) for increasingly anisotropic 2D Gaussian probability transition kernel $p(\mathbf{x}_t \mid \mathbf{x}_0)$. The covariance amplifies the magnitude of the conditional score field by its condition number $\kappa(\boldsymbol{\Sigma})$, and additionally rotates the direction towards the first principal subspace where there is higher density, while the $\mathbf{G}_t\mathbf{G}_t^\top \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t \mid \mathbf{x}_0)$ field remains stable in magnitude and directionally isotropic pointing towards the mean $\boldsymbol{\mu}_t$ of the probability path.

arbitrary Gaussian forward processes. Furthermore, Eq. 15 reveals that the conditional FM vector field $\mathbf{u}_t$ is a linear combination of our conditional WS function and the drift term (see Appendix B), highlighting that both WS and FM avoid covariance inversion by preconditioning the score function with $\mathbf{G}_t\mathbf{G}_t^\top$. This shared property enables principled modeling of flexible Gaussian probability paths.

3.2 Interpretation of WS

Concurrently with [4], [8] derived identical results for the reverse-time SDE, Eq. 2a, by decomposing the diffusion term of a reverse-time SDE into a unique sum of a zero-mean martingale and a *predictable process* $\mathbf{n}_t$, given as

$$\mathbf{n}_t = \frac{\sum_{i=1}^m \frac{\partial}{\partial x_t^i} \sum_{k=1}^m G_t^{ik}(\mathbf{x}_t, t) G_t^{\cdot k}(\mathbf{x}_t, t) p_t(\mathbf{x}_t)}{p_t(\mathbf{x}_t)}. \tag{14}$$

When $\mathbf{G}_t$ is independent of the state $\mathbf{x}_t$, $\mathbf{n}_t$ simplifies to $\mathbf{G}_t\mathbf{G}_t^\top \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)$. This process is conditionally deterministic with respect to the filtration of the reverse time flow, motivating modeling the complete predictable process instead of the score function in isolation.

Furthermore, multiplying the score with $\mathbf{G}_t\mathbf{G}_t^\top$ whitens its vector field, as seen in Fig. 3 leading to a two-fold effect. Firstly, the original score vector field is numerically unstable; its values are highly sensitive to small errors in the residual, characterized by the condition number $\kappa(\boldsymbol{\Sigma})$. This leads to unstable model training, as there is often noise amplified by the condition number. Multiplying the field with $\mathbf{G}\mathbf{G}^\top$, a scalar multiple of the transition kernel's covariance, preconditions the field.

Secondly, the score field rotates in the direction towards the major principal axis that contains most of the density for the noise *transition kernel* $p(\mathbf{x}_t \mid \mathbf{x}_0)$. For anisotropic Gaussian transition kernels, the score does not point towards the data distribution, but rather towards the major principal axis of the (correlated) *noise* from the forward-time SDE. Eq. 2 naturally re-orients the field towards the data mean, providing motivation for modeling the *complete predictable process* instead of solely the score function. Furthermore, by learning the predictable process, we enable a more general scheme for SDE-based DMs by developing a model that will always have isotropic reverse-time sample paths without needing to specify the diffusion matrix during sampling.

Connection to FM To connect WS DMs with FM and explain why training models with arbitrary Gaussian probability paths is achieved, we re-frame FM with the SDE framework and rewrite the conditional FM vector field expressed in terms of the VP-SDE variables in Eq. 8:

$$\mathbf{u}_t(\mathbf{x}_t \mid \mathbf{x}_0) = \mathbf{F}_t(2\mathbf{x}_t - \alpha_t\mathbf{x}_0) + \mathbf{G}_t\mathbf{G}_t^\top \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t \mid \mathbf{x}_0). \tag{15}$$

Eq. 15 reveals that the conditional FM vector $\mathbf{u}_t$ is a linear combination of our conditional WS function and the drift term (see Appendix B for derivation). The key property shared by FM and WS

Figure 4: Measurement noise with different covariance matrices shown in the bottom left of the measurement. The PSNR is shown in white text for a sample image in the CIFAR10 validation dataset and the CelebA (64×64) validation dataset. Compared to DMs trained only on isotropic Gaussian noise, our WS model is able to denoise correlated noise with superior PSNR. For uncorrelated noise (first row), our model has similar performance as conventional DMs.

DMs is that they avoid inverting the covariance matrix in the score by preconditioning it with $\mathbf{G}_t\mathbf{G}_t^\top$, enabling flexible modeling of arbitrary Gaussian probability paths.

3.3 WS diffusion priors for imaging inverse problems

We solve the imaging inverse problem using Eq. 7 with our WS diffusion prior and an approximation of the measurement likelihood score. Recall from Section 2.4 the myriad of methods developed to approximate the measurement likelihood score, all of which follow the template,

$$\nabla_{\mathbf{x}_t} \log p(\mathbf{y} \mid \mathbf{x}_t) \approx \boldsymbol{\Sigma}_{\mathbf{y}}^{-1} \nabla_{\mathbf{x}_t} \mathbf{r}(\mathbf{x}_t), \quad (16)$$

where $\nabla_{\mathbf{x}_t}\mathbf{r}(\mathbf{x}_t)$ is the gradient of the residual function that guides the update $\mathbf{x}_t$ towards regions where the observation $\mathbf{y}$ is more likely.

Algorithm 1 WS diffusion priors for imaging inverse problems

Require: T, $\mathbf{A}$, $\mathbf{y}$, $\{\beta_t\}_{t=0}^T$, $\mathbf{n}_{\boldsymbol{\theta}}$
1: Initialize $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
2: **for** $t = T$ to 0 **do**
3:      $\mathbf{x}_t' \leftarrow (2 - \sqrt{1 - \beta_t \Delta t})\mathbf{x}_t + \frac{\mathbf{n}_{\boldsymbol{\theta}}(\mathbf{x}_t, t)\Delta t}{2}$
4:      $\mathbf{x}_{t-1} \leftarrow \mathbf{x}_t' - \lambda_t \frac{\beta_t \mathbf{A}^H (\mathbf{y} - \mathbf{A}\mathbf{x}_t)}{2}$
5: **end for**
6: **return** $\mathbf{x}_0$

The reverse-time SDE framework aids inverse problems with correlated noise as the diffusion matrix $\mathbf{G}_t\mathbf{G}_t^\top$ preconditions the inverse measurement covariance in the likelihood score, when $\mathbf{G}_t\mathbf{G}_t^\top$ is designed to be proportional to the covariance matrix,

$$\mathbf{G}_t\mathbf{G}_t^\top \nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\mathbf{x}_t) \approx \underbrace{\mathbf{G}_t\mathbf{G}_t^\top \boldsymbol{\Sigma}_{\mathbf{y}}^{-1}}_{\propto \mathbf{I}} \nabla_{\mathbf{x}_t} \mathbf{r}(\mathbf{x}_t). \quad (17)$$

In designing our diffusion process, we set $\mathbf{G}_t\mathbf{G}_t^\top = \beta_t \mathbf{K}\mathbf{K}^\top + \gamma^2 \mathbf{I}$, where $\mathbf{K}\mathbf{K}^\top$ encompasses a large set of measurement noise covariances, and γ^2 is drawn uniformly between 0 and 1 in order to encourage the model to learn finer detailed features that reside in high spatial frequency subspaces.

In practice, a regularization term λ_t is important to balance the generative prior with the data likelihood. For proof-of-concept, we experiment with the likelihood-guided sampling via posterior score approximation in [22] due to its functional simplicity $\nabla_{\mathbf{x}_t} \log p(\mathbf{y} \mid \mathbf{x}_t) \approx \boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{A}^H(\mathbf{y} - \mathbf{A}\mathbf{x}_t)$. We also use the deterministic sampler of Eq. 7b. The resulting algorithm is shown in Algorithm 1.
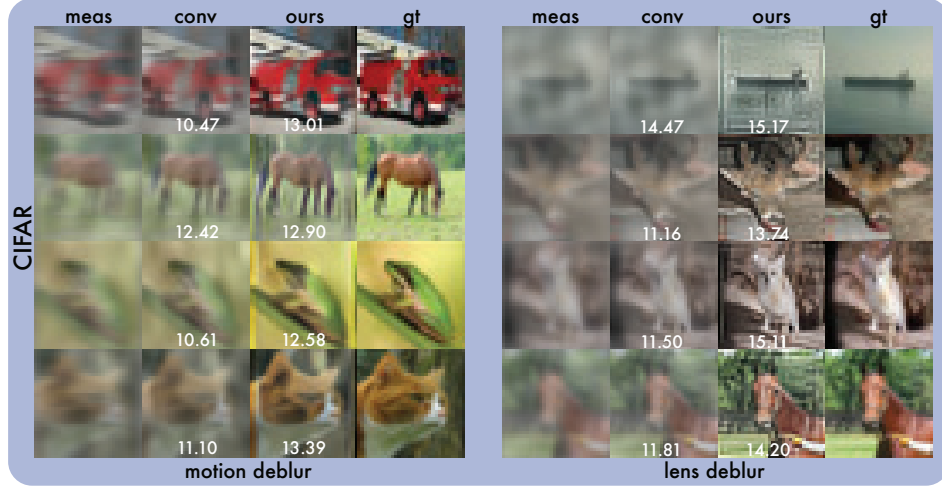
Figure 5: Motion and lens deblurring on CIFAR10 dataset with additive spatially correlated grayscale Gaussian noise of $\mathrm{std} = 2.5$ pixels. WS diffusion prior consistently removes correlated noise resulting in higher PSNR compared to DMs trained solely on isotropic Gaussian noise.

4 Experiments

4.1 Training details

For each dataset, we train two attention UNet models based on the architecture in [18] with three residual blocks in each downsampling layer, where one is for the conventional isotropic Gaussian SDE, and the other our anisotropic Gaussian SDE. We set the learning rate to $3\mathrm{e}^{-5}$ with a linear decay schedule. For CIFAR10 (32×32), the batch size is 128, for CelebA (64×64), the batch size is 16, and for CelebA-HQ (256×256), the batch size is 4. Models were trained on a single NVIDIA L40S GPU with 48GB of memory for two days. Our model is trained on the training sets of CIFAR-10 [29], CelebA (64×64) [36] and CelebA-HQ (256×256) [24] where $\mathbf{K}$ is a 2D Gaussian convolutional matrix characterized by an std. For CIFAR, the std that characterizes $\mathbf{K}$ is uniformly distributed between 0.1 and 3, between 0.1 and 5 for CelebA (64×64), and between 0.1 and 20 for CelebA-HQ (256×256) where $\mathrm{std} \leq 0.5$ equals the 2D delta function. The noise is also randomly grayscale or color with a 0.5 probability.

4.2 Imaging inverse problems with correlated noise

It is well-established that natural image spectra exhibit exponential decay [57], indicating the dominance of low-frequency components in representing images. When additive measurement noise occupies the same spectral subspace, especially at low frequencies, the computational imaging task becomes fundamentally more challenging. We show that our framework is beneficial as a generative prior for solving inverse problems with such structured noise by experimenting with a variety of computational imaging modalities that are known to be affected by structured noise.

The measurements in our experiments are corrupted by additive grayscale structured noise, designed to mimic real-world conditions frequently encountered in both computational photography—such as fog, haze, and atmospheric turbulence—and computational microscopy—including fluorescence background, laser speckle, and detector noise. We use Algorithm 1 with $T = 1000$ and $\beta_{min} = 0.01$ and $\beta_{max} = 20$ so that the SNR decays to 0 at $t = T$. For results with our WS prior, $\mathbf{x}_T$ was drawn from $\mathcal{N}(\mathbf{0}, \mathbf{K}\mathbf{K}^\top)$ with $\mathrm{std} = 3$ and $\mathrm{std} = 6$ for CIFAR and CelebA, respectively with grayscale color (all color channels have the same value). For conventional DM prior results $\mathbf{x}_T$ was drawn from $\mathcal{N}(\mathbf{0}, \mathbf{I})$. All evaluation was performed on unseen validation dataset sample images picked uniformly at random. The regularization parameter λ scales the magnitude of the likelihood step to be proportional to the magnitude of the prior step as was done by [22]. Line search was used to find an optimal λ that yielded a reconstruction with the highest PSNR, where PSNR is defined as
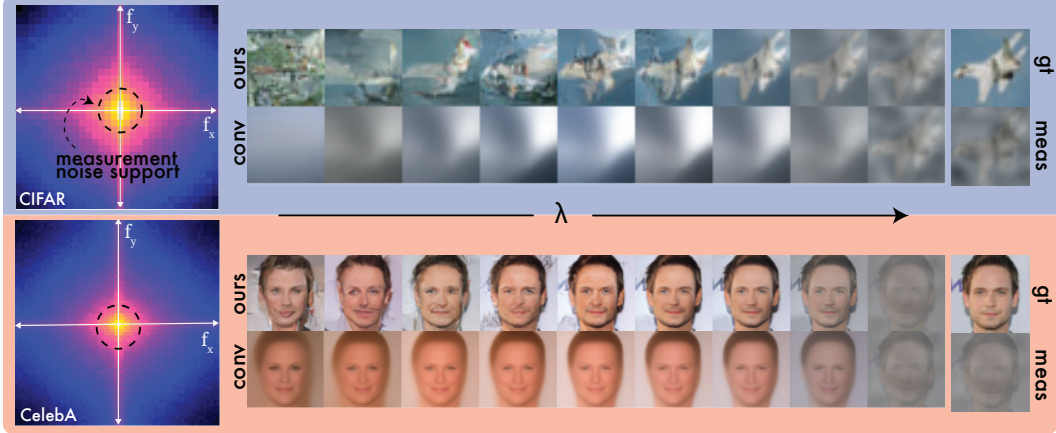
Figure 6: Effect of changing regularization parameter λ for denoising. The top figures are the average power spectral density of the images in CIFAR and CelebA, with a dotted red circle to denote the frequency support of the additive correlated noise. Changing the regularization weight λ for denoising affects the final reconstructions using our WS diffusion priors (top) and conventional diffusion priors (bottom). When λ is 0, it is equivalent to sampling form $p(\mathbf{x})$. As λ increases, the generative modeling effect is overpowered by the measurement fidelity term, that the reconstruction resembles the measurement $\mathbf{y}$.

$\text{PSNR} = 20 \cdot \log_{10}\left(\frac{1}{\text{MSE}}\right)$ where MSE is the mean squared error between the reconstruction and the ground truth.

Our results are demonstrated on a variety of computational imaging tasks such as imaging through fog, motion deblurring, lens deblurring, linear inverse scattering, and differential defocus. More details are in Appendix C.

Denoising correlated noise   To demonstrate the capabilities of our model as a generative prior for measurements corrupted by correlated noise, we explore the denoising problem and compare the results with that of a conventional score-based diffusion prior that was trained only on isotropic Gaussian forward diffusion. Fig. 2 shows the results on CIFAR and CelebA (64×64) test samples across a range of SNRs, where color is faithfully restored from fog-like corruption and likeness to the dataset is maintained due to the generative prior.

Generalize to different noise structures   Our model generalizes to different measurement noise covariance matrices with varying Gaussian noise distributions. Fig. 4 reveals that measurements corrupted by different distributions of spatially correlated Gaussian noise are restored with higher PSNR compared to conventional DM priors (conv). Conventional score-based priors change the higher level semantic features of the measurement, due to the model's inability to distinguish noisy features from target image features based on Fourier support. Specifically, the added correlated noise's low frequency support overlaps with that of the visual features in the data, seen in Fig. 6. This makes the reconstruction task more difficult.

Spectral inductive bias   WS diffusion priors more effectively distinguish structured noise from target features compared to conventional diffusion priors trained on isotropic Gaussian noise. As shown in Fig. 6, standard DMs tend to suppress high-frequency components in the measurement, assuming they originate from noise—a valid assumption only when the noise Fourier support extends beyond that of the data. For CIFAR, whose average signal spectrum extends beyond the noise's, this misclassification leads to undesired attenuation of image features. In contrast, for CelebA (64×64), where the average image spectrum lies within the noise support, conventional models better preserve image features.

WS DMs, trained on ensembles of Gaussian trajectories, learn to identify structured noise beyond simple spectral heuristics. This enables selective removal of low-frequency noise even when it spectrally overlaps with signal content, yielding improved denoising performance in the presence of correlated noise.

Computational imaging Using WS diffusion prior to solve inverse problems with non-identity forward operators outperforms traditional score-based diffusion priors in PSNR. Noticeably, our diffusion prior is able to maintain fidelity to the color distribution for restoring measurements corrupted by grayscale fog-like noise, while conventional score-based diffusion priors fail to remove the noise, as seen in Figs. 5 and 8 for deblurring inverse problems. Additional results for other imaging inverse problems on CIFAR, as well as on the CelebA (64×64) and CelebA-HQ (256×256) datasets, are presented in Appendix C and and Figs. 7, 8, 9, 10, and 11.

5 Conclusion

We introduced WS diffusion, a generalization of score-based methods that learns the Whitened Score, $\mathbf{G}_t \mathbf{G}_t^\top \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)$. This avoids noise covariance inversion, enabling arbitrary anisotropic Gaussian forward processes and bridging connections to FM. We demonstrate WS diffusion as robust generative priors for inverse problems involving correlated noise, common in computational imaging. Experiments consistently showed superior PSNR and visual reconstructions compared to conventional diffusion priors trained on isotropic noise, particularly in accurately handling structured noise while preserving image features. WS diffusion provides a principled approach for developing effective generative models tailored to structured noise, advancing their utility in computational imaging applications.

Limitations and future work A primary limitation of our approach lies in the computational cost of sampling. The current time discretization of the reverse-time SDE necessitates approximately 1000 denoising steps, which may be prohibitive for certain practical applications. Reducing the number of denoising steps through model distillation represents a promising direction for future work [44, 51].

Another limitation concerns the absence of an explicit mechanism to estimate the measurement noise covariance, which directly influences the specification of the diffusion matrix $\mathbf{G}_t$. A natural extension of this framework would involve parameterizing and learning $\mathbf{G}_t$ jointly with the model parameters. Such an approach would allow the diffusion process to adaptively capture data-dependent or task-specific noise structures, thereby enhancing the model's flexibility and representational capacity. This line of work connects to recent advances in vector-valued and multivariate diffusion models, which have demonstrated improved performance in scenarios characterized by complex or structured noise.

Finally, while our model exhibits strong performance as a denoising prior, additional research is needed for WS DMs to achieve competitive results in unconditional or conditional generation tasks. Promising directions include latent diffusion formulations and related techniques for improving generative efficiency and expressiveness [12, 42, 41].

References

[1] Michael S. Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants, 2023. URL `https://arxiv.org/abs/2209.15571`.

[2] Emma Alexander, Qi Guo, Sanjeev Koppal, Steven Gortler, and Todd Zickler. Focal Flow: Measuring Distance and Velocity with Defocus and Differential Motion. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, *Computer Vision – ECCV 2016*, pages 667–682, Cham, 2016. Springer International Publishing. ISBN 978-3-319-46487-9.

[3] Jeffrey Alido, Joseph Greene, Yujia Xue, Guorong Hu, Yunzhe Li, Mitchell Gilmore, Kevin J. Monk, Brett T. DiBenedictis, Ian G. Davison, and Lei Tian. Robust single-shot 3d fluorescence imaging in scattering media with a simulator-trained neural network. *Opt. Express*, 32(4):6241–6257, Feb 2024. doi: 10.1364/OE.514072. URL `https://opg.optica.org/oe/abstract.cfm?URI=oe-32-4-6241`.

[4] Brian D. O. Anderson. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3):313–326, May 1982. ISSN 0304-4149. doi: 10.1016/0304-4149(82)90051-5. URL `https://www.sciencedirect.com/science/article/pii/0304414982900515`.

[5] Benjamin Boys, Mark Girolami, Jakiw Pidstrigach, Sebastian Reich, Alan Mosca, and Omer Deniz Akyildiz. Tweedie moment projected diffusions for inverse problems, 2024. URL `https://openreview.net/forum?id=hDzjO41IOO`.

[6] Coleman Broaddus, Alexander Krull, Martin Weigert, Uwe Schmidt, and Gene Myers. Removing Structured Noise with Self-Supervised Blind-Spot Networks. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pages 159–163, April 2020. doi: 10.1109/ISBI45749.2020.9098336. URL `https://ieeexplore.ieee.org/document/9098336`. ISSN: 1945-8452.

[7] 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. URL `https://openreview.net/forum?id=nHESwXvxWK`.

[8] David Castanon. Reverse-time diffusion processes (Corresp.). *IEEE Transactions on Information Theory*, 28(6):953–956, November 1982. ISSN 1557-9654. doi: 10.1109/TIT.1982.1056571. URL `https://ieeexplore.ieee.org/abstract/document/1056571`. Conference Name: IEEE Transactions on Information Theory.

[9] Jooyoung Choi, Sungwon Kim, Yonghyun Jeong, Youngjune Gwon, and Sungroh Yoon. Ilvr: Conditioning method for denoising diffusion probabilistic models, 2021. URL `https://arxiv.org/abs/2108.02938`.

[10] Jooyoung Choi, Sungwon Kim, Yonghyun Jeong, Youngjune Gwon, and Sungroh Yoon. ILVR: Conditioning Method for Denoising Diffusion Probabilistic Models. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 14347–14356, October 2021. doi: 10.1109/ICCV48922.2021.01410. URL `https://ieeexplore.ieee.org/document/9711284`. ISSN: 2380-7504.

[11] Hyungjin Chung, Jeongsol Kim, Michael T. Mccann, Marc L. Klasky, and Jong Chul Ye. Diffusion Posterior Sampling for General Noisy Inverse Problems, May 2024. URL `http://arxiv.org/abs/2209.14687`. arXiv:2209.14687 [stat].

[12] Hyungjin Chung, Jong Chul Ye, Peyman Milanfar, and Mauricio Delbracio. Prompt-tuning latent diffusion models for inverse problems, 2024. URL `https://openreview.net/forum?id=ckzglrAMsh`.

[13] Giannis Daras, Mauricio Delbracio, Hossein Talebi, Alexandros G. Dimakis, and Peyman Milanfar. Soft Diffusion: Score Matching for General Corruptions, October 2022. URL `http://arxiv.org/abs/2209.05442`. arXiv:2209.05442 [cs].

[14] 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, September 2024. URL `http://arxiv.org/abs/2410.00083`. arXiv:2410.00083 [cs].

[15] Tim Dockhorn, Arash Vahdat, and Karsten Kreis. Score-based generative modeling with critically-damped langevin diffusion. *arXiv preprint arXiv:2112.07068*, 2021.

[16] 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.

[17] William Harvey, Saeid Naderiparizi, Vaden Masrani, Christian Weilbach, and Frank Wood. Flexible diffusion modeling of long videos. *Advances in neural information processing systems*, 35:27953–27965, 2022.

[18] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising Diffusion Probabilistic Models, December 2020. URL `http://arxiv.org/abs/2006.11239`. arXiv:2006.11239 [cs].

[19] Emiel Hoogeboom and Tim Salimans. Blurring diffusion models. *arXiv preprint arXiv:2209.05557*, 2022.

[20] Yuyang Hu, Albert Peng, Weijie Gan, Peyman Milanfar, Mauricio Delbracio, and Ulugbek S. Kamilov. Stochastic deep restoration priors for imaging inverse problems, 2025. URL `https://openreview.net/forum?id=O2aioX2Z2v`.

[21] David Huang, Eric A. Swanson, Charles P. Lin, Joel S. Schuman, William G. Stinson, Warren Chang, Michael R. Hee, Thomas Flotte, Kenton Gregory, Carmen A. Puliafito, and James G. Fujimoto. Optical Coherence Tomography. *Science*, 254(5035):1178–1181, November 1991. doi: 10.1126/science.1957169. URL `https://www.science.org/doi/10.1126/science.1957169`. Publisher: American Association for the Advancement of Science.

[22] Ajil Jalal, Marius Arvinte, Giannis Daras, Eric Price, Alexandros G. Dimakis, and Jonathan I. Tamir. Robust Compressed Sensing MRI with Deep Generative Priors, December 2021. URL `http://arxiv.org/abs/2108.01368`. arXiv:2108.01368 [cs].

[23] Thomas Jiralerspong, Berton Earnshaw, Jason Hartford, Yoshua Bengio, and Luca Scimeca. Shaping inductive bias in diffusion models through frequency-based noise control, 2025. URL `https://arxiv.org/abs/2502.10236`.

[24] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*, 2017.

[25] Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the Design Space of Diffusion-Based Generative Models. *Advances in Neural Information Processing Systems*, 35: 26565–26577, December 2022. URL `https://proceedings.neurips.cc/paper_files/paper/2022/hash/a98846e9d9cc01cfb87eb694d946ce6b-Abstract-Conference.html`.

[26] Tero Karras, Miika Aittala, Jaakko Lehtinen, Janne Hellsten, Timo Aila, and Samuli Laine. Analyzing and Improving the Training Dynamics of Diffusion Models, March 2024. URL `http://arxiv.org/abs/2312.02696`. arXiv:2312.02696 [cs].

[27] Bahjat Kawar, Gregory Vaksman, and Michael Elad. SNIPS: Solving Noisy Inverse Problems Stochastically, November 2021. URL `http://arxiv.org/abs/2105.14951`. arXiv:2105.14951 [eess].

[28] Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models, 2022. URL `https://arxiv.org/abs/2201.11793`.

[29] Alex Krizhevsky. Learning Multiple Layers of Features from Tiny Images.

[30] Tongyu Li, Jiabei Zhu, Yi Shen, and Lei Tian. Reflection-mode diffraction tomography of multiple-scattering samples on a reflective substrate from intensity images. *Optica*, 12(3):406–417, Mar 2025. doi: 10.1364/OPTICA.547372. URL `https://opg.optica.org/optica/abstract.cfm?URI=optica-12-3-406`.

[31] Huakang Lin and Chunling Luo. Demonstration of computational ghost imaging through fog. *Optics & Laser Technology*, 182:112075, April 2025. ISSN 0030-3992. doi: 10.1016/j.optlastec.2024.112075. URL `https://www.sciencedirect.com/science/article/pii/S0030399224015330`.

[32] Ruilong Ling, Waleed Tahir, Hsing-Ying Lin, Hakho Lee, and Lei Tian. High-throughput intensity diffraction tomography with a computational microscope. *Biomed. Opt. Express*, 9(5): 2130–2141, May 2018. doi: 10.1364/BOE.9.002130. URL `https://opg.optica.org/boe/abstract.cfm?URI=boe-9-5-2130`.

[33] Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow Matching for Generative Modeling, February 2023. URL `http://arxiv.org/abs/2210.02747`. arXiv:2210.02747 [cs].

[34] Yaron Lipman, Marton Havasi, Peter Holderrieth, Neta Shaul, Matt Le, Brian Karrer, Ricky T. Q. Chen, David Lopez-Paz, Heli Ben-Hamu, and Itai Gat. Flow matching guide and code, 2024. URL `https://arxiv.org/abs/2412.06264`.

[35] Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow, 2022. URL `https://arxiv.org/abs/2209.03003`.

[36] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proceedings of International Conference on Computer Vision (ICCV)*, December 2015.

[37] 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. URL `https://openreview.net/forum?id=1YO4EE3SPB`.

[38] Naoki Murata, Koichi Saito, Chieh-Hsin Lai, Yuhta Takida, Toshimitsu Uesaka, Yuki Mitsufuji, and Stefano Ermon. Gibbsddrm: A partially collapsed gibbs sampler for solving blind inverse problems with denoising diffusion restoration, 2023. URL `https://arxiv.org/abs/2301.12686`.

[39] Leonhard Möckl, Anish R. Roy, Petar N. Petrov, and W. E. Moerner. Accurate and rapid background estimation in single-molecule localization microscopy using the deep neural network BGnet. *Proceedings of the National Academy of Sciences*, 117(1):60–67, January 2020. doi: 10.1073/pnas.1916219117. URL `https://www.pnas.org/doi/full/10.1073/pnas.1916219117`. Publisher: Proceedings of the National Academy of Sciences.

[40] Kushagra Pandey, Maja Rudolph, and Stephan Mandt. Efficient integrators for diffusion generative models. *arXiv preprint arXiv:2310.07894*, 2023.

[41] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-Resolution Image Synthesis with Latent Diffusion Models, April 2022. URL `http://arxiv.org/abs/2112.10752`. arXiv:2112.10752 [cs].

[42] Litu Rout, Yujia Chen, Abhishek Kumar, Constantine Caramanis, Sanjay Shakkottai, and Wen-Sheng Chu. Beyond First-Order Tweedie: Solving Inverse Problems using Latent Diffusion, December 2023. URL `http://arxiv.org/abs/2312.00852`. arXiv:2312.00852 [cs].

[43] 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. URL `https://openreview.net/forum?id=XKBFdYwfRo`.

[44] Tim Salimans and Jonathan Ho. Progressive Distillation for Fast Sampling of Diffusion Models, June 2022. URL `http://arxiv.org/abs/2202.00512`. arXiv:2202.00512 [cs].

[45] Raghav Singhal, Mark Goldstein, and Rajesh Ranganath. Where to diffuse, how to diffuse, and how to get back: Automated learning for multivariate diffusions. *arXiv preprint arXiv:2302.07261*, 2023.

[46] 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. URL `https://openreview.net/forum?id=j8hdRqOUhN`.

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

[48] Yang Song and Stefano Ermon. Generative Modeling by Estimating Gradients of the Data Distribution, October 2020. URL `http://arxiv.org/abs/1907.05600`. arXiv:1907.05600 [cs].

[49] Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-Based Generative Modeling through Stochastic Differential Equations, February 2021. URL `http://arxiv.org/abs/2011.13456`. arXiv:2011.13456 [cs].

[50] 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. URL `https://openreview.net/forum?id=vaRCHVj0uGI`.

[51] Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency Models, May 2023. URL `http://arxiv.org/abs/2303.01469`. arXiv:2303.01469 [cs].

[52] Peng Sun, Yi Jiang, and Tao Lin. Unified continuous generative models, 2025. URL `https://arxiv.org/abs/2505.07447`.

[53] Yu Sun, Zihui Wu, Yifan Chen, Berthy T. Feng, and Katherine L. Bouman. Provable Probabilistic Imaging Using Score-Based Generative Priors. *IEEE Transactions on Computational Imaging*, 10:1290–1305, 2024. ISSN 2333-9403. doi: 10.1109/TCI.2024.3449114. URL `https://ieeexplore.ieee.org/abstract/document/10645293`.

[54] Simo Särkkä and Arno Solin. *Applied Stochastic Differential Equations*. Institute of Mathematical Statistics Textbooks. Cambridge University Press, Cambridge, 2019. ISBN 978-1-316-51008-7. doi: 10.1017/9781108186735. URL `https://www.cambridge.org/core/books/applied-stochastic-differential-equations/6BB1B8B0819F8C12616E4A0C78C29EAA`.

[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. URL `https://openreview.net/forum?id=6TxBxqNME1Y`.

[56] Takafumi Tsukui, Satoru Iguchi, Ikki Mitsuhashi, and Kenichi Tadaki. Estimating the statistical uncertainty due to spatially correlated noise in interferometric images. *Journal of Astronomical Telescopes, Instruments, and Systems*, 9(1):018001, 2023. doi: 10.1117/1.JATIS.9.1.018001. URL `https://doi.org/10.1117/1.JATIS.9.1.018001`. Publisher: SPIE.

[57] A. van der Schaaf and J.H. van Hateren. Modelling the power spectra of natural images: Statistics and information. *Vision Research*, 36(17):2759–2770, 1996. ISSN 0042-6989. doi: https://doi.org/10.1016/0042-6989(96)00002-8. URL `https://www.sciencedirect.com/science/article/pii/0042698996000028`.

[58] Pascal Vincent. A Connection Between Score Matching and Denoising Autoencoders. *Neural Computation*, 23(7):1661–1674, July 2011. ISSN 0899-7667, 1530-888X. doi: 10.1162/NECO_a_00142. URL `https://direct.mit.edu/neco/article/23/7/1661-1674/7677`.

[59] Laura Waller, Lei Tian, and George Barbastathis. Transport of Intensity phase-amplitude imaging with higher order intensity derivatives. *Optics Express*, 18(12):12552–12561, June 2010. ISSN 1094-4087. doi: 10.1364/OE.18.012552. URL `https://opg.optica.org/oe/abstract.cfm?uri=oe-18-12-12552`. Publisher: Optica Publishing Group.

[60] 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, 2024. URL `https://arxiv.org/abs/2405.16749`.

[61] Zihui Wu, Yu Sun, Yifan Chen, Bingliang Zhang, Yisong Yue, and Katherine L. Bouman. Principled probabilistic imaging using diffusion models as plug-and-play priors, 2024. URL `https://arxiv.org/abs/2405.18782`.

[62] Ali Zafari and Shirin Jalali. Bayesian Despeckling of Structured Sources, January 2025. URL `http://arxiv.org/abs/2501.11860`. arXiv:2501.11860 [cs].

[63] Bingliang Zhang, Wenda Chu, Julius Berner, Chenlin Meng, Anima Anandkumar, and Yang Song. Improving Diffusion Inverse Problem Solving with Decoupled Noise Annealing, December 2024. URL `http://arxiv.org/abs/2407.01521`. arXiv:2407.01521 [cs].

[64] Xingguang Zhang, Zhiyuan Mao, Nicholas Chimitt, and Stanley H. Chan. Imaging Through the Atmosphere Using Turbulence Mitigation Transformer. *IEEE Transactions on Computational Imaging*, 10:115–128, 2024. ISSN 2333-9403. doi: 10.1109/TCI.2024.3354421. URL `https://ieeexplore.ieee.org/abstract/document/10400926`.

[65] 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, January 2025. URL `http://arxiv.org/abs/2405.18816`. arXiv:2405.18816 [cs].

[66] Hongkai Zheng, Wenda Chu, Bingliang Zhang, Zihui Wu, Austin Wang, Berthy T. Feng, Caifeng Zou, Yu Sun, Nikola Kovachki, Zachary E. Ross, Katherine L. Bouman, and Yisong Yue. Inversebench: Benchmarking plug-and-play diffusion priors for inverse problems in physical sciences, 2025. URL `https://arxiv.org/abs/2503.11043`.

[67] Yuanzhi Zhu, Kai Zhang, Jingyun Liang, Jiezhang Cao, Bihan Wen, Radu Timofte, and Luc Van Gool. Denoising diffusion models for plug-and-play image restoration, 2023. URL `https://arxiv.org/abs/2305.08995`.

A Denoising Whitened Score matching

Lemma A.1. *(Generalized Tweedie's formula for non-diagonal covariance). Let:*

$$\mathbf{x}_t = \alpha_t \mathbf{x}_0 + \mathbf{z}, \tag{18}$$

where $\mathbf{x}_0 \sim p(\mathbf{x})$ and $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$. Then

$$\alpha_t \mathbb{E}[\mathbf{x}_0 \mid \mathbf{x}_t] = \mathbf{x}_t + \mathbf{\Sigma} \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t) \tag{19}$$

Proof.

$$
\begin{aligned}
\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t) &= \frac{\nabla_{\mathbf{x}_t} p_t(\mathbf{x}_t)}{p_t(\mathbf{x}_t)} \\
&= \frac{1}{p_t(\mathbf{x}_t)} \nabla_{\mathbf{x}_t} \int p_t(\mathbf{x}_t, \mathbf{x}_0) \, \mathrm{d}\mathbf{x}_0 && \text{(de-marginalize joint distribution)} \\
&= \frac{1}{p_t(\mathbf{x}_t)} \nabla_{\mathbf{x}_t} \int p_t(\mathbf{x}_t \mid \mathbf{x}_0) p_0(\mathbf{x}_0) \, \mathrm{d}\mathbf{x}_0 && \text{(factor joint via Bayes rule)} \\
&= \frac{1}{p_t(\mathbf{x}_t)} \int \nabla_{\mathbf{x}_t} p_t(\mathbf{x}_t \mid \mathbf{x}_0) p_0(\mathbf{x}_0) \, \mathrm{d}\mathbf{x}_0 && \text{(move gradient inside integral)} \\
&= \frac{1}{p_t(\mathbf{x}_t)} \int p_t(\mathbf{x}_t \mid \mathbf{x}_0) \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t \mid \mathbf{x}_0) p_0(\mathbf{x}_0) \, \mathrm{d}\mathbf{x}_0 && \text{(use identity } \nabla f = f \nabla \log f) \\
&= \int p_0(\mathbf{x}_0 \mid \mathbf{x}_t) \, \mathbf{\Sigma}^{-1} (\alpha_t \mathbf{x}_0 - \mathbf{x}_t) \, \mathrm{d}\mathbf{x}_0 && \text{(Gaussian conditional score)} \\
&= \mathbf{\Sigma}^{-1} \left(\alpha_t \, \mathbb{E}[\mathbf{x}_0 \mid \mathbf{x}_t] - \mathbf{x}_t \right) && \text{(expectation under posterior)}
\end{aligned}
$$

$\square$

Theorem A.2. *(Denoising Whitened Score matching)*
Our Whitened Score matching loss function, Eq. 13, copied here as:

$$\mathbb{E}_{t \sim U(0,1], \mathbf{x}_t \sim p(\mathbf{x}_t \mid \mathbf{x}_0), \mathbf{x}_0 \sim p(\mathbf{x})} \left\{ \|\mathbf{n}_\theta(\mathbf{x}_t, t) - \mathbf{G}_t \mathbf{G}_t^\top \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t \mid \mathbf{x}_0)\|_2^2 \right\} \tag{20}$$

is a denoising objective that uses the conditional probability to estimate $\mathbf{G}_t \mathbf{G}_t^\top \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)$. Here we prove that our loss function results in an estimator for $\mathbf{G}_t \mathbf{G}_t^\top \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)$.

Proof. Let $p(\mathbf{x}_t \mid \mathbf{x}_0)$ denote the Gaussian probability transition kernel associated with the forward-time SDE in Eq. 8. For a linear SDE in $\mathbf{x}_t$, the covariance $\mathbf{\Sigma}_t$ of the transition kernel is a scalar multiple of twice the diffusion matrix, $\mathbf{G}_t \mathbf{G}_t^\top$ ([54]) provided the initial conditions of $\boldsymbol{\mu}(0) = \mathbf{x}_0$ and $\mathbf{\Sigma}(0) = \mathbf{0}$ for $p(\mathbf{x}_t \mid \mathbf{x}_0)$:

$$\mathbf{\Sigma}_t = c \mathbf{G}_t \mathbf{G}_t^\top. \tag{21}$$

The Minimum Mean Squared Estimator (MMSE) $\mathbb{E}[\mathbf{x}_0 \mid \mathbf{x}_t]$ is achieved through optimizing the least squares objective:

$$\min_{\boldsymbol{\theta}} \mathbb{E}_{\mathbf{x}_t \sim p_t(\mathbf{x}_t \mid \mathbf{x}_0), \mathbf{x}_0 \sim p(\mathbf{x})} \left[\|\mathbf{h}_{\boldsymbol{\theta}}(\mathbf{x}_t) - \mathbf{x}_0\|_2^2 \right], \tag{22}$$

such that $\mathbf{h}_{\boldsymbol{\theta}^*}(\mathbf{x}_t) = \mathbb{E}[\mathbf{x}_0 \mid \mathbf{x}_t]$, for optimal network parameters $\boldsymbol{\theta}^*$. Tweedie's formula from Lemma A.1 gives us that,

$$\alpha_t \mathbf{h}_{\boldsymbol{\theta}^*}(\mathbf{x}_t) = \mathbf{x}_t + \mathbf{\Sigma} \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t). \tag{23}$$

Parameterizing $\mathbf{h}_{\boldsymbol{\theta}}(\mathbf{x}_t)$ as

$$\mathbf{h}_{\boldsymbol{\theta}}(\mathbf{x}_t) = \frac{\mathbf{x}_t + c \mathbf{n}_{\boldsymbol{\theta}}(\mathbf{x}_t, t)}{\alpha_t}, \tag{24}$$

with scalars $c, \alpha_t \in \mathbb{R}$ and $\mathbf{n}_{\boldsymbol{\theta}}(\mathbf{x}_t, t)$ our $\mathbf{G}_t \mathbf{G}_t^\top \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)$ model, the MMSE objective in Eq. 22 becomes

$$\min_{\boldsymbol{\theta}} \mathbb{E}_{\mathbf{x}_t \sim p_t(\mathbf{x}_t \mid \mathbf{x}_0), \mathbf{x}_0 \sim p(\mathbf{x})} \left[\|\mathbf{x}_t + c \mathbf{n}_{\boldsymbol{\theta}}(\mathbf{x}_t, t) - \alpha_t \mathbf{x}_0\|_2^2 \right], \tag{25}$$

which is equivalent to our objective in Eq. 20 through the closed form expression of $\mathbf{G}_t \mathbf{G}_t^\top \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t \mid \mathbf{x}_0)$ given in Eq. 12. Finding the optimal $\boldsymbol{\theta}^*$ implies

$$\mathbf{h}_{\boldsymbol{\theta}^*}(\mathbf{x}_t) = \mathbb{E}[\mathbf{x}_0 \mid \mathbf{x}_t] = \frac{\mathbf{x}_t + c\mathbf{n}_{\boldsymbol{\theta}^*}(\mathbf{x}_t, t)}{\alpha_t} = \frac{\mathbf{x}_t + \boldsymbol{\Sigma}\nabla_{\mathbf{x}_t}\log p_t(\mathbf{x}_t)}{\alpha_t}. \tag{26}$$

which proves that our model $\mathbf{n}_{\boldsymbol{\theta}}(\mathbf{x}_t, t)$ learns $\mathbf{G}_t\mathbf{G}_t^\top\nabla_{\mathbf{x}_t}\log p_t(\mathbf{x}_t)$ with objective Eq. 20.

$\square$

B Flow matching in SDE

Consider the probability path
$$p_t = \mathcal{N}(\mathbf{x}_t \mid \boldsymbol{\mu}_t(\mathbf{x}_0), \boldsymbol{\Sigma}_t(\mathbf{x}_0)), \tag{27}$$
and the corresponding continuous normalizing flow:

$$\phi_t(\mathbf{x}_0) = \boldsymbol{\mu}_t(\mathbf{x}_0) + \boldsymbol{\Sigma}_t^{\frac{1}{2}}(\mathbf{x}_0), \tag{28}$$

where we define $\boldsymbol{\Sigma}_t^{\frac{1}{2}}(\mathbf{x}_0)$ such that $\text{Cov}[\phi_t(\mathbf{x}_0)] = \boldsymbol{\Sigma}_t(\mathbf{x}_0) = \boldsymbol{\Sigma}_t^{\frac{1}{2}}(\mathbf{x}_0)(\boldsymbol{\Sigma}_t^{\frac{1}{2}}(\mathbf{x}_0))^\top$ and the initial condition $\boldsymbol{\mu}_0(\mathbf{x}_0) = \mathbf{x}_0$. This probability path is equivalent to the probability transition kernel in Eq. 9 defined by a linear SDE Eq. 8 with drift coefficient $\mathbf{F}_t$ and diffusion matrix $\mathbf{G}_t$. Therefore we may attain the time derivatives of the mean and covariance functions using Fokker-Planck (see Eqs. 6.2 in [54]) expressed as

$$\boldsymbol{\Sigma}_t'(\mathbf{x}_0) = 2\mathbf{F}_t\boldsymbol{\Sigma}_t(\mathbf{x}_0) + \mathbf{G}_t\mathbf{G}_t^\top, \tag{29}$$
$$\boldsymbol{\mu}_t'(\mathbf{x}_0) = \mathbf{F}_t\boldsymbol{\mu}_t(\mathbf{x}_0). \tag{30}$$

The FM conditional vector field for Gaussian probability paths is
$$\mathbf{u}_t(\mathbf{x}_t \mid \mathbf{x}_0) = \boldsymbol{\Sigma}_t'(\mathbf{x}_0)\boldsymbol{\Sigma}_t^{-1}(\mathbf{x}_0)(\mathbf{x}_t - \boldsymbol{\mu}_t(\mathbf{x}_0)) + \boldsymbol{\mu}_t'(\mathbf{x}_0). \tag{31}$$

Plugging in Eqs. 30 and 29 into Eq. 31 we have
$$\mathbf{u}_t(\mathbf{x}_t \mid \mathbf{x}_0) = \boldsymbol{\Sigma}_t'(\mathbf{x}_0)\boldsymbol{\Sigma}_t^{-1}(\mathbf{x}_0)(\mathbf{x}_t - \boldsymbol{\mu}_t(\mathbf{x}_0)) + \boldsymbol{\mu}_t'(\mathbf{x}_0) \tag{32}$$
$$= (2\mathbf{F}_t + \mathbf{G}_t\mathbf{G}_t^\top\boldsymbol{\Sigma}_t^{-1}(\mathbf{x}_0)(\mathbf{x}_t - \alpha_t\mathbf{x}_0)) + \mathbf{F}_t\boldsymbol{\mu}_t(\mathbf{x}_0) \tag{33}$$
$$= 2\mathbf{F}_t\mathbf{x}_t - \mathbf{F}_t\alpha_t\mathbf{x}_0 + \mathbf{G}_t\mathbf{G}_t^\top\boldsymbol{\Sigma}_t^{-1}(\mathbf{x}_0)(\mathbf{x}_t - \alpha_t\mathbf{x}_0) \tag{34}$$
$$= \mathbf{F}_t(2\mathbf{x}_t - \alpha_t\mathbf{x}_0) - \mathbf{G}_t\mathbf{G}_t^\top\nabla_{\mathbf{x}_t}\log p(\mathbf{x}_t \mid \mathbf{x}_0). \tag{35}$$

C Imaging inverse problems

Imaging through fog/turbulence   We simplify imaging through fog/turbulence as a denoising problem for correlated noise which we achieve by setting the imaging system $\mathbf{A} = \mathbf{I}$. Specifically, we demonstrate our model for grayscale low-pass filtered white Gaussian noise, where $\mathbf{K}$ is a 2D Gaussian kernel characterized by a std.

In our method, the model is trained to denoise correlated noise, and is able to distinguish target image features from anisotropic Gaussian noise features, even though both may share similar spatial frequency support. In contrast, the conventional DM trained to denoise only isotropic Gaussian noise, removes the images features for low enough λ, as it mistakes the correlated additive noise for the target features. As seen in Fig. 6, there are more image features on average outside the additive noise support for CIFAR, and for CelebA, the support of the image features are more closely overlapped with the noise support. Additional result on the CelebA-HQ (256×256) dataset is shown in Fig. 7 .

Motion deblurring   Motion blur is a common image degradation in computational photography. We experiment with a spatially invariant horizontal motion blur kernel of five pixels for CIFAR and seven for CelebA. The additive correlated noise is Gaussian-filtered grayscale WGN with a circular kernel of $\text{std} = 2.5$ pixels for CIFAR and 5 pixels for CelebA, each with $\text{SNR} = 0.493$. The result is shown in Figs. 5 and 8.
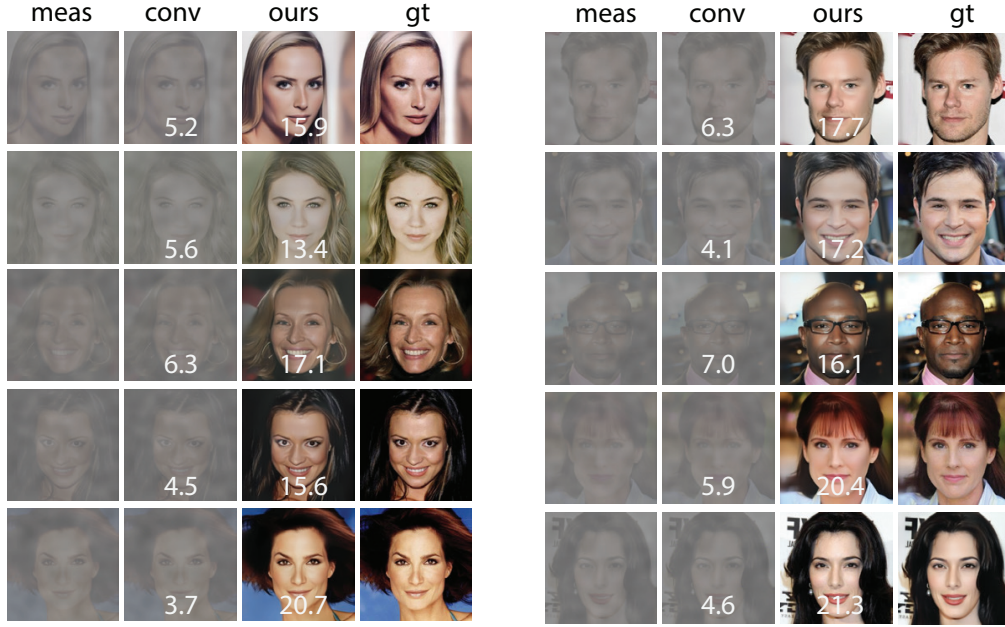
Figure 7: Denoising correlated noise on CelebA-HQ (256×256). We benchmark our WS DM trained on anisotropic Gaussian noise with the conventional DM (conv) trained on isotropic Gaussian noise. Results for measurements $\mathbf{y}$ with additive grayscale Gaussian noise of $\mathrm{std} = 5$ pixels.

Figure 8: Motion and lens deblurring on CIFAR10 dataset with additive spatially correlated grayscale Gaussian noise of $\mathrm{std} = 2.5$ pixels. Our diffusion prior is able to consistently remove correlated noise resulting in superior PSNR compared to DMs trained solely on isotropic Gaussian noise.
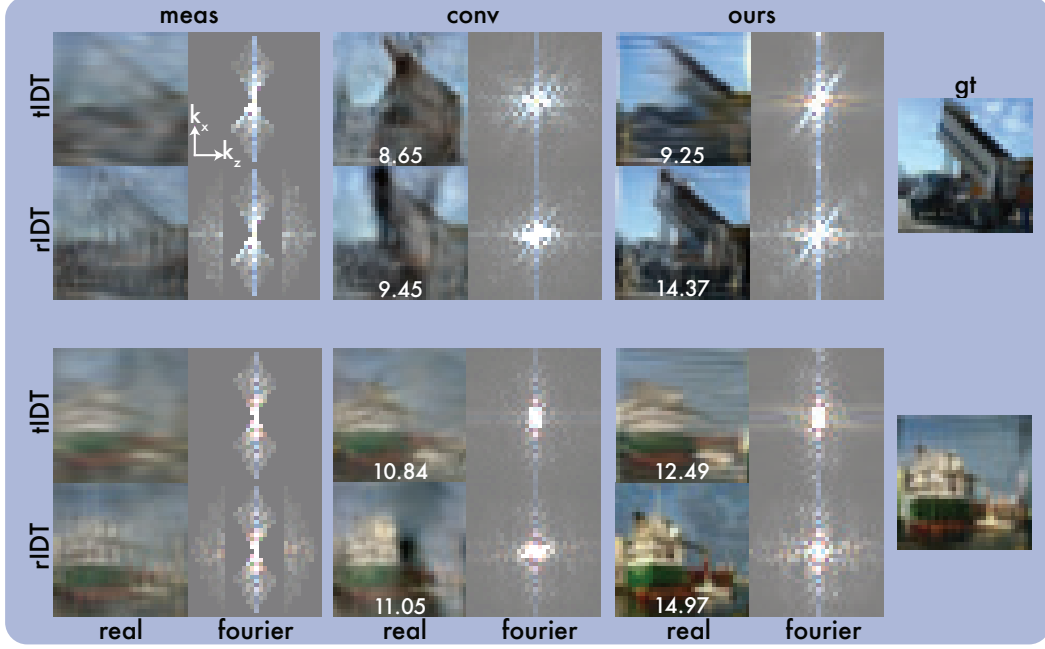
Figure 9: Linear inverse scattering CIFAR

Lens deblurring Lens blur is the loss of high spatial frequency information as a result of light rays being focused imperfectly due to the finite aperture size, causing rays from a point source to spread over a region in the image plane rather than converging to a single point. This can be effectively modeled as a convolution between a circular Gaussian kernel and the clean image. In Figs. 5 and 8, we demonstrate our WS diffusion prior on lens deblurring with a Gaussian blur kernel of $\mathrm{STD} = 0.8$ and 1.0 for CIFAR and CelebA, respectively. The additive correlated noise is Gaussian-filtered grayscale WGN with a circular kernel of $\mathrm{std} = 2.5$ pixels for CIFAR and 5 pixels for CelebA, each with $\mathrm{SNR} = 0.810$.

Linear inverse scattering Inverse scattering is a prevalent direction in optical imaging, to recover the permittivity field from measurements under angled illumination. Intensity diffraction tomography (IDT) is a powerful computational microscopy technique that can recover 3D refractive index distribution given a set of 2D measurements. The model can be linearized using the first Born approximation ([32]):

$$\mathbf{u}(\mathbf{r}) = \mathbf{u}_i(\mathbf{r}) + \int \mathbf{u}_i(\mathbf{r}')\mathbf{V}(\mathbf{r}')\mathbf{G}(\mathbf{r} - \mathbf{r}')\mathrm{d}\mathbf{r}', \tag{36}$$

for the field at the measurement plane $\mathbf{u}(\mathbf{r})$ and incident field $\mathbf{u}_i(\mathbf{r})$. The scattering potential $\mathbf{V}(\mathbf{r}) = \frac{1}{4\pi}k_0^2\Delta\epsilon(\mathbf{r})$ with permittivity contrast $\Delta\epsilon(\mathbf{r}) = \epsilon(\mathbf{r}) - \epsilon_0$ between the sample $\epsilon(\mathbf{r})$ and surrounding medium ϵ, and wavenumber $k_0 = \frac{2\pi}{\lambda}$ for illumination wavelength λ. Green's function $\mathbf{G}(\mathbf{r}) = \frac{\exp(ik|\mathbf{r}|)}{\mathbf{r}}$ where $k = \sqrt{\epsilon_0}k_0$.

When the illumination is transmissive, referred to as transmission intensity diffraction tomography (tIDT), meaning that the light passes through the sample, the linear operator, $\mathbf{A}$, results in a mask in the shape of a cross section of a torus that attenuates Fourier coefficients as seen in Figs. 9 and 10, leading to the well-known "missing cone" problem.

In reflection IDT (rIDT), placing the sample object on a specular mirror substrate causes light to reflect towards the camera, enabling the capture of additional axial frequency components and partially filling the missing cone, shown in Figs. 9 and 10.
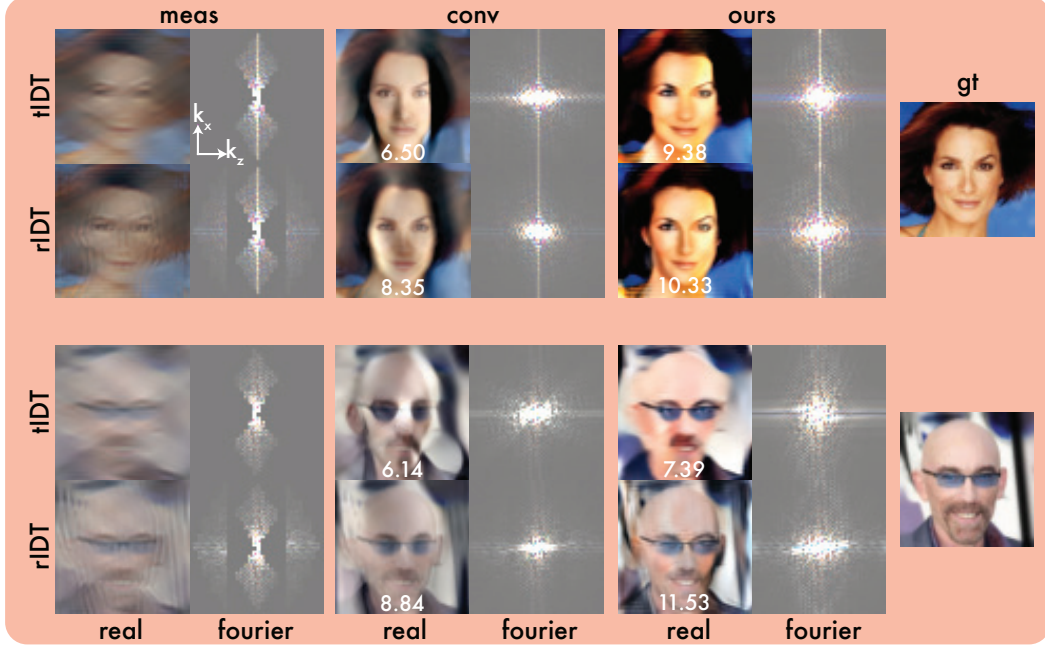
19

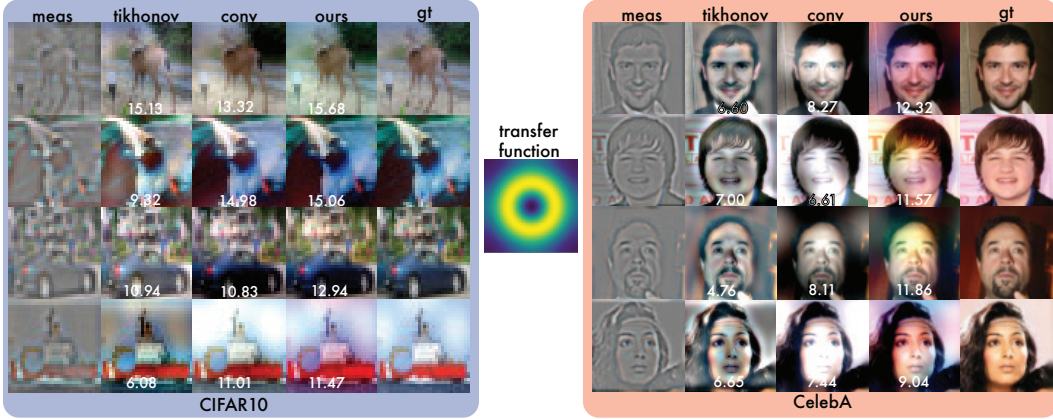Figure 10: Linear inverse scattering CelebA (64×64)

Figure 11: Laplace imaging, Transport of Intensity (TIE)

We experiment with both tIDT and rIDT with the measurements corrupted by low-pass filtered grayscale WGN to mimic background noise common in microscopy. The grayscale noise is similarly Gaussian-filtered with $\mathrm{std} = 2.5$ and 5, for CIFAR and CelebA, respectively with $\mathrm{SNR} = 0.632$.

Differential defocus Differential defocus is a computational imaging technique that aims to recover the depth map from a series of defocused measurements ([2]). The linear operator can be realized with a 2D Laplacian kernel, bandpassing mid-frequency components. In computational microscopy, this is also known as transport of intensity imaging ([59]) to recover the phase and amplitude of an object.

We demonstrate our framework on the differential defocus problem in Fig. 11. The noise is again grayscale WGN filtered with Gaussian kernels of $\mathrm{std} = 2.5$ and 5, for CIFAR and CelebA, respectively with $\mathrm{SNR} = 12.91$. We also compare with a Tikhonov regularization, which is an L_2 norm prior on the object to constrain the energy of the reconstruction.
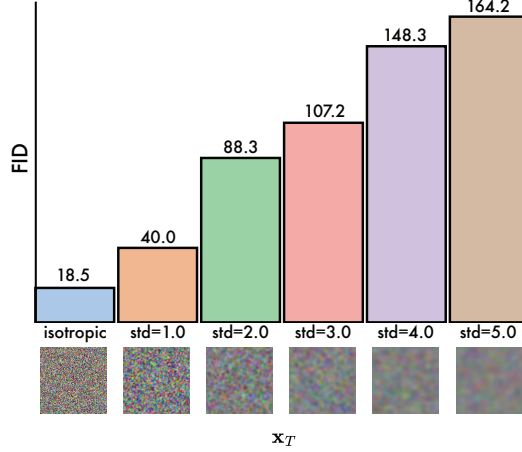
Figure 12: FID scores for different WS DMs trained on different std_{max}.

D   Generative modeling

We also train 6 different models where we vary the std of the maximum value of the Gaussian blur kernel, $\mathbf{K}$, from a delta function (isotropic Gaussian noise) to $\text{std}_{max} = 5$ pixels. During training, $\mathbf{K}$ varies uniformly from $\text{std} = 0.1$ to $\text{std} = \text{std}_{max}$.

Novel samples are produced by solving the probability flow ODE, replacing $\mathbf{G}_t \mathbf{G}_t^\top \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)$ with our optimized model $\mathbf{n}_\theta(\mathbf{x}_t, t)$ using Euler-Maruyama discretization with $T = 1000$ and $\beta_{min} = 0.01$ and $\beta_{max} = 20$. The initial noise condition, $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_{\text{std}_{max}} \mathbf{K}_{\text{std}_{max}}^\top)$. The Fréchet Inception Distance (FID) scores decrease as the spatial correlation range increases as seen in Fig. 12.

While WS DMs perform well as generative denoising priors for inverse problems, we leave to future work further investigation on their generative capabilities.

E   Forward Consistency Loss

The model $\mathbf{n}_\theta(\mathbf{x}_t, t)$ is trained to approximate the scaled noise component introduced in the forward stochastic differential equation (SDE) that perturbs $\mathbf{x}_0$ to yield $\mathbf{x}_t$. Accordingly, the score function $\mathbf{G}_t \mathbf{G}_t^\top \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t \mid \mathbf{x}_0)$ in Eq. 12 can be substituted with the model prediction. To enforce consistency with the forward diffusion process, we introduce an auxiliary loss term defined as:

$$L_2 = \mathbb{E}_{t \sim \mathcal{U}(0,1], \mathbf{x}_0 \sim p(\mathbf{x}), \mathbf{x}_t \sim p(\mathbf{x}_t | \mathbf{x}_0)} \left\{ \left\| \mathbf{x}_0 - \frac{\beta_t \mathbf{x}_t + (1 - \alpha_t^2) \mathbf{n}_\theta(\mathbf{x}_t, t)}{\beta_t \alpha_t} \right\|_2^2 \right\}. \qquad (37)$$

The term inside the expectation represents a reconstruction of $\mathbf{x}_0$ based on the noisy sample $\mathbf{x}t$ and the model prediction $\mathbf{n}_\theta$. Minimizing L_2 encourages the model to remain faithful to the generative process defined by the forward SDE.

Empirically, we find that including L_2 as an auxiliary objective—weighted equally with the primary loss term L—leads to improved training stability and faster convergence.

NeurIPS Paper Checklist

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 clearly state the core innovation—learning a whitened score to handle anisotropic Gaussian processes—and claims align well with the theoretical development and experimental validation presented in the paper.

   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 its suboptimal performance in generative tasks.

   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: [Yes]

Justification: We provide proofs in the appendix that clearly state assumptions and step-by-step proofs.

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: All details related to training and hyperparameter selection are mentioned in the Experiments section.

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: [Yes]

Justification: A link to the repository is provided and the datasets are public datasets.

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: All training and test details are described in the Experiment section.

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: [NA]

Justification: Experiments conducted are image reconstruction, which cannot include error bars.

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: The details of the compute resources are described in the Experiment section.

   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: Yes, the research adheres to the NeurIPS Code of Ethics, as it presents a methodological contribution without involving human subjects, sensitive data, or foreseeable societal harms.

   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: Not applicable—this paper presents a theoretical and algorithmic contribution in diffusion modeling without direct deployment implications, and does not discuss societal impacts.

    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: Not applicable, as the paper does not involve the release of models or data with high risk for misuse.

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: CIFAR-10 and CelebA datasets are cited.

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: No new assets are introduced.

    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: This work does not use crowdsourcing or 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: There are no 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 rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: LLMs are not used

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.