PaDIS-MRI: Patch-Based Diffusion for Data-Efficient, Radiologist-Preferred MRI Reconstruction

Rohan Sanda RSANDA@STANFORD.EDU
Department of Computer Science, Stanford University, USA

Asad Aali ASADAALI@STANFORD.EDU
Department of Radiology, Stanford University School of Medicine, USA

Andrew Johnston DREWJ32@STANFORD.EDU
Department of Radiology, Stanford University School of Medicine, USA

Eduardo Reis EDREIS@STANFORD.EDU
Stanford Center for Artificial Intelligence in Medicine and Imaging, USA

Gordon Wetzstein GORDONWZ@STANFORD.EDU
Department of Electrical Engineering, Stanford University, USA

Sara Fridovich-Keil SFK@GATECH.EDU
School of Electrical and Computer Engineering, Georgia Institute of Technology, USA

Abstract

Magnetic resonance imaging (MRI) requires long acquisition times, which raise costs, reduce accessibility, and increase susceptibility to motion artifacts. Diffusion probabilistic models that learn data-driven priors may reduce acquisition time by enabling reconstruction from undersampled k-space measurements. However, they typically require large training datasets that can be prohibitively expensive to collect. Patch-based diffusion models have shown promise in learning effective data-driven priors over small real-valued datasets, but have not yet demonstrated clinical value in MRI. We extend the Patch-based Diffusion Inverse Solver (PaDIS) to complex-valued, multi-coil MRI reconstruction, and compare it against a state-of-the-art whole-image diffusion baseline (FastMRI-EDM) for 7× undersampled MRI reconstruction on the FastMRI brain dataset. We show that PaDIS-MRI models trained on small datasets of as few as 25 k-space images outperform FastMRI-EDM on image quality metrics (PSNR, SSIM, NRMSE), pixel-level mask-induced variability, cross-contrast/-modality generalization, and robustness to severe k-space undersampling. In a blinded study with three radiologists, PaDIS-MRI reconstructions were chosen as diagnostically superior in 91.7% of cases, compared to baselines (i) FastMRI-EDM and (ii) classical convex reconstruction with wavelet sparsity. These findings highlight the potential of patch-based diffusion priors for high-fidelity MRI reconstruction in data-scarce clinical settings where diagnostic confidence matters.

Keywords: MRI reconstruction; diffusion models; patch-based priors; data efficiency; out-of-distribution robustness

Data and Code Availability. Code is available at: https://github.com/voilalab/PaDIS-MRI. Data is already public as part of FastMRI (Zbontar et al., 2019).

Institutional Review Board (IRB). This work uses the public, de-identified FastMRI dataset (Zbontar et al., 2019), with clinical analysis provided by radiologist members of the research team. IRB review was not required.

1. Introduction

Magnetic Resonance Imaging (MRI) is one of the most versatile and non-invasive diagnostic tools in modern medicine. Unlike other popular imaging modalities such as X-ray or computed tomography (CT) scans, MRI leverages magnetic fields and radiofrequency pulses to visualize soft tissue structures with high contrast without cell-damaging ionizing radiation (Grover et al., 2015). Despite its widespread use in clinical practice, the long acquisition times of MRI (Hollingsworth, 2015) contribute to increased cost, pa-

© 2025 R. Sanda, A. Aali, A. Johnston, E. Reis, G. Wetzstein & S. Fridovich-Keil.

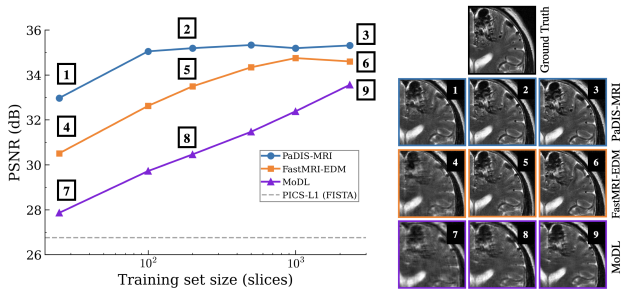

Figure 1: **Reconstruction quality across dataset sizes.** PaDIS-MRI (ours) consistently outperforms FastMRI-EDM (a full-image diffusion prior) and MoDL (a learned, supervised prior) across all training set sizes, even with as few as 25 images. PICS-L1 is untrained and thus independent of training set size. Tabulated metrics are in Appendix C.1.

tient discomfort, and degraded image quality from motion artifacts (Jaspan et al., 2015).

To decrease acquisition time, it is common practice for modern scanners to undersample k-space – the Fourier domain in which MRI data is captured (Ye, 2019). However, this introduces an ill-posed inverse problem of reconstructing a high-fidelity image from incomplete measurements. MRI reconstruction algorithms surmount this challenge by leveraging structural priors over medical images, ranging from classical compressed sensing priors like wavelet sparsity and limited total variation to modern data-driven priors that capture the statistics of training images (Ye, 2019). Among these approaches, generative models (e.g. GANs, VAEs, and diffusion models) have emerged as powerful tools for learning complex, high-dimensional priors directly from MRI data (Aali et al., 2023, 2024, 2025; Fan et al., 2024). Through a Bayesian framework, these models allow reconstruction to be posed as a posterior sampling optimization problem, where the goal is to find images that both explain the measured data and conform to the learned prior distribution of the anatomy observed in fully-sampled training images (Jalal et al., 2021).

However, generative models require notoriously large training datasets for reliable performance. This dependence presents multiple challenges to clinical use. First, acquiring high-fidelity MRI scans from a diverse set of patients is expensive, time-intensive, and often restricted by privacy constraints. Second, generative models trained on one anatomy (e.g., brain), pathology (e.g., tumor), or contrast (e.g., T1-weighted) may not generalize well to others, limiting real-world deployment. Data-efficient generative priors are thus critical to safe and clinically effective adoption of diffusion priors for MRI reconstruction.

To address these limitations, we extend the Patch-based Diffusion Inverse Solver (PaDIS) framework proposed by Hu et al. (2024b) to the setting of multi-coil undersampled MRI reconstruction. Rather than learning global anatomical priors over entire images, we train a diffusion model on complex-valued image patches drawn at random from the FastMRI dataset (Zbontar et al., 2019). This patch-based formulation enables the model to learn localized structural motifs, improving performance in low-data regimes commonly encountered in clinical settings.

We integrate this learned patch-based prior into the Diffusion Posterior Sampling (DPS) algorithm to reconstruct MRI images from undersampled k-space measurements, with our modifications specifically designed to handle complex-valued MRI. We benchmark our patch-based PaDIS-MRI prior against the whole-image diffusion model FastMRI-EDM developed by Aali et al. (2024). We find that the patch-based model consistently outperforms the whole-image prior across reconstruction quality metrics (PSNR, SSIM, and NRMSE) using small training datasets with as few as 25 images (see Figure 1), while also achieving better cross-contrast/-domain generalization and more consistent results across varying undersampling patterns.

To assess clinical utility, we conducted a blinded comparison analysis in which three radiologists examined 60 sets of images reconstructed by PaDIS-MRI, FastMRI-EDM, and a classical wavelet-sparse reconstruction and selected which image was most faithful to the ground truth fully-sampled reconstruction in terms of diagnostic quality. The majority of radiologists selected PaDIS-MRI as the most clinically accurate reconstruction in 55/60 cases (91.7%). To summarize, our core contributions include:

- Adapting PaDIS (Hu et al., 2024a) for training on two-channel (real/imaginary) complex-valued data in undersampled MRI reconstruction.
- Evaluating the quantitative reconstruction fidelity of PaDIS-MRI, which outperforms FastMRI-EDM and a wavelet-sparse baseline across training set sizes and k-space undersampling levels, and exhibits better cross-contrast/-domain generalization and lower pixel-wise mask-induced variability.
- Evaluating the potential diagnostic utility of PaDIS-MRI compared to FastMRI-EDM and a

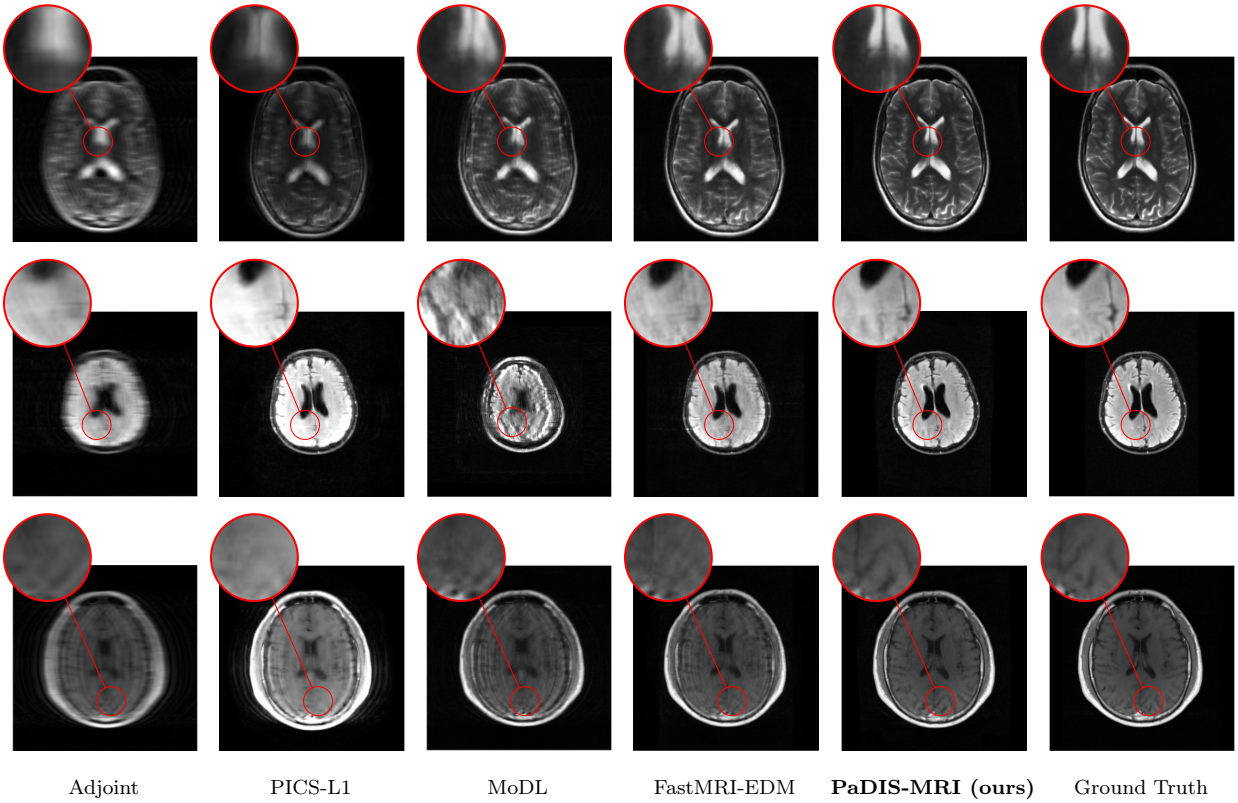

Figure 2: Example MRI magnitude reconstructions with $S = 25$ training slices at acceleration factor $R = 7$. PaDIS-MRI reconstructions are sharper and more faithful to the fully-sampled ground truth. Insets are shown to highlight detail.

wavelet-sparse baseline via a blinded comparison study with three radiologists, who overwhelmingly selected PaDIS-MRI reconstructions as having higher diagnostic quality and clinical accuracy with respect to the fully-sampled ground truth.

2. Related Work

2.1. Diffusion Models for MRI Reconstruction

Diffusion models have emerged as powerful generative priors for solving ill-posed inverse problems in MRI reconstruction. Song et al. (2021) first introduced continuous-time score-based generative modeling, where image generation is cast as reversing a stochastic differential equation (SDE) from noise back to data. Building on this foundation, Jalal et al. (2021) were among the first to apply score-based models to compressed-sensing MRI, showing that posterior sampling with a learned score prior can reconstruct high fidelity images from undersampled k-space. Chung and Ye (2022) adapted score-based diffusion to MRI denoising tasks and demonstrated significant gains over supervised baselines. Chung et al. (2024) then introduced the diffusion posterior sampling (DPS) algorithm by augmenting the reverse SDE with a measurement-consistency gradient to more directly approximate the intractable posterior distribution.

2.2. Patch-Based Diffusion

Wang et al. (2023) proposed training on image patches to improve efficiency, but still required full-image score estimation at inference time. Hu et al. (2024b) addressed this limitation with their PaDIS framework, which was demonstrated on a variety of inverse problems involving CT scans and natural images. By modeling an image as a Markov Random Field, where cliques correspond to patches of pixels, they leverage the Hammersley-Clifford theorem to express the score function of the entire image as a sum of score functions of patches. More recent work by Hu et al. (2024a) demonstrated that PaDIS also improves out-of-

distribution (OOD) generalization compared to whole-image diffusion models.

A few studies have demonstrated the feasibility of patch-based models for MRI tasks. Behrendt et al. (2023) introduced patched denoising diffusion probabilistic models for unsupervised anomaly detection in brain MRIs, demonstrating improvements in tumor segmentation performance. Separately, Huang et al. (2024) developed the PADMr framework that leverages patch-based inputs and a non-Markovian diffusion process to accelerate inference times for MRI reconstruction. Most recently, Roy et al. (2025) studied patch-based inference for generalized diffusion priors in MRI reconstruction, examining memory efficiency and performance on denoising and super-resolution tasks. Rather than focusing on computational benefits, we evaluate the potential clinical impact of the PaDIS framework in multi-coil undersampled MRI reconstruction, focusing on data efficiency, robustness (to contrast variation and degree of undersampling), and diagnostic fidelity as assessed by quantitative image metrics and radiologists.

3. Background

3.1. MRI Reconstruction

Multi-coil MRI is a linear inverse problem that aims to reconstruct a vectorized 2D anatomical ground-truth image $x \in \mathbb{C}^{N_x N_y}$ from undersampled k-space measurements y. Specifically, the forward operator $\mathcal{A}$ combines the coil sensitivity operator $S \in \mathbb{C}^{N_c N_x N_y \times N_x N_y}$, the Fourier Transform operator $F \in \mathbb{C}^{N_c N_x N_y \times N_c N_x N_y}$, and the undersampling pattern $P \in \mathbb{C}^{\frac{N_c N_x N_y}{R} \times N_c N_x N_y}$, where N_c is the number of coils, N_x, N_y denote spatial dimensions of the image, and R represents the undersampling (acceleration) factor (Aali et al., 2024). The resulting undersampled measurements are thus $y \in \mathbb{C}^{\frac{N_c N_x N_y}{R}}$ and we have:

$$y = PFSx + \epsilon \tag{1}$$

where ϵ denotes complex-valued additive noise from scanner imperfections or patient motion. Due to k-space undersampling ($R > 1$), reconstructing the ground truth clean image x from y constitutes an ill-posed inverse problem. Traditionally, this problem is formulated as a regularized optimization over x where $\mathcal{A} = PFS$:

$$\tilde{x} = \arg\min_x \|y - \mathcal{A}x\|_2^2 + \lambda \mathcal{R}(x). \tag{2}$$

From a Bayesian perspective, optimization problems like Equation (2) are often formulated as maximum a posteriori (MAP) estimation. However, DPS instead approximates posterior sampling from $p(x|y) \propto p(y|x)p(x)$ using a learned score-based prior $p_\theta(x)$, yielding diverse generations that reflect the full posterior distribution rather than collapsing to a single, deterministic point estimate.

3.2. Diffusion and DPS

Score-based diffusion models follow an SDE to progressively corrupt a clean image x. The forward SDE, defined by Song et al. (2021), can be expressed as:

$$dx = f(x,t)\,dt + g(t)\,dW_t, \tag{3}$$

where W_t denotes standard Brownian motion, and functions f and g define drift and diffusion coefficients, respectively chosen so that at time $t = T$, $x \sim \mathcal{N}(0, \sigma_T^2 I)$. Sampling from this model involves solving the reverse-time SDE (Song et al., 2021):

$$dx = \left[f(x,t) - g(t)^2 \nabla_x \log p_t(x)\right]\,dt + g(t)\,d\bar{W}_t, \tag{4}$$

where $\nabla_x \log p_t(x)$ is the learned score function estimated by the denoising neural network D_θ.

To solve linear inverse problems of the form $y = \mathcal{A}x + \epsilon$, Chung et al. (2024) augment the reverse SDE with a measurement-consistency likelihood gradient:

$$\begin{aligned} dx = &\left[f(x,t) - g(t)^2 \nabla_x \log p_t(x)\right]\,dt \\ &+ t\,\nabla_x \log p(y \mid x)\,dt + g(t)\,d\bar{W}_t. \end{aligned} \tag{5}$$

where $\nabla_x \log p(y \mid x) = -\nabla_x \|y - \mathcal{A}x\|^2$. In practice, to solve Equation (5) the Variance-Exploding (VE) discretization from Karras et al. (2022) can be used, parameterized by:

$$t_k = \left(\sigma_{\max}^{1/\rho} + \tfrac{k}{N-1}\left(\sigma_{\min}^{1/\rho} - \sigma_{\max}^{1/\rho}\right)\right)^\rho, \quad k = 0, \ldots, K-1 \tag{6}$$

where K is the number of steps, and ρ, $\sigma_{\min}$, and $\sigma_{\max}$ are hyperparameters for the noising schedule. In the VE-DPS algorithm, the following updates to the image x_k at step k are made:

$$\alpha_k = \tfrac{1}{2}\,t_k^2, \quad z_k \sim \mathcal{N}(0, I), \quad \mathrm{SSE}_k = \left\|y - \mathcal{A}\,\hat{x}_k\right\|_2^2,$$

$$\mathrm{score}_k = \frac{D_\theta(x_k, t_k) - x_k}{t_k^2}, \quad \eta_k = \begin{cases} \sqrt{\alpha_k}\,z_k, & k < K, \\ 0, & k = K, \end{cases}$$

$$x_k' = x_k - \frac{\zeta}{\sqrt{\mathrm{SSE}_k}}\,\nabla_{x_k}\,\mathrm{SSE}_k,$$

$$x_{k+1} = x_k' + \frac{\alpha_k}{2}\,\mathrm{score}_k + \eta_k,$$

where ζ controls the trade-off between data fidelity and prior consistency and y and $\mathcal{A}$ are as defined in Subsection 3.1.

3.3. Diffusion with Patches

The PaDIS framework (Hu et al., 2024b) employs a single neural network to learn a unified prior across all image patches, incorporating positional encodings to distinguish between patch locations within the image. During training, PaDIS adopts a flexible patching strategy where patches are extracted from zero-padded images, with both patch size and location randomized. Each patch is processed alongside its positional encoding, allowing the network to learn location-aware representations across multiple scales. At inference, we construct a fixed coverage grid and randomly choose a start offset at each diffusion iteration. Given an input image $x_0 \in \mathbb{R}^{N\times N}$ and a designated patch size P, we zero-pad the image to admit a $(k{+}1)\times(k{+}1)$ grid of $P\times P$ patches with stride P, where $k=\lfloor N/P\rfloor$. The padding width $M=(k{+}1)P-N$ ensures that the grid extends beyond the original image boundaries; it also equals the number of valid start offsets per axis, so there are M^2 possible partitionings.

The global score function of the entire image is then expressed as in Hu et al. (2024b):

$$\nabla_x \log p(x) = \sum_{i,j=1}^{M} \Big(s_{i,j,B}(x_{i,j,B}) + \sum_{r=1}^{(k+1)^2} s_{i,j,r}(x_{i,j,r}) \Big),$$

where each pair (i,j) indexes a start-offset partition, $x_{i,j,r}$ denotes the r-th patch under partition (i,j), $s_{i,j,r}$ is its corresponding score function, and $x_{i,j,B}$ is the zero-padded border region. For efficiency, at each inference iteration PaDIS samples a single partition (i,j) rather than evaluating all M^2; across iterations this Monte-Carlo averaging over start offsets mitigates patch-boundary artifacts.

4. Methods

We adapt the PaDIS framework (Hu et al., 2024b) for multi-coil undersampled MRI reconstruction, including both training of the patch-based score model and inference via our DPS sampler. Similar to PaDIS, we build on top of the EDM model architecture from Karras et al. (2022). Implementation details and hyperparameters may be found in Appendix A.

4.1. PaDIS-MRI Training

Training images of dimension $N\times N$ are zero-padded by $\frac{1}{4}N$ pixels on all sides (96 px when N=384). Rather than RGB or single-channel images, we train on complex coil-combined MR images. Each extracted $P\times P$ patch ($P\in\{16,32,64\}$, sampled with probabilities $\{0.2,0.3,0.5\}$) is represented as a 2–channel real tensor for the real/imag parts and processed by a Song-UNet-based denoiser (∼55M parameters). For each patch, we append a 2D positional encoding of its (x,y) location (normalized to $[-1,1]$), informing the denoiser of the absolute position within the full image. This position-aware training enables the model to learn location-specific features while maintaining coherent full-image reconstructions. As in PaDIS, the score-matching loss is computed on patches sampled from random locations across the padded image, allowing the model to learn the distribution of patches at all positions (Hu et al., 2024b).

4.2. PaDIS-MRI Inference

At inference time, we reconstruct image x from undersampled k-space y using our patch-based VE-DPS sampler with K=104 steps. We warm start with the adjoint $x_0=\mathcal{A}^\dagger(y)$, then symmetrically zero-padded by M=64 pixels per side. We can then follow the VE-DPS flow (Algorithm 1 in Appendix B) with patch-size P=64 to solve the MRI reconstruction problem.

For each step $k=0,\ldots,K-1$, we (1) inject noise $x\leftarrow x+t_k\varepsilon$ with $\varepsilon\sim\mathcal{N}(0,I)$, where $x\in\mathbb{C}^{(N+2M)\times(N+2M)}$ is the padded image; (2) extract a non-overlapping grid of $P\times P$ patches using a random starting offset, (3) convert each complex patch $p\in\mathbb{C}^{P\times P}$ into a 2-channel real tensor $p_{\text{real}}\in\mathbb{R}^{P\times P\times 2}$, (4) denoise via D_θ, and reassemble the denoised patches into $D\in\mathbb{C}^{(N+2M)\times(N+2M)}$; (5) form the estimated score $=(D-x)/t_k^2$; (6) crop D to the original field of view to obtain $\hat{x}\in\mathbb{C}^{N\times N}$, (7) compute the residual $r=y-\mathcal{A}(\hat{x})$ and SSE $=\|r\|^2$ in k-space, and (8) backpropagate and pad to get $\nabla_x\text{SSE}\in\mathbb{C}^{(N+2M)\times(N+2M)}$, and (9) perform the update

$$x \leftarrow \Big(x+\frac{\alpha_k}{2}\,\text{score}+\sqrt{\alpha_k}\,\varepsilon\Big)-\frac{\zeta}{\sqrt{\text{SSE}}}\,\nabla_x\,\text{SSE}.$$

After K iterations (with ζ=3.0) we remove the padding to obtain the final reconstruction $x_K\in\mathbb{C}^{N\times N}$.

4.3. Dataset

We use fully-sampled, multicoil k-space data from the NYU Langone Health FastMRI dataset (Zbontar et al., 2019), including FLAIR, native T1, pre-contrast T1, post-contrast T1, and axial T2-weighted brain scans (Zbontar et al., 2019). In total, the training set contains 446 k-space volumes, consisting of 40 FLAIR, 136 T1-weighted (91 post-contrast, 26 pre-contrast, 19 native), and 270 T2-axial. Our validation set contains 266 volumes (distinct from the training volumes) from which we sample an evaluation test set of center-slices to mirror the training contrast distribution, namely 50 T2-axial, 7 FLAIR, and 25 T1-weighted slices. For our cross-modality generalization experiment in Section 5.4, we evaluated on 30 randomly-selected, fat-suppressed (FS) knee MRI volumes (sampling the 21st slice) from the FastMRI dataset. All metrics are reported over this evaluation test set. Further preprocessing details are in Appendix A.

4.4. Blinded Radiologist Preference Study

We ran a blinded 3-way forced-choice preference study with three radiologists. For each of 60 cases (24 T1/FLAIR, 36 T2), each radiologist was shown the labeled ground truth and three reconstructions – PaDIS-MRI (ours, patch-based diffusion prior), FastMRI-EDM (whole-image diffusion prior from Aali et al. (2024)), and PICS-L1 (wavelet sparsity prior from Uecker et al. (2013)) – with method identities and ordering fully anonymized and randomized per case. The 60 cases were randomly split into two groups: in 30 cases the learned priors (PaDIS-MRI and FastMRI-EDM) were trained on 25 slices, and in 30 cases they were trained on 500 slices; each group preserved the same contrast mix (18 T2, 12 T1/FLAIR). Each radiologist selected a single "best" reconstruction per case, targeting maximal diagnostic quality. Each image therefore received three independent votes (one per radiologist). We report (i) per-radiologist preference rates for PaDIS-MRI, and (ii) an image-level majority outcome indicating whether $\geq$2 of 3 radiologists chose PaDIS-MRI for that case. We summarize radiologist preference over all 60 images and separately by contrast type (T1/FLAIR vs. T2). We report raw preference proportions as well as Wilson (binomial) 95% confidence intervals (CIs).

5. Experiments

We evaluate PaDIS-MRI and FastMRI-EDM (Aali et al., 2024) across several dimensions, against a convex untrained wavelet-sparse reconstruction PICS-L1 (Uecker et al., 2013). While PICS-L1 does not require any training data, its classical prior struggles to reconstruct fine details and sharp edges at the high undersampling rates we study ($R = 7$), and is outperformed by both diffusion models. We also include model-based deep learning (MoDL) as a supervised baseline (Aggarwal et al., 2019) in our data-efficiency experiments; see Appendix B.4 for our SENSE + CG unrolling and training setup.

We first assess quantitative reconstruction quality as a function of training set size, to validate the data efficiency of the patch-based PaDIS-MRI prior. We then examine reconstruction consistency by visualizing pixel-wise standard deviations across reconstructions from multiple k-space masks of the same slice. Finally, we test robustness to varying k-space undersampling levels and generalization to different contrast types. Across all quantitative experiments, we report Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Normalized Root Mean Square Error (NRMSE) compared to fully-sampled ground truth. Since there is large variation in metrics among individual images, instead of reporting dataset-wide standard deviation in metrics for each method, we report the mean and standard deviation of per-image pairwise differences between PaDIS-MRI and each baseline. These pairwise difference results show that PaDIS-MRI often provides consistent improvement across individual images. We evaluate diagnostic quality via expert visual comparison by a panel of three radiologists, who independently assessed each reconstruction in a blinded comparison study with access to ground truth images, but with algorithms anonymized and presented in random order.

5.1. Data Efficiency

We evaluate the performance of PaDIS-MRI, FastMRI-EDM, and MoDL when trained on dataset sizes $S = \{25, 100, 200, 500, 1000, 2330\}$ k-space slices and tested on our evaluation dataset with a fixed acceleration (undersampling) factor of $R = 7$. Figure 1 and Table 5 in Appendix C.1 summarize our findings. Across training set sizes 25 through 500, PaDIS-MRI consistently outperforms the baselines on all metrics. At larger dataset sizes ($S \geq 1000$), PaDIS-MRI maintains superior PSNR and NRMSE, while FastMRI-

EDM shows marginal improvement in SSIM. This small SSIM advantage is likely caused by the whole-image model's ability to capture global perceptual features when provided sufficient training examples, while PaDIS-MRI's patch-based approach continues to excel at preserving fine details and minimizing overall error (as reflected in better PSNR and NRMSE). PaDIS-MRI exhibits the most significant advantages as dataset size decreases, highlighting the improved data efficiency of the patch-based approach. Importantly, FastMRI-EDM (and PaDIS-MRI) outperforms MoDL, supporting the results in Aali et al. (2024). Full tabular results including performance stratified by contrast type are presented in Appendix C.1.

Figure 2 provides a visual comparison of the reconstruction quality of each method on example slices. In reconstructions from the $S = 25, R = 7$ models, FastMRI-EDM exhibits slightly more blurring and loss of fine structural details, particularly in regions with complex anatomical features. PaDIS-MRI, in contrast, preserves sharper boundaries and finer details. This enhanced data efficiency can be attributed to the patch-based prior's ability to learn localized structural motifs more effectively from limited examples. Figure 6 in the Appendix shows similar reconstructions at $S = 200, R = 7$ where FastMRI-EDM more closely approaches the performance of PaDIS-MRI but still struggles with highly detailed anatomy. At both dataset sizes, MoDL underperforms the diffusion priors. In Figure 2, MoDL's FLAIR reconstruction exhibits pronounced artifacts, likely reflecting reduced generalizability of a supervised prior when the training distribution underrepresents FLAIR.

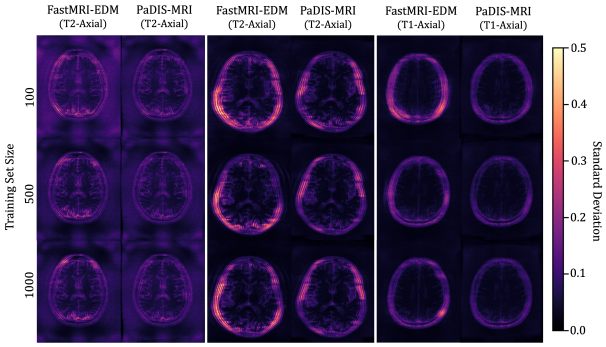

Figure 3: Pixel-wise standard deviations of reconstructions across 10 random k-space sampling masks ($R = 7$) of the same slice. *Top to Bottom*: models trained on 100, 500, and 1000 slices. PaDIS-MRI consistently exhibits lower pixel-wise variance (fewer bright regions) than FastMRI-EDM across training sizes and contrast types, indicating more stable reconstruction.

5.2. Mask-Induced Variability Analysis

We assess reconstruction consistency by evaluating variability across 10 random k-space undersampling patterns (generated with different random seeds) for the same slice. Since the ground truth image is the same across undersampling masks, lower pixel-wise standard deviation indicates more stable reconstruction. We emphasize this is a consistency metric rather than calibrated posterior uncertainty. Such reliability is of particular interest to clinical settings where higher model consistency paired with superior accuracy can enhance medical decision-making.

Figure 3 visualizes these standard deviation maps, with brighter regions indicating higher variability between reconstructions. Across the 100-slice, 500-slice, and 1000-slice training regimes, PaDIS-MRI exhibits noticeably lower pixel-wise standard deviations compared to FastMRI-EDM, with the largest differences occurring at the smaller training dataset sizes. This suggests that decomposing the reconstruction problem into localized patches enables more stable learning of anatomical features, resulting in more reliable and consistent predictions from limited training data.

Table 1: Pixel-wise mean standard deviation (↓) vs. training set size S. PICS-L1 is untrained and thus independent of S.

Algorithm	Training Set Size S					
	25	100	200	500	1000	2330
PaDIS–MRI	**0.070**	**0.064**	**0.063**	**0.062**	**0.064**	**0.062**
FastMRI–EDM	0.075	0.078	0.072	0.069	0.068	0.068
PICS–L1	0.111	0.111	0.111	0.111	0.111	0.111

Table 1 quantifies these observations over our evaluation dataset. Due to computational costs of reconstructing each validation image across 10 random seeds, we randomly subsample a smaller evaluation set of 17 images (10 T2-axial, 5 T1-axial, 2 FLAIR) that preserves the contrast-type distribution of the training dataset. We crop each image to its original 384×320 dimension to ignore deviations in our zero-padded regions, and report the pixel-wise standard deviation averaged across all test images. Again, PaDIS-MRI consistently achieves lower standard deviation values than FastMRI-EDM on average, with differences being most pronounced at smaller dataset sizes. These results further strengthen the case for patch-based diffusion priors in data-scarce settings. A breakdown of pixel-wise standard deviation averages by contrast type can be found in Appendix C.2.

Table 2: Image-level majority preference for PaDIS–MRI (breakdown of the 60 cases).

Cohort	Picks/ Total	Prop.	95% CI
Majority ($\geq$ 2 of 3 votes)	55/60	91.7%*	[81.9, 96.4]%
Breakdown by training size S (for data-driven priors)			
25-slice (n=30)	30/30	100.0%*	[88.7, 100.0]%
500-slice (n=30)	25/30	83.3%*	[66.4, 92.7]%
Breakdown by contrast type			
T1 and FLAIR (n=24)	21/24	87.5%*	[69.0, 95.7]%
T2 (n=36)	34/36	94.4%*	[81.9, 98.5]%
Radiologist 1	50/60	83.3%*	[72.0, 90.7]%
Radiologist 2	58/60	96.7%*	[88.6, 99.1]%
Radiologist 3	47/60	78.3%*	[66.4, 86.9]%

Proportions are Wilson 95% CIs. Stars (*) indicate $p < 0.001$ (one-sided exact binomial) significance. Image-level rows use $p_0 = 7/27 \approx 0.259$; per-radiologist rows use $p_0 = 1/3 \approx 0.333$.

5.3. Radiologist Verification

Results of our blinded radiologist preference comparison are summarized in Table 2. Radiologists consistently and significantly preferred PaDIS-MRI across contrast types and training set sizes, with particularly reliable preference for PaDIS-MRI at small training set sizes. Qualitatively, radiologists reported three criteria guiding their choices: (i) *diagnostic interchangeability* with the fully sampled ground truth image, penalizing motion and undersampling artifacts that would make an image unusable; (ii) *clinical accuracy*, no hallucinated structures and no missed fine anatomy (e.g., subtle gyri or small vessels); and (iii) *sharpness*. PaDIS-MRI met these criteria more consistently than either the whole-image FastMRI-EDM diffusion prior or the classical wavelet-sparse reconstruction, likely because PaDIS-MRI's patch-based prior more effectively learns local structural patterns from small datasets. This qualitative feedback from radiologists aligns with PaDIS-MRI's quantitative advantages and better reconstruction consistency, and supports PaDIS-MRI as a promising tool for accelerated MRI reconstruction with limited training data.

To further quantify the strong preference for PaDIS-MRI we compute the diagnostic agreement amongst radiologists. Despite high raw agreement (0.783), Fleiss' κ is only 0.105 because the expected-by-chance agreement is correspondingly high (0.758) – illustrating the known κ-paradox. We therefore interpret κ as largely uninformative about consensus; high raw agreement is of greater relevance.

5.4. Cross-Domain Generalization and Undersampling Robustness

In clinical settings, models must maintain performance across varying contrast types or modalities that may originate from a distribution different from the one used in training. Additionally, models must demonstrate robustness to different acceleration factors (undersampling rates) during reconstruction. We evaluate PaDIS-MRI in both scenarios.

First, we examine cross-contrast generalization by evaluating models trained exclusively on T2-weighted images when reconstructing T1 and FLAIR images at acceleration factor $R = 7$, for training set sizes $S \in \{100, 500\}$. Table 4 reveals a significant performance gap, with PaDIS-MRI outperforming both baselines in terms of reconstruction fidelity metrics. PaDIS-MRI also shows similar advantages when these brain-trained models are tested on FS knee MRI, with PaDIS-MRI maintaining strong advantages across all metrics. Notably, PaDIS-MRI exhibits a smaller performance drop between these out-of-distribution experiments and its in-distribution performance (Table 5 in Appendix C.1) compared to FastMRI-EDM, which appears to require closer alignment between training and evaluation distributions. These cross-domain experiments suggests that PaDIS-MRI's patch-based approach better captures generalizable anatomical features compared to the whole-image model of FastMRI-EDM, and mirrors similar findings by Hu et al. (2024a) in CT reconstruction.

Second, we assess robustness to varying undersampling ratios by testing models trained with $S = 100$ slices across acceleration factors $R \in \{2, 4, 6, 8, 10\}$. Table 3 demonstrates that while FastMRI-EDM performs comparably at low acceleration ($R = 2$), PaDIS-MRI shows increasingly superior performance as undersampling becomes more severe. This widening performance gap suggests that patch-based priors better preserve fine anatomical details when k-space information becomes severely limited, likely because they leverage consistency of local structures rather than relying on global patterns that become increasingly ambiguous under severe undersampling. Appendix C.2 contains a breakdown of undersampling rate performance by contrast type.

6. Discussion

We extended PaDIS (Hu et al., 2024b) to complex-valued, multi-coil undersampled MRI reconstruction

Table 3: Reconstruction quality metrics across undersampling ratios with $S = 100$ training slices.

R	PaDIS-MRI			FastMRI-EDM			PICS-L1		
	PSNR ↑	SSIM ↑	NRMSE ↓	PSNR ↑	SSIM ↑	NRMSE ↓	PSNR ↑	SSIM ↑	NRMSE ↓
2	40.34	0.898	0.062	**40.39**	**0.902**	**0.061**	35.45	0.748	0.106
4	**37.24**	0.862	**0.086**	36.26	**0.865**	0.096	29.55	0.631	0.205
6	**35.88**	**0.864**	**0.100**	33.87	0.856	0.126	27.38	0.596	0.262
8	**34.42**	**0.870**	**0.118**	31.61	0.846	0.164	26.25	0.582	0.298
10	**32.82**	**0.871**	**0.143**	30.09	0.839	0.194	25.74	0.582	0.315

Paired differences (mean±SD); Δ = PaDIS − EDM (left) and PaDIS − PICS (right).

R	Δ vs EDM			Δ vs PICS		
	ΔPSNR	ΔSSIM	ΔNRMSE	ΔPSNR	ΔSSIM	ΔNRMSE
2	-0.05 ± 0.56	-0.004 ± 0.019	$+0.001 \pm 0.004$	$+4.89 \pm 2.11$	$+0.150 \pm 0.054$	-0.045 ± 0.022
4	$+0.98 \pm 0.84$	-0.003 ± 0.021	-0.010 ± 0.009	$+7.69 \pm 2.39$	$+0.231 \pm 0.063$	-0.118 ± 0.038
6	$+2.00 \pm 1.19$	$+0.007 \pm 0.025$	-0.026 ± 0.018	$+8.50 \pm 2.48$	$+0.267 \pm 0.066$	-0.162 ± 0.052
8	$+2.81 \pm 1.09$	$+0.024 \pm 0.024$	-0.046 ± 0.025	$+8.17 \pm 2.43$	$+0.288 \pm 0.062$	-0.180 ± 0.058
10	$+2.73 \pm 1.26$	$+0.032 \pm 0.024$	-0.051 ± 0.028	$+7.08 \pm 2.51$	$+0.289 \pm 0.058$	-0.173 ± 0.062

Table 4: Cross-Domain generalization: T2-trained models reconstructing T1 and Knee (FS) images at $R = 7$. Arrows indicate direction of improvement. PICS-L1 is untrained.

Anatomy	Model	S	PSNR ↑	SSIM ↑	NRMSE ↓
T1	PaDIS-MRI	100	**33.90**	**0.835**	**0.128**
		500	**35.02**	**0.852**	**0.112**
	FastMRI-EDM	100	31.36	0.784	0.171
		500	31.69	0.780	0.166
	PICS-L1	–	27.73	0.587	0.261
Knee (FS)	PaDIS-MRI	100	**30.55**	**0.712**	**0.182**
		500	**30.92**	**0.716**	**0.175**
	FastMRI-EDM	100	28.58	0.647	0.227
		500	28.11	0.640	0.239
	PICS-L1	–	25.21	0.358	0.342

Paired Δ (mean±SD); Δ = $method_1$−$method_2$. T1: N=32; Knee: N=30.

Anatomy	Comparison	ΔPSNR	ΔSSIM	ΔNRMSE
T1	PaDIS–EDM (100)	$+2.53 \pm 0.74$	$+0.051 \pm 0.031$	-0.043 ± 0.015
	PaDIS–PICS (100)	$+6.17 \pm 2.64$	$+0.248 \pm 0.090$	-0.133 ± 0.059
	PaDIS–EDM (500)	$+3.32 \pm 1.35$	$+0.072 \pm 0.040$	-0.054 ± 0.026
	PaDIS–PICS (500)	$+7.29 \pm 2.40$	$+0.265 \pm 0.090$	-0.148 ± 0.054
Knee	PaDIS–EDM (100)	$+1.97 \pm 0.91$	$+0.065 \pm 0.027$	-0.044 ± 0.018
	PaDIS–PICS (100)	$+5.34 \pm 0.96$	$+0.354 \pm 0.044$	-0.159 ± 0.047
	PaDIS–EDM (500)	$+2.81 \pm 1.12$	$+0.076 \pm 0.033$	-0.064 ± 0.029
	PaDIS–PICS (500)	$+5.71 \pm 0.92$	$+0.358 \pm 0.043$	-0.167 ± 0.047

and observed consistent quantitative improvements in fidelity, reconstruction consistency, cross-domain generalization, and robustness compared to a whole-image diffusion prior FastMRI-EDM (Aali et al., 2024), especially in small-data settings. We then evaluated the accuracy of PaDIS-MRI with a blinded survey in which radiologists consistently found that PaDIS-MRI reconstructions, compared to FastMRI-EDM and a classical sparse-wavelet reconstruction (Uecker et al., 2013), are the most similar to fully sampled reference images and therefore more diagnostically useful.

Future work should examine ways to relax the conditional independence assumption inherent in patch-based models, which may limit their ability to capture long-range anatomical dependencies when large training datasets are available. Hybrid approaches that combine patch-based efficiency with mechanisms for modeling global anatomical structure (such as hierarchical patch representations or attention mechanisms across patches) could potentially preserve the data-efficiency advantages of patch-based learning while better leveraging large datasets when available. Balancing the sample-efficiency gains of local patch priors against the need to capture global anatomical context represents an important direction for scaling patch-based diffusion models to diverse clinical applications.

Limitations. Although PaDIS-MRI has fewer parameters than FastMRI-EDM (55M vs. 65M), it took roughly twice as long to train. We also observed diminishing marginal improvement with dataset size beyond ∼200 slices; this may reflect an earlier ceiling for patch-based priors or limited anatomical diversity from reusing adjacent slices when building larger datasets. These limitations may be addressed by future work to streamline the patch-based training pipeline and adapt it to better leverage large datasets when they are available.

Acknowledgments. We thank Dr. Jonathan Singh (Department of Radiology, Kaiser Permanente) for serving as a blinded reader in our study. This work was supported in part by the NSF Mathematical Sciences Postdoctoral Research Fellowship under award num-

ber 2303178 to SFK. AA is supported by NIH grant R01 HL167974 and ARPA-H contract AY2AX000045.

References

Asad Aali, Marius Arvinte, Sidharth Kumar, and Jonathan I Tamir. Solving inverse problems with score-based generative priors learned from noisy data. In *2023 57th Asilomar Conference on Signals, Systems, and Computers*, pages 837–843. IEEE, 2023.

Asad Aali, Marius Arvinte, Sidharth Kumar, Yamin I. Arefeen, and Jonathan I. Tamir. Robust multi-coil mri reconstruction via self-supervised denoising, 2024. URL https://arxiv.org/abs/2411.12919.

Asad Aali, Giannis Daras, Brett Levac, Sidharth Kumar, Alexandros G Dimakis, and Jonathan I Tamir. Ambient diffusion posterior sampling: Solving inverse problems with diffusion models trained on corrupted data. *International Conference on Learning Representations (ICLR)*, 2025.

Hemant K. Aggarwal, Merry P. Mani, and Mathews Jacob. Modl: Model-based deep learning architecture for inverse problems. *IEEE Transactions on Medical Imaging*, 38(2):394–405, 2019. doi: 10.1109/TMI.2018.2865356.

Finn Behrendt, Debayan Bhattacharya, Julia Krüger, Roland Opfer, and Alexander Schlaefer. Patched diffusion models for unsupervised anomaly detection in brain mri, 2023. URL https://arxiv.org/abs/2303.03758.

Hyungjin Chung and Jong Chul Ye. Score-based diffusion models for accelerated mri, 2022. URL https://arxiv.org/abs/2110.05243.

Hyungjin Chung, Jeongsol Kim, Michael T. Mccann, Marc L. Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems, 2024. URL https://arxiv.org/abs/2209.14687.

Yucheng Fan, Hongyi Liao, Shiyu Huang, Yiqiao Luo, Haoyu Fu, and Hairong Qi. A survey of emerging applications of diffusion probabilistic models in mri. *Meta-Radiology*, page 100082, 2024. doi: 10.1016/j.metrad.2024.100082.

Vijay P. Grover, James M. Tognarelli, Mary M. E. Crossey, I. Jane Cox, Simon D. Taylor-Robinson, and Mark J. W. McPhail. Magnetic resonance imaging: Principles and techniques: Lessons for clinicians. *Journal of Clinical and Experimental Hepatology*, 5(3):246–255, Sep 2015. doi: 10.1016/j.jceh.2015.08.001. PMID: 26628842; PMCID: PMC4632105.

Kelvin G Hollingsworth. Reducing acquisition time in clinical mri by data undersampling and compressed sensing reconstruction. *Physics in Medicine & Biology*, 60(21):R297–R322, Nov 2015. doi: 10.1088/0031-9155/60/21/R297. Epub 2015 Oct 8.

Jason Hu, Bowen Song, Jeffrey A. Fessler, and Liyue Shen. Patch-based diffusion models beat whole-image models for mismatched distribution inverse problems, 2024a. URL https://arxiv.org/abs/2410.11730.

Jason Hu, Bowen Song, Xiaojian Xu, Liyue Shen, and Jeffrey A. Fessler. Learning image priors through patch-based diffusion models for solving inverse problems, 2024b. URL https://arxiv.org/abs/2406.02462.

Mengting Huang, Thanh Nguyen-Duc, Martin Soellradl, Daniel F. Schmidt, and Roland Bammer. Padmr: Patch-based denoising diffusion probabilistic model for magnetic resonance imaging reconstruction. In *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 1–5, 2024. doi: 10.1109/ISBI56570.2024.10635278.

Ajil Jalal, Marius Arvinte, Giannis Daras, Eric Price, Alexandros G. Dimakis, and Jonathan I. Tamir. Robust compressed sensing mri with deep generative priors, 2021. URL https://arxiv.org/abs/2108.01368.

Ona N. Jaspan, Roman Fleysher, and Michael L. Lipton. Compressed sensing mri: a review of the clinical literature. *The British Journal of Radiology*, 88 (1056):20150487, 2015. doi: 10.1259/bjr.20150487.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models, 2022. URL https://arxiv.org/abs/2206.00364.

Saikat Roy, Mahmoud Mostapha, Radu Miron, Matt Holbrook, and Mariappan Nadar. Investigating the feasibility of patch-based inference for generalized diffusion priors in inverse problems for medical images. *arXiv preprint arXiv:2501.15309*, 2025. URL https://arxiv.org/abs/2501.15309.

Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations, 2021. URL https://arxiv.org/abs/2011.13456.

Martin Uecker, Patrick Virtue, Frank Ong, Mark J Murphy, Marcus T Alley, Shreyas S Vasanawala, and Michael Lustig. Software toolbox and programming library for compressed sensing and parallel imaging. In *ISMRM Workshop on Data Sampling and Image Reconstruction*, page 41. ISMRM, 2013.

Zhendong Wang, Yifan Jiang, Huangjie Zheng, Peihao Wang, Pengcheng He, Zhangyang Wang, Weizhu Chen, and Mingyuan Zhou. Patch diffusion: Faster and more data-efficient training of diffusion models, 2023. URL https://arxiv.org/abs/2304.12526.

Jong Chul Ye. Compressed sensing mri: a review from signal processing perspective. *BMC Biomedical Engineering*, 1(1):8, 2019. doi: 10.1186/s42490-019-0008-7.

Jure Zbontar, Florian Knoll, Anuroop Sriram, Tullie Murrell, Zhengnan Huang, Matthew J. Muckley, Aaron Defazio, Ruben Stern, Patricia Johnson, Mary Bruno, Marc Parente, Krzysztof J. Geras, Joe Katsnelson, Hersh Chandarana, Zizhao Zhang, Michal Drozdzal, Adriana Romero, Michael Rabbat, Pascal Vincent, Nafissa Yakubova, James Pinkerton, Duo Wang, Erich Owens, C. Lawrence Zitnick, Michael P. Recht, Daniel K. Sodickson, and Yvonne W. Lui. fastmri: An open dataset and benchmarks for accelerated mri, 2019. URL https://arxiv.org/abs/1811.08839.

Appendix A. Experimental Setup

Datasets. All experiments use fully sampled k-space data from the NYU Langone Health FastMRI dataset (Zbontar et al., 2019), including FLAIR, native T1, pre-contrast T1, post-contrast T1, and axial T2-weighted brain scans. We use the central slice (and additional adjacent slices for larger dataset sizes as discussed below). Following Aali et al. (2024), we first apply the inverse Fourier transform to convert the raw multi-coil k-space data into coil images. We estimate the multi-coil noise covariance from a 30 $\times$ 30 background patch of coil images and perform pre-whitening of the k-space data using the Berkeley Advanced Reconstruction Toolbox (BART) (Uecker et al., 2013). We then normalize the pre-whitened k-space by reconstructing a 24 $\times$ 24 centered ACS region using root-sum-of-squares (RSS) and scaling by its 99th percentile. We estimate coil sensitivities from the normalized, pre-whitened central k-space using BART's ESPIRiT. The images are then zero-padded to 384 $\times$ 384 pixels (N = 384). For our undersampled validation sets, we apply retrospective Cartesian masks at acceleration factors of R $\in$ 2-10, fully sample the same 24 $\times$ 24 ACS region, randomly select remaining phase-encode lines, and pair each masked k-space with the precomputed coil sensitivity maps. Ground truth images are pre-whitened and normalized, and no extra noise is added beyond the measurement noise intrinsic to FastMRI.

For creating training datasets of various sizes, we sampled slices systematically from randomly selected volumes. For smaller datasets (25, 100, and 200 slices), we extracted only the center slice from each volume to maximize data diversity. For 500 slices, we used two adjacent slices (center and one adjacent) from 250 volumes, while for 1000 slices, we selected four adjacent slices (center plus three adjacent) from 250 volumes. The full dataset of 2330 slices is comprised of five slices from each of the 446 training volumes. All volumes were selected randomly from the training pool without weighting by contrast type. There is no overlap between training and validation volumes from the FastMRI dataset (Zbontar et al., 2019).

In total, the training set contains 446 k-space volumes, consisting of 40 FLAIR, 136 T1-weighted (91 post-contrast, 26 pre-contrast, 19 native), and 270 T2-axial. Our validation set contains 266 volumes (distinct from the training volumes) from which we sample an evaluation test set of center-slices to mirror the training contrast distribution, namely 50 T2-axial, 7 FLAIR, and 25 T1-weighted slices. We also use an evaluation set consisting of 30 knee MRI volumes from the FastMRI dataset (each using the 21st slice) for the cross-modality generalization experiment in Section 5.4. All metrics are reported over these evaluation test sets.

Models. The PaDIS-MRI model consists of approximately 55M parameters and was trained with a padding width of 96 pixels on each side. Patches of sizes $P \in \{16, 32, 64\}$ were sampled with probabilities $p \in \{0.2, 0.3, 0.5\}$, respectively. We describe further specific training parameters in Appendix B.1. We benchmark against the FastMRI-EDM model from Aali et al. (2024), a 65M parameter, whole-image diffusion (EDM) model built on the Song-UNet denoiser. Unlike PaDIS-MRI, which trains on zero-padded patches, FastMRI-EDM is evaluated on unpadded, full-resolution images using its original hyperparameters tuned on FastMRI, which align closely with the hyperparameters of PaDIS-MRI (see Appendix B.1). By preserving each method's published training regimen, we ensure a fair "out-of-the-box" comparison against a high-performing baseline. Both models were trained until the unconditional generation quality had visually stabilized. We observed that the PaDIS-MRI model took approximately twice as long as the whole-image FastMRI-EDM model to train. We provide detailed training and inference hyperparameters in Section B.

Appendix B. Training and Inference Details

Algorithm 1 provides detailed pseudocode for our patch-based VE-DPS algorithm that builds on the PaDIS framework (Hu et al., 2024b) and implements the steps described in Section 4.2. It operates primarily in the padded-image space, only cropping x to the original field of view when computing the measurement consistency gradient in k-space. The final returned x_K is also cropped down to the original image dimensions. Note that our implementation of Algorithm 2 follows the implementation of Hu et al. Hu et al. (2024b).

B.1. Training Hyperparameters

The PaDIS-MRI model uses a Song-UNet architecture with standard encoder and decoder components (Song et al., 2021). It uses 128 base feature channels with

Algorithm 1 Patch-based Diffusion Posterior Sampling (`VE DPS`)

Require: Denoiser $D_\theta : \mathbb{R}^{2\times N'\times N'} \times \mathbb{R}_+ \to \mathbb{R}^{2\times N'\times N'}$, positional encoding pos_enc, forward operator $\mathcal{A}$, adjoint $\mathcal{A}^\dagger$, measurement $y \in \mathbb{C}^{N^2}$, steps K, noise bounds $\sigma_{\min}, \sigma_{\max}$, exponent ρ, data weight ζ, pad M, patch size P, inner loops $L = 10$, $N' = N + 2M$

Initialize:
  $x_0 \leftarrow \mathcal{A}^\dagger(y)$ // $x_0 \in \mathbb{C}^{N\times N}$
  $x_0 \leftarrow \mathrm{Pad}(x_0, M)$ // $x_0 \in \mathbb{C}^{N'\times N'}$
Generate positional encodings:
  $x_{\text{grid}}, y_{\text{grid}} \leftarrow \mathrm{linspace}(-1, 1, N')$ // Create normalized coordinate grids
  $x_{\text{pos}} \leftarrow$ reshape and repeat x_{grid} to size $N' \times N'$
  $y_{\text{pos}} \leftarrow$ reshape and repeat y_{grid} to size $N' \times N'$
  pos_enc $\leftarrow [x_{\text{pos}}, y_{\text{pos}}]$
Precompute timesteps:

$$t_k \leftarrow \left(\sigma_{\max}^{1/\rho} + \frac{k}{K-1}(\sigma_{\min}^{1/\rho} - \sigma_{\max}^{1/\rho})\right)^\rho, \quad \alpha_k \leftarrow \frac{1}{2}t_k^2 \quad (k = 0, \ldots, K)$$

for $k = 0$ **to** $K - 1$ **do**
  $x \leftarrow x_k$ // $x \in \mathbb{C}^{N'\times N'}$
  for $j = 1$ **to** L **do**
    Draw $\varepsilon \sim \mathcal{N}(0, I)$ // Complex Gaussian noise
    $\tilde{x} \leftarrow x + t_k\,\varepsilon$ // VE noise injection
    $\tilde{x}_{\text{real}} \leftarrow \mathrm{toRealChannels}(\tilde{x})$ // $\tilde{x}_{\text{real}} \in \mathbb{R}^{2\times N'\times N'}$
    Randomly sample offset $(a, b) \in [0, M-1]^2$ // For patch extraction
    $D_{\text{real}} \leftarrow \mathrm{PatchDenoising}(D_\theta, \tilde{x}_{\text{real}}, t_k, \text{pos_enc}, P, (a, b))$
    $D \leftarrow \mathrm{toComplex}(D_{\text{real}})$ // $D \in \mathbb{C}^{N'\times N'}$
    score $\leftarrow (D - x)/t_k^2$ // Score function
    $\hat{x} \leftarrow \mathrm{Crop}(D_{\text{real}}, M)$ // $\hat{x} \in \mathbb{R}^{2\times N\times N}$
    $r \leftarrow y - \mathcal{A}(\hat{x})$
    SSE $\leftarrow \|r\|^2$
    $g \leftarrow \nabla_x \mathrm{SSE}$ // Compute grad and re-pad back to $g \in \mathbb{C}^{N'\times N'}$
    $x \leftarrow x - (\zeta/\sqrt{\mathrm{SSE}})\, g$ // Measurement consistency update
    if $k < K - 1$ **then**
      $x \leftarrow x + (\alpha_k/2)\,\mathrm{score} + \sqrt{\alpha_k}\,\varepsilon$
    else
      $x \leftarrow x + (\alpha_k/2)\,\mathrm{score}$
    end if
  end for
  $x_{k+1} \leftarrow x$
end for
Return $\mathrm{Crop}(x_K, M)$ // $\in \mathbb{C}^{N\times N}$

multipliers [2,2,2] and a dropout rate of 0.05. Training was performed with a batch size of 4 using the Adam optimizer (learning rate 1e-4, betas [0.9, 0.999]) and was trained using FP32 precision. The FastMRI-EDM model was trained on full-size brain MR images using the same Song-UNet architecture type but with a deeper network structure following Aali et al. (2024). It uses 128 base channels with multipliers [1,1,2,2,2,2,2], a dropout rate of 0.05, and a batch size of 8. It was trained with a lower learning rate of 5e-5 (with the Adam optimizer and same betas as above) because 1e-4 lead to a higher-variance loss curve. Like PaDIS-MRI, we use FP32 precision. Training was performed on a cluster of NVIDIA RTX A6000s, with

Algorithm 2 Patch-based Image Denoising (Hu et al., 2024b)

Require: Neural network D_θ, noisy image $x \in \mathbb{R}^{1\times 2\times N'\times N'}$, noise level t, positional encoding pos_enc, patch size P, sampling offset (a, b)

Initialize:
  output $\leftarrow$ zeros_like(x) // Initialize output tensor
Compute grid points: grid $\leftarrow \{0, P, 2P, \ldots, (n-1)P\}$ where $n = \lceil N'/P \rceil$
Generate patch indices using grid points and offset (a, b):
  indices $\leftarrow \{(i+a, i+a+P, j+b, j+b+P) \mid i, j \in \text{grid}\}$
patches $\leftarrow$ |indices| // Number of patches
Initialize $x_{\text{input}} \in \mathbb{R}^{\text{patches}\times 2\times P\times P}$
Initialize pos_input $\in \mathbb{R}^{\text{patches}\times 2\times P\times P}$
for $i = 0$ **to** patches $- 1$ **do**
  $z \leftarrow$ indices$[i]$ // $z = (z_1, z_2, z_3, z_4)$ are patch coordinates
  $x_{\text{input}}[i] \leftarrow x[0, :, z_1 : z_2, z_3 : z_4]$ // Extract image patch
  pos_input$[i] \leftarrow$ pos_enc$[0, :, z_1 : z_2, z_3 : z_4]$ // Extract positional encodings
end for
denoised_patches $\leftarrow D_\theta(x_{\text{input}}, t, \text{pos_input})$ // Apply denoiser to all patches
for $i = 0$ **to** patches $- 1$ **do**
  $z \leftarrow$ indices$[i]$
  output$[0, :, z_1 : z_2, z_3 : z_4]$ += denoised_patches$[i]$ // Accumulate denoised patches
  output$[0, :, z_1 : z_2, z_3 : z_4]$ −= $x[0, :, z_1 : z_2, z_3 : z_4]$ // Subtract noisy patches
end for
$x_{\text{denoised}} \leftarrow x +$ output // Add residuals to input
Return x_{denoised}

each model trained on a single GPU. As discussed in Section B.3, we train models until the unconditional generation visually stabilizes, which is around checkpoint 650 (roughly 2 days) for FastMRI-EDM and checkpoint 1451 (roughly 4.5 days) for PaDIS-MRI.

Figure 4 illustrates how the patch-based training works. The input to our denoising network is a single complex-valued patch that is represented as a two-channel real-valued tensor, with one channel for the real component and one channel for the imaginary component. Following Hu et al. (2024b), we pass this tensor to the model along with a 2-dimensional, normalized positional encoding to inform the model as to which part of the full image is represented in the patch.

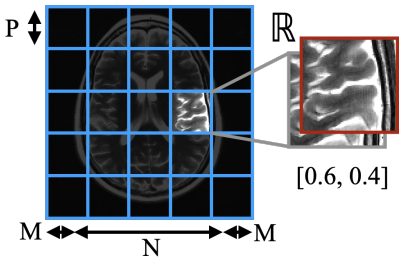

Figure 4: The inputs to our denoising network consist of a two-channel, real-valued representation of a complex-valued patch from the image, along with a positional encoding normalized to [-1, 1].

B.2. Inference Hyperparameters

Our VE-DPS algorithms for both FastMRI-EDM and PaDIS-MRI use a shared noise schedule over $K = 104$ steps, with bounds $\sigma_{min} = 0.003$ to $\sigma_{max} = 10.0$ and no churn ($S_{\text{churn}} = 0$). In the patch-based implementation, we explicitly loop through the 104 noise levels with 10 inner iterations per level, while in the whole-image implementation, we equivalently construct a 1040-step trajectory where each noise level is repeated 10 times. For the measurement update step, we use a data fidelity weight $\zeta = 3.0$, pad by $M = 64$, and use patch size $P = 64$. Inference was also run on our NVIDIA RTX A6000 GPUs. Examples of each model's unconditional generations can be found in Appendix B.3.

Table 5: Combined T1/T2/FLAIR reconstruction results for varying training set sizes at undersampling rate $R = 7$. Top: Mean ± SD across $N = 82$ slices. Bottom: Paired differences (Δ = PaDIS-MRI minus comparison method).

S	PaDIS-MRI			FastMRI-EDM			PICS-L1			MoDL		
	PSNR↑	SSIM↑	NRMSE↓	PSNR↑	SSIM↑	NRMSE↓	PSNR↑	SSIM↑	NRMSE↓	PSNR↑	SSIM↑	NRMSE↓
25	**32.97±3.4**	**.847±.06**	**.143±.05**	30.51±2.9	.817±.07	.185±.05	26.76±2.1	.587±.07	.281±.04	27.86±2.1	.726±.06	.247±.05
100	**35.05±2.9**	**.866±.06**	**.110±.03**	32.62±3.1	.853±.07	.146±.04	26.76±2.3	.587±.08	.281±.04	29.72±1.7	.752±.06	.199±.03
200	**35.19±3.1**	**.866±.07**	**.109±.03**	33.49±3.0	.861±.07	.132±.03	26.76±2.4	.587±.07	.281±.04	30.46±1.7	.782±.05	.184±.03
500	**35.33±2.9**	.869±.06	**.107±.03**	34.34±2.9	**.873±.06**	.119±.03	26.76±2.5	.587±.07	.281±.03	31.46±1.5	.812±.05	.165±.04
1000	**35.19±2.9**	.866±.06	**.108±.03**	34.75±3.0	**.876±.06**	.114±.03	26.76±2.5	.587±.07	.281±.03	32.38±1.3	.833±.05	.147±.02
2330	**35.31±3.0**	.865±.07	**.107±.03**	34.60±2.8	**.873±.06**	.116±.03	26.76±2.5	.587±.07	.281±.03	33.55±1.3	.859±.05	.129±.02

Paired differences (mean±SD over $N = 82$ slices); Δ = PaDIS-MRI minus comparison method.

S	Δ vs EDM			Δ vs PICS			Δ vs MoDL		
	ΔPSNR	ΔSSIM	ΔNRMSE	ΔPSNR	ΔSSIM	ΔNRMSE	ΔPSNR	ΔSSIM	ΔNRMSE
25	2.46±1.4	.031±.03	−.042±.03	6.21±3.4	.261±.07	−.138±.08	5.11±2.9	.121±.06	−.104±.06
100	2.43±1.5	.012±.03	−.036±.03	8.28±2.4	.279±.06	−.171±.05	5.33±1.7	.114±.06	−.089±.03
200	1.70±1.3	.005±.03	−.023±.02	8.42±2.5	.279±.06	−.172±.05	4.73±1.7	.084±.05	−.075±.03
500	0.99±0.8	−.004±.02	−.013±.01	8.57±2.4	.282±.06	−.175±.05	3.87±1.5	.057±.05	−.058±.04
1000	0.44±1.5	−.010±.02	−.006±.03	8.42±2.4	.279±.06	−.173±.05	2.81±1.3	.033±.05	−.039±.02
2330	0.71±0.7	−.008±.02	−.009±.01	8.55±2.5	.279±.07	−.174±.05	1.76±1.3	.006±.05	−.022±.02

Table 6: T1/FLAIR MRI reconstruction results for varying training set sizes at undersampling rate $R = 7$. Top: Mean ± SD across $N = 32$ slices. Bottom: Paired differences (Δ = PaDIS-MRI minus comparison method).

S	PaDIS-MRI			FastMRI-EDM			PICS-L1			MoDL		
	PSNR↑	SSIM↑	NRMSE↓	PSNR↑	SSIM↑	NRMSE↓	PSNR↑	SSIM↑	NRMSE↓	PSNR↑	SSIM↑	NRMSE↓
25	**33.46±2.5**	**.840±.06**	**.138±.04**	31.82±2.6	.803±.07	.162±.04	27.72±2.2	.587±.08	.261±.04	29.07±2.1	.709±.07	.221±.04
100	**36.00±2.4**	**.859±.06**	**.101±.03**	34.01±2.8	.850±.07	.128±.03	27.72±2.2	.587±.08	.261±.04	30.77±2.3	.718±.08	.182±.03
200	**36.31±2.5**	**.861±.06**	**.097±.03**	34.81±2.8	.855±.07	.116±.03	27.72±2.2	.587±.08	.261±.04	31.87±2.5	.760±.08	.160±.03
500	**36.45±2.5**	.865±.06	**.096±.03**	35.43±2.7	**.869±.06**	.108±.03	27.72±2.2	.587±.08	.261±.04	32.88±2.5	.790±.08	.143±.03
1000	**36.31±2.5**	.863±.06	**.097±.03**	35.66±2.7	**.867±.06**	.105±.03	27.72±2.2	.587±.08	.261±.04	33.59±2.6	.812±.08	.132±.02
2330	**36.40±2.5**	.859±.07	**.096±.03**	35.69±2.7	**.866±.06**	.104±.03	27.72±2.2	.587±.08	.261±.04	34.86±2.6	.841±.08	.114±.02

Paired differences (mean±SD over $N = 32$ slices); Δ = PaDIS-MRI minus comparison method.

S	Δ vs EDM			Δ vs PICS			Δ vs MoDL		
	ΔPSNR	ΔSSIM	ΔNRMSE	ΔPSNR	ΔSSIM	ΔNRMSE	ΔPSNR	ΔSSIM	ΔNRMSE
25	1.64±1.1	.037±.03	−.025±.02	5.74±3.1	.254±.09	−.123±.07	4.39±2.1	.131±.05	−.084±.04
100	1.99±1.5	.009±.03	−.028±.02	8.28±2.6	.272±.09	−.160±.05	5.23±1.7	.140±.07	−.081±.03
200	1.50±1.1	.006±.04	−.019±.02	8.59±2.6	.274±.09	−.164±.05	4.45±1.6	.101±.07	−.063±.02
500	1.02±0.9	−.004±.02	−.012±.01	8.77±2.5	.278±.09	−.165±.05	3.57±1.5	.074±.06	−.047±.02
1000	0.65±0.7	−.001±.02	−.008±.01	8.58±2.5	.276±.09	−.164±.05	2.72±1.2	.051±.06	−.035±.02
2330	0.71±0.6	−.000±.02	−.008±.01	8.67±2.6	.273±.09	−.165±.06	1.54±1.2	.019±.06	−.018±.01

B.3. Unconditional Generation Quality

Figure 5 shows sampled unconditional generations for both PaDIS-MRI and FastMRI-EDM, at model convergence. As discussed previously, we trained models by qualitatively evaluating the change in unconditional generation quality over time. The FastMRI-EDM model is able to capture more detailed anatomy in its unconditional generations, whereas PaDIS-MRI tends to generate more uniformly gray ovals with limited detail. We hypothesize that the relatively poorer-looking unconditional generations of PaDIS-MRI may actually make it more robust during inference, or conditional generation. Rather than capturing intricate global features, the relatively simple structure may provide a form of implicit regularization that favors locally consistent features over potentially spurious or overfitted global patterns.

The training dynamics also differ substantially between the two approaches, with PaDIS-MRI requiring approximately 1450 checkpoints to stabilize compared to 650 for FastMRI-EDM. We expect further development of patch-based models may reduce this overhead in training time. Nonetheless, this gap in training time does not explain the PaDIS-MRI model's superior performance at small dataset sizes. Several experiments were done with the FastMRI-EDM at checkpoints of 1200 and 1450; the performance of

Table 7: T2-weighted MRI reconstruction results for varying training set sizes at undersampling rate $R = 7$. Top: Mean ± SD across $N = 50$ slices. Bottom: Paired differences (Δ = PaDIS-MRI minus comparison method).

S	PaDIS-MRI			FastMRI-EDM			PICS-L1			MoDL		
	PSNR↑	SSIM↑	NRMSE↓	PSNR↑	SSIM↑	NRMSE↓	PSNR↑	SSIM↑	NRMSE↓	PSNR↑	SSIM↑	NRMSE↓
25	**32.66±3.2**	**.852±.06**	**.146±.05**	29.67±2.8	.826±.06	.200±.06	26.15±2.0	.587±.07	.294±.04	27.09±1.7	.738±.06	.264±.04
100	**34.44±2.8**	**.869±.05**	**.116±.04**	31.73±2.8	.856±.06	.157±.05	26.15±2.0	.587±.07	.294±.04	29.05±1.9	.772±.07	.211±.03
200	**34.47±2.9**	**.869±.05**	**.117±.04**	32.64±2.8	.865±.06	.141±.04	26.15±2.0	.587±.07	.294±.04	29.56±1.9	.796±.06	.199±.03
500	**34.62±2.7**	.872±.05	**.113±.03**	33.65±2.7	**.875±.05**	.126±.03	26.15±2.0	.587±.07	.294±.04	30.56±2.0	.826±.06	.178±.03
1000	**34.47±2.7**	.869±.05	**.116±.04**	34.17±2.8	**.881±.05**	.120±.04	26.15±2.0	.587±.07	.294±.04	31.60±2.0	.846±.06	.157±.03
2330	**34.62±2.8**	.869±.05	**.114±.04**	33.90±2.7	**.879±.05**	.123±.04	26.15±2.0	.587±.07	.294±.04	32.72±1.9	.872±.06	.138±.02

Paired differences (mean±SD over $N = 50$ slices); Δ = PaDIS-MRI minus comparison method.

S	Δ vs EDM			Δ vs PICS			Δ vs MoDL		
	ΔPSNR	ΔSSIM	ΔNRMSE	ΔPSNR	ΔSSIM	ΔNRMSE	ΔPSNR	ΔSSIM	ΔNRMSE
25	2.99±1.4	.026±.02	−.054±.03	6.51±3.5	.265±.05	−.148±.08	5.57±3.2	.115±.06	−.117±.07
100	2.71±1.4	.014±.02	−.041±.03	8.29±2.3	.283±.04	−.178±.05	5.39±1.7	.098±.04	−.095±.03
200	1.83±1.4	.004±.02	−.025±.02	8.32±2.4	.282±.04	−.178±.05	4.91±1.8	.073±.04	−.082±.03
500	0.98±0.8	−.004±.01	−.013±.01	8.47±2.3	.285±.04	−.181±.05	4.06±1.5	.046±.03	−.064±.02
1000	0.31±1.9	−.013±.02	−.005±.03	8.32±2.3	.281±.04	−.179±.05	2.87±1.3	.022±.03	−.042±.02
2330	0.71±0.8	−.009±.01	−.009±.01	8.47±2.3	.282±.04	−.181±.05	1.89±1.3	−.002±.03	−.025±.02

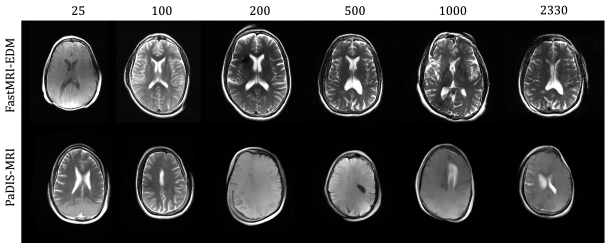

Figure 5: Unconditional image generation quality comparison between FastMRI-EDM (top row) and PaDIS-MRI (bottom row) across different training dataset sizes (25 to 2330 slices, indicated above each column). PaDIS-MRI produces less visually intricate unconditional generations, yet provides a more reliable prior for conditional generation in undersampled MRI reconstruction.

FastMRI-EDM was worse at these checkpoints than its performance at checkpoint 650. Thus, we do not believe that FastMRI-EDM was undertrained relative to PaDIS-MRI.

B.4. Supervised Baseline (MoDL) Implementation

We implement MoDL following Aggarwal et al. (2019). The forward operator uses SENSE with precomputed coil sensitivity maps, and each unrolled iteration alternates (i) a U-Net prior module and (ii) a conjugate-gradient (CG) data-consistency (DC) update. We unroll six DC steps. The DC weight λ is a learned scalar initialized to 0.1. Models are trained on magnitude images with an NRMSE loss for 10 epochs using batch size 1; other optimizer and preprocessing choices match Appendix A. At test time, MoDL is evaluated with the same k-space masks and metrics as our diffusion baselines. This implementation follows that used in Aali et al. (2024).

Appendix C. Additional Experimental Results

C.1. Dataset Size

We present the tabular form of Figure 1 here in Table 5. Across reconstruction quality metrics, PaDIS-MRI outperforms FastMRI-EDM, MoDL, and PICS-L1, especially at small dataset sizes. Moreover, the paired difference statistics confirm that the quantitative improvements of PaDIS-MRI are consistent across images in the dataset.

Tables 6 and 7 provide a contrast-specific breakdown of reconstruction performance across training dataset sizes, complementing the combined results presented in Figure 1 and Table 5. For T1 and FLAIR contrasts (Table 6), PaDIS-MRI consistently outperforms FastMRI-EDM across all dataset sizes, with particularly pronounced advantages at smaller dataset sizes. For T2-weighted images (Table 7), we observe a similar trend with PaDIS-MRI showing substantial improvements at smaller dataset sizes. Interestingly, while PaDIS-MRI maintains higher PSNR and lower NRMSE at larger dataset sizes, FastMRI-EDM achieves marginally better SSIM scores for T2 reconstructions at $S \geq 500$. This suggests that the whole-image approach may better preserve certain

structural features in T2 images when sufficient training data is available, despite its overall lower fidelity in terms of pixel-wise error.

C.2. Mask-Induced Variability

Table 8 presents the pixel-wise standard deviation results broken down by contrast type. For T2-weighted images, PaDIS-MRI consistently demonstrates lower pixel-wise standard deviation than FastMRI-EDM across all dataset sizes, with differences being most pronounced at smaller dataset sizes. For T1 and FLAIR contrasts, the advantage of PaDIS-MRI becomes evident at $S \geq 100$, with substantial reductions in standard deviation compared to FastMRI-EDM. Interestingly, at the smallest dataset size ($S = 25$), FastMRI-EDM shows slightly lower uncertainty for T1/FLAIR contrasts despite its poorer reconstruction quality, perhaps because at this level its errors are dominated by bias rather than variance.

C.3. Robustness

Tables 9 and 10 provide contrast-specific performance across different undersampling ratios for models trained with $S = 100$ slices. For T1 and FLAIR contrasts (Table 9), FastMRI-EDM shows marginally better performance at the lowest acceleration factor ($R = 2$), but PaDIS-MRI quickly establishes dominance as the acceleration factor increases.

A similar pattern emerges for T2-weighted images (Table 10), where FastMRI-EDM performs slightly better at low acceleration factors ($R = 2$) and maintains a higher SSIM at $R = 4$ despite lower PSNR. However, as undersampling becomes more severe ($R \geq 6$), PaDIS-MRI clearly outperforms FastMRI-EDM across all metrics.

These contrast-specific results reinforce our finding that patch-based priors offer improved robustness to severe undersampling, with the advantage becoming increasingly pronounced at higher acceleration factors as the reconstruction relies more heavily on the diffusion prior. This pattern holds true across different anatomical contrasts, though the magnitude of improvement varies between T1 and T2 weighted images.

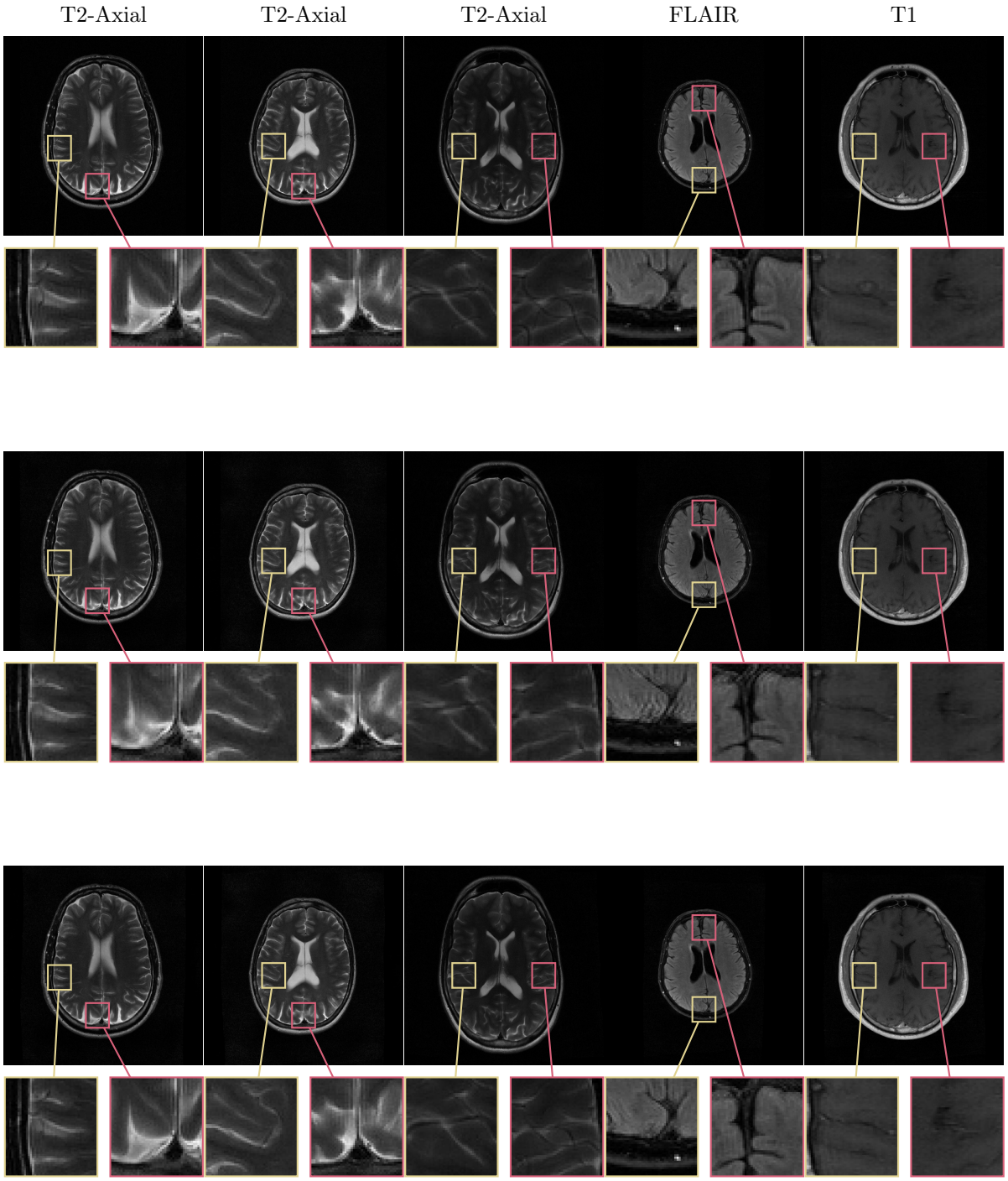

Figure 6: **Dataset Size $S = 200$, Undersampling Rate $R = 7$ Inset Comparisons.** The top row shows the ground truth images. The middle row shows the FastMRI-EDM reconstructions. The bottom row shows the PaDIS-MRI reconstructions, which exhibit slight perceptual improvements at the insets.

Table 8: Pixel-wise standard deviation (↓) by contrast type across dataset sizes at undersampling rate $R = 7$.

Dataset Size (S)	PaDIS-MRI		FastMRI-EDM		PICS-L1	
	T2	T1	T2	T1	T2	T1
25	**0.069**	0.073	0.078	**0.070**	0.114	0.107
100	**0.070**	**0.056**	0.086	0.065	0.114	0.107
200	**0.070**	**0.053**	0.080	0.061	0.114	0.107
500	**0.069**	**0.053**	0.077	0.058	0.114	0.107
1000	**0.072**	**0.054**	0.076	0.056	0.114	0.107
2330	**0.069**	**0.052**	0.076	0.056	0.114	0.107

Table 9: T1/FLAIR reconstruction quality metrics across undersampling ratios with $S = 100$ training slices.

R	PaDIS-MRI			FastMRI-EDM			PICS–L1		
	PSNR ↑	SSIM ↑	NRMSE ↓	PSNR ↑	SSIM ↑	NRMSE ↓	PSNR ↑	SSIM ↑	NRMSE ↓
2	40.92	**0.888**	0.059	**40.94**	0.887	**0.058**	35.85	0.746	0.105
4	**37.88**	**0.852**	**0.082**	37.19	0.851	0.089	30.06	0.618	0.201
6	**36.57**	**0.854**	**0.095**	34.84	0.845	0.116	28.28	0.593	0.243
8	**35.42**	**0.866**	**0.108**	33.10	0.847	0.141	27.29	0.587	0.273
10	**33.95**	**0.871**	**0.129**	31.54	0.845	0.169	26.73	0.593	0.291

Paired differences (mean±SD; N=32); Δ = PaDIS − EDM (left) and PaDIS − PICS (right).

R	Δ vs EDM			Δ vs PICS		
	ΔPSNR	ΔSSIM	ΔNRMSE	ΔPSNR	ΔSSIM	ΔNRMSE
2	-0.02 ± 0.71	$+0.001 \pm 0.025$	$+0.001 \pm 0.005$	$+5.07 \pm 2.54$	$+0.142 \pm 0.070$	-0.046 ± 0.028
4	$+0.70 \pm 0.90$	$+0.001 \pm 0.028$	-0.006 ± 0.008	$+7.82 \pm 2.41$	$+0.234 \pm 0.089$	-0.118 ± 0.041
6	$+1.72 \pm 1.39$	$+0.009 \pm 0.033$	-0.022 ± 0.021	$+8.28 \pm 2.28$	$+0.261 \pm 0.091$	-0.148 ± 0.043
8	$+2.32 \pm 1.25$	$+0.019 \pm 0.032$	-0.033 ± 0.019	$+8.13 \pm 2.40$	$+0.279 \pm 0.084$	-0.166 ± 0.056
10	$+2.41 \pm 1.35$	$+0.026 \pm 0.029$	-0.040 ± 0.026	$+7.21 \pm 2.74$	$+0.278 \pm 0.079$	-0.161 ± 0.062

Table 10: T2 reconstruction quality metrics across undersampling ratios with $S = 100$ training slices.

R	PaDIS-MRI			FastMRI-EDM			PICS-L1		
	PSNR ↑	SSIM ↑	NRMSE ↓	PSNR ↑	SSIM ↑	NRMSE ↓	PSNR ↑	SSIM ↑	NRMSE ↓
2	39.97	0.905	0.063	**40.03**	**0.911**	**0.062**	35.19	0.750	0.107
4	**36.83**	0.869	**0.089**	35.67	**0.874**	0.101	29.23	0.639	0.208
6	**35.44**	**0.870**	**0.103**	33.25	0.863	0.131	26.80	0.598	0.274
8	**33.78**	**0.873**	**0.124**	30.65	0.846	0.179	25.58	0.578	0.314
10	**32.10**	**0.871**	**0.151**	29.17	0.835	0.210	25.11	0.575	0.331

Paired differences (mean±SD; N=50); Δ = PaDIS − EDM (left) and PaDIS − PICS (right).

R	Δ vs EDM			Δ vs PICS		
	ΔPSNR	ΔSSIM	ΔNRMSE	ΔPSNR	ΔSSIM	ΔNRMSE
2	-0.06 ± 0.44	-0.007 ± 0.012	$+0.001 \pm 0.004$	$+4.77 \pm 1.77$	$+0.155 \pm 0.041$	-0.044 ± 0.018
4	$+1.16 \pm 0.75$	-0.006 ± 0.015	-0.012 ± 0.008	$+7.60 \pm 2.38$	$+0.230 \pm 0.038$	-0.119 ± 0.035
6	$+2.19 \pm 1.00$	$+0.007 \pm 0.017$	-0.029 ± 0.015	$+8.64 \pm 2.58$	$+0.271 \pm 0.043$	-0.172 ± 0.056
8	$+3.12 \pm 0.85$	$+0.027 \pm 0.016$	-0.055 ± 0.025	$+8.20 \pm 2.44$	$+0.294 \pm 0.041$	-0.190 ± 0.057
10	$+2.93 \pm 1.16$	$+0.036 \pm 0.019$	-0.059 ± 0.026	$+6.99 \pm 2.35$	$+0.297 \pm 0.037$	-0.180 ± 0.062