Training-free Medical Image Inverses via Bi-level Guided Diffusion Models

Hossein Askari, **Fred Roosta**, **Hongfu Sun**
University of Queensland, Brisbane, Australia
{h.askari, fred.roosta, hongfu.sun}@uq.edu.au

Supplementary Material

A.1. Related Works

A solution to the inverse problem $\mathbf{y} = \mathcal{A}\mathbf{x} + \mathbf{n}$, can be probabilistically derived via the maximum likelihood estimation (MLE), defined as $\mathbf{x}_{\text{ML}} = \arg\max_{\mathbf{x}} \log p(\mathbf{y}|\mathbf{x})$, where $p(\mathbf{y}|\mathbf{x}) := \mathcal{N}(\mathcal{A}\mathbf{x}, \sigma_{\mathbf{y}}^2)$ represents the likelihood of observation $\mathbf{y}$, ensuring data consistency. Nevertheless, if the forward operator $\mathcal{A}$ is singular, e.g., when $m < n$, the problem is ill-posed. In such cases, it is fundamentally infeasible to uniquely recover the signal set $\mathcal{X}$ using only the observed measurements $\mathcal{Y}$, even in the noiseless scenario where $\mathcal{Y} = \mathcal{A}\mathcal{X}$. This challenge arises due to the nontrivial nature of the null space of $\mathcal{A}$.

To mitigate the ill-posedness, it is therefore essential to incorporate an additional assumption based on *prior* knowledge to constrain the space of possible solutions. A predominantly adopted framework that offers a more meaningful solution is Maximum a Posteriori (MAP) estimation which is formulated as $\mathbf{x}_{\text{MAP}} = \arg\max_{\mathbf{x}}[\log p(\mathbf{y}|\mathbf{x}) + \log p(\mathbf{x})]$, where the term $\log p(\mathbf{x})$ encapsulates the prior information of the clean image $\mathbf{x}$.

The concept of priors in solving inverse problems has evolved considerably over time. Classically, many methodologies relied on hand-crafted priors, which are analytically defined constraints such as sparsity [10,31], low-rank [14, 16], total variation [9], to name but a few, to enhance reconstruction. With the advent of deep learning models, priors have transitioned to being data-driven, yielding significant gains in reconstruction quality [1,2,7,17,34]. These priors, whether learned in a supervised or unsupervised fashion, have been integrated within the MAP framework to address ill-posed inverse problems. In the supervised paradigm, the reliance on the availability of paired original images and observed measurements also can potentially limit the model's generalizability. As a result, the trend has shifted towards an increasing interest in unsupervised approaches, where priors are learned implicitly or explicitly using deep generative models.

The strategies within the unsupervised learning paradigm vary based on how the learned priors (a.k.a. generative priors) are imposed. For instance, generators $\mathcal{G}_{\theta}$ in pre-trained generative models such as Generative Adversarial Networks (GANs) [7, 18], Variational Autoencoders (VAEs) [1], and Normalizing Flows (NFs) [2], are employed as priors to identify the latent code that explains the measurements, as described by the optimization problem $\hat{\mathbf{z}} = \arg\max_{\mathbf{z}} \log p(\mathbf{y}|\mathcal{G}_{\theta}(\mathbf{z})) + \log p(\mathbf{z})$. In such a way, the solution $\hat{\mathbf{z}}$ is constrained to be within the domain of the generative model. This approach, however, suffers from critical restrictions. In the first place, the low dimensionality of the latent space is a major concern, as it hampers the reconstruction of images that lie outside their manifold. Additionally, it demands computationally expensive iterative updates, given the complexity of generator $\mathcal{G}_{\theta}$. Crucially, the deterministic nature of the recovered solutions hinders the assessment of the reliability of reconstruction. In fact, MAP inference fails to fully capture the entire range of the solution spectrum, particularly when solving an ill-posed problem that might hold multiple solutions aligned closely with both the observed measurements and prior assumptions.

To account for the variety within the solution domain and to measure reconstruction certainty, the inverse problem is tackled from a Bayesian inference standpoint. Bayesian inference yields a posterior distribution, $p(\mathbf{x}|\mathbf{y})$, from which multiple conditional samples can be extracted [6, 8]. Recently, pre-trained diffusion models [19, 27] are utilized as a powerful generative prior (a.k.a denoiser), in a zero-shot manner, to effectively sample from the conditional posterior [15, 21, 28]. The strategies for posterior (conditional) sampling via diffusion models fall into two distinct approaches. In the first approach, diffusion models are trained conditionally, directly embedding the conditioning information $\mathbf{y}$ during the training phase [19, 26, 28]. However, conditional training tends to require: (i) the assembly of a massive amount of paired data and its corresponding conditioning $(\mathbf{x}, \mathbf{y})$, and (ii) retraining when testing on new conditioning tasks, highlighting the adaptability issues. In the second approach, unconditionally pre-trained diffusion models are employed as generative prior (a.k.a denoiser) to perform conditional sampling for certain tasks. A primary

difficulty, however, is how to impose data consistency between measurements and the generated images in each iteration [11, 32]. For a comprehensive discussion on the various approaches to this challenge, refer to [25].

A.2. Proposition

Proof. Consider an iteration of gradient descent, initialized from $\mathbf{x}^{(0)}$, on the least squares problem

$$\mathbf{x}^{(t+1)} = \mathbf{x}^{(t)} + \alpha \mathcal{A}^T (\mathbf{y} - \mathcal{A}\mathbf{x}^{(t)}).$$

Defining $\mathbf{r}^{(t)} = \mathbf{y} - \mathcal{A}\mathbf{x}^{(t)}$, it follows that

$$\mathbf{r}^{(t+1)} = \left(\mathbf{I} - \alpha\mathcal{A}\mathcal{A}^T\right)\mathbf{r}^{(t)} = \ldots = \left(\mathbf{I} - \alpha\mathcal{A}\mathcal{A}^T\right)^{t+1}\mathbf{r}^{(0)}.$$

Hence,

$$\begin{aligned}
\mathbf{x}^{(t+1)} &= \mathbf{x}^{(t)} + \alpha\mathcal{A}^T\left(\mathbf{I} - \alpha\mathcal{A}\mathcal{A}^T\right)^t\mathbf{r}^{(0)} \\
&= \mathbf{x}^{(0)} + \alpha\mathcal{A}^T\sum_{i=0}^{t}\left(\mathbf{I} - \alpha\mathcal{A}\mathcal{A}^T\right)^i\mathbf{r}^{(0)} \\
&= \mathbf{x}^{(0)} + \alpha\sum_{i=0}^{t}\left(\mathbf{I} - \alpha\mathcal{A}^T\mathcal{A}\right)^i\mathcal{A}^T\mathbf{r}^{(0)}.
\end{aligned}$$

Subsequently, as long as $0 < \alpha < 1/\|\mathcal{A}\|^2$, from [5, Theorem 16], we get

$$\begin{aligned}
\lim_{t\to\infty}\mathbf{x}^{(t)} &= \mathbf{x}^{(0)} + \alpha\sum_{i=0}^{\infty}\left(\mathbf{I} - \alpha\mathcal{A}^T\mathcal{A}\right)^i\mathcal{A}^T\mathbf{r}^{(0)} \\
&= \mathbf{x}^{(0)} + \mathcal{A}^\dagger\mathbf{r}^{(0)}.
\end{aligned}$$

This concludes the proof. □

A.3. Closed-form solutions

Consider the following optimization problem in Eq. (2)

$$\hat{\mathbf{x}}_{0|t} = \arg\min_{\mathbf{x}}\frac{1}{2}\|\mathbf{y} - \mathcal{A}\mathbf{x}\|_2^2 + \frac{\lambda}{2}\|\mathbf{x} - \mathbf{x}_{0|t}\|_2^2.$$

For the MRI reconstruction task, we express $\mathcal{A}\mathbf{x} = \mathcal{M} \odot (\mathcal{F}\mathbf{x}) = \mathcal{M} \odot \mathbf{w}$, where $\mathcal{M}$ represents the Cartesian equispaced mask, $\mathcal{F}$ is the Fourier matrix, and $\odot$ signifies element-wise multiplication. Given this definition, and considering the identity $\arg\min_{\mathbf{x}}\|\mathbf{x} - \mathbf{x}_{0|t}\|_2^2 = \arg\min_{\mathbf{x}}\|\mathcal{F}\mathbf{x} - \mathcal{F}\mathbf{x}_{0|t}\|_2^2$, then the optimization problem in terms of $\mathbf{w}$ can be redefined as

$$\hat{\mathbf{w}}_{0|t} = \arg\min_{\mathbf{w}}\frac{1}{2}\|\mathcal{M} \odot \mathbf{w} - \mathbf{y}\|_2^2 + \frac{\lambda}{2}\|\mathbf{w} - \mathbf{w}_{0|t}\|_2^2.$$

By expanding the L2-norm terms, we obtain

$$\hat{\mathbf{w}}_{0|t} = \arg\min_{\mathbf{w}}\sum_{i=1}^{n}(m_iw_i - y_i)^2 + \lambda\sum_{i=1}^{n}(w_i - w_{0|t}^i)^2.$$

The solution for $\hat{\mathbf{w}}_{0|t}$ is

$$\hat{\mathbf{w}}_{0|t} = \frac{\mathcal{M}\mathbf{y} + \lambda\mathbf{w}_{0|t}}{\mathcal{M} + \lambda}.$$

Given the relation $\hat{\mathbf{x}}_{0|t} = \mathcal{F}^{-1}\hat{\mathbf{w}}_{0|t}$, we can then deduce

$$\hat{\mathbf{x}}_{0|t} = \mathcal{F}^{-1}\left(\frac{\mathcal{M}\mathbf{y} + \lambda\mathcal{F}\mathbf{x}_{0|t}}{\mathcal{M} + \lambda}\right)$$

Consider the range-null space decomposition defined in Eq. (1) $\hat{\mathbf{x}}_{0|t} = \mathcal{A}^\dagger\mathbf{y} + \left(\mathbf{I} - \mathcal{A}^\dagger\mathcal{A}\right)\mathbf{x}_{0|t}$, where $\mathcal{A}^\dagger$ denotes the pseudo-inverse of matrix $\mathcal{A}$ and $\mathbf{I}$ is the identity matrix. For MRI, the forward operator is modelled as $\mathcal{A} = \mathcal{M}\mathcal{F}$. An important property that arises is $\mathcal{A}\mathcal{A}\mathcal{A} \equiv \mathcal{A}$, which suggests that $\mathcal{A}$ itself can be represented as its pseudo-inverse $\mathcal{A}^\dagger$. With this property, the pseudo-inverse is then expressed as $\mathcal{A}^\dagger = \mathcal{F}^{-1}\mathcal{M}$. Substituting this representation into our original expression, we obtain

$$\hat{\mathbf{x}}_{0|t} = \mathcal{F}^{-1}\mathcal{M}\mathbf{y} + \left(\mathbf{I} - \mathcal{F}^{-1}\mathcal{M}\mathcal{F}\right)\mathbf{x}_{0|t}.$$

Using the Fourier identity $\mathcal{F}^{-1}\mathcal{F} = \mathbf{I}$, we can further simplify this to:

$$\hat{\mathbf{x}}_{0|t} = \mathcal{F}^{-1}\left(\mathcal{M}\mathbf{y} + (\mathbf{I} - \mathcal{M})\mathcal{F}\mathbf{x}_{0|t}\right)$$

A.4. Posterior mean

A.4.1 Posterior mean with additional measurement for VPSDE

A notable SDE with an analytic transition probability is the variance-Preserving SDE (VPSDE) [22, 30], which considers $\mathbf{f}(\mathbf{x}_t, t) = -\frac{1}{2}\beta(t)\mathbf{x}_t$ and $g(t) = \sqrt{\beta(t)}$, where $\beta(t) = \beta_{min} + t(\beta_{max} - \beta_{min})$; and its transition probability follows a Gaussian distribution of $p_{0t}(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \boldsymbol{\mu}_t\mathbf{x}_0, \boldsymbol{\sigma}_t^2\mathbf{I})$ with $\boldsymbol{\mu}_t = \exp\{-\frac{1}{2}\int_0^t\beta(s)s\}$ and $\boldsymbol{\sigma}_t^2 = 1 - \exp\{-\int_0^t\beta(s)s\}$. Given such transition probability, we seek to derive the corresponding posterior mean with additional measurement.

Begin by representing the distribution $p(\mathbf{x}_t|\mathbf{y})$ as marginalizing out $\mathbf{x}_0$ conditioned on $\mathbf{y}$:

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

Differentiate w.r.t. $\mathbf{x}_t$ on both sides

$$\nabla_{\mathbf{x}_t}p(\mathbf{x}_t|\mathbf{y}) = \int_{\mathbf{x}_0}p(\mathbf{x}_0|\mathbf{y})\nabla_{\mathbf{x}_t}p(\mathbf{x}_t|\mathbf{x}_0, \mathbf{y})d\mathbf{x}_0.$$

With our new probability distribution model, the gradient becomes

$$\nabla_{\mathbf{x}_t}\log p(\mathbf{x}_t|\mathbf{x}_0) = \frac{(\boldsymbol{\mu}_t\mathbf{x}_0 - \mathbf{x}_t)}{\boldsymbol{\sigma}_t^2}.$$

Inserting this into our previous equation, we have

$$\nabla_{\mathbf{x}_t} p(\mathbf{x}_t|\mathbf{y}) = \int_{\mathbf{x}_0} p(\mathbf{x}_0|\mathbf{y})p(\mathbf{x}_t|\mathbf{x}_0,\mathbf{y})\frac{(\boldsymbol{\mu}_t\mathbf{x}_0 - \mathbf{x}_t)}{\boldsymbol{\sigma}_t^2} d\mathbf{x}_0.$$

Simplifying the above equation, we get:

$$\nabla_{\mathbf{x}_t} p(\mathbf{x}_t|\mathbf{y}) = \frac{1}{\boldsymbol{\sigma}_t^2}\left[\int_{\mathbf{x}_0} p(\mathbf{x}_0|\mathbf{y})p(\mathbf{x}_t|\mathbf{x}_0,\mathbf{y})\boldsymbol{\mu}_t\mathbf{x}_0 d\mathbf{x}_0 \right.$$
$$\left. - \int_{\mathbf{x}_0} p(\mathbf{x}_0|\mathbf{y})p_t(\mathbf{x}_t|\mathbf{x}_0,\mathbf{y})\mathbf{x}_t d\mathbf{x}_0\right].$$

Using Bayes' rule and recognizing the marginalization, we get:

$$\nabla_{\mathbf{x}_t} p(\mathbf{x}_t|\mathbf{y}) = \frac{1}{\boldsymbol{\sigma}_t^2}\left[\int_{\mathbf{x}_0} \boldsymbol{\mu}_t\mathbf{x}_0 p(\mathbf{x}_t|\mathbf{y})p(\mathbf{x}_0|\mathbf{x}_t,\mathbf{y}) d\mathbf{x}_0 \right.$$
$$\left. - \mathbf{x}_t p(\mathbf{x}_t|\mathbf{y})\right].$$

$$\nabla_{\mathbf{x}_t} p(\mathbf{x}_t|\mathbf{y}) = \frac{1}{\boldsymbol{\sigma}_t^2}\left[\boldsymbol{\mu}_t p(\mathbf{x}_t|\mathbf{y})\mathbb{E}[\mathbf{x}_0|\mathbf{x}_t,\mathbf{y}] - \mathbf{x}_t p(\mathbf{x}_t|\mathbf{y})\right].$$

$$\frac{\nabla_{\mathbf{x}_t} p(\mathbf{x}_t|\mathbf{y})}{p(\mathbf{x}_t|\mathbf{y})} = \frac{1}{\boldsymbol{\sigma}_t^2}\left[\boldsymbol{\mu}_t\mathbb{E}[\mathbf{x}_0|\mathbf{x}_t,\mathbf{y}] - \mathbf{x}_t\right].$$

Using the identity property of logarithm $\nabla_{\mathbf{x}} \log p(\mathbf{x}) = \nabla_{\mathbf{x}} p(\mathbf{x})/p(\mathbf{x})$, we can rewrite:

$$\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t|\mathbf{y}) = \frac{1}{\boldsymbol{\sigma}_t^2}\left[\boldsymbol{\mu}_t\mathbb{E}[\mathbf{x}_0|\mathbf{x}_t,\mathbf{y}] - \mathbf{x}_t\right].$$

From this, the posterior mean becomes:

$$\mathbb{E}[\mathbf{x}_0|\mathbf{x}_t,\mathbf{y}] = \frac{\mathbf{x}_t + \boldsymbol{\sigma}_t^2 \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t|\mathbf{y})}{\boldsymbol{\mu}_t}.$$

This shows that the posterior mean of $\mathbf{x}_0$ conditioned on $\mathbf{x}_t$ and $\mathbf{y}$ now incorporates a scaling by $\boldsymbol{\mu}_t$. By considering $\boldsymbol{\mu}_t = \sqrt{\overline{\alpha}_t}$ and $\boldsymbol{\sigma}_t^2 = 1 - \overline{\alpha}_t$, we have then

$$\mathbb{E}[\mathbf{x}_0|\mathbf{x}_t,\mathbf{y}] = \frac{1}{\sqrt{\overline{\alpha}_t}}(\mathbf{x}_t + (1 - \overline{\alpha}_t)\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t|\mathbf{y})).$$

A.4.2 Approximated Conditional Posterior Mean

$$\mathbb{E}[\mathbf{x}_0|\mathbf{x}_t,\mathbf{y}] = \frac{1}{\sqrt{\overline{\alpha}_t}}(\mathbf{x}_t + (1 - \overline{\alpha}_t)\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t|\mathbf{y}))$$

Considering Bayes' rule we have

$$\mathbb{E}[\mathbf{x}_0|\mathbf{x}_t,\mathbf{y}] = \frac{1}{\sqrt{\overline{\alpha}_t}}\big(\mathbf{x}_t + (1 - \overline{\alpha}_t)(\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t)$$
$$+ \nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\mathbf{x}_t))\big)$$

By knowing that $\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t) \simeq \frac{-1}{\sqrt{1-\overline{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)$, we get

$$\mathbb{E}[\mathbf{x}_0|\mathbf{x}_t,\mathbf{y}] \simeq \frac{1}{\sqrt{\overline{\alpha}_t}}\bigg(\mathbf{x}_t + (1 - \overline{\alpha}_t)\Big(\frac{-1}{\sqrt{1-\overline{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)$$
$$+ \nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\mathbf{x}_t)\Big)\bigg)$$

which can be simplified further as

$$\mathbb{E}[\mathbf{x}_0|\mathbf{x}_t,\mathbf{y}] \simeq \frac{1}{\sqrt{\overline{\alpha}_t}}\big(\mathbf{x}_t - \sqrt{1-\overline{\alpha}_t}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)$$
$$+ (1 - \overline{\alpha}_t)\nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\mathbf{x}_t))\big)$$

From approximation made by DPS [11], that is, $\nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\mathbf{x}_t) \simeq -1/\boldsymbol{\sigma}_{\mathbf{y}}^2 \nabla_{\mathbf{x}_t}\|\mathbf{y} - \mathcal{A}(\mathbf{x}_{0|t})\|_2^2$, we then get

$$\tilde{\mathbf{x}}_{0|t} \simeq \frac{1}{\sqrt{\overline{\alpha}_t}}\Big[\mathbf{x}_t - \sqrt{1-\overline{\alpha}_t}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t) - \zeta\nabla_{\mathbf{x}_t}\|\mathbf{y} - \mathcal{A}\mathbf{x}_{0|t}\|_2^2\Big].$$

A.5. Theoretical Insight

Consider the outer-level objective in Eq. (2). This objective can be regarded as a method for estimating this expectation $\mathbb{E}[\mathbf{x}_0|\mathbf{x}_t,\mathbf{y}]$, though it comes from different conceptual frameworks. This is achieved by using an isotropic Gaussian approximation of the denoising posterior.

Let us define the isotropic Gaussian approximation as $q_t(\mathbf{x}_0|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_0|\mathbf{x}_{0|t}, r_t^2\mathbf{I})$. We can derive that the approximate distribution $q_t(\mathbf{x}_0|\mathbf{x}_t,\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x}_0)q_t(\mathbf{x}_0|\mathbf{x}_t)$ for $p_t(\mathbf{x}_0|\mathbf{x}_t,\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x}_0)p_t(\mathbf{x}_0|\mathbf{x}_t)$ is also Gaussian. The mean of this distribution, $\mathbb{E}_q[\mathbf{x}_0|\mathbf{x}_t,\mathbf{y}]$, can be obtained by solving the optimization problem:

$$\mathbb{E}_q[\mathbf{x}_0|\mathbf{x}_t,\mathbf{y}] = \arg\max_{\mathbf{x}_0} \log q_t(\mathbf{x}_0|\mathbf{x}_t, y)$$
$$= \arg\max_{\mathbf{x}_0}\left[\log p(\mathbf{y}|\mathbf{x}_0) + \log p_t(\mathbf{x}_0|\mathbf{x}_t)\right]$$
$$= \arg\min_{\mathbf{x}_0}\left[\|\mathbf{y} - \mathcal{A}\mathbf{x}_0\|^2 + \frac{\sigma_t^2}{r_t^2}\|\mathbf{x}_0 - \mathbf{x}_{0|t}\|^2\right]$$

Furthermore, we know that in scenarios when $\sigma_{\mathbf{y}} = 0$, the following condition holds:

$$\mathbb{E}_q[\mathbf{x}_0|\mathbf{x}_t,\mathbf{y}] = \mathbb{E}_q[\mathcal{A}^\dagger \mathbf{A}\mathbf{x}_0 + (\mathbf{I} - \mathcal{A}^\dagger\mathbf{A})\mathbf{x}_0|\mathbf{x}_t,\mathbf{y}]$$
$$= \mathcal{A}^\dagger\mathbf{y} + (\mathbf{I} - \mathcal{A}^\dagger\mathbf{A})\mathbb{E}_q[\mathbf{x}_0|\mathbf{x}_t,\mathbf{y}].$$

Now, given the following relation,

$$\mathbb{E}[\mathbf{x}_0|\mathbf{x}_t,\mathbf{y}] = \mathbb{E}[\mathbf{x}_0|\mathbf{x}_t] + \frac{1 - \overline{\alpha}_t}{\sqrt{\overline{\alpha}_t}}\nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\mathbf{x}_t)$$

we then have

$$\mathbb{E}[\mathbf{x}_0|\mathbf{x}_t,\mathbf{y}] = \mathcal{A}^\dagger\mathbf{y} + (\mathbf{I} - \mathcal{A}^\dagger\mathcal{A})\Big(\mathbb{E}[\mathbf{x}_0|\mathbf{x}_t]$$
$$+ \frac{1 - \overline{\alpha}_t}{\sqrt{\overline{\alpha}_t}}\nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\mathbf{x}_t)\Big)$$
$$= \mathcal{A}^\dagger\mathbf{y} + (\mathbf{I} - \mathcal{A}^\dagger\mathcal{A})\Big(\mathbb{E}[\mathbf{x}_0|\mathbf{x}_t] - \zeta\nabla_{\mathbf{x}_t}\|\mathbf{y} - \mathcal{A}\mathbf{x}_{0|t}\|^2\Big)$$

Table 1. Comparison of R-BGDM against various supervised methods across multiple datasets.

Method	BraTS-MRI				fastMRI				LIDC-CT	
	8× ACR		24× ACR		4× ACR		8× ACR		23 **Proj**	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
DuDoRNet [35]	37.88	0.985	18.46	0.662	33.46	0.856	29.65	0.777	–	–
SIN-4c-PRN [33]	–	–	–	–	–	–	–	–	30.48	0.895
R-BGDM	**38.46**	**0.964**	**30.04**	**0.887**	**34.73**	**0.875**	**32.74**	**0.835**	**35.82**	**0.911**

Table 2. Computation Times (Rounded to the nearest integer) to process 1000 samples for various algorithms across each task.

Algorithm	Super-Resolution (fastMRI brain)	Undersampled MRI Rec (BraTS)	Undersampled MRI Rec (fastMRI Knee)	Sparse-view CT Rec (LIDC)
DPS [11]	6788 s	7112 s	9100 s	—
DDNM [32]	3753 s	4467 s	4773 s	48011
DDS [12]	3921 s	4610 s	5162 s	—
BGDM	6792 s	71 72s	9128 s	47987 s
R-BGDM	6944 s	7309 s	9334 s	48426 s

Two remarks regarding the last result from the equation presented above are worth mentioning.

- **Balanced Approach**: This comprehensive approach harnesses both the structural advantages of linear projections and the adaptive capabilities of non-linear optimization, providing a robust framework for tackling complex inverse problems with high accuracy.

- **Synergistic Effect**: The corrections from the projection and gradient guidance can be mutually reinforcing, where the initial projection provides a good starting point that is refined by the gradient steps, leading to potentially faster convergence and more accurate reconstruction.

A.6. Limitations

A few limitations remain that deserve further examination.

- Despite achieving superior reconstruction results compared to other methods [11, 29, 32] and demonstrating more efficient sampling for medical imaging applications [12, 13, 20, 29], BGDM requires the tuning of ζ and R-BGDM remains sensitive to both hyper-parameters ζ and γ. Therefore, exploring a more general hyperparameter tuning approach, such as Bayesian optimization, would be beneficial.

- The BGDM algorithm performs similarly to ΠGDM and DPS in terms of computational efficiency, as shown in 2. However, the R-BGDM variant, while more computationally intensive and requiring additional memory, significantly speeds up the sampling process, as shown in our experiments. This highlights a trade-off between increased computational resources and faster sampling times.

- It should be noted that our CT simulation adheres to the 2D parallel beam geometry assumption, aligning with the baseline models used in other studies for direct comparison. This differs from the more complex 3D cone-beam CT or helical CT simulations [24].

In future work, we plan to enhance the method for compatibility with 3D simulations and, adaptability to distributional shifts [3, 4].

A.7. Additional Results

A.7.1 Comparing R-BGDM with Supervised Methods

Similar to other zero-shot inverse problem solvers [11, 23, 32], R-BGDM is superior to existing supervised methods [33, 35] in these dimensions:

- R-BGDM can be a zero-shot solver for diverse tasks, while supervised methods need to train separate models for each task and sampling patterns.

- R-BGDM demonstrates robustness to patterns of undersampling and sparsification, whereas supervised techniques exhibit weak generalizability.

- R-BGDM, akin to ScoreMed [29] and Score-MRI [13], achieves notably enhanced results on medical datasets compared to supervised methods.

These claims are substantiated by the experimental results in Table 1. The results are reported from [13, 29].

A.7.2 Additional Visual Results
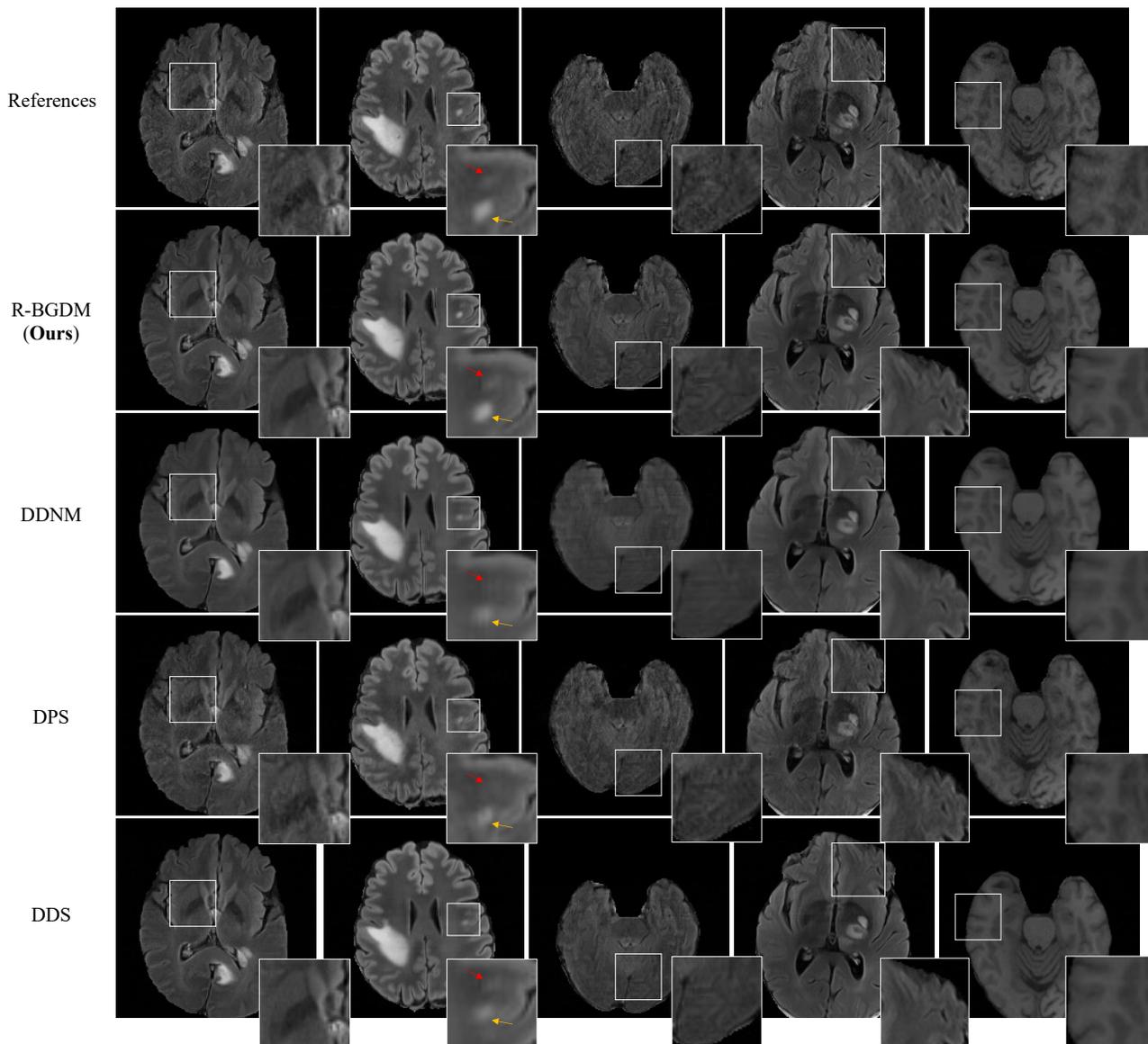
Figure 1. Additional results from undersampled MRI reconstruction on Brats at **8x** acceleration rate.
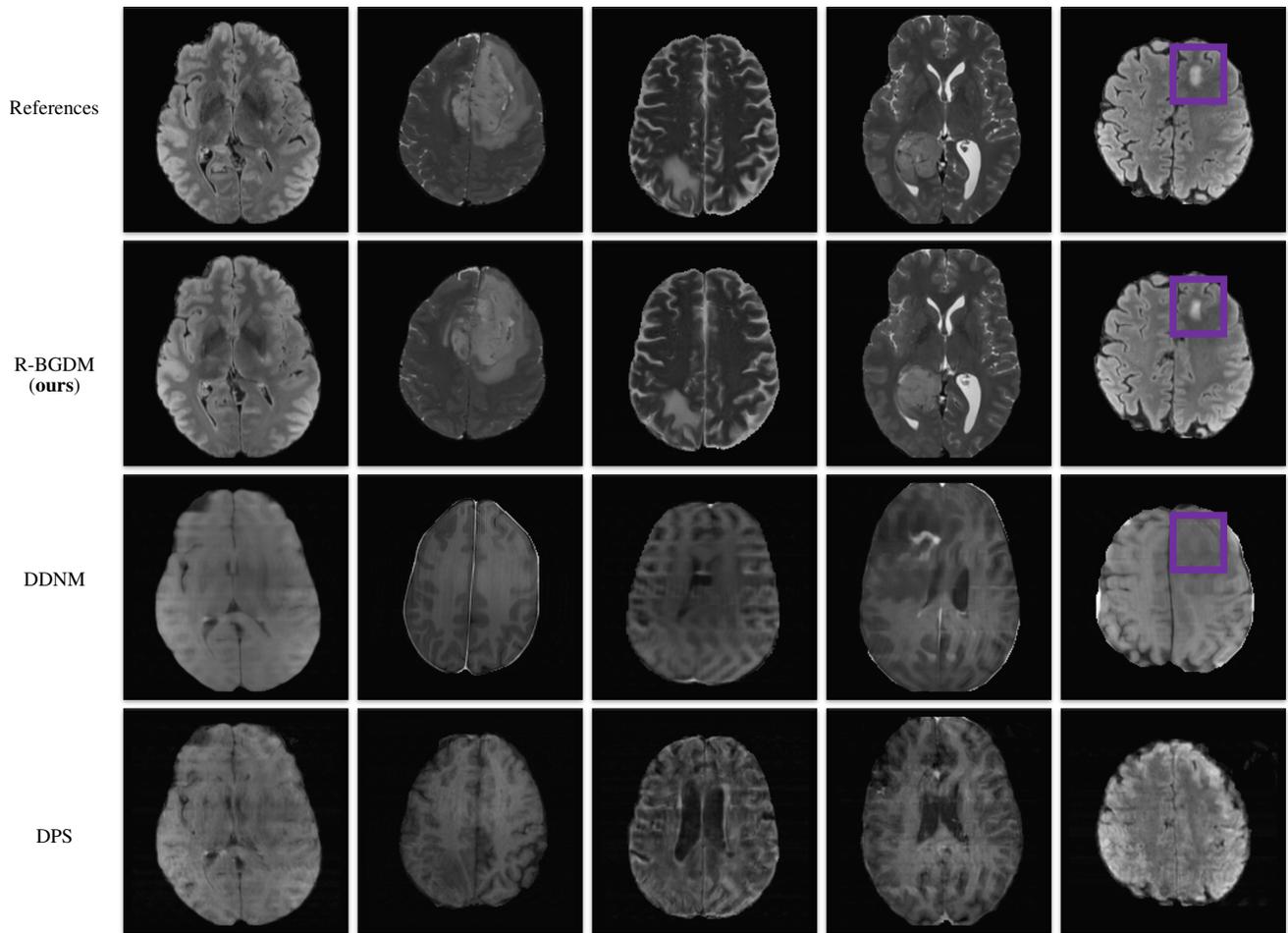
Figure 2. Additional results from undersampled MRI reconstruction on Brats at **24x** acceleration rate.
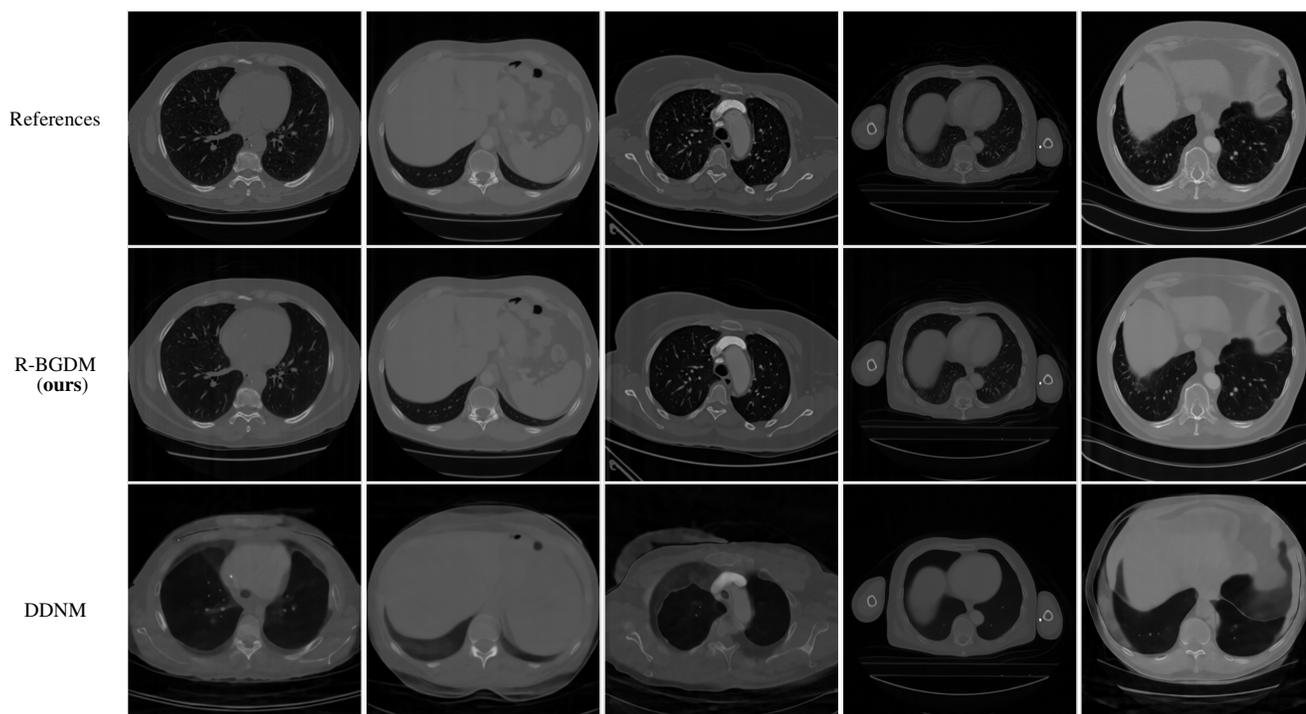
Figure 3. Additional results from sparse-view CT reconstruction on LIDC dataset with **23** projections.

References

R-BGDM
(**Ours**)

DDNM

DPS

DDS
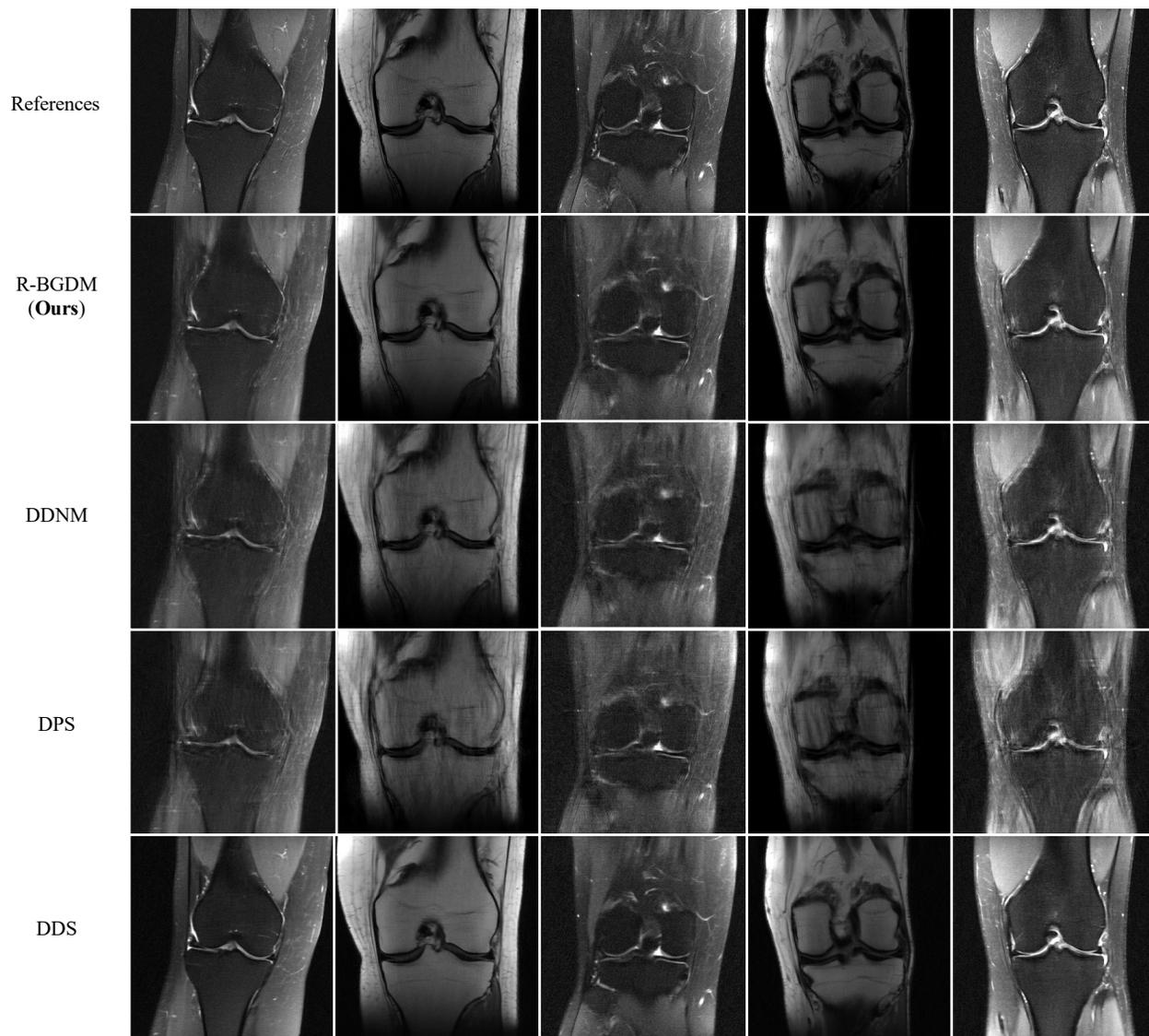
Figure 4. Additional reconstruction results for undersampled knee fastMRI at **4x** acceleration rate.
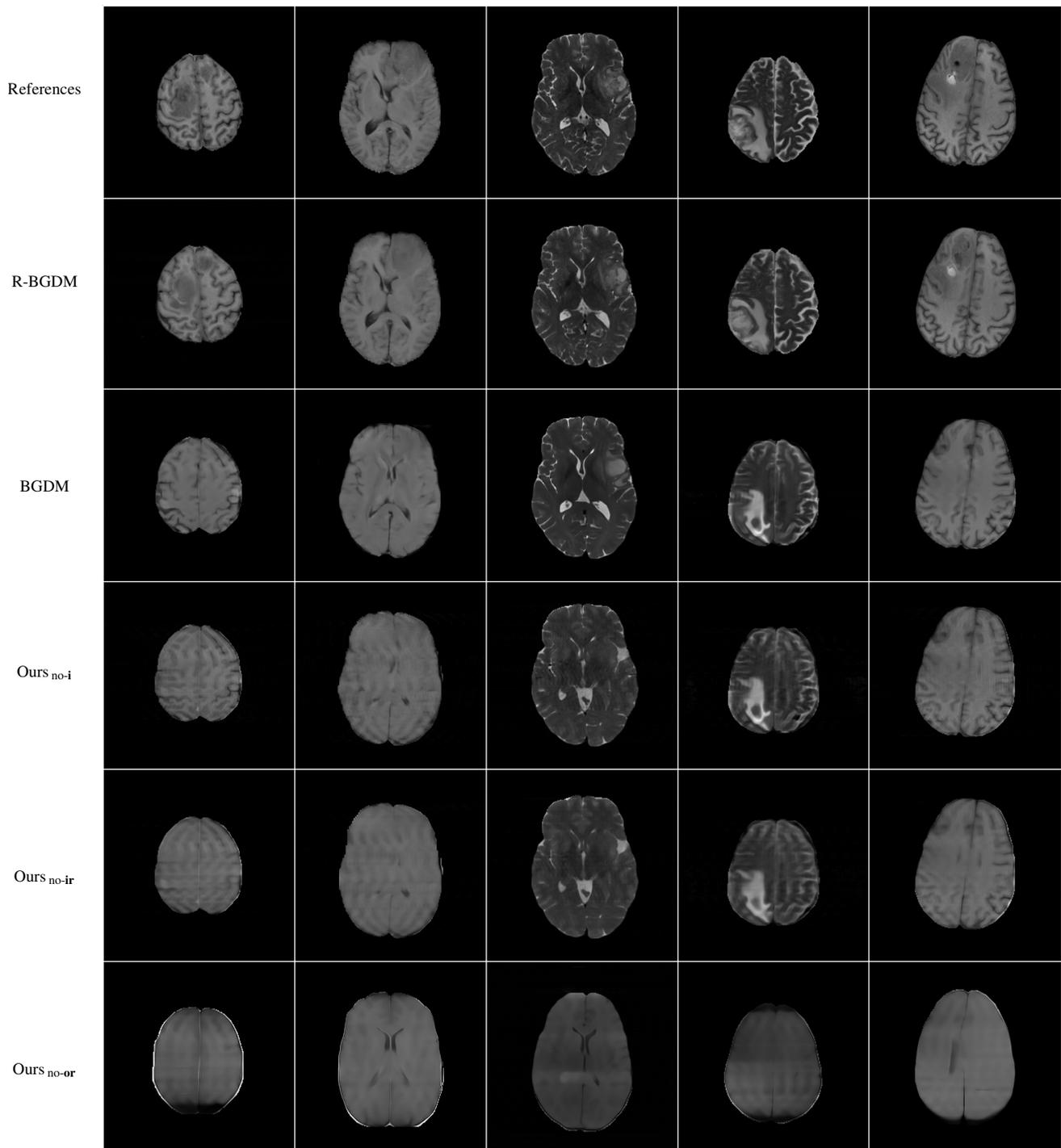
Figure 5. Additional results of our ablation study from undersampled MRI reconstruction on Brats at **24x** acceleration rate.
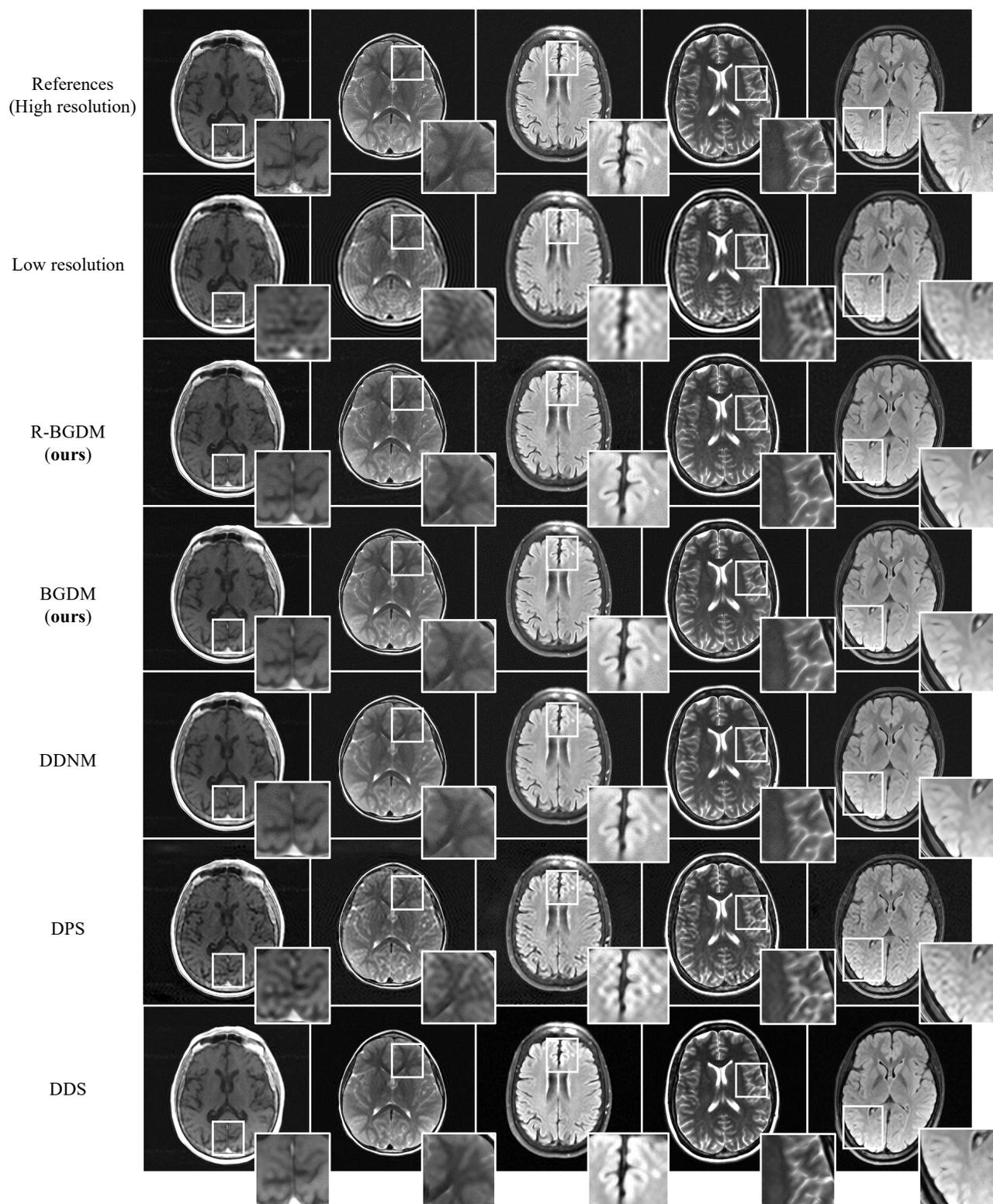
Figure 6. Additional results for super-resolution fastMRI at **16x** acceleration rate.

References

[1] Lynton Ardizzone, Jakob Kruse, Carsten Rother, and Ullrich Köthe. Analyzing inverse problems with invertible neural networks. In *International Conference on Learning Representations*, 2019. 1

[2] Muhammad Asim, Max Daniels, Oscar Leong, Ali Ahmed, and Paul Hand. Invertible generative models for inverse problems: mitigating representation error and dataset bias. In *International Conference on Machine Learning*, pages 399–409. PMLR, 2020. 1

[3] Hossein Askari, Yasir Latif, and Hongfu Sun. Mapflow: latent transition via normalizing flow for unsupervised domain adaptation. *Machine Learning*, 112(8):2953–2974, 2023. 4

[4] Riccardo Barbano, Alexander Denker, Hyungjin Chung, Tae Hoon Roh, Simon Arrdige, Peter Maass, Bangti Jin, and Jong Chul Ye. Steerable conditional diffusion for out-of-distribution adaptation in imaging inverse problems. *arXiv preprint arXiv:2308.14409*, 2023. 4

[5] Adi Ben-Israel and A Charnes. Contributions to the theory of generalized inverses. *Journal of the Society for Industrial and Applied Mathematics*, 11(3):667–699, 1963. 2

[6] David M Blei, Alp Kucukelbir, and Jon D McAuliffe. Variational inference: A review for statisticians. *Journal of the American statistical Association*, 112(518):859–877, 2017. 1

[7] Ashish Bora, Ajil Jalal, Eric Price, and Alexandros G Dimakis. Compressed sensing using generative models. In *International conference on machine learning*, pages 537–546. PMLR, 2017. 1

[8] Steve Brooks, Andrew Gelman, Galin Jones, and Xiao-Li Meng. *Handbook of markov chain monte carlo*. CRC press, 2011. 1

[9] Emmanuel J Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on information theory*, 52(2):489–509, 2006. 1

[10] Emmanuel J Candès and Michael B Wakin. An introduction to compressive sampling. *IEEE signal processing magazine*, 25(2):21–30, 2008. 1

[11] Hyungjin Chung, Jeongsol Kim, Michael Thompson McCann, Marc Louis Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *The Eleventh International Conference on Learning Representations*, 2023. 2, 3, 4

[12] Hyungjin Chung, Suhyeon Lee, and Jong Chul Ye. Decomposed diffusion sampler for accelerating large-scale inverse problems. In *The Twelfth International Conference on Learning Representations*, 2024. 4

[13] Hyungjin Chung and Jong Chul Ye. Score-based diffusion models for accelerated mri. *Medical image analysis*, 80:102479, 2022. 4

[14] Tiangang Cui, James Martin, Youssef M Marzouk, Antti Solonen, and Alessio Spantini. Likelihood-informed dimension reduction for nonlinear inverse problems. *Inverse Problems*, 30(11):114015, 2014. 1

[15] Giannis Daras, Yuval Dagan, Alex Dimakis, and Constantinos Daskalakis. Score-guided intermediate level optimization: Fast langevin mixing for inverse problems. In *International Conference on Machine Learning*, pages 4722–4753. PMLR, 2022. 1

[16] Maryam Fazel, Emmanuel Candes, Benjamin Recht, and Pablo Parrilo. Compressed sensing and robust recovery of low rank matrices. In *2008 42nd Asilomar Conference on Signals, Systems and Computers*, pages 1043–1047. IEEE, 2008. 1

[17] Hwan Goh, Sheroze Sheriffdeen, Jonathan Wittmer, and Tan Bui-Thanh. Solving bayesian inverse problems via variational autoencoders. In *Mathematical and Scientific Machine Learning*, pages 386–425. PMLR, 2022. 1

[18] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT Press, 2016. 1

[19] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020. 1

[20] Ajil Jalal, Marius Arvinte, Giannis Daras, Eric Price, Alexandros G Dimakis, and Jon Tamir. Robust compressed sensing mri with deep generative priors. *Advances in Neural Information Processing Systems*, 34:14938–14954, 2021. 4

[21] Zahra Kadkhodaie and Eero Simoncelli. Stochastic solutions for linear inverse problems using the prior implicit in a denoiser. *Advances in Neural Information Processing Systems*, 34:13242–13254, 2021. 1

[22] Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. *Advances in Neural Information Processing Systems*, 35:26565–26577, 2022. 2

[23] Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. *Advances in Neural Information Processing Systems*, 35:23593–23606, 2022. 4

[24] Donghwan Kim, Sathish Ramani, and Jeffrey A Fessler. Combining ordered subsets and momentum for accelerated x-ray ct image reconstruction. *IEEE transactions on medical imaging*, 34(1):167–178, 2014. 4

[25] Xin Li, Yulin Ren, Xin Jin, Cuiling Lan, Xingrui Wang, Wenjun Zeng, Xinchao Wang, and Zhibo Chen. Diffusion models for image restoration and enhancement–a comprehensive survey. *arXiv preprint arXiv:2308.09388*, 2023. 2

[26] Guan-Horng Liu, Arash Vahdat, De-An Huang, Evangelos A Theodorou, Weili Nie, and Anima Anandkumar. I2sb: image-to-image schrödinger bridge. In *Proceedings of the 40th International Conference on Machine Learning*, pages 22042–22062, 2023. 1

[27] Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International Conference on Machine Learning*, pages 8162–8171. PMLR, 2021. 1

[28] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022. 1

[29] Yang Song, Liyue Shen, Lei Xing, and Stefano Ermon. Solving inverse problems in medical imaging with score-based generative models. In *International Conference on Learning Representations*, 2022. 4

[30] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021. 2

[31] Jie Tang, Brian E Nett, and Guang-Hong Chen. Performance comparison between total variation (tv)-based compressed sensing and statistical iterative reconstruction algorithms. *Physics in Medicine & Biology*, 54(19):5781, 2009. 1

[32] Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-shot image restoration using denoising diffusion null-space model. In *The Eleventh International Conference on Learning Representations*, 2023. 2, 4

[33] Haoyu Wei, Florian Schiffers, Tobias Würfl, Daming Shen, Daniel Kim, Aggelos K Katsaggelos, and Oliver Cossairt. 2-step sparse-view ct reconstruction with a domain-specific perceptual network. *arXiv preprint arXiv:2012.04743*, 2020. 4

[34] Jay Whang, Qi Lei, and Alex Dimakis. Solving inverse problems with a flow-based noise model. In *International Conference on Machine Learning*, pages 11146–11157. PMLR, 2021. 1

[35] Bo Zhou and S Kevin Zhou. Dudornet: learning a dual-domain recurrent network for fast mri reconstruction with deep t1 prior. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4273–4282, 2020. 4