Self-Supervised Diffusion MRI Denoising via Iterative and Stable Refinement

Chenxu Wu [1,2] Qingpeng Kong [1,2] Zihang Jiang [1,2,†] S. Kevin Zhou [1,2,3,4,†]

[1]School of Biomedical Engineering, Division of Life Sciences and Medicine, USTC [2]MIRACLE Center, Suzhou Institute for Advance Research, USTC

[3]State Key Laboratory of Precision and Intelligent Chemistry, USTC [4]Key Laboratory of Intelligent Information Processing of CAS, Institute of Computing Technology, CAS [†]Corresponding authors

Background

MRI, including diffusion MRI (dMRI), serves as a "microscope" for anatomical structures.

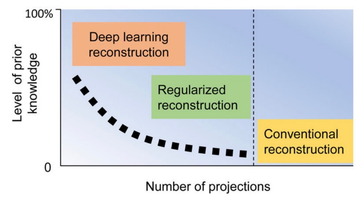

Figure 1. Different image-reconstruction schemes in the context of prior knowledge and projection sampling [1].

dMRI routinely mitigates the influence of low SNR scans by compromising temporal or spatial resolution. However, these fail to meet clinical demands for both efficiency and precision.

Noise2Noise [2] performs image restoration solely by observing corrupted measurements:

$$\arg\min_{\theta} \left\{ \mathbb{E} \left\| f_\theta\left(x'\right) - x \right\|^2 \right\} \approx \arg\min_{\theta} \left\{ \mathbb{E} \left\| f_\theta\left(x'\right) - y \right\|^2 + \mathbb{E} \|x - y\|^2 \right\}, \tag{1}$$

DDPM [3, 4]. In the forward process, noise with specified levels is introduced to the data,

$$q(x_t|x_0) := \mathcal{N}\left(x_t; \sqrt{\bar{\alpha}_t}x_0, (1-\bar{\alpha}_t)\mathbf{I}\right), \tag{2}$$

where $\bar{\alpha}_t := \Pi_{s=1}^{t}\alpha_s$, $\alpha_t := 1 - \beta_t$ and β_t is a variance of noise level. For the reverse process,

$$q(x_{t-1}|x_t, x_0) := \mathcal{N}\left(x_{t-1}; \tilde{\boldsymbol{\mu}}_t(x_t, x_0), \sigma_t^2\mathbf{I}\right), \tag{3}$$

where $\tilde{\boldsymbol{\mu}}_t(x_t, x_0) := \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}x_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}x_t$ and $\sigma_t^2 := \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t$.

Key Features in our methods

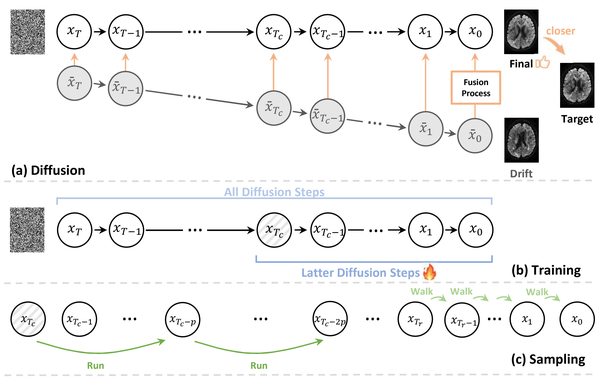

Figure 2. **(a)** Fusion process aligns $\{\bar{x}_t\}_1^T$ to $\{x_t\}_1^T$ and avoids drift ("Drift" means drifted results, "Final" means the denoised version of "Target"); **(b)** Training the latter diffusion steps imposes restrictions on the diversity of diffusion models and decreases uncertainty; **(c)** *Run-Walk* accelerated sampling accelerates the entire sampling process.

Methods

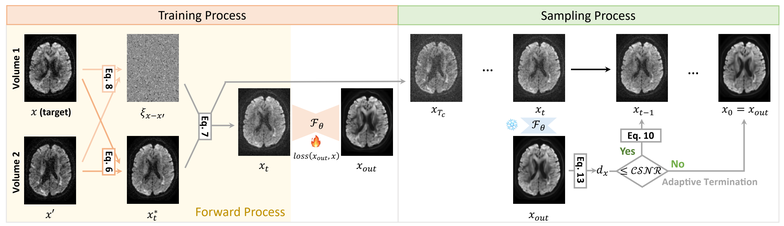

Figure 3. Overview of our single-stage **Di-Fusion (See following for details)**. The training process does not involve any extra model training apart from $\mathcal{F}_\theta$, and the sampling process offers adaptive and controllable results.

Input 4D data: $X \in \mathbb{R}^{w\times h\times d\times l}$, $x = X_{*,*,i,j}, x' = X_{*,*,i,j-1}$

Eq.6: $x_t^* = \lambda_1^t x + \lambda_2^t x'$ (**Fusion process** interpolates between two endpoints)

Eq.8: $\xi_{x-x'} = mess\left((x - x') - \mu_{x-x'}\right)$ (**Di- process** characterizes real-world noise)

Eq.7: $x_t = \sqrt{\bar{\alpha}_t}x_t^* + \sqrt{1-\bar{\alpha}_t}\xi_{x-x'}$ (Diffusion forward process)

Training target: $L_{\text{simple}}(\theta) := \mathbb{E}_{t,x_t^*,\xi_{x-x'}}\left[\left\|x - \mathcal{F}_\theta(\sqrt{\bar{\alpha}_t}x_t^* + \sqrt{1-\bar{\alpha}_t}\xi_{x-x'}, t)\right\|^2\right]$

Latter diffusion steps: $x_{T_c} \to x_0, T_c \leq T$.

Eq.13: $d_x = \|x - x_{out}\|^2 \times b_x$, here b_x accounts for the ratio of brain tissue to the entire image.

Eq.10: Diffusion sampling process. Our sampling schedule in Figure 2 (c).

Experiments

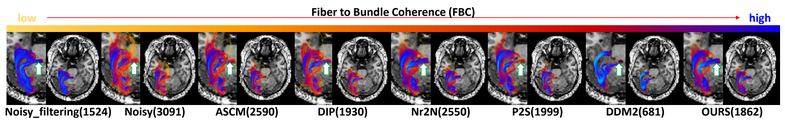

Figure 4. Density map of FBC projected on the streamlines of the OR bundles. The numbers in parentheses represent the number of streamlines. **OURS generates the minimal number of streamlines while maintaining high FBCs (consider "Noisy_filtering" as references for high FBCs)**.

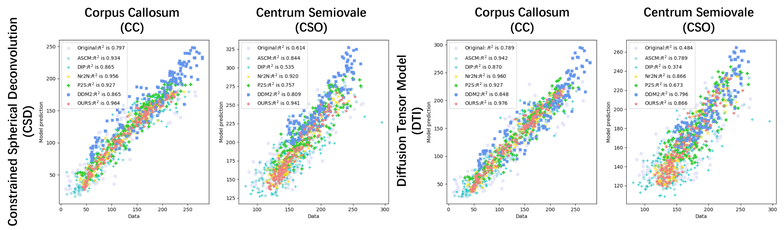

Figure 5. Scatter plots of the microstructure model predictions against input data. The top-left of each plot shows the quantitative R^2 metric computed from each model fit on the corresponding data. OURS data points are more concentrated (higher R^2), which means **OURS aids in the characterization of the microstructure**.

Metric	Noisy	ASCM	MPPCA	Nr2N	P2S	DDM2	OURS
PSNR (FA)	23.47	22.63	26.37	29.47	24.18	26.77	**30.79**
PSNR (MD)	31.79	21.81	35.28	38.42	32.16	37.53	**40.26**
PSNR (RD)	33.56	24.70	36.35	38.48	30.62	37.39	**40.35**
PSNR (AD)	26.39	20.89	29.63	34.87	33.27	33.21	**35.30**
SSIM (FA)	87.60	88.97	90.48	93.54	90.90	90.41	**94.50**
SSIM (MD)	96.35	81.88	97.51	98.51	97.08	98.72	**99.31**
SSIM (RD)	96.37	89.21	97.80	98.65	96.32	98.48	**99.23**
SSIM (AD)	92.47	81.02	95.34	97.12	96.77	96.10	**97.63**

Table 1. Quantitative results for DTI diffusion signal estimates. PSNR (dB) and SSIM (%) are reported. FA (fractional anisotropy), MD (mean diffusivity), RD (radial diffusivity), and AD (axial diffusivity). Best results are in bold. **OURS provides better evidence for the diffusion signal estimates**.

Method	Simulation 1		Simulation 2		Simulation 3		Simulation 4		Simulation 5	
	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR
Noisy	11.38	13.72	20.65	16.09	37.41	20.38	52.02	23.76	64.62	26.63
P2S	24.53	11.20	43.94	17.63	65.34	24.91	78.61	**30.26**	86.13	**33.87**
Nr2N	23.72$_{0.93}$	16.78$_{0.22}$	50.83$_{2.27}$	22.13$_{0.94}$	71.77$_{6.94}$	26.80$_{1.69}$	70.32$_{7.99}$	26.40$_{3.17}$	87.31$_{4.70}$	32.82$_{0.72}$
DDM2	26.02$_{2.63}$	17.92$_{1.06}$	49.57$_{15.6}$	21.76$_{1.64}$	59.64$_{8.47}$	23.25$_{4.05}$	77.96$_{2.16}$	29.14$_{1.01}$	81.43$_{2.23}$	31.35$_{2.23}$
OURS	**39.05**$_{1.42}$	**19.91**$_{0.32}$	**62.02**$_{1.11}$	**23.60**$_{0.05}$	**77.36**$_{0.61}$	**26.96**$_{0.37}$	**83.43**$_{1.39}$	28.69$_{0.67}$	**89.52**$_{0.18}$	30.63$_{0.27}$

Table 2. Quantitative results on simulated data. PSNR (dB) and SSIM (%) are reported. Numbers are presented as mean value with standard deviation. Best results are in bold. **OURS holds a tremendous potential for generalization and applicability for its stable performance and better performance under high noise intensity**

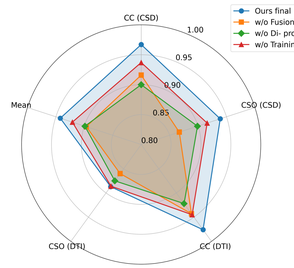

Figure 6. Ablation of our methods.

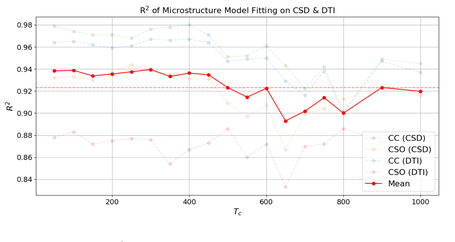

Figure 7. T_c vs. R^2. When $T_c < 500$, the performance is consistent.

Limitations: **(i)** Long inference duration. **(ii)** Possible hallucinations. **(iii)** Noise-artifact conflation [5] and additive noise assumptions [6]. While our assumption enables tractable solutions, it may limit effectiveness for signal-dependent noise (better characterized as artifacts).

References

[1] L. Shen, W. Zhao, and L. Xing, "Patient-specific reconstruction of volumetric computed tomography images from a single projection view via deep learning," *Nature biomedical engineering*, vol. 3, no. 11, pp. 880–888, 2019.

[2] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, "Noise2noise: Learning image restoration without clean data," *arXiv preprint arXiv:1803.04189*, 2018.

[3] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in neural information processing systems*, vol. 33, pp. 6840–6851, 2020.

[4] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, and S. Ganguli, "Deep unsupervised learning using nonequilibrium thermodynamics," in *International conference on machine learning*, pp. 2256–2265, PMLR, 2015.

[5] J. Veraart, D. S. Novikov, D. Christiaens, B. Ades-Aron, J. Sijbers, and E. Fieremans, "Denoising of diffusion mri using random matrix theory," *Neuroimage*, vol. 142, pp. 394–406, 2016.

[6] L. Cordero-Grande, D. Christiaens, J. Hutter, A. N. Price, and J. V. Hajnal, "Complex diffusion-weighted image estimation via matrix recovery under general noise models," *Neuroimage*, vol. 200, pp. 391–404, 2019.

[7] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.