Uncertainty-Propelled Physics-MAE Fusion for Self-Supervised Diffusion-Weighted Image Denoising

Zeyu Deng [1], Lihui Wang[1]*, Xi Tao[1], Qijian Chen[1], Ying Cao[1], XuLinHu[1], YingFeng Ou[1],

[1] Key Laboratory of Advanced Medical Imaging and Intelligent Computing in Guizhou Province,Engineering Research Center of Text Computing & Cognitive Intelligence, Ministry of Education, College of Computer Science and Technology,Guizhou University,Guiyang,China
gs.zydeng21@gzu.edu.cn, lhwang2@gzu.edu.cn

Abstract

The inherently low signal-to-noise ratio (SNR) in diffusion-weighted (DW) imaging fundamentally impedes precise tissue microstructure characterization, rendering effective noise suppression a persistent challenge. Existing denoising methods frequently suffer from over-smoothing or distortion of microstructure information when handling spatially correlated or severe noise. To address these limitations, we propose UP^2-MAE fusion model, a self-supervised DWI denoising method based on Uncertainty-Propelled Physics and Masked Auto-Encoder (MAE) fusion. This framework integrates two complementary branches: one leverages MAE to suppress noise through robust feature learning, while the other constructs uncorrelated noisy pairs using diffusion tensor imaging (DTI) physics and denoises them via a Noise2Noise approach, which can preserve texture details by exploiting angular relationships between DW images along diffusion encoding directions. To fully integrate the strengths of both branches, an uncertainty-propelled fusion strategy based on maximum likelihood estimation is proposed to derive the final denoised output. In addition, to further promote the performance, uncertainty-guided reconstruction and consistency loss are presented. Evaluations against state-of-the-art denoising methods on both simulated and acquired DW datasets confirm the efficacy of our approach.

Code — https://github.com/strawberry1996/Up2-MAE

Introduction

Diffusion magnetic resonance imaging (dMRI) is an essential technique for non-invasively mapping the microstructure of in vivo tissues (Basser, Mattiello, and LeBihan 1994; Le Bihan 2003). However, diffusion-weighted (DW) images acquired using this technique are inherently susceptible to noise contamination, which compromises the reliability of subsequent microstructural analysis (Le Bihan et al. 2006). Consequently, developing effective denoising methods for DW images is critical for achieving reliable microstructure characterization. Conventional denoising approaches for DW images, such as non-local means filtering (Wiest-Daesslé et al. 2007; Coupé et al. 2012; Chen et al. 2016) and principal component analysis (PCA) (Manjón et al. 2013; Ramos-Llordén et al. 2021; Mosso et al. 2022;

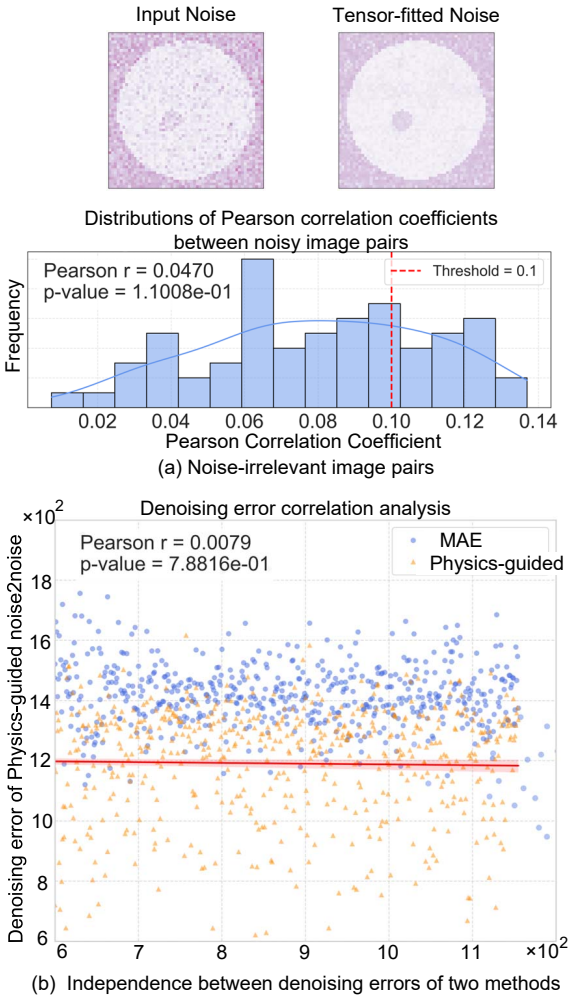


Figure 1: Motivation of UP^2-MAE. (a) Noise independence between raw and DTI-fitted DW images; (b) Uncorrelated denoising errors across methods, with red lines indicating the correlation curve.

*Lihui Wang is the corresponding author.

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Olesen et al. 2023), primarily exploit local spatial coherence and inherent data redundancy within the dMRI data to denoise. While capable of reducing some noise, their effectiveness is highly dependent on careful hyperparameter tuning. Furthermore, under low signal-to-noise ratio (SNR) conditions, these methods frequently result in excessive smoothing or artifacts.

State-of-the-art techniques for denoising DW images primarily employ either supervised or self-supervised learning models. Supervised approaches, such as JD-CNN (Wang et al. 2019) and DnCNN (Zhang et al. 2017), typically generate clean reference images by averaging multiple acquisitions to train convolutional neural networks (CNNs). Although these methods exhibit strong denoising performance, their clinical utility is restricted due to the impractical requirement of repeated acquisitions, which substantially prolongs scan times. Self-supervised learning (SSL) methods offer a promising alternative to overcome the limitations of supervised approaches by eliminating the need for clean reference data. These techniques exploit various noise modeling strategies, including noisy image pairs (Lehtinen et al. 2018; Huang et al. 2021; Mansour and Heckel 2023; Jiang et al. 2024), J-invariance theory (Batson and Royer 2019; Quan et al. 2020; Krull, Buchholz, and Jug 2019; Batson and Royer 2019; Tian et al. 2022b; Wang et al. 2022; Jang et al. 2023; Linhai et al. 2025), generative models (Xiang et al. 2023; Vasylechko, Afacan, and Kurugol 2023), and physical models (Tian et al. 2020, 2022a) to estimate and remove noise effectively. While SSL-based denoisers outperform conventional supervised methods, they still encounter critical challenges in preserving directional microstructure information, especially under spatially correlated or high noise levels. Disruption of angular relationships between DW images along different diffusion encoding directions lead to biased estimations of microstructural parameters, significantly impacting subsequent quantitative analyses.

Motivated by two key observations: noise independence between original DW images and diffusion tensor model fits (Figure 1(a)), and uncorrelated denoising errors across different self-supervised denoising methods (Figure 1(b)), we propose UP^2-MAE, an uncertainty-propelled fusion model to address the aforementioned challenges in DW image denoising, the main contributions are summarized as follows:

1) An uncertainty-aware fusion framework is established through maximum likelihood estimation (MLE), providing a theoretical foundation for multi-branch information integration.
2) The proposed UP^2-MAE architecture integrates two complementary denoising pathways: a physics-guided Noise2Noise pathway that maintains angular fidelity by exploiting diffusion tensor imaging (DTI) principles through noise-irrelevant image pair training, and a masked autoencoder (MAE) pathway learns noise-robust feature representations through self-supervised reconstruction.
3) Within the uncertainty-aware fusion framework, reconstruction outputs from both pathways are dynamically weighted according to their respective uncertainty maps, achieving optimal balance between effective noise reduction and preservation of structural details.
4) To further boost pathway performance, an uncertainty-weighted consistency loss enabling implicit cross-branch guidance is proposed, which enhances both denoising robustness and anatomical accuracy without requiring explicit feature-level interaction.
5) Comprehensive experiments across multiple datasets validate the superiority of our proposed approach over existing methods.

Related work

Current self-supervised denoising approaches can be broadly classified into four paradigms: those leveraging noisy image pairs, methods grounded in J-invariance theory, and approaches incorporating physical models.

Methods based on Noisy Image Pairs: The Noise2Noise (N2N) paradigm (Lehtinen et al. 2018) and its subsequent extensions (Huang et al. 2021; Mansour and Heckel 2023) have opened new possibilities for training denoisers using noisy image pairs that share structural similarity while containing independent noise. Drawing on this framework, Yuan et al. (Yuan et al. 2023) developed SSECNN, a method for denoising DW images through identification of structurally similar pairs in Q-space. Nevertheless, its effectiveness is substantially influenced by both the number of diffusion encoding directions and the degree of structural correspondence between image pairs. Following this work, Jiang et al. (Jiang et al. 2024) introduced the SAN2N approach, which implements Noise2Noise training through localized patch matching in the joint spatial-angular domain of diffusion MRI data. However, similar to SSECNN, this technique demonstrates sensitivity to the angular resolution of the input data. Most recently, Tian et al. (Tian et al. 2025) developed a multi-scale denoising framework that utilizes super-resolution generated image pairs, demonstrating promising performance while remaining constrained by SR reconstruction artifacts. These methods reveals that constructing reliable image pairs that perfectly match in structure while maintaining noise independence is the fundamental challenge in Noise2Noise-based approaches.

Methods based on J-invariance Theory: To address the limitation of noisy image pairs, self-supervised methods based on J-invariance theory (Batson and Royer 2019) (e.g., Self2Self (Quan et al. 2020), Noise2Void (Krull, Buchholz, and Jug 2019), Noise2Self (Batson and Royer 2019), Noise2SR (Tian et al. 2022b), Blind2UNBlind (Wang et al. 2022)) create training pairs by masking or replacing pixels within individual noisy images. However, their reliance on spatial noise independence limits their application to spatially correlated noise, which is prevalent in DW imaging due to parallel imaging, multi-shot acquisition, and k-space sampling (St-Jean, Coupé, and Descoteaux 2016; Henriques et al. 2023). Recent approaches like AP-BSN (Lee, Son, and Lee 2022), PUCA (Jang et al. 2023), Complementary-BSN (Fan et al. 2024) and Replace2Self (Linhai et al. 2025)

handle spatial correlations but fail to account for orientational correlations across diffusion directions, affecting microstructure accuracy. Patch2Self (Fadnavis, Batson, and Garyfallidis 2020) and its variants (Fadnavis et al. 2024) overcome this by using cross-directional information, regressing target signals from other directions while avoiding spatial noise correlations. However, it ignores reference-direction information, potentially biasing intensity restoration.

Methods based on dMRI Physics: DeepDTI (Tian et al. 2020), an early representative method based on dMRI physics, first estimates DTI from acquired multi-directional DW images and identifies six optimal diffusion gradient directions through transformation matrix condition number minimization. This tensor and optimized directions are then used to synthesize clean DW images as training targets. A UNet-like architecture is subsequently trained to map noisy inputs to these synthetic targets. Building upon DeepDTI, SDnDTI(Tian et al. 2022a) employs an ensemble learning strategy, averaging multiple predictions generated by DeepDTI models trained on different subsets of DW images. A significant limitation shared by both DeepDTI and SDnDTI, however, is the potential for substantial deviations of the synthesized targets from a fixed tensor. This deviation becomes particularly pronounced under extremely low SNR conditions.

To address these limitations, we integrate these three complementary approaches into a unified framework that leverages their respective strengths, while employing uncertainty-aware fusion to optimize denoising performance.

Method

Problem Formulation

Let I_n^i denote the noisy DW image along the i-th diffusion encoding direction, and I^i the corresponding noise-free image which is unavailable. The proposed method intends to integrate a data-driven (MAE) and a physics-driven (physics-guided Noise2Noise) branch to denoise the DW image. Assuming that the reconstruction error of data- and physics-driven branch (ϵ_d/ϵ_p) conforms to zero-mean Gaussian distribution, with variance determined by the uncertainty map (σ_d/σ_p), that means:

$$\begin{aligned} \epsilon_d &= I^i - \hat{I}_d^i \propto \mathcal{N}(0, \sigma_d^2) \\ \epsilon_p &= I^i - \hat{I}_p^i \propto \mathcal{N}(0, \sigma_p^2) \end{aligned} \tag{1}$$

where $\hat{I}_d^i$ and $\hat{I}_p^i$ represents the denoising results of data-driven and physics-driven branch from the noisy image I_n^i, respectively. Accordingly, the noise-free image I^i conforms the following distributions,

$$\begin{aligned} p(I^i|\hat{I}_d^i, \sigma_d) &\propto \mathcal{N}(I^i; \hat{I}_d^i, \sigma_d^2) \\ p(I^i|\hat{I}_p^i, \sigma_p) &\propto \mathcal{N}(I^i; \hat{I}_p^i, \sigma_p^2) \end{aligned} \tag{2}$$

Based on the idea of MLE, the reconstruction loss for these two single branches can be written as:

$$\mathcal{L}_{\text{single}} = \sum_{k \in \{d,p\}} \left(\frac{\|I^i - \hat{I}_k^i\|^2}{2\sigma_k^2} + \frac{1}{2} \log \sigma_k^2 \right) \tag{3}$$

Assuming that the estimation error of these two branches are independent (as shown in Figure 1(b)), then we have:

$$p(\epsilon_d, \epsilon_p) = p(\epsilon_d)p(\epsilon_p) = \frac{1}{2\pi\sigma_d\sigma_p} \exp\left(-\frac{\epsilon_d^2}{2\sigma_d^2} - \frac{\epsilon_p^2}{2\sigma_p^2} \right) \tag{4}$$

Take $\epsilon_k = I^i - \hat{I}_k^i$ into Eq.(4), we get:

$$p(I_i|\hat{I}_d^i, \hat{I}_p^i) = \frac{1}{2\pi\sigma_d\sigma_p} \exp\left(-\frac{(I^i - \hat{I}_d^i)^2}{2\sigma_d^2} - \frac{(I^i - \hat{I}_p^i)^2}{2\sigma_p^2} \right) \tag{5}$$

Completing the square transform Eq.(5) it to Gaussian form:

$$p(I^i|\hat{I}_d^i, \hat{I}_p^i) \propto \frac{1}{\sqrt{2\pi}\sigma_{\text{fuse}}} \exp\left(-\frac{(I^i - \hat{I}_{\text{fuse}}^i)^2}{2\sigma_{\text{fuse}}^2} \right) \tag{6}$$

with the predicted mean fusion image $\hat{I}_{\text{fuse}}^i$ expressed as:

$$\hat{I}_{\text{fuse}}^i = \frac{\hat{I}_p^i/\sigma_p^2 + \hat{I}_d^i/\sigma_d^2}{1/\sigma_p^2 + 1/\sigma_d^2} \tag{7}$$

The corresponding estimation uncertainty is:

$$\sigma_{\text{fuse}}^2 = \frac{\sigma_d^2\sigma_p^2}{\sigma_d^2 + \sigma_p^2} \tag{8}$$

The complete mathematical derivation for Eq.(7) and Eq.(8) is provided in the appendix.

Based on the idea of MLE, the loss function for the fused results can be formulated as:

$$\mathcal{L}_{\text{fuse}} = \left(\frac{\|I^i - \hat{I}_{\text{fuse}}^i\|^2}{2\sigma_{\text{fuse}}^2} + \frac{1}{2} \log \sigma_{\text{fuse}}^2 \right) \tag{9}$$

Note that, Eq.(3) and Eq.(9) requires the noise-free image I^i to optimize, however in DW image denoising, which is usually not available. How to use the data itself and the physical principle of dMRI to design the self-supervised learning target and guarantee the dependence of estimation error of two branches are challenging.

Architecture of UP^2-MAE and Optimization

To realize the self-supervised denoising with DW image data and diffusion MRI physics, we proposed a UP^2-MAE network, its overall architecture is illustrated in Figure 2. It integrates two complementary denoising branches: a MAE branch that learns noise-robust representations through self-supervised masked image reconstruction, and a physics-constrained branch that maintains diffusion angular relationships by combining dMRI principles with Noise2Noise methodology. Each branch independently reconstructs a preliminary denoised image while simultaneously deriving an uncertainty map from its own feature via a simple network. These uncertainties are then used as confidence weights to

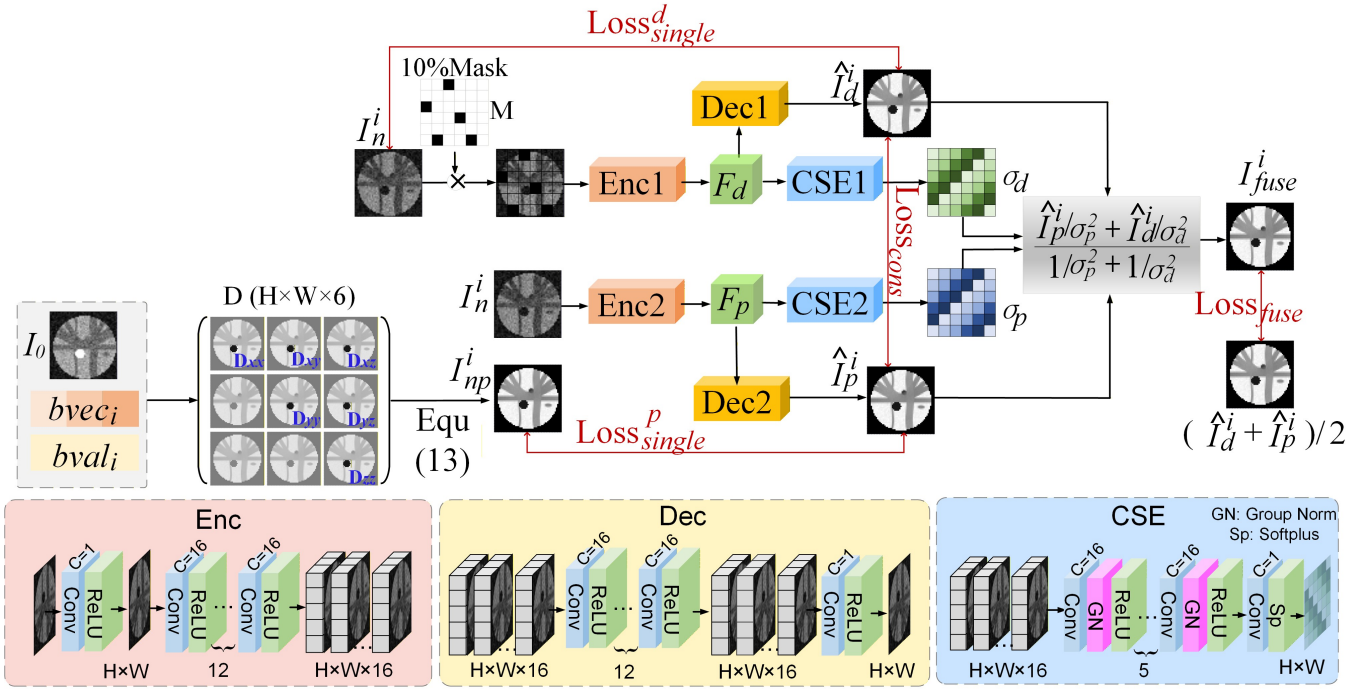


Figure 2: Overview of the proposed framework UP^2-MAE.

fuse the two reconstructed images, yielding the final denoised result.

Given the noisy DW image along the i-th diffusion encoding direction I^i_n, it was randomly masked with a mask M and then input into an encoder of the MAE branch to extract the features F^d,

$$F_d = \mathbf{Enc1}(I^i_n \cdot M), \tag{10}$$

At the same time, I^i_n was also input into an encoder of the physic branch to extract the features F_p,

$$F_p = \mathbf{Enc2}(I^i_n), \tag{11}$$

To fully utilize both branches, F_d and F_p passed then into a decoder and confidence sore estimation (CSE) block to derive the reconstruction results ($\hat{I}^i_d$/ $\hat{I}^i_p$) and uncertainty map (σ_d/σ_p) respectively.

$$\begin{aligned} \hat{I}^i_d &= \mathbf{Dec1}(F_d), \hat{I}^i_p = \mathbf{Dec2}(F_p) \\ \sigma_d &= \mathbf{CSE1}(F_d), \sigma_p = \mathbf{CSE2}(F_p) \end{aligned} \tag{12}$$

The detailed structure of CSE module can be found in the Figure 2. The final denoised DW image is derived with Eq.(7). To constrain the physic branch, the DW image I^i_{np} along the same diffusion encoding direction was fitted with the diffusion tensor model as labels for the physic branch.

$$I^i_{np} = I_0 \exp\left(-b_i \mathbf{g}_i^\top \mathbf{D} \mathbf{g}_i\right), \tag{13}$$

where $\mathbf{g}_i = [g_{ix}, g_{iy}, g_{iz}]^\top$ denotes the i-th diffusion encoding direction, b_i the diffusion weighting factor, I_0 the b0 image, and $\mathbf{D} \in \mathbb{R}^{W\times H\times L\times 3\times 3}$ the diffusion tensor map, with each voxel in $\mathbf{D}$ being a positive symmetric matrix, therefore the diffusion tensor map can be decomposed with six maps with size of $W \times H \times L$,

$$\mathbf{D} = \{D_{xx}, D_{xy}, D_{xz}, D_{yy}, D_{yz}, D_{zz}\} \tag{14}$$

The value of diffusion tensor map $\mathbf{D}$ was estimated from the noisy DW images using least square fitting method. Accordingly, for each single branch, the corresponding reconstruction loss can be formulated as:

$$\begin{aligned} \mathcal{L}^p_{\text{single}} &= \frac{\|I^i_{np} - \hat{I}^i_p\|^2}{2\sigma_p^2} + \frac{1}{2}\log\sigma_p^2 \\ \mathcal{L}^d_{\text{single}} &= \frac{\|(1-M)(I^i_d - \hat{I}^i_n)\|^2}{2\sigma_d^2} + \frac{1}{2}\log\sigma_d^2 \end{aligned} \tag{15}$$

To optimize the final denoised output, we define the fusion loss function under the ideal condition where both branches produce equally accurate denoised images:

$$\mathcal{L}_{\text{fuse}} = \frac{\|I^i_{\text{fuse}} - \frac{\hat{I}^i_p + \hat{I}^i_d}{2}\|^2}{2\sigma^2_{\text{fuse}}} + \frac{1}{2}\log\sigma^2_{\text{fuse}} \tag{16}$$

To enhance inter-branch consistency while preserving uncertainty-guided confidence, we propose an uncertainty-weighted consistency loss that optimally balances agreement between branches with their respective reliability estimates, specifically,

$$\mathcal{L}_{\text{cons}} = \left\| \frac{\hat{I}^i_d}{\sigma_d} - \frac{\hat{I}^i_p}{\sigma_p} \right\|^2 \tag{17}$$

As evidenced by Eq.(17), regions with lower uncertainty yield more reliable reconstructions, confirming the expected inverse relationship between uncertainty and result confidence.

The training objective is finally given by:

$$\mathcal{L}_{\text{total}} = L^{d}_{\text{single}} + L^{p}_{\text{single}} + L_{\text{fuse}} + L_{\text{cons}} \tag{18}$$

Experiments

Datasets and Metrics

We evaluate the proposed method on three different datasets. One was simulation DW image dataset of virtual Phantomαs provided by ISBI challenge (Caruyer et al. 2014), with the simulated noise conforming to zero-mean Gaussian, Rician and non-central Chi distribution (noise level=10%). The second dataset comes from the HCP WU-Minn-On Consortium dataset (Van Essen et al. 2013), which was acquired uniformly from 12 diffusion encoding directions among 90 diffusion encoding directions with a b-value of 3000 s/mm^2. The dimension of the DW image is 145×174×145, and the spatial resolution is 1.25×1.25×1.25 mm^3. The third dataset was cardiac dataset collected from (https://med.stanford.edu/cmrgroup/data/myofiber_data.html) which was imaged at end systole phases with b-value = 350 s/mm^2 along 12 diffusion encoding directions.

The denoising performance was quantitatively assessed through several metrics, including Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), ΔR^2 coefficient measuring the fitting ability with diffusion tensor model, Average Angular Error (AAE) describing the fiber orientation accuracy. The detailed calculation of each metric can be found in the appendix.

Implement Details

To demonstrate the effectiveness of the proposed method, we compared it with several established DW image denoising techniques, including MPPCA, Patch2Self, DDM2, and Replace2Self. Among these, MPPCA is conventional methods implemented using publicly available Matlab-based software, with a search radius set to 2. Patch2Self was executed via the DIPY (Garyfallidis et al. 2014) API, employing a patch size of 1. All other experiments were conducted using Pytorch on an Nvidia Tesla P40 GPU.

Our model training was optimized with the ADAM optimizer, utilizing a learning rate of 1×10^{-4}. The batch size and the number of epochs were configured to 1 and 3000, respectively. In UP^2-MAE, all encoders and decoders contain 12 convolutional layers with ReLU activations, and the number of feature channels is set to 16. The CSE module consists of 5 blocks of convolution, GroupNorm, and ReLU, followed by a single-channel output for structural estimation. The full architectural details and configuration parameters are provided in the appendix. Additionally, our implementation code has been made publicly available in the Appendix to facilitate reproducibility and future research.

Results and Analysis

Comparison with existing methods

Results on Simulation Dataset We evaluated all models on simulated DW images with 10% Gaussian, Rician, and Nc-Chi noise. Figure 3 presents denoising results for Rician noise (Gaussian and Nc-Chi results are in the Appendix). The top three rows of Figure 3 compare DW images processed by various denoising methods through zoomed-in residual maps. While MPPCA and Replace2Self preserve some structural information, they show residual noise artifacts. Patch2Self maintains structural consistency but suffers from incomplete noise suppression, while DDM2 introduces significant intensity deviations from ground truth. Our method produces the cleanest residuals with optimal detail preservation. Quantitative metrics in Table 1 confirm this advantage, achieving superior PSNR (31.48 ± 1.06 dB) and SSIM (0.98 ± 0.00) (Table 1). The ground-truth versus denoised voxel intensity scatterplot (Figure 4) further validates our method's accuracy, showing tighter clustering around the identity line ($y = x$) than competing techniques.

The fiber orientation and tractography results (Figure 3, rows 4-7) demonstrate that our method most closely matches the ground truth, achieving superior angular coherence as evidenced by the lowest AAE (2.25 ± 3.16, Table 1). Comparative analysis shows that MPPCA and DDM2 exhibit noticeable fiber loss and Patch2Self produces disorganized streamlines with residual noise, Replace2Self yields comparable results but suffers from peripheral fiber clustering. Notably, our approach consistently preserves structural integrity while accurately reconstructing fiber bundles.

FA/MD maps and residual errors are shown in the bottom rows of Figure 3. Patch2Self and DDM2 exhibit the worst performance, with severe FA underestimation and large errors. Replace2Self produces smoother maps but systematically underestimates FA. MPPCA approaches ground truth but suffers from localized signal loss. In contrast, our method achieves the most accurate FA/MD estimates (Table 1), better preserving microstructure. The above comprehensive evaluations confirm our method's superior balance between noise suppression and structural preservation across all tested metrics.

Results on HCP Dataset Figure 5 compares method performance on in-vivo HCP data reconstructed from 12 diffusion directions. While DDM2 reduces noise, it oversmooths boundaries, leaving anatomical artifacts in residuals. MPPCA shows milder smoothing, particularly in the corticospinal (CC) tract (red arrow). Patch2Self introduces jagged textures, whereas Replace2Self better preserves white-matter patterns but remains noise. In contrast, our method achieves optimal balance, residuals are weakest and homogeneous, indicating effective noise suppression with anatomical preservation.

These advantages extend to tractography. For CC fiber reconstruction, DDM2 causes fiber discontinuities; Patch2Self shows fiber deletions (yellow arrows); Replace2Self generates aberrant pseudo-fibers. Our method yields coherent bundles with natural orientation. For FA maps, DDM2

	Metrics	Noisy	MPPCA	Patch2Self	DDM2	Replace2Self	Our
DW image	PSNR(dB)	16.26±0.68	29.77±1.44	24.56±0.99	22.50±1.34	29.25±0.93	**31.48±1.06**
	SSIM	0.79±0.02	0.97±0.00	0.95±0.01	0.92±0.01	0.97±0.00	**0.98±0.00**
DTI	FA Residuals	0.21±0.03	0.07±0.02	0.14±0.02	0.17±0.03	0.09±0.00	**0.06±0.00**
	MD Residuals	0.10±0.00	0.06±0.00	0.10±0.00	0.34±0.00	0.05±0.00	**0.03±0.00**
	AAE	4.44±5.92	3.23±16.34	2.31±2.78	5.04±25.33	2.68±2.85	**2.25±3.16**

Table 1: Quantitative comparison of different denoising methods on DW images and DTI-derived metrics under 10% Rician noise. MD residuals are scaled by 10^{-3} for better readability.

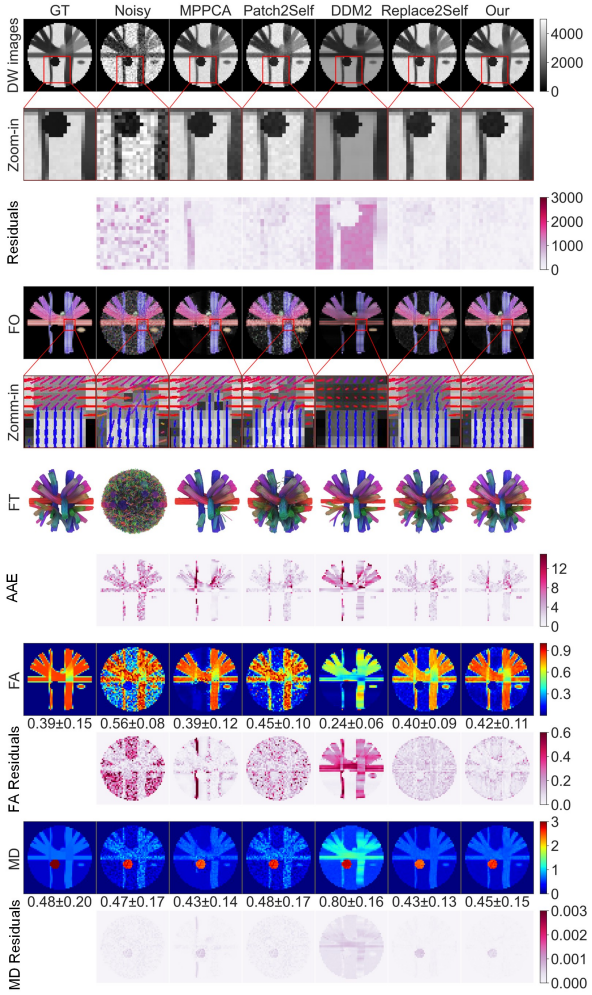


Figure 3: Denoised DW images and diffusion metric maps obtained by different methods on simulation dataset (b=1000 s/mm^2) with Rician noise distributions and a noise level of 10%.

creates striping artifacts and reduces CC contrast; MPPCA blurs fine structures; Patch2Self/Replace2Self underestimate callosal FA values. Our approach maintains clear WM/GM contrast while removing noise. MD maps appear globally similar across methods.

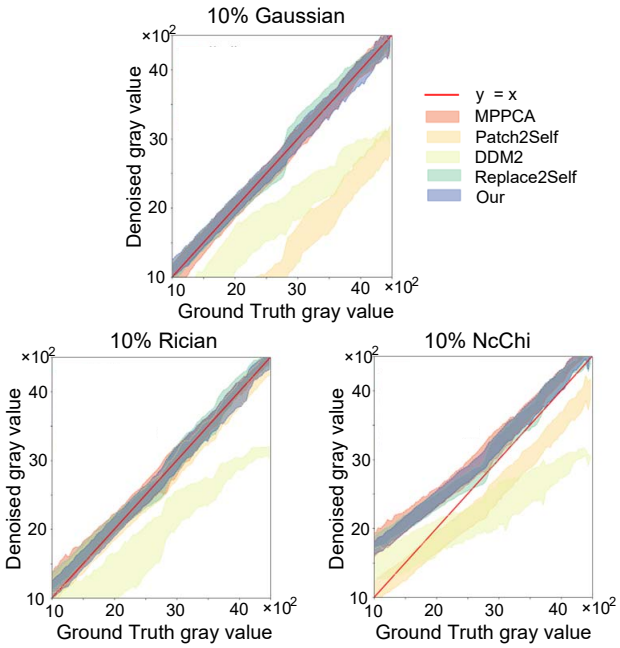


Figure 4: Pixel-wise comparison of denoised intensity values versus ground truth under three noise types: Gaussian, Rician, and NcChi. Each shaded band represents the standard deviation across samples for each method.

Results on Cardiac Dataset For human cardiac imaging where ground truth is unavailable, pseudo-ground truth references were generated by averaging three independent acquisitions. Evaluation was conducted through residual mapping and DTI-derived metrics (Figure 6). Residual analysis revealed distinct performance patterns: Patch2Self and Replace2Self preserved anatomical structures in residuals (indicating biased denoising), while DDM2 partially removed noise but introduced edge artifacts. Both our method and MPPCA yielded relatively homogeneous residuals, demonstrating effective noise suppression without structural distortion. Notably, our method achieved comparable performance using only one single-acquisition data, unlike MPPCA which requires multiple acquisitions.

DTI metric evaluation showed that MPPCA, Patch2Self, and Replace2Self introduced directional discontinuities along the epicardial border due to residual noise. In contrast,

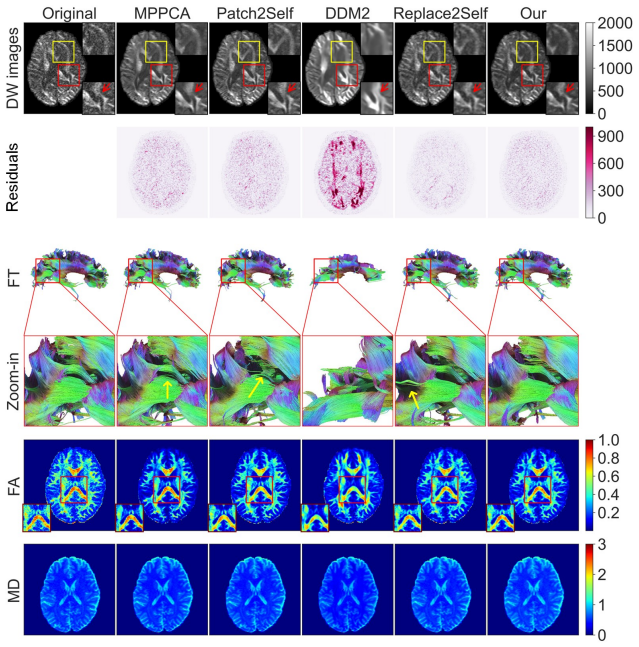


Figure 5: Denoised DW images and diffusion metric maps obtained by different methods on HCP dataset (b=3000 s/mm^2).

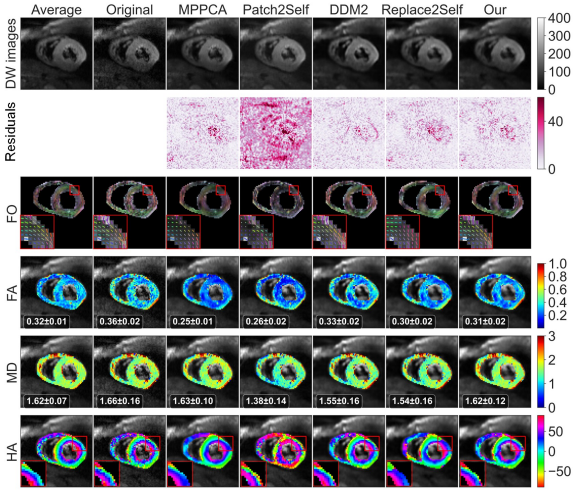


Figure 6: Denoised DW images and Diffusion metric maps obtained by different methods on Cardiac dataset (b=350 s/mm^2) .

our method preserved the smooth transmural helical pattern characteristic of myocardial architecture (as indicated by red rectangle). Quantitative analysis further differentiated the methods: both MPPCA (FA=0.25 ± 0.01) and Patch2Self (FA=0.26 ± 0.02) exhibited oversmoothing relative to the averaged reference (FA=0.32 ± 0.01). While DDM2 and Replace2Self yielded more realistic FA/MD values, their residual structural artifacts undermined reliability. Our approach achieved optimal agreement with established literature values (FA=0.31 ± 0.02, MD=$1.62\pm0.12\times10^{-3}$ mm^2/s) while maintaining coherent helix angle transitions.

Figure 7 quantitatively compare the distribution of ΔR^2 values obtained by different methods on both HCP and heart datasets. We observe that our method achieves the largest ΔR^2 on both HCP and heart datasets, confirming that the denoised signal can be explained more accurately by the diffusion tensor model.

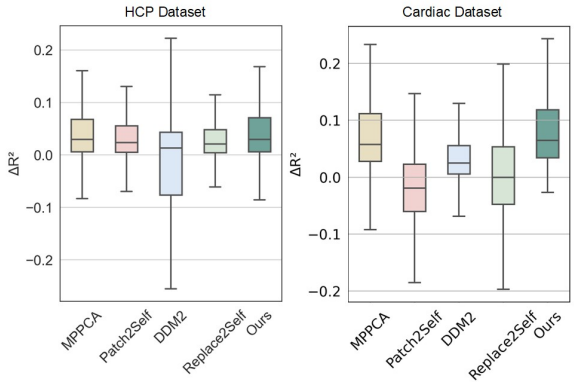


Figure 7: Box-whisker plot of ΔR^2 across 12 directions on HCP dataset and Cardiac dataset.The improvement in ΔR^2 for each box is calculated by subtracting the model's fit score on the noisy data from the R^2 value of each denoised output fit.

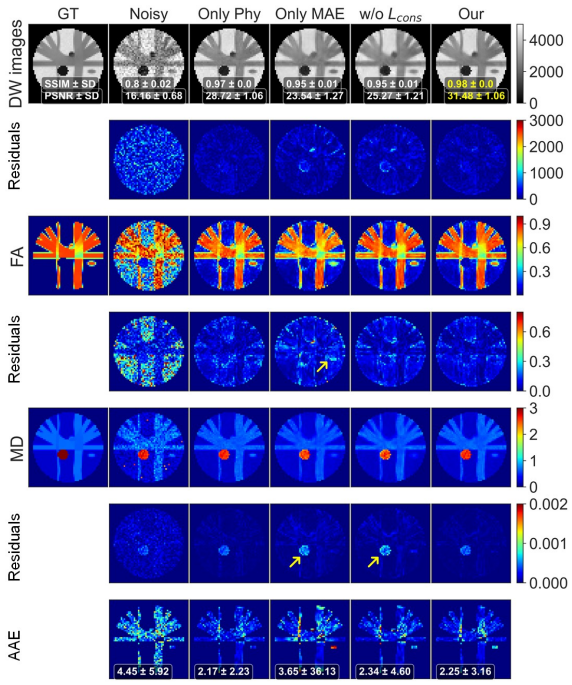
Ablation results and Analysis

We conducted three ablation studies on Phantomαs dataset with 10% Rician noise: (1) Physics branch only (Phy), (2) MAE branch only (MAE), and (3) model without consistency loss constraint (w/o $\mathcal{L}_{\text{cons}}$). Qualitative results with the quantitative metrics are shown in Figure 8(a).

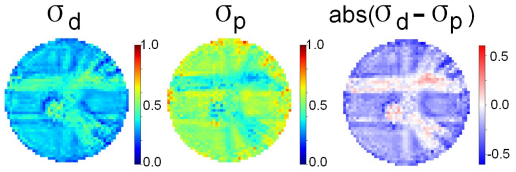
Phy: While effectively removing random noise and recovering near-GT FA values and closest fiber orientations (smallest AAE=2.17 ± 2.23), background noise still remains, revealing that physical model can maintain the angular relationship between DW images along different directions, but it cannot deal with the noise in isotropic regions (i.e. background, with high uncertainty (σ_p in Figure 8(b))).

MAE: The MAE branch effectively denoised homogeneous regions but underperformed in anisotropic fiber areas (opposite to the physics branch) resulting in FA underestimation, elevated MD residuals, and higher AAE. This demonstrates that while data-driven methods learn robust uniform-region features, they cannot preserve directional relationships in heterogeneous fiber structures (with high uncertainty in the fiber regions, σ_d in Figure 8(b)).

w/o $\mathcal{L}_{\text{cons}}$: The absence of consistency loss $\mathcal{L}_{\text{cons}}$ caused residual increases in both isotropic and anisotropic regions, resulting in degraded PSNR/SSIM and underestimated FA/MD values. These findings confirm the efficacy of $\mathcal{L}_{\text{cons}}$ loss in enabling cross-component guidance for improved denoising.



(a) Results for ablation of different components.



(b) Uncertainty map of two branches.

Figure 8: Ablations for different components. Our complete model (rightmost) demonstrates optimal noise suppression while preserving structural fidelity.

Conlusion

In this work, we presented UP^2-MAE, a novel self-supervised DWI denoising framework that synergistically combines physics-based and data-driven (MAE) approaches to overcome limitations in diffusion MRI denoising. By fusing noise-robust feature learning (MAE) and directional noise modeling (physics-based N2N) via an uncertainty-propelled fusion strategy, our method achieves robust noise suppression while preserving angular relationship of DW images along different directions. The introduction of uncertainty-guided consistency loss further enhances denoising performance by ensuring coherent predictions across complementary branches. Extensive validation on both simulated and real DWI data demonstrates that UP^2-MAE outperforms existing methods.

Limitations: While UP^2-MAE achieves effective denoising, its reliance on DTI's Gaussian diffusion assumption limits performance in regions with complex fiber configurations or non-Gaussian diffusion properties. Future improvements could incorporate more advanced microstructure models.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No.62571150), and the Guizhou Provincial Basic Research Program (QianKeHe ZK [2023] 058).

References

Basser, P. J.; Mattiello, J.; and LeBihan, D. 1994. MR diffusion tensor spectroscopy and imaging. *Biophysical journal*, 66(1): 259–267.

Batson, J.; and Royer, L. 2019. Noise2self: Blind denoising by self-supervision. In *International conference on machine learning*, 524–533. PMLR.

Caruyer, E.; Daducci, A.; Descoteaux, M.; Houde, J.-C.; Thiran, J.-P.; and Verma, R. 2014. Phantomas: a flexible software library to simulate diffusion MR phantoms. In *Ismrm*.

Chen, G.; Wu, Y.; Shen, D.; and Yap, P.-T. 2016. XQ-NLM: denoising diffusion MRI data via x-q space non-local patch matching. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 587–595. Springer.

Coupé, P.; Manjón, J. V.; Robles, M.; and Collins, D. L. 2012. Adaptive multiresolution non-local means filter for three-dimensional magnetic resonance image denoising. *IET image Processing*, 6(5): 558–568.

Fadnavis, S.; Batson, J.; and Garyfallidis, E. 2020. Patch2Self: Denoising Diffusion MRI with Self-Supervised Learning. *Advances in Neural Information Processing Systems*, 33: 16293–16303.

Fadnavis, S.; Chowdhury, A.; Batson, J.; Drineas, P.; and Garyfallidis, E. 2024. Patch2self2: Self-supervised denoising on coresets via matrix sketching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 27641–27651.

Fan, L.; Cui, J.; Li, H.; Yan, X.; Liu, H.; and Zhang, C. 2024. Complementary blind-spot network for self-supervised real image denoising. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(10): 10107–10120.

Garyfallidis, E.; Brett, M.; Amirbekian, B.; Rokem, A.; Van Der Walt, S.; Descoteaux, M.; Nimmo-Smith, I.; and Contributors, D. 2014. Dipy, a library for the analysis of diffusion MRI data. *Frontiers in neuroinformatics*, 8: 8.

Henriques, R. N.; Ianuş, A.; Novello, L.; Jovicich, J.; Jespersen, S. N.; and Shemesh, N. 2023. Efficient PCA denoising of spatially correlated redundant MRI data. *Imaging Neuroscience*, 1: 1–26.

Huang, T.; Li, S.; Jia, X.; Lu, H.; and Liu, J. 2021. Neighbor2neighbor: Self-supervised denoising from single noisy images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 14781–14790.

Jang, H.; Park, J.; Jung, D.; Lew, J.; Bae, H.; and Yoon, S. 2023. Puca: patch-unshuffle and channel attention for enhanced self-supervised image denoising. *Advances in Neural Information Processing Systems*, 36: 19217–19229.

Jiang, H.; Zhang, S.; Wen, X.; Cui, H.; Lu, J.; Rekik, I.; Ma, J.; and Chen, G. 2024. Self-supervised denoising of diffusion mri data via spatio-angular noise2noise. In *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, 1–5. IEEE.

Krull, A.; Buchholz, T.-O.; and Jug, F. 2019. Noise2void-learning denoising from single noisy images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2129–2137.

Le Bihan, D. 2003. Looking into the functional architecture of the brain with diffusion MRI. *Nature reviews neuroscience*, 4(6): 469–480.

Le Bihan, D.; Poupon, C.; Amadon, A.; and Lethimonnier, F. 2006. Artifacts and pitfalls in diffusion MRI. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 24(3): 478–488.

Lee, W.; Son, S.; and Lee, K. M. 2022. Ap-bsn: Self-supervised denoising for real-world images via asymmetric pd and blind-spot network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 17725–17734.

Lehtinen, J.; Munkberg, J.; Hasselgren, J.; Laine, S.; Karras, T.; Aittala, M.; and Aila, T. 2018. Noise2Noise: Learning image restoration without clean data. *arXiv preprint arXiv:1803.04189*.

Linhai, W.; Wang, L.; Deng, Z.; Zhu, Y.; and Wei, H. 2025. Replace2Self: Self-Supervised Denoising based on Voxel Replacing and Image Mixing for Diffusion MRI. *IEEE Transactions on Medical Imaging*.

Manjón, J. V.; Coupé, P.; Concha, L.; Buades, A.; Collins, D. L.; and Robles, M. 2013. Diffusion weighted image denoising using overcomplete local PCA. *PloS one*, 8(9): e73021.

Mansour, Y.; and Heckel, R. 2023. Zero-shot noise2noise: Efficient image denoising without any data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 14018–14027.

Mosso, J.; Simicic, D.; Şimşek, K.; Kreis, R.; Cudalbu, C.; and Jelescu, I. O. 2022. MP-PCA denoising for diffusion MRS data: promises and pitfalls. *NeuroImage*, 263: 119634.

Olesen, J. L.; Ianus, A.; Østergaard, L.; Shemesh, N.; and Jespersen, S. N. 2023. Tensor denoising of multidimensional MRI data. *Magnetic resonance in medicine*, 89(3): 1160–1172.

Quan, Y.; Chen, M.; Pang, T.; and Ji, H. 2020. Self2self with dropout: Learning self-supervised denoising from single image. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 1890–1898.

Ramos-Llordén, G.; Vegas-Sánchez-Ferrero, G.; Liao, C.; Westin, C.-F.; Setsompop, K.; and Rathi, Y. 2021. SNR-enhanced diffusion MRI with structure-preserving low-rank denoising in reproducing kernel Hilbert spaces. *Magnetic resonance in medicine*, 86(3): 1614–1632.

St-Jean, S.; Coupé, P.; and Descoteaux, M. 2016. Non Local Spatial and Angular Matching: Enabling higher spatial resolution diffusion MRI datasets through adaptive denoising. *Medical image analysis*, 32: 115–130.

Tian, Q.; Bilgic, B.; Fan, Q.; Liao, C.; Ngamsombat, C.; Hu, Y.; Witzel, T.; Setsompop, K.; Polimeni, J. R.; and Huang, S. Y. 2020. DeepDTI: High-fidelity six-direction diffusion tensor imaging using deep learning. *NeuroImage*, 219: 117017.

Tian, Q.; Li, Z.; Fan, Q.; Polimeni, J. R.; Bilgic, B.; Salat, D. H.; and Huang, S. Y. 2022a. SDnDTI: Self-supervised deep learning-based denoising for diffusion tensor MRI. *Neuroimage*, 253: 119033.

Tian, X.; Wu, J.; Lao, G.; Du, C.; Jiang, C.; Li, Y.; Zhang, J. L.; Wei, H.; and Zhang, Y. 2025. Self-supervised denoising for high-dimensional magnetic resonance image. *Biomedical Signal Processing and Control*, 104: 107451.

Tian, X.; Wu, Q.; Wei, H.; and Zhang, Y. 2022b. Noise2sr: Learning to denoise from super-resolved single noisy fluorescence image. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 334–343. Springer.

Van Essen, D. C.; Smith, S. M.; Barch, D. M.; Behrens, T. E.; Yacoub, E.; Ugurbil, K.; Consortium, W.-M. H.; et al. 2013. The WU-Minn human connectome project: an overview. *Neuroimage*, 80: 62–79.

Vasylechko, S.; Afacan, O.; and Kurugol, S. 2023. Self Supervised Denoising Diffusion Probabilistic Models for Abdominal DW-MRI. In *International Workshop on Computational Diffusion MRI*, 80–91. Springer.

Wang, H.; Zheng, R.; Dai, F.; Wang, Q.; and Wang, C. 2019. High-field mr diffusion-weighted image denoising using a joint denoising convolutional neural network. *Journal of Magnetic Resonance Imaging*, 50(6): 1937–1947.

Wang, Z.; Liu, J.; Li, G.; and Han, H. 2022. Blind2unblind: Self-supervised image denoising with visible blind spots. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2027–2036.

Wiest-Daesslé, N.; Prima, S.; Coupé, P.; Morrissey, S. P.; and Barillot, C. 2007. Non-local means variants for denoising of diffusion-weighted and diffusion tensor MRI. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 344–351. Springer.

Xiang, T.; Yurt, M.; Syed, A. B.; Setsompop, K.; and Chaudhari, A. 2023. DDM2: Self-supervised diffusion MRI denoising with generative diffusion models. *arXiv preprint arXiv:2302.03018*.

Yuan, N.; Wang, L.; Ye, C.; Deng, Z.; Zhang, J.; and Zhu, Y. 2023. Self-supervised structural similarity-based convolutional neural network for cardiac diffusion tensor image denoising. *Medical Physics*, 50(10): 6137–6150.

Zhang, K.; Zuo, W.; Chen, Y.; Meng, D.; and Zhang, L. 2017. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE transactions on image processing*, 26(7): 3142–3155.