DDM^2: SELF-SUPERVISED DIFFUSION MRI DENOISING WITH GENERATIVE DIFFUSION MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

Magnetic resonance imaging (MRI) is a common and life-saving medical imaging technique. However, acquiring high signal-to-noise ratio MRI scans requires long scan times, resulting in increased costs and patient discomfort, and decreased throughput. Thus, there is great interest in denoising MRI scans, especially for the subtype of diffusion MRI scans that are severely SNR-limited. While most prior MRI denoising methods are supervised in nature, acquiring supervised training datasets for the multitude of anatomies, MRI scanners, and scan parameters proves impractical. Here, we propose **D**enoising **D**iffusion **M**odels for **D**enoising **D**iffusion **M**RI (DDM^2), a self-supervised denoising method for MRI denoising using diffusion denoising generative models. Our three-stage framework integrates statistic-based denoising theory into diffusion models and performs denoising through conditional generation. During inference, we represent input noisy measurements as a sample from an intermediate posterior distribution within the diffusion Markov chain. We conduct experiments on 4 real-world *in-vivo* diffusion MRI datasets and show that our DDM^2 demonstrates superior denoising performances ascertained with clinically-relevant visual qualitative and quantitative metrics.

1 INTRODUCTION

Magnetic resonance imaging (MRI) is a non-invasive clinical imaging modality that can provide life-saving diagnostic information. Diffusion MRI is a subtype of MRI commonly used in oncologic and neurologic disorders (Bihan, 2003; Bihan et al., 2006), which can quantitatively assess microstructural anatomical details. However, diffusion MRI scans suffer from severe signal to noise ratio (SNR) deficits, hindering diagnostic and quantitative accuracy. Image SNR can be improved either by lowering image resolution, which further reduces diagnostic utility, or by increasing the total scan time, which already can require 10+ minutes of time in the MRI scanner. Thus, there is large interest in decreasing diffusion MRI scan times to improve patient throughout in hospitals and the patient experience. While image SNR is governed by the underlying MRI physics, applying post-processing denoising techniques to fast and low-SNR MRI acquisitions can improve overall image SNR. Developing such methods to improve SNR of diffusion MRI scans is an unsolved problem that may improve the efficacy of the millions of such scans performed in routine clinical practice annually.

Supervised machine learning techniques have previously been proposed for MRI denoising, however, such methods are limited by their clinical feasibility. It is clinically impractical to acquire paired high- and low-SNR diffusion MRI scans across various anatomies (e.g., brain, abdomen, etc), diffusion weighting factors (and resultant SNR variations), MRI field strength and vendors, and clinical use-cases. Such large distributional shifts across heterogeneous use-cases leads to fundamental drop in model performance (Darestani et al., 2021). The diversity of data and the need for effective denoising methods motivates the use of unsupervised denoising techniques for diffusion MRI.

To address these challenges, the major contributions of our work are three-fold: *(i)* We propose DDM^2 for unsupervised denoising of diffusion MRI scans using diffusion denoising models. Our three-stage self-supervised approach couples statistical self-denoising techniques into the diffusion models (shown in Figure 1 Left); *(ii)* DDM^2 allows representing noisy inputs as samples from an intermediate state in the diffusion Markov chain to generate fine-grained denoised images without requiring ground truth references; *(iii)* We evaluate our method on four real-world diffusion MRI datasets that encompass the latest and longer established MRI acquisition methods. DDM^{2}demonstrates

state-of-the-art denoising performance that outperforms second best runner-up by an average of 3.2 SNR and 3.1 contrast-to-noise ratio (CNR) points, across this large range of MRI distributions.

2 Related Work and Background

2.1 Diffusion MRI

Diffusion MRI sensitizes the MR signal to the movement of protons within an object being imaged using tailored imaging gradients (small magnetic fields steerable in the all directions). Given that both magnitude (overall duration that gradients are turned on, termed as *b-value*) and directions (linear combination of the three independent gradient directions) can be steered, the diffusion weighting is considered as a vector. A typical diffusion MRI scan for resolving tissue microstructure is consequently 4D in nature, with 3 dimensions of spatial coordinates and 1 dimension of diffusion vectors. Depending on the clinical use-case, the number of diffusion vectors can range from 2 to upwards of 100. Prior approaches for improving diffusion MRI sequences exist; however, these tasks are solved with separate supervised tools for very specific image anatomies and acquisitions (Kaye et al., 2020; Gibbons et al., 2018). Such supervised methods, while powerful, cannot generalizable to the variety of anatomies, MRI scanners, and image parameters of diffusion MRI scans.

2.2 Statistic-Based Image Denoising

Given a noisy measurement $\mathbf{x}$, an inverse problem is to recover the clean image $\mathbf{y}$ defined by:

$$\mathbf{x} = \lambda_1 \mathbf{y} + \epsilon, \tag{1}$$

where λ_1 is a linear coefficient and ϵ denotes the additive noise. Without losing generality, one usually represent $\epsilon = \lambda_2 \mathbf{z}$ as a sample from the Gaussian distribution, $\mathcal{N}(\mathbf{0}, \lambda_2^2 \mathbf{I})$. When paired ground truth images $\mathbf{y}$ are available, a neural network can then be trained to approximate $\mathbf{y}$ through direct supervisions. However, supervised training proves infeasible when $\mathbf{y}$ are missing in the datasets.

Attempts have been made to relax the reliance on supervised signal from clean images, by using noisy images alone. Based on the assumption that the additive noise ϵ is pixel-wise independent, Noise2Noise (Lehtinen et al., 2018) claimed that to denoise an noisy measurement $\mathbf{x}$ towards another measurement $\mathbf{x}'$ is statistically equivalent to the supervised training on $\mathbf{y}$ up to a constant:

$$\mathcal{L}(\Phi(\mathbf{x}'), \mathbf{x}) = ||\Phi(\mathbf{x}') - \mathbf{x}||^2 \approx ||\Phi(\mathbf{x}') - \mathbf{y}||^2 + \texttt{const}. \tag{2}$$

Based on the above idea, Noise2Self (Batson & Royer, 2019) further designed the $\mathcal{J}$-Invariance theory that achieves self-supervised denoising by using the input $\mathbf{x}$ itself. With the statistical independence characteristic of noise, Noise2Void (Krull et al., 2019) and Laine *et al*. (Laine et al., 2019) extended the self-supervised strategy to a so-called 'blind-spot' technique that masked out image patches and designed a particular neural networks to predict the unmasked pixels. Neighbor2Neighbor (Huang et al., 2021) creates noisy image pairs by dividing sub-samples and uses them supervise each other. Without losing pixel information, Recorrupted2Recorrupted (Pang et al., 2021) was proposed to keep all pixels intact and achieve self-supervised denoising by deriving two additional corrupted signals. To better utilize 4D MRI characteristics, Patch2Self (Fadnavis et al., 2020) was proposed to utilize multiple diffusion vector volumes as input $\{\mathbf{x}'\}$ and learn volume-wise denoising.

Building on the statistical theory studied above, in this work we achieve unsupervised MRI denoising through a generative approach. Compared to Patch2Self that requires a large number of volumes (e.g. > 60) as a guarantee to denoise a single volume, DDM^2 can efficiently denoise MRI scans acquired with very few diffusion directions and limited number of volumes (e.g. < 5). This is clinically relevant since common clinical diffusion MRI scans use fewer than 10 directions. Moreover, unlike Patch2Self, our trained denoiser is universally applicable for the entire 4D sequence and no repeated training is needed for denoising volumes in different directions.

2.3 Diffusion Generative Model

A diffusion model (Sohl-Dickstein et al., 2015) denotes a parameterized Markov chain with T discretized states $\mathrm{S}_{1,\cdots,T}$ that can be trained to generate samples to fit a given data distribution. Transition of this chain is bi-directional and is controlled by a pre-defined noise schedule $\beta_{1,\cdots,T}$.

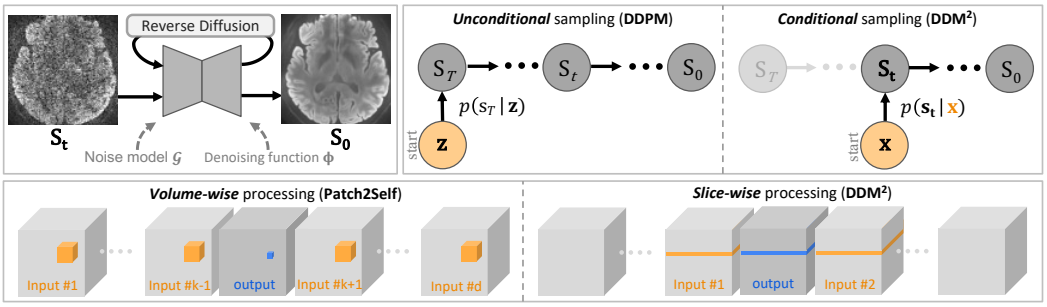

Figure 1: **Top Left:** DDM^2 generates clean MRI through reverse diffusion. **Top Right:** *Unconditional* sampling process (Ho et al., 2020) (starts from a noisy prior $\mathbf{z}$) $v.s.$ our *conditional* sampling process (starts from a noisy measurement $\mathbf{x}$). **Bottom:** *Volume-wise* processing (Fadnavis et al., 2020) (requires a large number of input volumes) $v.s.$ our *slice-wise* processing (requires very few input volumes). Gray blocks indicate accessible 3D volumes in a MRI sequence, orange blocks indicate the input priors $\{\mathbf{x}'\}$ and black blocks indicate the one-pass output.

The forward process or the *diffusion process* $q(\mathrm{S}_t|\mathrm{S}_{t-1})$ gradually adds Gaussian noise to the posterior samples at every state until the signal is completely corrupted. The reverse of the above diffusion process or the *reverse process* tries to recover the prior signal with a parameterized neural network $\mathcal{F}$:

$$p_{\mathcal{F}}(\mathrm{S}_{0,\cdots,T}) := p(\mathrm{S}_T)\prod_{t=1}^{T} p_{\mathcal{F}}(\mathrm{S}_{t-1}|\mathrm{S}_t), \qquad p_{\mathcal{F}}(\mathrm{S}_{t-1}|\mathrm{S}_t) := \mathcal{N}(\mathrm{S}_{t-1}; \mathcal{F}(\mathrm{S}_t, *), \sigma_t^2\mathbf{I}). \tag{3}$$

Starting from $\mathrm{S}_T = \mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, the reverse transition continues till the first state S_0: a valid sample from the given data distribution. Following (Ho et al., 2020), we set $\sigma_t^2 = \beta_t$ and hold $\beta_{1,\cdots,T}$ as hyperparameters. In what follows, we refer this kind of set up as the baseline diffusion model.

Recent works have extended diffusion models to improve sampling quality or efficiency. Song *et al.* (Song & Ermon, 2019) investigated the density of data distribution and incorporated Langevin dynamics and score matching methods (Hyvärinen, 2005) into diffusion model. In a follow-up work, Song *et al.* (Song et al., 2021b) formulated score based generative models as solutions to Stochastic Differential Equation (SDE). Recently, Noise2Score (Kim & Ye, 2021) also trained a score function and in a manner similar to ours to perform score-based diffusion model for blind image denoising. Rather than transiting through the diffusion Markov chain, the authors validated its ability as a denoiser by following the Tweedie's formula (Efron, 2011).

On the other hand, Denoising Diffusion Implicit Model (DDIM) (Song et al., 2021a) presented a non-Markovian acceleration technique for the reverse process that enables the sampling of a data in only few states. Similarly, Watson *et al.* (Watson et al., 2022) proposed a generalized family of probabilistic samplers called Generalized Gaussian Diffusion Processes (GGDP) to reduce the number of states without compromising generation quality. By using knowledge distillation, Salimans *et al.* proved that faster reverse process can be learnt from a slower one in the student-teacher manner iteratively (Salimans & Ho, 2022). Instead of starting from S_T, our method initiates the reverse process from an intermediate state that also accelerates generation. DDM^2 is compatible with the above advances that can be utilized for further accelerations.

Prior works also adopted the diffusion model to recover data across modalities. WaveGrad (Chen et al., 2021), DiffWave (Kong et al., 2021), and BBDM (Lam et al., 2022) are for speech/audio generation. Luo *et al.* (Luo & Hu, 2021), Zhou *et al.* (Zhou et al., 2021), and GeoDiff (Xu et al., 2022) are for graph-structure data generation. Chung *et al.* (Chung et al., 2021) and Song *et al.* (Song et al., 2022) are for medical image inverse problem solving. Chung *et al.* (Chung et al., 2022) concurrently explored sampling noisy inputs as posteriors in the diffusion Markov chain for image denoising and super resolution using coarse noise approximations via image eigenvalues. In contrast, we integrate statistical theory and self-supervised noise estimation into diffusion models to account for noise distribution shifts. We note that there was no open source code available for (Chung et al., 2022) for using it as a comparison method.

3 METHODS

3.1 PROBLEM DEFINITION AND OVERVIEW

In this section, we show that clean image approximations $\bar{\mathbf{y}}$ can be generated through a diffusion model with a known noise model $\mathcal{G}$. To condition diffusion sampling on the noisy input $\mathbf{x}$, our idea is to represent $\mathbf{x}$ as a sample from a posterior at some data-driven intermediate state $\mathrm{S}_\mathbf{t}$ in the Markov chain and initiate the sampling process directly from $\mathbf{x}$ through $p(\mathrm{S}_\mathbf{t}|\mathbf{x})$. A comparison between the *unconditional* sampling process (Ho et al., 2020) that starts from $p(\mathrm{S}_T|\mathbf{z})$, $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and our *conditional* sampling process is outlined in Figure 1 Right.

There are now three questions need answers: *(i)* How to find a noise model $\mathcal{G}$ in an unsupervised manner? *(ii)* How to match $\mathbf{x}$ to an intermediate state in the Markov chain? and *(iii)* How to train the diffusion model as a whole when both noise model and the best matched states are available? To mitigate parameter tuning, DDM^2 is designed in three sequential stages as solutions w.r.t the above questions: **(`Stage I`)** Noise model learning; **(`Stage II`)** Markov chain state matching; and **(`Stage III`)** Diffusion model training. The overall framework is outlined in Figure 2. All experiments below are performed on 2D slices, however 3D volume results are shown in the supplementary.

3.2 STAGE I: NOISE MODEL LEARNING

It is a common assumption that noisy signals can be obtained by corrupting the clean signals with noise sampled from an explicit noise model (Prakash et al., 2021). Without prior knowledge on the corruption process, it is difficult to find an effective mapping between $\mathbf{x}$ and $\mathbf{y}$. Following the $\mathcal{J}$-Invariance theory, one can approximate a solution by learning a denoising neural network from the noisy images themselves. To be more specific, with some prior signals $\{\mathbf{x}'\}$ as input, a denoising function Φ can be trained to generate a clean approximation $\bar{\mathbf{y}}$ of $\mathbf{x} \notin \{\mathbf{x}'\}$ such that $\mathbf{y} \approx \bar{\mathbf{y}} = \Phi(\{\mathbf{x}'\})$.

For a 4D MRI sequence $X \in \mathcal{R}^{w \times h \times d \times l}$ with l 3D volumes (acquired at diverse gradient directions) of size $w \times h \times d$, Patch2Self learns Φ by aggregating $l-1$ volume patches as $\{\mathbf{x}'\}$ and regressing one clean voxel (3D pixel) $\bar{\mathbf{y}} \in \mathcal{R}^{1\times1\times1}$ for the remaining volume (Fadnavis et al., 2020). This training process repeats for all of the 3D volumes in X to obtain l volume-specific denoisers. There are three major drawbacks to such a denoising strategy: *(i)* the denoising function Φ needs to be trained for each volume individually, which inevitably demands long and repeated training process for MRI sequences with several volumes. *(ii)* the denoising quality heavily relies on the size of $\{\mathbf{x}'\}$—number of accessible volumes in the 4D sequence (see studies in section 4.3); *(iii)* when inspecting in 2D slices, the denoised results yield significant spatial inconsistencies due to its volume-wise regression.

Inspired by the performance of Patch2Self, but considering its challenges, we learn a slice-to-slice mapping function for the entire sequence instead of a patch-to-voxel mapping. This enables Φ to regress an entire 2D slice at a time by taking also slices to form $\{\mathbf{x}'\}$. Such a simple modification overcomes the above three drawbacks: *(i)* only a single Φ needs to be trained for the entire sequence; *(ii)* denoising quality remains stable when forming $\{\mathbf{x}'\}$ with very few volumes (see studies in section 4.3); *(iii)* better spatial consistency can be observed in the denoised slices. A comparison between our modification and original Patch2Self is briefly shown in Figure 1 Bottom. We retain this slice-wise processing scheme throughout all three stages in our framework.

3.3 STAGE II: MARKOV CHAIN STATE MATCHING

When an optimal mapping function is learned by Φ, the noise model $\mathcal{G}$ can be obtained by fitting the approximated residual noise $\bar{\epsilon}$ to a Gaussian $\mathcal{N}(\sigma^2\mathbf{I})$ with zero mean and a variable standard deviation σ. Without any restrictions, $\bar{\epsilon}$ not necessarily has a mean value of zero and a direct fitting will lead to distribution mean shift. In this way, we propose to explicitly complement $\bar{\epsilon}$'s mean value $\boldsymbol{\mu}_{\bar{\epsilon}} = \frac{1}{||\bar{\epsilon}||}\sum \bar{\epsilon}$ to be zero and calibrate the denoising result $\bar{\mathbf{y}}$ accordingly w.r.t Equation 1:

$$\bar{\epsilon} := \bar{\epsilon} - \boldsymbol{\mu}_{\bar{\epsilon}}, \qquad \bar{\mathbf{y}} := \bar{\mathbf{y}} + \frac{\lambda_2 \boldsymbol{\mu}_{\bar{\epsilon}}}{\lambda_1}. \tag{4}$$

The complemented $\bar{\epsilon}$ can be then used to fit the $\mathcal{G}$ and estimate σ. Recall that in the diffusion model, a noise schedule $\beta_{1,\dots,T}$ is pre-defined that represents the noise level at every state in the Markov

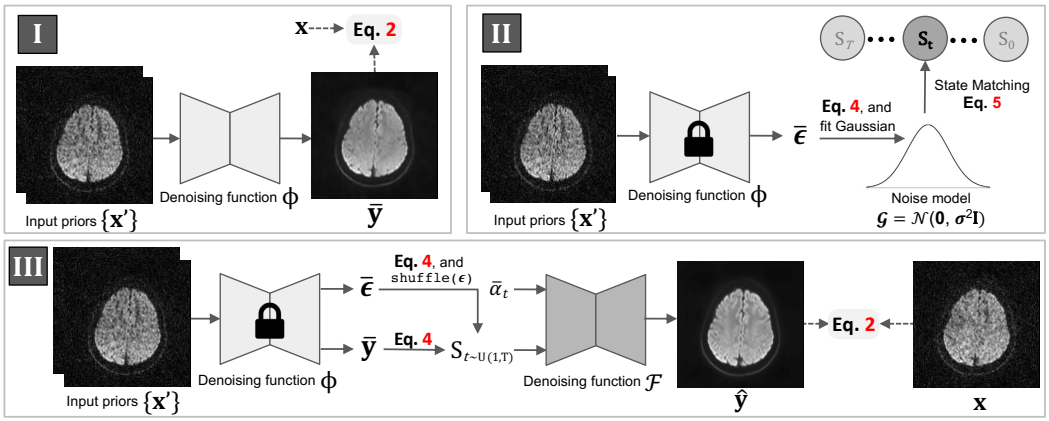

Figure 2: **Our three stage DDM^2.** In `Stage I`, we train a denoising function Φ in a self-supervised manner to estimate an initial noise distribution. In `Stage II`, the estimated noise by Φ is used to fit a Gaussian noise model $\mathcal{G}$. We match standard deviation of this noise model σ to the diffusion sampling noise schedule β and represent $\mathbf{x}$ as a sample at $\mathrm{S}_\mathbf{t}$. With both $\mathcal{G}$ and $\mathbf{t}$, in `Stage III`, we train another denoising function $\mathcal{F}$ in the diffusion model for the reverse process $p_\mathcal{F}(\mathrm{S}_0|\mathbf{x})$. All three stages take as inputs and returns 2D slices as discussed in section 3.2.

chain. We can determine the closest matching state $\mathrm{S}_\mathbf{t}$ of $\mathbf{x}$ by comparing $\mathcal{G}$ and posterior $p(\mathrm{S}_t)$—σ and $\sqrt{\beta_t}$. Specifically, a state is matched when a time stamp $\mathbf{t}$ is found that minimizes the distance:

$$\arg\min_t ||\sqrt{\beta_t} - \sigma||^p, \tag{5}$$

where $||\cdot||^p$ denotes the p-norm distance. Since t is a discrete integer within a finite interval: $\{1, \cdots, T\}$, for simplicity, we cast the optimization problem into a surrogate searching problem.

A match at state $\mathrm{S}_\mathbf{t}$ indicates that, with this particular noise schedule β, there exists at least one possible sample from the posterior at state $\mathrm{S}_\mathbf{t}$ in the baseline unconditional generation process that is sufficiently close to the given input $\mathbf{x}$. In this way, a denoised image can be sampled at state S_0 through the reverse process iteratively $p(\mathrm{S}_0|\mathrm{S}_t)$.

3.4 `Stage III`: Diffusion Model Training

In diffusion models, during the reverse process, an additional denoising function $\mathcal{F}$ is trained to predict $\bar{\epsilon}$ or $\bar{\mathbf{y}}$ at each transition step. As inputs to $\mathcal{F}$, clean images $\mathbf{y}$ are first corrupted through a prior sampling $q(\mathrm{S}_t|\mathbf{y}) = \lambda_1\mathbf{y} + \lambda_2\mathbf{z}$, where $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ (Equation 1). Following DDPM and its hyper-parameters, we rewrite $\lambda_1 = \sqrt{\bar{\alpha}_t}$ and $\lambda_2 = \sqrt{1-\bar{\alpha}_t}$ in terms of β_t: $\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i, \alpha_t = 1 - \beta_t$. In unsupervised denoising with $\mathbf{y}$ unavailable, denoised images $\bar{\mathbf{y}}$ are approximated by Φ.

Unlike the supervised training schema that can train $\mathcal{F}$ on clean ground truth $\mathbf{y}$ directly, it is sub-optimal to supervise $\mathcal{F}$ with the `Stage I` estimated $\bar{\mathbf{y}}$ in the unsupervised denoising scenario: $\mathcal{F}$ would degenerate to become an inverse function of Φ that traps $\mathcal{F}$ to interpolate only within the solution space of Φ. To enable $\mathcal{F}$ extrapolation to a wider and more accurate solution space and to potentially improve denoising, we introduce two simple modifications as below.

Noise shuffle. First, we modify the sampling protocol for the diffusion process from $q(\mathrm{S}_t|\bar{\mathbf{y}}) = \sqrt{\bar{\alpha}_t}\bar{\mathbf{y}} + \sqrt{1-\bar{\alpha}_t}\mathbf{z}$ to $q(\mathrm{S}_t|\bar{\mathbf{y}}) = \sqrt{\bar{\alpha}_t}\bar{\mathbf{y}} + \texttt{shuffle}(\bar{\epsilon})$, where $\texttt{shuffle}(\cdot)$ denotes the spatial shuffling operation and $\bar{\epsilon}$ is the residual noise predicted by Φ. Under the assumption of noise independence, such noise shuffle operation forces $\mathcal{F}$ to learn from the implicit noise distribution modelled by Φ rather than the explicit noise model $\mathcal{G}$ to reduce the distribution disparity in between.

$\mathcal{J}$-Invariance optimization. Second, we modify the loss function to train $\mathcal{F}$. Borrowing the objective used in `Stage I`, we let $\mathcal{F}$ optimize towards the original noisy input $\mathbf{x}$ instead of $\bar{\mathbf{y}}$. The objective therefore becomes $\arg\min_\mathcal{F} ||\mathcal{F}(\mathrm{S}_t, \bar{\alpha}_t) - \mathbf{x}||^2$. Such a modification not only minimizes the distance to the statistical 'ground truth' but also the approximation error brought by Φ.

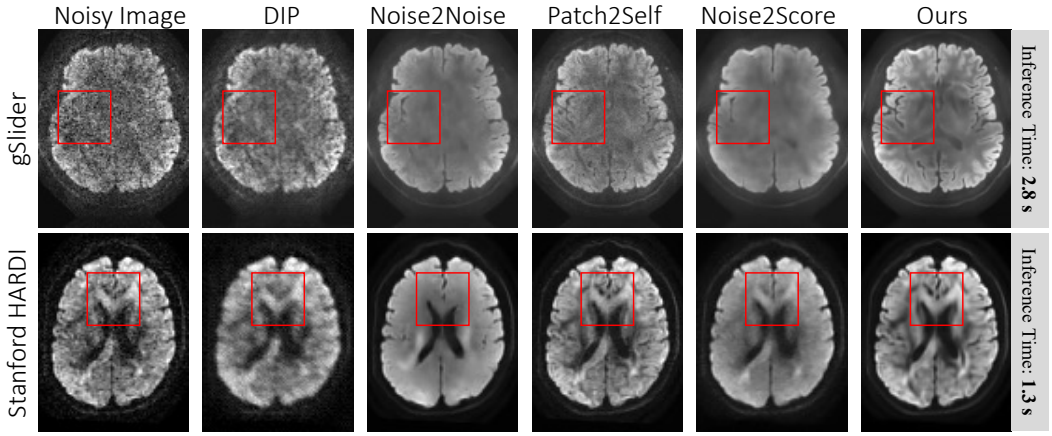

Figure 3: **Qualitative results on the most challenging datasets.** Inference time per slice is reported alongside. DDM^2 not only suppresses noise the most but also restores anatomical structures the best.

4 EXPERIMENTS

4.1 DATASETS AND EVALUATIONS

Datasets. Experiments were first conducted on an in-house dataset of brain diffusion MRI (gSlider) that has 0.5mm isotropic resolution, which was acquired under institutional review board approval using a recent advanced diffusion MRI sampling technique (Setsompop et al., 2017). The resulting gSlider dataset has dimensions of $128 \times 128 \times 160 \times 60$ with 60 being gradient directions among which 10 volumes have b-values of 0 and the rest have a b-value of 1000. To evaluate the generalizability of DDM^2, additional experiments were done on 3 other publicly-available brain diffusion MRI datasets acquired with different protocols with less advanced MRI encoding for image SNR and resolution: *(i)* Sherbrooke 3-Shell dataset (dimensions = $128 \times 128 \times 69 \times 193$) with 64 directions and denoising the b-value$= 2000$ images due to its clinical feasibility (Garyfallidis et al., 2014); *(ii)* Stanford HARDI (dimensions = $106 \times 81 \times 76 \times 160$) with 150 directions with b-value$= 2000$ (Rokem, 2016); *(iii)* Parkinson's Progression Markers Initiative (PPMI) dataset (dimensions = $116 \times 116 \times 72 \times 65$) with 64 directions with non-zero b-values (Marek et al., 2011).

Evaluations. We considered 4 state-of-the-art methods for comparisons with DDM^2: *(i)* Deep Image Prior (DIP), a classic self-supervised denoising method (Ulyanov et al., 2018); *(ii)* Noise2Noise (N2N), a statistic-based denoising method (Lehtinen et al., 2018); *(iii)* Patch2Self (P2S), state-of-the-art MRI denoising method; *(iv)* Noise2Score (N2S), a recent score-based denoising method.

We note that assessing perceptual MRI quality is a challenging research problem (Mittal et al., 2011), as *there is a lack of consensus on image quality metrics, especially in an unsupervised reference-free settings* (Chaudhari et al., 2019) (Woodard & Carley-Spencer, 2006). Metrics such as PSNR and SSIM require a reference ground truth image, yet they still have poor concordance with perceptual image quality and consequently, evaluating on downstream clinical metrics of interest provides a better measure of real-world utility (Mason et al., 2020; Adamson et al., 2021). Thus, we utilized Signal-to-Noise Ratio (SNR) and Contrast-to-Noise Ratio (CNR) to quantify denoising performance and report mean and standard deviation scores for the full 4D volumes. Both metrics require segmentation of regions of interest from the image to delineate foreground and background signals. Rather than manually selecting them as in (Chung et al., 2022), we follow the official guideline provided in (Garyfallidis et al., 2014) (a diffusion MRI community standard) to automatically segment Corpus Callosum (CC) using quantitative and clinically-relevant Fractional-Anisotropy (CFA) maps.

Implementation details. A typical noise schedule (Ho et al., 2020; Saharia et al., 2021) follows a 'warm-up' scheduling strategy. However, for most of MRI acquisitions, noise level is mild and the matched Markov chain state $\mathbf{S_t}$ will be close to S_0. This leads to very few transitions in the reverse process and eventually poor generation quality. We therefore modify β to follow a reverse 'warm-up' strategy such that β remains at $5e^{-5}$ for the first 300 iterations and then linearly increases to $1e^{-2}$ between $(300, 1000]$ iterations. *See supplementary materials for more details.*

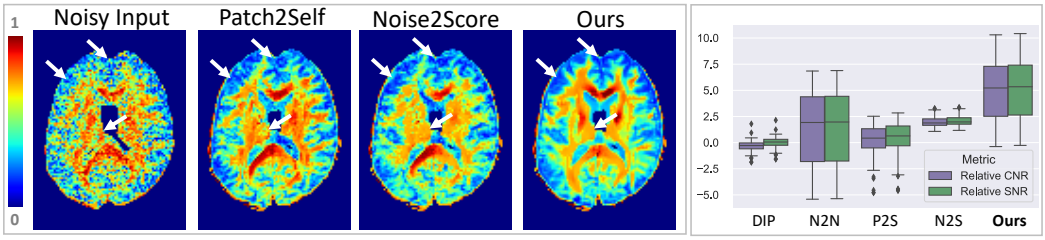

Figure 4: **Left:** Quantitative fractional anisotropy map comparisons (major differences highlighted with arrows). **Right:** DDM^2 demonstrates statistically significant SNR/CNR improvements (two-sided t-test, $< 1e^{-4}$ p-values) over all competing methods.

4.2 Main Results

Qualitative results. In Figure 3, we visualize denoising results on axial 2D slices for each comparison method for the gSlider and HARDI datasets (additional dataset images in supplementary). Compared to DIP, all other methods demonstrate better MRI denoising abilities. Although the simplest Noise2Noise model and Noise2Score models achieve promising denoising, both suffer from excessive blurring of pertinent details in the underlying anatomical structures. Utilizing the information provided in all different volumes, Patch2Self is therefore able to restore most of the details. Unfortunately, this method is not robust to heavy noise and cannot produce clean estimations for the noisier Sherbrooke and gSlider datasets. Noise2Score, as a score-based denoising method, yields intermediate denoising quality between Noise2Noise and Patch2Self but with excessive blurring. By integrating statistical denoising into diffusion models, DDM^2 not only suppresses noise the most but also restores anatomical details. *Two neuroradiologists reviewed the DDM^{2}images and commented that there were no new hallucinated lesions or regions with excessive blurring.*

Quantitative results. We calculated SNR/CNR on the entire Stanford HARDI dataset. The segmentation quality of the corpus callosum regions impacts metric calculation and an explicit tuning of RoIs would lead to unfair comparisons. To avoid this, we used scripts provided by (Garyfallidis et al., 2014) for RoI segmentation and metric calculation. In Figure 4 Right, we show box plots of relative SNR/CNR by subtracting the scores of the input noisy images from the denoised ones. Compared to DIP, N2N, and P2S, our method shows significant statistical improvements on both metrics (two-sided t-test, p-value $< 1^{-4}$). Compared to the most advanced score-based denoising method N2S, our method demonstrates an average of 0.95/0.93 improvements on SNR/CNR respectively.

The SNR/CNR measures depict how DDM^{2}improves best-case results, depicting its excellent denoising capabilities, while maintaining worst-case results compared to the comparison methods. We do not that N2S did have improved worst-case results than DDM^{2}but this was likely due to excessive image blurring induced by the method (Figure 3). Note that without ground truth references, SNR is used as approximate indications of image quality.

As a diffusion model based approach, iterative inferences are necessary for DDM^{2}Ṫhis by its nature becomes a disadvantage from the perspective of inference time, especially when comparing to single forward pass-based CNN/MLP based methods. However, due to the novel design of Markov chain state matching, DDM^2 yields 5-10 times inference speed up than DDPMs. Moreover, since many diffusion MRI applications are performed in post-processing, improved SNR at the expense of a few minutes of inference can still benefit downstream clinical tasks.

Fractional anisotropy comparisons. Besides quantitative and qualitative evaluations with DDM^2 we examine how the denoising quality translates to downstream clinical tasks such as creation of Fractional Anisotropy (FA) maps using the various image denoising methods. To calculate FA, we used the volumes acquired by the first 6 diffusion directions along with the 10 b-value=0 volumes. Representative results displayed in Figure 4 Left indicate that the proposed DDM^2 yields enhanced and less noisy capture of diffusion directions of fiber tracts.

4.3 Extensive Ablation Studies

Modifications at `Stage I`. In section 3.2, several upgrades were made upon Patch2Self for relaxing the reliance on large number of volumes (expensive to obtain during MRI acquisition) and for better

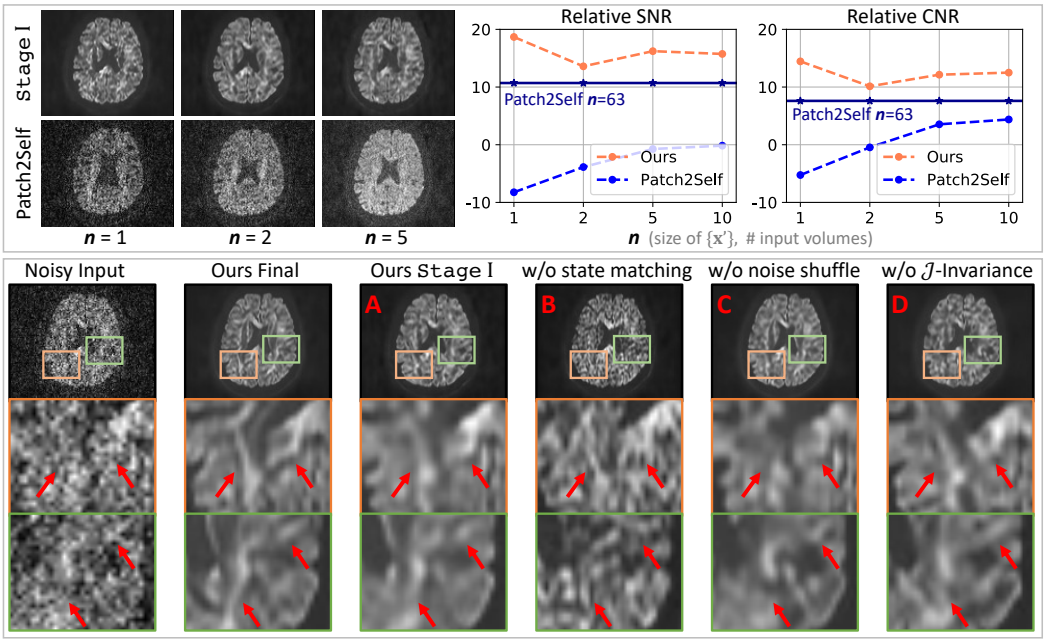

Figure 5: **Top:** Ablative results on number of accessible volumes n used as inputs. **Bottom:** Qualitative results of the ablation studies. Results obtained by different ablation settings are labeled by different identifiers. Major differences are zoomed in and are highlighted by red arrows.

spatial consistency in the denoised results. In Figure 5 Top, we compare our `Stage I` against Patch2Self by inputting limited number of volumes n (size of $\{\mathbf{x}'\}$). Patch2Self can achieve decent denoising performance when using a large number of volumes (e.g. n>60). When n is small, Patch2Self fails to denoise the data properly. However, our simple modifications leverage spatial priors and achieve impressive denoising result even for $n = 1$. With more accessible volumes, our `Stage I` maintains its denoising performance at a high level. Considering both denoising quality and clinical applicability, we adopted $n = 2$ as our final setting.

Necessity of the multi-stage design. With the learned noise model from `Stage I`, our method learns an additional diffusion model to restore more subtle details and reduce the approximated statistical error. To compare whether our three-stage method performs better than the `Stage I` alone, we compare our final results to the denoised results obtained at `Stage I` (Figure 5 **A**, *more comparisons can be found in supplementary materials*). Due to the upgrades made in section 3.2, our `Stage I` already achieves satisfactory denoising results. However, similar to other statistic-based methods, `Stage I` cannot restore high-frequency details and always converges to smooth solutions. This emphasized the novelty of using a diffusion model to generate self-consistent details to further restore the underlying anatomy.

Necessity of `Stage II`. In section 3.3, we propose to represent $\mathbf{x}$ as a sample from an intermediate state $\mathrm{S}_\mathbf{t}$ in the Markov chain. This is done by minimizing the distance between the fitted standard deviation and the pre-defined noise schedule β. Instead of finding the most matched state, we evaluate whether it is possible to naively represent $\mathbf{x}$ as the middle state $\mathrm{S}_{t=\frac{T}{2}}$ regardless of β. Given that the reverse process starts from $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, representing $\mathbf{x}$ as $\mathrm{S}_{t=\frac{T}{2}}$ may overestimate the noise level. Consequently, excessive signals will be generated and be imposed into the denoising results as shown in Figure 5 **B**. These hallucinations are not valid restorations of the underlying anatomical structure even though they may lead to high SNR/CNR scores.

Modifications at `Stage III`. As mentioned in section 3.4, two modifications are designed to assist diffusion model training. We visualize the denoising results by disabling the noise shuffle operation (Figure 5 **C**) and the $\mathcal{J}$-Invariance optimization (Figure 5 **D**). By learning from an explicit noise model and optimizing towards the `Stage I` results, we observe significant structural degradation in the denoised images. Moreover, both of the results differ considerably from the `Stage I` results (Figure 5 **A**), which further validates the need of our modifications to be used in `Stage III`.

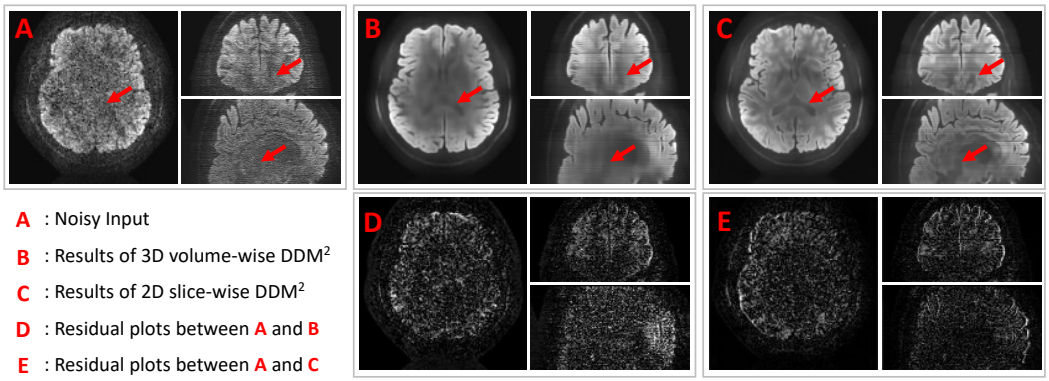

Figure 6: Comparisons between 3D volume-wise DDM^2 and 2D slice-wise DDM^2. Results are shown on the transverse, Coronal, and Sagittal planes. Major differences are highlighted by arrows.

3D volume-wise DDM^2. Other than 2D slice-wise processing schema of DDM^2, 3D volume-wise schema could be another natural design. However, we show its inferior denoising performances via experiments as demonstrated in Figure 6. The residuals between the original and denoised images maintain more structure in the 3D methods than 2D methods, which depicts subpar denoising performance. To match computational burden of 2D slice-wise experiments, we experimented with the largest possible patch size of $16 \times 16 \times 16$. *In supplementary materials, we discussed the impact of patch size and presented additional results.* There are three drawbacks can be observed: *(i)* Volume-wise DDM^2 requires larger networks with $3\times$ model parameters; *(ii)* Contextual dependencies in the longitudinal dimension (across slices) might be weaker for datasets acquired with thick slices due to varying noise characteristics within and across slices. This limits 3D networks, whereas 2D networks can capture the information within the isotropic slice; *(iii)* Most advances in diffusion models focus on 2D sequential acquisitions, which can be easily adopted to improve DDM^2.

5 Discussions

MRI noise considerations. Noise in MRI typically follows a Rician distribution for background regions and transitions to Gaussian in foreground regions where the SNR$\geq$3 (Andersen, 1996; Robson et al., 2008). Moreover, linear image reconstructions result in complex Gaussian noise remaining Gaussian in the image domain also (Robson et al., 2008). Modern image acceleration techniques simply perturb the magnitude of noise spatially (based on coil and subject patterns) but to not modify the Gaussian distributions (Pruessmann et al., 1999). Consequently, using DDM^{2}for denoising relevant foreground image regions is consistent with the underlying image noise statistics. For future work in non-linear image reconstructions (such as learning-based reconstructions (Chaudhari et al., 2020)), our framework still allows data-driven fitting of different noise profiles with a corresponding noise model and score function suggested by Noise2Score.

Limitations. As with most DDPM applications, DDM^2 is primarily limited by slow inference durations. However, one can switch to faster networks or explore acceleration techniques as in (Song et al., 2021a). Moreover, DDM^2 may hallucinate actual lesions due to its generation nature. While this was not seen in our results, it can be mitigated by enforcing data consistency during inference. Systematic radiologist-driven studies may also assess anatomic integrity and expert quality metrics.

6 Conclusion

In this paper, we present DDM^2, a self-supervised denoising method for MRI. Our three-stage method achieves MRI denoising via a generative approach that not only reduces noisy signals but also restores anatomical details. In `Stage I`, a denoising function is trained to capture an initial noise distribution. In `Stage II`, we fit a noise model to match the noisy input with an intermediate state in the diffusion Markov chain. In `Stage III`, we train a diffusion model to generate clean image approximations in an unsupervised manner. In experiments, we demonstrate that DDM^2 outperforms existing state-of-the-art denoisers for diffusion MRI methods both qualitatively and quantitatively.

ETHICS STATEMENT

All datasets used in this study were acquired under institutional review board (IRB) approval for human subjects research. No additional metadata is provided to enable re identification of the individuals involved. The proposed denoising technique has the ability of reducing MRI scan time or enabling noisier data acquisitions on older MRI scanners. This presents a large opportunity to make life-saving MRI technology available and more accessible to patient populations that would not ordinarily be exposed to it due to initial infrastructure and maintenance costs of expensive MRI scanners. Further clinical validation of the proposed technique will be necessary prior to routine usage in these scenarios, however, in order to ensure lack of visual hallucinations and comparable diagnostic outcomes compared to the current standards of care.

REPRODUCIBILITY STATEMENT

Our proposed method is fully reproducible with respect to the main results as reported in section 4.2. For methodology & implementation details, readers are referred to our source codes, which are also submitted as parts of the supplementary materials. These will be hosted on a public GitHub repository following publication of the work. Our analyses are performed on three publicly available datasets and we will open source the fourth dataset (gSlider) upon publication of the work.

REFERENCES

Philip M Adamson, Beliz Gunel, Jeffrey Dominic, Arjun D Desai, Daniel Spielman, Shreyas Vasanawala, John M. Pauly, and Akshay Chaudhari. SSFD: Self-supervised feature distance as an MR image reconstruction quality metric. In *NeurIPS 2021 Workshop on Deep Learning and Inverse Problems*, 2021. URL https://openreview.net/forum?id=dgMvTzf6M_3.

Anders H Andersen. On the rician distribution of noisy mri data. *Magnetic resonance in medicine*, 36(2):331–332, 1996.

Joshua Batson and Loïc Royer. Noise2self: Blind denoising by self-supervision. In Kamalika Chaudhuri and Ruslan Salakhutdinov (eds.), *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, volume 97 of *Proceedings of Machine Learning Research*, pp. 524–533. PMLR, 2019. URL http://proceedings.mlr.press/v97/batson19a.html.

Denis Le Bihan. Looking into the functional architecture of the brain with diffusion MRI. *Nature Reviews Neuroscience*, 4(6):469–480, June 2003. doi: 10.1038/nrn1119. URL https://doi.org/10.1038/nrn1119.

Denis Le Bihan, Cyril Poupon, Alexis Amadon, and Franck Lethimonnier. Artifacts and pitfalls in diffusion MRI. *Journal of Magnetic Resonance Imaging*, 24(3):478–488, 2006. doi: 10.1002/jmri.20683. URL https://doi.org/10.1002/jmri.20683.

Akshay S. Chaudhari, Kathryn J. Stevens, Jeff P. Wood, Amit K. Chakraborty, Eric K. Gibbons, Zhongnan Fang, Arjun D. Desai, Jin Hyung Lee, Garry E. Gold, and Brian A. Hargreaves. Utility of deep learning super-resolution in the context of osteoarthritis MRI biomarkers. *Journal of Magnetic Resonance Imaging*, 51(3):768–779, July 2019. doi: 10.1002/jmri.26872. URL https://doi.org/10.1002/jmri.26872.

Akshay S. Chaudhari, Christopher M. Sandino, Elizabeth K. Cole, David B. Larson, Garry E. Gold, Shreyas S. Vasanawala, Matthew P. Lungren, Brian A. Hargreaves, and Curtis P. Langlotz. Prospective deployment of deep learning in mri: A framework for important considerations, challenges, and recommendations for best practices. *Journal of Magnetic Resonance Imaging*, 54 (2):357–371, August 2020. doi: 10.1002/jmri.27331. URL https://doi.org/10.1002/jmri.27331.

Nanxin Chen, Yu Zhang, Heiga Zen, Ron J. Weiss, Mohammad Norouzi, and William Chan. Wavegrad: Estimating gradients for waveform generation. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021. URL https://openreview.net/forum?id=NsMLjcFaO8O.

Hyungjin Chung, Byeongsu Sim, and Jong Chul Ye. Come-closer-diffuse-faster: Accelerating conditional diffusion models for inverse problems through stochastic contraction. *CoRR*, abs/2112.05146, 2021. URL https://arxiv.org/abs/2112.05146.

Hyungjin Chung, Eun Sun Lee, and Jong Chul Ye. MR image denoising and super-resolution using regularized reverse diffusion. *CoRR*, abs/2203.12621, 2022. doi: 10.48550/arXiv.2203.12621. URL https://doi.org/10.48550/arXiv.2203.12621.

Mohammad Zalbagi Darestani, Akshay S Chaudhari, and Reinhard Heckel. Measuring robustness in deep learning based compressive sensing. In Marina Meila and Tong Zhang (eds.), *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 2433–2444. PMLR, 18–24 Jul 2021. URL https://proceedings.mlr.press/v139/darestani21a.html.

Bradley Efron. Tweedie's formula and selection bias. *Journal of the American Statistical Association*, 106(496):1602–1614, 2011. doi: 10.1198/jasa.2011.tm11181.

Shreyas Fadnavis, Joshua Batson, and Eleftherios Garyfallidis. Patch2self: Denoising diffusion MRI with self-supervised learning8203. In Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin (eds.), *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/bc047286b224b7bfa73d4cb02de1238d-Abstract.html.

Eleftherios Garyfallidis, Matthew Brett, Bagrat Amirbekian, Ariel Rokem, Stéfan van der Walt, Maxime Descoteaux, and Ian Nimmo-Smith. Dipy, a library for the analysis of diffusion MRI data. *Frontiers Neuroinformatics*, 8:8, 2014. doi: 10.3389/fninf.2014.00008. URL https://doi.org/10.3389/fninf.2014.00008.

Eric K. Gibbons, Kyler K. Hodgson, Akshay S. Chaudhari, Lorie G. Richards, Jennifer J. Majersik, Ganesh Adluru, and Edward V.R. DiBella. Simultaneous NODDI and GFA parameter map generation from subsampled q-space imaging using deep learning. *Magnetic Resonance in Medicine*, 81(4):2399–2411, November 2018. doi: 10.1002/mrm.27568. URL https://doi.org/10.1002/mrm.27568.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin (eds.), *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/4c5bcfec8584af0d967f1ab10179ca4b-Abstract.html.

Tao Huang, Songjiang Li, Xu Jia, Huchuan Lu, and Jianzhuang Liu. Neighbor2neighbor: Self-supervised denoising from single noisy images. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021, virtual, June 19-25, 2021*, pp. 14781–14790. Computer Vision Foundation / IEEE, 2021. URL https://openaccess.thecvf.com/content/CVPR2021/html/Huang_Neighbor2Neighbor_Self-Supervised_Denoising_From_Single_Noisy_Images_CVPR_2021_paper.html.

Aapo Hyvärinen. Estimation of non-normalized statistical models by score matching. *J. Mach. Learn. Res.*, 6:695–709, 2005. URL http://jmlr.org/papers/v6/hyvarinen05a.html.

Elena A. Kaye, Emily A. Aherne, Cihan Duzgol, Ida Häggström, Erich Kobler, Yousef Mazaheri, Maggie M. Fung, Zhigang Zhang, Ricardo Otazo, Hebert A. Vargas, and Oguz Akin. Accelerating prostate diffusion-weighted MRI using a guided denoising convolutional neural network: Retrospective feasibility study. *Radiology: Artificial Intelligence*, 2(5):e200007, August 2020. doi: 10.1148/ryai.2020200007. URL https://doi.org/10.1148/ryai.2020200007.

Kwanyoung Kim and Jong Chul Ye. Noise2score: Tweedie's approach to self-supervised image denoising without clean images. *CoRR*, abs/2106.07009, 2021. URL https://arxiv.org/abs/2106.07009.

Zhifeng Kong, Wei Ping, Jiaji Huang, Kexin Zhao, and Bryan Catanzaro. Diffwave: A versatile diffusion model for audio synthesis. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021. URL https://openreview.net/forum?id=a-xFK8Ymz5J.

Alexander Krull, Tim-Oliver Buchholz, and Florian Jug. Noise2void - learning denoising from single noisy images. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*, pp. 2129–2137. Computer Vision Foundation / IEEE, 2019. doi: 10.1109/CVPR.2019.00223. URL http://openaccess.thecvf.com/content_CVPR_2019/html/Krull_Noise2Void_-_Learning_Denoising_From_Single_Noisy_Images_CVPR_2019_paper.html.

Samuli Laine, Tero Karras, Jaakko Lehtinen, and Timo Aila. High-quality self-supervised deep image denoising. In Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d'Alché-Buc, Emily B. Fox, and Roman Garnett (eds.), *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pp. 6968–6978, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/2119b8d43eafcf353e07d7cb5554170b-Abstract.html.

Max W. Y. Lam, Jun Wang, Dan Su, and Dong Yu. BDDM: Bilateral denoising diffusion models for fast and high-quality speech synthesis. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=L7wzpQttNO.

Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2noise: Learning image restoration without clean data. In Jennifer G. Dy and Andreas Krause (eds.), *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*, volume 80 of *Proceedings of Machine Learning Research*, pp. 2971–2980. PMLR, 2018. URL http://proceedings.mlr.press/v80/lehtinen18a.html.

Shitong Luo and Wei Hu. Diffusion probabilistic models for 3d point cloud generation. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021, virtual, June 19-25, 2021*, pp. 2837–2845. Computer Vision Foundation / IEEE, 2021. URL https://openaccess.thecvf.com/content/CVPR2021/html/Luo_Diffusion_Probabilistic_Models_for_3D_Point_Cloud_Generation_CVPR_2021_paper.html.

Kenneth Marek, Danna Jennings, Shirley Lasch, Andrew Siderowf, Caroline Tanner, Tanya Simuni, Chris Coffey, Karl Kieburtz, Emily Flagg, Sohini Chowdhury, Werner Poewe, Brit Mollenhauer, Paracelsus-Elena Klinik, Todd Sherer, Mark Frasier, Claire Meunier, Alice Rudolph, Cindy Casaceli, John Seibyl, Susan Mendick, Norbert Schuff, Ying Zhang, Arthur Toga, Karen Crawford, Alison Ansbach, Pasquale De Blasio, Michele Piovella, John Trojanowski, Les Shaw, Andrew Singleton, Keith Hawkins, Jamie Eberling, Deborah Brooks, David Russell, Laura Leary, Stewart Factor, Barbara Sommerfeld, Penelope Hogarth, Emily Pighetti, Karen Williams, David Standaert, Stephanie Guthrie, Robert Hauser, Holly Delgado, Joseph Jankovic, Christine Hunter, Matthew Stern, Baochan Tran, Jim Leverenz, Marne Baca, Sam Frank, Cathi-Ann Thomas, Irene Richard, Cheryl Deeley, Linda Rees, Fabienne Sprenger, Elisabeth Lang, Holly Shill, Sanja Obradov, Hubert Fernandez, Adrienna Winters, Daniela Berg, Katharina Gauss, Douglas Galasko, Deborah Fontaine, Zoltan Mari, Melissa Gerstenhaber, David Brooks, Sophie Malloy, Paolo Barone, Katia Longo, Tom Comery, Bernard Ravina, Igor Grachev, Kim Gallagher, Michelle Collins, Katherine L. Widnell, Suzanne Ostrowizki, Paulo Fontoura, Tony Ho, Johan Luthman, Marcel van der Brug, Alastair D. Reith, and Peggy Taylor. The parkinson progression marker initiative (ppmi). *Progress in Neurobiology*, 95(4):629–635, 2011. ISSN 0301-0082. doi: https://doi.org/10.1016/j.pneurobio.2011.09.005. URL https://www.sciencedirect.com/science/article/pii/S0301008211001651. Biological Markers for Neurodegenerative Diseases.

Allister Mason, James Rioux, Sharon E. Clarke, Andreu Costa, Matthias Schmidt, Valerie Keough, Thien Huynh, and Steven Beyea. Comparison of objective image quality metrics to expert radiologists' scoring of diagnostic quality of MR images. *IEEE Transactions on Medical Imaging*, 39(4):1064–1072, April 2020. doi: 10.1109/tmi.2019.2930338. URL https://doi.org/10.1109/tmi.2019.2930338.

Anish Mittal, Anush K. Moorthy, and Alan C. Bovik. Blind/referenceless image spatial quality evaluator. In Michael B. Matthews (ed.), *Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers, ACSCC 2011, Pacific Grove, CA, USA, November 6-9, 2011*, pp. 723–727. IEEE, 2011. doi: 10.1109/ACSSC.2011.6190099. URL https://doi.org/10.1109/ACSSC.2011.6190099.

Tongyao Pang, Huan Zheng, Yuhui Quan, and Hui Ji. Recorrupted-to-recorrupted: Unsupervised deep learning for image denoising. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021, virtual, June 19-25, 2021*, pp. 2043–2052. Computer Vision Foundation / IEEE, 2021. URL https://openaccess.thecvf.com/content/CVPR2021/html/Pang_Recorrupted-to-Recorrupted_Unsupervised_Deep_Learning_for_Image_Denoising_CVPR_2021_paper.html.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32: 8026–8037, 2019.

Mangal Prakash, Alexander Krull, and Florian Jug. Fully unsupervised diversity denoising with convolutional variational autoencoders. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021. URL https://openreview.net/forum?id=agHLCOBM5jP.

Klaas P Pruessmann, Markus Weiger, Markus B Scheidegger, and Peter Boesiger. Sense: sensitivity encoding for fast mri. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 42(5):952–962, 1999.

Philip M Robson, Aaron K Grant, Ananth J Madhuranthakam, Riccardo Lattanzi, Daniel K Sodickson, and Charles A McKenzie. Comprehensive quantification of signal-to-noise ratio and g-factor for image-based and k-space-based parallel imaging reconstructions. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 60(4):895–907, 2008.

Ariel Rokem. Stanford HARDI surfaces. 10 2016. doi: 10.6084/m9.figshare.4033746.v1. URL https://figshare.com/articles/dataset/Stanford_HARDI_surfaces/4033746.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241. Springer, 2015.

Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J. Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. *CoRR*, abs/2104.07636, 2021. URL https://arxiv.org/abs/2104.07636.

Tim Salimans and Jonathan Ho. Progressive distillation for fast sampling of diffusion models. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=TIdIXIpzhoI.

Kawin Setsompop, Qiuyun Fan, Jason Stockmann, Berkin Bilgic, Susie Huang, Stephen F. Cauley, Aapo Nummenmaa, Fuyixue Wang, Yogesh Rathi, Thomas Witzel, and Lawrence L. Wald. High-resolution in vivo diffusion imaging of the human brain with generalized slice dithered enhanced resolution: Simultaneous multislice (gSlider-SMS). *Magnetic Resonance in Medicine*, 79(1): 141–151, March 2017. doi: 10.1002/mrm.26653. URL https://doi.org/10.1002/mrm.26653.

Jascha Sohl-Dickstein, Eric A. Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In Francis R. Bach and David M. Blei (eds.), *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, volume 37 of *JMLR Workshop and Conference Proceedings*, pp. 2256–2265. JMLR.org, 2015. URL http://proceedings.mlr.press/v37/sohl-dickstein15.html.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021a. URL https://openreview.net/forum?id=St1giarCHLP.

Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. In Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d'Alché-Buc, Emily B. Fox, and Roman Garnett (eds.), *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pp. 11895–11907, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/3001ef257407d5a371a96dcd947c7d93-Abstract.html.

Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021b. URL https://openreview.net/forum?id=PxTIG12RRHS.

Yang Song, Liyue Shen, Lei Xing, and Stefano Ermon. Solving inverse problems in medical imaging with score-based generative models. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=vaRCHVj0uGI.

Dmitry Ulyanov, Andrea Vedaldi, and Victor S. Lempitsky. Deep image prior. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018*, pp. 9446–9454. Computer Vision Foundation / IEEE Computer Society, 2018. doi: 10.1109/CVPR.2018.00984. URL http://openaccess.thecvf.com/content_cvpr_2018/html/Ulyanov_Deep_Image_Prior_CVPR_2018_paper.html.

Daniel Watson, William Chan, Jonathan Ho, and Mohammad Norouzi. Learning fast samplers for diffusion models by differentiating through sample quality. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=VFBjuF8HEp.

Jeffrey P. Woodard and Monica P. Carley-Spencer. No-reference image quality metrics for structural MRI. *Neuroinformatics*, 4(3):243–262, 2006. doi: 10.1385/NI:4:3:243. URL https://doi.org/10.1385/NI:4:3:243.

Minkai Xu, Lantao Yu, Yang Song, Chence Shi, Stefano Ermon, and Jian Tang. Geodiff: A geometric diffusion model for molecular conformation generation. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=PzcvxEMzvQC.

Linqi Zhou, Yilun Du, and Jiajun Wu. 3d shape generation and completion through point-voxel diffusion. In *2021 IEEE/CVF International Conference on Computer Vision, ICCV 2021, Montreal, QC, Canada, October 10-17, 2021*, pp. 5806–5815. IEEE, 2021. doi: 10.1109/ICCV48922.2021.00577. URL https://doi.org/10.1109/ICCV48922.2021.00577.

A CODE AND DATA

Code. We provide our implementations of DDM^2 as part of the supplementary materials for: *(i)* training DDM^{2}on a dataset; *(ii)* using DDM^2 for denoising a dataset. By following the instructions provided in README.md, users are guaranteed to reproduce our main results as reported in section 4.2.

Data. *Note that there is a lack of consensus on image quality metrics, especially in an unsupervised settings. Therefore, no single quantitative metric can be adopted to reflect on the superiority of DDM^2.* We hence provide direct access to the noisy input data and the denoised data for all of the four datasets: gSlider, Sherbrooke, Stanford HARDI, and PPMI through an anonymous link. *Readers are encouraged to download the data and view them through any on-line viewing tools or apply any quantitative measurements as they wish.*

LINK TO DATA: `https://figshare.com/s/6275f40c32f67e3b6083`

B TABLE OF SYMBOLS

For notation simplicity, we adopted alphabetic symbols in this paper to represent essential components in our framework. For better symbol-name correspondences, here we justify implications of all symbols used in the paper in Table 1 to help readers comprehend.

Symbols	Implication
$\mathbf{x}$	target noisy input
$\mathbf{x}'$	noisy input prior
$\mathbf{y}$	clean ground truth of $\mathbf{x}$
ϵ	residual noise
$\bar{\mathbf{y}}$	Stage I estimated clean image of $\mathbf{x}$
$\bar{\epsilon}$	Stage I estimated residual noise
$\mu_{\bar{\epsilon}}$	arithmetic mean of $\bar{\epsilon}$
$\phi(\cdot)$	denoising function in Stage I
$\mathcal{F}(\cdot)$	denoising function in Stage III
S	state in the Markov chain
$\mathcal{G}$	noise model
σ	standard deviation of noise distribution
β	noise schedule in diffusion models

Table 1: Table of symbols used in this paper.

C EXPERIMENTAL DETAILS

Implementation details. We implement both denoising functions Φ and $\mathcal{F}$ via U-Net (Ronneberger et al., 2015) with modifications suggested in (Song et al., 2021b; Saharia et al., 2021). In `Stage I` we set the size of $\{\mathbf{x}'\}$ $n = 2$. In `Stage III`, inspired by (Song & Ermon, 2019; Chen et al., 2021), we train $\mathcal{F}$ to condition on $\bar{\alpha}_t, t \sim U(1, T)$. The Adam optimizer was used to optimize both networks with a fixed learning rate of $1e^{-4}$ and a batch size of 32. We trained Φ for $1e^4$ steps and $\mathcal{F}$ for $1e^5$ steps from scratch. All experiments were performed on RTX GeForce 2080-Ti GPUs in PyTorch (Paszke et al., 2019).

Evaluation details. Our DDM^{2}is compared against 4 state-of-the-art unsupervised denoising methods. *For fair comparisons, we aligned the total number of parameters in DDM^2 and all comparison methods: For U-Net based methods, we adopted the exact architecture as in DDM^2; For MLP based methods, we aligned the number of parameters. (i)* Deep Image Prior (Ulyanov et al., 2018) trains a neural network on random inputs to overfit a specific noisy image. The network is claimed to restore low-level consistencies exist in the target image at an intermediate state of the training process. In our experiments, since there is no reliable quantitative metrics like PSNR, we pause network training based on SNR and the visual quality of the reconstructions; *(ii)* Noise2Noise (Lehtinen et al., 2018)

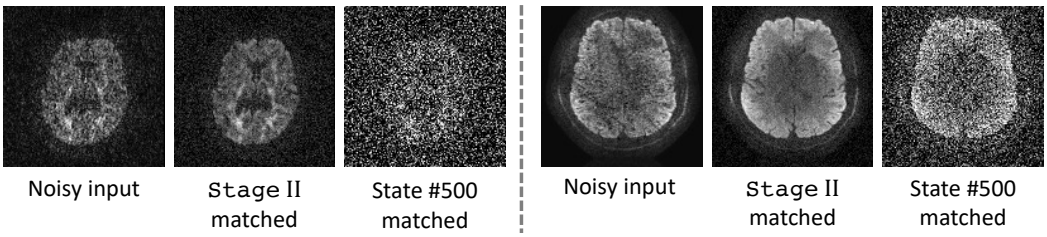

Figure 7: Visual demonstration of different matches in the Markov chain. Our `Stage II` matched state generates samples that look sufficiently close to the noisy input, while arbitrarily chosen states (e.g. state #500) cannot match the noisy input well.

constructs training pairs of two independent noisy measurements of the same target and trains a network to transform one measurement to the other. 4D MRI by their nature own independent noisy samples acquired at different gradient directions. In our experiments, we train Noise2Noise model by using the same slices at different volumes; *(iii)* Patch2Self (Fadnavis et al., 2020) generalizes Noise2Noise and Noise2Self for patch-wise MRI denoising. In our experiments, we follow their official implementation and adopted MLP as the regressor and patch radiance 0 for their best performance; *(iv)* Noise2Score (Kim & Ye, 2021) trains a neural network to estimate a noise-dependent score function to find the clean posterior distribution of noise inputs. In our experiments, we estimate the score function under the same assumption of Gaussian noise.

D INTUITIVE DEMONSTRATION OF MARKOV CHAIN STATE MATCHING

In section 3.3, we claim that a matched state at $\mathbf{t}$ denotes the minimum p-norm distance achieved between $\sqrt{\beta_t}$ and σ (the standard deviation of the Gaussian fitted on the input image $\mathbf{x}$). To verify the match intuitively, we show visual comparisons between $\mathbf{x}$ and the noise corrupted `stage I` result $\mathbf{c} = q(\mathrm{S}_\mathbf{t}|\mathrm{S}_0) = \sqrt{\bar{\alpha}_\mathbf{t}}\bar{\mathbf{y}} + \epsilon$, $\epsilon \sim \mathcal{N}(\mathbf{0}, \beta_\mathbf{t}\mathbf{I})$ as a prior sample at the state $\mathrm{S}_\mathbf{t}$. A successful match leads to close image distributions ($\mathbf{x}$ and $\mathbf{c}$ in similar appearances), while $\mathbf{x}$ and $\mathbf{c}$ may look considerably different in unsuccessful matches.

As shown in Figure 7, compared to a naive representation at the middle Markov chain (state #500, as studied in section 4.3), our `stage II` matches the input in the Markov chain better. The successful match indicates there exists at least one possible sample drawn from the posterior distribution at the matched state that is sufficiently close to the noisy input. Therefore our claim in section 3.1 can be well proved. Without a proper state matching strategy, the final generation results will be in poor quality (section 4.3 **B**).

E NOISE SCHEDULE WITH REVERSE WARM-UP

Noise levels in most of MRI datasets are small. Matching them in `Stage II` with commonly used 'warm-up' based noise schedule β leads to few transitions in the sampling process and eventually poor generation quality. Considering this, we adopted a 'reverse warm-up' setting for β.

We compare $\lambda_1^{(t)} = \sqrt{\prod_{i=1}^{t}(1-\beta_t)}$ of the warm-up schedule ($\beta_1 = 1e^{-2}$, $\beta_T = 5e^{-5}$, ratio$= 0.7$) and our reverse warm-up schedule ($\beta_1 = 5e^{-5}, \beta_T = 1e^{-2}$, ratio$= 0.7$) in Figure 8. Recall that $\mathbf{x} = \lambda_1\mathbf{y} + \lambda_2\mathbf{z}$ (Eq. 1). During the sampling process, our reverse warm-up based strategy raises the value of λ_1 faster than the baseline strategy, hence reveals the underlying anatomical structures in MRIs much faster. This allows us to represent noisy inputs at a later state (e.g. state #200) in the Markov chain to enable more reverse process transitions. The baseline warm-up based schedule, on the other hand, represents noisy inputs at a much earlier state (e.g. state #5). In this way, the final denoising result will be generated in very few transitions and eventually in inferior quality.

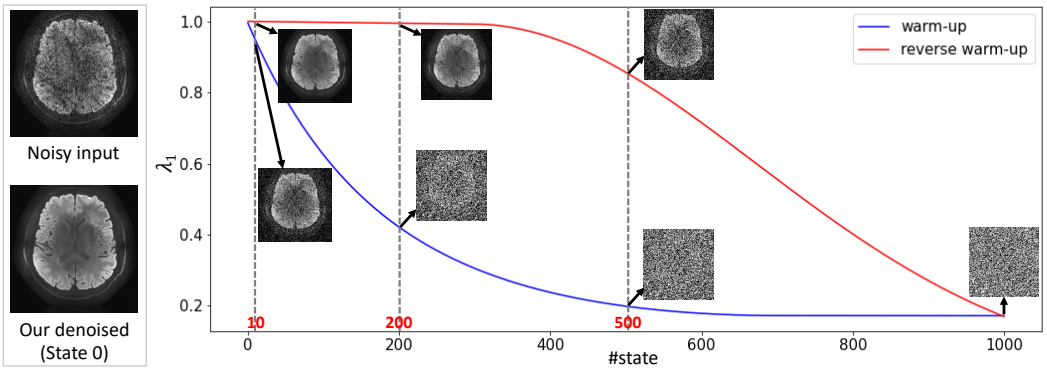

Figure 8: Comparison of the baseline warm-up based schedule and our reverse warm-up based schedule. Our modified schedule allows noisy inputs to be represented at a later state than the baseline, which leads to more transition steps and better denoising quality.

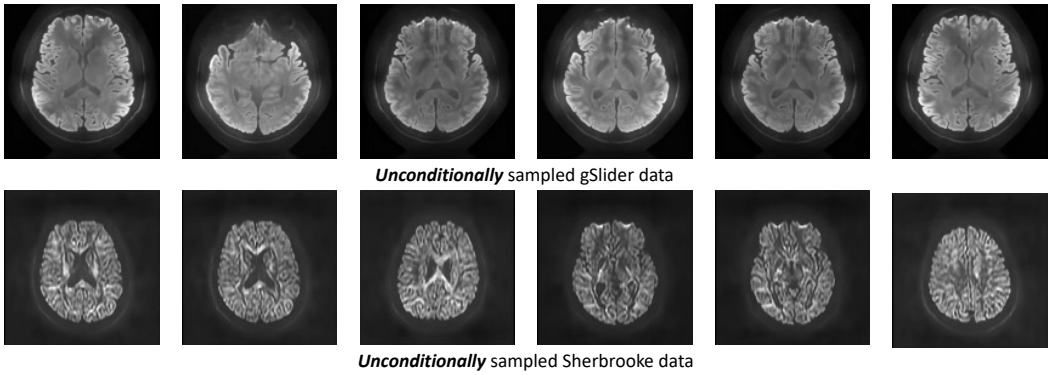

Figure 9: Unconditional sampling of DDM^2 on our gSlider and the Sherbrooke datasets. All of the generations are realistic due to the restored anatomical details.

F Unconditional Sampling of DDM^2

Differing from all competing methods, besides functioning as a denoiser, our diffusion model based method is capable of generating new (nonexistent) images that fit a given dataset. We here conducted experiments to generate realistic MRI scans through *unconditional sampling* from our trained diffusion model. The generation results for our gSlider dataset and the Sherbrooke dataset are shown in Figure 9.

These forged images successfully restored essential anatomical details in the MRI scans, which can be used to enlarge the dataset further to assist down-streaming tasks.

G Outlier Detection

As a diffusion model based method, the denoising performance of DDM^2 is relatively input-dependent. In order to detect on which cases DDM^2 performs the best and which the worst, we here present a simple technique that can be adopted during DDM^2 inference to detect poorly denoised outliers.

Specifically, we compute RMSE scores between the initial noisy input ($\mathbf{S_t}$) and every image ($\mathbf{S_{t\to 0}}$) generated in the reverse diffusion process. For good denoising, RMSE values will increase significantly over iterations (represents denoised results dissimilar to inputs). For poorly denoised outliers, RMSE values will change slowly (represents denoising results similar to the original noisy measurement).

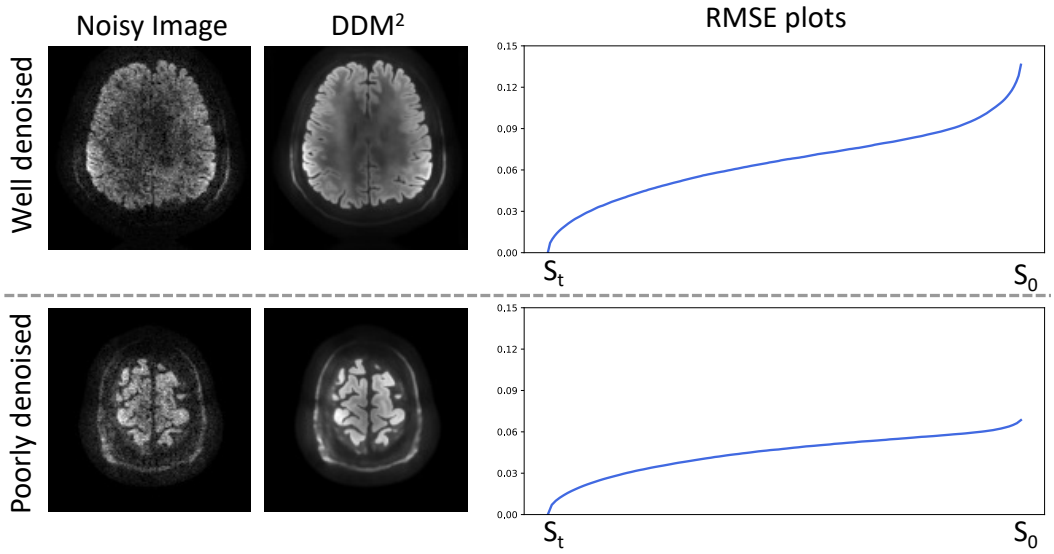

Figure 10: RMSE plot comparisons between well denoised and poorly denoised slices.

We provide demo results in Figure 10. Clearly, the poorly denoised slice (bottom slice) cannot achieve as high RMSE value as the well denoised one (top slice). This technique is easy in implementation and can be adopted to detect poorly denoised outliers by DDM^2 effectively.

H Additional Results

Qualitative results. Additional qualitative comparisons on all of the four datasets are presented in Figure 11. Obviously, DDM^2 not only reduces noise the most but also restores anatomical details the best.

3D volume-wise DDM^2. In section 3.2, we present our slice-wise improvements over the patch-wise processing (Fadnavis et al., 2020). To validate our design choice, in section 5, we presented an alternative design of DDM^{2}that functions on 3D volume patches rather than 2D slices. An intuitive comparison between the two designs is outlined in Figure 12.

The best results for 3D volume-wise DDM^2 were achieved at patch size $16 \times 16 \times 16$—also the largest volume patches we can afford in our GPU memory. Empirically, we found that when functioning on small-size patches, diffusion models fail to denoise but instead overfit and return the exact inputs. Comparisons at the transverse, coronal, and sagittal planes between the 2D slice-wise DDM^{2}and 3D volume-wise DDM^{2}are shown in Figure 6 and Figure 13.

`Stage I` Comparisons. In Figure 14 we present additional qualitative comparisons between the `Stage I` results and our final results. Clearly, `Stage I` of DDM^2 alone cannot restore indicative geometries and textures of the brain. With our `Stage II` and `Stage III`, the diffusion model is able to generate such missing contents for more promising denoising results.

Noise residual comparisons. In Figure 15 we present noise residual comparisons between Patch2Self and DDM^2. Our method suppresses more noise and does not show any anatomical structures in the corresponding residual plots.

Recall that in section A, we provide access to the source denoised results. One can have better sense of the superiority of DDM^2 by scrutinizing the denoised data directly.

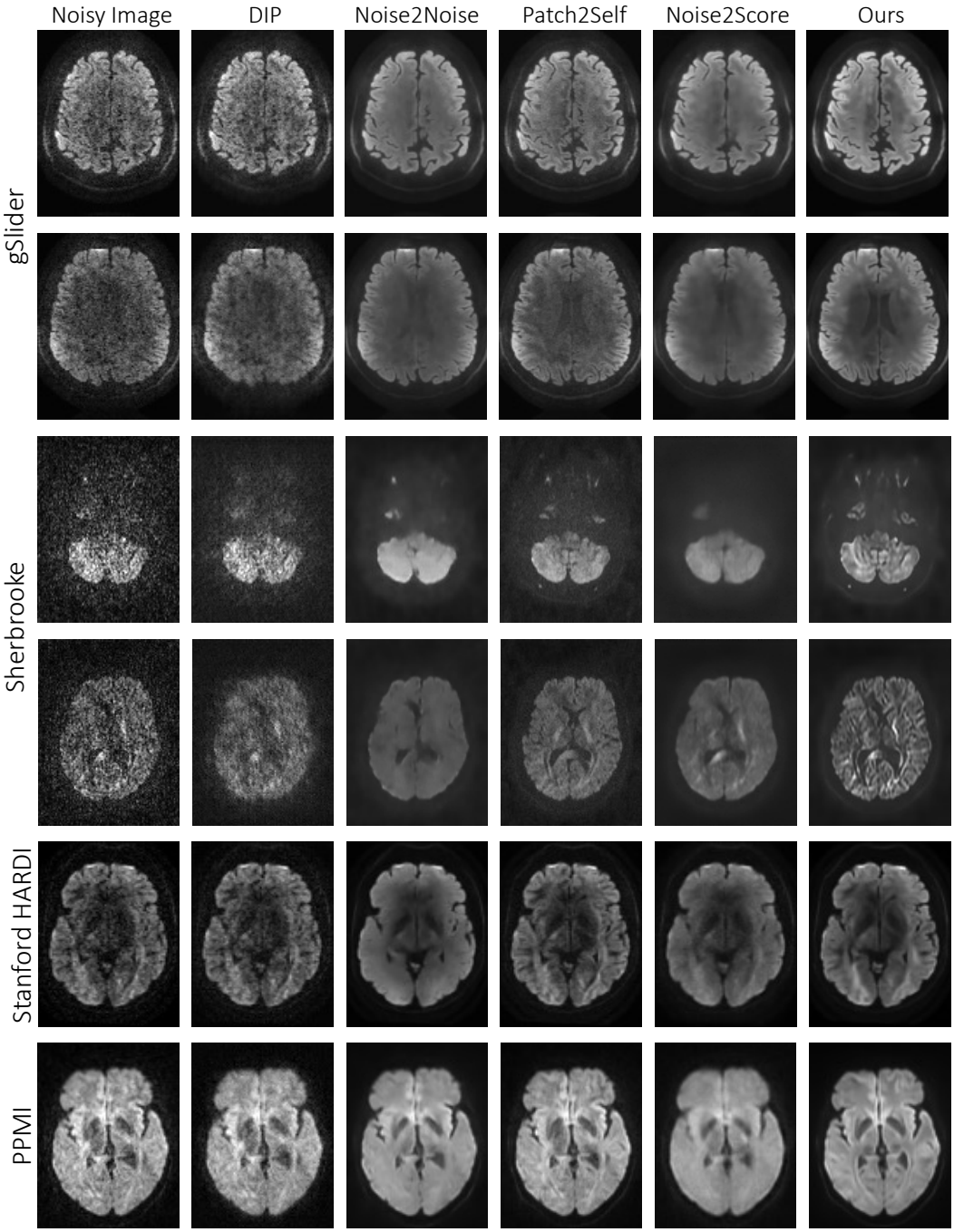

Figure 11: Additional qualitative results. DDM^2 demonstrates superior denoising quality in both image restorations and faithful reproduction of the original image contrasts without image hallucination or suppression. Best viewed in zoomed in.

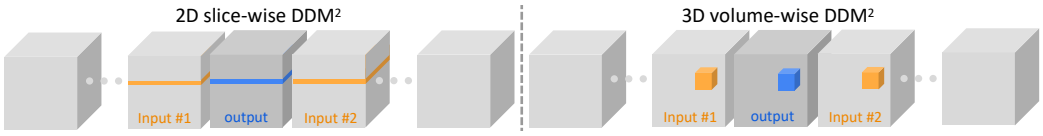

Figure 12: Intuitive comparisons between **Left:** 2D slice-wise DDM^2(adopted) and **Right:** 3D volume-wise DDM^2.

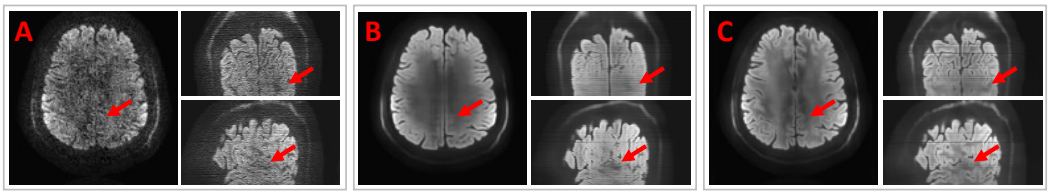

Figure 13: Additional comparisons at the transverse, Coronal, and Sagittal planes. **A:** Noisy input. **B:** results of volume-wise DDM^2. **C:** results of slice-wise DDM^2.

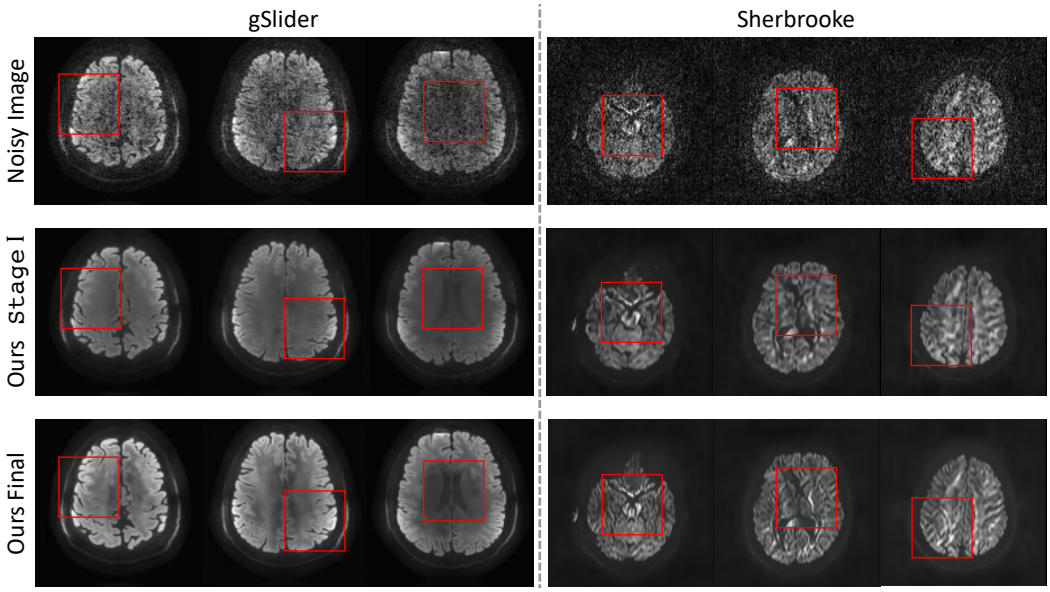

Figure 14: Additional comparisons between DDM^2 `Stage I` results and DDM^2 final results on our gSlider dataset and the Sherbrooke dataset. Major differences are highlighted.

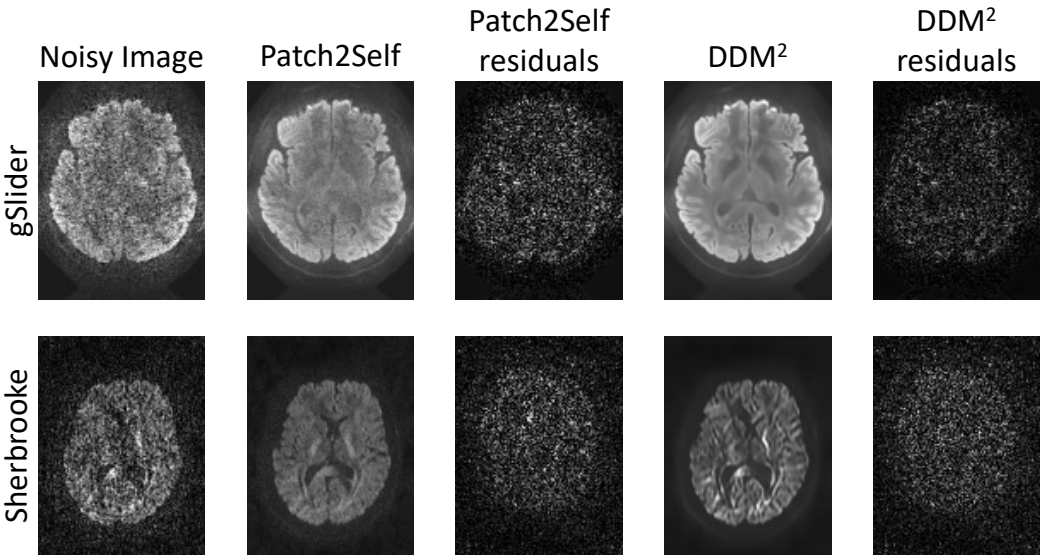

Figure 15: Residual noise comparisons between Patch2Self and DDM^2.