Deep Diffusion Image Prior for Efficient OOD Adaptation in 3D Inverse Problems

Hyungjin Chung⊙ and Jong Chul Ye⊙

KAIST
{hj.chung,jong.ye}@kaist.ac.kr

Abstract. Recent inverse problem solvers that leverage generative diffusion priors have garnered significant attention due to their exceptional quality. However, adaptation of the prior is necessary when there exists a discrepancy between the training and testing distributions. In this work, we propose deep diffusion image prior (DDIP), which generalizes the recent adaptation method of SCD [3] by introducing a formal connection to the deep image prior. Under this framework, we propose an efficient adaptation method dubbed D3IP, specified for 3D measurements, which accelerates DDIP by orders of magnitude while achieving superior performance. D3IP enables seamless integration of 3D inverse solvers and thus leads to coherent 3D reconstruction. Moreover, we show that meta-learning techniques can also be applied to yield even better performance. We show that our method is capable of solving diverse 3D reconstructive tasks from the generative prior trained only with phantom images that are vastly different from the training set, opening up new opportunities of applying diffusion inverse solvers even when training with gold standard data is impossible. Code: https://github.com/HJ-harry/DDIP3D

Keywords: Diffusion models · Inverse problems · OOD Adaptation

1 Introduction

Diffusion models (DM) have shown remarkable performance as general inverse problem solvers [6, 31, 33]. The advantage naturally comes from the powerful capabilities of DMs to accurately model the prior distribution effectively learned from the collected training data. Unfortunately, there are numerous cases where the collection of high-quality gold standard data is not possible. For instance, it is difficult and expensive to collect large-scale data in medical imaging [18, 61], black hole imaging [11] and cryo-EM imaging [23, 63], etc. Consequently, one either has to resort to phantoms [11, 19] for generative modeling, or leverage implicit priors [63]. In these challenging cases, one faces an out-of-distribution (OOD) problem arising from mismatched priors [47], as the training data will be sufficiently different from the underlying true data distribution.

While it has been shown that DM-based inverse problem solvers (DIS) are less prone to distribution shifts [10, 29], the performance is significantly compromised [3, 47], leading to large gaps in performance. In the reconstructed images,

the discrepancy in the distributions is characterized as artifacts and hallucinations. As there exist provable bounds in the performance [47] when considering standard DIS with fixed parameters, the goal is to *adapt* the parameters of the diffusion model so that it better covers the true distribution, even when all we have access to is a degraded measurement.

Recently, SCD [3] was proposed as the first work to demonstrate that one can adapt the parameters of the diffusion models during the reverse diffusion sampling process to steer the distribution towards the unknown. While it has been demonstrated that SCD is effective for boosting performance on several medical imaging inverse problems, there exist clear limitations of the work. First, the inverse problems considered such as computed tomography (CT) and magnetic resonance imaging (MRI) reconstruction problems are inherently volumetric. However, SCD requires different adapted parameters for every different slice of the 3D volume. As

Fig. 1: OOD inverse problem setting. Pre-trained diffusion model learns $p_\theta(\boldsymbol{x})$, but at test time we only have $\boldsymbol{y}_{\text{out}}$ obtained from unknown OOD distributions, and aim to sample from $p_{\text{out}}(\boldsymbol{x}|\boldsymbol{y}_{\text{out}})$.

the adaptation framework slows down the inference time of DIS by $\times 5 \sim 30$ factor [3], the required reconstruction time for a 256^3 volume requires more than 6 hours of inference time on a single commodity GPU. On top of computation and memory requirements growing with $\mathcal{O}(N)$ where N is the number of 2D slices[1], it is also counter-intuitive that one needs a different set of parameters for adjacent slices, as the components are very similar. Ideally, the parameters should get better adapted as we increase the number of measurements, learning the common characteristics from the group. In contrast, we observe that the performance of SCD does not improve even if we try to gradually adapt the parameters with multiple samples.

In this work, we propose a novel framework that clarifies and solves all the aforementioned problems. Our contributions can be summarized as follows:

1. We show a formal link between deep image prior (DIP) [55] and the adaptation framework introduced in SCD [3]: when we correct for some factors in SCD, it is a multi-scale DIP constructed on the probability-flow ODE path [53]. Under this view, SCD can be generalized as a deep diffusion image prior (**DDIP**).
2. An efficient adaptation method specified for 3D inverse problem solving [6, 38], called **D3IP**, is proposed. It requires $\mathcal{O}(1)$ memory and compute, achieving significant acceleration while reaching performance on par with DDIP.

---

[1] This only considers the complexity along the slice dimension. The exact complexity will be presented in detail in Sec. 4.
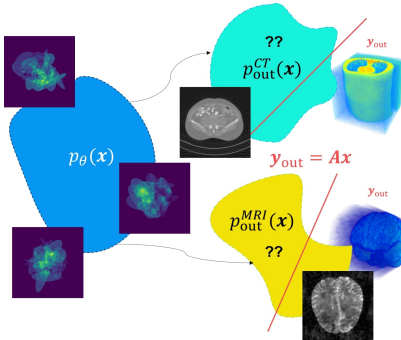
3. We show that D3IP can also benefit from using a more advanced solver suited for 3D reconstruction by seamlessly combining it with 3D DIS (e.g. DiffusionMBIR [6]). This lets us sample reconstructions that are spatially coherent across all the dimensions.
4. We present a meta-learning [20, 45] algorithm for initializing a 3D adaptation model, and later fine-tuning to specific 2D slices, achieving higher performance by benefiting both from the volume statistics while also being able to fit a specific characteristic.
5. Experimentally, we focus on the problem settings where we train a diffusion model on phantoms that can be generated on-the-fly, and hence can be considered fully unsupervised. This highlights the applicability to challenging imaging applications.

2 Background and Related Works

2.1 Diffusion models for inverse problems

In inverse problems, we aim to retrieve the signal $\boldsymbol{x}$ from the measurement $\boldsymbol{y}$, where

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n}, \quad \boldsymbol{y} \in \mathbb{R}^m, \, \boldsymbol{x} \in \mathbb{R}^n, \, \boldsymbol{A} \in \mathbb{R}^{m \times n}, \, \boldsymbol{n} \sim \mathcal{N}(0, \sigma_y^2 \boldsymbol{I}). \qquad (1)$$

As the problem is ill-posed, a natural way to solve the problem is through Bayesian inference, aiming to sample from the posterior distribution $p(\boldsymbol{x}|\boldsymbol{y}) \propto p_{\text{data}}(\boldsymbol{x})p(\boldsymbol{y}|\boldsymbol{x})$ by defining a suitable prior $p_{\text{data}}(\boldsymbol{x})$. A powerful way to do so is through generative modeling [22, 25, 36], a predominant modern choice being diffusion models (DM) [25, 49, 53].

In DMs, a time variable $t \in [0, T]$ is introduced, where $p_0(\boldsymbol{x}_0) = p_{\text{data}}(\boldsymbol{x})$, and $p_T(\boldsymbol{x}_T)$ approaches an isotropic normal distribution $\mathcal{N}$ as $t \to T$ through a Gaussian noising process $p(\boldsymbol{x}_t|\boldsymbol{x}_0) = \mathcal{N}(\boldsymbol{x}_0, t^2\boldsymbol{I})$. Interestingly, the reverse of such a process can be characterized as a stochastic differential equation (SDE) or a probability-flow ODE (PF-ODE)

$$d\boldsymbol{x}_t = -t\nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{x}_t)\, dt = \frac{\boldsymbol{x}_t - \mathbb{E}[\boldsymbol{x}_0|\boldsymbol{x}_t]}{t}\, dt, \quad p(\boldsymbol{x}_T) \sim \mathcal{N}, \qquad (2)$$

where the second equality holds from Tweedie's formula $\mathbb{E}[\boldsymbol{x}_0|\boldsymbol{x}_t] = \boldsymbol{x}_t + t^2\nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{x}_t)$ [17]. Due to this equivalence, one can train a denoiser D_θ through denoising score matching [52, 56], which estimates the denoised $\boldsymbol{x}_0$ from $\boldsymbol{x}_t$

$$\theta^* = \arg\min_\theta \mathbb{E}_{t\sim\text{Unif}(\varepsilon,T),\boldsymbol{x}_t\sim p(\boldsymbol{x}_t|\boldsymbol{x}_0),\boldsymbol{x}_0\sim p(\boldsymbol{x}_0)} \left[\|D_\theta(\boldsymbol{x}_t) - \boldsymbol{x}_0\|_2^2\right], \qquad (3)$$

leading to $D_{\theta^*}(\boldsymbol{x}_t) \approx \mathbb{E}[\boldsymbol{x}_0|\boldsymbol{x}_t]$. Solving Eq. (2) numerically by plugging D_{θ^*} in Eq. (2) leads to sampling from the prior distribution $p(\boldsymbol{x}_0)$.

When aiming for *posterior* sampling, one can additionally condition the PF-ODE with $\boldsymbol{y}$, which reads

$$d\boldsymbol{x}_t = -t\nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{x}_t|\boldsymbol{y})\, dt = \frac{\boldsymbol{x}_t - \mathbb{E}[\boldsymbol{x}_0|\boldsymbol{x}_t, \boldsymbol{y}]}{t}\, dt, \quad p(\boldsymbol{x}_T) \sim \mathcal{N}. \qquad (4)$$

As $\mathbb{E}[\boldsymbol{x}_0|\boldsymbol{x}_t, \boldsymbol{y}]$ (equivalently $\nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{x}_t|\boldsymbol{y})$) is intractable, various methods have been proposed to approximate the posterior sampling process [6,7,57,64]. DPS [6] approximates

$$\mathbb{E}[\boldsymbol{x}_0|\boldsymbol{x}_t, \boldsymbol{y}] = \mathbb{E}[\boldsymbol{x}_0|\boldsymbol{x}_t] + t^2 \nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{y}|\boldsymbol{x}_t) \approx \hat{\boldsymbol{x}}_{0|t} + \frac{t^2}{2\sigma_y^2} \nabla_{\boldsymbol{x}_t} \|\boldsymbol{y} - \boldsymbol{A}\hat{\boldsymbol{x}}_{0|t}\|_2^2, \quad (5)$$

with $\hat{\boldsymbol{x}}_{0|t} := D_{\theta^*}(\boldsymbol{x}_t)^2$. As taking backprop w.r.t. $\boldsymbol{x}_t$ through the network is expensive and unstable [46], a way to avoid this was proposed in DDS [7], which uses

$$\mathbb{E}[\boldsymbol{x}_0|\boldsymbol{x}_t, \boldsymbol{y}] \approx \arg\min_{\boldsymbol{x}_0} \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}_0\|_2^2 + \frac{\gamma}{2}\|\boldsymbol{x}_0 - \hat{\boldsymbol{x}}_{0|t}\|_2^2, \qquad (6)$$

solved with M-step conjugate gradient (CG), which we simply denote as $\texttt{CG}(\hat{\boldsymbol{x}}_{0|t}, M)$. Existing DIS can be considered as different ways of approximating the *empirical* conditional posterior, which we denote as $\mathbb{E}[\boldsymbol{x}_0|\boldsymbol{x}_t, \boldsymbol{y}] \approx D_\theta(\boldsymbol{x}_t|\boldsymbol{y})$ (See Appendix A for details and a review of existing methods). Note that while we present an introduction to DMs in a variance exploding (VE) framework as in [32], extension to the variance preserving (VP) framework is straightforward [25]. In the following sections and all the experiments, we leverage the VP framework with standard notations adopted from [25]. In most DIS, the predicted conditional posterior mean is used along with DDIM sampling [50]

$$\boldsymbol{x}_{t-1} = \text{DDIM}_\theta(\boldsymbol{x}_t, \eta) := \sqrt{\bar{\alpha}_{t-1}}D_\theta(\boldsymbol{x}_t|\boldsymbol{y}) + \sqrt{1 - \bar{\alpha}_{t-1}}\left(\eta\boldsymbol{\epsilon} + (1 - \eta)\boldsymbol{\epsilon}^\theta\right), \quad (7)$$

where $\eta \in [0, 1]$ controls the stochasticity and $\boldsymbol{\epsilon}^\theta$ is the predicted noise.

2.2   Steerable Conditional Diffusion (SCD)

While solving Eq. (4) with DIS approximations lets us perform posterior sampling, this is only true when the underlying prior distribution, and the distribution that D_{θ^*} models, are close enough (i.e. in-distribution). When $\boldsymbol{y}_{\text{out}}$ is a measurement obtained from Eq. (1) with $\boldsymbol{x}_{\text{out}} \sim p_{\text{out}} \neq p_\theta$, it was shown that one can close the gap by running the following iteration

$$\text{for } t = T, \ldots, 1 : \theta_{t-1} \leftarrow \arg\min_{\theta_t} \|\boldsymbol{y}_{\text{out}} - \boldsymbol{A}\texttt{CG}(\hat{\boldsymbol{x}}_{0|t}(\boldsymbol{x}_t; \theta_t), 1)\|_2^2, \qquad (8)$$

$$\boldsymbol{x}_{t-1} \leftarrow \text{DDIM}_{\theta_{t-1}}(\texttt{CG}(\hat{\boldsymbol{x}}_{0|t}(\boldsymbol{x}_t; \theta_{t-1}), M), \eta) \qquad (9)$$

where additional LoRA [27] parameters are introduced for adaptation to keep the original network parameters intact, and one-step CG was used to move the Tweedie denoised estimate closer towards $\boldsymbol{y}_{\text{out}}$. Adaptation of the network parameters in Eq. (8) is performed in between every DDS sampling step in Eq. (9)

---

[2] This notation is taken for simplicity. When emphasizing the dependence, we denote $\hat{\boldsymbol{x}}_{0|t}(\boldsymbol{x}_t; \theta^*)$

for a fixed number of optimization steps L, which enables adaptation of the parameters on-the-fly during the sampling process.

Nevertheless, one major limitation of SCD was the lack of understanding on why the algorithm works well in practice: Which design components are the key? What can we improve? Moreover, SCD requires running Eq. (8) for *every* different measurement $\boldsymbol{y}_{\text{out}}^i$, where $i = 1, \dots, N$ denotes the slice index across a single volume with a total of N slices. Re-initializing the parameters $\theta^i, \dots, \theta^N$, the process is N times slower, memory expensive, and does not synergistically leverage information from adjacent slices.

2.3   Related Works

When the training data is unavailable, one of the most standard approaches in inverse imaging is the use of deep image prior (DIP) [55]

$$\theta^* = \arg\min_{\theta} \|\boldsymbol{y} - \boldsymbol{A}G_{\theta}(\boldsymbol{z})\|_2^2, \quad \boldsymbol{z} \sim \mathcal{N}(0, \boldsymbol{I}), \tag{10}$$

where G_{θ} produces the reconstruction, and some means of early stopping is used to prevent the network from too much overfitting. As neural networks favor output signals that lie in the natural data manifold [55], and it can be considered a natural signal representation analogous to Fourier or Wavelets [26], optimizing for Eq. (10) leads to a decent reconstruction without any direct ground truth supervision. Over the years, there have been numerous advances in each of these components: design of the loss function, early stopping criterion, network parametrization, and initialization [2, 4, 24, 30, 40]. Despite the advances, DIP is still hard to optimize, and computationally demanding. For instance, applying DIP to a relatively small 3D data (167^3 resolution) requires about a day of training on a single RTX 3090 GPU [4].

In this work, we extend the idea of DIP into the realm of diffusion inverse solvers, showing that we can yield a superior algorithm in speed, robustness, and quality by leveraging the merits of diffusion models. Two works that are perhaps the most related are Educated DIP [4] (EDIP) and Baguer *et al.* [2]. EDIP [4] shows that initializing G_{θ} with a network trained for reconstruction helps in convergence. [2] proposes to use some initial reconstruction as an input to G_{θ}, rather than some random vector $\boldsymbol{z}$. As will be later seen, our method naturally leverages both of these techniques by pivoting along the PF-ODE path.

On the other hand, there have been efforts to train a diffusion prior from corrupted measurements alone [13, 34]. However, this is only possible when the measurements satisfy strict conditions, and one needs abundant training data obtained from these identical conditions. Our work is orthogonal from such approaches as we are free from conditions, and our goal is not to train a new generative prior but to adapt an OOD prior for reconstruction.

Solver	DDS	DDNM	DPS	DDIP							
$\mathbb{E}[\boldsymbol{x}_0\|\boldsymbol{x}_t,\boldsymbol{y}]$	DDS	DDNM	DPS	DDS			DDNM				DPS
K	-	-	-	1 (SCD [3])	3	**5 (ours)**	1	5	10	**20 (ours)**	**1 (ours)**
PSNR	31.37	30.69	23.49	35.43	35.83	**36.31**	34.92	35.50	35.98	**36.07**	**31.32**
SSIM	0.887	0.870	0.608	0.911	0.921	**0.922**	0.902	0.911	0.920	**0.920**	**0.841**

Table 1: Sparse-view CT reconstruction results on ELLIPSES → AAPM OOD reconstruction by varying the solver, $\mathbb{E}[\boldsymbol{x}_0|\boldsymbol{x}_t,\boldsymbol{y}]$ approximation method, and the number of CG iterations K. Standard non-adapted DIS , SCD [3] , Proposed DDIP method .

3 Proposed Method

3.1 Deep Diffusion Image Prior as a generalization of SCD

Recall that DIP optimizes the network parameters with the fidelity loss in Eq. (10). As the optimization is held specific to $\boldsymbol{y}$, Eq. (10) aims to implicitly recover the posterior mean $\mathbb{E}[\boldsymbol{x}_0|\boldsymbol{z},\boldsymbol{y}]$. On the other hand, as pointed out in Sec. 2.1, DIS at time t during sampling, produces some estimate of the conditional posterior mean $\mathbb{E}[\boldsymbol{x}_0|\boldsymbol{x}_t,\boldsymbol{y}] \approx D_\theta(\boldsymbol{x}_t|\boldsymbol{y})$. As $\boldsymbol{x}_t$ is equivalent to $\boldsymbol{z}$ at $t = T$ (initial noise), we can generalize DIP in Eq. (10) to multi-scale DIP over multiple noise scales

$$\text{for } t = T,\dots,1 : \theta_{t-1} \leftarrow \underset{\theta_t}{\arg\min} \|\boldsymbol{y} - \boldsymbol{A}D_{\theta_t}(\boldsymbol{x}_t|\boldsymbol{y})\|_2^2, \tag{11}$$

$$\boldsymbol{x}_{t-1} \leftarrow \text{DDIM}_{\theta_{t-1}}(D_{\theta_{t-1}}(\boldsymbol{x}_t|\boldsymbol{y}),\eta) \tag{12}$$

where Eq. (12) denotes some diffusion inverse problem solver that proceeds by leveraging the adapted parameters θ_{t-1}. It is easy to see that Eq. (11) is equivalent to Eq. (10) when $t = T$, but it is beneficial because 1) the optimization steps are performed in a coarse-to-fine manner starting from large noise scale, 2) the trajectory is pivoted along the original PF-ODE trajectory, providing a good initialization[3]. We define a general method presented in Eq. (11), Eq. (12) as deep diffusion image prior (DDIP).

Revisiting Eq. (8) from this perspective immediately reveals that there exists a discrepancy between the conditional posterior mean estimation used for adaptation (with 1 CG iteration), and the conditional posterior mean estimation used for sampling (with $K > 1$ CG iterations). In Tab. 1, we show that simply rectifying this error by using K CG iterations also for adaptation significantly improves the performance, highlighting the advantage of our interpretation. Moreover, it now becomes clear that the parametrization of Eq. (8) is merely a design choice: any choice of producing the posterior mean can be used. To emphasize that DDIP is agnostic to the choice of the approximation, we show in Tab. 1 that DDIP consistently improves performance on OOD reconstruction regardless of the choices made.

---

[3] As pointed out in Sec. 2.3, our approach introduces a better initialization in the network parameters as in [4], and also a better initialization in the input to the network [2] for $t < T$.

Algorithm 1 DDIP

Require: $\theta, D^t, N, T', \eta, \boldsymbol{y}$
1: $\mathcal{L}(\boldsymbol{x}, \boldsymbol{y}, D_\theta) := \|\boldsymbol{y} - \boldsymbol{A}D_\theta(\boldsymbol{x}|\boldsymbol{y})\|_2^2$
2: **for** $i = 1$ **to** N **do**
3: $\quad \theta_{T'}^{(i)} \leftarrow \theta$
4: $\quad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \boldsymbol{I})$
5: $\quad \boldsymbol{x}_T^{(i)} \leftarrow \sqrt{\bar{\alpha}_{T'}}\boldsymbol{A}^\dagger\boldsymbol{y} + \sqrt{1 - \bar{\alpha}_{T'}}\boldsymbol{\epsilon}$
6: $\quad$ **for** $t = T'$ **to** 1 **do**
7: $\qquad \theta_{t-1}^i \leftarrow \arg\min_{\theta_t^i}\mathcal{L}(\boldsymbol{x}_t^i, \boldsymbol{y}^i, D_{\theta_t^i})$
8: $\qquad \hat{\boldsymbol{x}}_{0|t}^i \leftarrow D_{\theta_{t-1}^i}^t(\boldsymbol{x}_t^i|\boldsymbol{y}^i)$
9: $\qquad \boldsymbol{x}_{t-1}^i \leftarrow \text{DDIM}(\hat{\boldsymbol{x}}_{0|t}^i, \eta)$
10: $\quad$ **end for**
11: $\quad$ **return** $\boldsymbol{x}_0^i$
12: **end for**

Algorithm 2 D3IP

Require: $\theta, D^t, N, T', K, \eta, \boldsymbol{Y}$
1: $\mathcal{L}(\boldsymbol{x}, \boldsymbol{y}, D_\theta) := \|\boldsymbol{y} - \boldsymbol{A}D_\theta(\boldsymbol{x}|\boldsymbol{y})\|_2^2$
2: $\theta_{T'} \leftarrow \theta$
3: $\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_N \sim \mathcal{N}(0, \boldsymbol{I})$
4: $\boldsymbol{\epsilon} \leftarrow \text{slerp}(\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_N, \frac{i}{N})$
5: $\boldsymbol{X}_{T'} \leftarrow \sqrt{\bar{\alpha}_{T'}}\boldsymbol{A}^\dagger\boldsymbol{Y} + \sqrt{1 - \bar{\alpha}_{T'}}\boldsymbol{\epsilon}$
6: **for** $t = T'$ **to** 1 **do**
7: $\quad \boldsymbol{x}_t^{\{i\}}, \boldsymbol{y}^{\{i\}} \sim \text{MC}((\boldsymbol{X}_t, \boldsymbol{Y}), K)$
8: $\quad \theta_{t-1} \leftarrow \arg\min_{\theta_t}\mathcal{L}(\boldsymbol{x}_t^{\{i\}}, \boldsymbol{y}^{\{i\}}, D_{\theta_t})$
9: $\quad \hat{\boldsymbol{X}}_{0|t} \leftarrow D_{\theta_{t-1}}^t(\boldsymbol{X}_t|\boldsymbol{Y})$
10: $\quad \boldsymbol{X}_{t-1} \leftarrow \text{DDIM}(\hat{\boldsymbol{X}}_{0|t}, \eta)$
11: **end for**
12: **return** $\boldsymbol{X}_0$

3.2 Extending DDIP to 3D

Let $\boldsymbol{X} \in \mathbb{R}^{n \times N} = [\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}]$ a stacked 3D tensor with N slices, and $\boldsymbol{Y}$ its corresponding measurement. In [3], it was proposed to independently run Eq. (8) for every different slice, which requires a prohibitively long amount of time for adaptation/sampling. Concretely, with a 256^3-sized volume that we consider in this work, it requires 90 seconds per slice, which is > 6 hours per volume.

Eq. (11) is a sequential optimization problem that minimizes the fidelity loss in a multi-grid fashion for every slice i. In SCD, the authors solve this optimization independently across all slices, as illustrated in Fig. 2 (a). Instead, we propose to optimize Eq. (11) so that the adaptation can be done synergistically in expectation (See Fig. 2 (b) for an illustration)

$$\text{for } t = T, \ldots, 1 : \min_\theta \mathbb{E}_i\left[\|\boldsymbol{y}^i - \boldsymbol{A}D_\theta^t(\boldsymbol{x}_t^i|\boldsymbol{y}^i)\|_2^2\right] \approx \frac{1}{K}\sum_{i=1}^K \|\boldsymbol{y}^i - \boldsymbol{A}D_\theta^t(\boldsymbol{x}_t^i|\boldsymbol{y}^i)\|_2^2, \tag{13}$$

where i denotes the index across the slices, and we can use K Monte Carlo samples to compute the expectation. In practice, we find even using a small K suffices for stable optimization and the performance plateaus for $K > 6$[4], yielding a compute-effective algorithm. The formulation of Eq. (13) lets us adapt the parameters that are suited for the whole 3D volume so that the memory and computation are reduced to $\mathcal{O}(1)$. We name our method D3IP (base), which will be shown to be improvable in specific settings, as we investigate in the following sections. We highlight the difference of D3IP against DDIP in Alg. 1,2, where $\text{MC}(\cdot, K)$ in L6 of Alg. 2 represents K-Monte Carlo sampling, and $\boldsymbol{x}^{\{i\}}$ denotes the sampled vectors stacked into a single tensor. Surprisingly, not only is D3IP

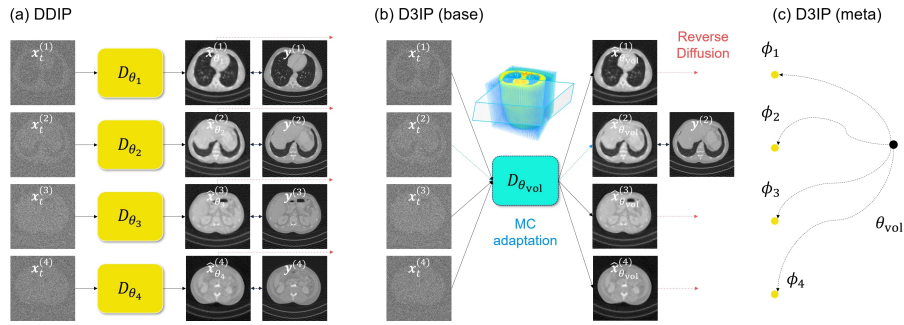---

[4] See App. D for further discussion.

Fig. 2: OOD adaptation schemes in DIS. (a) DDIP/SCD performs *independent* adaptation across slices and requires $\mathcal{O}(N)$ compute & memory. (b) D3IP (base) performs joint adaptation with stochastic gradients from MC sampling (blue dotted line) and requires $\mathcal{O}(1)$ compute & memory. (c) θ_{vol} adapted from D3IP (base) can be used as a meta-parameter to be further adapted to specific slices.

an order of magnitude cheaper and faster than DDIP, but it performs *better* than DDIP. This can be attributed to D3IP learning from a broader and richer context, whereas DDIP is only allowed to learn from a single slice of information. **Incorporating 3D DIS to D3IP.** Another big advantage of D3IP is that we can now seamlessly integrate 3D DIS methods [8,38] into our framework. As [38] would require adapting two diffusion models operating on different orientations, we choose DiffusionMBIR [8] accelerated with DDS [7] as our approximator (simply denoted DiffusionMBIR hereafter). Concretely, the approximation reads

$$D_\theta^t(\boldsymbol{X}_t|\boldsymbol{Y}) = \underset{\boldsymbol{X}_0}{\arg\min} \ \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{X}_0\|_2^2 + \frac{\gamma}{2}\|\boldsymbol{T}\boldsymbol{X}_0\|_1, \tag{14}$$

where $\boldsymbol{T}$ computes the finite difference along the stacked slice dimension, and the optimization is solved with a few-step ADMM [5] initialized with $\hat{\boldsymbol{X}}_{0|t}(\boldsymbol{X}_t; \theta_t)$ to induce proximal regularization.

One caveat is that Eq. (14) requires computing total variation (TV) regularization for adjacent slices, which cannot be done if we randomly sample K slices as done in D3IP (base). Instead, this can be effectively approximated by sampling $K < N$ neighboring slices, i.e. modifying L6 of Alg. 2 to

$$\boldsymbol{x}_t^{i:i+K-1}, \boldsymbol{y}^{i:i+K-1} \sim \text{MC}((\boldsymbol{X}_t, \boldsymbol{Y}), K). \tag{15}$$

We refer to this method with D3IP (mbir). The regularization in Eq. (14) is already useful for the OOD setting as it is independent of training data, and 3D samplers typically provide a better estimate in 3D inverse problem settings. Consequently, we observe improved performance in the adapted setting by simply incorporating a 3D DIS.

Notice that for D3IP, we keep a single set of updated neural network parameters that are shared across the volume to save memory and compute. When one is willing to max out the performance by trading off more compute/memory,

we show that one can do so by leveraging meta-learning. In fact, we reveal an interesting connection of D3IP to the widely used Reptile algorithm [45] and show that one can easily leverage meta-learning into our framework. Details can be found in Appendix B.

3.3   Technical Advances

On top of the fundamental innovations, we propose several technical advances to the adaptation algorithm that yield faster and more robust optimization. These advances are orthogonal to the contributions that are proposed in the rest of the section, and can be applied to all adaptation methods.

Constraining the optimization horizon. It is known that the empirical score functions exhibit problematic behavior in the end regimes [32, 35] as the estimation tends to be inaccurate and volatile. Motivated from the score distillation sampling literature [46, 58], we truncate the regime where we optimize for Eq. (11) or Eq. (13), so that for $t \notin [\zeta, T - \zeta]$, we only run standard DIS, with $\zeta = 40$ unless specified otherwise. We find that including adaptation outside this regime deteriorates the performance.

Initialization strategy. It is standard practice to initialize DIS with random Gaussian noise [6, 33, 51, 57]. However, it is also known that due to the non-zero terminal SNR of diffusion models [39], the low-frequency component of the initialization is carried out to the sample. Due to this property, some works propose to initialize the Gaussian noise with low-frequency part replaced from a similar sample [59]. In the case of inverse problems, this often comes for free. For instance, one can use the pseudo-inverse of the measurement. Moreover, motivated by the initialization strategy for DIP [60] and diffusion models [21] for video, which has correlations across time frames, we propose to sample two random noise vectors for the end slices and leverage the spherically interpolated (slerp) vector for the initialization noise. Summing up, our initialization reads

$$\boldsymbol{x}_{T'}^{(i)} \leftarrow \sqrt{\bar{\alpha}_{T'}}\boldsymbol{A}^\dagger\boldsymbol{y}^{(i)} + \sqrt{1 - \bar{\alpha}_{T'}}\text{slerp}(\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_N, \frac{i}{N}), \quad \boldsymbol{\epsilon}_0, \boldsymbol{\epsilon}_N \sim \mathcal{N}(0, \boldsymbol{I}), \quad (16)$$

with $T' < T$ typically set to 980. This not only lets us start from a correlated initial noise vector with low-frequency components consistent with the measurement but also lets us avoid any instabilities arising from the non-zero terminal SNR problem [39].

4   Experiments

We consider three canonical inverse problems in medical imaging: 3D sparse-view CT reconstruction (3D SV-CT), 3D MRI reconstruction (3D MRI), and multi-coil MRI reconstruction (CS-MRI), as these are the cases where the different measurement slices originate from the same volume. For the CT reconstruction task, we have a few hundred test slices originating from the same volume. For

Method	3D SV-CT (Ellipses → AAPM)			3D MRI (Ellipses → BRATS)			Mult. coil CS-MRI (BRAIN → KNEE)			Compute	Memory
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	$^*K \ll T < N$	
ADMM-TV	25.35	0.783	0.220	23.89	0.508	0.441	24.64	0.689	0.340	-	-
DDNM [57]	23.55	0.765	0.241	14.59	0.281	0.753	a19.35	0.354	0.493	$\mathcal{O}(T)$	$\mathcal{O}(1)$
DPS [6]	17.85	0.470	0.463	27.30	0.394	0.410	27.26	0.732	0.312	$\mathcal{O}(T)$	$\mathcal{O}(1)$
DDS [7]	27.65	0.805	0.188	24.59	0.532	0.339	28.36	0.741	0.278	$\mathcal{O}(T)$	$\mathcal{O}(1)$
DiffusionMBIR [8]	28.92	0.845	0.199	26.97	0.620	0.299	-	-	-	$\mathcal{O}(T)$	$\mathcal{O}(1)$
SCD [3]	32.91	0.904	0.184	26.00	0.561	0.338	29.01	0.752	0.269	$\mathcal{O}(KNT)$	$\mathcal{O}(N)$
DDIP	33.13	0.903	0.164	27.46	0.647	0.289	29.11	0.775	0.246	$\mathcal{O}(KNT)$	$\mathcal{O}(N)$
D3IP (base)	31.73	0.908	0.141	30.59	0.859	0.152	29.00	**0.789**	0.233	$\mathcal{O}(KT)$	$\mathcal{O}(1)$
D3IP (mbir)	33.69	**0.919**	**0.133**	**33.89**	**0.907**	**0.103**	-	-	-	$\mathcal{O}(KT)$	$\mathcal{O}(1)$
D3IP (meta)	**33.96**	0.917	0.136	32.60	0.877	0.126	**29.52**	0.779	**0.216**	$\mathcal{O}(KNT)$	$\mathcal{O}(N)$

Table 2: Quantitative measure of OOD Inverse problem solving on 3 main tasks. Non-adapted standard DIS , the family of proposed methods .

the first two tasks, we use a diffusion prior trained only on Ellipses phantoms [1] generated on-the-fly, which are completely irrelevant to the target data distribution. We focus on such setting this leads to a completely unsupervised approach for a reconstructive task, requiring no collection of gold standard data.

For the MRI reconstruction task, we consider two different realistic cases: when the acquisition scheme is truly 3D and the volume consists of a few hundred slices, and when the acquisition scheme is 2D but the slices originate from the same volume, yielding a few ten slices. For all tasks, we consider variance preserving (VP) diffusion models trained under the ADM [15] framework unless specified otherwise.

4.1 Experimental settings

3D SV-CT. Following [3], we take a diffusion model trained on the Ellipses dataset [1] and test it on a volume of the American Association of Physicists in Medicine (AAPM) grand challenge [44] dataset, following the settings of [8, 9]. This is a particularly useful and interesting setting, as the Ellipses dataset are phantoms that can be easily generated on-the-fly, requiring *no* collection of data, a realistic condition when acquiring high-quality ground truth images is impossible. We consider parallel CT geometry with 60-angle measurements to be consistent with [3].

3D MRI reconstruction We take the same Ellipses diffusion model used for 3D SV-CT, and adapt it to the multimodal brain tumor image segmentation benchmark (BRATS) 2018 [43]. The test set consists of 5 T1 contrast / 5 T2 contrast volumes. We consider variable density (VD) Poisson disc sampling pattern [16] (×8 acceleration) for 3D volume measurements in a single-coil setting.

2D Multi-coil MRI reconstruction. A diffusion model trained on fastMRI [61] BRAIN data was taken from [7][5]. The evaluation was done on fastMRI KNEE data, consisting of 10 test volumes and 294 slices. The measurement was simulated using the uniform 1D subsampling (×4 acceleration) with 8% Auto Calibrating

---

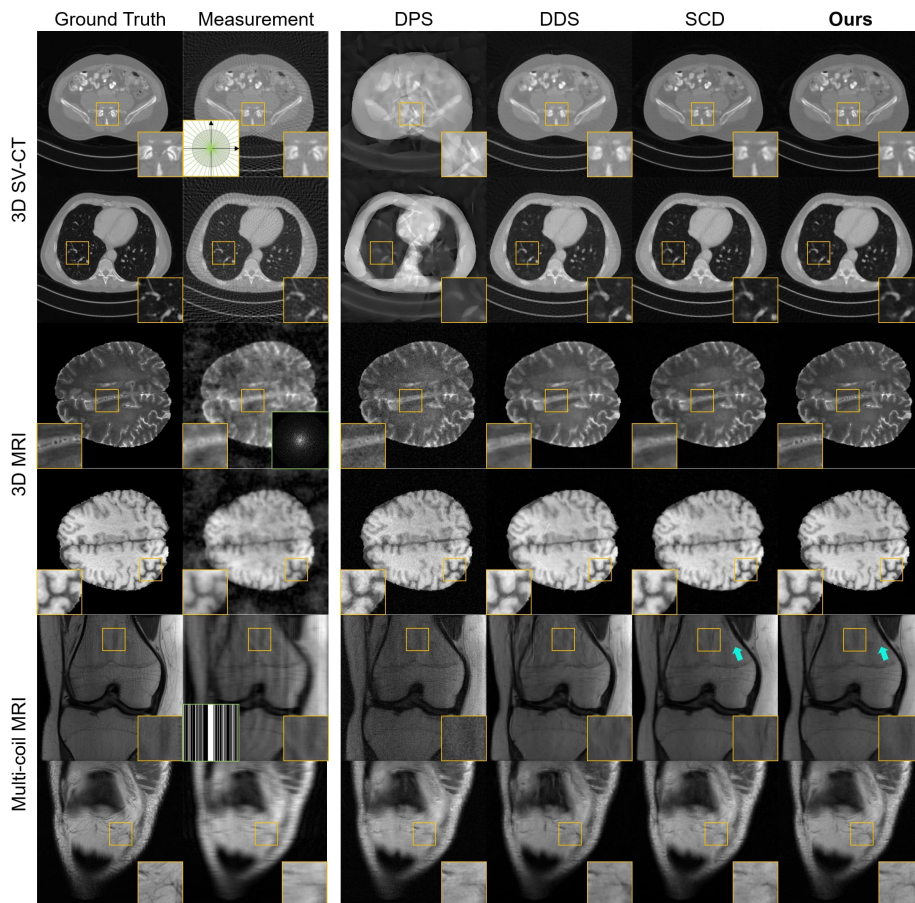[5] https://github.com/HJ-harry/DDS

Fig. 3: Representative results on 3 different tasks. (row 1-2): 3D SV-CT, (row 3-4): 3D MRI, (row 5-6): CS-MRI. Comparison against DPS [6], DDS [7], and SCD [3]. **Ours**: D3IP (base). Cyan arrows indicate regions of remaining artifacts even after adaptation with SCD. Green boxes illustrate the acquisition scheme of the measurement (acquisition angle, sub-sampling pattern).

Signal (ACS) region, as proposed in [61]. Since we consider multi-coil measurements in this task, the coil sensitivity maps are estimated using ESPiRiT [54].

For all inverse problems considered, we add $\sigma_y = 0.01$ measurement noise when simulating with forward operators. Further details can be found in App. C.

4.2 Diffusion samplers

For all DIS methods including the proposed method, we consider 50 NFE DDIM sampling with $\eta = 0.85$ unless specified otherwise. One exception is DPS [6], where we take 1000 NFE with $\eta = 1.0$ (i.e. DDPM sampling) to achieve satisfactory results.
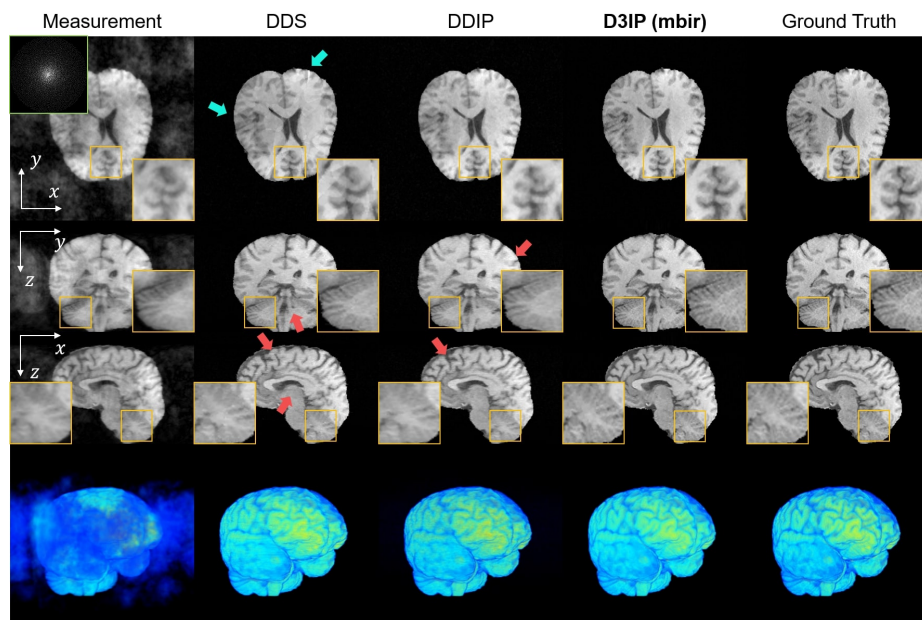
Fig. 4: 3D-MRI reconstruction with DDS [7], DDIP, D3IP (mbir). Cyan and red arrows indicate artifacts from prior mismatch and slice-wise independent reconstruction, respectively. 1-4$^{\text{th}}$ row: xy, yz, xz slice, and 3D rendering.

Adaptation. For the family of the proposed method (DDIP, D3IP), we take DDS [7] as our estimator D_θ for $\mathbb{E}[\boldsymbol{x}_0|\boldsymbol{x}_t, \boldsymbol{y}]$, which runs 5 CG optimization step per diffusion denoising step. In 3D SV-CT and 3D MRI tasks, K is set to 6, and for CS-MRI, $K = 3$. For D3IP (mbir), we use the approximation in Eq. (14), which is solved with 5 ADMM iterations per diffusion step. D3IP (meta) runs D3IP in Alg. 2 with linearly decreasing step size from $\alpha = 1.0$ to $\alpha = 0.5$ in Eq. (30). The obtained meta-parameter θ_{vol} is then fine-tuned with respect to each slice using DDIP in Alg. 1 initialized from this meta-parameter.

Comparison methods. We consider several strong DIS baselines: DDNM [57], DPS [6], DDS [7], DiffusionMBIR [8], and SCD [3] for comparison. In order to apply DDNM, one needs access to the pseudo-inverse operator, which is often hard to compute, or numerically unstable. We circumvent this by leveraging CG updates motivated from [7]. We note that DiffusionMBIR is the only 3D-aware DIS, and SCD is the only adaptation-based DIS among the comparison methods. See App. C for details.

Method	PSNR ↑	SSIM ↑
DDIP(baseline)	35.62	0.908
+ constrained horizon	35.98	0.916
+ init. strategy	**36.31**	**0.922**

Table 3: Improvements from configurations introduced in Sec. 3.3.

4.3 Results

Ablation study on the technical advances. Before diving into the main results, we first conduct an ablation study introduced in Sec. 3.3, as these techniques can be leveraged by all adaptation samplers. In Tab. 3, it is evident that the techniques provide improvements in the metrics, hence we keep the final configuration through all our experiments, including our proof-of-concept experiment presented in Tab. 1.

Main Results. In Tab. 2, we present a thorough comparison study on all three tasks that we consider in this work. Here, we see that D3IP not only dramatically reduces the computation cost of the adaptation, but also often results in a *better* reconstruction quality than DDIP. Specifically, this effect is most evidently seen in the 3D MRI task, where there is more than 3 dB difference in PSNR, and the perceptual LPIPS metric improves also by a large margin. Among all the tasks and the metrics considered, only the PSNR values of 3D SV-CT and Multi-coil CS-MRI fall short of DDIP. For the rest of the cases, D3IP outperforms DDIP, thanks to the richer information provided during the optimization process. In App. D, we further show the intriguing properties of D3IP: robustness to the choice of hyperparameters, and higher capacity.

In Fig. 3, we compare the qualitative results of our proposed D3IP (base) solver against SOTA DIS - DPS, DDS, and the previous adapted sampler SCD. It is clear that the results obtained with DPS and DDS are contaminated with artifacts and hallucinations that originate from the training data (elliptical phantoms). While SCD greatly alleviates these artifacts, the details are typically blurred out, erasing structures that can even be seen from the measurement (e.g. 1st row of Fig. 3). Notably, this happens even without any TV regularization that was used in [3]. Introducing such regularization yields even blurrier results. In contrast, the proposed method is capable of restoring crisp features without such blurring. Moreover, D3IP further eliminates leftover artifacts that are still apparent in SCD, as pointed out with cyan arrows.

It should be noted that all these advances take place while also being much faster and cheaper in memory. On a single 3090 RTX GPU, for a 256^3 volume, D3IP (base) requires 40 minutes, whereas SCD requires $\sim$6.2 hours, as it should be conducted slice-wise. Moreover, by using D3IP, we can reduce the memory requirements of the LoRA parameters from 2.9 Gb to 14.5 Mb as we only need to keep a single set of parameters for the whole volume, rather than specific parameters for each slice.
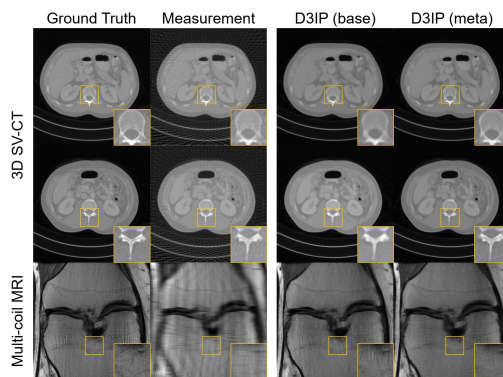

Fig. 5: Comparison of reconstructions with D3IP (base) and D3IP (meta).

Incorporating DiffusionMBIR. While D3IP (base) enables the use of average gradients computed from multiple slices, it does not guarantee smooth reconstruction across slices, which can be effectively counteracted by leveraging 3D solvers. In Fig. 4, we visualize the 3D rendering as well as the coronal and the sagittal slices of the reconstruction, highlighting the advantage of the proposed method by allowing a plug-and-play incorporation of 3D DIS.

Meta-learning further improves D3IP. By initializing a meta-parameter θ_{vol} and further adapting the parameters toward specific slices, we see in Fig. 5 and Tab. 2 that we can achieve even better performance by trading off compute. In contrast, gradually adapting the parameters across different slices with SCD does not lead to noticeable improvements, compared to the case where the weights are re-initialized for every slice.

5   Conclusion

In this work, we proposed DDIP, a method of adapting diffusion priors trained on OOD distribution for solving inverse imaging tasks. We clarified that DDIP is a generalization of DIP constructed on the PF-ODE trajectory of diffusion, which makes the method much stabler than standard DIP. Focusing on 3D inverse problems, we proposed D3IP, which significantly reduces the computation cost while achieving better performance. We showed that D3IP can be further enhanced by incorporating 3D inverse problem solvers, or by leveraging meta-learning to induce a meta-parameter, then fine-tuning to specific 2D measurements. Notably, our work improves previous approaches in reconstruction quality, closing the gap against in-distribution DIS; it accelerates the speed of the algorithm to a regime where it will be practical for real-world usage; it enhances the interpretability of the algorithm and clarifies why the method works. We believe that our work can become the ground for many inverse imaging applications where unsupervised reconstruction remains crucial, such as in biomedical imaging and astronomy.

Acknowledgements

References

1. Adler, J., Öktem, O.: Learned primal-dual reconstruction. IEEE transactions on medical imaging **37**(6), 1322–1332 (2018) 10
2. Baguer, D.O., Leuschner, J., Schmidt, M.: Computed tomography reconstruction using deep image prior and learned reconstruction methods. Inverse Problems **36**(9), 094004 (2020) 5, 6
3. Barbano, R., Denker, A., Chung, H., Roh, T.H., Arrdige, S., Maass, P., Jin, B., Ye, J.C.: Steerable conditional diffusion for out-of-distribution adaptation in imaging inverse problems. arXiv preprint arXiv:2308.14409 (2023) 1, 2, 6, 7, 10, 11, 12, 13, 21, 24
4. Barbano, R., Leuschner, J., Schmidt, M., Denker, A., Hauptmann, A., Maass, P., Jin, B.: An educated warm start for deep image prior-based micro ct reconstruction. IEEE Transactions on Computational Imaging **8**, 1210–1222 (2022) 5, 6
5. Boyd, S., Parikh, N., Chu, E.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Now Publishers Inc (2011) 8, 20
6. 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), https://openreview.net/forum?id=OnD9zGAGT0k 1, 2, 3, 4, 9, 10, 11, 12, 19, 20, 24
7. Chung, H., Lee, S., Ye, J.C.: Decomposed diffusion sampler for accelerating large-scale inverse problems. In: International Conference on Learning Representations (2024) 4, 8, 10, 11, 12, 19, 20, 24, 25
8. Chung, H., Ryu, D., Mccann, M.T., Klasky, M.L., Ye, J.C.: Solving 3d inverse problems using pre-trained 2d diffusion models. IEEE/CVF Conference on Computer Vision and Pattern Recognition (2023) 8, 10, 12, 20
9. Chung, H., Sim, B., Ryu, D., Ye, J.C.: Improving diffusion models for inverse problems using manifold constraints. In: Oh, A.H., Agarwal, A., Belgrave, D., Cho, K. (eds.) Advances in Neural Information Processing Systems (2022), https://openreview.net/forum?id=nJJjv0JDJju 10
10. Chung, H., Ye, J.C.: Score-based diffusion models for accelerated mri. Medical Image Analysis p. 102479 (2022) 1
11. Collaboration, E.H.T., et al.: First m87 event horizon telescope results. iv. imaging the central supermassive black hole. arXiv preprint arXiv:1906.11241 (2019) 1
12. Crawshaw, M.: Multi-task learning with deep neural networks: A survey. arXiv preprint arXiv:2009.09796 (2020) 21
13. Daras, G., Shah, K., Dagan, Y., Gollakota, A., Dimakis, A.G., Klivans, A.: Ambient diffusion: Learning clean distributions from corrupted data. arXiv preprint arXiv:2305.19256 (2023) 5
14. Désidéri, J.A.: Multiple-gradient descent algorithm (mgda) for multiobjective optimization. Comptes Rendus Mathematique **350**(5-6), 313–318 (2012) 21
15. Dhariwal, P., Nichol, A.Q.: Diffusion models beat GANs on image synthesis. In: Beygelzimer, A., Dauphin, Y., Liang, P., Vaughan, J.W. (eds.) Advances in Neural Information Processing Systems (2021) 10
16. Dwork, N., Baron, C.A., Johnson, E.M., O'Connor, D., Pauly, J.M., Larson, P.E.: Fast variable density poisson-disc sample generation with directional variation for compressed sensing in mri. Magnetic resonance imaging **77**, 186–193 (2021) 10
17. Efron, B.: Tweedie's formula and selection bias. Journal of the American Statistical Association **106**(496), 1602–1614 (2011) 3, 19

18. Fabian, Z., Heckel, R., Soltanolkotabi, M.: Data augmentation for deep learning based accelerated mri reconstruction with limited data. In: International Conference on Machine Learning. pp. 3057–3067. PMLR (2021) 1
19. Feng, B.T., Smith, J., Rubinstein, M., Chang, H., Bouman, K.L., Freeman, W.T.: Score-based diffusion models as principled priors for inverse imaging. arXiv preprint arXiv:2304.11751 (2023) 1
20. Finn, C., Abbeel, P., Levine, S.: Model-agnostic meta-learning for fast adaptation of deep networks. In: International conference on machine learning. pp. 1126–1135. PMLR (2017) 3, 21
21. Ge, S., Nah, S., Liu, G., Poon, T., Tao, A., Catanzaro, B., Jacobs, D., Huang, J.B., Liu, M.Y., Balaji, Y.: Preserve your own correlation: A noise prior for video diffusion models. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 22930–22941 (2023) 9
22. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. Advances in neural information processing systems **27** (2014) 3
23. Gupta, H., McCann, M.T., Donati, L., Unser, M.: Cryogan: A new reconstruction paradigm for single-particle cryo-em via deep adversarial learning. IEEE Transactions on Computational Imaging **7**, 759–774 (2021) 1
24. Heckel, R., Hand, P.: Deep decoder: Concise image representations from untrained non-convolutional networks. arXiv preprint arXiv:1810.03982 (2018) 5
25. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems **33**, 6840–6851 (2020) 3, 4
26. Hoyer, S., Sohl-Dickstein, J., Greydanus, S.: Neural reparameterization improves structural optimization. arXiv preprint arXiv:1909.04240 (2019) 5
27. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W.: Lora: Low-rank adaptation of large language models. arXiv preprint arXiv:2106.09685 (2021) 4, 22
28. Hyvärinen, A., Dayan, P.: Estimation of non-normalized statistical models by score matching. Journal of Machine Learning Research **6**(4) (2005) 19
29. Jalal, A., Arvinte, M., Daras, G., Price, E., Dimakis, A.G., Tamir, J.: Robust compressed sensing mri with deep generative priors. Advances in Neural Information Processing Systems **34** (2021) 1, 19
30. Jo, Y., Chun, S.Y., Choi, J.: Rethinking deep image prior for denoising. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5087–5096 (2021) 5
31. Kadkhodaie, Z., Simoncelli, E.: Stochastic solutions for linear inverse problems using the prior implicit in a denoiser. In: Advances in Neural Information Processing Systems. vol. 34, pp. 13242–13254. Curran Associates, Inc. (2021) 1
32. Karras, T., Aittala, M., Aila, T., Laine, S.: Elucidating the design space of diffusion-based generative models. In: Proc. NeurIPS (2022) 4, 9
33. Kawar, B., Elad, M., Ermon, S., Song, J.: Denoising diffusion restoration models. In: Oh, A.H., Agarwal, A., Belgrave, D., Cho, K. (eds.) Advances in Neural Information Processing Systems (2022), https://openreview.net/forum?id=kxXvopt9pWK 1, 9
34. Kawar, B., Elata, N., Michaeli, T., Elad, M.: Gsure-based diffusion model training with corrupted data. arXiv preprint arXiv:2305.13128 (2023) 5
35. Kim, D., Shin, S., Song, K., Kang, W., Moon, I.C.: Soft truncation: A universal training technique of score-based diffusion model for high precision score estimation. International conference on machine learning (2022) 9

36. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114 (2013) 3
37. Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A.A., Milan, K., Quan, J., Ramalho, T., Grabska-Barwinska, A., et al.: Overcoming catastrophic forgetting in neural networks. Proceedings of the national academy of sciences **114**(13), 3521–3526 (2017) 23
38. Lee, S., Chung, H., Park, M., Park, J., Ryu, W.S., Ye, J.C.: Improving 3D imaging with pre-trained perpendicular 2D diffusion models. arXiv preprint arXiv:2303.08440 (2023) 2, 8
39. Lin, S., Liu, B., Li, J., Yang, X.: Common diffusion noise schedules and sample steps are flawed. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 5404–5411 (2024) 9
40. Liu, J., Sun, Y., Xu, X., Kamilov, U.S.: Image restoration using total variation regularized deep image prior. In: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 7715–7719. Ieee (2019) 5
41. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017) 21
42. McCloskey, M., Cohen, N.J.: Catastrophic interference in connectionist networks: The sequential learning problem. In: Psychology of learning and motivation, vol. 24, pp. 109–165. Elsevier (1989) 23
43. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor image segmentation benchmark (brats). IEEE transactions on medical imaging **34**(10), 1993–2024 (2014) 10
44. Moen, T.R., Chen, B., Holmes III, D.R., Duan, X., Yu, Z., Yu, L., Leng, S., Fletcher, J.G., McCollough, C.H.: Low-dose ct image and projection dataset. Medical physics **48**(2), 902–911 (2021) 10
45. Nichol, A., Schulman, J.: Reptile: a scalable metalearning algorithm. arXiv preprint arXiv:1803.02999 **2**(3), 4 (2018) 3, 9, 21
46. Poole, B., Jain, A., Barron, J.T., Mildenhall, B.: Dreamfusion: Text-to-3d using 2d diffusion. In: The Eleventh International Conference on Learning Representations (2023), https://openreview.net/forum?id=FjNys5c7VyY 4, 9
47. Renaud, M., Liu, J., de Bortoli, V., Almansa, A., Kamilov, U.S.: Plug-and-play posterior sampling under mismatched measurement and prior models. arXiv preprint arXiv:2310.03546 (2023) 1, 2
48. Ruiz, N., Li, Y., Jampani, V., Pritch, Y., Rubinstein, M., Aberman, K.: Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 22500–22510 (2023) 23
49. Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., Ganguli, S.: Deep unsupervised learning using nonequilibrium thermodynamics. In: International Conference on Machine Learning. pp. 2256–2265. PMLR (2015) 3
50. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. In: 9th International Conference on Learning Representations, ICLR (2021) 4
51. Song, J., Vahdat, A., Mardani, M., Kautz, J.: Pseudoinverse-guided diffusion models for inverse problems. In: International Conference on Learning Representations (2023), https://openreview.net/forum?id=9_gsMA8MRKQ 9
52. Song, Y., Ermon, S.: Generative modeling by estimating gradients of the data distribution. In: Advances in Neural Information Processing Systems. vol. 32 (2019) 3

53. Song, Y., Sohl-Dickstein, J., Kingma, D.P., Kumar, A., Ermon, S., Poole, B.: Score-based generative modeling through stochastic differential equations. In: 9th International Conference on Learning Representations, ICLR (2021) 2, 3

54. Uecker, M., Lai, P., Murphy, M.J., Virtue, P., Elad, M., Pauly, J.M., Vasanawala, S.S., Lustig, M.: ESPIRiT—an eigenvalue approach to autocalibrating parallel MRI: where SENSE meets GRAPPA. Magnetic resonance in medicine **71**(3), 990–1001 (2014) 11

55. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Deep image prior. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 9446–9454 (2018) 2, 5, 22

56. Vincent, P.: A connection between score matching and denoising autoencoders. Neural computation **23**(7), 1661–1674 (2011) 3

57. Wang, Y., Yu, J., Zhang, J.: Zero-shot image restoration using denoising diffusion null-space model. In: The Eleventh International Conference on Learning Representations (2023), https://openreview.net/forum?id=mRieQgMtNTQ 4, 9, 10, 12, 19

58. 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. arXiv preprint arXiv:2305.16213 (2023) 9

59. Wu, T., Si, C., Jiang, Y., Huang, Z., Liu, Z.: Freeinit: Bridging initialization gap in video diffusion models. arXiv preprint arXiv:2312.07537 (2023) 9

60. Yoo, J., Jin, K.H., Gupta, H., Yerly, J., Stuber, M., Unser, M.: Time-Dependent Deep Image Prior for Dynamic MRI. IEEE Transactions on Medical Imaging (2021) 9

61. Zbontar, J., Knoll, F., Sriram, A., Murrell, T., Huang, Z., Muckley, M.J., Defazio, A., Stern, R., Johnson, P., Bruno, M., et al.: fastMRI: An open dataset and benchmarks for accelerated MRI. arXiv preprint arXiv:1811.08839 (2018) 1, 10, 11

62. Zhang, L., Rao, A., Agrawala, M.: Adding conditional control to text-to-image diffusion models. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3836–3847 (2023) 23

63. Zhong, E.D., Bepler, T., Berger, B., Davis, J.H.: Cryodrgn: reconstruction of heterogeneous cryo-em structures using neural networks. Nature methods **18**(2), 176–185 (2021) 1

64. Zhu, Y., Zhang, K., Liang, J., Cao, J., Wen, B., Timofte, R., Van Gool, L.: Denoising diffusion models for plug-and-play image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1219–1229 (2023) 4, 19