Regularization by Denoising Diffusion Process for MRI Reconstruction

Batu Ozturkler[1], **Morteza Mardani**[2], **Arash Vahdat**[2], **Jan Kautz**[2], **John Pauly**[1]
Department of Electical Engineering,
Stanford University[1]
NVIDIA[2]
ozt@stanford.edu,{mmardani,avahdat,jkautz}@nvidia.com,
pauly@stanford.edu

Abstract

Diffusion models have recently delivered state-of-the-art performance for MRI reconstruction with improved robustness. However, these models fail when there is a large distribution shift, and their long inference times impede their clinical utility. Recently, regularization by denoising diffusion process (RED-diff) was introduced for solving general inverse problems. RED-diff uses a variational sampler based on a measurement consistency loss and a score matching regularization. In this paper, we extend RED-diff to MRI reconstruction. RED-diff formulates MRI reconstruction as stochastic optimization, and outperforms diffusion baselines in PSNR/SSIM with $3\times$ faster inference while using the same amount of memory. The code is publicly available at `https://github.com/NVlabs/SMRD`.

1 Introduction

Magnetic Resonance Imaging (MRI) is a widely used non-invasive imaging technique due to its ability to generate high-quality images. However, acquiring clinical MRI data requires long scan times. Imaging can be accelerated by using multiple receiver coils, and by reducing the amount of captured data with Fourier domain (k-space) undersampling [1, 2]. Recently, unrolled methods that alternate between measurement consistency and a neural-network based regularization have shown superior performance as a data-driven approach. [3, 4, 5]

Generative diffusion models gained popularity for MRI reconstruction due to their high sample quality, improving robustness over unrolled methods under distribution shifts [6, 7, 8, 9]. Diffusion models can be pretrained for MRI to serve as the data prior and the pretrained model can be used in a plug-and-play fashion by incorporating the forward model at inference time. This approach allows for universally solving downstream reconstruction tasks without the need for re-training or fine-tuning. However, diffusion models still fail dramatically under large distribution shifts such as scan parameter change, or anatomy change between training and testing [9]. Furthermore, inference time for diffusion models is much larger than end-to-end approaches due to the sequential denoising procedure during reverse diffusion, impeding their clinical utility [7].

Recently, [10] proposed regularization by denoising diffusion (RED-diff) for solving generic inverse problems. RED-diff uses a variational sampler based on a measurement consistency loss and a score matching regularization. In this paper, for the first time, we propose to extend RED-diff for MRI reconstruction.

Contributions. Our contributions can be summarized as follows:

NeurIPS 2023 Workshop on Deep Learning and Inverse Problems.

Algorithm 1 RED-diff: regularization by denoising diffusion process for MRI reconstruction

Input: k-space data y; acquisition model $A = \Omega FS$; $\{\alpha_t, \sigma_t, \lambda_t\}_{t=1}^T$
Initialize: $\mu = x_{zf} = A^{-1}y$
for $t = T, ..., 1$ **do**
  $\epsilon \sim \mathcal{N}(0, I)$
  $x_t = \alpha_t \mu + \sigma_t \epsilon$
  $loss = \|A\mu - y\|^2 + \lambda_t(\text{sg}[\epsilon_\theta(x_t; t) - \epsilon])^T \mu$
  $\mu \leftarrow \text{OptimizerStep}(loss)$
end for
return μ

- We propose regularization by denoising diffusion processes for MRI reconstruction (RED-diff), a variational inference method for MRI reconstruction using pre-trained diffusion models.
- We evaluate RED-diff for MRI reconstruction on FastMRI and Mridata, and show that it achieves state-of-the-art performance across different acceleration rates and anatomies.
- RED-diff achieves $3\times$ faster inference while using the same amount of memory and improving reconstruction performance.

2 Background

2.1 Diffusion Models

Diffusion models are a recent class of generative models showing remarkable sample quality for computer vision tasks [11]. Diffusion models consist of two processes: a forward process that gradually adds noise to input images and a reverse process that learns to generate images by iterative denoising. A popular class of diffusion models uses the variance preserving stochastic differential equation (VP-SDE) [12]. The forward and reverse process is characterized by the noise schedule $\beta(t)$ with $t \in [0, T]$ where t is the timestep. $\beta(t)$ is designed such that the final distribution of x_T at the end of the process converges to a standard Gaussian distribution.

The reverse generative process requires estimating the score function $\nabla_{x_t} \log p(x_t)$, which denotes the score function of diffused data at time t. $\nabla_{x_t} \log p(x_t)$ can be estimated by training a joint neural network, denoted as $\epsilon_\theta(x_t; t)$, via denoising score matching [13]. For denoising score matching, diffused samples are generated by

$$x_t = \alpha_t x_0 + \sigma_t \epsilon \tag{1}$$

where $\epsilon \sim \mathcal{N}(0, I)$, $x_0 \sim p_{\text{data}}$ is the data distribution, $\sigma_t = 1 - e^{-\int_0^t \beta(s)ds}$, and $\alpha_t = \sqrt{1 - \sigma_t^2}$, and $\epsilon_\theta(x_t; t) \approx -\sigma_t \nabla_{x_t} \log p(x_t)$.

2.2 Diffusion Models for Accelerated MRI Reconstruction

The forward model for accelerated MRI with compressed sensing [1] and parallel imaging [2] is given by

$$y = \Omega FSx_0 + \nu \tag{2}$$

where y is the measurement, x_0 is the real image, S are sensitivity maps, F is the Fourier transform, Ω is the subsampling mask, $\nu \sim \mathcal{N}(0, \sigma_\nu^2 I)$, and $A = \Omega FS$ is the forward model.

In MRI reconstruction, the real image x from measurements y can be reconstructed by sampling from the posterior distribution $p(x_0|y)$. This can be achieved by leveraging diffusion models as data priors assuming the conditional score function $p(x_t|y)$ is available [7]. The conditional score function can be obtained using Bayes rule:

$$\nabla_{x_t} \log p(x_t|y) = \nabla_{x_t} \log p(y|x_t) + \nabla_{x_t} \log p(x_t) \tag{3}$$

where $\nabla_{x_t} \log p(x_t)$ is estimated by the diffusion model as discussed in Section 2.1. However, the likelihood term $\nabla_{x_t} \log p(y|x_t)$ is often intractable to estimate. Due to the intractability of the likelihood, previous works often resort to approximations [7] or projections onto the measurement subspace [14].

Anatomy	Brain	Knee		Timing
R	$R=4$	$R=12$	$R=16$	(sec/iter)
Zero-filled	27.8/0.81	24.5/0.63	24.0/0.60	-
CSGM-Langevin	36.3/0.78	31.4/**0.82**	31.8/**0.79**	0.344
RED-diff	**37.1/0.83**	**33.2**/0.78	**32.7**/0.77	**0.114**

Table 1: Reconstruction PSNR/SSIM for fastMRI brain and Mridata knee dataset.

3 RED-diff for MRI Reconstruction

To sidestep the challenges of approximating the likelihood term $\nabla_{x_t} \log p(y|x_t)$, [10] recently proposed RED-diff, a variational inference approach based on KL minimization for solving general inverse problems. RED-diff considers the following KL objective:

$$KL(q(x_0|y)\|p(x_0|y)) = -\mathbb{E}_{q(x_0|y)}[\log p(y|x_0)] + KL(q(x_0|y)\|p(x_0)) + \log p(y) \tag{4}$$

where $q := \mathcal{N}(\mu, \sigma^2 I)$ is a variational distribution. Upon inspection of Equation 4, it is observed that the objective is a composition of the variational bound that is often used for training variational autoencoders [15], and the observation likelihood $\log p(y)$ which is constant with respect to q, and thus can be ignored. Then, it can be shown that performing KL minimization amounts to minimizing the following score-matching loss [10]:

Proposition 1. *The KL minimization with respect to q in Eq. 4 is equivalent to minimizing the score matching loss:*

$$\min_{\{\mu,\sigma\}} \mathbb{E}_{q(x_0|y)} \left[\frac{\|y - Ax_0\|_2^2}{2\sigma_\nu^2} \right] + \int_0^T \frac{\beta(t)}{2} \mathbb{E}_{q(x_t|y)} \left[\left\| \nabla_{x_t} \log q(x_t|y) - \nabla_{x_t} \log p(x_t) \right\|_2^2 \right] dt, \tag{5}$$

where $q(x_t|y) = \mathcal{N}(\alpha_t \mu, (\alpha_t^2 \sigma^2 + \sigma_t^2) I)$ produces samples x_t by drawing x_0 from $q(x_0|y)$ and applying the forward process in Eq. 1.

Equation 5 is composed of a measurement consistency loss obtained by the definition of $p(y|x_0)$, and a score-matching regularization term imposed by the diffusion prior. The score-matching term is obtained by expanding the KL term in terms of the score-matching objective as shown in [16]. The integral is evaluated on a diffused trajectory, namely $x_t \sim q(x_t|y)$ for $t \in [0, T]$, which is the forward diffusion process applied to $q(x_0|y)$. As $q(x_0|y)$ admits a Gaussian form, it can be shown that $q(x_t|y)$ is also a Gaussian in the form $q(x_t|y) = \mathcal{N}(\alpha_t \mu, (\alpha_t^2 \sigma^2 + \sigma_t^2) I)$ [17]. As a result, the conditional score function $\nabla_{x_t} \log q(x_t|y)$ can be computed analytically. We refer the reader to [10] for further details.

For MRI reconstruction, RED-diff corresponds to minimizing a measurement consistency loss equipped with a score-matching regularization term. Thus, asssuming the variance of the variational distribution is small near zero (i.e., $\sigma \approx 0$) we can consider the following minimization problem:

$$\min_{\mu} \|A\mu - y\|^2 + \mathbb{E}_{t,\epsilon}[w(t)\|\epsilon_\theta(x_t; t) - \epsilon\|_2^2] \tag{6}$$

where $x_t = \alpha_t \mu + \sigma_t \epsilon$, and $w(t)$ is a time-dependent weighting mechanism. To search for μ, we use first-order stochastic optimization. We define the loss per timestep based on the instantaneous gradient by detaching it at each timestep. Then, we can form the loss at time step t as

$$\|A\mu - y\|^2 + \lambda_t (\text{sg}[\epsilon_\theta(x_t; t) - \epsilon])^T \mu \tag{7}$$

where λ_t is the weighting term, and sg denotes stopped-gradient, indicating that score is not differentiated during the optimization. We set $\lambda_t = \lambda \sigma_t / \alpha_t$, where λ is a hyperparameter. Our full method is described in Algorithm 1. Intuitively, solving the optimization problem in Equation 6 will find an image μ that reconstructs the observation y given the forward model A, while having a high likelihood under the prior as imposed by the regularization. A small regularization term implies that either the diffusion reaches the fixed point, $\epsilon_\theta(x_t; t) = \epsilon$, or the residual only contains noise with no contribution left from the image.

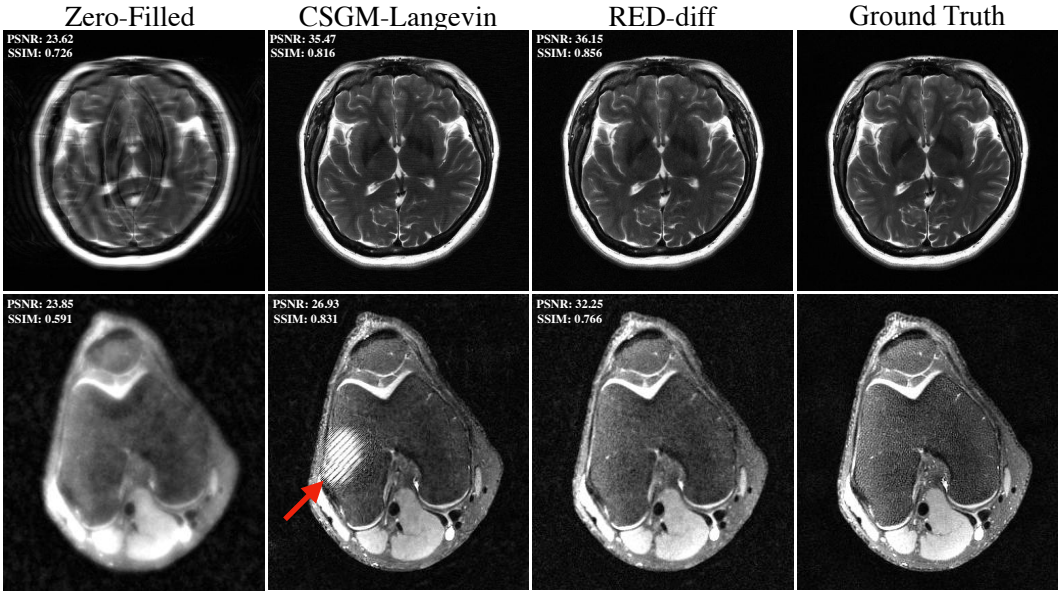

Figure 1: Example reconstruction for brain at $R = 4$, and knee at $R = 12$.

4 Results

We perform experiments with PyTorch on a NVIDIA Tesla V100 GPU [18]. We use the multi-coil fastMRI brain dataset [19] with 1D equispaced undersampling. The matrix size for each scan is 384×384 with 15 coils. Sensitivity maps are estimated using ESPIRiT with a kernel width of 8 and a calibration region of 12×12 [20]. Additionally, we use the fully-sampled 3D fast-spin echo multi-coil knee MRI dataset from [21] with 2D Poisson Disc undersampling mask, as in [7]. Each 3D volume has a matrix size of $320 \times 320 \times 256$ with 8 coils. Sensitivity map estimation is performed in SigPy [22] using JSENSE [23] with kernel width of 8 for each volume. We use 6 validation volumes for fastMRI, and 3 volumes for Mridata by selecting 32 middle slices from each volume. Both datasets have a total of 96 test slices.

For RED-diff, we use linear schedule for $\beta(t)$ from 0.0001 to 0.02, and $T = 1000$. We adopt Adam optimizer with initial learning rate 0.1 and no weight decay regularization, and set the momentum to $(0.9, 0.99)$ where $\lambda = 0.25$. We compare RED-diff with CSGM-Langevin [7]. For CSGM-Langevin and RED-diff, we use the score function from [7] which was trained on a subset of the FastMRI multi-coil brain dataset.

We evaluate the methods in i) the in-distribution setting on brain at $R = 4$, ii) the out-of-distribution setting with knee at $R = \{12, 16\}$. Table 1 shows comparison of reconstruction methods for FastMRI brain, and Mridata knee datasets. RED-diff outperforms CSGM-Langevin in most cases, with a PSNR improvement of $+0.7$dB for brain, $+1.8$dB for knee, and an SSIM improvement of $+0.05$ for brain, while having $3\times$ faster inference time using same amount of memory. Figure 1 shows example reconstructions for brain at $R = 4$, and knee at $R = 12$. RED-diff produces higher quality reconstruction in both cases. Crucially, it is observed that CSGM-Langevin is sensitive in the out-of-distribution setting and produces hallucination artifacts, whereas RED-diff mitigates these artifacts and produces a reconstruction with no hallucinations.

5 Conclusion

In this paper, we presented RED-diff for MRI reconstruction, a variational approach based on KL minimization for sampling from diffusion models that sidesteps posterior approximation. Our experiments on multiple datasets, anatomies and acceleration factors demonstrate that RED-diff improves reconstruction quality and robustness for MRI reconstruction. In addition to improving performance, RED-diff speeds up inference by at least $3\times$ while using the same inference memory.

References

[1] Michael Lustig, David Donoho, and John M Pauly. Sparse mri: The application of compressed sensing for rapid mr imaging. *Magnetic resonance in medicine*, 58(6):1182—1195, December 2007.

[2] Klaas P. Pruessmann, Markus Weiger, Markus B. Scheidegger, and Peter Boesiger. Sense: Sensitivity encoding for fast mri. *Magnetic Resonance in Medicine*, 42(5):952–962, 1999.

[3] Kerstin Hammernik, Teresa Klatzer, Erich Kobler, Michael P Recht, Daniel K Sodickson, Thomas Pock, and Florian Knoll. Learning a variational network for reconstruction of accelerated mri data. *Magnetic resonance in medicine*, 79(6):3055–3071, 2018.

[4] Zalan Fabian, Berk Tinaz, and Mahdi Soltanolkotabi. Humus-net: Hybrid unrolled multi-scale network architecture for accelerated mri reconstruction. *Advances in Neural Information Processing Systems*, 35:25306–25319, 2022.

[5] Batu Ozturkler, Arda Sahiner, Tolga Ergen, Arjun D Desai, Christopher M Sandino, Shreyas Vasanawala, John M Pauly, Morteza Mardani, and Mert Pilanci. Gleam: Greedy learning for large-scale accelerated mri reconstruction. *arXiv preprint arXiv:2207.08393*, 2022.

[6] Hyungjin Chung and Jong Chul Ye. Score-based diffusion models for accelerated MRI. *arXiv*, 2021.

[7] Ajil Jalal, Marius Arvinte, Giannis Daras, Eric Price, Alexandros G Dimakis, and Jonathan I Tamir. Robust compressed sensing mri with deep generative priors. *Advances in Neural Information Processing Systems*, 2021.

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

[9] Batu Ozturkler, Chao Liu, Benjamin Eckart, Morteza Mardani, Jiaming Song, and Jan Kautz. SMRD: Sure-based robust MRI reconstruction with diffusion models. In *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pages 199–209. Springer, 2023.

[10] Morteza Mardani, Jiaming Song, Jan Kautz, and Arash Vahdat. A variational perspective on solving inverse problems with diffusion models. *arXiv preprint arXiv:2305.04391*, 2023.

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

[12] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020.

[13] Pascal Vincent. A connection between score matching and denoising autoencoders. *Neural computation*, 23(7):1661–1674, 2011.

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

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

[16] Yang Song, Conor Durkan, Iain Murray, and Stefano Ermon. Maximum likelihood training of score-based diffusion models. *Advances in Neural Information Processing Systems*, 34:1415–1428, 2021.

[17] Arash Vahdat, Karsten Kreis, and Jan Kautz. Score-based generative modeling in latent space. In *Neural Information Processing Systems (NeurIPS)*, 2021.

[18] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

[19] Jure Zbontar, Florian Knoll, Anuroop Sriram, Tullie Murrell, Zhengnan Huang, Matthew J Muckley, Aaron Defazio, Ruben Stern, Patricia Johnson, Mary Bruno, et al. fastMRI: An open dataset and benchmarks for accelerated mri. *arXiv preprint arXiv:1811.08839*, 2018.

[20] Martin Uecker, Peng Lai, Mark J Murphy, Patrick Virtue, Michael Elad, John M Pauly, Shreyas S Vasanawala, and Michael Lustig. Espirit—an eigenvalue approach to autocalibrating parallel mri: where sense meets grappa. *Magnetic resonance in medicine*, 71(3):990–1001, 2014.

[21] F Ong, S Amin, S Vasanawala, and M Lustig. Mridata. org: An open archive for sharing mri raw data. In *Proc. Intl. Soc. Mag. Reson. Med*, volume 26, 2018.

[22] F Ong and M Lustig. Sigpy: a python package for high performance iterative reconstruction. In *Proceedings of the ISMRM 27th Annual Meeting, Montreal, Quebec, Canada*, volume 4819, 2019.

[23] Leslie Ying and Jinhua Sheng. Joint image reconstruction and sensitivity estimation in sense (jsense). *Magnetic Resonance in Medicine*, 57(6):1196–1202, 2007.