

Self-Supervised Adversarial Diffusion Models for Fast MRI Reconstruction

Mojtaba Safari¹, Zach Eidex¹, Shaoyan Pan¹, Richard L.J. Qiu¹, Xiaofeng Yang, PhD^{1,‡}

¹Department of Radiation Oncology and Winship Cancer Institute, Emory University,
Atlanta, GA 30322, United States of America

[‡]Corresponding Author

Author to whom correspondence should be addressed. email: xiaofeng.yang@emory.edu

Abstract

Background: Magnetic resonance imaging (MRI) offers excellent soft tissue contrast essential for diagnosis and treatment, but its long acquisition times can cause patient discomfort and motion artifacts.

Purpose: To propose a self-supervised deep learning-based compressed sensing MRI (DL-based CS-MRI) method named “Self-Supervised Adversarial Diffusion for MRI Accelerated Reconstruction (SSAD-MRI)” to accelerate data acquisition without requiring fully sampled datasets.

Materials and Methods: We used the fastMRI multi-coil brain axial T₂-weighted (T₂-w) dataset from 1,376 cases and single-coil brain quantitative magnetization prepared 2 rapid acquisition gradient echoes (MP2RAGE) T₁ maps from 318 cases to train and test our model. Robustness against domain shift was evaluated using two out-of-distribution (OOD) datasets: multi-coil brain axial postcontrast T₁-weighted (T₁c) dataset from 50 cases and axial T₁-weighted (T₁-w) dataset from 50 patients. Data were retrospectively subsampled at acceleration rates $R \in \{2\times, 4\times, 8\times\}$. SSAD-MRI partitions a random sampling pattern into two disjoint sets, ensuring data consistency during training. We compared our method with ReconFormer Transformer and SS-MRI, assessing performance using normalized mean squared error (NMSE), peak signal-to-noise ratio (PSNR), and structural similarity index (SSIM). Statistical tests included one-way analysis of variance (ANOVA) and multi-comparison Tukey’s Honesty Significant Difference (HSD) tests.

Results: SSAD-MRI preserved fine structures and brain abnormalities visually better than comparative methods at $R = 8\times$ for both multi-coil and single-coil

datasets. It achieved the lowest NMSE at $R \in \{4\times, 8\times\}$, and the highest PSNR and SSIM values at all acceleration rates for the multi-coil dataset. Similar trends were observed for the single-coil dataset, though SSIM values were comparable to ReconFormer at $R \in \{2\times, 8\times\}$. These results were further confirmed by the voxel-wise correlation scatter plots. OOD results showed significant ($p \ll 10^{-5}$) improvements in undersampled image quality after reconstruction.

Conclusions: SSAD-MRI successfully reconstructs fully sampled images without utilizing them in the training step, potentially reducing imaging costs and enhancing image quality crucial for diagnosis and treatment.

keywords: k -space sampling, fastMRI, accelerated MRI, reconstruction, adaptive partitioning

1 Introduction

Magnetic Resonance Imaging (MRI) provides excellent soft tissue contrast, playing a vital role in diagnosis, treatment, and follow-up. However, prolonged acquisition times can lead to patient discomfort and increase the likelihood of motion artifacts, which compromise image quality. Additionally, the requirement for highly sampled k -space data to achieve high-resolution MRI images inherently extends acquisition times, reducing imaging throughput and limiting patient access to MRI services. This issue is particularly significant in low- and middle-income countries, where the recent Lancet Oncology Commission highlighted severe shortages of MRI and other medical imaging devices, potentially resulting in 2.5 million preventable deaths worldwide [1, 2]. Globally, only about seven MRI scanners are installed per million people, largely due to the high costs of installation, operation, and maintenance. Therefore, developing techniques that can accelerate MRI acquisition without sacrificing image quality is crucial for improving accessibility, reducing operational costs, and enhancing patient care worldwide [3].

The MRI acquisition can be accelerated by reducing the sampled k -space data, but this is limited by the Nyquist criteria. Compressed sensing (CS) and parallel imaging (PI) techniques aim to recover fully sampled images from under-sampled images by exploiting data in a sparse transformed space and redundant data acquired using uncorrelated radiofrequency coils, respectively [4, 5]. However, at high acceleration rates, PI and CS methods suffer from noise amplification [6] and residual artifacts [7], respectively.

Deep learning (DL) algorithms have been extensively used to reconstruct accelerated high-resolution MRI images. DL-based compressed sensing MRI (CS-MRI) approaches often blend data-driven learning with physics-guided modeling [8]. Data-

driven methods focus on learning mappings from under-sampled k -space data to fully-sampled images or k -space representations [9, 10, 11, 12], while physics-guided methods incorporate knowledge of the MRI acquisition process, such as coil sensitivity profiles and under-sampling patterns, to enforce data consistency and address the ill-posed nature of the inverse problem.

Many state-of-the-art methods integrate both approaches, leveraging the strengths of data-driven learning and physics-based constraints. For instance, unrolled optimization methods [13, 14, 15] mimic iterative reconstruction algorithms by alternating between data consistency enforcement and learned regularization in a fixed number of iterations. These methods often incorporate data consistency layers (DC layers) within deep neural networks to ensure that the reconstructed images adhere to the acquired k -space measurements [16, 17, 18].

Classic works such as Deep ADMM-Net [19] and the variational network [20] exemplify this integration by unrolling iterative algorithms and embedding physics-based constraints within DL frameworks. Similarly, approaches like deep density priors [21] and deep Bayesian estimation [22, 23] combine data-driven learning with probabilistic modeling to capture complex image priors while respecting the underlying physics of MRI acquisition.

These methods are typically trained under supervised frameworks where the reference fully sampled images were utilized to train a model. However, obtaining the fully sampled images might be impractical in imaging scenarios such as cardiovascular MRI due to excessive involuntary movements, or diffusion MRI with echo planar imaging due to quick T_2^* signal decay [24]. Additionally, acquiring high-resolution anatomical brain MRI images can be prohibitively long.

In this study, we propose a self-supervised adversarial diffusion model to reconstruct fully sampled images without requiring them in the training step. Our proposed SSAD-MRI model is based on an adversarial mapper to reconstruct fully sampled MRI images. Our proposed method was evaluated using both single-coil and multi-coil MRI data, as well as two out-of-distribution (OOD) datasets. Our method leverages a recently proposed ReconFormer transformer [25] as a generator and is compared with two state-of-the-art models. Our contributions are as follows:

- To our knowledge, SSAD-MRI is the first study proposing a self-supervised method using an adversarial mapper.
- The proposed method performs the backward diffusion process in smaller steps that improve sampling efficiency.
- The proposed method’s robustness against domain shift was evaluated at the test time,

- To our knowledge, It is the first self-supervised method aimed at reconstructing fully sampled quantitative magnetization prepared 2 rapid acquisition gradient echoes (MP2RAGE) T_1 map,

2 Materials and Methods

2.1 Compressed sensing MRI

Let $Y \in \mathbb{C}^N$ denote the observed subsampled k -space measurement corresponding to the subsampled image $y \in \mathbb{C}^N$, while $x \in \mathbb{C}^M$ represents the unobserved fully sampled data. Hereafter, we adopt the convention that lowercase letters denote data in the image domain, whereas uppercase letters represent data in the Fourier (k -space) domain. The compressed sensing formulation is expressed as follows:

$$Y = \mathcal{A}_\Omega x + \delta \quad (1)$$

where $\delta \in \mathbb{C}^N$ is the additive acquisition noise and $\mathcal{A}_\Omega \in \mathbb{C}^{N \times M} \rightarrow \mathbb{C}^N$ give $N \ll M$ represents the encoding operator. \mathcal{A}_Ω is composed of a coil sensitivity map \mathcal{S} , a Fourier transform \mathcal{F} , and a sampling map with the specified pattern Ω controlling the acceleration rate (R). The mathematical expression for the encoding operator is $\mathcal{A}_\Omega = \Omega \odot \mathcal{F} \odot \mathcal{S}$. The MRI reconstruction is formulated as an unconstrained optimization problem as follows [15, 18, 26],

$$\arg \min_{x, \varphi} \| x - f_{\text{SSAD-MRI}}(y | \varphi) \|_2^2 + \lambda \| Y - \mathcal{A}_\Omega x \|_2^2 \quad (2)$$

where $f_{\text{SSAD-MRI}}$ represents our proposed model parametrized by φ . The first and second terms represent the regularization and data consistency, and $\lambda > 0$ is a scalar regularization weight that balances between the data consistency and regularization terms.

2.2 Diffusion model

The diffusion model inspired by nonequilibrium thermodynamics aims to approximate complex and intractable distributions with a tractable one like normal Gaussian [27]. It consists of two process: the forward process and the reverse process.

Forward process:

The forward process utilizes a noise scheduler to add Gaussian noise to the noise-free y_0 using a first-order Markov process $q(y_t|y_{t-1})$ in a large number of steps T ,

eventually converting y_T to a normal multivariate Gaussian $y_T \sim \mathcal{N}(y_T; \mathbf{0}, \mathbf{I})$. This step employs the first-order Markov process to calculate $q(y_t|y_{t-1})$ as follows [28]:

$$\begin{aligned} y_t &= \sqrt{1 - \beta_t}y_{t-1} + \sqrt{\beta_t}\epsilon \\ q(y_t | y_{t-1}) &= \mathcal{N}(y_t; \sqrt{1 - \beta_t}y_{t-1}, \beta_t\mathbf{I}) \end{aligned} \quad (3)$$

where $\epsilon \sim \mathcal{N}(\epsilon; \mathbf{0}, \mathbf{I})$ and $\beta_t \in (0, 1)$ with $\beta_1 = 10^{-4}$ is the noise variance. Assuming additive Gaussian noise, sampling y_t in an arbitrary step t can be calculated in closed form as follows:

$$q(y_t | y_0) = \mathcal{N}(y_t; \sqrt{\bar{\alpha}_t}y_0, (1 - \bar{\alpha}_t)\mathbf{I}) \quad (4)$$

where $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{s=1}^t \alpha_s$.

Reverse process:

The reverse process gradually learns to remove the added noise in y_T to recover noise-free y_0 . This process of training a network p_φ to generate y_0 from Gaussian noise y_T . The reverse process will follow the forward process trajectories but in the reverse direction for small β values as follows [27, 29, 30]:

$$\arg \min_{\varphi} \sum_{t=2}^T \mathbb{E}_{q(y_t|y_0)} [D_{\text{KL}}(q(y_{t-1} | y_t, y_0) \| p_\varphi(y_{t-1} | y_t))] \quad (5)$$

This is a denoising matching term where it learns the desired denoising transition step $p_\varphi(y_{t-1} | y_t)$ as an approximator to tractable, ground-truth denoising transition step $q(y_{t-1} | y_t, y_0)$ given in (6), where it is modeled as a Gaussian.

$$\begin{aligned} q(y_{t-1} | y_t, y_0) &= \mathcal{N}(y_{t-1}; \mu_q(y_t, y_0), \sigma_q(t)\mathbf{I}) \\ \mu_q(y_t, y_0) &= \frac{\sqrt{\bar{\alpha}_t}(1 - \bar{\alpha}_{t-1})y_t + \sqrt{\bar{\alpha}_{t-1}}y_0}{1 - \bar{\alpha}_t} \\ \sigma_q(t) &= \frac{(1 - \alpha_t)(1 - \bar{\alpha}_{t-1})}{1 - \bar{\alpha}_t} \end{aligned} \quad (6)$$

Furthermore, all the α terms are frozen at each timestep, it was shown that the loss function given in (5) becomes as follows:

$$\mathbf{L} = \arg \min_{\varphi} \frac{1}{2\sigma_q^2(t)} [\| \mu_q(y_t, y_0) - \mu_\varphi(y_t) \|_2^2] \quad (7)$$

where $\mu_\varphi(y_t)$ is the estimated average recovered image as follows:

$$\mu_\varphi(y_t) = \frac{\sqrt{\bar{\alpha}_t}(1 - \bar{\alpha}_{t-1})y_t + \sqrt{\bar{\alpha}_{t-1}}y_\varphi(y_t)}{1 - \bar{\alpha}_t} \quad (8)$$

where $y_\varphi(y_t)$ is parameterized by our DL model to recover y_0 from noisy image y_t at a given step t .

We employed an adversarial mapper to implicitly capture the conditional distribution for the reverse process steps. The generator $f_{\text{SSAD-MRI}}$ is used to sample $\hat{y}_t \sim p_\varphi(y_t | y_{t+k})$. At the same time, a discriminator \mathcal{D}_θ differentiates between samples \hat{y}_t and actual sample y_t sampled from the true denoising distribution $q(y_t | y_{t+k})$. Our discriminator was coupled with a gradient penalty to improve learning [16, 30]:

To improve the efficiency of modeling the reverse diffusion process, we propose an adversarial mapper that implicitly learns the conditional distribution between time steps without the need for an explicit formulation. This mapper consists of a generator and a discriminator that train simultaneously. The generator, denoted as $f_{\text{SSAD-MRI}}$, is designed to estimate the conditional distribution $p_\varphi(y_t | y_{t+k})$ by generating samples \hat{y}_t that closely approximate the true intermediate states y_t , given the future states y_{t+k} . This process effectively facilitates the mapping from y_{t+k} back to y_t .

The discriminator, represented as D_θ , is trained to differentiate between the outputs produced by the generator, denoted as \hat{y}_t , and the true samples y_t drawn from the denoising distribution $q(y_t | y_{t+k})$. Its purpose is to evaluate the authenticity of the generated samples and provide essential feedback to the generator, thereby improving its capability to produce realistic approximations of y_t .

The adversarial training process is formulated as a minimax optimization between the generator and the discriminator. The generator aims to produce samples that the discriminator cannot distinguish from real data, while the discriminator strives to accurately classify the generated samples as synthetic and the true samples as real. To stabilize the training and promote convergence, we incorporate a gradient penalty term into the discriminator’s loss function. This penalty enforces smoothness in the discriminator’s output with respect to its input. The gradient penalty is formulated following techniques proposed in recent studies [16, 30], thereby bolstering learning efficiency and improving the model’s robustness.

$$\begin{aligned} L_D = & \sum_{t \geq 0} (\mathbb{E}_{q(y_0, y_t)} \mathbb{E}_{q(y_{t+k}|y_t)} [-\log(\mathcal{D}_\theta(y_t, y_{t+k}))]) \\ & + \mathbb{E}_{q(y_{t+k})} \mathbb{E}_{\mathcal{N}_\varphi(\mu_\varphi, \sigma_\varphi)} [-\log(1 - \mathcal{D}_\theta(\hat{y}_t, y_{t+k}))] \\ & + \mathbb{E}_{q(y_0, y_t)} \mathbb{E}_{q(y_{t+k}|y_t)} \left[\frac{1}{2} \|\nabla_{y_t} \mathcal{D}_\theta(\hat{y}_t, y_{t+k})\|_2^2 \right] \end{aligned} \quad (9)$$

where $\mathcal{N}_\varphi(\mu_\varphi, \sigma_\varphi)$ is our generator parameterized by φ to reconstruct mean and vari-

ance. The generator loss becomes:

$$L_G = \sum_{t \geq 0} \mathbb{E}_{q(y_{t+k})} \mathbb{E}_{\mathcal{N}_\varphi(\mu_\varphi, \sigma_\varphi)} [-\log(\mathcal{D}_\theta(y_t, y_{t+k}))] \quad (10)$$

2.3 Proposed self-supervised framework and training details

Acquiring fully sampled data can be impractical due to constraints such as voluntary and involuntary motions, lengthy acquisition times, and signal decay. These constraints hinder the application of supervised DL-based CS-MRI approaches. Thus, we proposed a self-supervised approach that randomly divided the sampling pattern Ω into two sets \aleph and Υ as given in (11). These two sets have no elements in common except the center of k -space.

$$\Omega = \aleph \vee \Upsilon \quad (11)$$

We used an under-sampling pattern \aleph to train our proposed model and define our DC layer as follows.

$$\hat{Y}_t^{\text{SSAD-MRI}}(k) = \begin{cases} \hat{Y}_t^{\text{SSAD-MRI}}(k), & \text{if } k \in \aleph \\ X(k), & \text{if } k \notin \aleph \end{cases} \quad (12)$$

where capital letters refer to the Fourier transform of the corresponding parameters. In other words, our method updates the k -space lines that were under-sampled and keeps the original k -space lines that were not sampled during image acquisition.

Finally, our proposed method defined the loss function given in (7) using a sampling pattern Υ as follows:

$$\mathbf{L}^\Upsilon = \arg \min_{\varphi} \frac{1}{2\sigma_q^2(t)} \frac{(1 - \alpha_t)^2}{(1 - \bar{\alpha}_t)\alpha_t} [\| \epsilon - \hat{\epsilon}_\varphi^\Upsilon(y_t) \|_2^2] \quad (13)$$

The final loss function is composed of discriminator loss, generator loss, and reconstruction loss as follows:

$$L_{\text{final}} = \mathbf{L}^\Upsilon + \lambda(L_D + L_G) \quad (14)$$

where $\lambda = 0.1$ was used to scale and control the ratio between the losses. Our proposed method used the indices specified by \aleph to reconstruct k -space at indices given by Υ . Figure 1 illustrates the flowchart of our proposed method. Then the loss function was calculated at the location indicated by the Υ pattern. In other words, our self-supervised network was trained to decrease the discrepancy between the predicted images y^\aleph and the acquired measurement y^Υ that was not seen in the

training. At the inference step, the test unseen under -sampled data using a sampling pattern Ω was used to reconstruct the fully sampled data.

This self-supervised scenario is similar to the cross-validation concept used to reduce bias and the likelihood of overfitting. Cross-validation partitions the dataset into at least two sets where one set is used to train a model and another set is used to evaluate the model. However, unlike cross-validation in machine learning, which performs partitioning once for all data, our method performs random partitioning per image slice.

In the inference step, we did not start from complete noise. Instead, we set the Gaussian noise covariance to 0.1 and then used them directly as initial inputs. This approach has been previously employed to reconstruct high-resolution MR images from under-sampled k -space data and remove MRI motion artifacts [31, 32].

We used the original implementation of the recently proposed transformer, named ReconFormer*, as a generator. It incorporates the pyramid structures, enabling scale processing at each pyramid unit, while the globally columnar structure maintains high-resolution information [25]. The discriminator consisted of four convolution layers with kernel size (3×3) and padding one. Each convolution layer was followed by a ReLU activation layer and a batch normalization layer.

Our proposed method was trained using the Adam optimizer with a learning rate of 2×10^{-4} to minimize the loss function, employing a batch size of four over 25 epochs. Training was performed on our server using a single NVIDIA A100 GPU with CUDA Toolkit 12.2 and the PyTorch framework version 2.1.2 [33]. The total training time was approximately 175 hours. The inference time for reconstructing a single image slice was about 14 seconds. While we used an A100 GPU for training, the model can also be trained on GPUs with less memory, such as an NVIDIA RTX 3090, by adjusting the batch size to accommodate the available GPU memory.

2.4 Sampling masks

We employed a 1d random sampling pattern Ω where the k -space center was excluded from sampling. The center fraction excluded from sampling was set to 4% of the k -space lines in the horizontal direction. Our proposed SSAD-MRI divides the acquired sampling mask Ω into two disjoint sets \aleph and Υ with except four k -space lines at the center of dataset. Figure S1 in the supplementary document presents the Ω , \aleph , and Υ sampling masks across different values of $\rho \in 0.3, 0.5, 0.7$ with a fixed acceleration rate of $R = 4\times$.

The sampling pattern Υ was randomly sampled for each different slice to train the model. Therefore, our subsampled data using Υ will be able to simulate the ghosting

*<https://github.com/guopengf/ReconFormer>

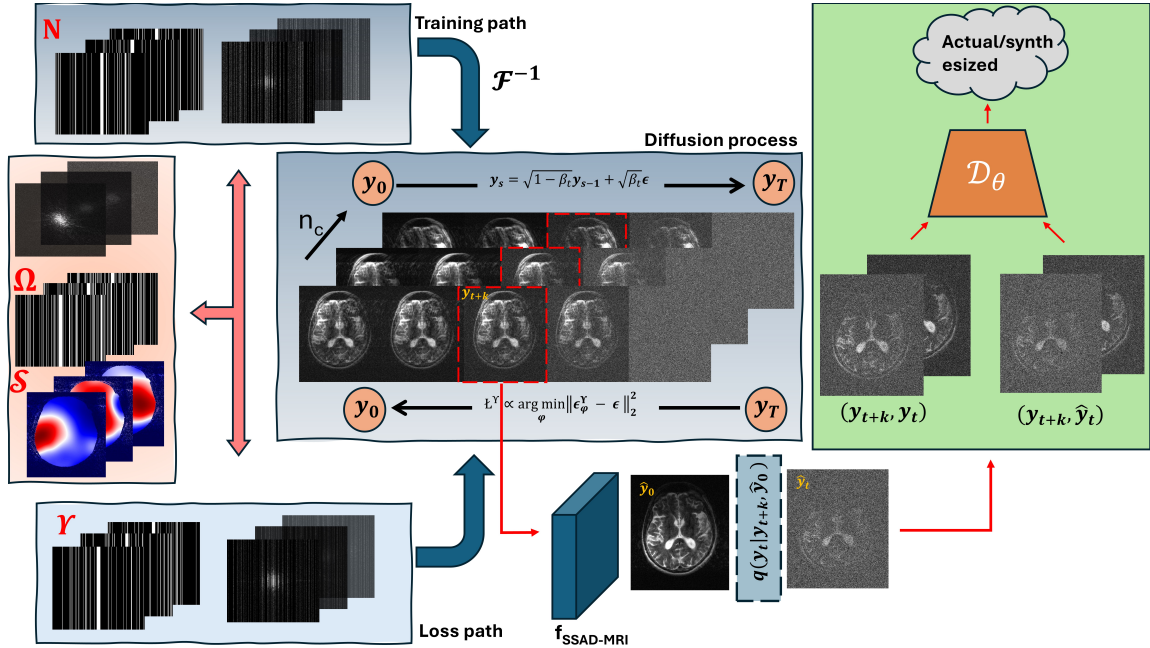


Figure 1: The flowchart of our proposed self-supervised adaptive diffusion model is illustrated. The coil sensitivity \mathcal{S} , random sampling mask Ω , and k -space were illustrated as inputs. Firstly, the random sampling pattern Ω is divided into two non-overlapping sampling masks \aleph and Υ that were used in the training path and loss path as given in (13), respectively. Then, the adversarial mapper was trained as given in (9) and (10) using the data sampled in step $t + k$. The sample y_{t+k} used to recover \hat{y}_0 where was used to calculate \hat{y}_t in a given step t Equation (4).

that is present in zero-filled datasets retrospectively subsampled using Ω masks. We investigated a uniformly random selection among elements of Ω to create subsampled patterns \aleph and Υ . The sampling ratio $\rho = \frac{|\aleph|}{|\Upsilon|} \in \{0.3, 0.5, 0.7\}$ was used to train and test the proposed model, where $|\bullet|$ is the number of elements. In other words, $\rho = 0.3$ means 30% of Ω was randomly assigned to \aleph to train the model and the rest were assigned to Υ sampling pattern to define loss function as given in Equation (13).

2.5 Dataset

We utilized publicly available single-coil and multi-coil brain MRI datasets. Both datasets were retrospectively under-sampled using a random sample provided in the fastMRI database, with acceleration rates (R) equal to 2, 6, and 8 [34, 35].

2.5.1 Single-coil data:

The brain MPI-Leipzig Mind-Brain-Body dataset [36, 37] was used to train and test our proposed model. We utilized high-resolution MP2RAGE T1 maps of 318 patients that were split into two non-overlapping sets: training (250 patients) and testing (68 patients). The sagittal acquisition orientation of volumetric MP2RAGE T₁ maps with 176 slices were acquired with the imaging parameters: T_R = 5000 ms, T_R = 2.92 ms, T_R = 700 ms, T_I₂ = 2500 ms, FA₁ = 4°, FA₂ = 5°, pre-scan normalization, echo spacing = 6.9 ms, bandwidth = 240 Hz/pixel, FOV = 256 mm, voxel size = 1 mm isotropic, GRAPPA acceleration factor = 3, slice order = interleaved, duration = 8 min 22 s.

2.5.2 Multi-coil data:

The brain datasets used in the preparation of this article were obtained from the NYU fastMRI Initiative database[†] that was approved by the NYU School of Medicine Institutional Review Board [34, 35]. We utilized T₂-weighted (T₂-w) images of 1051 patients’ data for training and 325 patients’ unseen data for testing. In addition, we evaluated the robustness of our model to domain shift using two out-of-distribution (OOD) unseen datasets, T₁-weighted (T₁-w) and postcontrast T₁-w (T1c), using 50 patients’ data for each MRI sequence.

The sensitivity maps \mathcal{S} were generated from 24 × 24 center of k -space using ES-PIRiT[‡] [38] with a kernel size 6 × 6 and the calibration matrix and eigenvalue decomposition threshold 0.02 and 0.95, respectively.

2.6 Quantitative analysis

To evaluate the performance of the proposed SSAD-MRI model, we compared it against two benchmark models: SS-MRI [39] and ReconFormer [25], which were trained under our proposed sampling approach.

Three quantitative metrics were employed: normalized mean square error (NMSE), peak signal-to-noise ratio (PSNR), and structural similarity index (SSIM), utilizing the PyTorch Image Quality library[§] [40]. The NMSE compared reconstructed \hat{y} with ground truth y as follows:

$$\text{NMSE}(y, \hat{y}) = \frac{\|y - \hat{y}\|_2^2}{\|y\|_2^2} \quad (15)$$

[†]fastmri.med.nyu.edu

[‡]<https://mrirecon.github.io/bart/>

[§]<https://piq.readthedocs.io/en/latest/index.html>

where $\|\bullet\|_2$ is the squared Euclidean distance. Lower NMSE values indicate better image reconstruction. However, it might be in favor of blurry images [34].

The PSNR defined below utilizes a logarithmic scaling that makes the quantification results more aligned with human perception [41].

$$\text{PSNR}(y, \hat{y}) = \frac{y_{\max}^2}{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (16)$$

where y_{\max} is the maximum signal intensity of ground truth images. Higher PSNR indicates better image reconstruction.

The SSIM quantifies the structural similarity between the reconstructed \hat{y} and ground truth y is defined as follows:

$$\text{SSIM}(y, \hat{y}) = \frac{(2\mu_y\mu_{\hat{y}} + c_1)(2\sigma_{y\hat{y}} + c_2)}{(\mu_y^2 + \mu_{\hat{y}}^2 + c_1)(\sigma_y^2 + \sigma_{\hat{y}}^2 + c_2)} \quad (17)$$

where μ_y and $\mu_{\hat{y}}$ are the average voxel values of y and \hat{y} images, σ_y $\sigma_{\hat{y}}$ are the variance, and $\sigma_{y\hat{y}}$ is the covariance between y and \hat{y} images. The constants c_1 and c_2 stabilize the division, we used $c_1 = 0.01y_{\max}$ and $c_2 = 0.03y_{\max}$. SSIM ranges between -1 and +1, with the best similarity achieved by an SSIM equal to +1.

2.7 Statistical analysis

The quantitative metrics were compared using one-way analysis of variance (ANOVA) to evaluate the null hypothesis that the mean values of each method were the same. The differences with $p < 0.05$ was considered statistically significant. Additionally, a multi-comparison Tukey’s honestly test difference (HSD) was performed to evaluate pairwise differences between the methods, with $p < 0.05$ indicating statistical significance.

We reported the average values of the quantitative metrics. In addition, we calculated the 95% confidence intervals (CIs) on the average values using the percentile bootstrap method (with $n = 10000$ iterations) with the bias-adjusted and accelerated bootstrap method [42].

3 Results

In this section, we present the comprehensive evaluation of our proposed SSAD-MRI model for MRI reconstruction. The results are organized to highlight the model’s performance across different datasets and under various conditions.

We assess the within-domain multi-coil T_2 -w reconstruction and single-coil MP2RAGE T_1 map reconstruction capabilities of SSAD-MRI using both qualitative and quantitative metrics. The model’s ability to preserve fine structures and reduce artifacts is demonstrated through visual comparisons and statistical analysis. The impact of loss mask partitioning on the performance of the proposed method is demonstrated.

We also evaluate the robustness of SSAD-MRI against domain shifts using OOD datasets. This analysis illustrates the model’s adaptability to new and unseen data.

3.1 Within-domain reconstruction

Our proposed SSAD-MRI was evaluated for within-domain reconstruction at $R = 2\times$, $4\times$, and $8\times$ with a partitioning rate $\rho = 0.5$. We conducted both qualitative and quantitative comparisons with the ReconFormer Transformer and SS-MRI models.

3.1.1 Qualitative Results

Multi-coil dataset: Figures 2 and S2 present the qualitative results of within-domain T_2 -w multi-coil images for two subjects. Figure 2(a) shows results for a healthy volunteer, highlighting that our method recovers more details, particularly in the gold boxed regions, compared to the ReconFormer model trained with our framework. Difference maps in Figure 2(b) illustrate that our method better preserves gray matter and edges, indicated by the gold and red boxes.

Figure S2 displays results for a patient with brain abnormalities. Our proposed method could recover the abnormality boundary shown by a red box close to the ground truth. Furthermore, our method reconstructed details better (gold boxes) compared with the SS-MRI method.

Single-coil dataset: Figure 3 presents the qualitative results for high-resolution single-coil MP2RAGE T_1 quantitative map for a healthy volunteer at $R = 2\times$, $4\times$, and $8\times$ with a partitioning rate $\rho = 0.5$. Our method could preserve the small structure shown by gold boxes as well as the Putamen and Caudate nuclei shown by red boxes with better spatial contrast than the comparative models, as shown in Figure 3(a) and the difference map shown in Figure 3(b).

3.1.2 Quantitative Results

Tables 1 and 2 summarize the quantitative metrics for multi-coil and single-coil datasets, respectively, at different acceleration rates. The ANOVA test indicated statistically significant differences ($p \ll 10^{-5}$) between the average values of the

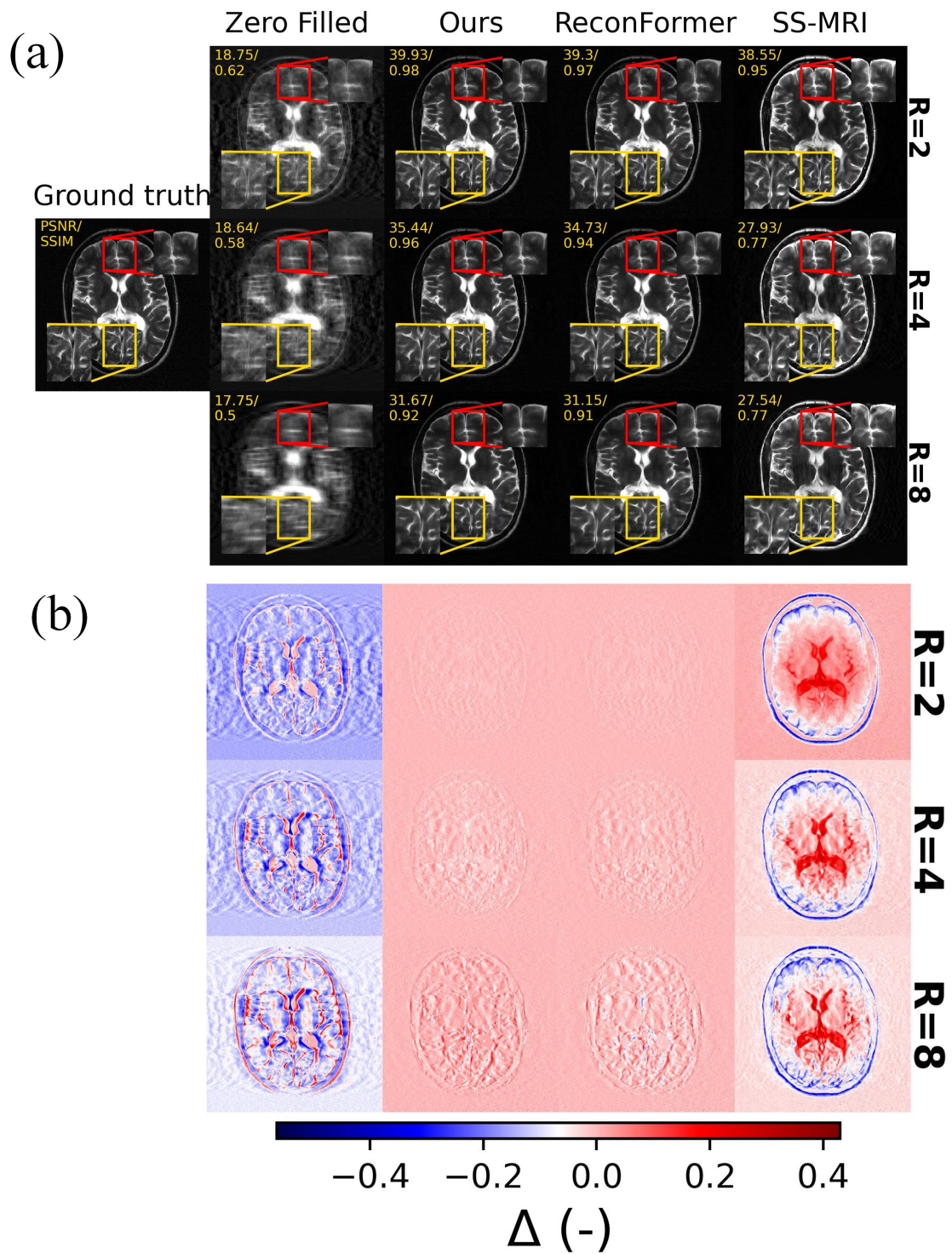


Figure 2: Within-domain axial T_2 -w image reconstruction at $R = 2\times, 4\times,$ and $8\times$ and $\rho = 0.5$ are illustrated. (a) illustrates the results for a healthy subject and (b) illustrates the difference map between the reconstructed and ground truth images.

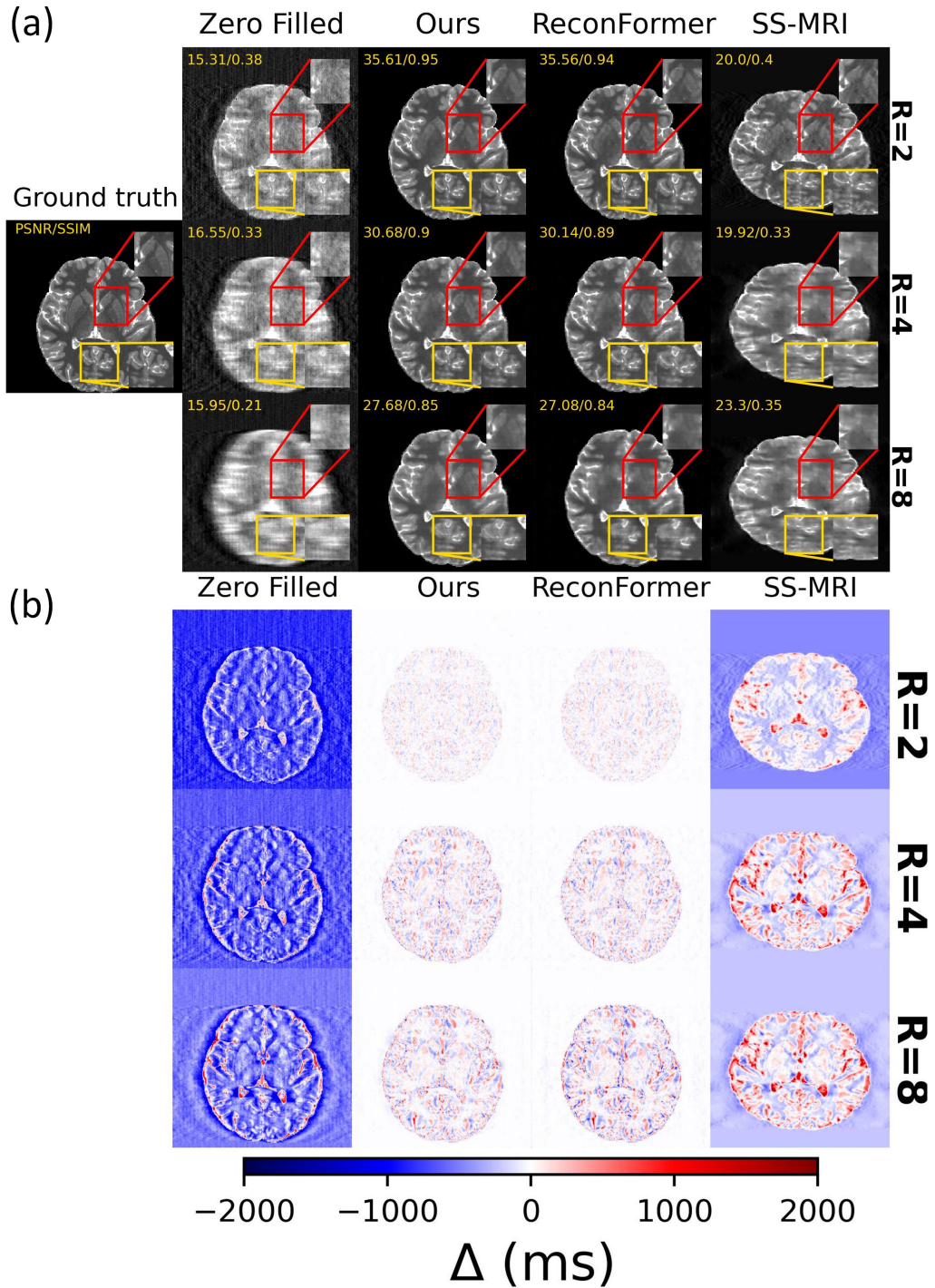


Figure 3: Reconstruction of within-domain MP2RAGE T_1 quantitative maps at acceleration factors $R \in \{2\times, 4\times, 8\times\}$ with $\rho = 0.5$. (a) Reconstructed images for a representative subject. (b) Corresponding pixel-wise difference maps between the reconstructed images and the ground truth images.

methods. The results of the Tukey’s HSD multi-comparison are presented in the text and tables.

Multi-coil dataset: The quantitative results are listed in Table 1 at $R = 2\times$, $4\times$, and $8\times$ and $\rho = 0.5$. The ANOVA test indicated $p \ll 10^{-5}$ for all metrics indicating that there are statistically significant differences between average values. Our method got the lowest NMSE for all acceleration rates except for $R = 2\times$ and $R = 4\times$. The ReconFormer method achieved the lowest NMSE value at $R = 2\times$, nonetheless, it was not statistically significantly different from our method ($p = 0.65$). Although our method achieved the lowest NMSE value for $R = 8\times$, but it was not statistically significantly different from ReconFormer ($p = 0.83$). Our proposed self-supervised method achieved the highest PSNR and SSIM values for all acceleration rates, demonstrating the lowest remaining spatial distortion, like ghosting, and the highest structural similarities between the ground truth and reconstructed images, respectively.

Single-coil dataset: The quantitative metrics for the within-domain single-coil high-resolution MP2RAGE T1 map are summarized in Table 2. One-way ANOVA tests indicated significant differences ($p \ll 10^{-5}$) between methods for all metrics. Tukey’s HSD test confirmed that SSAD-MRI performed significantly better than comparative methods for most metrics. Our method achieved the lowest NMSE values that were statistically different from the comparative methods for all acceleration rates except with $R = 2$ where our method achieved a performance that did differ statistically significant (p -value = 0.97) from ReconFormer. Our method achieved the highest PSNR values for all acceleration rates that were statistically significantly ($p < 0.05$) from comparative methods. Furthermore, our proposed method achieved the

Table 1: Within-domain performance for multi-coil axial T_2 -w fastMRI at $R \in \{2\times, 4\times, 8\times\}$ and $\rho = 0.5$ are provided. The arrows indicate directions of better performance.

| | | Zero filled | Reconformer | SS-MRI | Ours |
|----------------------|----|-----------------------|-----------------------|-----------------------|-----------------------|
| NMSE (95% CI) [%] ↓ | 2× | 14.99 (12.95 - 17.14) | 0.47 (0.45 - 0.50)* | 0.52(0.50 - 0.56)* | 0.51 (0.48 - 0.54) |
| | 4× | 26.38 (24.43 - 28.41) | 1.51 (1.42 - 1.64) | 4.87 (4.76 - 4.98) | 1.26 (1.18 - 1.35) |
| | 8× | 30.78 (28.49 - 33.11) | 3.34 (3.14 - 3.63)* | 5.16 (5.06 - 5.26) | 3.26 (3.08 - 3.51) |
| PSNR (95% CI) [dB] ↑ | 2× | 17.75 (17.64 - 17.86) | 38.80 (38.65 - 38.96) | 39.05 (38.90 - 39.20) | 39.93 (39.78 - 40.09) |
| | 4× | 17.64 (17.53 - 17.77) | 34.23 (34.10 - 34.37) | 27.93 (27.79 - 28.06) | 35.44 (35.31 - 35.58) |
| | 8× | 18.75 (18.62 - 18.87) | 27.54 (27.43 - 27.65) | 30.65 (30.52 - 30.76) | 31.67 (31.55 - 31.79) |
| SSIM (95% CI) [%] ↑ | 2× | 81.53(81.06 - 81.98) | 96.10(96.02 - 96.17) | 96.33 (96.15 - 96.49) | 98.14 (98.06 - 98.21) |
| | 4× | 70.89(70.44 - 71.36) | 93.36(93.19 - 93.51) | 76.72 (76.32 - 77.15) | 95.55 (95.38 - 95.69) |
| | 8× | 65.17(64.65 - 65.68) | 89.72 (89.49 - 89.93) | 77.01 (76.63 - 77.39) | 91.67 (91.45 - 91.89) |

* indicates p -value > 0.05 of Tukey’s HSD.

highest SSIM at $R = 4\times$ that differed statistically significantly higher ($p < 0.05$) than the other methods. Nonetheless, our method and ReconFormer performed similarly in terms of the SSIM index at $R = 2\times$ and $8\times$ with p of 0.15 and 0.61, respectively.

Table 2: Within-domain performance for single-coil high resolution MP2RAGE T1 map at $R \in \{2\times, 4\times, 8\times\}$ and $\rho = 0.5$ are provided. The arrows indicate directions of better performance.

| | | Zero filled | Reconformer | SS-MRI | Ours |
|----------------------|----|-----------------------|-----------------------|-----------------------|----------------------|
| NMSE (95% CI) [%] ↓ | 2× | 15.37 (13.46 - 17.49) | 0.53 (0.52 - 0.55)* | 5.18 (4.97 - 5.40) | 0.56 (0.54 - 0.57) |
| | 4× | 22.43 (20.65 - 24.33) | 2.04 (1.99 - 2.09) | 8.75 (8.36 - 9.18) | 1.87 (1.82 - 1.93) |
| | 8× | 33.17 (31.79 - 34.72) | 3.99 (3.90 - 4.10) | 10.76 (10.42 - 11.13) | 3.65 (3.56 - 3.75) |
| PSNR (95% CI) [dB] ↑ | 2× | 15.46 (15.39 - 15.53) | 35.42(35.31 - 35.53) | 23.07(22.98 - 23.16) | 36.15(36.03 - 36.26) |
| | 4× | 15.35 (15.28 - 15.43) | 29.46(29.37 - 29.56) | 21.31(21.21 - 21.40) | 30.51(30.42 - 30.61) |
| | 8× | 16.31 (16.23 - 16.40) | 26.54(26.45 - 26.65) | 20.75(20.66 - 20.85) | 27.94(27.84 - 28.04) |
| SSIM (95% CI) [%] ↑ | 2× | 73.83(73.43 - 74.21) | 95.75(95.68 - 95.81)* | 85.12(84.57 - 85.65) | 95.31(95.25 - 95.38) |
| | 4× | 68.62(68.29 - 68.94) | 89.11(88.96 - 89.26) | 80.47(80.06 - 80.89) | 89.80(89.63 - 89.95) |
| | 8× | 66.13(65.83 - 66.42) | 84.55(84.32 - 84.78)* | 77.17(76.75 - 77.62) | 84.77(84.53 - 84.99) |

* indicates p -value > 0.05 of Tukey’s HSD.

3.1.3 Voxel-wise correlation

We visualized the scatter plot for 80 randomly selected within-domain multi-coil and single-coil data to visualize and quantify the agreement between reconstructed images at $R = 8\times$ and $\rho = 0.5$ with the ground truth reference images (see Figure S3). For the single-coil dataset, our method reconstructed images shown in Figure S3(c) with better conformity with the references markedly better than SS-MRI shown in Figure S3(a) and comparable to ReconFormer shown in Figure S3(b). The same results were achieved for multi-coil data where the SS-MRI performed better than itself trained using the single-coil dataset. Still there is a noticeable gap between our proposed (see Figure S3(f)) method and SS-MRI (see Figure S3(d)). Furthermore, ground truth voxel-wise shown in Figure S3(e) and (f) confirms that our method might be able to generate images that are more similar to the ground truth.

3.1.4 Effect of Sampling Mask Ratio

Our proposed method split the sampling mask Ω randomly into two non-overlapping masks \aleph and Υ , which were used in train and loss paths, respectively (see Figure 1). The ratio $\rho = \frac{|\aleph|}{|\Upsilon|}$ plays a role in the network’s performance. We trained and tested the network for varying $\rho \in \{0.3, 0.5, 0.7\}$ (see Figure S1 in the supplementary for the sampling masks). The effect of network training with different ρ s on the quantitative metrics NMSE, PSNR, and SSIM are shown in Figure S4 at three acceleration rates $R \in \{2, 4, 8\}$ that were acquired using multi-coil axial T2-w images. Our method

achieved better performance at $\rho = 0.5$ in terms of PSNR and SSIM at $R = 2\times$ and $R = 4\times$. However, it performed similarly to $\rho = 0.7$ at $R = 8\times$. Furthermore, NMSE metrics indicate that our method achieved the best performance when trained using $\rho = 0.5$ than other splitting ratios $\rho \in \{0.3, 0.7\}$.

3.2 Out-of-domain reconstruction

We evaluated our model performance in OOD reconstructions where our proposed method was trained using multi-coil axial T_2 -w fastMRI images and then tested using axial T_1 c and T_1 -w shown, respectively, in first and second rows of Figure 4 at $R \in \{2\times, 4\times, 8\times\}$. Our method statistically significant ($p \ll 10^{-5}$) improved quantitative metrics.

Our method achieved significantly lower NMSE values ($p \ll 10^{-5}$), indicating high voxel-wise similarity between the reconstructed and ground truth images for T_1 -w and T_1 c illustrated in Figures 4(a) and (d), respectively. In addition, high PSNR values confirm minimal remaining spatial distortion in the reconstructed images for T_1 -w and T_1 c illustrated in Figures 4(b) and (e), respectively. Finally, The high SSIM values demonstrate a remarkable resemblance between the reconstructed and ground truth images (see Figures 4(c) and (f)).

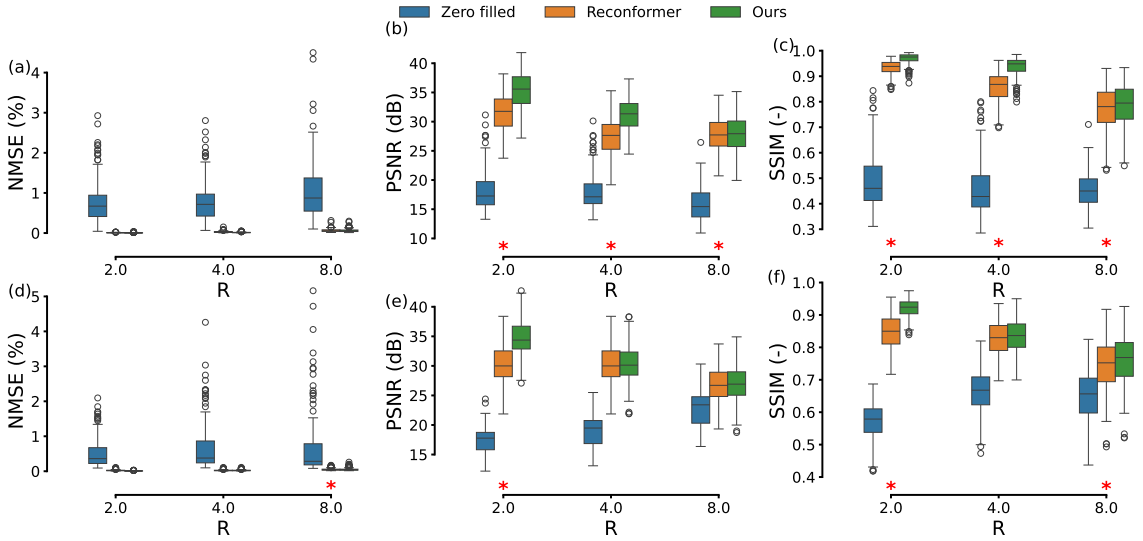


Figure 4: Out-of-distribution reconstruction quantitative results are illustrated for multi-coil axial T_1 -w and T_1 c in the first and second rows, respectively, for three difference acceleration rates. The red stars indicate statistically significant differences ($p < 0.05$).

4 Discussion

This study introduces the Self-supervised Adversarial Diffusion (SSAD-MRI) model, a novel self-supervised deep learning-based compressed sensing MRI (DL-based CS-MRI) method designed to mitigate the challenges associated with prolonged MRI acquisition times and the necessity for fully sampled datasets. Our approach leverages the synergy between CS and DL to accelerate MRI acquisition without compromising image quality.

The SSAD-MRI model’s primary innovation lies in its use of an adversarial mapper within a self-supervised framework, eliminating the dependence on fully sampled training datasets. This advancement is particularly significant for clinical scenarios where acquiring fully sampled data is impractical. By integrating a diffusion model, our method enhances sampling efficiency and reconstruction quality. The backward diffusion process, executed in smaller steps, contributes to robust and efficient image sampling. Extensive testing on OOD datasets demonstrated significant improvements in NMSE, PSNR, and SSIM metrics. These results highlight the model’s ability to generalize across various MRI sequences and patient-specific conditions, underscoring its versatility for clinical use.

The majority of DL-based CS-MRI methods use supervised learning to train networks [15, 17, 18, 25, 31]. However, acquiring fully sampled can be challenging in some practical applications due to the long acquisition time, physiological constraints, and signal decay. For instance, fully sampled high-resolution brain MP2RAGE T₁ maps can take around 24 minutes to acquire, which is impractical for large-scale studies and may lead to patient discomfort unless using acceleration approaches. Such a long acquisition time increases the likelihood of patient movements, which could substantially reduce the image quality. Thus, being able to use self-supervised DL-based CS MRI approaches is imperative to broaden their applications where acquiring such data are challenging. By integrating a diffusion model, our method enhances reconstruction quality. The backward diffusion process, executed in smaller steps, contributes to robust and efficient image sampling. Extensive testing on within domain and OOD datasets demonstrated significant improvements in NMSE, PSNR, and SSIM metrics. These results highlight the model’s ability to generalize across various MRI sequences and patient-specific conditions, underscoring its versatility for clinical use.

Several self-supervised studies have been proposed to train models without using fully sampled data. For instance, a data-driven method of de-aliasing was proposed for single-coil data that performed an image-to-image translation [43]. However, it did not encode the operator and used similar sampling patterns for training and loss that increased noise during test time. Alternatively, a self-supervised study was proposed, assuming data acquired using two different sampling patterns [44]. Furthermore, physics-driven methods that unroll the training process were also proposed [26].

However, the unrolling nature of the method might increase the test burden and time.

Our self-supervised method uses a DC layer that was trained end-to-end and utilizes the sampling pattern that is feasible to implement clinically. Our experiment using the multi-coil and single-coil datasets enabled us to recover target ground truth images at markedly lower acquisition times (see Figures 2 - 3). Specifically, our method could recover fine details highlighted by gold and red boxes better than the comparative models. The negligible amount of remaining aliasing in the reconstructed images was confirmed by Figures 2 - 3. Furthermore, this observation was confirmed by the low NMSE and high PSNR achieved in Tables 1 and 2.

Although our method could achieve the best performance for $\rho = 0.5$, Figure S4 suggests that our method might be resilient to variations in the hyper-parameter ρ around $\rho = 0.5$. In addition, out-of-domain reconstruction indicated in Figure 4 with significantly better performance ($p \ll 10^{-5}$) indicates reasonable robustness against domain shifts. However, the results indicate that our method performed better on T1-w (first row in Figure 4) images than when tested on T1c (second row in Figure 4). That might be due to contrast agent enhancement in T1c in the image regions that are not evident in T2-w images.

Despite the promising outcomes, several limitations need to be addressed. Our model was not tested on prospectively undersampled raw k -space datasets acquired under parallel imaging frameworks. Future work should explore the application of SSAD-MRI to these datasets to validate its clinical utility further. In addition, we did not train our method using raw multi-coil high-resolution 3D MRI, such as T₁ magnetization-prepared rapid acquisition gradient echoes MRI images because they are not readily available to the end-user. Expanding the training dataset to include such data could enhance the model’s performance.

Addressing these limitations will enhance not only the model’s performance but also its clinical applicability. By reducing acquisition times through high acceleration factors—potentially up to 8-fold. Integrating our method into existing clinical workflows is feasible because it does not require fully sampled training datasets; it can be incorporated into the reconstruction pipelines of current MRI systems without major alterations to imaging protocols. Focusing on model generalizability, computational efficiency, and adherence to regulatory standards will be crucial for clinical deployment. Collaboration between researchers, clinicians, and industry partners will facilitate the optimization of our method for image reconstruction and ensure its effectiveness across diverse clinical scenarios. Ultimately, this approach has the potential to lead to improved patient outcomes and enhanced diagnostic capabilities by providing high-quality images more efficiently.

5 Conclusions

Our self-supervised adversarial diffusion model significantly improves the quality of reconstructed MRI images from undersampled data without requiring fully sampled training datasets, offering a promising solution for accelerating MRI acquisition. The proposed method has the potential to enhance clinical MRI practices by reducing scan times and improving image quality, which is crucial for accurate diagnosis and treatment planning. By decreasing imaging time and the likelihood of motion artifacts, our approach may benefit a wide range of applications where rapid and reliable MRI is essential.

Conflicts of interest

There are no conflicts of interest declared by the authors.

Acknowledgement

This research is supported in part by the National Institutes of Health under Award Number R56EB033332, R01EB032680, and R01CA272991.

Data availability

Brain MRI data were obtained from the New York University fastMRI initiative database (<https://fastmri.med.nyu.edu/>) and the OpenNeuro MPI-Leipzig Mind-Brain-Body dataset (<https://openneuro.org/datasets/ds000221/versions/00002>). The datasets were acquired with the relevant institutional review board approvals.

References

- [1] H. Hricak, M. Abdel-Wahab, R. Atun, et al., Medical imaging and nuclear medicine: a Lancet Oncology Commission, *The Lancet Oncology* **22**, e136–e172 (2021).
- [2] Y. Liu, A. T. Leong, Y. Zhao, et al., A low-cost and shielding-free ultra-low-field brain MRI scanner, *Nature communications* **12**, 7238 (2021).
- [3] M. Safari, X. Yang, C.-W. Chang, R. L. Qiu, A. Fatemi, and L. Archambault, Unsupervised MRI motion artifact disentanglement: introducing MAUDGAN, *Physics in Medicine and Biology* (2024).

- [4] J. P. Haldar, D. Hernando, and Z.-P. Liang, Compressed-sensing MRI with random encoding, *IEEE transactions on Medical Imaging* **30**, 893–903 (2010).
- [5] A. Deshmane, V. Gulani, M. A. Griswold, and N. Seiberlich, Parallel MR imaging, *Journal of Magnetic Resonance Imaging* **36**, 55–72 (2012).
- [6] R. Otazo, D. Kim, L. Axel, and D. K. Sodickson, Combination of compressed sensing and parallel imaging for highly accelerated first-pass cardiac perfusion MRI, *Magnetic resonance in medicine* **64**, 767–776 (2010).
- [7] Y. Chang, D. Liang, and L. Ying, Nonlinear GRAPPA: A kernel approach to parallel MRI reconstruction, *Magnetic Resonance in Medicine* **68**, 730–740 (2012).
- [8] M. Safari, Z. Eidex, C.-W. Chang, R. L. Qiu, and X. Yang, Fast MRI Reconstruction Using Deep Learning-based Compressed Sensing: A Systematic Review, arXiv preprint [arXiv:2405.00241](https://arxiv.org/abs/2405.00241) (2024).
- [9] Z. Jin and Q.-S. Xiang, Improving accelerated MRI by deep learning with sparsified complex data, *Magnetic Resonance in Medicine* **89**, 1825–1838 (2023).
- [10] C. Cao, Z.-X. Cui, Y. Wang, et al., High-Frequency Space Diffusion Model for Accelerated MRI, *IEEE Transactions on Medical Imaging* (2024).
- [11] D. Narnhofer, K. Hammernik, F. Knoll, and T. Pock, Inverse GANs for accelerated MRI reconstruction, in *Wavelets and Sparsity XVIII*, volume 11138, pages 381–392, SPIE, 2019.
- [12] Z. Eidex, M. Safari, R. L. Qiu, et al., Temporal-Aware Cardiac MRI Reconstruction with a Deep Residual-Transformer Network, in *AAPM 66th Annual Meeting & Exhibition*, AAPM, 2024.
- [13] A. Kofler, M.-C. Pali, T. Schaeffter, and C. Kolbitsch, Deep supervised dictionary learning by algorithm unrolling—Application to fast 2D dynamic MR image reconstruction, *Medical Physics* **50**, 2939–2960 (2023).
- [14] B. Qu, J. Zhang, T. Kang, et al., Radial magnetic resonance image reconstruction with a deep unrolled projected fast iterative soft-thresholding network, *Computers in Biology and Medicine* **168**, 107707 (2024).
- [15] M. Safari, X. Yang, and A. Fatemi, MRI data consistency guided conditional diffusion probabilistic model for MR imaging acceleration, in *Medical Imaging 2024: Clinical and Biomedical Imaging*, volume 12930, pages 202–205, SPIE, 2024.

- [16] A. Güngör, S. U. Dar, Ş. Öztürk, et al., Adaptive diffusion priors for accelerated MRI reconstruction, *Medical Image Analysis* **88**, 102872 (2023).
- [17] M. L. Terpstra, M. Maspero, J. J. Verhoeff, and C. A. van den Berg, Accelerated respiratory-resolved 4D-MRI with separable spatio-temporal neural networks, *Medical physics* **50**, 5331–5342 (2023).
- [18] J. Schlemper, J. Caballero, J. V. Hajnal, A. N. Price, and D. Rueckert, A deep cascade of convolutional neural networks for dynamic MR image reconstruction, *IEEE transactions on Medical Imaging* **37**, 491–503 (2017).
- [19] J. Sun, H. Li, Z. Xu, et al., Deep ADMM-Net for compressive sensing MRI, *Advances in neural information processing systems* **29** (2016).
- [20] K. Hammernik, T. Klatzer, E. Kobler, et al., Learning a variational network for reconstruction of accelerated MRI data, *Magnetic resonance in medicine* **79**, 3055–3071 (2018).
- [21] K. C. Tezcan, C. F. Baumgartner, R. Luechinger, K. P. Pruessmann, and E. Konukoglu, MR image reconstruction using deep density priors, *IEEE transactions on medical imaging* **38**, 1633–1642 (2018).
- [22] G. Luo, N. Zhao, W. Jiang, E. S. Hui, and P. Cao, MRI reconstruction using deep Bayesian estimation, *Magnetic resonance in medicine* **84**, 2246–2261 (2020).
- [23] G. Luo, M. Blumenthal, M. Heide, and M. Uecker, Bayesian MRI reconstruction with joint uncertainty estimation using diffusion models, *Magnetic Resonance in Medicine* **90**, 295–311 (2023).
- [24] K. Setsompop, R. Kimmlingen, E. Eberlein, et al., Pushing the limits of in vivo diffusion MRI for the Human Connectome Project, *Neuroimage* **80**, 220–233 (2013).
- [25] P. Guo, Y. Mei, J. Zhou, S. Jiang, and V. M. Patel, ReconFormer: Accelerated MRI reconstruction using recurrent transformer, *IEEE transactions on medical imaging* (2023).
- [26] B. Yaman, S. A. H. Hosseini, S. Moeller, J. Ellermann, K. Uğurbil, and M. Akçakaya, Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data, *Magnetic resonance in medicine* **84**, 3172–3191 (2020).

- [27] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, and S. Ganguli, Deep unsupervised learning using nonequilibrium thermodynamics, in *International conference on machine learning*, pages 2256–2265, PMLR, 2015.
- [28] J. Ho, A. Jain, and P. Abbeel, Denoising diffusion probabilistic models, *Advances in neural information processing systems* **33**, 6840–6851 (2020).
- [29] C. Luo, Understanding diffusion models: A unified perspective, arXiv preprint [arXiv:2208.11970](https://arxiv.org/abs/2208.11970) (2022).
- [30] S. H. Chan, Tutorial on Diffusion Models for Imaging and Vision, arXiv preprint [arXiv:2403.18103](https://arxiv.org/abs/2403.18103) (2024).
- [31] J. Huang, A. I. Aviles-Rivero, C.-B. Schönlieb, and G. Yang, Cdiffmr: Can we replace the gaussian noise with k-space undersampling for fast mri?, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 3–12, Springer, 2023.
- [32] M. Safari, X. Yang, A. Fatemi, and L. Archambault, MRI motion artifact reduction using a conditional diffusion probabilistic model (MAR-CDPM), *Medical Physics* **51**, 2598–2610 (2024).
- [33] A. Paszke, S. Gross, F. Massa, et al., Pytorch: An imperative style, high-performance deep learning library, *Advances in neural information processing systems* **32** (2019).
- [34] J. Zbontar, F. Knoll, A. Sriram, et al., fastMRI: An Open Dataset and Benchmarks for Accelerated MRI, 2018.
- [35] F. Knoll, J. Zbontar, A. Sriram, et al., fastMRI: A publicly available raw k-space and DICOM dataset of knee images for accelerated MR image reconstruction using machine learning, *Radiology: Artificial Intelligence* **2**, e190007 (2020).
- [36] A. Babayan, B. Baczkowski, R. Cozatl, et al., ”MPI-Leipzig-Mind-Brain-Body”, 2018.
- [37] A. Babayan, M. Erbey, D. Kumral, et al., A mind-brain-body dataset of MRI, EEG, cognition, emotion, and peripheral physiology in young and old adults, *Scientific data* **6**, 1–21 (2019).
- [38] M. Uecker, P. Lai, M. J. Murphy, et al., ESPIRiT—an eigenvalue approach to autocalibrating parallel MRI: where SENSE meets GRAPPA, *Magnetic resonance in medicine* **71**, 990–1001 (2014).

- [39] C. Hu, C. Li, H. Wang, Q. Liu, H. Zheng, and S. Wang, Self-supervised learning for mri reconstruction with a parallel network training framework, in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VI 24*, pages 382–391, Springer, 2021.
- [40] S. Kastruyulin, J. Zakirov, D. Prokopenko, and D. V. Dylov, PyTorch Image Quality: Metrics for Image Quality Assessment, 2022.
- [41] Z. Eidex, J. Wang, M. Safari, et al., High-resolution 3T to 7T ADC map synthesis with a hybrid CNN-transformer model, *Medical Physics* (2024).
- [42] B. Efron and R. J. Tibshirani, Confidence intervals based on bootstrap percentiles, in *An introduction to the bootstrap*, pages 168–177, Springer, 1993.
- [43] O. Senouf, S. Vedula, T. Weiss, A. Bronstein, O. Michailovich, and M. Zibulevsky, Self-supervised learning of inverse problem solvers in medical imaging, in *Domain Adaptation and Representation Transfer and Medical Image Learning with Less Labels and Imperfect Data: First MICCAI Workshop, DART 2019, and First International Workshop, MIL3ID 2019, Shenzhen, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 13 and 17, 2019, Proceedings 1*, pages 111–119, Springer, 2019.
- [44] J. Liu, Y. Sun, C. Eldeniz, W. Gan, H. An, and U. S. Kamilov, RARE: Image reconstruction using deep priors learned without groundtruth, *IEEE Journal of Selected Topics in Signal Processing* **14**, 1088–1099 (2020).