HIGH-RESOLUTION QUANTITATIVE MRI USING CONTRAST-DOMAIN MULTISCALE ENERGY (MuSE) REGULARIZATION

Rushdi Rusho[1]*, Jyothi Rikhab Chand*[1]*, Steve R. Kecskemeti*[2]*, James H. Holmes*[3]*, Mathews Jacob*[1]

[1]University of Virginia, Charlottesville, VA, USA
[2]University of Wisconsin–Madison, Madison, WI, USA
[3]University of Iowa, Iowa City, IA, USA

arXiv:2611.05132v1 [eess.IV] 3 Nov 2026

ABSTRACT

Quantitative MRI methods can offer tissue parameter maps (e.g. T_1) that are comparable across scanners, sessions and subjects, together with multiple contrast-weighted images from a single scan. However, most data are acquired at 1 mm, which limits their use in neurodegenerative diseases that affect fine brain structures. We introduce, to our knowledge, the first diffusion-model-based reconstruction for quantitative MRI at 0.5 mm isotropic resolution from scans of less than 5 minutes, which has eightfold smaller voxel volume than clinical protocols. The key idea is to apply the prior to the synthesized contrast images rather than to the subspace coefficients. The prior is then independent of the temporal basis and the sequence, which improves generalization and allows a 2-D multiscale energy (MuSE) model trained on grayscale CelebA face images to be used without any MRI training data. The explicit energy makes reconstruction MAP estimation by a step-size-free majorize–minimize algorithm that regularizes a rotating subset of contrasts along different axes, which makes 3-D multi-contrast reconstruction tractable. For 3-D radial MPnRAGE at 7 T in 4.3 min, MuSE reduced the phantom NRMSE from 0.133 (CG-SENSE) and 0.091 (L1-wavelet) to 0.074 and the WM T_1 interquartile range from 117 and 47 ms to 38 ms, and it was demonstrated in vivo at 7 T.

***Index Terms*—** Quantitative MRI, MPnRAGE, majorize–minimize, MAP estimation, energy-based models, generic image prior

1. INTRODUCTION

Quantitative MRI estimates tissue parameters such as T_1 that are more readily comparable across scanners, sessions and subjects. Methods such as MPnRAGE fit a signal model to images acquired during magnetization recovery [1]. In neurodegenerative diseases, early changes affect thin structures, such as the cortical ribbon and the hippocampal subfields, whose dimensions approach 1 mm [2]. However, a 0.5 mm isotropic voxel has one eighth of the volume of a 1 mm voxel; because the signal-to-noise ratio scales with voxel volume and the square root of scan time, preserving it would require 64 times the scan time.

Subspace models are the most widely used approach to accelerate quantitative MRI [3]. They constrain the time profile of each voxel to a few principal components of Bloch-simulated signal evolutions, which leaves a few spatial factors to be estimated. These spatial factors can be further regularized by wavelet sparsity [4] or locally low-rank (LLR) penalties [5]. These methods degrade at high acceleration and low SNR (Sec. 2.3). Diffusion models, which learn multiscale denoisers from images, are now emerging as powerful priors for MRI reconstruction [6]. Most perform posterior sampling, which typically needs many network evaluations; extending them to 3-D quantitative imaging with hundreds of contrast volumes is challenging. In addition, fully sampled training data are impractical to acquire at this resolution.

We introduce, to our knowledge, the first diffusion-model-based reconstruction for quantitative MRI at 0.5 mm resolution. It rests on three design choices. First, we use multiscale energy (MuSE) models [7, 8], diffusion models whose denoiser is the gradient of an explicit energy. The explicit energy poses reconstruction as maximum a posteriori (MAP) estimation, which we solve by majorize–minimize (MM) with quadratic, step-size-free updates. This avoids the posterior sampling used by approaches such as DPS and DAPS [9, 10]. Second, to minimize the risk of overfitting limited, basis-specific training data, we apply the prior to the contrast images synthesized from the subspace model [11] rather than to the spatial factors. The prior is then independent of the basis and the sequence, so a generic 2-D model learned from grayscale face images (CelebA) can be used without MRI training data; such generic priors were recently shown to generalize well across imaging domains [12]. Third, we apply this 2-D prior to slices along all three orientations [13], and each stochastic MM iteration regularizes a rotating subset of contrasts along different axes, which makes hundreds of 3-D contrasts tractable.

We evaluate MuSE on 3-D radial MPnRAGE data at 7 T (8.6 min, and a 2× subset with 4.3 min) against CG-SENSE, L1-wavelet and LLR, on a phantom and in vivo.

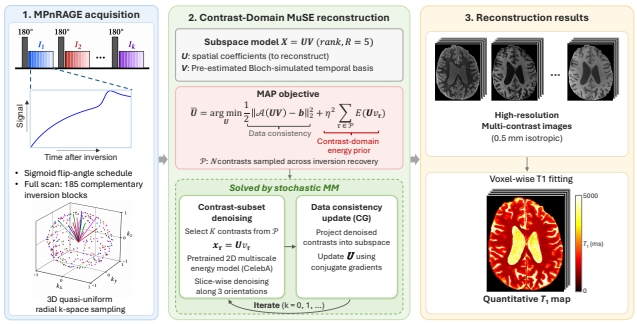


Fig. 1. Overview of MuSE for 0.5 mm quantitative MRI. (1) 3-D radial MPnRAGE acquisition. (2) Each MM iteration denoises a subset of synthesized contrasts slice by slice with a 2-D MuSE model trained only on grayscale CelebA faces, then enforces data consistency by CG in the rank-five subspace. (3) Voxel-wise T_1 fitting at 0.5 mm.

2. BACKGROUND

2.1. MPnRAGE acquisition

We use an MPnRAGE [1] sequence, on a 7 T scanner, which repeats an inversion block (Fig. 1, step 1). Each block consists of an inversion pulse, 461 3-D radial spokes (TR 4.94 ms, 2.3 s; the first is discarded) and a 499-ms recovery delay, about 2.8 s in total. The flip angle rises from 4° to 8° along a sigmoid. Spokes with the same index in different blocks share the same time since inversion, so each of the $T = 460$ spoke indices defines one contrast τ, sampled by the spokes of all blocks. The 185 blocks take 8.6 min; the last 92 blocks give the $2\times$ scan (4.3 min).

2.2. Subspace model

The contrast images $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_T] \in \mathbb{C}^{M\times T}$ follow $\mathbf{x}_\tau(m) = \rho_m\, s_\tau(\vartheta_m)$, where ρ_m is the proton density and $s_\tau(\vartheta_m)$ the magnetization at time τ. In this work, ϑ_m is a function of T_1, B_1^+ and the inversion efficiency (IE) [1]. The measurements are $\mathbf{b}_\tau = \mathcal{A}_\tau \mathbf{x}_\tau + \mathbf{n}_\tau$, where $\mathcal{A}_\tau$ includes the coil sensitivities, the trajectory and rigid motion, and $\mathbf{n}_\tau$ is Gaussian noise of variance ν^2. Subspace methods write $\mathbf{X} = \mathbf{U}\mathbf{V}$ [3], where $\mathbf{V} \in \mathbb{R}^{R\times T}$ has orthonormal rows estimated from Bloch simulations and $\mathbf{U} \in \mathbb{C}^{M\times R}$ holds the spatial coefficients, so that $\mathbf{x}_\tau = \mathbf{U}\mathbf{v}_\tau$.

2.3. Classical regularizers

L1-wavelet regularization penalizes the ℓ_1 norm of the wavelet detail coefficients of the columns of $\mathbf{U}$ [4]. LLR soft-thresholds the singular values of the $P \times R$ matrix of coefficients in each block of P voxels [5]. If the block has local rank $k \le R$, the rank-k approximation retains a fraction of about k/R of the noise energy for $P \gg R$, so with $R = 5$ the denoising is limited unless k is well below R.

2.4. Multiscale energy models

The prior of 2-D images $\mathbf{u}$ at noise level σ is $p_w(\mathbf{u};\sigma) \propto \exp(-\phi_w(\mathbf{u};\sigma)/\sigma^2)$, with $\phi_w(\mathbf{u};\sigma) = \frac{1}{2}\|\mathbf{u} - \psi_w(\mathbf{u};\sigma)\|_2^2$ and a convolutional network ψ_w [7, 8]. Its gradient defines the denoiser

$$D_\sigma(\mathbf{u}) = \mathbf{u} - \nabla_{\mathbf{u}}\phi_w(\mathbf{u};\sigma), \tag{1}$$

trained by multiscale denoising score matching [14], which makes D_σ approximate the MMSE denoiser. The explicit energy allows classical optimization, such as the majorize–minimize scheme below, instead of posterior sampling. Annealing σ from large to small further improves convergence: early iterations smooth the energy and fix coarse structure, and later ones recover detail [7, 8].

3. PROPOSED METHOD

3.1. Multiscale MAP formulation using contrast priors

Contrast prior. Unlike SS-MuSE [11], which learns a prior directly on the spatial factors $\mathbf{U}$, we learn it indirectly through the contrast images $\mathbf{x}_\tau = \mathbf{U}\mathbf{v}_\tau$, $\tau = 1, \dots, T$, which are linear functions of $\mathbf{U}$. A prior on $\mathbf{U}$ requires training data specific to the basis functions $\mathbf{V}$, which unfortunately is not available in our high-resolution setting. A prior on contrast images can instead be learned from generic images, and such priors were recently shown to generalize well to MRI [12].

Slice-wise energy. A 3-D energy would need much 3-D training data and memory, so we apply the 2-D model to slices. For a volume $\mathbf{x} \in \mathbb{C}^{n\times n\times n}$ and an axis $a \in \{x, y, z\}$, the operator $\mathcal{S}_a$ extracts the n 2-D slices perpendicular to a, each an $n \times n$ image, and the real and imaginary parts are treated as separate images [13, 11]. The energy of the volume along a is the sum of the energies of its slices, $e_\sigma(\mathcal{S}_a\mathbf{x}) = \sum_j e_\sigma(\mathbf{x}_{a,j})$, where $e_\sigma = -\log p_w(\cdot;\sigma)$ and $\mathbf{x}_{a,j}$ is the jth slice.

MAP estimation. We estimate $\mathbf{U}$ by minimizing

$$\begin{aligned} J(\mathbf{U}) &= \tfrac{1}{2}\|\mathcal{A}(\mathbf{U}\mathbf{V}) - \mathbf{b}\|_2^2 + \eta^2 E_\sigma(\mathbf{U}), \\ E_\sigma(\mathbf{U}) &= \sum_{\tau\in\mathcal{P}} \sum_{a\in\{x,y,z\}} e_\sigma\big(\mathcal{S}_a(\mathbf{U}\mathbf{v}_\tau)\big), \end{aligned} \tag{2}$$

where η is a regularization parameter and $\mathcal{P}$ is a pool of N frames spread over the inversion recovery. Since $R \ll T$, a pool much smaller than T suffices. By (1), $\nabla e_\sigma(\mathbf{x}) = (\mathbf{x} - D_\sigma(\mathbf{x}))/\sigma^2$, so E_σ is never evaluated explicitly.

3.2. Majorization

For an ideal MMSE denoiser, the Hessian of e_σ is $\sigma^{-2}\big(I - \sigma^{-2}\mathrm{Cov}[\mathbf{u}\,|\,\tilde{\mathbf{u}}]\big) \preceq \sigma^{-2} I$ [15], so at $\mathbf{x}_\tau^k = \mathbf{U}_k\mathbf{v}_\tau$ each term is bounded by a quadratic majorizer whose minimization requires no step size (for the trained network, this bound holds

approximately),

$$e_\sigma(\mathbf{x}) \le \tfrac{1}{2\sigma^2}\left\|\mathbf{x} - D_\sigma(\mathbf{x}_\tau^k)\right\|_2^2 + c_\tau^k. \quad (3)$$

3.3. Rotating contrast subsets

Evaluating all $3N$ terms at every iteration is expensive, so at iteration k we keep the terms of a subset $\mathcal{K}_k \subseteq \mathcal{P}$ of $K \le N$ frames, each assigned one axis a_τ, and scale each kept term by $3N/K$ so that the sum approximates the full sum over $3N$ terms. The subset and the axis assignments rotate between iterations, so that MuSE is applied along different axes and, over the iterations, to all frames of the pool. The resulting majorizer is stochastic: it is an unbiased surrogate of (3) summed over all terms if the frames and axes are drawn uniformly [16], and rotating through interleaved subsets approximates this. Each iteration costs K 2-D passes over the volume, and the descent is not monotone. The selected contrasts are denoised, $\hat{\mathbf{x}}_\tau = D_\sigma(\mathbf{x}_\tau^k)$ along a_τ, and mapped back to the subspace by least squares,

$$\tilde{\mathbf{U}}_k = \mathbf{U}_k + \sum_{\tau\in\mathcal{K}_k} \left(\hat{\mathbf{x}}_\tau - \mathbf{x}_\tau^k\right)\mathbf{v}_\tau^H \left(\mathbf{V}_{\mathcal{K}_k}\mathbf{V}_{\mathcal{K}_k}^H + \epsilon I\right)^{-1}, \quad (4)$$

where $\mathbf{V}_{\mathcal{K}_k}$ holds the columns of $\mathbf{V}$ of the selected frames and $\epsilon > 0$ is a small constant that stabilizes the inverse. Approximating $\|(\mathbf{U}-\tilde{\mathbf{U}}_k)\mathbf{V}_{\mathcal{K}_k}\|_F$ by $\|\mathbf{U}-\tilde{\mathbf{U}}_k\|_F$ and absorbing constant scales into η gives

$$\mathbf{U}_{k+1} = \arg\min_{\mathbf{U}} \ \tfrac{1}{2}\|\mathcal{A}(\mathbf{U}\mathbf{V}) - \mathbf{b}\|_2^2 + \tfrac{\lambda_k}{2}\|\mathbf{U} - \tilde{\mathbf{U}}_k\|_F^2, \quad (5)$$

with $\lambda_k = \eta^2/\sigma_k^2$, solved by a few warm-started conjugate gradient (CG) iterations.

3.4. Continuation and renoising

As σ_k is annealed, λ_k increases, so the weight on mildly denoised anchors grows. The denoiser is trained on inputs of the form $\mathbf{u} + \sigma\mathbf{z}$, whereas the iterate $\mathbf{x}_\tau^k$ carries residual aliasing and noise that are not Gaussian at level σ_k. To reduce this train–test mismatch, as in D-MOD [17], the anchor can be averaged over L renoised inputs, $\hat{\mathbf{x}}_\tau = \frac{1}{L}\sum_l D_\sigma(\mathbf{x}_\tau^k + \sigma\mathbf{z}_l)$ with $\mathbf{z}_l \sim \mathcal{N}(0, I)$; $L = 0$ means no renoising.

4. EXPERIMENTS

4.1. Validation strategy and BrainWeb phantom

No ground truth is available for 0.5 mm quantitative MRI in vivo, so we validated the methods on a numerical phantom with known T_1 maps and contrasts. The BrainWeb anatomical model [18] provides 1 mm segmentation maps, which we interpolated to 0.5 mm. We assigned representative T_1, B_1^+, inversion efficiency (IE) and proton density (PD) values to gray matter, white matter, CSF and skull, which gives the parameter maps ϑ_m and ρ_m of the signal model in Sec. 2. The ground-truth contrasts ($T = 460$) were generated by Bloch simulation of the MPnRAGE sequence with the blocks, spoke timing and flip-angle schedule of the in vivo scan. The k-space data were then simulated with the forward model $\mathcal{A}_\tau$, that is, the in vivo radial trajectory and coil sensitivities, without motion, with in vivo-matched noise added.

4.2. In vivo data

One healthy volunteer was scanned on a 7 T GE scanner with a 32-channel head coil (Sec. 2). Images were reconstructed at 0.5 mm isotropic resolution (512^3, 256 mm FOV) from 16 compressed virtual coils, with per-block rigid motion included in $\mathcal{A}_\tau$.

4.3. Training and use of the prior

The network ψ_w is a 3.4 M-parameter SongUNet with EDM preconditioning [19]. It was trained by denoising score matching on 100 k grayscale CelebA face images, with no MRI data. To apply it to MRI, each synthesized contrast volume is cut into 2-D slices along the chosen axis. Because the network is trained on real-valued images, the real and imaginary parts of each slice are denoised separately as two grayscale images and then recombined into the complex slice, $\hat{\mathbf{x}} = D_\sigma(\mathrm{Re}\,\mathbf{x}) + i\,D_\sigma(\mathrm{Im}\,\mathbf{x})$.

4.4. Compared methods, tuning and metrics

All methods share the basis ($R = 5$), the forward model (scaled to $\lambda_{\max}(\mathcal{A}^H\mathcal{A}) = 1$) and the coil maps. CG-SENSE [20] uses 50 CG iterations and initializes the regularized methods. MuSE uses a pool of $N = 8$ frames (frames 10–420, TI 69–2095 ms) with MuSE applied along rotating axes (here $K = N$, so the subset is the full pool), no renoising, and the first seven levels of a ten-level EDM schedule from σ_0 (the noise level of the CG-SENSE images) to $0.08\,\sigma_0$. Each MuSE iteration uses 20 CG iterations. L1-wavelet uses a six-level orthonormal Daubechies-4 transform and ADMM (10 iterations of 5 CG steps); LLR uses shifted 8^3 blocks, the threshold $c\,(\sqrt{P} + \sqrt{R})$ in noise units and ADMM (10 iterations). At $1\times$ and $2\times$, the MuSE weight is $\eta/\sigma_0 = 0.2$ and 0.15, the wavelet weight is $\lambda = 3.2 \times 10^{-3}$ and 1.6×10^{-3}, and the LLR constant is $c = 0.25$ and 0.15.

The parameters of all methods were tuned on the phantom to minimize the NRMSE in the brain, averaged over the WM-nulled, MPRAGE-like and CSF-nulled frames (TI 415/1398/855 ms). In vivo, the noise-relative c and η were kept, and the wavelet λ was scaled by the ratio of noise levels and data gains (0.69 at $1\times$, 0.82 at $2\times$). Figures show contrasts synthesized with $B_1^+ = \mathrm{IE} = 1$. Fine detail was assessed by the high-frequency error norm (HFEN; 3-D LoG, $\sigma = 1.5$ voxels) in the brain, and T_1 by the median bias and interquartile range (IQR) in WM and GM on three orthogonal slices.

Table 1. Phantom accuracy (lower is better). NRMSE and high-frequency error (HFEN) over the brain volume, averaged over the WM-nulled, MPRAGE-like and CSF-nulled contrasts, at $1\times$ (8.6 min) and $2\times$ (4.3 min). T_1 rows: median bias/IQR (ms) at $2\times$. MuSE has the lowest NRMSE and HFEN in every row, and its $2\times$ NRMSE is below the $1\times$ L1-wavelet value.

	CG-SENSE	LLR	L1-wavelet	MuSE
NRMSE, $1\times$	0.119	0.105	0.080	**0.058**
HFEN, $1\times$	0.197	0.217	0.253	**0.146**
NRMSE, $2\times$	0.133	0.110	0.091	**0.074**
HFEN, $2\times$	0.317	0.249	0.274	**0.244**
T_1 bias/IQR, WM	-42/117	-6/66	**-2**/47	-19/**38**
T_1 bias/IQR, GM	-34/325	-34/251	-30/193	**-18/159**

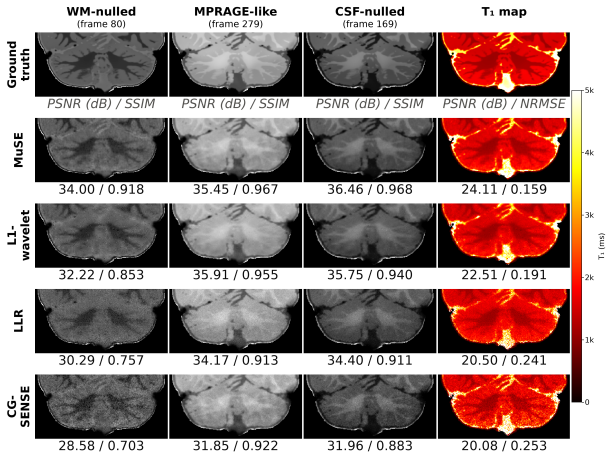


Fig. 2. BrainWeb validation at $2\times$ (4.3 min), cerebellar crop, compared with the ground truth. Columns: three synthesized contrasts (frames 80, 279, 169) and the T_1 map; metrics under each panel are PSNR/SSIM (contrasts) and PSNR/NRMSE (T_1) over the full-slice brain mask.

5. RESULTS AND DISCUSSION

Phantom comparison. MuSE gave the lowest whole-brain NRMSE at both scan times (Table 1); at $2\times$, it was 44% lower than CG-SENSE and 19% lower than L1-wavelet, and even below the $1\times$ L1-wavelet result (0.074 versus 0.080). It also had the lowest HFEN and, at $2\times$, the narrowest T_1 IQR in WM (38 ms, versus 47–117) and GM (159 ms, versus 193–325). L1-wavelet was the closest competitor, and the two methods are nearly equal on the high-SNR MPRAGE-like frame (NRMSE 0.0426 versus 0.0431; PSNR 35.9 versus 35.5 dB). The difference lies in the low-SNR images (Fig. 2). In the WM-nulled frame, L1-wavelet leaves a mottled, patchy texture, whereas MuSE is smoother and keeps the thin cerebellar white matter branches (PSNR 34.0 versus 32.2 dB; SSIM 0.918 versus 0.853). The T_1 map shows the same pattern: L1-wavelet is speckled, whereas MuSE is homogeneous

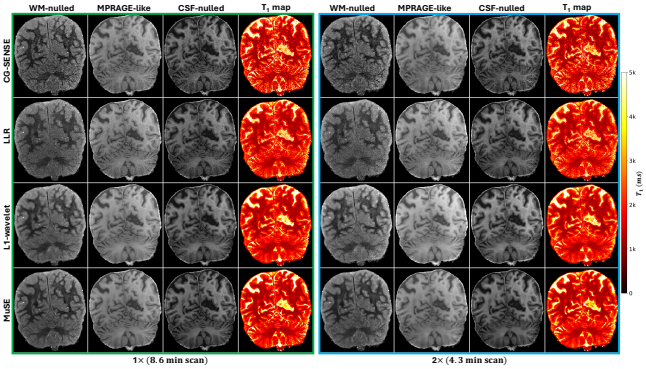


Fig. 3. In vivo reconstructions at $1\times$ (8.6 min) and $2\times$ (4.3 min), layout as in Fig. 2; no ground truth is available.

with sharper folia boundaries (NRMSE 0.159 versus 0.191). MuSE had a WM T_1 bias of -19 ms at $2\times$.

In vivo. Fig. 3 shows the in vivo reconstructions of the four methods at $1\times$ and $2\times$. Since no ground truth exists, we compare them only visually, and a quantitative in vivo evaluation is left for future work.

6. LIMITATIONS

This study is preliminary: it uses one subject and one phantom, the in vivo comparison is visual, memory use is high (about 115 GB on each of two GPUs at $1\times$), and the face-trained prior is untested in pathology. More elaborate studies will follow.

7. CONCLUSION

MuSE regularizes subspace quantitative MRI with a multi-scale energy prior on synthesized contrasts, learned from face images and minimized by a step-size-free stochastic MM algorithm. At 0.5 mm, it outperformed the other methods on a phantom and was demonstrated in vivo at 7 T.

8. COMPLIANCE WITH ETHICAL STANDARDS

This study was performed in line with the principles of the Declaration of Helsinki. In vivo data were acquired at the University of Iowa with approval of the Institutional Review Board and written informed consent from the subject.

9. ACKNOWLEDGMENTS

Supported by NIH grants R01-AG067078, R01-EB031169 and R01-EB019961. We thank UVA Research Computing (`rc.virginia.edu`) for computational resources. Claude (Anthropic) was used to refine the text of this paper. The authors reviewed the content and take full responsibility for it.

10. REFERENCES

[1] Steven Kecskemeti, Alexey Samsonov, Samuel A. Hurley, Douglas C. Dean, Aaron Field, and Andrew L. Alexander, "MPnRAGE: A technique to simultaneously acquire hundreds of differently contrasted MPRAGE images with applications to quantitative T1 mapping," *Magnetic Resonance in Medicine*, vol. 75, no. 3, pp. 1040–1053, 2016.

[2] Laura E. M. Wisse, Gaël Chételat, Ana M. Daugherty, et al., "Hippocampal subfield volumetry from structural isotropic 1 mm^3 MRI scans: A note of caution," *Human Brain Mapping*, vol. 42, no. 2, pp. 539–550, 2021.

[3] Zhi-Pei Liang, "Spatiotemporal imaging with partially separable functions," in *Proc. IEEE Int. Symp. Biomed. Imaging (ISBI): From Nano to Macro*, 2007, pp. 988–991.

[4] Michael Lustig, David Donoho, and John M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.

[5] Tao Zhang, John M. Pauly, and Ives R. Levesque, "Accelerating parameter mapping with a locally low rank constraint," *Magnetic Resonance in Medicine*, vol. 73, no. 2, pp. 655–661, 2015.

[6] Hyungjin Chung and Jong Chul Ye, "Score-based diffusion models for accelerated MRI," *Medical Image Analysis*, vol. 80, pp. 102479, 2022.

[7] Jyothi Rikhab Chand and Mathews Jacob, "Multiscale energy (MuSE) framework for inverse problems in imaging," *IEEE Transactions on Computational Imaging*, vol. 10, pp. 1250–1265, 2024.

[8] Jyothi Rikhab Chand and Mathews Jacob, "Annealed Langevin posterior sampling (ALPS): A rapid algorithm for image restoration with multiscale energy models," *arXiv preprint arXiv:2601.02594*, 2026.

[9] Hyungjin Chung, Jeongsol Kim, Michael T. McCann, Marc L. Klasky, and Jong Chul Ye, "Diffusion posterior sampling for general noisy inverse problems," in *International Conference on Learning Representations (ICLR)*, 2023.

[10] Bingliang Zhang, Wenda Chu, Julius Berner, Chenlin Meng, Anima Anandkumar, and Yang Song, "Improving diffusion inverse problem solving with decoupled noise annealing," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025, pp. 20895–20905.

[11] Yan Chen, Jyothi Rikhab Chand, Steve R. Kecskemeti, James H. Holmes, and Mathews Jacob, "Accelerating quantitative MRI using subspace multiscale energy model (SS-MuSE)," in *Proc. IEEE 22nd International Symposium on Biomedical Imaging (ISBI)*, 2025, pp. 1–5.

[12] Lukas Glaszner and Martin Zach, "Bigger isn't always better: Towards a general prior for medical image reconstruction," *arXiv preprint arXiv:2501.07376*, 2025.

[13] Suhyeon Lee, Hyungjin Chung, Minyoung Park, Jonghyuk Park, Wi-Sun Ryu, and Jong Chul Ye, "Improving 3D imaging with pre-trained perpendicular 2D diffusion models," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, pp. 10710–10720.

[14] Pascal Vincent, "A connection between score matching and denoising autoencoders," *Neural Computation*, vol. 23, no. 7, pp. 1661–1674, 2011.

[15] Hila Manor and Tomer Michaeli, "On the posterior distribution in denoising: Application to uncertainty quantification," in *International Conference on Learning Representations (ICLR)*, 2024, arXiv:2309.13598.

[16] Julien Mairal, "Stochastic majorization-minimization algorithms for large-scale optimization," in *Advances in Neural Information Processing Systems*, C. J. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Weinberger, Eds. 2013, vol. 26, Curran Associates, Inc.

[17] Md Kawsar Ahmed and Mathews Jacob, "D-MOD: Diffusion mode estimation for accelerated MRI," *IEEE Transactions on Medical Imaging*, 2026, submitted.

[18] Chris A. Cocosco, Vasken Kollokian, Remi K.-S. Kwan, and Alan C. Evans, "BrainWeb: Online interface to a 3D MRI simulated brain database," *NeuroImage*, vol. 5, no. 4, part 2/4, pp. S425, 1997.

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

[20] Klaas P. Pruessmann, Markus Weiger, Peter Börnert, and Peter Boesiger, "Advances in sensitivity encoding with arbitrary k-space trajectories," *Magnetic Resonance in Medicine*, vol. 46, no. 4, pp. 638–651, 2001.