# Diffusion Prior-Based Amortized Variational Inference for Noisy Inverse Problems

Sojin Lee[1⋆], Dogyun Park[1⋆], Inho Kong[1], and Hyunwoo J. Kim[1†]

Korea University, Seoul, Republic of Korea
{sojin_lee,gg933,inh212,hyunwoojkim}@korea.ac.kr

**Abstract.** Recent studies on inverse problems have proposed posterior samplers that leverage the pre-trained diffusion models as powerful priors. These attempts have paved the way for using diffusion models in a wide range of inverse problems. However, the existing methods entail computationally demanding iterative sampling procedures and optimize a separate solution for each measurement, which leads to limited scalability and lack of generalization capability across unseen samples. To address these limitations, we propose a novel approach, Diffusion prior-based Amortized Variational Inference (DAVI) that solves inverse problems with a diffusion prior from an amortized variational inference perspective. Specifically, instead of separate measurement-wise optimization, our amortized inference learns a function that directly maps measurements to the implicit posterior distributions of corresponding clean data, enabling a single-step posterior sampling even for unseen measurements. Extensive experiments on image restoration tasks, *e.g.*, Gaussian deblur, 4× super-resolution, and box inpainting with two benchmark datasets, demonstrate our approach's superior performance over strong baselines. Code is available at `https://github.com/mlvlab/DAVI`.

**Keywords:** Inverse Problems · Diffusion Models · Posterior Sampling

## 1 Introduction

Noisy inverse problems are important and have a wide range of real-world applications such as image restoration [16, 43], medical imaging [25, 34], and astronomy [11, 37]. Typically, noisy inverse problems involve estimating the original clean data $\mathbf{x}_0$ from a noisy observation $\mathbf{y}$. The forward (measurement) model is formulated as

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

where a known degradation matrix $\mathbf{H} \in \mathbb{R}^{d_{\mathbf{y}} \times d_{\mathbf{x}_0}}$ is applied to the clean signal $\mathbf{x}_0 \in \mathbb{R}^{d_{\mathbf{x}_0}}$ with i.i.d white Gaussian noise $\mathbf{n} \in \mathbb{R}^{d_{\mathbf{y}}}$. Then, the likelihood of the measurement is defined as $p(\mathbf{y}|\mathbf{x}_0) = \mathcal{N}(\mathbf{y}|\mathbf{H}\mathbf{x}_0, \sigma_{\mathbf{y}}^2\mathbf{I})$. However, it is challenging to accurately estimate the solution $\mathbf{x}_0$ due to its ill-posed nature [28] where there exist multiple solutions $\mathbf{x}$ for a measurement $\mathbf{y}$, *i.e.*, many-to-one $\mathbf{x} \mapsto \mathbf{y}$.

⋆ equal contribution, †corresponding author

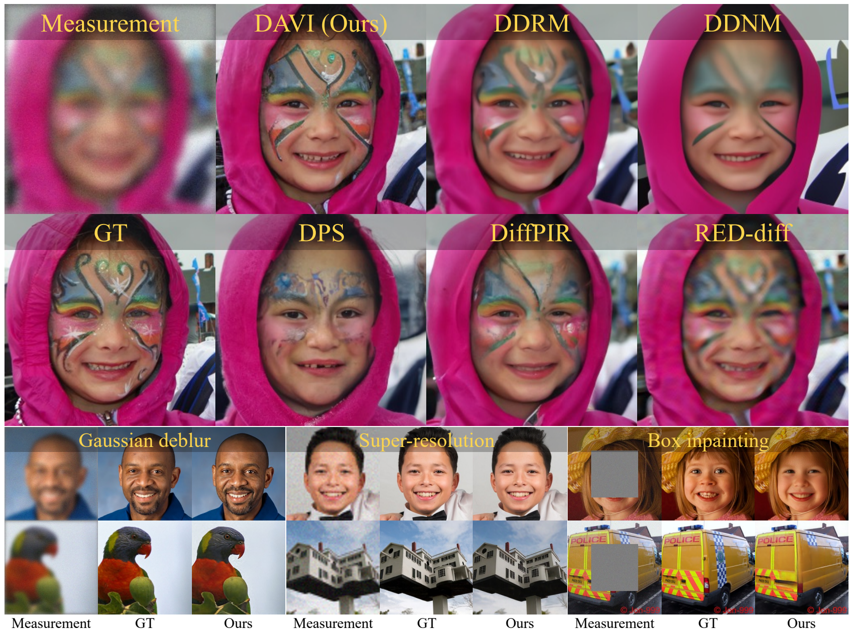


**Fig. 1: Representative results of Diffusion prior-based Amortized Variational Inference (DAVI)**. The top row demonstrates the qualitative comparison between our method and baselines. The bottom two rows showcase that DAVI provides robust solutions with fine-grained details across various image restoration tasks, achieved with a *single neural network evaluation*.

Diffusion models have achieved remarkable success across various applications, including image [20,42], video [2,41], 3D [27,40], and domain-agnostic [26] generation, encompassing 2D, 3D, and video. Inspired by these impressive results, recent approaches [3, 4, 6, 8, 18, 32, 39, 45] have leveraged a pre-trained diffusion model [13,15,31,33] as a powerful prior for solving image inverse problems. Existing diffusion-based methods alter the reverse sampling process of the pre-trained diffusion model either by approximating the posterior score function via Bayes' rule [4, 32] or running the reverse process in the decomposed space [18,39]. Mardani [23] proposes a variational approach relying on the diffusion sampling process to approximate the mode of the true posterior distribution. Despite the promising results of these previous methods, their necessity for the *iterative sampling procedure* limits their scalability. This makes it difficult to deploy the diffusion-based methods on commodity devices for real-time applications. In addition, *independent optimization* for each sample is suboptimal, leading to poor generalization on unseen samples. Arguably, solving the inverse problems for multiple measurements together is more effective if a method seeks a generalizable function beyond one clean sample $\mathbf{x}_0$. The estimated function is readily applicable to unseen samples even without any iterative procedures.

Thus, we introduce **D**iffusion prior-based **A**mortized **V**ariational **I**nference (DAVI) that addresses inverse problems using a diffusion prior within an amortized variational inference framework. Unlike previous approaches, DAVI learns a function that associates a measurement with the implicit posterior distribution of the corresponding clean data. This enables efficient single-step posterior sampling for both seen and unseen measurements. Our method optimizes this function by minimizing the Kullback-Leibler (KL) divergence between the implicit and the true posterior distributions for multiple measurements, employing objectives from variational inference.

Our **contributions** are summarized as follows:

- We propose a novel approach, Diffusion prior-based Amortized Variational Inference (DAVI), which solves inverse problems with a diffusion prior from an amortized variational inference perspective.
- Our framework enables efficient posterior sampling by a single evaluation of a neural network and generalization for both seen and unseen measurements without any optimization at test time.
- We propose a novel Perturbed Posterior Bridge that provides intermediary measurements to further enhance the generalization capabilities.
- Our extensive experiments demonstrate the effectiveness of our proposed method in image restoration tasks.

## 2 Related Works

### 2.1 Diffusion models for inverse problems

Several recent studies [4, 7, 8, 18, 23, 24, 32, 45] have focused on solving inverse problems using pre-trained diffusion models as priors due to their strong ability to model complex distributions. Specifically, DDRM [18] proposes running the diffusion model in a spectral domain through the singular value decomposition (SVD) of the signal space. However, computing SVD can be computationally expensive, especially for high-dimensional signals, and is not feasible for complex degradation operators such as motion blur. While DDNM [39] and $\Pi$GDM [32] introduce pseudo-inverse guidance during the reverse diffusion process without computing SVD, these methods often struggle with noisy measurements due to the inherent error in estimating the pseudo-inverse with a noisy measurement. DPS [4] proposes sampling from a diffusion posterior $p(\mathbf{x}_0|\mathbf{y})$ by an approximation of the intractable time-dependent likelihood $p(\mathbf{y}|\mathbf{x}_t)$ using Tweedie's formula. FPS [8] connects Bayesian posterior sampling and Bayesian filtering in diffusion models, where multiple samples share a single measurement. Like our method, RED-diff [23] explores the variational perspective of inverse problems by optimizing in pixel space. However, these methods heavily depend on reverse diffusion trajectories for inference, which inherently requires multiple neural network evaluations. In contrast, our work aims to solve the inverse problem with a single step, which is significantly more efficient.

### 2.2 Amortized variational inference

Variational inference [9, 14, 19] (VI) approximates the posterior with a parameterized variational distribution family $\mathcal{Q}$ and finds the optimal distribution from $\mathcal{Q}$ that minimizes the KL divergence from the true posterior. Several VI approaches [36], such as the mean-field family, optimize an independent set of variational distributions for each data sample, causing the parameters to scale with the dataset size. This process, inherently memoryless [10], can pose significant computational inefficiency, particularly for large datasets. On the other hand, Amortized VI (AVI) amortizes the optimization by using a stochastic function, typically a neural network, to represent the entire variational distribution family $\mathcal{Q}$. AVI memorizes past inferences by optimizing a single neural network across multiple samples rather than optimizing each sample independently. The optimized network can generalize to unseen samples by using information from previous samples without additional optimization costs. This allows for computational efficiency compared to per-data optimization. In our work, we adopt amortized VI to address the computationally intensive sampling processes in previous methods, offering a more efficient and scalable solution.

## 3 Preliminary

Diffusion models [13, 35] define stochastic differential equations (SDEs) for the diffusion forward process $\{\mathbf{x}_t\}_{t=0}^{T}$, where $\mathbf{x}_t$ is the perturbed data at time $t$. The data distribution is defined at $t = 0$, *i.e.*, $\mathbf{x}_0 \sim p(\mathbf{x}_0)$ and the prior distribution is achieved at $t = T$, following the standard normal distribution, *i.e.*, $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. The forward SDE is defined as

$$d\mathbf{x} = f_t \mathbf{x}_t \, dt + g_t \, d\mathbf{w}, \tag{2}$$

where $f_t : \mathbb{R} \to \mathbb{R}$ is a drift coefficient, $g_t \in \mathbb{R} \to \mathbb{R}$ is a diffusion coefficient, and $\mathbf{w}$ represents a standard Wiener process. Song et.al [35] define reverse SDE that achieves the data distribution from the prior distribution using the score function $\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t)$ as

$$d\mathbf{x} = \left[ f_t \mathbf{x}_t - g_t^2 \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t) \right] dt + g_t \, d\bar{\mathbf{w}}, \tag{3}$$

where $dt$ is an infinitesimal negative timestep and $\bar{\mathbf{w}}$ is reverse-time standard Wiener process [1]. The time-dependent score function $\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t)$ is estimated by a neural network $s_\theta(\mathbf{x}_t, t)$, *i.e.*, $\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t) \approx s_\theta(\mathbf{x}_t, t)$, by minimizing the following denoising score matching objective:

$$\mathbb{E}_{p(\mathbf{x}_t|\mathbf{x}_0), t \sim \mathrm{U}[0,T]} \left[ \lambda(t) \| \mathbf{s}_\theta(\mathbf{x}_t, t) - \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t|\mathbf{x}_0) \|_2^2 \right], \tag{4}$$

where $p(\mathbf{x}_t|\mathbf{x}_0)$ is a Gaussian transition kernel from time 0 to $t$, and $\lambda(t)$ is a time-dependent weighting function. For Variance Preserving (VP) SDE [35] or DDPM [13], the transition kernel $q(\mathbf{x}_t|\mathbf{x}_0)$ is defined as $\mathbf{x}_t = \sqrt{\alpha_t}\mathbf{x}_0 + \sqrt{1 - \alpha_t}\epsilon$, $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, where $\alpha_t = \prod_{s=1}^{t}(1 - \beta_t)$ and $\beta_t = g_t^2$.

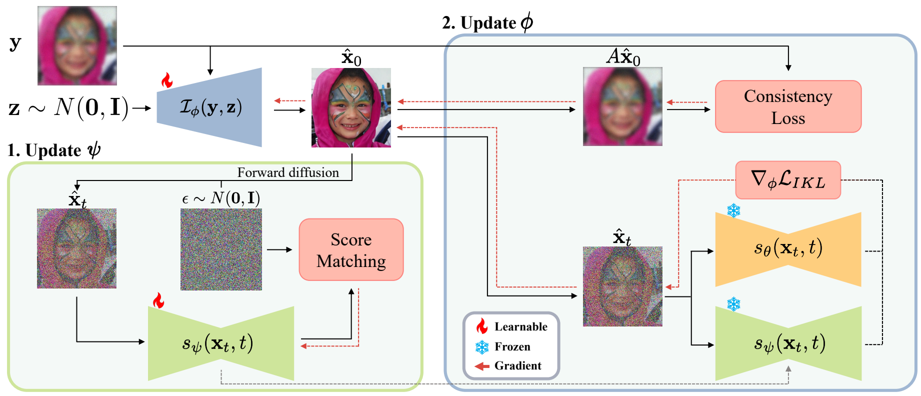


**Fig. 2: Illustration of Diffusion prior-based Amortized Variational Inference (DAVI).** Our proposed method employs an alternative optimization procedure between the score function $\mathbf{s}_\psi$ and the neural network $\mathcal{I}_\phi$ to minimize the KL divergence between the implicit posterior distribution $q_\phi(\mathbf{x}_0|\mathbf{y})$ and the true posterior distribution $p(\mathbf{x}_0|\mathbf{y})$, where the true posterior distribution is approximated by the likelihood $p(\mathbf{y}|\mathbf{x}_0)$ and the diffusion prior $\mathbf{s}_\theta$.

## 4 Proposed Method

We propose a novel approach, Diffusion prior-based Amortized Variational Inference (DAVI), that tackles inverse problems by leveraging a pre-trained diffusion model as a prior distribution from an amortized variational inference perspective [9, 30]. To be specific, our framework learns a neural network that directly maps a measurement $\mathbf{y}$ to the implicit posterior distribution $p(\mathbf{x}_0|\mathbf{y})$ of the corresponding clean data $\mathbf{x}_0$. This enables efficient posterior sampling via a single evaluation of the neural network and generalization over unseen measurements (*i.e.*, zero-shot inference). In Section 4.1 and 4.2, we introduce training objectives that minimize the KL divergence between the implicit distribution and the true posterior distribution based on variational inference. In Section 4.3, we propose a novel Perturbed Posterior Bridge (PPB) to provide intermediary measurements to further enhance the generalization power of the proposed method.

### 4.1 Diffusion prior-based Amortized Variational Inference (DAVI)

Our framework approximates the posterior distribution $p(\mathbf{x}_0|\mathbf{y})$, which is the distribution of the clean sample $x_0$ given the measurement $y$, using an implicit distribution $\mathcal{I}_\phi$ parameterized by a neural network. Sampling from the approximated distribution is performed by the reparameterization trick with a random Gaussian noise $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ as follows:

$$\hat{\mathbf{x}} = \mathcal{I}_\phi(\mathbf{y} + h \cdot \mathbf{z}), \quad \mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \tag{5}$$

where $\hat{x}$ indicates the sample from the implicit distribution and $h \in \mathbb{R}_{++}$ is a hyperparameter that determines the variance of the noise. Variational optimization between our implicit distribution and the true posterior distribution is formulated as

$$\phi^\star = \arg\min_{\phi} \left[ D_{KL}(q_\phi(\mathbf{x}_0|\mathbf{y}) \parallel p(\mathbf{x}_0|\mathbf{y})) \right], \tag{6}$$

where $q_\phi(\mathbf{x}_0|\mathbf{y})$ and $p(\mathbf{x}_0|\mathbf{y})$ are the implicit distribution and the true posterior, respectively. By the definition of KL-divergence and dropping the constant term '$\log p(\mathbf{y})$', the objective function can be rewritten as

$$-\mathbb{E}_{q_\phi(\mathbf{x}_0|\mathbf{y})} \left[ \log p(\mathbf{y}|\mathbf{x}_0) \right] + D_{KL}(q_\phi(\mathbf{x}_0|\mathbf{y}) \parallel p(\mathbf{x}_0)). \tag{7}$$

The constant term '$\log p(\mathbf{y})$' can be ignored since it is independent of the parameter $\phi$. Then, the first term is the likelihood of the measurement $\mathbf{y}$, *i.e.*, $\log p(\mathbf{y}|\mathbf{x}_0)$, which we refer to as a data consistency loss. The second term, the KL-divergence between the approximated posterior distribution $q_\phi(\mathbf{x}_0|\mathbf{y})$ and the prior $p(\mathbf{x}_0)$, is represented by the pre-trained diffusion model $p_\theta(\mathbf{x}_0)$. We discuss the terms below in further detail.

**Data consistency loss.** The data consistency loss is

$$\mathcal{L}_C = \mathbb{E}_{q_\phi(\mathbf{x}_0|\mathbf{y})} \left[ \frac{\|\mathbf{y} - \mathbf{H}\mathbf{x}_0\|_2^2}{2\sigma_{\mathbf{y}}^2} \right], \tag{8}$$

since the likelihood $p(\mathbf{y}|\mathbf{x}_0)$ can be analytically calculated by Eq. (1). Note that the constant term is omitted. By minimizing the data consistency loss, we incorporate measurement information into the implicit distribution.

**Integral KL divergence (IKL).** Optimizing the KL divergence between approximated posterior distribution and diffusion prior poses a risk of the KL divergence potentially diverging towards infinity, especially when the supports of two distributions are misaligned, leading to unstable and suboptimal results. To mitigate this issue, Luo et al. [22] have proposed the Integral KL divergence (IKL). Motivated by the IKL divergence, we upper bound the second term with the integral of KL divergence between $q_\phi(\mathbf{x}_t|\mathbf{y})$ and $p_\theta(\mathbf{x}_t)$ over $t$, and denote it as $\mathcal{L}_{IKL}$. It is formulated as:

$$\mathcal{L}_{IKL} = \int_{t=0}^{T} w(t) D_{KL}(q_\phi(\mathbf{x}_t|\mathbf{y}) \parallel p_\theta(\mathbf{x}_t)) dt, \tag{9}$$

where we achieve $\mathbf{x}_t$ by the forward SDE in Eq. (2) and $w(t)$ is a positive weighting function, notably with $w(0) = 1$. Roughly speaking, the forward SDE can be viewed as a set of Gaussian filters with various bandwidths. It transforms the two distributions into smoother distributions with infinite supports, alleviating the disjoint support problem. As the perturbation or $t$ increases, the densities overlap more, leading to more robust convergence than the original KL divergence (see Fig. 3). This loss can be interpreted as the KL-divergence between two sets

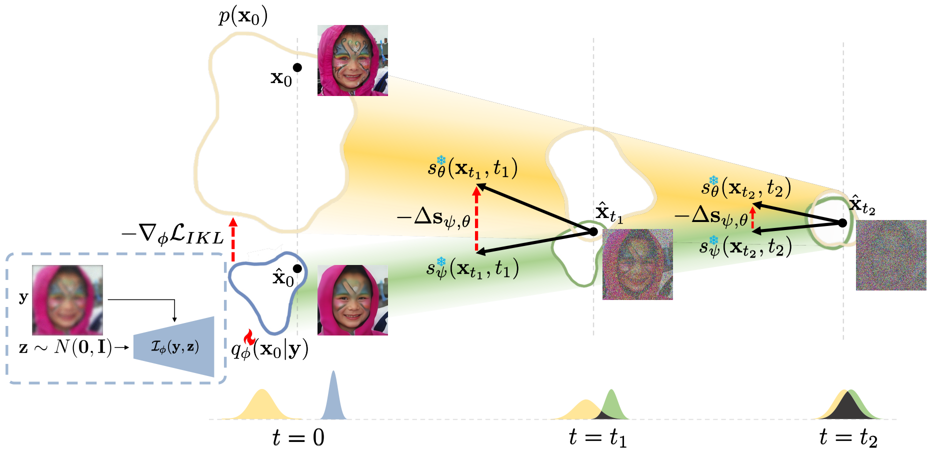


**Fig. 3: Visualization of integral KL divergence.** IKL loss perturbs the implicit distribution into a smoother distribution using forward SDE to alleviate the disjoint support problem (see $t = t_1, t = t_2$, and $t_1 < t_2$). The gradient of IKL loss, $\nabla_\phi \mathcal{L}_{IKL}$, updates the parameters of the implicit distribution in a direction that minimizes the $\triangle\ \mathbf{s}_{\psi,\theta}$, which leads to minimizing the discrepancy between $q_\phi(\mathbf{x}_0|\mathbf{y})$ and $p(\mathbf{x}_0)$.

of distributions after smoothing by Gaussian kernels with various bandwidths (*i.e.*, the forward diffusion process at various time points). As $\mathcal{L}_{IKL}$ decreases, the samples from the implicit posterior distribution increasingly resemble the cleaner images of the diffusion prior.

### 4.2 Score distillation gradient

Here, we present the procedure to learn the implicit distribution with the IKL loss. The gradient of IKL loss with respect to $\phi$, *i.e.*, $\nabla_\phi \mathcal{L}_{IKL}$, can be analytically derived as

$$\int_{t=0}^{T} w(t)\mathbb{E}_{q_\phi(\mathbf{x}_t|\mathbf{y})}[\nabla_{\mathbf{x}_t} \log q_\phi(\mathbf{x}_t|\mathbf{y}) - \nabla_{\mathbf{x}_t} \log p_\theta(\mathbf{x}_t)]\, \frac{\partial \mathbf{x}_t}{\partial \phi} dt, \tag{10}$$

where $q_\phi(\mathbf{x}_t|\mathbf{y}) = \int q(\mathbf{x}_t|\mathbf{x}_0) \cdot q_\phi(\mathbf{x}_0|\mathbf{y}) d\mathbf{x}_0$. Here, $q(\mathbf{x}_t|\mathbf{x}_0)$ is the forward diffusion kernel that is identically defined as $p(\mathbf{x}_t|\mathbf{x}_0)$ (see supplement Section A.1 for detailed derivation). Eq. (10) requires the score estimation for both distributions, $p_\theta$ and $q_\phi$, respectively. We utilize a pre-trained diffusion model $\mathbf{s}_\theta$ to approximate $\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t)$; however, a corresponding score function for $q_\phi(\mathbf{x}_t|\mathbf{y})$ is not tractable for the implicit distribution. To address this, we introduce an implicit score function, denoted as $\mathbf{s}_\psi$, and train it to approximate $\nabla_{\mathbf{x}_t} \log q_\phi(\mathbf{x}_t|\mathbf{y})$. Then, Eq. (10) is approximated as

$$\nabla_\phi \mathcal{L}_{IKL} \approx \mathbb{E}_{q_\phi(\mathbf{x}_t|\mathbf{y}),t}\left[w(t)(\triangle\ \mathbf{s}_{\psi,\theta} \cdot \frac{\partial \mathbf{x}_t}{\partial \phi})\right], \tag{11}$$

**Algorithm 1** Training

**Require:** $\mathbf{H}, K, T, \mathcal{I}_{\phi,a}, \mathbf{s}_\psi, \mathbf{s}_\theta, w, \gamma, h$

**for** $k = 1, ..., K$ **do**

  $\mathbf{x}_0 \sim p(\mathbf{x}_0)$

  $\mathbf{y} \sim p(\mathbf{y}|\mathbf{x}_0)$

  $\mathbf{y}_a = (1-\sigma_a)\mathbf{y} + \sigma_a \mathbf{x}_0 + h\bar{\sigma}_a \mathbf{z}$, where $a \sim p(a)$, $z \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$

  $\hat{\mathbf{x}}_0 = \mathcal{I}_{\phi,a}(\mathbf{y}_a)$

  $t \sim \mathcal{U}(0, T)$

  Draw $\hat{\mathbf{x}}_t \sim q(\mathbf{x}_t|\mathbf{x}_0)$ using Forward SDE (Eq. (2))

  $\mathcal{L}_S = \|\mathbf{s}_\psi(\mathbf{x}_t, t) - \nabla_{\mathbf{x}_t} \log q_\phi(\mathbf{x}_t|\mathbf{y})\|_2^2$ (Eq. (12))

  Update $\psi$ with $\nabla_\psi \mathcal{L}_S$

  $\mathcal{L}_C = \gamma\|\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}_0\|_2^2|$ (Eq. (8))

  $\nabla_\phi \mathcal{L}_{IKL} = w(t)(\mathbf{s}_\psi(\mathbf{x}_t, t) - \mathbf{s}_\theta(\mathbf{x}_t, t))\frac{\partial \mathbf{x}_t}{\partial \phi}$ (Eq. (11))

  Update $\phi$ with $\nabla_\phi(\mathcal{L}_C + \mathcal{L}_{IKL})$

**end for**

**return** $\mathcal{I}_{\phi,a}$

**Algorithm 2** Single-step Inference

**Require:** $\mathbf{y}, \mathcal{I}_{\phi,a}$

$\hat{\mathbf{x}}_0 = \mathcal{I}_{\phi,1}(\mathbf{y} + h\mathbf{z})$, $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$

**return** $\hat{\mathbf{x}}_0$

where $\triangle\, \mathbf{s}_{\psi,\theta} \coloneqq \mathbf{s}_\psi(\mathbf{x}_t, t) - \mathbf{s}_\theta(\mathbf{x}_t, t)$ and $\mathbf{s}_\psi$ indicates the score function of $q_\phi$. Roughly speaking, IKL updates $\phi$ in the direction that minimizes the discrepancy $\triangle\, \mathbf{s}_{\psi,\theta}$ between $\mathbf{s}_\psi$ and $\mathbf{s}_\theta$ and this eventually minimizes the discrepancy between $q_\phi(\mathbf{x}|\mathbf{y})$ and $p_\theta(\mathbf{x})$, as demonstrated in Fig. 3.

Lastly, we train the score function $\mathbf{s}_\psi$ by the denoising score matching loss described in Eq. (4) to approximate the score function of $q_\phi$:

$$\mathcal{L}_S = \mathbb{E}_{q_\phi(\mathbf{x}_t|\mathbf{y}),t}\left[\|\mathbf{s}_\psi(\mathbf{x}_t, t) - \nabla_{\mathbf{x}_t} \log q_\phi(\mathbf{x}_t|\mathbf{y})\|_2^2\right]. \tag{12}$$

**Alternating optimization.** Since the score of $q_\phi(\mathbf{x}_t|\mathbf{y})$ evolves based on the implicit distribution $q_\phi(\mathbf{x}_0|\mathbf{y})$ as the optimization progresses, we need to estimate the score function of $q_\phi$ accordingly. Therefore, we propose an alternating optimization approach for training two parameters $\phi$ and $\psi$ with separate objectives (Eq. (7) and Eq. (12)) as described in Algorithm 1 and Fig. 2.

### 4.3 Perturbed Posterior Bridge

To further improve the generalization power of our framework, we propose a Perturbed Posterior Bridge (PPB), which acts as an intermediary set of trajectories between $\mathbf{y}$ and $\mathbf{x}$, to facilitate a more guided and effective training of the

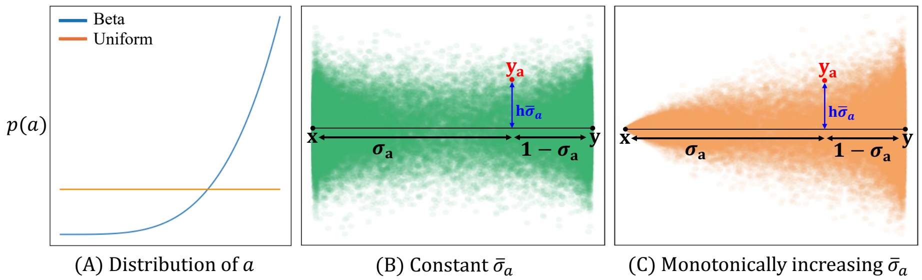


**Fig. 4: Perturbed Posterior Bridge (PPB).** (A) shows sampling distributions of $a$. (B) and (C) illustrate the PPB between two 1D samples, *e.g.*, $\mathbf{x} = 0$ and $\mathbf{y} = 1$, with different perturbation schedule $\bar{\sigma}_a$. We plot the perturbation along the $y$ axis.

neural network $\mathcal{I}_\phi$. The sample $\mathbf{y}_a$ drawn from PPB is defined as:

$$\mathbf{y}_a = (1 - \sigma_a)\mathbf{y} + \sigma_a \mathbf{x} + h\bar{\sigma}_a \mathbf{z}, \quad \mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \tag{13}$$

where $a \in [0, 1]$, $\sigma_a = \frac{\int_a^1 \beta_t dt}{\int_0^a \beta_t dt + \int_a^1 \beta_t dt}$, and $\beta_t$ is a diffusion noise schedule [13]. The function $\sigma_a$ is defined to ensure the conditions $\sigma_a \in [0, 1]$, $\sigma_0 = 1$, and $\sigma_1 = 0$. This creates a bridge, modulated by $\sigma_a$, with an additional perturbation term $h\bar{\sigma}_a$. The $\bar{\sigma}_a$ determines the deviation from the linear bridge between $\mathbf{y}$ and $\mathbf{x}$ (see Fig. 4 (B) and (C)). Our experiments in Tab. 5 show that the PPB with monotonically increasing $\bar{\sigma}_a$ as approaching $\mathbf{y}$ in Fig. 4 (C) is more effective than a constant perturbation Fig. 4 (B).

Now, we extend the implicit distribution $\mathcal{I}_\phi$ with PPB and $a$ as:

$$\hat{\mathbf{x}} = \mathcal{I}_{\phi,a}(\mathbf{y}_a) = \begin{cases} \mathcal{I}_{\phi,0}(\mathbf{x}), & \text{if } a = 0 \\ \mathcal{I}_{\phi,a}((1 - \sigma_a)\mathbf{y} + \sigma_a \mathbf{x} + (\sigma_a(1 - \sigma_a) + h\bar{\sigma}_a)\,\mathbf{z}), & \text{elif } 0 < a < 1 \\ \mathcal{I}_{\phi,1}(\mathbf{y} + h\mathbf{z}), & \text{otherwise} \end{cases} \tag{14}$$

where $a$ is encoded using the sinusoidal positional embedding [38] and fed into each block of the neural network $\mathcal{I}_{\phi,a}$. Note that when $a = 1$, Eq. 14 is reduced to the implicit distribution defined in Eq. 5. This approach allows us to optimize the implicit distribution with more diverse samples drawn from PPB between $\mathbf{y}$ and $\mathbf{x}$. We observed it effectively enhances the generalization to unseen measurements. During the training stage, we randomly sample $a$ from pre-defined distribution $p(a)$, *i.e.*, $a \sim p(a)$, and the overall training procedure is described in Alg. 1.

**Inference.** Our single step inference is described in Alg. 2. Our framework generalizes well on unseen measurements $\mathbf{y}$. We sample a random noise $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, and obtain the posterior sample via a single evaluation of a neural network as $\hat{\mathbf{x}} = \mathcal{I}_{\phi,1}(\mathbf{y} + h\mathbf{z})$.

# 5 Experiments

## 5.1 Experimental Setups

**Baselines.** To demonstrate the efficacy of our method for solving noisy inverse problems, we compare DAVI with state-of-the-art methods such as DDRM [18], $\text{DDNM}^{+}$ [39], DPS [4], $\Pi$GDM [32], DiffPIR [45], and RED-diff [23]. The comparisons are conducted on the FFHQ [17] 1K and ImageNet [29] 1K validation datasets at a resolution of $256\times256$.
**Tasks.** We focus on three challenging image restoration tasks: *Gaussian deblurring*, $4\times$ *super-resolution*, and *box inpainting* with $128\times128$ size of the mask. Specifically, we employ a Gaussian blur kernel with $61\times61$ size, a standard deviation of 3.0 for deblurring, and an average-pooling operation for super-resolution. Additionally, Gaussian noise with a standard deviation ($\sigma_{\mathbf{y}}$) of 0.05 is introduced to the images to simulate measurement noise unless stated otherwise. In Section B.1 of the supplement, we provide additional experiments such as *denoising* and *colorization* tasks, and adopt Poisson noise results.
**Evaluation metrics.** For evaluation, we focus on two aspects: 1) *consistency* with the original measurement and 2) the *realism* of the restored images. We employ the Peak Signal-to-Noise Ratio (PSNR) to assess measurement fidelity; however, it is noteworthy that PSNR prefers blurry images [44]. To address this, we complement it with the Learned Perceptual Image Patch Similarity (LPIPS) score [44], which measures structural fidelity. To evaluate the realism of restored images, we employ the Frechet Inception Distance (FID) [12], which better aligns with human visual perception. We also report the number of function evaluations (NFEs) to compare the method's efficiency.
**Implementation details.** All methods utilize a pre-trained unconditional diffusion model $\mathbf{s}_\theta$ used in [4], for the prior $p_\theta(\mathbf{x}_0)$. We employ the same architecture for the implicit distribution $\mathcal{I}_\phi$ and the implicit score function $\mathbf{s}_\psi$, and initialize $\mathcal{I}_\phi$ and $\mathbf{s}_\psi$ with the same pre-trained diffusion model for better convergence. In the optimization stage, we use the dataset that trained the pre-trained diffusion model $\mathbf{s}_\theta$ to obtain clean and degraded pairs. We set $p(a)$ as a beta distribution and $\bar{\sigma}_a$ as a monotonically increasing function with $a$. Note that we measure the performance on a *distinct* validation set. For further details, refer to the supplement Section C.1.

## 5.2 Quantitative Results

We demonstrate the quantitative results on FFHQ and ImageNet in Tab. 1 and 2, respectively. The results clearly show that our proposed method, DAVI, surpasses the leading works in all scenarios, especially in FID and LPIPS, even in a *single-step*. For example, on the super-resolution task, DAVI achieves a remarkable improvement in FID score, with a gain of 12.33 points on FFHQ and 7.36 points on ImageNest compared to DPS, which records the best FID score among the baselines. Moreover, on the box inpainting task, DAVI enhances the FID score by 16.96 points on FFHQ and 5.34 points on ImageNet compared to

**Table 1: Comparative results for noisy inverse problems (Gaussian deblur, super-resolution, and box inpainting with centered 128x128 mask) on the FFHQ 256×256**. The highest performance is highlighted in bold, while the second-highest is underlined.

| | | Gaussian deblur | | | 4× Super-resolution | | | Box Inpainting | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Method | NFE ↓ | PSNR ↑ | LPIPS ↓ | FID ↓ | PSNR ↑ | LPIPS ↓ | FID ↓ | PSNR ↑ | LPIPS ↓ | FID ↓ |
| DDRM [18] | 20 | 26.26 | 0.269 | 56.82 | 28.09 | 0.228 | 48.19 | 22.27 | 0.177 | 37.62 |
| DDNM$^+$ [39] | 100 | 24.19 | 0.348 | 91.48 | 28.17 | 0.244 | 59.09 | 23.79 | 0.218 | 48.82 |
| DPS [4] | 1000 | 21.88 | 0.315 | 34.47 | 25.55 | 0.255 | 36.29 | 21.94 | 0.302 | 42.45 |
| $\Pi$GDM [32] | 100 | 23.05 | 0.320 | 52.72 | 27.73 | 0.225 | 49.86 | 23.04 | 0.220 | 32.99 |
| DiffPIR [45] | 100 | 24.41 | 0.299 | 33.91 | 25.32 | 0.341 | 40.35 | 23.97 | 0.195 | 31.27 |
| RED-diff [23] | 1000 | **26.44** | 0.324 | 46.55 | 26.75 | 0.379 | 92.82 | 24.16 | 0.216 | 35.80 |
| **DAVI (Ours)** | **1** | 25.46 | **0.225** | **29.92** | **28.23** | **0.171** | **23.96** | **26.25** | **0.112** | **14.31** |

**Table 2: Comparative results for noisy inverse problems (Gaussian deblur, super-resolution, and box inpainting with centered 128x128 mask) on the ImageNet 256×256**. The highest performance is highlighted in bold, while the second-highest is underlined.

| | | Gaussian deblur | | | 4× Super-resolution | | | Box Inpainting | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Method | NFE ↓ | PSNR ↑ | LPIPS ↓ | FID ↓ | PSNR ↑ | LPIPS ↓ | FID ↓ | PSNR ↑ | LPIPS ↓ | FID ↓ |
| DDRM [18] | 20 | **23.96** | 0.384 | 69.59 | 25.83 | 0.300 | 48.47 | 18.10 | 0.260 | 75.32 |
| DDNM$^+$ [39] | 100 | 22.16 | 0.486 | 120.16 | 26.37 | 0.278 | 46.05 | 20.17 | 0.285 | 92.34 |
| DPS [4] | 1000 | 19.86 | 0.444 | 64.42 | 24.92 | 0.311 | 43.63 | 18.87 | 0.391 | 72.48 |
| $\Pi$GDM [32] | 100 | 21.72 | 0.443 | 81.32 | 25.53 | 0.321 | 60.61 | 18.83 | 0.233 | 71.42 |
| DiffPIR [45] | 100 | 22.10 | 0.400 | **62.48** | 23.08 | 0.385 | 57.39 | 20.26 | 0.259 | 68.38 |
| RED-diff [23] | 1000 | 23.66 | 0.448 | 100.80 | 24.82 | 0.406 | 84.72 | 20.26 | 0.276 | 74.38 |
| **DAVI (Ours)** | **1** | 23.73 | **0.343** | 63.29 | **26.58** | **0.242** | **36.27** | **21.96** | **0.207** | **63.04** |

DiffPIR. On the Gaussian deblurring task, we improve the LPIPS score by 0.044 points on FFHQ and 0.041 points on ImageNet compared to DDRM. DAVI also exhibits comparable or superior PSNR compared to the other methods. In short, these quantitative analyses highlight that DAVI is more effective than the baselines in achieving favorable solutions for inverse problems regarding consistency and realism.

**Robustness to unknown noise scale $\sigma_{\mathbf{y}}$.** In our main experiments, we initially set the measurement noise to $\sigma_{\mathbf{y}} = 0.05$. However, to account for the variability of noise levels encountered in real-world applications, we extend our evaluation by sampling $\sigma_{\mathbf{y}}$ from a uniform distribution $\mathcal{U}(0.01, 0.1)$ for Gaussian deblurring task. This approach aims to emulate more challenging scenarios. Tab. 3 demonstrates the robustness of DAVI across various scales of $\sigma_{\mathbf{y}}$ for noisy inverse problems without knowing the noise scale in the measurements. On the other hand, $\Pi$GDM, DDRM, and DDNM$^+$ show significant performance degradations compared to Tab. 2 since these approaches require the noise scale $\sigma_{\mathbf{y}}$ for their algorithms and leverage this information to compensate for error in estimating pseudo-inverse or SVD.

**Table 3: Robustness to the unknown noise scale $\sigma_{\mathbf{y}}$.** Comparative results for Gaussian deblurring on the ImageNet 256×256. The best results are highlighted in bold, while the second-highest are underlined. "Required" indicates that the noise scale $\sigma_{\mathbf{y}}$ is utilized for the method's operation.

| Method | **DAVI** (Ours) | RED-diff [23] | DPS [4] | ΠGDM [32] | DDRM [18] | DDNM$^+$ [39] |
|---|---|---|---|---|---|---|
| Noise scale $\sigma_{\mathbf{y}}$ | - | - | - | Required | Required | Required |
| PSNR ↑ | **23.47** | <u>23.25</u> | 19.81 | 21.80 | 16.09 | 19.25 |
| LPIPS ↓ | **0.347** | 0.470 | 0.447 | <u>0.446</u> | 0.595 | 0.543 |
| FID ↓ | **63.37** | 111.08 | <u>64.00</u> | 82.98 | 157.38 | 147.30 |

**Table 4: Inference speed (sec/image).** Wall-clock time on a single TITAN RTX GPU for 4× Super-resolution on FFHQ 256×256. The highest performance is highlighted in bold, followed by the second-best in underline.

| Method | **DAVI** (Ours) | DDRM [18] | DDNM$^+$ [39] | DPS [4] | ΠGDM [32] | DiffPIR [45] | RED-diff [23] |
|---|---|---|---|---|---|---|---|
| time(sec/img) ↓ | **0.04** | <u>1.7</u> | 4.2 | 77.7 | 8.3 | 4.3 | 41.1 |
| FID ↓ | **23.96** | 48.19 | 59.09 | <u>36.29</u> | 49.86 | 40.35 | 92.82 |

### 5.3 Qualitative Results

Fig. 5 shows qualitative comparisons of our model (DAVI) against baseline methods, selected based on their performance in Tab. 1 and 2. The results reveal that DAVI outperforms the baseline methods in terms of both measurement fidelity and realism. For instance, the microphone in the 2nd row of Fig. 5 gets blurry in DDRM, DPS, and RED-diff. In contrast, DPS and DiffPIR generate realistic and vivid images similar to ours; however, they often remove or change the details in the measurement. For example, in the sock and crow image (1st and 3rd row, respectively), DPS and DiffPIR struggle to capture small details in measurements, like eyes or green spots. For the box inpainting task, DAVI provides the most realistic results in the masked areas. This analysis aligns with the quantitative results, where DPS and DiffPIR, despite ranking high in FID scores for realism, score low in PSNR for consistency. On the other hand, DAVI generates realistic details in a *single step* and maintains consistency with the measurements. Please refer to the supplement Section E for further qualitative results.

### 5.4 Analysis

**Inference speed.** Our efficient single-step posterior sampling achieves an inference time of 0.04 seconds per image, outperforming the fastest baseline DDRM by 49.7% as demonstrated in Tab. 4. This clearly showcases the efficiency of our framework compared to the baselines. Moreover, DAVI demonstrates exceptional FID scores, indicating its high-quality performance in solving inverse problems. In contrast, other methodologies require iterative optimization for each measurement, leading to increased inference times as the number of inferences grows.

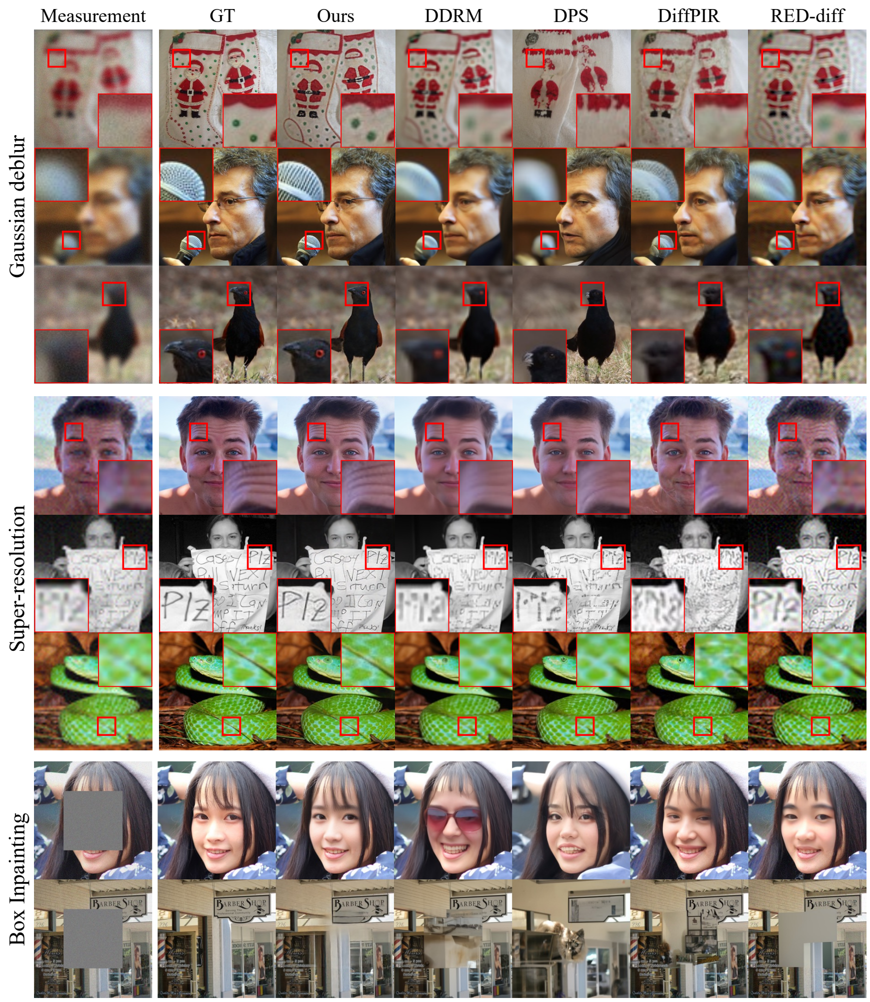


**Fig. 5: Qualitative comparison.** DAVI shows the most vivid and realistic solutions while maintaining intricate details of measurement, in contrast to baselines, which struggle to satisfy both aspects, as highlighted in red boxes.

**Ablation study.** Here, we present an ablation study to investigate the contributions of the proposed components in DAVI. Specifically, we examine the improvement by 1) integral KL divergence (IKL) and 2) Perturbed Posterior Bridge (PBB). The configurations and results are outlined in Tab. 5. We first assess the role of IKL in enhancing the realism of the restored images by minimizing the discrepancy between $q_\phi(\mathbf{x}_0|\mathbf{y})$ and $p(\mathbf{x})$. Comparing configurations A (without IKL, setting $T = 0$ in Alg. 1) and B (with IKL), we observe a significant FID score improvement in B by 13.08 points, underscoring IKL's effectiveness. From configuration C to F, we evaluate the importance of the Perturbed Posterior

**Table 5: Ablation study.** We conduct an ablation study on the Gaussian deblurring task on the FFHQ dataset. *Const.* indicates the constant perturbation scale, and *M.I* indicates a monotonically increasing schedule for $a$. Details are in the supplement.

| Config. | IKL | PPB | $\bar{\sigma}_a$ | $h$ | $p(a)$ | PSNR ↑ | LPIPS ↓ | FID ↓ |
|---|---|---|---|---|---|---|---|---|
| (A) | ✗ | ✗ | - | - | - | 21.47 | 0.443 | 50.69 |
| (B) | ✓ | ✗ | - | - | - | 24.70 | 0.236 | 37.61 |
| (C) | ✓ | ✓ | Const. | 0 | Uniform | **26.71** | 0.247 | 49.37 |
| (D) | ✓ | ✓ | Const. | 0.01 | Uniform | <u>26.17</u> | 0.239 | 47.41 |
| (E) | ✓ | ✓ | Const. | 0.01 | Beta | 25.99 | <u>0.232</u> | <u>34.34</u> |
| (F) | ✓ | ✓ | M.I | 0.01 | Beta | 25.46 | **0.225** | **29.92** |

Bridge, including the perturbation schedule $\bar{\sigma}_a$ and sampling distribution $p(a)$. Configuration C transforms the implicit distribution to a non-variational one by eliminating perturbation ($h = 0$). This transforms the PBB into a line segment between $\mathbf{x}_0$ and $\mathbf{y}$. This results in a notable degradation in image quality metrics (FID) compared to B. Conversely, introducing perturbations ($h = 0.01$) in D improves the LPIPS and FID, demonstrating the critical role of the PBB in enhancing high-quality image restoration. Notably, configuration F shows the most superior performance, with the FID improvement of 17.49 points compared to D, underscoring the significance of $\bar{\sigma}_a$ and $p(a)$ for optimizing the implicit distribution. This suggests that making the PPB generate intermediary measurements closer to $\mathbf{y}$ is more effective.

In supplement Sections B.2 and B.3, we evaluate zero-shot capabilities on out-of-distribution data and an unknown degradation model, and train with generated data. In supplement Section B.4, we compare our method with image-to-image translation methods like I$^2$SB [21] and CDDB [5]. For more details, please refer to the supplementary materials.

## 6 Conclusion

In our paper, we propose Diffusion prior-based Amortized Variational Inference (DAVI), a novel framework for solving noisy inverse problems based on amortized variational inference. DAVI optimizes an implicit distribution parameterized by a neural network by minimizing the KL divergence between the implicit distribution and the true posterior, enabling efficient single-step posterior sampling from implicit distribution by eliminating measurement-wise optimization. We introduce a novel Perturbed Posterior Bridge (PPB) to enhance the generalizability of our implicit distribution. Extensive experiments on image restoration tasks show that DAVI outperforms previous state-of-the-art diffusion-based methods. Lastly, our work uses the human dataset, which may raise ethical concerns that the restored images might resemble real individuals. We are fully aware of such ethical concerns and emphasize the critical need for ethical mindfulness and the development of robust strategies to address potential concerns.

## Acknowledgement

This work was supported by ICT Creative Consilience Program through the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT)(IITP-2024-2020-0-01819, 10%), the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2023R1A2C2005373, 30%), the National Supercomputing Center with supercomputing resources including technical support (KSC-2023-CRE-0325, 30%), and Artificial intelligence industrial convergence cluster development project funded by the Ministry of Science and ICT (MSIT, Korea)&Gwangju Metropolitan City (30%).

## References

1. Anderson, B.D.: Reverse-time diffusion equation models. In: Stochastic Processes and their Applications (1982)
2. Chen, H., Zhang, Y., Cun, X., Xia, M., Wang, X., Weng, C., Shan, Y.: Videocrafter2: Overcoming data limitations for high-quality video diffusion models. In: Conference on Computer Vision and Pattern Recognition) (2024)
3. Choi, J., Kim, S., Jeong, Y., Gwon, Y., Yoon, S.: ILVR: conditioning method for denoising diffusion probabilistic models. In: International Conference on Computer Vision (2021)
4. Chung, H., Kim, J., Mccann, M.T., Klasky, M.L., Ye, J.C.: Diffusion posterior sampling for general noisy inverse problems. In: International Conference on Learning Representations (2023)
5. Chung, H., Kim, J., Ye, J.C.: Direct diffusion bridge using data consistency for inverse problems. In: Advances in Neural Information Processing Systems (2024)
6. Chung, H., Sim, B., Ryu, D., Ye, J.C.: Improving diffusion models for inverse problems using manifold constraints. In: Advances in Neural Information Processing Systems (2022)
7. Chung, H., Sim, B., Ye, J.C.: Come-closer-diffuse-faster: Accelerating conditional diffusion models for inverse problems through stochastic contraction. In: Conference on Computer Vision and Pattern Recognition (2022)
8. Dou, Z., Song, Y.: Diffusion posterior sampling for linear inverse problem solving: A filtering perspective. In: International Conference on Learning Representations (2024)
9. Ganguly, A., Jain, S., Watchareeruetai, U.: Amortized variational inference: A systematic review. In: Journal of Artificial Intelligence Research (2023)
10. Gershman, S., Goodman, N.: Amortized inference in probabilistic reasoning. In: Proceedings of the annual meeting of the cognitive science society (2014)
11. Guilloteau, C., Oberlin, T., Berné, O., Dobigeon, N.: Hyperspectral and multispectral image fusion under spectrally varying spatial blurs–application to high dimensional infrared astronomical imaging. In: IEEE Transactions on Computational Imaging (2020)
12. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: Gans trained by a two time-scale update rule converge to a local nash equilibrium. In: Advances in neural information processing systems (2017)
13. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. In: Advances in Neural Information Processing Systems (2020)

14. Jordan, M.I., Ghahramani, Z., Jaakkola, T.S., Saul, L.K.: An introduction to variational methods for graphical models. In: Machine learning (1999)
15. Karras, T., Aittala, M., Aila, T., Laine, S.: Elucidating the design space of diffusion-based generative models. In: Advances in Neural Information Processing Systems (2022)
16. Karras, T., Aittala, M., Laine, S., Härkönen, E., Hellsten, J., Lehtinen, J., Aila, T.: Alias-free generative adversarial networks. In: Advances in Neural Information Processing Systems (2021)
17. Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. In: Conference on Computer Vision and Pattern Recognition (2019)
18. Kawar, B., Elad, M., Ermon, S., Song, J.: Denoising diffusion restoration models. In: Advances in Neural Information Processing Systems (2022)
19. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. In: International Conference on Learning Representations (2014)
20. Ko, J., Kong, I., Park, D., Kim, H.J.: Stochastic conditional diffusion models for robust semantic image synthesis. In: International Conference on Machine Learning (2024)
21. Liu, G.H., Vahdat, A., Huang, D.A., Theodorou, E.A., Nie, W., Anandkumar, A.: $I^2$sb: Image-to-image schrodinger bridge. In: International Conference on Machine Learning (2023)
22. Luo, W., Hu, T., Zhang, S., Sun, J., Li, Z., Zhang, Z.: Diff-instruct: A universal approach for transferring knowledge from pre-trained diffusion models. In: Advances in Neural Information Processing Systems (2023)
23. Mardani, M., Song, J., Kautz, J., Vahdat, A.: A variational perspective on solving inverse problems with diffusion models. In: International Conference on Learning Representations (2023)
24. Mokady, R., Hertz, A., Aberman, K., Pritch, Y., Cohen-Or, D.: Null-text inversion for editing real images using guided diffusion models. In: Conference on Computer Vision and Pattern Recognition) (2023)
25. Özbey, M., Dalmaz, O., Dar, S.U., Bedel, H.A., Özturk, Ş., Güngör, A., Çukur, T.: Unsupervised medical image translation with adversarial diffusion models. In: IEEE Transactions on Medical Imaging (2023)
26. Park, D., Kim, S., Lee, S., Kim, H.J.: Ddmi: Domain-agnostic latent diffusion models for synthesizing high-quality implicit neural representations. In: International Conference on Learning Representations (2024)
27. Raj, A., Kaza, S., Poole, B., Niemeyer, M., Ruiz, N., Mildenhall, B., Zada, S., Aberman, K., Rubinstein, M., Barron, J., et al.: Dreambooth3d: Subject-driven text-to-3d generation. In: International Conference on Computer Vision (2023)
28. Richardson, W.H.: Bayesian-based iterative method of image restoration. In: JoSA (1972)
29. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L.: ImageNet Large Scale Visual Recognition Challenge. In: International Journal of Computer Vision (2015)
30. Shu, R., Bui, H.H., Zhao, S., Kochenderfer, M.J., Ermon, S.: Amortized inference regularization. In: Advances in Neural Information Processing Systems (2018)
31. Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., Ganguli, S.: Deep unsupervised learning using nonequilibrium thermodynamics. In: International Conference on Machine Learning (2015)

32. Song, J., Vahdat, A., Mardani, M., Kautz, J.: Pseudoinverse-guided diffusion models for inverse problems. In: International Conference on Learning Representations (2023)
33. Song, Y., Ermon, S.: Generative modeling by estimating gradients of the data distribution. In: Advances in Neural Information Processing Systems (2019)
34. Song, Y., Shen, L., Xing, L., Ermon, S.: Solving inverse problems in medical imaging with score-based generative models. In: International Conference on Learning Representations (2022)
35. Song, Y., Sohl-Dickstein, J., Kingma, D.P., Kumar, A., Ermon, S., Poole, B.: Score-based generative modeling through stochastic differential equations. In: International Conference on Learning Representations (2021)
36. Steyvers, M., Griffiths, T.: Probabilistic topic models. Handbook of latent semantic analysis (2007)
37. Takeishi, N., Kalousis, A.: Physics-integrated variational autoencoders for robust and interpretable generative modeling. In: Ranzato, M., Beygelzimer, A., Dauphin, Y.N., Liang, P., Vaughan, J.W. (eds.) Advances in Neural Information Processing Systems (2021)
38. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. In: Advances in neural information processing systems (2017)
39. Wang, Y., Yu, J., Zhang, J.: Zero-shot image restoration using denoising diffusion null-space model. In: International Conference on Learning Representations (2023)
40. Wang, Z., Lu, C., Wang, Y., Bao, F., Li, C., Su, H., Zhu, J.: Prolificdreamer: High-fidelity and diverse text-to-3d generation with variational score distillation. In: Advances in Neural Information Processing Systems (2023)
41. Wei, Y., Zhang, S., Qing, Z., Yuan, H., Liu, Z., Liu, Y., Zhang, Y., Zhou, J., Shan, H.: Dreamvideo: Composing your dream videos with customized subject and motion. In: Conference on Computer Vision and Pattern Recognition) (2024)
42. Yin, T., Gharbi, M., Zhang, R., Shechtman, E., Durand, F., Freeman, W.T., Park, T.: One-step diffusion with distribution matching distillation. In: Conference on Computer Vision and Pattern Recognition (2024)
43. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H.: Restormer: Efficient transformer for high-resolution image restoration. In: Conference on Computer Vision and Pattern Recognition (2022)
44. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: Conference on Computer Vision and Pattern Recognition (2018)
45. Zhu, Y., Zhang, K., Liang, J., Cao, J., Wen, B., Timofte, R., Gool, L.V.: Denoising diffusion models for plug-and-play image restoration. In: Conference on Computer Vision and Pattern Recognition Workshops (2023)