SMRD: SURE-based Robust MRI Reconstruction with Diffusion Models

Batu Ozturkler[1], Chao Liu[2], Benjamin Eckart[2], Morteza Mardani[2], Jiaming Song[2], Jan Kautz[2]

[1] Stanford University, Stanford, CA 94305, USA
`ozt@stanford.edu`
[2] NVIDIA Corporation, Santa Clara, CA 95051, USA
`{chaoliu, beckart, mmardani, jkautz}@nvidia.com`
`jiaming.tsong@gmail.com`

Abstract. Diffusion models have recently gained popularity for accelerated MRI reconstruction due to their high sample quality. They can effectively serve as rich data priors while incorporating the forward model flexibly at inference time, and they have been shown to be more robust than unrolled methods under distribution shifts. However, diffusion models require careful tuning of inference hyperparameters on a validation set and are still sensitive to distribution shifts during testing. To address these challenges, we introduce SURE-based MRI Reconstruction with Diffusion models (SMRD), a method that performs test-time hyperparameter tuning to enhance robustness during testing. SMRD uses Stein's Unbiased Risk Estimator (SURE) to estimate the mean squared error of the reconstruction during testing. SURE is then used to automatically tune the inference hyperparameters and to set an early stopping criterion without the need for validation tuning. To the best of our knowledge, SMRD is the first to incorporate SURE into the sampling stage of diffusion models for automatic hyperparameter selection. SMRD outperforms diffusion model baselines on various measurement noise levels, acceleration factors, and anatomies, achieving a PSNR improvement of up to 6 dB under measurement noise. The code is publicly available at https://github.com/NVlabs/SMRD.

Keywords: MRI Reconstruction · Diffusion models · Measurement Noise · Test-time Tuning.

1 Introduction

Magnetic Resonance Imaging (MRI) is a widely applied imaging technique for medical diagnosis since it is non-invasive and able to generate high-quality clinical images without exposing the subject to radiation. The imaging speed is of vital importance for MRI [21] in that long imaging limits the spatial and temporal resolution and induces reconstruction artifacts such as subject motion during the imaging process. One possible way to speed up the imaging process is to use
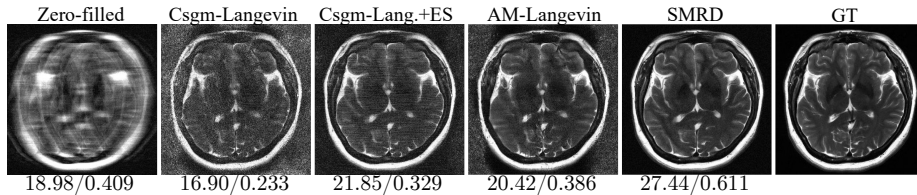
Zero-filled	Csgm-Langevin	Csgm-Lang.+ES	AM-Langevin	SMRD	GT
18.98/0.409	16.90/0.233	21.85/0.329	20.42/0.386	27.44/0.611	

Fig. 1: Reconstruction of a brain slice from the FastMRI dataset with acceleration rate $R = 8$, and measurement noise level $\sigma = 0.005$. SMRD is robust to the measurement noise. Metrics are reported as *PSNR/SSIM*.

multiple receiver coils, and reduce the amount of captured data by subsampling the k-space and exploiting the redundancy in the measurements [29,21].

In recent years, Deep-Learning (DL) based methods have shown great potential as a data-driven approach in achieving faster imaging and better reconstruction quality. DL-based methods can be categorized into two families: unrolled methods that alternate between measurement consistency and a regularization step based on a feed-forward network [41,13,27]; conditional generative models that use measurements as guidance during the generative process [23,16]. Compared with unrolled methods, generative approaches have recently been shown to be more robust when test samples are out of the training distribution due to their stronger ability to learn the data manifold [7,16]. Among generative approaches, diffusion models have recently achieved the state of the art performance [8,7,6,39,9,35,20,26].

However, measurements in MRI are often noisy due to the imaging hardware and thermal fluctuations of the subject [31]. As a result, DL-based methods fail dramatically when a distribution shift due to noise or other scanning parameters occurs during training and testing [3,19]. Although diffusion models were shown to be robust against distribution shifts in noisy inverse problems [32,5] and MRI reconstruction, the hyperparameters that balance the measurement consistency and prior are tuned manually during validation where ground truth is available [16]. These hyperparameters may not generalize well to test settings as shown in Fig.1, and ground truth data may not be available for validation.

In this paper, we propose a framework with diffusion models for MRI reconstruction that is robust to measurement noise and distribution shifts. To achieve robustness, we perform test-time tuning when ground truth data is not available at test time, using Stein's Unbiased Risk Estimator (SURE) [36] as a surrogate loss function for the true mean-squared error (MSE). SURE is used to tune the weight of the balance between measurement consistency and the learned prior for each diffusion step such that the measurement consistency adapts to the measurement noise level during the inference process. SURE is then used to perform early stopping to prevent overfitting to measurement noise at inference.

We evaluate our framework on FastMRI [43] and Mridata [24], and show that it achieves state-of-the-art performance across different noise levels, acceleration rates and anatomies without fine-tuning the pre-trained network or performing

validation tuning, with no access to the target distribution. In summary, our contributions are three-fold:

– We propose a test-time hyperparameter tuning algorithm that boosts the robustness of the pre-trained diffusion models against distribution shifts;
– We propose to use SURE as a surrogate loss function for MSE and incorporate it into the sampling stage for test-time tuning and early stopping without access to ground truth data from the target distribution;
– SMRD achieves state-of-the-art performance across different noise levels, acceleration rates, and anatomies.

2   Related Work

SURE for MRI Reconstruction. SURE has been used for tuning parameters of compressed sensing [15], as well as for unsupervised training of DL-based methods for MRI reconstruction [2]. In addition, two concurrent works propose to incorporate SURE as a loss function for diffusion model training for MRI reconstruction [17,1]. To the best of our knowledge, SURE has not yet been applied to the sampling stage of diffusion models in MRI reconstruction or in another domain.

Adaptation in MRI Reconstruction. Several proposals have been made for adaptation to a target distribution in MRI reconstruction using self-supervised losses for unrolled models [40,10]. Later, [4] proposed single-shot adaptation for test-time tuning of diffusion models by performing grid search over hyperparameters with a ground truth from the target distribution. However, this search is computationally costly, and the assumption of access to samples from the target distribution is limiting for imaging cases where ground truth is not available.

3   Method

3.1   Accelerated MRI Reconstruction using Diffusion Models

The sensing model for accelerated MRI can be expressed as

$$y = \Omega F S x + \nu \tag{1}$$

where y is the measurements in the Fourier domain (k-space), x is the real image, S are coil sensitivity maps, F is the Fourier transform, Ω is the undersampling mask, ν is additive noise, and $A = \Omega F S$ denotes the forward model.

Diffusion models are a recent class of generative models showing remarkable sample fidelity for computer vision tasks [14]. A popular class of diffusion models is score matching with Langevin dynamics [33]. Given $i.i.d.$ training samples from a high-dimensional data distribution (denoted as $p(x)$), diffusion models can estimate the scores of the noise-perturbed data distributions. First, the data distribution is perturbed with Gaussian noise of different intensities with standard deviation β_t for various timesteps t, such that $p_{\beta_t}(\tilde{x}|x) = \mathcal{N}(\tilde{x}|x, \beta_t^2 I)$ [33],

---
Algorithm 1 SMRD: Test-time Tuning for Diffusion Models via SURE

Input: measurement y, forward model A, initial λ_0, window size w, learning rate α
1: sample $x_0 \sim \mathcal{N}(0, I)$
2: **for** $t \in 0, ..., T-1$ **do**
3:     $x_t^+ = x_t + \eta_t f(x_t; t) + \sqrt{2\eta_t}\zeta_t$
4:     $x_{t+1} = h(x_t, \lambda_t) = (A^H A + \lambda_t I)^{-1}(x_{zf} + \lambda_t x_t^+)$
5:     sample $\mu \sim \mathcal{N}(0, I)$
6:     $\text{SURE}(t) = \|h(x_t, \lambda_t) - x_{\text{zf}}\|^2 \mu^T (h(x_t + \epsilon\mu, \lambda_t) - h(x_t, \lambda_t))/N\epsilon$
7:     $\lambda_{t+1} = \lambda_t - \alpha\nabla_{\lambda_t}\text{SURE}(t)$
8:     **if** $mean(\text{SURE}[t-w:t]) > mean(\text{SURE}[t-2w:t-w])$ **then**
9:         **break**
10: **return** x_{t+1}

---

leading to a series of perturbed data distributions $p(x_t)$. Then, the score function $\nabla_{x_t} \log p(x_t)$ can be estimated by training a joint neural network, denoted as $f(x_t; t)$, via denoising score matching [38]. After training the score function, annealed Langevin Dynamics can be used to generate new samples [34]. Starting from a noise distribution $x_0 \sim \mathcal{N}(0, I)$, annealed Langevin Dynamics is run for T steps

$$x_{t+1} = x_t + \eta_t \nabla_{x_t} \log p(x_t) + \sqrt{2\eta_t}\zeta_t \tag{2}$$

where η_t is a sampling hyperparameter, and $\zeta_t \sim \mathcal{N}(0, I)$. For MRI reconstruction, measurement consistency can be incorporated via sampling from the posterior distribution $p(x_t|y)$ [16]:

$$x_{t+1} = x_t + \eta_t \nabla_{x_t} \log p(x_t|y) + \sqrt{2\eta_t}\zeta_t \tag{3}$$

The form of $\nabla_{x_t} \log p(x_t|y)$ depends on the specific inference algorithm [8,7].

3.2   Stein's Unbiased Risk Estimator (SURE)

SURE is a statistical technique which serves as a surrogate for the true mean squared error (MSE) when the ground truth is unknown. Given the ground truth image x, the zero-filled image can be formulated as $x_{\text{zf}} = x + z$, where z is the noise due to undersampling. Then, SURE is an unbiased estimator of $\text{MSE} = \|\hat{x} - x\|_2^2$ [36] and can be calculated as:

$$\text{SURE} = \|\hat{x} - x_{\text{zf}}\|_2^2 - N\sigma^2 + \sigma^2 \operatorname{tr}\left(\frac{\partial\hat{x}}{\partial x_{\text{zf}}}\right) \tag{4}$$

where x_{zf} is the input of a denoiser, $\hat{x}$ is the prediction of a denoiser, N is the dimensionality of $\hat{x}$. In practical applications, the noise variance σ^2 is not known a priori. In such cases, it can be assumed that the reconstruction error is not large, and the sample variance between the zero-filled image and the reconstruction can be used to estimate the noise variance, where $\sigma^2 \approx \|\hat{x} - x_{\text{zf}}\|_2^2/N$ [12]. Then, SURE can be rewritten as:

$$\text{SURE} \approx \frac{\|\hat{x} - x_{\text{zf}}\|^2}{N} \operatorname{tr}\left(\frac{\partial\hat{x}}{\partial x_{\text{zf}}}\right) \tag{5}$$

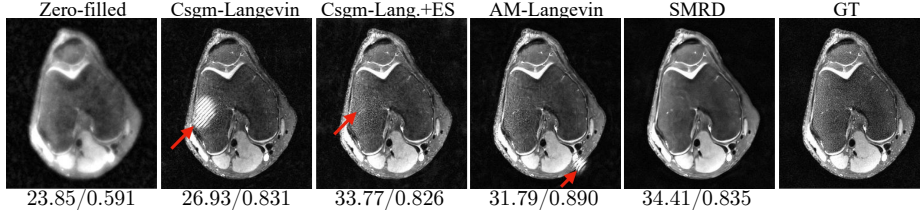| Zero-filled	Csgm-Langevin	Csgm-Lang.+ES	AM-Langevin	SMRD	GT
23.85/0.591	26.93/0.831	33.77/0.826	31.79/0.890	34.41/0.835	

Fig. 2: Knee reconstruction with $R = 12$. Metrics are reported as *PSNR/SSIM*.

A key assumption behind SURE is that the noise process that relates the zero-filled image to the ground truth is i.i.d. normal, namely $z \sim \mathcal{N}(0, \sigma^2 I)$. However, this assumption does not always hold in the case of MRI reconstruction due to undersampling in k-space that leads to structured aliasing. In this case, density compensation can be applied to enforce zero-mean residuals and increase residual normality [12].

3.3   SURE-based MRI Reconstruction with Diffusion Models

Having access to SURE at test time enables us to monitor it as a proxy for the true MSE. Our goal is to optimize inference hyperparameters in order to minimize SURE. To do this, we first consider the following optimization problem at time-step t:

$$x_{t+1} = \arg\min_x \|Ax - y\|_2^2 + \lambda_t \|x - x_t\|_2^2 \tag{6}$$

where we introduce a time-dependent regularization parameter λ_t. The problem in Eq. 6 can be solved using alternating minimization (which we call AM-Langevin):

$$x_t^+ = x_t + \eta_t f(x_t; t) + \sqrt{2\eta_t}\zeta_t \tag{7}$$

$$x_{t+1} = \arg\min_x \|Ax - y\|_2^2 + \lambda_t \|x - x_t^+\|_2^2 \tag{8}$$

Eq. 8 can be solved using the conjugate gradient (CG) algorithm, with the iterates

$$x_{t+1} = h(x_t, \lambda_t) = (A^H A + \lambda_t I)^{-1}(x_{\text{zf}} + \lambda_t x_t^+) \tag{9}$$

where A^H is the Hermitian transpose of A, $x_{\text{zf}} = A^H y$ is the zero-filled image, and h denotes the full update including the Langevin Dynamics and CG. This allows us to explicitly control the balance between the prior through the score function and the regularization through λ_t.

Monte-Carlo SURE. Calculating SURE requires evaluating the trace of the Jacobian $\text{tr}(\frac{\partial x_{t+1}}{\partial x_{\text{zf}}})$, which can be computationally intensive. Thus, we approximate this term using Monte-Carlo SURE [30]. Given an N-dimensional noise vector μ from $\mathcal{N}(0, I)$ and the perturbation scale ϵ, the approximation is:

$$\text{tr}(\frac{\partial x_{t+1}}{\partial x_{\text{zf}}}) \approx \mu^T (h(x_t + \epsilon\mu, \lambda_t) - h(x_t, \lambda_t))/\epsilon \tag{10}$$

Table 1: FastMRI brain dataset results. Metrics are reported as *PSNR/SSIM*.

R	$R = 4$			$R = 8$		
$\sigma \; (\times 10^{-3})$	$\sigma = 0$	$\sigma = 2.5$	$\sigma = 5$	$\sigma = 0$	$\sigma = 2.5$	$\sigma = 5$
Zero-filled	27.8/0.81	27.1/0.63	25.3/0.43	23.2/0.69	23.1/0.61	22.7/0.43
Csgm-Langevin	36.3/0.78	25.1/0.36	18.5/0.16	**34.7**/0.79	21.3/0.32	15.8/0.15
Csgm-Lang.+ES	35.9/0.79	26.1/0.40	20.8/0.23	32.3/0.69	26.1/0.41	20.7/0.25
AM-Langevin	**39.7/0.95**	28.7/0.53	22.4/0.29	**34.7/0.93**	25.7/0.52	19.0/0.30
SMRD	36.5/0.89	**33.5/0.78**	**29.4/0.59**	32.4/0.80	**31.4/0.75**	**28.6/0.61**

This approximation is typically quite tight for some small value ϵ; see e.g., [12]. Then, at time step t, SURE is given as

$$\text{SURE}(t) = \frac{\|h(x_t, \lambda_t) - x_{\text{zf}}\|^2}{N\epsilon} \mu^T (h(x_t + \epsilon\mu, \lambda_t) - h(x_t, \lambda_t)) \qquad (11)$$

where $h(x_t, \lambda_t)$ is the prediction which depends on the input x_{zf}, shown in Eq. 9.

Tuning λ_t. By allowing λ_t to be a learnable, time-dependent variable, we can perform test-time tuning (TTT) for λ_t by updating it in the direction that minimizes SURE. As both $\text{SURE}(t)$ and λ_t are time-dependent, the gradients can be calculated with backpropagation through time (BPTT). In SMRD, for the sake of computation, we apply truncated BPTT, and only consider gradients from the current time step t. Then, the λ_t update rule is:

$$\lambda_{t+1} = \lambda_t - \alpha\nabla_{\lambda_t}\text{SURE}(t) \qquad (12)$$

where α is the learning rate for λ_t.

Early Stopping (ES). Under measurement noise, it is critical to prevent overfitting to the measurements. We employ early-stopping (ES) by monitoring the moving average of SURE loss with a window size w at test time. Intuitively, we perform early stopping when the SURE loss does not decrease over a certain window. We denote the early-stopping iteration as T_{ES}. Our full method is shown in Alg. 1.

4   Experiments

Experiments were performed with PyTorch on a NVIDIA Tesla V100 GPU [28]. For baselines and SMRD, we use the score function from [16] which was trained on a subset of the FastMRI multi-coil brain dataset. We refer the reader to [16] for implementation details regarding the score function. For all AM-Langevin variants, we use 5 CG steps. In SMRD, for tuning λ_t, we use the Adam optimizer [18] with a learning rate of $\alpha = 0.2$ and $\lambda_0 = 2$. In the interest of inference speed, we fixed λ_t after $t = 500$, as convergence was observed in earlier iterations. For SURE early stopping, we use window size $w = 160$. For evaluation,

Table 2: Cross-dataset results with the Mridata knee dataset. Metrics are reported as *PSNR/SSIM*.

R	$R = 12$			$R = 16$		
$\sigma\ (\times 10^{-3})$	$\sigma = 0$	$\sigma = 2.5$	$\sigma = 5$	$\sigma = 0$	$\sigma = 2.5$	$\sigma = 5$
Zero-filled	24.5/0.63	24.5/0.61	24.1/0.54	24.0/0.60	24.1/0.59	23.9/0.55
Csgm-Langevin	31.4/0.82	24.2/0.49	21.7/0.29	31.8/0.79	24.4/0.50	22.1/0.32
Csgm-Lang.+ES	34.0/0.82	28.8/0.60	23.6/0.37	33.6/0.81	28.4/0.60	23.2/0.37
AM-Langevin	34.0/**0.87**	29.9/0.69	23.5/0.38	**34.3**/**0.85**	29.3/0.66	22.7/0.37
SMRD	**35.0**/0.84	**33.5**/**0.82**	**29.4**/**0.67**	34.1/0.82	**32.8**/**0.80**	**29.2**/**0.67**

Table 3: Ablation study for different components of our method on the Mridata multi-coil knee dataset where $R = 12$. Metrics are reported as *PSNR/SSIM*.

$\sigma\ (\times 10^{-3})$	$\sigma = 0$	$\sigma = 2.5$	$\sigma = 5$	$\sigma = 7.5$
AM-Langevin	34.0/**0.87**	29.9/0.69	23.5/0.38	20.5/0.24
AM-Lang.+TTT	34.7/0.86	31.0/0.75	25.1/0.45	21.5/0.29
AM-Lang.+ES	**35.3**/0.85	32.9/0.80	28.3/0.61	25.2/0.45
AM-Lang.+TTT+ES (SMRD)	35.0/0.84	**33.5**/**0.82**	**29.4**/**0.67**	**26.3**/**0.52**

we used the multi-coil fastMRI brain dataset [43] and the fully-sampled 3D fast-spin echo multi-coil knee MRI dataset from mridata.org [24] with 1D equispaced undersampling and a 2D Poisson Disc undersampling mask respectively, as in [16]. We used 6 volumes from the validation split for fastMRI, and 3 volumes for Mridata where we selected 32 middle slices from each volume so that both datasets had 96 test slices in total.

Noise Simulation. The noise source in MRI acquisition is modeled as additive complex-valued Gaussian noise added to each acquired k-space sample [22]. To simulate measurement noise, we add masked complex-gaussian noise to the masked k-space $\nu \sim \mathcal{N}(0, \sigma)$ with standard deviation σ, similar to [11].

5   Results and Discussion

We compare with three baselines: (1) Csgm-Langevin [16], using the default hyperparameters that were tuned on two validation brain scans at $R = 4$; (2) Csgm-Langevin with early stopping using SURE loss with window size $w = 25$; (3) AM-Langevin where $\lambda_t = \lambda_0$ and fixed throughout the inference. For tuning AM-Langevin, we used a brain scan for validation at $R = 4$ in the noiseless case ($\sigma = 0$) similar to Csgm-Langevin, and the optimal value was $\lambda_0 = 2$.

We evaluate the methods across different measurement noise levels, acceleration rates and anatomies using the same pretrained score function from [16]. Table 5 shows a comparison of reconstruction methods applied to the FastMRI brain dataset where $R = \{4, 8\}$, and $\sigma = \{0, 0.0025, 0.005\}$. SMRD performs best except for $\sigma = 0$, which is the training setting for the score function. Example

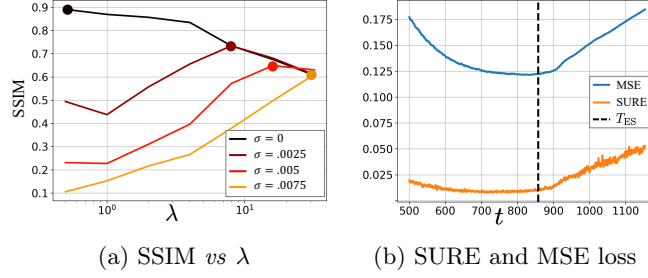(a) SSIM *vs* λ (b) SURE and MSE loss

Fig. 3: (a) SSIM on a validation scan (knee) with varying noise levels σ and regularization parameter λ parameter choices. As σ increases, the optimal λ value shifts to the right. (b) SURE and true MSE over sampling iterations. MSE and SURE start increasing at similar T due to overfitting to measurement noise.
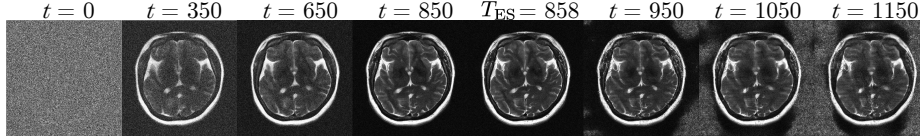

Fig. 4: Iterations of inference for an example brain slice where $R = 8$, $\sigma = 0.0075$.

reconstructions are shown in Fig.1 where $R = 8$, $\sigma = 0.005$. Under measurement noise, reconstructions of baseline methods diverge, where only SMRD can produce an acceptable reconstruction. Table 2 shows a comparison of reconstruction methods in the cross-dataset setup with the Mridata knee dataset where $R = \{12, 16\}$. SMRD outperformed baselines across every R and σ, and is on par with baselines when $\sigma = 0$. Fig.2 shows example reconstructions where $R = 12$, $\sigma = 0$. Hallucination artifacts are visible in baselines even with no added measurement noise, whereas SMRD mitigates these artifacts and produces a reconstruction with no hallucinations. Fig.3a shows SSIM on a knee validation scan with varying noise levels σ and λ. As σ increases, the optimal λ value increases as well. Thus, hyperparameters tuned with $\sigma = 0$ do not generalize well under measurement noise change, illustrating the need for test-time tuning under distribution shift. Fig.3b shows true MSE vs SURE for an example brain slice where $\sigma = 0.0075$, $R = 8$. SURE accurately estimates MSE, and the increase in loss occurs at similar iterations, enabling us to perform early stopping before true MSE increases. The evolution of images across iterations for this sample is shown in Fig.4. SMRD accurately captures the correct early stopping point. As a result of early stopping, SMRD mitigates artifacts, and produces a smoother reconstruction. Table 3 shows the ablation study for different components of SMRD. TTT and ES both improve over AM-Langevin, where SMRD works best for all σ while being on par with others on $\sigma = 0$.

6   Conclusion

We presented SMRD, a SURE-based TTT method for diffusion models in MRI reconstruction. SMRD does not require ground truth data from the target distribution for tuning as it uses SURE for estimating the true MSE. SMRD surpassed baselines across different shifts including anatomy shift, measurement noise change, and acceleration rate change. SMRD could be helpful to improve the safety and robustness of diffusion models for MRI reconstruction used in clinical settings. While we applied SMRD to MRI reconstruction, future work could explore the application of SMRD to other inverse problems and diffusion sampling algorithms and can be used to tune their inference hyperparameters.

References

1. Aali, A., Arvinte, M., Kumar, S., Tamir, J.I.: Solving inverse problems with score-based generative priors learned from noisy data. arXiv preprint arXiv:2305.01166 (2023)
2. Aggarwal, H.K., Pramanik, A., John, M., Jacob, M.: Ensure: A general approach for unsupervised training of deep image reconstruction algorithms. IEEE Transactions on Medical Imaging (2022)
3. Antun, V., Renna, F., Poon, C., Adcock, B., Hansen, A.C.: On instabilities of deep learning in image reconstruction and the potential costs of ai. Proceedings of the National Academy of Sciences **117**(48), 30088–30095 (2020)
4. Arvinte, M., Jalal, A., Daras, G., Price, E., Dimakis, A., Tamir, J.I.: Single-shot adaptation using score-based models for mri reconstruction. In: International Society for Magnetic Resonance in Medicine, Annual Meeting (2022)
5. Chung, H., Kim, J., Mccann, M.T., Klasky, M.L., Ye, J.C.: Diffusion posterior sampling for general noisy inverse problems. In: The Eleventh International Conference on Learning Representations (2023)
6. Chung, H., Ryu, D., McCann, M.T., Klasky, M.L., Ye, J.C.: Solving 3d inverse problems using pre-trained 2d diffusion models. arXiv preprint arXiv:2211.10655 (2022)
7. Chung, H., Sim, B., Ryu, D., Ye, J.C.: Improving diffusion models for inverse problems using manifold constraints. arXiv preprint arXiv:2206.00941 (2022)
8. Chung, H., Ye, J.C.: Score-based diffusion models for accelerated MRI. arXiv (2021)
9. Dar, S.U., Öztürk, Ş., Korkmaz, Y., Elmas, G., Özbey, M., Güngör, A., Çukur, T.: Adaptive diffusion priors for accelerated mri reconstruction. arXiv preprint arXiv:2207.05876 (2022)
10. Darestani, M.Z., Liu, J., Heckel, R.: Test-time training can close the natural distribution shift performance gap in deep learning based compressed sensing. In: International Conference on Machine Learning. pp. 4754–4776. PMLR (2022)
11. Desai, A.D., Ozturkler, B.M., Sandino, C.M., Boutin, R., Willis, M., Vasanawala, S., Hargreaves, B.A., Ré, C., Pauly, J.M., Chaudhari, A.S.: Noise2recon: Enabling snr-robust mri reconstruction with semi-supervised and self-supervised learning. Magnetic Resonance in Medicine (2023)
12. Edupuganti, V., Mardani, M., Vasanawala, S., Pauly, J.: Uncertainty quantification in deep mri reconstruction. IEEE Transactions on Medical Imaging **40**(1), 239–250 (2020)
13. Hammernik, K., Klatzer, T., Kobler, E., Recht, M.P., Sodickson, D.K., Pock, T., Knoll, F.: Learning a variational network for reconstruction of accelerated mri data. Magnetic resonance in medicine **79**(6), 3055–3071 (2018)
14. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems **33**, 6840–6851 (2020)
15. Iyer, S., Ong, F., Setsompop, K., Doneva, M., Lustig, M.: Sure-based automatic parameter selection for espirit calibration. Magnetic Resonance in Medicine **84**(6), 3423–3437 (2020)
16. Jalal, A., Arvinte, M., Daras, G., Price, E., Dimakis, A.G., Tamir, J.I.: Robust compressed sensing mri with deep generative priors. Advances in Neural Information Processing Systems (2021)
17. Kawar, B., Elata, N., Michaeli, T., Elad, M.: Gsure-based diffusion model training with corrupted data. arXiv preprint arXiv:2305.13128 (2023)

18. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization (2017)
19. Knoll, F., Hammernik, K., Kobler, E., Pock, T., Recht, M.P., Sodickson, D.K.: Assessment of the generalization of learned image reconstruction and the potential for transfer learning. Magnetic Resonance in Medicine **81**(1), 116–128 (2019)
20. Levac, B., Jalal, A., Tamir, J.I.: Accelerated motion correction for mri using score-based generative models. arXiv preprint arXiv:2211.00199 (2022)
21. Lustig, M., Donoho, D., Pauly, J.M.: Sparse mri: The application of compressed sensing for rapid mr imaging. Magnetic resonance in medicine **58**(6), 1182—1195 (December 2007)
22. Macovski, A.: Noise in mri. Magnetic resonance in medicine **36**(3), 494–497 (1996)
23. Mardani, M., Gong, E., Cheng, J.Y., Vasanawala, S., Zaharchuk, G., Alley, M., Thakur, N., Han, S., Dally, W., Pauly, J.M., et al.: Deep generative adversarial networks for compressed sensing automates mri. arXiv preprint arXiv:1706.00051 (2017)
24. Ong, F., Amin, S., Vasanawala, S., Lustig, M.: Mridata. org: An open archive for sharing mri raw data. In: Proc. Intl. Soc. Mag. Reson. Med. vol. 26 (2018)
25. Ong, F., Lustig, M.: Sigpy: a python package for high performance iterative reconstruction. In: Proceedings of the ISMRM 27th Annual Meeting, Montreal, Quebec, Canada. vol. 4819 (2019)
26. Ozturkler, B., Mardani, M., Vahdat, A., Kautz, J., Pauly, J.M.: Regularization by denoising diffusion process for MRI reconstruction. In: Medical Imaging with Deep Learning, short paper track (2023), https://openreview.net/forum?id=M1V498MXelq
27. Ozturkler, B., Sahiner, A., Ergen, T., Desai, A.D., Sandino, C.M., Vasanawala, S., Pauly, J.M., Mardani, M., Pilanci, M.: Gleam: Greedy learning for large-scale accelerated mri reconstruction. arXiv preprint arXiv:2207.08393 (2022)
28. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al.: Pytorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems **32** (2019)
29. Pruessmann, K.P., Weiger, M., Scheidegger, M.B., Boesiger, P.: Sense: Sensitivity encoding for fast mri. Magnetic Resonance in Medicine **42**(5), 952–962 (1999)
30. Ramani, S., Blu, T., Unser, M.: Monte-carlo sure: A black-box optimization of regularization parameters for general denoising algorithms. IEEE Transactions on Image Processing **17**(9), 1540–1554 (2008)
31. Redpath, T.W.: Signal-to-noise ratio in mri. The British Journal of Radiology **71**(847), 704–707 (1998)
32. 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
33. Song, Y., Ermon, S.: Generative modeling by estimating gradients of the data distribution. Advances in neural information processing systems **32** (2019)
34. Song, Y., Ermon, S.: Improved techniques for training score-based generative models. Advances in neural information processing systems **33**, 12438–12448 (2020)
35. Song, Y., Shen, L., Xing, L., Ermon, S.: Solving inverse problems in medical imaging with score-based generative models. In: International Conference on Learning Representations (2022), https://openreview.net/forum?id=vaRCHVjOuGI
36. Stein, C.M.: Estimation of the mean of a multivariate normal distribution. Annals of Statistics **9**, 1135–1151 (1981)

37. 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)
38. Vincent, P.: A connection between score matching and denoising autoencoders. Neural computation **23**(7), 1661–1674 (2011)
39. Xie, Y., Li, Q.: Measurement-conditioned denoising diffusion probabilistic model for under-sampled medical image reconstruction. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part VI. pp. 655–664. Springer (2022)
40. Yaman, B., Hosseini, S.A.H., Akcakaya, M.: Zero-shot self-supervised learning for MRI reconstruction. In: International Conference on Learning Representations (2022), https://openreview.net/forum?id=085y6YPaYjP
41. Yang, Y., Sun, J., Li, H., Xu, Z.: Deep admm-net for compressive sensing mri. In: Lee, D., Sugiyama, M., Luxburg, U., Guyon, I., Garnett, R. (eds.) Advances in Neural Information Processing Systems. vol. 29. Curran Associates, Inc. (2016)
42. Ying, L., Sheng, J.: Joint image reconstruction and sensitivity estimation in sense (jsense). Magnetic Resonance in Medicine **57**(6), 1196–1202 (2007)
43. 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)

7   Supplementary Material

Mridata Multi-Coil Knee Dataset. We used the fully-sampled 3D fast-spin echo (FSE) multi-coil knee MRI dataset publicly available on mridata.org [24]. Each 3D volume had a matrix size of $320 \times 320 \times 256$ with 8 coils. Sensitivity map estimation was performed in SigPy [25] using JSENSE [42] with kernel width of 8 for each volume. Fully-sampled references were retrospectively undersampled using a 2D Poisson Disc undersampling mask at $R = \{12, 16\}$, where R denotes acceleration factor. For evaluation, we used 3 volumes, and selected 32 middle slices from each volume for a total of 96 test slices.

FastMRI Brain Multi-Coil Dataset. We used the multi-coil brain MRI dataset from fastMRI [43]. The matrix size for each scan was 384×384 with 15 coils. Sensitivity maps were estimated using ESPIRiT with a kernel width of 8 and a calibration region of 12×12 [37]. At acceleration factor $R = 8$, this calibration region is 4% of the autocalibration region. Fully-sampled references were retrospectively undersampled using 1D equispaced undersampling, as in [16]. For evaluation, we select 6 volumes from the validation split, and evaluate all methods on 96 slices.

Anatomy	Brain		Knee	
R	$R = 4$	$R = 8$	$R = 12$	$R = 16$
Zero-filled	23.4/0.31	22.0/0.32	23.8/0.48	23.4/0.48
Csgm-Langevin	16.0/0.08	14.8/0.09	18.7/0.16	19.6/0.20
Csgm-Langevin+ES	18.0/0.16	17.7/0.17	20.5/0.25	20.6/0.26
AM-Langevin	18.7/0.18	15.9/0.19	20.5/0.24	19.2/0.23
SMRD	**26.3/0.44**	**26.4/0.49**	**26.3/0.52**	**26.1/0.52**

Table 4: Reconstruction methods applied to the FastMRI multi-coil brain dataset and the Mridata multi-coil knee dataset at $\sigma = 0.0075$. Metrics are reported as *PSNR/SSIM*. In high noise regimes, SMRD outperforms all baselines.

Inference Metrics	Time per iteration (sec/iter)	Number of Iterations	Total Time	Memory (GB)
Csgm-Langevin	0.34	1155	6 min. 38 sec.	5.3
Csgm-Langevin+ES	0.69	974	11 min. 16 sec.	5.3
AM-Langevin	0.80	1155	15 min. 26 sec.	5.3
SMRD	1.14	902	14 min. 54 sec.	5.3

Table 5: Inference time and memory for different reconstruction methods.

| | | | | $T_{ES} = 886$ | | | |
| $t=0$ | $t=350$ | $t=650$ | $t=850$ | | $t=950$ | $t=1050$ | $t=1150$ |

11.65/0.011 14.74/0.063 21.28/0.275 26.67/0.552 27.44/0.611 27.81/0.667 24.65/0.547 22.40/0.470

| $t=0$ | $t=350$ | $t=650$ | $t=850$ | $T_{ES}=902$ | $t=950$ | $t=1050$ | $t=1150$ |

11.58/0.010 14.31/0.067 21.73/0.290 26.96/0.520 27.15/0.545 25.27/0.489 22.57/0.391 21.29/0.326
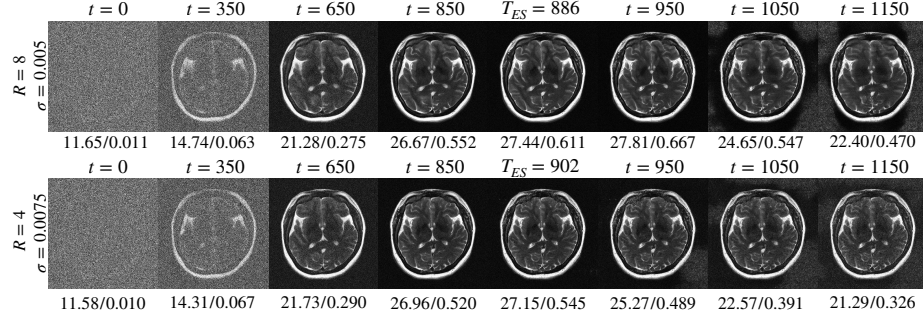
Fig. 5: Iterations of inference for an example brain slice where $R=8$, $\sigma=0.005$ (top), and $R=4$, $\sigma=0.0075$ (bottom). Metrics are reported as *PSNR/SSIM*. The time step t where divergence is observed for the inference process changes for different R and σ. As a result, T_{ES} changes.

Inference Metrics	Total Time	Number of Iterations
$\sigma=0$	16 min. 47 sec.	1007
$\sigma=0.0025$	16 min. 22 sec.	985
$\sigma=0.005$	15 min. 38 sec.	942
$\sigma=0.0075$	14 min. 54 sec	902

Table 6: Inference time for SMRD at different σ. For different noise levels, T_{ES} for SMRD changes, reducing the total reconstruction time under high noise regimes.
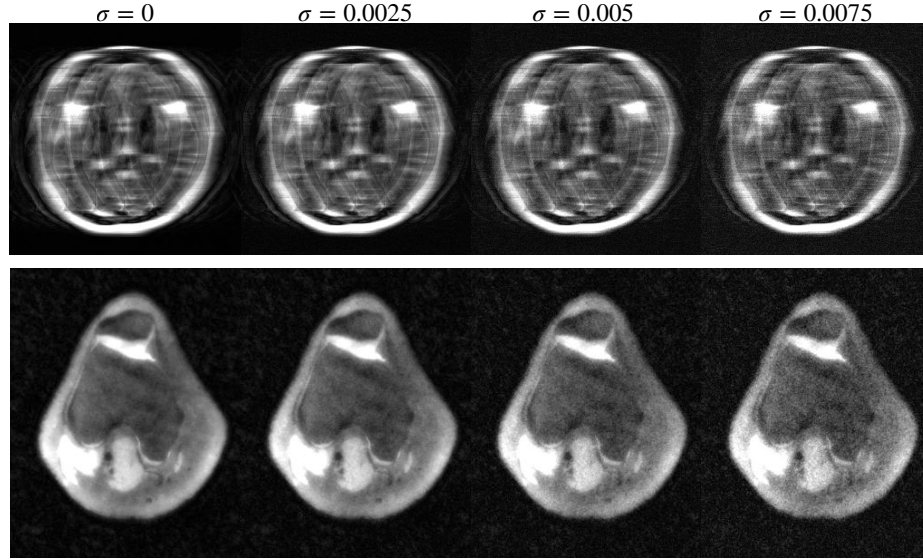

Fig. 6: Example zero-filled images for different measurement noise levels for brain at $R=8$ (top) and for knee at $R=12$ (bottom).