Score-Based Diffusion Models as Principled Priors for Inverse Imaging

Berthy T. Feng[1]   Jamie Smith[2]   Michael Rubinstein[2]   Huiwen Chang[2]
Katherine L. Bouman[1]   William T. Freeman[2]
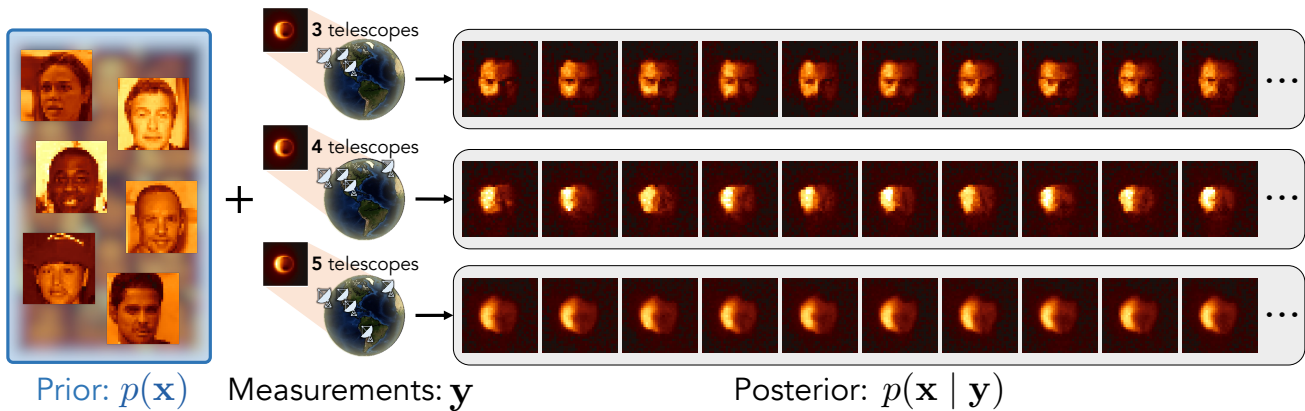[1]California Institute of Technology   [2]Google Research

Figure 1. A score-based prior is a **hyperparameter-free, probabilistic prior** that is also **expressive and data-driven**. Paired with a set of measurements, the prior can be used for principled inference of a full posterior. In this example, a score-based prior was trained on face images ("Prior" shows samples from the learned prior). The inverse problem is interferometic imaging of a synthetic black hole. We simulated interferometric measurements from the actual telescope array used to capture the first black-hole image [17] and sampled images from the posterior via variational inference. From the top to bottom row, the posterior stably moves away from the prior given more constraining measurements. With measurements from only three telescopes, the posterior shows strong influence from the prior and contains images resembling faces that are brighter on the left half. As more telescopes (measurements) are added, the posterior reveals the ring-like structure of the underlying image. Our framework finds the proper relative strengths of the prior and measurements automatically.

Abstract

Priors are essential for reconstructing images from noisy and/or incomplete measurements. The choice of the prior determines both the quality and uncertainty of recovered images. We propose turning score-based diffusion models into principled image priors ("score-based priors") for analyzing a posterior of images given measurements. Previously, probabilistic priors were limited to handcrafted regularizers and simple distributions. In this work, we empirically validate the theoretically-proven probability function of a score-based diffusion model. We show how to sample from resulting posteriors by using this probability function for variational inference. Our results, including experiments on denoising, deblurring, and interferometric imaging, suggest that score-based priors enable principled inference with a sophisticated, data-driven image prior.

1. Introduction

Priors are crucial for solving inverse problems in computational imaging, which tend to be ill-posed due to noisy and limited sensors. When many different images agree with observed measurements, a prior helps constrain solutions according to desired image statistics. How to incorporate a sophisticated prior, however, is not straightforward. Our work addresses the problem of incorporating a rich prior into principled approaches to inverse problems.

Previous work poses a tradeoff: using principled methods requires simple priors, while using deep-learned priors precludes precise analysis. On the principled side, Bayesian-inference methods model the posterior distribution of images $\mathbf{x}$ conditioned on measurements $\mathbf{y}$:

$$p(\mathbf{x} \mid \mathbf{y}) \propto p(\mathbf{y} \mid \mathbf{x}) \, p(\mathbf{x}).$$

This Bayesian framework supports a modular approach to inverse problems where the likelihood $p(\mathbf{y} \mid \mathbf{x})$ is defined by an expert based on knowledge of how measurements are obtained, and the prior $p(\mathbf{x})$ is defined independently. Furthermore, it allows for principled solutions. Maximum a posteriori (MAP) estimation can be done by optimizing the posterior probability. Posterior sampling, which is useful

for uncertainty quantification, can be done with MCMC or variational inference. But since such methods require the value or gradient of $p(\mathbf{x})$, they have been limited to simple priors (e.g., Gaussian) and weighted regularizers (e.g., total variation). In practice, the relative weights of the prior and likelihood terms are usually tuned by hand, introducing a human bias that is unsatisfactory for scientific applications.

On the deep-learning side, solutions leveraging an implicit, deep-learned prior may look convincing but do not lend themselves to principled analysis. For example, a convolutional neural network (CNN) can be trained in a supervised way to output images given measurements, but its prior cannot be probed and does not generalize to new tasks. Recent work shows how to condition a diffusion model — a type of generative model whose prior is captured in a learned image denoiser — on arbitrary measurements [13, 14, 15, 16, 27, 66, 69], but the methods depend on hand-tuned hyperparameters and do not sample from a true posterior except in auspicious cases. To get the best of both worlds (traditional Bayesian inference and modern deep learning), we need a way to incorporate the expressive prior of a deep-learned model into a traditional, principled Bayesian-inference approach.

We propose employing a diffusion model as the prior in Bayesian inference for imaging. A *score-based prior* is the distribution under a score-based diffusion model [70], which has been proven to allow for exact probabilities (a feature under-explored in practice). In this paper, we first review related work and then investigate the probability function, including empirical validation of its accuracy.

The main contribution of this paper is establishing score-based priors as an interface between modern deep-learning and traditional inverse problem-solving, giving proven, principled approaches direct access to learned, rich priors. Under our framework, we train a score-based prior once on a dataset of images. Paired with any likelihood $p(\mathbf{y} \mid \mathbf{x})$, this prior can be plugged into any inference algorithm that uses the value or gradient of the posterior. We demonstrate this with an existing variational-inference approach for posterior sampling and show results for three inverse problems: denoising, a version of deblurring, and interferometry. Interferometry is used for black-hole imaging (Fig. 1) and highlights the benefits of score-based priors for scientific applications, which call for exact posterior sampling given standalone priors to accurately quantify uncertainty.

2. Related Work

2.1. Inverse Problems in Imaging

The goal of an imaging inverse problem is to recover a hidden image $\mathbf{x}^* \in \mathbb{R}^D$ from measurements $\mathbf{y} \in \mathbb{R}^M$:

$$\mathbf{y} = \mathbf{f}(\mathbf{x}^*) + \epsilon. \tag{1}$$

We usually assume the forward model $\mathbf{f}$ is known and the measurement noise ϵ is a random variable with a known distribution. When $\mathbf{y}$ is missing information about the underlying image, solving for $\mathbf{x}^*$ is ill-posed.

Bayesian inference. The Bayesian approach considers the posterior distribution $p(\mathbf{x} \mid \mathbf{y})$, which decomposes an inverse problem explicitly into a likelihood and a prior:

$$\log p(\mathbf{x} \mid \mathbf{y}) = \log p(\mathbf{y} \mid \mathbf{x}) + \log p(\mathbf{x}) + \text{const.} \tag{2}$$

This quantity is often re-interpreted as a data-fidelity term plus a weighted regularizer. Possible solutions to an inverse problem include the maximum a posteriori (MAP), which is the mode of the posterior, and unbiased samples from the posterior. Samples are more informative than the MAP and allow for uncertainty quantification. Principled approaches for posterior sampling include Markov chain Monte Carlo (MCMC) [8], which generates a Markov chain whose stationary distribution is the posterior, and variational inference [6], which finds the best approximation of the posterior within a family of parameterized distributions. Such algorithms require either the value or gradient of the posterior density function, which is especially difficult to determine for images. Assuming a known likelihood, the challenge is defining a prior on images that reflects their complicated statistics. As such, image priors are usually regularizers enforcing some simple property of images. Examples include total variation (TV) and total squared variation (TSV) for spatial smoothness [7, 47] and L1 norm for sparsity [9]. The weightings of these regularizers are typically set by hand.

Deep learning for inverse problems. Deep neural networks can learn complex image distributions, but their implicit priors are difficult to analyze. One could train a neural network on a paired dataset of images and measurements [56, 35, 79, 81, 80, 78, 62, 19, 18], but this requires re-training for new tasks, and uncertainty cannot be analyzed under the learned prior. Bayesian networks [24] can account for uncertainty, but their priors are still implicit, and they have not been demonstrated on complicated posteriors. Deep Image Prior [74] showed that the inductive bias of a CNN can act as an implicit prior. Other methods like Plug-and-Play (PnP) [76] and Regularization by Denoising (RED) [59] use an image denoiser as an implicit regularizer to provide a point solution. PnP/RED-style methods that provide samples [49, 37, 40] have not been shown to sample from true posteriors based on the denoiser's prior.

2.2. Data-Driven Generative Models

Generative models learn a probability distribution of images, but besides score-based diffusion models, they have been either limited in complexity or unable to provide exact image probabilities. Classical examples are Gaussian mixture models [82, 83] and independent components analysis (ICA) [5, 34], which give probabilities but over-simplify

image distributions. As for deep generative models, generative adversarial networks (GANs) [26] do not have tractable probabilities, and variational autoencoders (VAEs) [45, 58] only provide a lower-bound on probabilities. Flow-based models like discrete normalizing flows [43, 22, 29, 28, 4, 21, 52] and autoregressive flows [48, 25, 75, 55] support exact log-probability computation [3] but are restricted to certain network architectures and do not generalize well outside of training data [46, 52]. Diffusion models [30, 44, 77, 65, 54] are a promising alternative, with the score-based interpretation [70] providing a theoretical framework for deriving exact probabilities [51]. This capability has gone underexamined in previous work, which mostly focuses on diffusion models as unconditional samplers [20, 31, 60] and conditional samplers [61, 53, 57, 32, 39].

2.2.1 Diffusion models

A diffusion model transforms a simple distribution into a complex one, with image generation usually done by gradually denoising a sample from $\mathcal{N}(\mathbf{0}_D, \mathbf{I}_D)$ until it becomes a clean image. Denoising diffusion probabilistic models (DDPMs) [64, 30] treat this transformation as a discrete-time process with a fixed number of denoising steps. Score-based generative models [68] generalize to continuous time.

Score-based diffusion models. In the continuous-time setting, a stochastic differential equation (SDE) describes the data-transformation process and lends itself to a theoretical framework for analyzing the induced sequence of data distributions. In particular, a *forward-time SDE* defines the diffusion process of an image $\mathbf{x}_0$ from $t = 0$ to $t = T$:

$$d\mathbf{x}_t = \mathbf{f}(\mathbf{x}_t, t)dt + g(t)d\mathbf{w}. \tag{3}$$

$\mathbf{f}(\cdot, t) : \mathbb{R}^D \to \mathbb{R}^D$ is the drift coefficient and defines the deterministic evolution of $\mathbf{x}_t$. $\mathbf{w} \in \mathbb{R}^D$ denotes Brownian motion, and the diffusion coefficient $g(\cdot) : \mathbb{R} \to \mathbb{R}$ controls the rate of diffusion. This SDE gives rise to a time-dependent probability distribution p_t, where p_0 is the original data distribution, and p_T is the standard normal.

Sampling an image from p_0 requires reversing the diffusion process. The *reverse-time SDE* [2] is given by

$$d\mathbf{x}_t = \left[\mathbf{f}(\mathbf{x}_t, t) - g(t)^2 \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)\right] dt + g(t)d\bar{\mathbf{w}}. \tag{4}$$

The problematic term here is $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$, which is the (Stein) score of $\mathbf{x}$ under p_t. In words, undoing diffusion is difficult because it requires knowing how to nudge each perturbed distribution p_t closer to the clean data distribution. Being the only data-dependent component of this SDE, the score function is learned by a neural network $\mathbf{s}_\theta(\mathbf{x}, t)$ with parameters θ [70, 67]. Because each p_t is p_0 perturbed by Gaussian noise, the time-dependent score model $\mathbf{s}_\theta(\mathbf{x}, t)$

can be thought-of as an image denoiser: it takes a noisy image as input and estimates the noise, and t indicates the level of noise applied (higher t means more noise). To sample a new image, a point $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}_D, \mathbf{I}_D)$ is drawn whose final state $\mathbf{x}_0$ is given by solving the reverse-time SDE, essentially through many denoising steps [70].

Diffusion models for inverse problems. Previous methods attempt to sample from a posterior using an unconditional diffusion model but are not guaranteed to sample from the true posterior or a close approximation of it. Most methods incorporate measurements into the reverse-diffusion process. Measurement-consistency is enforced by either projecting images onto a measurement subspace [69, 15, 16, 12, 14] or following a gradient toward higher measurement likelihood [13, 36, 27, 38, 1]. These methods (1) require careful hyperparameter-tuning to obtain reasonable samples and (2) fail to sample from the true posterior no matter the hyperparameter values. We put forth a new perspective: applying the hyerparameter-free probability function of a score-based diffusion model to principled inference algorithms that simply require a differentiable image prior.

3. Score-Based Priors

We propose score-based priors as differentiable image priors that can be trained on any dataset and employed for principled inverse imaging. For example, one can learn a prior on face images by training a score model $\mathbf{s}_\theta(\mathbf{x}, t)$ on CelebA [50]. Then one can model any posterior with a face-image prior by appealing to the function $\log p_\theta(\mathbf{x})$ (Fig. 1).

3.1. Log-Probability Computation

Our method leverages previous work that shows how to compute image probabilities under the SDE framework [70]. This feature has been mostly discussed in theory but is yet to be demonstrated in practice. In this section, we discuss log-probability computation under a score-based diffusion model and empirically validate it.

Probability flow ODE. Computing probabilities requires inverting the sampling process: the probability of an image $\mathbf{x}_0$ depends on the probability of the $\mathbf{x}_T$ that would have resulted in that image through reverse diffusion. There is an ordinary differential equation (*probability flow ODE*) [70] that makes both the forward and reverse SDEs (Eqs. 3, 4) invertible, inducing the same time-dependent probability distribution p_t but without Brownian motion. The ODE is the same forward and backward in time, defining a bijective mapping between p_t and $p_{t'}$ for any two times $t, t' \in [0, T]$.[1] For a score model $\mathbf{s}_\theta(\mathbf{x}, t) \approx \nabla_{\mathbf{x}} \log p_t(\mathbf{x})$, the learned probability flow ODE is given by

---

[1]The implicit assumption is that the score model is well-trained such that $p_0 \approx p_{\text{data}}$, where p_{data} is the true distribution of the training data.

$$\frac{\mathrm{d}\mathbf{x}_t}{\mathrm{d}t} = \mathbf{f}(\mathbf{x}_t, t) - \frac{1}{2}g(t)^2 \mathbf{s}_\theta(\mathbf{x}_t, t) =: \tilde{\mathbf{f}}_\theta(\mathbf{x}_t, t). \quad (5)$$

Log-probability formula. By the continuous-time change-of-variables formula [11], the log-probability of an image $\mathbf{x} = \mathbf{x}_0$ under the p_0 distribution is given by the log-probability of $\mathbf{x}_T$ under the p_T Gaussian, plus a normalization factor accounting for the change in probability density from $\mathbf{x}_0$ to $\mathbf{x}_T$. We compute the log-probability under the learned ODE (Eq. 5) by solving an initial-value problem:

$$\log p_0(\mathbf{x}_0) = \log p_T(\mathbf{x}_T) + \int_0^T \nabla \cdot \tilde{\mathbf{f}}_\theta(\mathbf{x}_t, t)\mathrm{d}t, \quad (6)$$

where $\mathbf{x}_0 = \mathbf{x}$. The divergence $\nabla \cdot \tilde{\mathbf{f}}_\theta(\mathbf{x}_t, t)$ quantifies the instantaneous change in log-probability of $\mathbf{x}_t$ caused by applying $\tilde{\mathbf{f}}_\theta(\mathbf{x}_t, t)$ in either time direction. It can be estimated with Hutchinson-Skilling estimation of the trace of $\frac{\partial}{\partial \mathbf{x}_t}\tilde{\mathbf{f}}_\theta(\mathbf{x}_t, t)$ [70]. We denote the log-probability function of a score-based prior as $\log p_\theta := \log p_0$.

3.2. Log-Probability Validation

In several experiments (Figs. 2, 3, 5), we compare a learned score-based prior to a known ground-truth distribution. The ground-truth is a Gaussian distribution of 16×16 grayscale images, whose mean and preconditioned covariance were fit to CelebA face images. The score model was trained on samples from this Gaussian. Fig. 2 shows our empirical analysis of log-probabilities under the score-based prior versus the true prior. We verify accuracy on both in-distribution and out-of-distribution images.
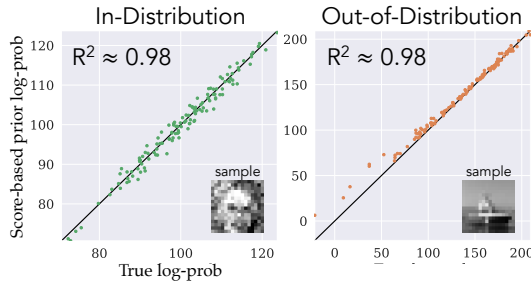
Figure 2. Log-probabilities of the score-based prior vs. ground-truth. Black line indicates perfect agreement. **In-Distribution.** The log-probabilities of 128 samples from the Gaussian ground-truth distribution were evaluated (shown as scatter points). Score-based log-probabilities are strongly correlated with ground-truth log-probabilities ($R^2 \approx 0.98$). **Out-of-Distribution.** The log-probabilities of test images from CIFAR-10 (scaled to 16×16) are shown. The score-based prior generalizes well out of distribution.

Gradients. As many Bayesian-inference approaches require a *differentiable* prior, we validate that the gradient function $\nabla_\mathbf{x} \log p_\theta(\mathbf{x})$ provides more accurate gradients
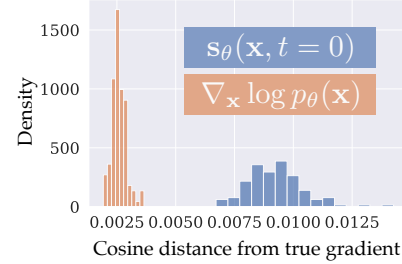
Figure 3. ODE gradients vs. score-model outputs. Histogram shows density of cosine distance between the estimated gradient and true gradient for 128 samples from the ground-truth Gaussian. $\mathbf{s}_\theta(\mathbf{x}, t = 0)$ is the score-model-approximated gradient ($t = 0.001$ is actually used for numerical stability). $\nabla_\mathbf{x} \log p_\theta(\mathbf{x})$ is computed numerically according to the probability flow ODE.

than the score model $\mathbf{s}_\theta(\mathbf{x}, t = 0) \approx \nabla_\mathbf{x} \log p_0(\mathbf{x})$. It is tempting to apply the score model as a cheap gradient approximator since it is designed in theory as such, but in practice, it does not generalize to out-of-distribution data. It is more reliable to compute the gradient of the entire ODE solve of $\log p_\theta(\mathbf{x})$ with respect to $\mathbf{x}$. Fig. 3 shows that given samples from the ground-truth Gaussian, gradients computed according to the ODE are closer to the true gradients (in terms of cosine similarity) than score-model outputs. Fig. 5 (appearing in Sec. 4) shows that using score-model outputs as gradients leads to an incorrect posterior.

Implementation details. Our code[2] is written in JAX and Diffrax [42] to make it easy to just-in-time (JIT) compile log-probability computations, select ODE solvers, and compute autograds. For experiments in this paper, we used 5th-order Runge-Kutta solvers [23, 73] and Hutchinson-Skilling trace estimation [33, 63]. We approximated gradients with the continuous adjoint method [41]. The supplementary text includes further practical details about solvers, trace estimation, and our experimental setups.

4. Posterior Sampling with Score-Based Priors

With score-based priors, we can solve inverse problems by sampling from rich posteriors given any image prior (assuming sufficient training data) and any measurements (assuming a known forward model). We accomplish posterior sampling by plugging the score-based prior into a variational-inference method. This both highlights the applicability of score-based priors to established optimization methods and provides a solution to the open problem of posterior sampling with unconditional diffusion models. In this section, we describe our chosen approach, *Deep Probabilistic Imaging* (DPI), which converges faster than MCMC and provides efficient sampling [71].

---

[2]Website: http://imaging.cms.caltech.edu/score_prior

4.1. Main Idea for Posterior Sampling

Our main idea for unlocking the rich prior of a diffusion model is to remove the concept of diffusion and consider the prior as a fixed distribution just like the likelihood. This lets us directly model the posterior $\log p(\mathbf{x}|\mathbf{y})$.

Previous methods treat the diffusion model as an implicit prior and entangle measurements with diffusion, which inhibits true posterior sampling. Posterior reverse diffusion requires the *posterior* score function

$$\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t \mid \mathbf{y}) = \nabla_{\mathbf{x}_t} \log p_t(\mathbf{y} \mid \mathbf{x}_t) + \mathbf{s}_\theta(\mathbf{x}, t)$$

for all $t \in [0, T]$. The score model gives the prior score, but the likelihood score is only defined for $t = 0$. For every diffusion time t, the true likelihood is defined by an intractable integral over all $\mathbf{x}_0$:

$$p_t(\mathbf{y} \mid \mathbf{x}_t) = \int_{\mathbf{x}_0} p(\mathbf{y} \mid \mathbf{x}_0) p(\mathbf{x}_0 \mid \mathbf{x}_t) d\mathbf{x}_0.$$

Previous methods either abandon the true measurement uncertainty [12, 14, 15, 16, 27, 69, 1] or strongly approximate $p_t(\mathbf{y} \mid \mathbf{x}_t)$ [13, 36, 38, 66]. This necessitates hyperparameter(s) for determining the importance of measurements versus the prior, making the estimated posterior more of a *conditional* distribution rather than a principled posterior. Different hyperparameter settings have drastic effects on the estimated posterior: if measurements are under-weighted, then the posterior is overly-biased toward the prior and may contain misleading data; if measurements are over-weighted, then the samples may collapse onto a subspace that does not make sense under the prior. Fig. 4 shows these outcomes on a 2D example. It also shows that, even with ideally-tuned hyperparameters (which require knowledge of ground-truth), previous methods cannot capture the true posterior. We describe the baseline methods in Sec. 5.

4.2. Selected Approach: Variational Inference with a Normalizing Flow

Our approach for directly modeling the posterior is rooted in variational inference. Following the method proposed in DPI [71, 72], we define a family of distributions q_ϕ via a RealNVP [22] normalizing flow with parameters ϕ, which we optimize to approximate the desired posterior. For a score-based prior with parameters θ, the objective is

$$\phi^* = \arg \min_\phi D_{\mathrm{KL}} \left(q_\phi \| p_\theta(\cdot \mid \mathbf{y}) \right) \tag{7}$$

$$= \arg \min_\phi \mathbb{E}_{\mathbf{x} \sim q_\phi} \left[-\log p(\mathbf{y} \mid \mathbf{x}) - \log p_\theta(\mathbf{x}) + \log q_\phi(\mathbf{x}) \right].$$

This variational objective includes the log-posterior and an entropy term, $\mathbb{E}_{\mathbf{x} \sim q_\phi} [\log q_\phi(\mathbf{x})]$, which can be tractably
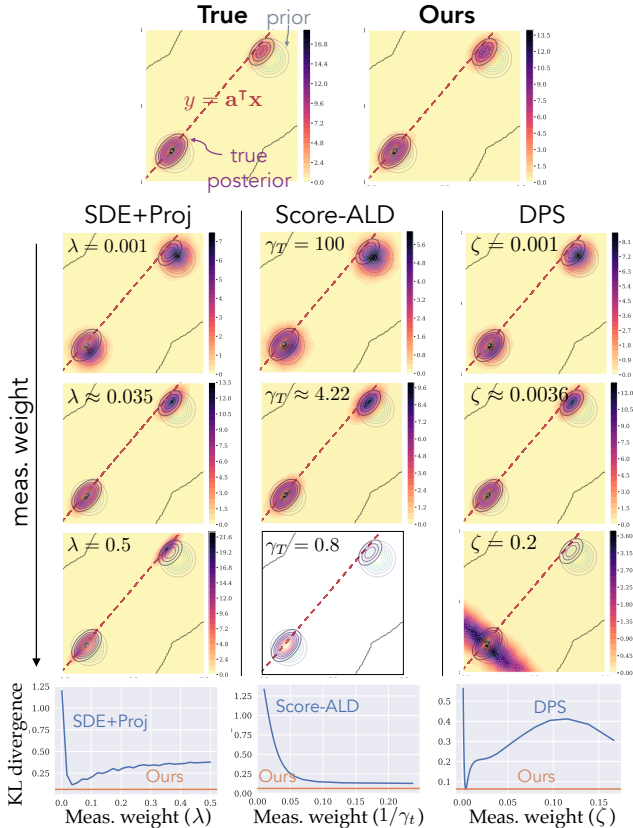

Figure 4. Baseline methods [69, 36, 13] do not sample true posterior. Heatmap depicts $p(\mathbf{x}|y)$ approximated from samples; contour lines depict true posterior. All methods use the same (true) score function. For baselines, we did a grid search to find the optimal hyperparameter weight (ALD and DPS hyperparameters were distilled into a global hyperparameter). The **KL divergence** from the estimated posterior to the true posterior was approximated for each hyperparameter value. No matter the value, baselines do not get more accurate than our hyperparameter-free method. For instance, with "Best meas. weight", all baselines sample from both modes equally, even though the lower-left mode should have more density. A poor hyperparameter setting can even lead to unstable sampling (see "High meas. weight" of Score-ALD, where samples became NaNs due to slightly over-weighted measurements).

computed under the RealNVP.[3] During fitting, the expectation is Monte-Carlo approximated with a batch of samples from the RealNVP. Note that the **absence of hyperparameters** in this objective is a feature of the score-based prior (not DPI). Because our prior is truly probabilistic, there is no need for a hand-tuned weight.

As evidence that our method correctly samples the posterior, Fig. 5 compares an estimated posterior to ground-

---

[3]As discussed in Sec. 2, discrete normalizing flows allow for exact probabilities, but this feature is not reliable on out-of-distribution data. We therefore only use $\log q_\phi(\mathbf{x})$ during fitting and then once ϕ is fixed, only use the RealNVP for sampling.
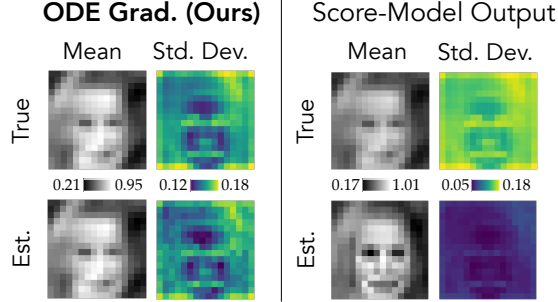
Figure 5. True vs. estimated posterior. Measurements are the 6.25% lowest DFT spatial-frequencies of an image from the prior and have i.i.d. noise with $\sigma = |1|$. The true mean and variance were derived analytically since the prior is the Gaussian ground-truth distribution, and the likelihood distribution is also Gaussian. The mean and variance were estimated from 10240 samples.

truth. The score-based prior was trained on the ground-truth Gaussian used in Sec. 3, and the task was to deblur an image from the true prior. The estimated mean and covariance closely agree with the analytical mean and covariance. Additionally, we find that directly using the score-model output leads to an incorrect posterior.

We note that accurate posterior sampling is dependent on accurate log-probabilities in the first place. We verify that log-probabilities under a score-based prior well-approximate those under a ground-truth Gaussian prior in Fig. 2, but log-probabilities are difficult to validate for complex priors (which motivates our work).

5. Posterior Sampling Results

5.1. Baseline Methods

We include comparisons to three previous methods for posterior sampling with an unconditional diffusion model. Fig. 4 compares methods for a 2D example.

SDE+Proj (Song et al. 2022) [69]. The image $\mathbf{x}_t$ is projected onto a measurement subspace at each t throughout reverse diffusion. A hyperparameter λ determines measurement weight. This approach is purported to give plausible solutions rather than exact posterior samples. It is restricted to compressed-sensing linear inverse problems where the forward matrix has fewer rows than columns.

Score-based annealed Langevin dynamics (Score-ALD) (Jalal & Arvinte et al. 2021) [36]. The authors propose ALD with approximate posterior scores:

$$\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t \mid \mathbf{y}) \approx \mathbf{s}_\theta(\mathbf{x}_t, t) + \frac{\mathbf{A}^H(\mathbf{y} - \mathbf{A}\mathbf{x}_t)}{\sigma^2 + \gamma_t^2}, \quad (8)$$

where $\frac{\mathbf{A}^H(\mathbf{y} - \mathbf{A}\mathbf{x}_t)}{\sigma^2 + \gamma_t^2}$ is the log-likelihood gradient assuming linear measurements ($\mathbf{y} = \mathbf{A}\mathbf{x} + \epsilon$) and Gaussian noise ($\epsilon \sim \mathcal{N}(\mathbf{0}_M, \sigma^2 \mathbf{I}_M)$). γ_t is a hyperparameter for the weight

of the log-likelihood term at step t, thus calling for a hand-tuned approximation of the likelihood score at each t. One way to automatically adjust the magnitude of the likelihood score is to renormalize it to have the same magnitude as the prior score, which is what the authors do in their experiments. Without this trick, we find image posterior sampling highly sensitive to the γ_t annealing schedule.

Diffusion posterior sampling (DPS) (Chung & Kim et al. 2023) [13]. The authors propose an approximation of $\log p_0(\mathbf{y} \mid \mathbf{x}_t)$ throughout reverse diffusion. A hyper-paramter ζ_t determines the magnitude of the log-likelihood gradient at each t. Similar to ALD, the authors use a trick to determine the magnitude of the likelihood score, setting $\zeta_t := \zeta / \|\mathbf{y} - \mathbf{f}(\hat{\mathbf{x}}_0(\mathbf{x}_t))\|$, where $\mathbf{f}$ is the measurement forward operator, and $\mathbf{x}_0(\mathbf{x}_t) := \mathbb{E}[\mathbf{x}_0 \mid \mathbf{x}_t]$.

5.2. Imaging Inverse Problems

We demonstrate image posterior sampling on denoising, a version of deblurring, and interferometric imaging of a simulated black hole. There are two comparisons to perform: (1) score-based priors vs. other *priors* and (2) our posterior-sampling approach vs. other *posterior-sampling approaches* using an unconditional score model. (1) is done with a denoising experiment, and (2) is done with a deblurring experiment, although the findings are not task-dependent. We then focus on black-hole imaging as an endeavor that could benefit from score-based priors as tools in the scientific process.

5.2.1 Validation of prior (denoising)

We denoise images corrupted by i.i.d. Gaussian noise with standard deviation $\sigma = 0.2$ (20% of the dynamic range). Fig. 6 shows denoised posteriors given different priors: (1) **TV** regularization, (2) PCA-Gaussian (**PCA-G**), (3) Real-NVP normalizing flow (**NF**), and (4) our score-based prior. DPI was used to approximate the posterior for each prior except PCA-G, which has an analytical Gaussian posterior.

Score-based priors provide more informative posteriors than traditional priors do. This is reflected in sample quality: as Tab. 1 shows, the average SSIM and PSNR of posterior samples are highest when using a score-based prior. Score-based priors also provide richer posteriors: as shown by the empirical standard deviations in Fig. 6, our priors result in full posteriors with a data-driven uncertainty (e.g., uncertainty is higher on facial features like the eyes, ears, and hair of the CelebA image, whereas other priors are more unaware of facial structure). The RealNVP NF performed poorly as an image prior, perhaps because it struggles to generalize to non-training images and thus leads to unstable optimization of the variational posterior. Please refer to the supplementary text for more details about this experiment.

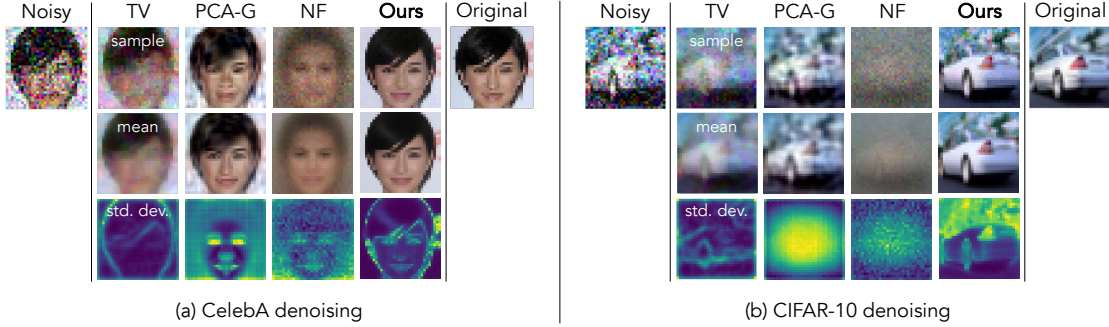(a) CelebA denoising

(b) CIFAR-10 denoising

Figure 6. Denoising with different priors. The data-driven priors (PCA-G, NF, Ours) were trained on the CelebA training set in (a) and the CIFAR-10 training set in (b). **TV** regularization weight is 0.00025. **PCA-G** is a Gaussian based on the top 512 principal components of the training data. **NF** is a RealNVP with 64 affine-couple layers. TV, PCA-G, and NF are relatively simple priors that do not give rich posteriors. Score-based priors (**Ours**) capture complex spatial correlations (see sample quality and data-driven std. dev. maps).

Prior / Image	CelebA (Ours)	CIFAR (Ours)	Same dataset (PCA-G)	Same dataset (NF)	TV
CelebA	SSIM: **0.88**	0.71	0.76	0.31	0.53
	PSNR: **25.1**	21.6	20.5	11.3	17.4
CIFAR	SSIM: 0.79	**0.85**	0.74	0.22	0.57
	PSNR: 19.0	**21.0**	19.4	13.4	17.6

Table 1. Denoising metrics. Row refers to test image in Fig. 6. Column refers to prior used for denoising. Average SSIM and PSNR were computed across 128 posterior samples for each (test image, prior) pair. The "correct" score-based prior (Ours) performs best for each test image, while even the "incorrect" score-based prior performs well compared to the other priors.

5.2.2 Validation of posterior sampling (deblurring)

We consider the task of reconstructing an image from measurements of the lowest DFT spatial frequencies, which we call "deblurring" to simplify terminology. In our experiments, we observed the lowest 6.25% DFT spatial frequencies with complex-valued measurement noise with standard deviation $|\sigma| = 1.0$ (~0.2% of the magnitude of the zero-frequency component). Figs. 7, 8 show results for a CelebA and CIFAR-10 source image, respectively, comparing our method, Score-ALD, DPS, and SDE+Proj. We used the same two score models (one trained on CelebA and one trained on CIFAR-10) for all methods. Our method outperforms others in terms of MSE, PSNR, and SSIM (e.g., in Fig. 7(a), our posterior samples have an average PSNR of 24.75, while DPS, the best-performing method in terms of PSNR, achieved an average of 20.37). Full metrics for Figs. 7, 8 are provided in the supplementary text.

Our posterior-sampling method is much more robust to mismatched priors than baselines are. Consider the CelebA prior applied to measurements of a CIFAR-10 image in Fig. 8(b). When the measurement weight in baselines is lower, they hallucinate faces, resulting in a posterior that lies within the prior. When it is higher, they introduce unnatural artifacts to fit the measurements. Hyperparameter-dependent methods make it easy to mistakenly over-bias the
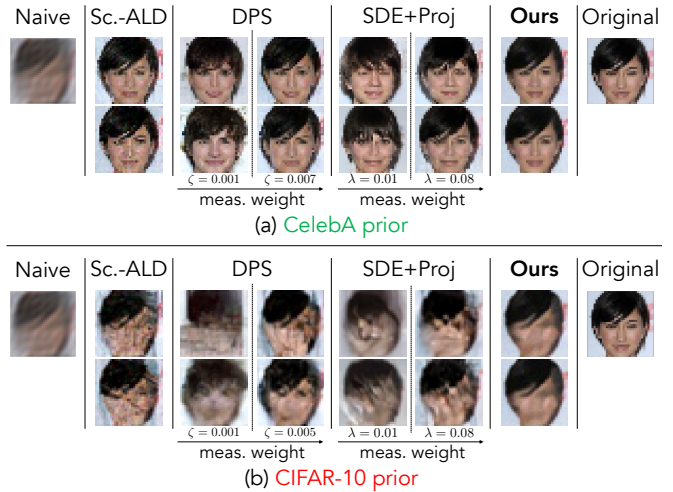

(a) CelebA prior

(b) CIFAR-10 prior

Figure 7. Deblurring (CelebA). Two samples from each method are shown. The source is a CelebA test image (Original). Score-ALD uses the likelihood-gradient renormalization trick. **(a) Using the "correct" prior.** All methods recover plausible posterior samples when the prior includes the true source image. With baselines, posterior variance depends on the meas. weight, which is difficult to set without knowing the ground-truth posterior. **(b) Using an "incorrect" prior.** All methods (expectedly) are not able to recover a face. But baseline methods introduce heavy artifacts, while samples from Ours still look natural.

prior or the measurements. Without access to a ground-truth, one cannot reliably interpret a posterior if it is not the result of principled Bayesian inference.

5.2.3 Interferometric imaging

The scientific venture of black-hole imaging calls for principled inference with priors that are minimally hand-tuned. Radio interferometry is a technique for imaging astronomical targets with high angular resolution by using a distributed array of radio telescopes. An interferometer col-
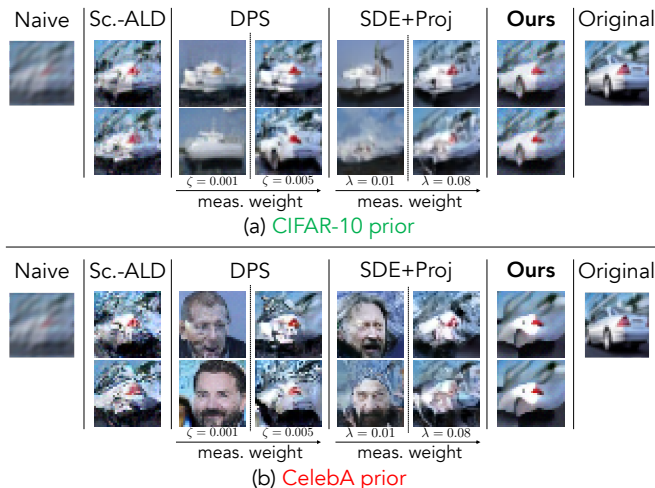
Figure 8. Deblurring (CIFAR-10). **(a) Using the "correct" prior.** For DPS and SDE+Proj, both weights achieve plausible samples, but variance differs drastically. **(b) Using an "incorrect" prior.** ALD, DPS, and SDE+Proj with high meas. weight struggle to produce natural images when faced with out-of-distribution measurements. With lower meas. weight, DPS and SDE+Proj hallucinate faces. This is also troubling since the posterior should not lie inside the prior in this case.
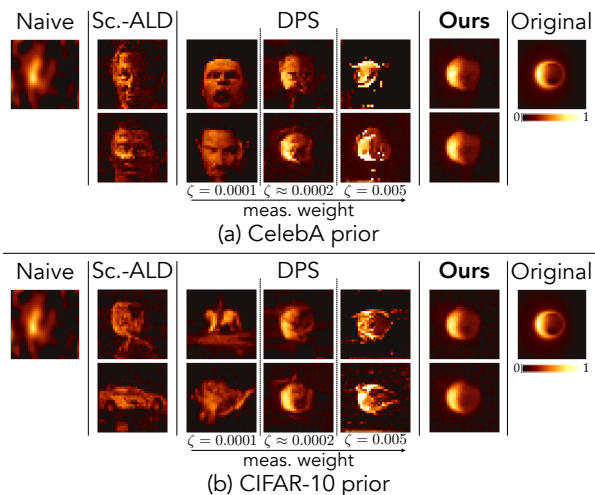

Figure 9. Interferometric imaging of a synthetic black-hole image. Two random samples from each method are shown (images are all shown with the same color scale). Both baselines struggle to balance measurements and prior, hallucinating faces in (a) and a car and horse in (b). Posterior samples become unstable as meas. weight increases, as evidenced by the unnatural structures resulting from the highest meas. weight in DPS. Regardless of meas. weight, variance of baseline samples far exceeds ours (ALD max. std. dev. is 0.202; DPS max. std. devs. are 0.237, 0.242, 0.444, resp., as meas. weight increases; our max. std. dev. is 0.036). **Ours** produces samples that agree with the true structure, automatically balancing likelihood and prior.

lects sparse spatial-frequency measurements of the sky's image. The Event Horizon Telescope (EHT) notably used this technique to take the first image of a black hole [17]. In this work, we simulated interferometric measurements from the EHT telescope array using the `ehtim` package [10].[4]

The first image of a black hole, while the result of carefully-obtained measurements from the EHT array, was only possible with image priors (technically formulated as regularizers). EHT scientists handcrafted many priors, each bringing different biases to the image reconstruction. Only the common structure found between these priors could be reliably interpreted, such as the diameter of the photon ring surrounding the black hole [17]. Score-based priors could streamline this process as principled, hyperparameter-free priors that can be easily trained on different image distributions. Plugged into the imaging algorithm, they provide a collection of posteriors that incorporate different image statistics while maintaining measurement consistency.

It is important to remember that a posterior exists for *any* combination of a prior and measurements, no matter how far the prior is from the source image. Faithfully modeling the posterior that arises from a given prior is especially crucial in a task like black-hole imaging. Since it is impossible to train a prior on real black holes, any data-driven prior likely does not perfectly agree with the source image. Similar to our deblurring experiments (Figs. 7, 8), Fig. 9 highlights that score-based priors are robust to mismatches

between the prior and true underlying distribution. Fig. 9 shows results using score-based priors trained on CelebA and CIFAR-10. The simulated measurements, from all five telescopes in the array, contain enough information for both priors to recover the underlying image structure. We compare our results to those of Score-ALD and DPS (SDE+Proj does not support this type of forward model).[5] Unlike these methods, which introduce either prior-related features (e.g., face) or unstructured artifacts (e.g., random pixels to agree with measurements), our hyperparameter-free method automatically balances the measurements and prior.

As shown in Fig. 1, a score-based prior visibly affects the posterior wherever measurements are not sufficient to constrain image structure. For instance, when applying a score-based prior trained on CelebA, as measurements are removed, more facial features appear in the posterior images. Given enough measurements, the recovered structure in the posterior samples can be reliably analyzed for scientific interpretation. With our framework, it is also possible to train a collection of score-based priors and look for common features between the posteriors that arise from the different priors. These common features are more likely to reflect the true underlying image.

---

[4]These simulated measurements contain thermal noise but exclude realistic atmospheric noise that results in additional phase corruption.

[5]The EHT forward matrix has more rows than columns, but it is ill-conditioned because measurements are highly correlated.

6. Limitations

A score-based prior is a learned approximation of a desired image distribution, with its correctness tied to the correctness of the score model. There is also a tradeoff between the complexity of the prior and compute time/memory. To obtain exact log-probabilities with a score-based prior, it is necessary to solve an ODE (often repeatedly throughout an optimization algorithm). Theoretical and practical improvements can be made to reduce this cost, such as making the score model more generalizable to quickly estimate $\nabla_{\mathbf{x}} \log p_\theta(\mathbf{x})$. We also note that our approach for posterior sampling approximates the true posterior with a parameterized variational distribution, whose expressiveness determines the quality of the approximation.

7. Conclusion

We have shown how to turn a score-based diffusion model into a principled prior, specifically demonstrating a variational approach for posterior sampling. Results on 2D data establish the soundness of the approach, while our image denoising and deblurring experiments show the modularity of the score-based prior (applying the same prior across multiple tasks without any hand-tuning), its expressiveness (comparing to traditional priors), and its robustness (comparing to diffusion-based approaches to inverse problems). Score-based priors are especially useful for scientific imaging, such as interferometric imaging of black holes. Our work opens a new direction in computational imaging that merges data-driven priors and principled inference.

Acknowledgments

References

[1] Alexandre Adam, Adam Coogan, Nikolay Malkin, Ronan Legin, Laurence Perreault-Levasseur, Yashar Hezaveh, and Yoshua Bengio. Posterior samples of source galaxies in strong gravitational lenses with score-based priors. *arXiv preprint arXiv:2211.03812*, 2022. 3, 5

[2] Brian D.O. Anderson. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3):313–326, 1982. 3

[3] Muhammad Asim, Max Daniels, Oscar Leong, Ali Ahmed, and Paul Hand. Invertible generative models for inverse problems: mitigating representation error and dataset bias. In *International Conference on Machine Learning*, pages 399–409. PMLR, 2020. 3

[4] Jens Behrmann, Will Grathwohl, Ricky TQ Chen, David Duvenaud, and Jörn-Henrik Jacobsen. Invertible residual networks. In *International Conference on Machine Learning*, pages 573–582. PMLR, 2019. 3

[5] Anthony J Bell and Terrence J Sejnowski. The "independent components" of natural scenes are edge filters. *Vision research*, 37(23):3327–3338, 1997. 2

[6] David M Blei, Alp Kucukelbir, and Jon D McAuliffe. Variational inference: A review for statisticians. *Journal of the American statistical Association*, 112(518):859–877, 2017. 2

[7] Charles Bouman and Ken Sauer. A generalized gaussian image model for edge-preserving map estimation. *IEEE Transactions on image processing*, 2(3):296–310, 1993. 2

[8] Steve Brooks, Andrew Gelman, Galin Jones, and Xiao-Li Meng. *Handbook of markov chain monte carlo*. CRC press, 2011. 2

[9] Emmanuel Candes and Justin Romberg. Sparsity and incoherence in compressive sampling. *Inverse problems*, 23(3):969, 2007. 2

[10] Andrew A Chael, Michael D Johnson, Katherine L Bouman, Lindy L Blackburn, Kazunori Akiyama, and Ramesh Narayan. Interferometric imaging directly with closure phases and closure amplitudes. *The Astrophysical Journal*, 857(1):23, 2018. 8

[11] Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. *NeurIPS*, 31, 2018. 4

[12] Jooyoung Choi, Sungwon Kim, Yonghyun Jeong, Youngjune Gwon, and Sungroh Yoon. Ilvr: Conditioning method for denoising diffusion probabilistic models. In *ICCV*. IEEE, 2021. 3, 5

[13] Hyungjin Chung, Jeongsol Kim, Michael Thompson Mccann, Marc Louis Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *The Eleventh International Conference on Learning Representations*, 2023. 2, 3, 5, 6

[14] Hyungjin Chung, Byeongsu Sim, Dohoon Ryu, and Jong Chul Ye. Improving diffusion models for inverse problems using manifold constraints. *arXiv preprint arXiv:2206.00941*, 2022. 2, 3, 5

[15] Hyungjin Chung, Byeongsu Sim, and Jong Chul Ye. Come-closer-diffuse-faster: Accelerating conditional diffusion models for inverse problems through stochastic contraction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12413–12422, 2022. 2, 3, 5

[16] Hyungjin Chung and Jong Chul Ye. Score-based diffusion models for accelerated mri. *Medical Image Analysis*, 80:102479, 2022. 2, 3, 5

[17] Event Horizon Telescope Collaboration et al. First m87 event horizon telescope results. iv. imaging the central supermas-

sive black hole. *arXiv preprint arXiv:1906.11241*, 2019. 1, 8

[18] Mauricio Delbracio and Peyman Milanfar. Inversion by direct iteration: An alternative to denoising diffusion for image restoration. *arXiv preprint arXiv:2303.11435*, 2023. 2

[19] M. Delbracio, H. Talebi, and P. Milanfar. Projected distribution loss for image enhancement. In *2021 IEEE International Conference on Computational Photography (ICCP)*, pages 1–12, Los Alamitos, CA, USA, may 2021. IEEE Computer Society. 2

[20] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in Neural Information Processing Systems*, 34:8780–8794, 2021. 3

[21] Laurent Dinh, David Krueger, and Yoshua Bengio. Nice: Non-linear independent components estimation. *arXiv preprint arXiv:1410.8516*, 2014. 3

[22] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using real nvp. *arXiv preprint arXiv:1605.08803*, 2016. 3, 5

[23] J. R. Dormand and P. J. Prince. A family of embedded Runge–Kutta formulae. *J. Comp. Appl. Math*, 6:19–26, 1980. 4

[24] Yarin Gal and Zoubin Ghahramani. Bayesian convolutional neural networks with bernoulli approximate variational inference. *arXiv preprint arXiv:1506.02158*, 2015. 2

[25] Mathieu Germain, Karol Gregor, Iain Murray, and Hugo Larochelle. Made: Masked autoencoder for distribution estimation. In *International conference on machine learning*, pages 881–889. PMLR, 2015. 3

[26] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020. 3

[27] Alexandros Graikos, Nikolay Malkin, Nebojsa Jojic, and Dimitris Samaras. Diffusion models as plug-and-play priors. In *Thirty-Sixth Conference on Neural Information Processing Systems*, 2022. 2, 3, 5

[28] Will Grathwohl, Ricky TQ Chen, Jesse Bettencourt, Ilya Sutskever, and David Duvenaud. Ffjord: Free-form continuous dynamics for scalable reversible generative models. *arXiv preprint arXiv:1810.01367*, 2018. 3

[29] Jonathan Ho, Xi Chen, Aravind Srinivas, Yan Duan, and Pieter Abbeel. Flow++: Improving flow-based generative models with variational dequantization and architecture design. In *International Conference on Machine Learning*, pages 2722–2730. PMLR, 2019. 3

[30] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020. 3

[31] Jonathan Ho, Chitwan Saharia, William Chan, David J Fleet, Mohammad Norouzi, and Tim Salimans. Cascaded diffusion models for high fidelity image generation. *J. Mach. Learn. Res.*, 23:47–1, 2022. 3

[32] Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*, 2022. 3

[33] Michael F Hutchinson. A stochastic estimator of the trace of the influence matrix for laplacian smoothing splines. *Communications in Statistics-Simulation and Computation*, 18(3):1059–1076, 1989. 4

[34] Aapo Hyvärinen and Erkki Oja. Independent component analysis: algorithms and applications. *Neural networks*, 13(4-5):411–430, 2000. 2

[35] Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa. Globally and locally consistent image completion. *ACM Transactions on Graphics (ToG)*, 36(4):1–14, 2017. 2

[36] Ajil Jalal, Marius Arvinte, Giannis Daras, Eric Price, Alexandros G Dimakis, and Jonathan I Tamir. Robust compressed sensing mri with deep generative priors. *NeurIPS*, 2021. 3, 5, 6

[37] Zahra Kadkhodaie and Eero Simoncelli. Stochastic solutions for linear inverse problems using the prior implicit in a denoiser. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 13242–13254. Curran Associates, Inc., 2021. 2

[38] Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. In *Advances in Neural Information Processing Systems*, 2022. 3, 5

[39] Bahjat Kawar, Roy Ganz, and Michael Elad. Enhancing diffusion-based image synthesis with robust classifier guidance. *arXiv preprint arXiv:2208.08664*, 2022. 3

[40] Bahjat Kawar, Gregory Vaksman, and Michael Elad. Snips: Solving noisy inverse problems stochastically. *NeurIPS*, 34:21757–21769, 2021. 2

[41] Patrick Kidger. *On Neural Differential Equations*. PhD thesis, University of Oxford, 2021. 4

[42] Patrick Kidger. On neural differential equations. *arXiv preprint arXiv:2202.02435*, 2022. 4

[43] Durk P Kingma and Prafulla Dhariwal. Glow: Generative flow with invertible 1x1 convolutions. *Advances in neural information processing systems*, 31, 2018. 3

[44] Diederik P. Kingma, Tim Salimans, Ben Poole, and Jonathan Ho. Variational diffusion models. *arXiv preprint arXiv:2107.00630*, 2021. 3

[45] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013. 3

[46] Polina Kirichenko, Pavel Izmailov, and Andrew G Wilson. Why normalizing flows fail to detect out-of-distribution data. *Advances in neural information processing systems*, 33:20578–20589, 2020. 3

[47] Kazuki Kuramochi, Kazunori Akiyama, Shiro Ikeda, Fumie Tazaki, Vincent L Fish, Hung-Yi Pu, Keiichi Asada, and Mareki Honma. Superresolution interferometric imaging with sparse modeling using total squared variation: application to imaging the black hole shadow. *The Astrophysical Journal*, 858(1):56, 2018. 2

[48] Hugo Larochelle and Iain Murray. The neural autoregressive distribution estimator. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 29–37. JMLR Workshop and Conference Proceedings, 2011. 3

[49] Rémi Laumont, Valentin De Bortoli, Andrés Almansa, Julie Delon, Alain Durmus, and Marcelo Pereyra. Bayesian imaging using plug & play priors: When langevin meets tweedie. *SIAM Journal on Imaging Sciences*, 15(2):701–737, 2022. 2

[50] 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. 3

[51] Cheng Lu, Kaiwen Zheng, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. Maximum likelihood training for score-based diffusion odes by high order denoising score matching. In *International Conference on Machine Learning*, pages 14429–14460. PMLR, 2022. 3

[52] Eric Nalisnick, Akihiro Matsukawa, Yee Whye Teh, Dilan Gorur, and Balaji Lakshminarayanan. Do deep generative models know what they don't know? *arXiv preprint arXiv:1810.09136*, 2018. 3

[53] Alex Nichol, Prafulla Dhariwal, Aditya Ramesh, Pranav Shyam, Pamela Mishkin, Bob McGrew, Ilya Sutskever, and Mark Chen. Glide: Towards photorealistic image generation and editing with text-guided diffusion models. *arXiv preprint arXiv:2112.10741*, 2021. 3

[54] Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International Conference on Machine Learning*, pages 8162–8171. PMLR, 2021. 3

[55] George Papamakarios, Theo Pavlakou, and Iain Murray. Masked autoregressive flow for density estimation. *Advances in neural information processing systems*, 30, 2017. 3

[56] Deepak Pathak, Philipp Krahenbuhl, Jeff Donahue, Trevor Darrell, and Alexei A Efros. Context encoders: Feature learning by inpainting. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2536–2544, 2016. 2

[57] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125*, 2022. 3

[58] Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. Stochastic backpropagation and approximate inference in deep generative models. In *International conference on machine learning*, pages 1278–1286. PMLR, 2014. 3

[59] Yaniv Romano, Michael Elad, and Peyman Milanfar. The little engine that could: Regularization by denoising (red). *SIAM Journal on Imaging Sciences*, 10(4):1804–1844, 2017. 2

[60] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10684–10695, 2022. 3

[61] Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily Denton, Seyed Kamyar Seyed Ghasemipour, Burcu Karagol Ayan, S Sara Mahdavi, Rapha Gontijo Lopes, et al. Photorealistic text-to-image diffusion models with deep language understanding. *arXiv preprint arXiv:2205.11487*, 2022. 3

[62] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022. 2

[63] John Skilling. The eigenvalues of mega-dimensional matrices. *Maximum Entropy and Bayesian Methods: Cambridge, England, 1988*, pages 455–466, 1989. 4

[64] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *Int. Conf. Machine Learning*, pages 2256–2265. PMLR, 2015. 3

[65] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020. 3

[66] Jiaming Song, Arash Vahdat, Morteza Mardani, and Jan Kautz. Pseudoinverse-guided diffusion models for inverse problems. In *International Conference on Learning Representations*, 2023. 2, 5

[67] Yang Song, Conor Durkan, Iain Murray, and Stefano Ermon. Maximum likelihood training of score-based diffusion models. In *Thirty-Fifth Conference on Neural Information Processing Systems*, 2021. 3

[68] Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. In *NeurIPS*, pages 11895–11907, 2019. 3

[69] Yang Song, Liyue Shen, Lei Xing, and Stefano Ermon. Solving inverse problems in medical imaging with score-based generative models. In *ICLR*, 2022. 2, 3, 5, 6

[70] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *ICLR*, 2021. 2, 3, 4

[71] He Sun and Katherine L Bouman. Deep probabilistic imaging: Uncertainty quantification and multi-modal solution characterization for computational imaging. In *AAAI*, pages 2628–2637, 2021. 4, 5

[72] He Sun, Katherine L Bouman, Paul Tiede, Jason J Wang, Sarah Blunt, and Dimitri Mawet. alpha-deep probabilistic inference (alpha-dpi): efficient uncertainty quantification from exoplanet astrometry to black hole feature extraction. *arXiv preprint arXiv:2201.08506*, 2022. 5

[73] Ch Tsitouras. Runge–kutta pairs of order 5 (4) satisfying only the first column simplifying assumption. *Computers & Mathematics with Applications*, 62(2):770–775, 2011. 4

[74] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *CVPR*, pages 9446–9454, 2018. 2

[75] Aäron Van Den Oord, Nal Kalchbrenner, and Koray Kavukcuoglu. Pixel recurrent neural networks. In *International conference on machine learning*, pages 1747–1756. PMLR, 2016. 3

[76] Singanallur V Venkatakrishnan, Charles A Bouman, and Brendt Wohlberg. Plug-and-play priors for model based reconstruction. In *2013 IEEE Global Conference on Signal and Information Processing*, pages 945–948. IEEE, 2013. 2

[77] Lilian Weng. What are diffusion models? *lilianweng.github.io*, Jul 2021. 3

[78] Tianwei Yin, Zihui Wu, He Sun, Adrian V Dalca, Yisong Yue, and Katherine L Bouman. End-to-end sequential sampling and reconstruction for mri. *arXiv preprint arXiv:2105.06460*, 2021. 2

[79] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. Free-form image inpainting with gated convolution. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4471–4480, 2019. 2

[80] Jinwei Zhang, Hang Zhang, Alan Wang, Qihao Zhang, Mert Sabuncu, Pascal Spincemaille, Thanh D Nguyen, and Yi Wang. Extending loupe for k-space under-sampling pattern optimization in multi-coil mri. In *International Workshop on Machine Learning for Medical Image Reconstruction*, pages 91–101. Springer, 2020. 2

[81] Kaihao Zhang, Wenqi Ren, Wenhan Luo, Wei-Sheng Lai, Björn Stenger, Ming-Hsuan Yang, and Hongdong Li. Deep image deblurring: A survey. *International Journal of Computer Vision*, 130(9):2103–2130, 2022. 2

[82] Daniel Zoran and Yair Weiss. From learning models of natural image patches to whole image restoration. In *ICCV*, pages 479–486. IEEE, 2011. 2

[83] Daniel Zoran and Yair Weiss. Natural images, gaussian mixtures and dead leaves. *NeurIPS*, 25, 2012. 2