



Research Article

High-Resolution Diffusion-Weighted Imaging with Self-Gated Self-Supervised Unrolled Reconstruction

Submission ID d9d3bd30-9320-4191-a7db-f45df8965854

Submission Version Initial Submission

PDF Generation 09 Sep 2025 08:57:36 EST by Atypon ReX

Authors

Zhengguo Tan
Corresponding Author
Submitting Author

Affiliations

- Michigan Institute for Imaging Technology and Translation (MIITT), Department of Radiology, University of Michigan, Ann Arbor, Michigan, USA

Patrick A Liebig

Affiliations

- Ultra-High Field Predevelopment Team, Siemens Healthcare GmbH, Erlangen, Bavaria, Germany

Annika Hofmann

Affiliations

- Artificial Intelligence in Biomedical Engineering, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Bavaria, Germany

Frederik B Laun

Affiliations

- Institute of Radiology, University Hospital Erlangen, Erlangen, Bavaria, Germany

Florian Knoll

Affiliations

- Artificial Intelligence in Biomedical Engineering, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Bavaria, Germany

Additional Information

Research Focus

Neurological
Neurological / Brain
Brain

Research Type

Technical Research / Image processing/Image analysis / Machine Learning/Deep Learning
Machine Learning/Deep Learning

Technique Development / Diffusion
Diffusion

Research Topics

Imaging Methodology

Files for peer review

All files submitted by the author for peer review are listed below. Files that could not be converted to PDF are indicated; reviewers are able to access them online.

Name	Type of File	Size	Page
MRM-AMA.pdf	Main Document - LaTeX PDF	6.0 MB	Page 4
responses.pdf	Additional File for Review but Not for Publication	4.5 MB	Page 17

Received xx xxx 2025; | Revised xx xxx 2025; | Accepted xx xxx 2025

DOI: xxxx/xxxx

RESEARCH ARTICLE

High-Resolution Diffusion-Weighted Imaging with Self-Gated Self-Supervised Unrolled Reconstruction

Zhengguo Tan¹ | Patrick A Liebig² | Annika Hofmann³ | Frederik B Laun⁴ | Florian Knoll³

¹Michigan Institute for Imaging Technology and Translation (MIITT), Department of Radiology, University of Michigan, Ann Arbor, Michigan, USA

²Ultra-High Field Predevelopment Team, Siemens Healthcare GmbH, Erlangen, Bavaria, Germany

³Artificial Intelligence in Biomedical Engineering, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Bavaria, Germany

⁴Institute of Radiology, University Hospital Erlangen, Erlangen, Bavaria, Germany

Correspondence

Zhengguo Tan, Room 3111, Medical Science Unit I, 1301 Catherine St, Ann Arbor, MI 48109. Email: zgtan@med.umich.edu

Funding Information

German Research Foundation (DFG); projects 513220538, 512819079, project 50088779 in the Research Unit RU5534 for MR biosignatures at UHF. National Institutes of Health (NIH); grants R01EB024532 and P41EB017183.

Abstract

Purpose: High-resolution diffusion-weighted imaging (DWI) is clinically demanding. The purpose of this work is to develop an efficient self-supervised algorithm unrolling technique for submillimeter-resolution DWI.

Methods: We developed submillimeter DWI acquisition utilizing multi-band multi-shot EPI with diffusion shift encoding. We unrolled the alternating direction method of multipliers (ADMM) to perform scan-specific selfgated self-supervised DeepDWI learning for multi-shot echo planar imaging with difusion shift encoding on a clinical 7T scanner.

Results: We demonstrate that (1) ADMM unrolling is generalizable across slices, (2) ADMM unrolling outperforms multiplexed sensitivity-encoding (MUSE) and compressed sensing with locally-low rank (LLR) regularization in terms of image sharpness, tissue continuity, and motion robustness, (3) ADMM unrolling enables clinically feasible inference time.

Conclusion: Our proposed ADMM unrolling enables whole brain DWI of 21 diffusion volumes at 0.7 mm isotropic resolution and 10 minutes scan, and shows higher signal-to-noise ratio (SNR), clearer tissue delineation, and improved motion robustness, which makes it plausible for clinical translation.

KEYWORDS:

diffusion weighted imaging, submillimeter resolution, image reconstruction, machine learning, self-supervised learning, algorithm unrolling

WORD COUNT: 3828

1 | INTRODUCTION

Diffusion-weighted imaging (DWI)¹ has been an important imaging modality in neuro-scientific research and

clinical diagnosis and staging of tumors. However, clinical DWI, based on single-shot echo planar imaging (EPI)², poses challenges in the pursuit of high spatial, temporal, and angular resolution. Until now, the search for precise neuro imaging has fostered significant advances in DWI, including multi-shot EPI (interleaved^{3,4,5} and readout-segmented^{6,7}), field inhomogeneity and eddy

Part of this work has been presented at the ISMRM 2025, Honolulu, USA.

current correction⁸, simultaneous multi-slice⁹, reconstruction techniques such as parallel imaging^{10,11,12} and compressed sensing^{13,14,15}, as well as diffusion-weighted image denoising^{16,17}.

To achieve submillimeter isotropic resolution DWI, advanced k -space encoding strategies have been proposed. For example, Setsompop et al.¹⁸ developed generalized slice dithered enhanced resolution (gSlider) to boost SNR per slice. gSlider excites one thick slab multiple times with complementary slice encoding schemes (e.g., Hadamard encoding). The thin slices are then reconstructed by solving a linear least square problem given the complementary slab signal. However, it has been reported that gSlider has stricter requirements on B_0 and B_1 field homogeneities and shows residual slab boundary artifacts (appearing as striping artifacts)¹⁹. Another advanced k -space encoding strategy, rotating-view echo planar time-resolved imaging (Romer-EPTI)²⁰, has recently been developed. Romer-EPTI acquires thick-slice volumes with different slice orientations. In addition, a rigid-motion transformation matrix is extracted from the multi-shot data and incorporated into the super-resolution linear least square problem to disentangle motion-free thin-slice images. Together with low-rank subspace modeling and reconstruction^{21,22,23}, Romer-EPI achieves submillimeter TE-specific distortion-free DWI. However, Romer-EPTI requires relatively long scan time.

Compressed sensing image reconstruction techniques have also been proposed to push the boundary of DWI. Mani et al.¹⁴ developed MUSSELS, in which structural low rankness is enforced in multi-shot k -space. After reconstruction, one diffusion-weighted image is calculated via root sum of square of all shot images. MUSSELS bypasses the shot-to-shot phase variation correction. Hu et al.¹⁵ developed SPA-LLR, which employs locally-low rank (LLR) regularization in joint k - q -space reconstruction. Further, Tan et al.²⁴ extended LLR to accelerated multi-band multi-shot multi-shell reconstruction.

Recently, emerging deep learning techniques show great promise in deep DWI reconstruction. Cho et al.²⁵ proposed to learn a k -space regularization function with ResNet²⁶ and zero-shot self-supervised learning^{27,28}. This approach learns an unrolled algorithm utilizing only the acquired data itself, and thus requires no extra training data and is scan specific. Similar to MUSSELS, this approach addresses the multi-shot EPI reconstruction problem for a single diffusion encoding, and does not perform joint reconstruction that explores q -space redundancy. Mani et al.²⁹ proposed to learn a denoising autoencoder (DAE) model³⁰ from a physics-informed simulated dictionary. This learned DAE model is subsequently utilized as a q -space regularizer, in combination

with a total-variation spatial regularizer, in the joint k - q -space reconstruction, as solved by the alternating direction method of multipliers (ADMM)³¹.

To address the challenges of achieving submillimeter isotropic DWI at clinically feasible scan times, we propose a novel solution that leverages an ADMM unrolling method with self-supervised learning for multi-band, multi-shot DWI with diffusion shift encoding. Unlike previous methods that require extensive phase variation correction or long scan times (e.g., gSlider's strict field homogeneity requirements or Romer-EPTI's lengthy acquisition times), our approach incorporates complementary k - q -space sampling and jointly reconstructs multiple slices and all diffusion-weighted images. Moreover, by training the ADMM unrolling model using a single multi-band slice, we enable self-gated joint reconstruction, significantly reduce training time, and avoid the need for large-scale dictionaries or extra training data, as seen in methods like supervised learning approaches^{32,33,34}. Our method not only achieves high-resolution DWI at 0.7 mm isotropic resolution with 21 diffusion-encoding directions but also does so in under 10 minutes of scan time and approximately 1 minute of reconstruction time per slice. This provides a clinically viable, efficient solution to the submillimeter resolution DWI challenge.

2 | METHODS

2.1 | In Vivo Acquisition and Reconstruction

Table 1 lists two acquisition protocols implemented on a clinical 7 T MR system (MAGNETOM Terra, Siemens Healthineers, Erlangen, Germany) equipped with a 32-channel head coil (Nova Medical, Wilmington, MA, USA) and the XR-gradient system (maximum gradient strength 80 mT/m and a peak slew rate 200 T/m/s). Protocol #1 with 1 mm isotropic resolution serves as the reference with in-plane fully sampling and the multi-band factor 3. This reference 4-shot data is retrospectively undersampled to only 2 shots (i.e., 2-fold in-plane undersampling) and then trained and tested with the proposed self-supervised learning. Protocols #2 and #3 realize high resolution mesoscale DWI with isotropic resolution 0.7 mm. Two-fold acceleration is used in both the in-plane and in-slice directions. Every diffusion encoding is acquired by three shots in an interleaved manner and is shifted with respect to its former, resulting in a 6×2 -fold acceleration per shot. It is noteworthy that the total scan time can be reduced to about 10 minutes (Protocol #3) when switching off navigator acquisition.

TABLE 1 Acquisition protocols

Protocol ¹	#1 (1.0 mm)	#2 (0.7 mm NAV)	#3 (0.7 mm)
FOV (mm ²)		200	
Matrix size	200 × 200 × 114	286 × 286 × 176	
Voxel (mm ³)	1.0 × 1.0 × 1.0	0.7 × 0.7 × 0.7	
Shots	4	3	
Acceleration	1 × 3	2 × 2	
Partial Fourier	5/8	5/8	
Bandwidth (Hz/Pixel)	1086	972	
ESP (ms)	1.04	1.17	
Navigator	No	Yes	No
TE (ms)	66	58/98.3	58
TR (ms)	5400	15000	8900
Acquisition (min)	7 : 52	16 : 27	9 : 57

¹ All protocols employed the MDDW diffusion acquisition mode with monopolar diffusion encoding gradients, 1 b_0 volume and 20 diffusion-weighted volumes with the b -value of 1000 s/mm².

Three young healthy volunteers with written informed consent approved by the local ethics committee participated in this study. All reconstructions in this work were done on a single A100 SXM4/NVLink GPU with 80 GB memory (NVIDIA, Santa Clara, CA, USA).

2.2 | Multi-Band Multi-Shot DWI with Diffusion-Shift Encoding

Our previous work²⁴ demonstrated the joint k - q -slice reconstruction for multi-band multi-shot navigator-based interleaved EPI (NAViEPI) DWI acquisition with diffusion shift encoding. As shown in Figure 1, the starting line k_y for a diffusion encoding is shifted with respect to its adjacent line to create a complementary k - q -slice sampling pattern. In the joint reconstruction, the forward model maps the multi-slice multi-diffusion-weighted images (\mathbf{x}) to their corresponding k -space,

$$\mathcal{A}(\mathbf{x}) = \mathbf{P}\Sigma\Theta\mathbf{F}\mathbf{S}\Phi\mathbf{x} \quad (1)$$

Here, the images \mathbf{x} are point-wise multiplied with the precomputed shot-to-shot phase variation maps (Φ) and coil sensitivity maps (\mathbf{S}). The output images are then converted to k -space via the two-dimensional fast Fourier transform (\mathbf{F}), multiplied point-wise with the multi-band phases (Θ), summed along the slice dimension (Σ), and then multiplied by the k -space undersampling mask (\mathbf{P}).

In Equation (1), one challenge is to accurately estimate the shot-to-shot phase variation. Multiplexed sensitivity-encoding (MUSE)-type reconstruction techniques^{4,35,5,36} achieved the self-gating strategy, where the k -space data of each shot were used to reconstruct its corresponding shot image followed by a phase smoothing approach (i.e., the phase variation operator Φ).

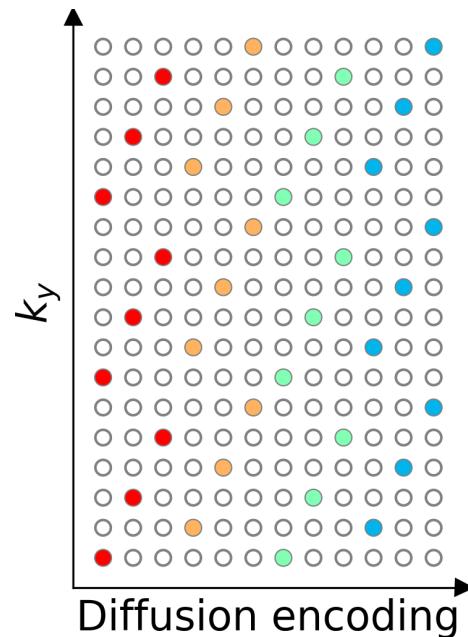


FIGURE 1 Three-shot DWI with diffusion shift encoding. This work employs three-shot per diffusion encoding and each shot has an in-plane undersampling factor of 6. Every three columns assemble one diffusion encoding and thus are colored the same. The starting k_y line is shifted between adjacent diffusion encoding to create complementary k - q -space sampling.

Self-gated shot phase estimation does not require the acquisition of phase navigator data, thereby rendering short scan time. In the imaging scenario of submillimeter resolution, usually many shots are needed. As a result, this increases the acceleration factor per shot and thus necessitates the use of navigators. The drawback of

adding navigators is the increase of scan time. Therefore, this work aims to develop an efficient DWI protocol that can achieve submillimeter resolution while retaining short scan time.

With the operator \mathcal{A} , the joint reconstruction reads,

$$\operatorname{argmin}_{\mathbf{x}} \|\mathbf{y} - \mathcal{A}(\mathbf{x})\|_2^2 + \lambda \mathcal{R}(\mathbf{x}) \quad (2)$$

where \mathbf{y} is the measured k -space data. The first term in Equation (2) presents the data-consistency term, and the second term presents the regularization function $\mathcal{R}(x)$ with the regularization strength λ . When using the Tikhonov regularization, i.e. $\mathcal{R}(\mathbf{x}) = \|\mathbf{x}\|_2^2$, Equation (2) can be solved via the conjugate gradient (CG) method. For nonlinear regularization functions, such as the locally-low rank (LLR) regularization²⁴ or neural networks with nonlinear activation functions. ADMM was implemented in PyTorch to solve for Equation (2).

2.3 | Image Reconstruction via Self-Supervised ADMM Unrolling

Instead of the two-step alternating minimization unrolling scheme as used in MoDL³⁴, we employed the ADMM unrolling to solve the self-supervised learning reconstruction in Equation (2). The update rule of ADMM unrolling reads

$$\begin{cases} \mathbf{x}^{(k+1)} = \operatorname{argmin}_{\mathbf{x}^{(k)}} \|\mathbf{y} - \mathcal{A}(\mathbf{x}^{(k)})\|_2^2 \\ \quad + \frac{\rho}{2} \|\mathbf{x}^{(k)} - \mathbf{v}^{(k)} + \mathbf{u}^{(k)}\|_2^2 \\ \mathbf{v}^{(k+1)} = (\lambda/\rho) \cdot \mathcal{D}_\omega(\mathbf{x}^{(k+1)} + \mathbf{u}^{(k)}) \\ \mathbf{u}^{(k+1)} = \mathbf{u}^{(k)} + \mathbf{x}^{(k+1)} - \mathbf{v}^{(k+1)} \end{cases} \quad (3)$$

ADMM updates the variables \mathbf{x} , \mathbf{v} , and \mathbf{u} in an alternating scheme. It splits the unrolled reconstruction into three steps, as shown in Equation (3) and in the pseudo code of Algorithm 1. First, the updating step for \mathbf{x} is solved by conjugate gradient. Second, the variable \mathbf{v} is then updated via the forward pass of the neural network \mathcal{D}_ω with the input as the sum of current estimates of \mathbf{x} and \mathbf{u} . Third, the variable \mathbf{u} is updated by adding its current estimate to the difference between \mathbf{x} and \mathbf{v} .

Every training epoch consists of 12 looping repetitions. In each repetition, the data sampling mask \mathbf{P} is split into three disjoint sets: the training mask \mathbf{T} for the data consistency term, the training loss mask \mathbf{L} for the loss function calculation, and the validation loss mask \mathbf{V} , as shown in Figure 2. Each repetition has different masks. In each training epoch, the corresponding masks of the given repetition is used in order to update the ResNet parameters ω (Figure 2 (B)). Plus, the validation step is performed after every training epoch to update the minimal validation loss. If the validation loss does not

Algorithm 1 Self-Supervised ADMM Unrolling

```

1: Initialization:
2:   split sampling mask  $\mathbf{P}$  into 12 repetitions, each of
   which consists of three disjoint sets  $\mathbf{T}$ ,  $\mathbf{L}$ , and  $\mathbf{V}$ 
3:    $p \leftarrow 0$  and  $N_{\text{epoch}} \leftarrow 100$ 
4:    $\mathcal{D}_\omega$  set as ResNet
5:    $\rho \leftarrow 0.05$  and  $\lambda \leftarrow 0.05$ 
6:    $\text{Loss}_{\text{valid}} \leftarrow \inf$  and  $\text{trace} \leftarrow 0$ 
7:   function ADMM(mask)
8:      $\mathcal{A}_{\text{mask}} \leftarrow$  set the mask in the forward operator  $\mathcal{A}$ 
9:      $\mathbf{x}^{(0)} \leftarrow \mathcal{A}_{\text{mask}}^H(\mathbf{y})$ 
10:     $\mathbf{v}^{(0)} \leftarrow \mathbf{x}^{(0)}$  and  $\mathbf{u}^{(0)} \leftarrow \mathbf{0}$ 
11:     $k \leftarrow 0$  and  $N_{\text{unroll}} \leftarrow 8$ 
12:    while  $k < N_{\text{unroll}}$  do
13:       $\mathbf{x}^{(k+1)} \leftarrow$  conjugate gradient with 6 iterations
14:       $\mathbf{v}^{(k+1)} \leftarrow (\lambda/\rho) \cdot \mathcal{D}_\omega(\mathbf{x}^{(k+1)} + \mathbf{u}^{(k)})$ 
15:       $\mathbf{u}^{(k+1)} \leftarrow \mathbf{u}^{(k)} + \mathbf{x}^{(k+1)} - \mathbf{v}^{(k+1)}$ 
16:       $k \leftarrow k + 1$ 
17:    end while
18:    return  $\mathbf{x}^{(k+1)}$ 
19:  end function
20: Training:
21: while  $p < N_{\text{epoch}}$  or  $\text{trace} \leq 12$  do
22:    $\mathbf{x}_t \leftarrow \text{ADMM}(\mathbf{T})$ 
23:    $\text{Loss}_{\text{train}} \leftarrow \mathcal{L}(\mathbf{Ly}, \mathcal{A}_{\mathbf{L}}(\mathbf{x}_t))$ 
24:   update  $\omega$  via ADAM
25: Validation:
26:    $\mathbf{x}_t \leftarrow \text{ADMM}(\mathbf{T} \cup \mathbf{L})$ 
27:    $\text{Loss}_{\text{temp}} \leftarrow \mathcal{L}(\mathbf{Vy}, \mathcal{A}_{\mathbf{V}}(\mathbf{x}_t))$ 
28:   if  $\text{Loss}_{\text{temp}} \leq \text{Loss}_{\text{valid}}$  then
29:      $\text{Loss}_{\text{valid}} \leftarrow \text{Loss}_{\text{temp}}$ 
30:      $\text{trace} \leftarrow 0$ 
31:   else
32:      $\text{trace} \leftarrow \text{trace} + 1$ 
33:   end if
34: end while

```

reduce for 12 consecutive epochs or if 100 epochs are reached, the training is terminated. The use of three disjoint masks is inline with the zero-shot self-supervised learning approach^{27,28} for scan-specific parallel imaging reconstruction. In contrast, Self-supervised learning via data undersampling (SSDU)³⁸ splits the sampling mask into only two sub-masks, but requires multiple datasets for training.

The index k in Equation (3) denotes the unrolling iteration, and \mathcal{D}_ω denotes the ResNet²⁶ parameterized by ω . In this work, 2D convolution was employed to construct the ResNet layers. In PyTorch, 2D convolution requires four-dimensional tensors as input and output. For instance, a matrix with the size (N, C, H, W) is acceptable for the 'conv2d' function in PyTorch.

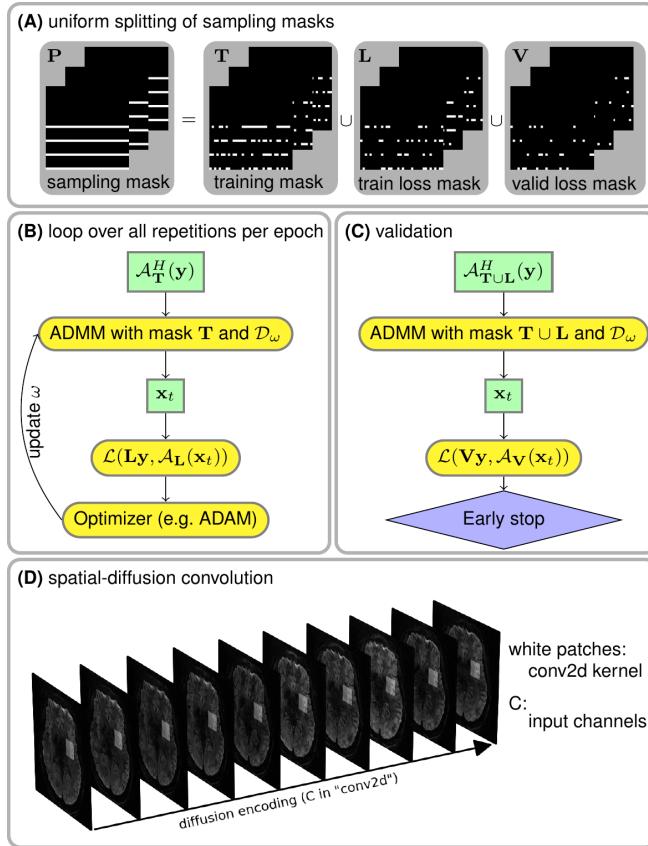


FIGURE 2 Illustration of the key components in ADMM unrolling. **(A)** The sampling mask P in Equation (1) was uniformly split into three disjoint sets: the training mask T used for the data consistency term during training, the train loss mask L used for the loss function calculation during training, and the validation loss mask V used for the loss function calculation during validation. **(B)** and **(C)** show the flowchart for the training and the validation of an unrolled ADMM model, respectively. Note that the ResNet parameters ω are updated via ADAM³⁷ during training, but remain fixed during the validation step. **(D)** A stack of diffusion-weighted images is input into ResNet during ADMM unrolling.

Here, W and H denote the width and height of the convolution kernel, C denotes the number of channels, and N denotes the batch size. However, the diffusion-weighted images (\mathbf{x}) to be reconstructed have the size $(N_{\text{diff}}, N_Z, N_Y, N_X, 2)$, where 2 stands for the real and imaginary part of the complex-valued diffusion-weighted images, N_X and N_Y are the width and the height of diffusion-weighted images, N_Z is the number of slices (identical to the multi-band factor), and N_{diff} is the number of diffusion encoding. To train a ResNet based on 2D convolution, the diffusion-weighted images were reshaped and permuted as $(N_Z, 2 \cdot N_{\text{diff}}, N_Y, N_X)$, as illustrated in Figure 2 (D). In this manner, 2D convolution kernels in combination with ReLU activation

functions loop through the varying diffusion-weighted contrast to learn the key features of the high-dimensional data and to reduce noisy and aliasing artifacts in unrolled reconstruction.

2.4 | Model Generalizability

Volumetric whole brain DWI acquisition consists of many multi-band slices, and the training of algorithm unrolling models on all slices requires hundreds of GPU computing hours. To investigate the model generalizability and to accelerate reconstruction, we performed two training and inference strategies. First, we trained the ADMM unrolling model with only one multi-band slice data, and subsequently tested the model on all remaining multi-band slices. We called this approach "single-slice training". Second, we trained and tested every multiband slice individually, which was referred to as "slice-by-slice training". The single-slice training strategy saves tremendous computing time, as its model is learned from one single slice and the inference time per slice is only about one minute. By comparing these two training strategies, we aim at demonstrating the model generalizability and its applicability to other slices which are "unseen" in single-slice training.

2.5 | Comparison of Regularization Techniques

This work compared the reconstruction performance of three different regularization techniques, Tikhonov ℓ^2 regularization (as used in MUSE), LLR regularization, and ADMM unrolling with a learned regularization. Note that MUSE is a simultaneous multi-slice (SMS) parallel imaging method and poses no regularization along the diffusion dimension, effectively solving each DWI reconstruction independently. In contrast, the other two regularized reconstructions fall into the joint reconstruction regime. They jointly reconstruct all diffusion-weighted images and impose regularization terms exploring spatial-diffusion redundancies. For example, LLR enforces the low rankness of local spatial-diffusion matrices from diffusion-weighted images, whereas ADMM unrolling learns a regularization function composed by neural networks based on spatial-diffusion convolution kernels while enforcing data consistency during the unrolled training process.

3 | RESULTS

3.1 | Ablation Study

Figure 3 validates the proposed self-supervised ADMM unrolling reconstruction method with the 4-shot fully-sampled 1.0 mm dataset (Protocol #1 in Table 1). MUSE on the 2-shot undersampled data exhibits noticeable image quality degradations, as confirmed by the visual inspections as well as the SSIM and PSNR quantities. Both LLR and ADMM unrolling are capable of reconstructing high quality diffusion-weighted images without significant loss of image details and SNR. The computed quantitative metrics show ADMM unrolling performs slightly better than LLR.

3.2 | Model Generalizability

Figure 4 demonstrates the generalizability of the proposed ADMM unrolling approach, i.e., an unrolled ADMM model trained on one single multi-band slice is applicable to all remaining "unseen" slices. Single-direction diffusion-weighted images from both the slice-by-slice training and the single-slice training strategies are displayed. The absolute difference between these two images shows no residual structural information, but mainly noise. Moreover, we plotted the mean and standard deviation within the selected region-of-interest (colored boxes in Figure 4) along all diffusion encoding directions. This again proves the cross-slice generalization of the proposed selfgated self-supervised ADMM unrolling method. The plotted curves show quantitatively similar values between the two training strategies. With this, the following results were obtained based upon the single-slice training strategy.

3.3 | Self-Gated ADMM Unrolling

Figure 5 demonstrates the efficacy of the self-gated self-supervised ADMM unrolling reconstruction by comparing with the navigated reconstruction on the first volunteer. Both MUSE and ADMM unrolling reconstructions were performed. Data were acquired using the NAViEPI sequence, as listed in Protocol #2 in Table 1. The single-direction diffusion-weighted images are displayed.

The diffusion-weighted images from navigated reconstructions show spatially varying phase. The reason is that the shot-to-shot phase variations were estimated from the second echo in NAViEPI, i.e., the navigator, whose echo time is different from the first echo. The echo time difference results in residual phases in the combined

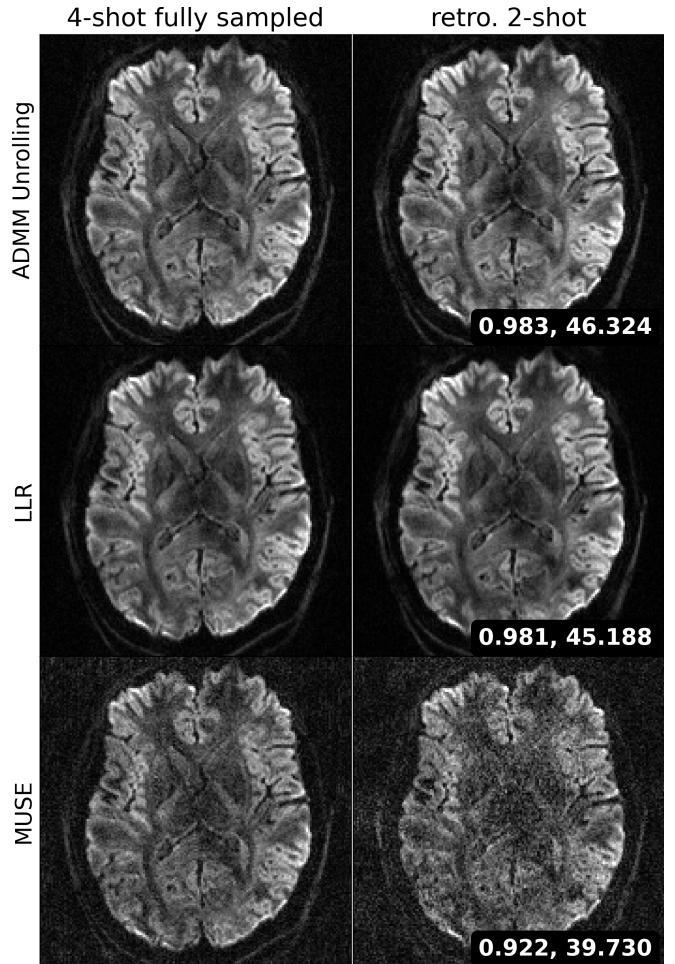


FIGURE 3 Ablation study with the fully-sampled reference data acquired by Protocol #1. The first column displays one diffusion-weighted image from the 4-shot fully-sampled data reconstructed via (from top to bottom) the proposed self-supervised ADMM unrolling, LLR, and MUSE. The second column displays the diffusion-weighted image from the retrospectively undersampled 2-shot data reconstructed via the afore-mentioned methods. Two image metrics, structural similarity index measure (SSIM) and peak signal-to-noise ratio (PSNR) are computed between the 4-shot and the 2-shot reconstructions.

diffusion-weighted images. On the contrary, selfgated reconstructions show only subtle phase, because shot-to-shot phase variations were estimated from the first echoes themselves. The reduced phase variation in the selfgated reconstruction leads to less phase ambiguity. This is beneficial in the ADMM unrolling reconstruction, where convolutions were performed in both the real and imaginary channels. Reduced phase ambiguity fosters the learning procedure. Consequently, compared to MUSE with the MPPCA denoiser³⁹, the selfgated ADMM unrolling reconstruction achieves strong denoising and resolves clear tissue details.

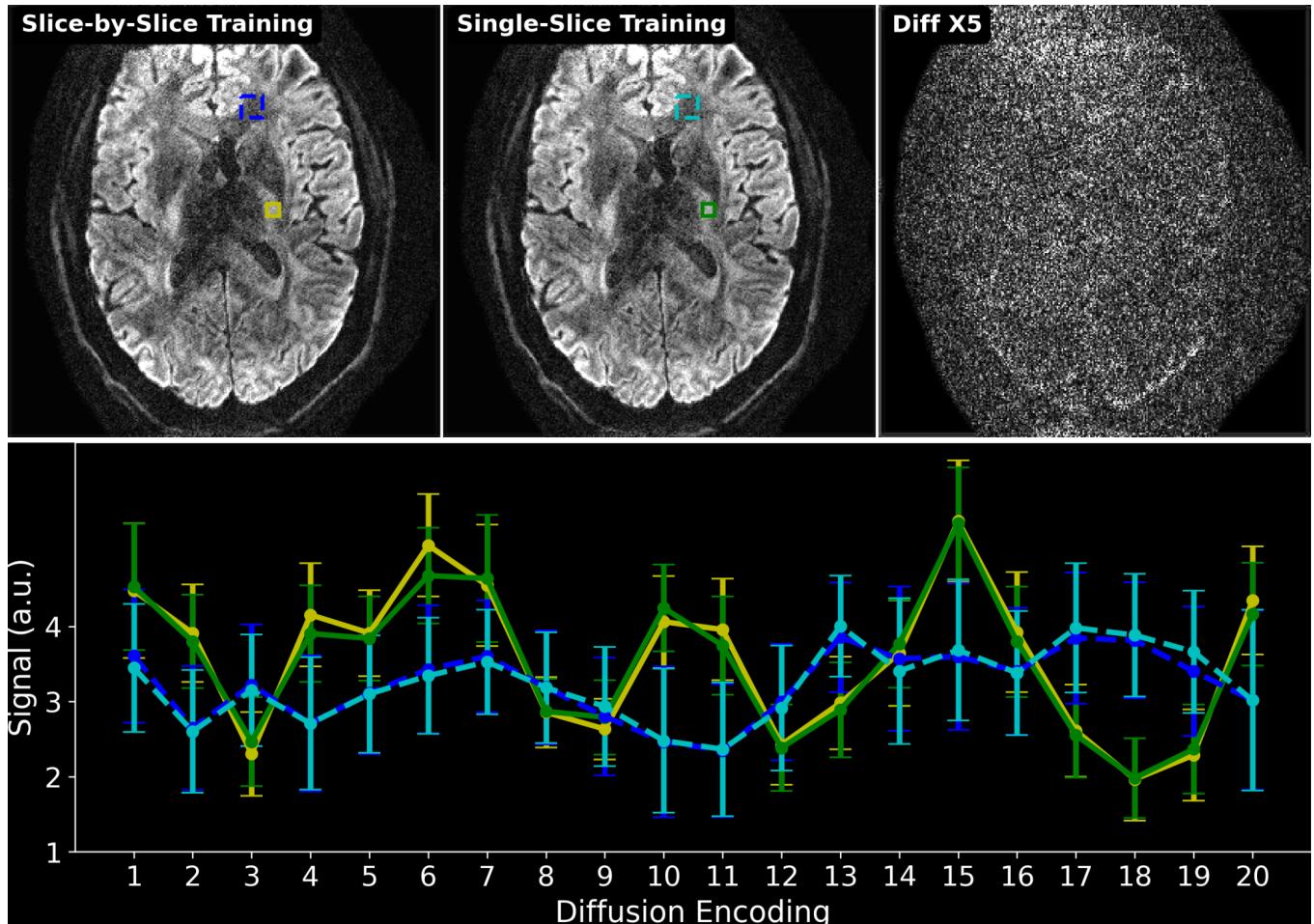


FIGURE 4 Comparison of two training strategies: (1) slice-by-slice training, where every slice is trained and tested individually; (2) single-slice training, where the unrolled ADMM model is trained on only one slice and tested on all remaining slices. The top-right image shows the absolute difference between the reconstructed diffusion-weighted images at the 10th diffusion direction between (1) and (2). The bottom panel plots the mean and standard deviation of the signal within two sets of rectangles in the slice-by-slice training and the single-slice training, respectively. No major qualitative or quantitative difference can be seen between the two training strategies.

The advantage of the proposed ADMM unrolling for high resolution DWI with accelerated acquisition is further evident in Figure 6. The mean diffusion-weighted image from ADMM unrolling shows clear delineation of the claustrum, which is a thin sheet of neurons and is important to consciousness. In contrast, MUSE with the MPPCA denoiser shows noisy and blurred boundaries of the claustrum.

Figure 7 shows coronal- and sagittal-view diffusion-weighted images with the same diffusion encoding as in Figure 5. As mentioned in Section 2.4, the unrolled ADMM model was trained using only one slice and then inferred on all remaining slices. Again, the single-slice model generalizes well across slices. The inference of every slice takes only about one minute, whereas the LLR reconstruction takes about 48 minutes per slice. More importantly, the self-gated LLR reconstruction

exhibits residual motion-induced stripping artifacts⁴⁰, whereas the self-gated ADMM unrolling approach substantially removes these artifacts and supplies high-quality diffusion-weighted images without the need of navigators. Both reconstructions show B_1 field inhomogeneities in the cerebellum region as well as off-resonance induced spatial distortion in the frontal brain region. These artifacts, however, are beyond the scope of this work.

3.4 | Diffusion Tensor Imaging (DTI)

Figures 4 and 8 utilize the Protocol #3 acquired data from the same volunteer. Here, Figure 8 displays the cFA maps based on the reconstructed diffusion-weighted images by MUSE with denoiser, LLR, and ADMM unrolling, respectively. Given the 2×2 -fold acceleration

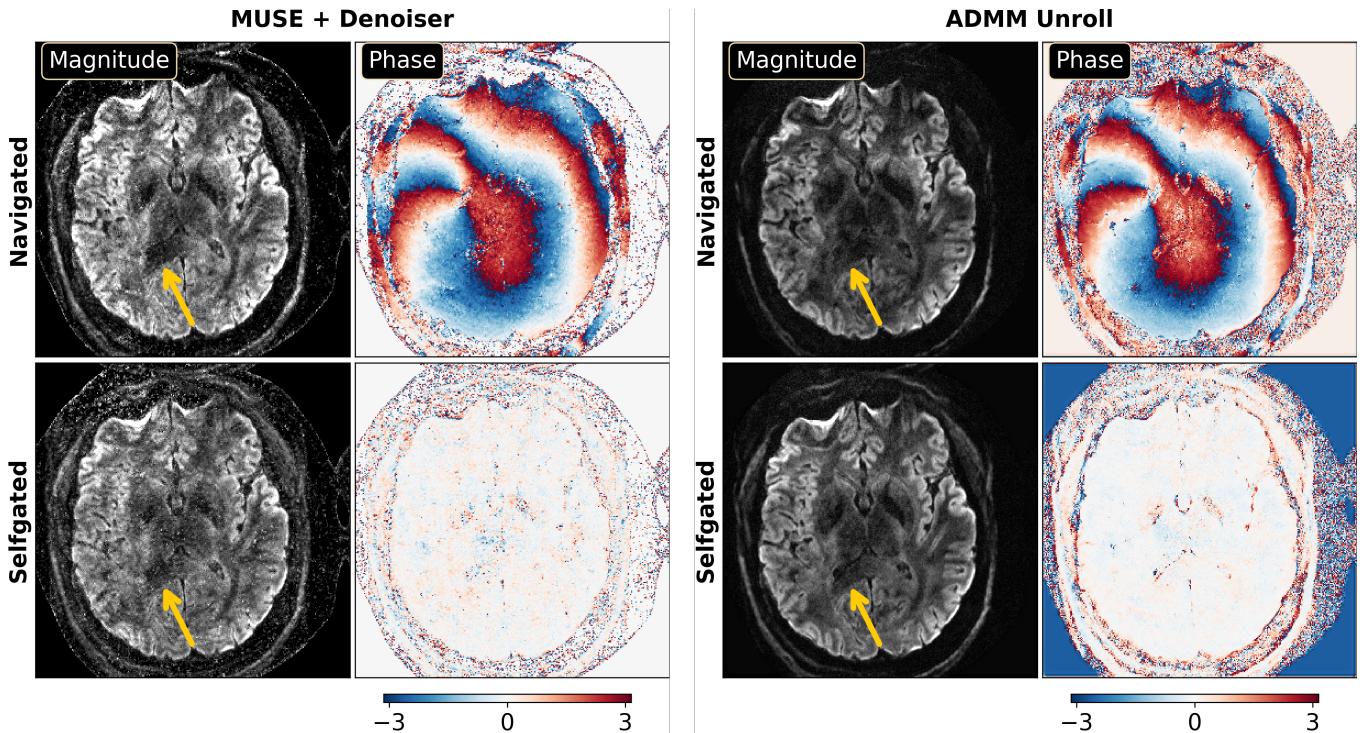


FIGURE 5 Validation of the proposed self-gated ADMM unrolling reconstruction with data acquired by Protocol #2 in Table 1. Both MUSE and ADMM unrolling were performed with navigated and selfgated shot-to-shot phase maps, respectively. Compared to MUSE, the ADMM unrolled reconstruction excels in denoising while maintaining structural details. Selfgated ADMM unrolling shows improved image quality in terms of tissue delineation than navigated reconstruction.

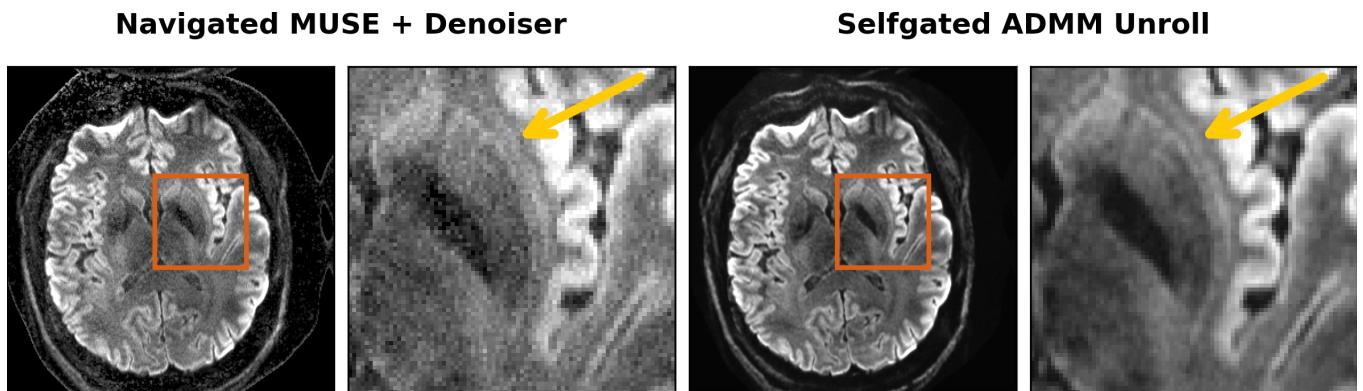


FIGURE 6 0.7 mm isotropic resolution DWI with the proposed selfgated ADMM unrolling enables the visualization of the tiny structure claustrum, whereas the MUSE reconstruction shows only blurred appearance. Displayed images are the mean diffusion-weighted image from 20 directions and its zoomed-in region.

and the submillimeter spatial resolution (0.7 mm), the MPPCA denoiser applied onto MUSE is insufficient to supply sharp fiber orientations. Although LLR shows improvements when compared to the MUSE approach, but still shows overall blurring in the cFA map, especially within the gray matter region. The proposed self-supervised ADMM unrolling is able to resolve thin fibers within gray matters, as pointed by the color arrows.

Figure 9 displays the training and validation loss as well as the learned regularization strength along epochs. It can be seen that 100 epochs are sufficient for the convergence of ADMM unrolling. The model converges well along epochs, and does not show any over-fitting behavior (The validation loss decays similarly as the training loss). In addition, the regularization strength converges to the value of about 0.027. Note that the validation loss is slightly low than the training loss. This is because more

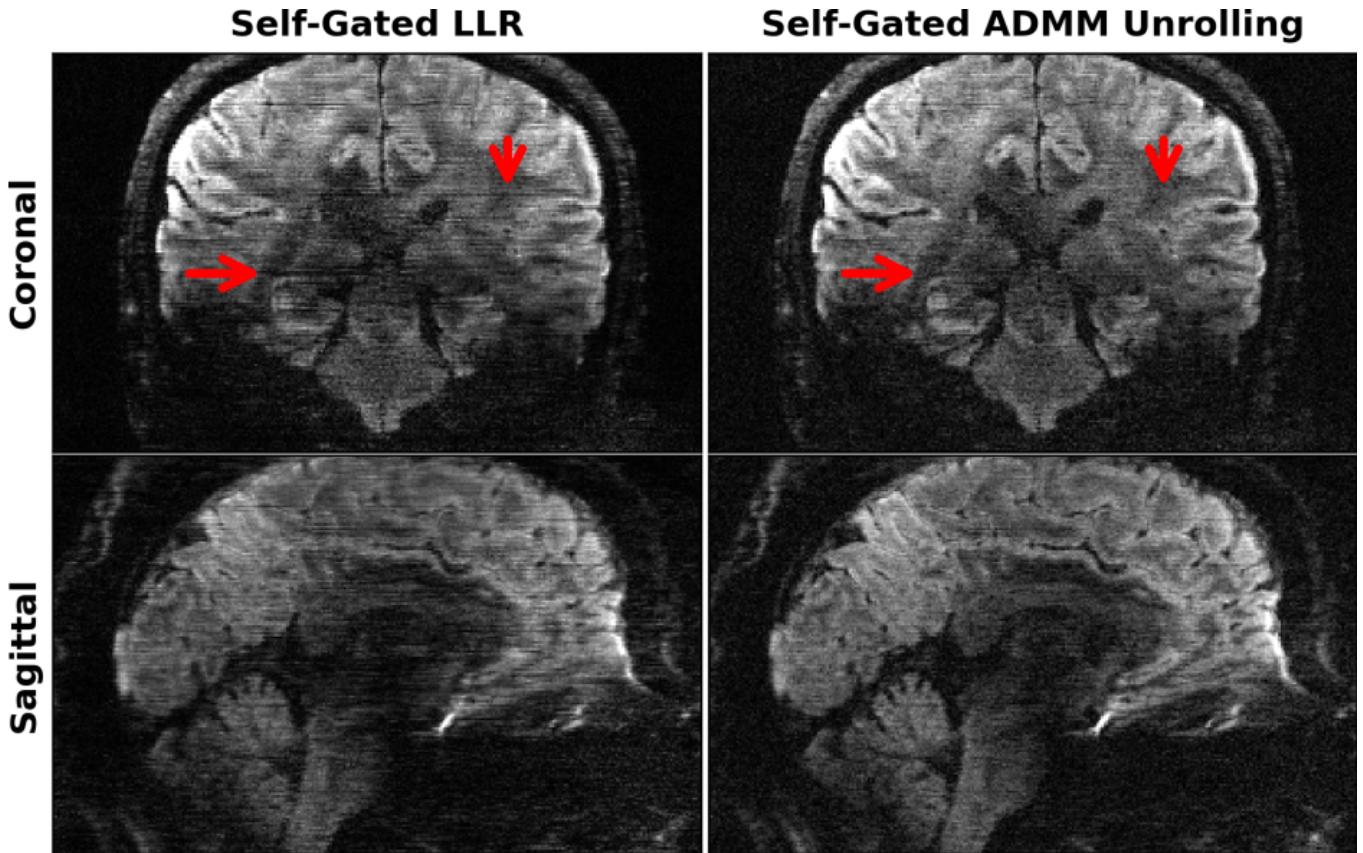


FIGURE 7 Single-direction diffusion-weighted images at 0.7 mm isotropic resolution as reconstructed by retrospectively self-gated (left) LLR and (right) ADMM unrolling in (top) the coronal and (bottom) the sagittal views, respectively. The same diffusion direction as in Figure 5 is chosen for display. ADMM unrolling reduces phase ambiguities in the shot-combined reconstruction, thereby rendering clearer tissue delineation and reducing stripping artifacts (as indicated by the red arrows).

data is split into the training mask than the validation mask.

4 | DISCUSSION

This work reports a novel self-gated self-supervised learning approach based on ADMM unrolling for multi-shot multi-band undersampled iEPI acquisition and high-resolution DWI reconstruction. The self-gated ADMM unrolling achieves whole brain DWI with 21 diffusion volumes and a b -value of 1000s/mm^2 at 0.7 mm isotropic resolution, all within a scan time of less than 10 minutes. Our proposed ADMM unrolling approach has several advantages. (1) Inline with the previous approaches for single image recovery²⁷ and parallel imaging²⁸, our approach trains an unrolled reconstruction network with only one dataset utilizing the concept of data splitting^{27,38,28}. Therefore, our approach is scan-specific and does not require large-scale datasets for training. (2) Our approach explores the joint k - q -space redundancy with the use of spatial-diffusion convolutions and is

also constrained by the physics-based data consistency. Therefore, our approach is versatile to downstream diffusion model analysis (e.g., DTI). (3) We observe that the ADMM unrolling model can be trained from one single multi-band slice and is generalizable to other "unseen" multi-band slices. This substantially reduces the required training time. Furthermore, given that unrolled reconstructions require much shorter inference time than conventional iterative regularized reconstructions such as compressed sensing, our approach is feasible for clinical translation.

This work demonstrated the capability of self-gated ADMM unrolling in reconstructing 0.7 mm isotropic resolution 3-shot iEPI DWI with (6×2) -fold acceleration per shot. However, we also observed that the self-gated approach failed to recover aliasing-free diffusion-weighted images in the case of higher acceleration factors (e.g. the $0.5 \times 0.5 \times 2.0 \text{ mm}^3$ DWI data with an acceleration of 15×2 per shot). To address this issue, acquiring shot-to-shot phase navigators helps with the shot-combined DWI reconstruction²⁴. Therefore, the utilization of navigator acquisition and advanced deep

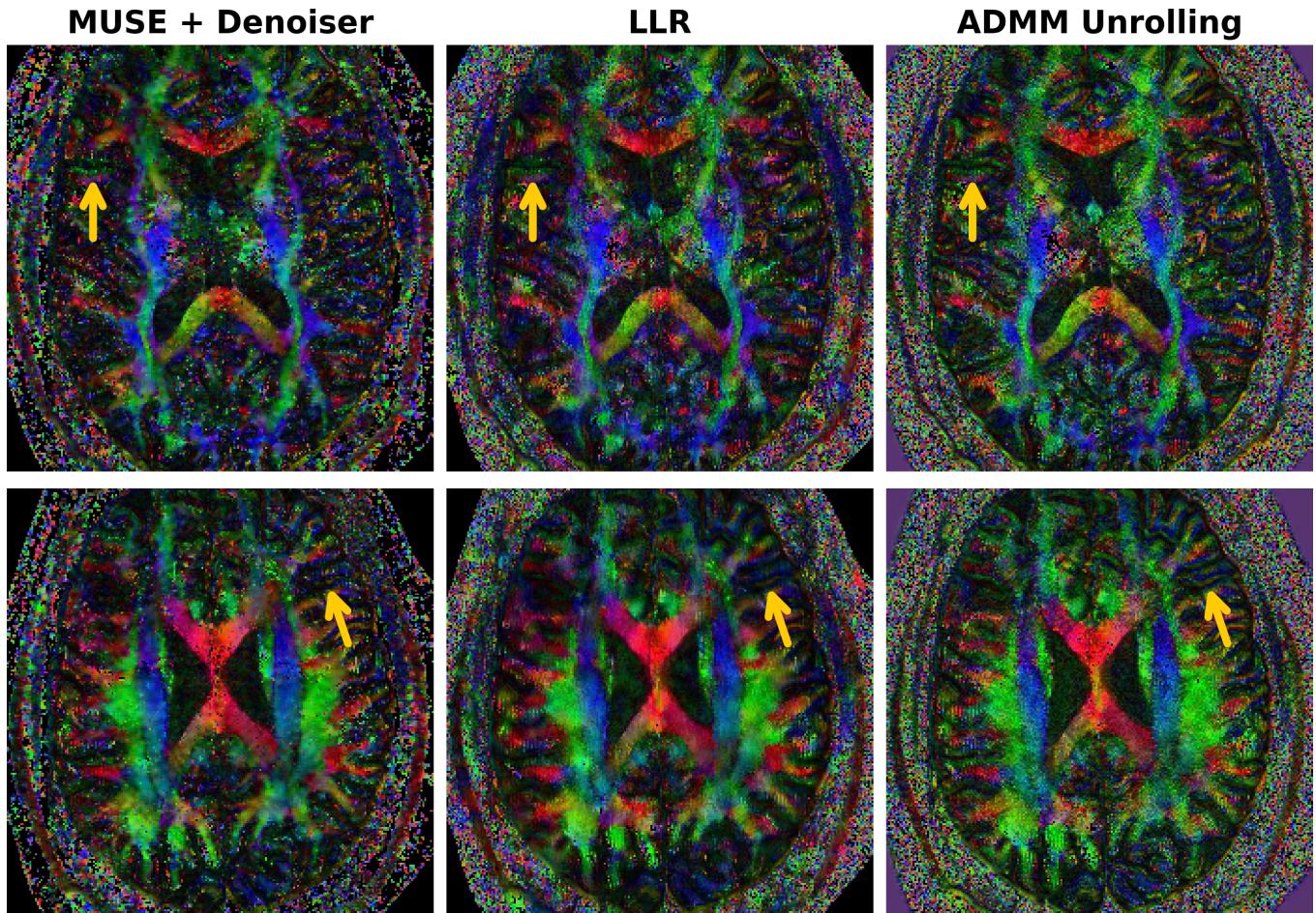


FIGURE 8 Diffusion tensor imaging (DTI) model derived colored fractional anisotropy (cFA) maps based on the diffusion-weighted images as reconstructed by MUSE, LLR, and ADMM unrolling, respectively. Two slices are displayed in the top and the bottom row, respectively. Even with limited scan time (about 10 minutes at 0.7 mm spatial resolution) and limited diffusion directions (20), the proposed self-gated self-supervised ADMM unrolling reconstruction delivers clearer fiber orientations, as indicated by the maize-color arrows.

learning reconstruction should be application oriented. Ultra-high spatial resolution, which requires many shots, necessitates the use of navigator shots. For the 0.7 mm resolution with 3 shots, as shown in this work, the self-gated acquisition is beneficial of reducing scan time, given the superior performance of the proposed ADMM unrolling reconstruction. Alternatively, employing optimized trajectories with a more densely sampled k -space central region could help better estimate shot phase variations⁴.

We observed that stripping-type motion artifacts occurred more frequently in the sub-millimeter isotropic resolution DWI regime. This makes sense, as scans with reduced slice thickness are more susceptible to shot-to-shot phase variations. In addition, sub-millimeter isotropic voxel resulted in higher noise in diffusion-weighted images. Since the primary aim of this work is to develop an efficient self-supervised learning technique for sub-millimeter DWI, we did not explore other advanced

sampling strategies such as gSlider. However, because unrolled algorithms are flexible to MR physics modeling (e.g., the forward operator \mathcal{A}), the proposed ADMM unrolling is extendable to incorporate with the gSlider encoding model for enhanced SNR performance.

This work does have several limitations. (1) This work did not incorporate off-resonance correction in the reconstruction. As a logic extension, the multi-shot sequence can be modified to encode dynamic B_0 field variation, which can then be employed in the SENSE-based forward operator and reconstruction. An established approach is known as the blip-up/down encoding⁴¹. This approach requires the acquisition of two images with opposing phase-encoding polarities (i.e., blip-up and blip-down) for the computation of B_0 field maps. An alternative approach is to iteratively update B_0 field based on the phase difference among acquired multiple echoes⁴². This approach does not require the

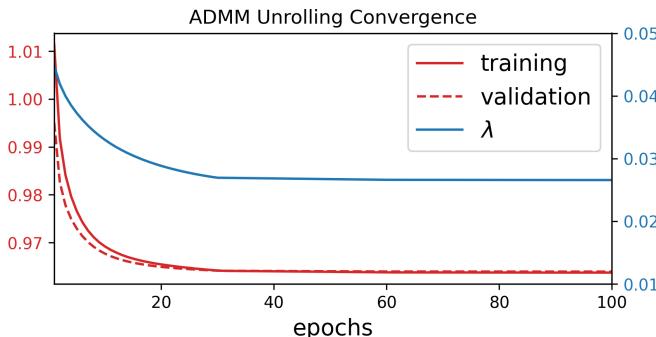


FIGURE 9 Convergence analysis along the ADMM unrolling training and validation epochs. Displayed curves are (red solid) the training loss, (red dashed) the validation loss, and (blue solid) the learned regularization strength λ , respectively. All parameters converge sufficiently and show no over-fitting.

pre-determination of B_0 field, but poses higher computational demand in the inversion course of phase increments from every echo. (2) As this work primarily focused on the development of selfgated self-supervised unrolled reconstruction for high-resolution DWI, only three volunteers were recruited. A pilot study with a large number of volunteers and even patients is beyond the scope of this work. (3) Given the small sample size, it is unlikely that we compare our proposed approach with other semi-self-supervised approaches such as SSDU³⁸. Only MUSE and LLR are chosen for comparison in this work. However, we believe that MUSE and LLR are competitive and representative methods to be compared with, as the former one has already been translated to clinical practice and the latter one has been widely used for multi-contrast compressed sensing image reconstruction.

5 | CONCLUSIONS

In this work, we proposed a self-gated self-supervised learning reconstruction framework based on ADMM unrolling for high-resolution and motion-robust diffusion-weighted imaging at ultra-high field. Based on the mechanism of data splitting (cross validation), our proposed ADMM unrolling requires only one single multi-band slice for training and is generalized cross-slice. Plus, ADMM unrolling renders ultra-short inference / reconstruction time, and is thus feasible for clinical translation.

ACKNOWLEDGMENTS

This work was supported in part by German Research Foundation (DFG) under projects 513220538 and 512819079, project 500888779 in the Research Unit RU5534 for MR biosignatures at UHF, and by the National Institutes of Health (NIH) under grants R01EB024532 and P41EB017183. The authors are grateful to scientific support and HPC resources provided by the Erlangen National High Performance Computing Center (NHR) of Friedrich-Alexander-University Erlangen-Nuremberg (FAU) under the NHR project b143dc. NHR is funded by federal and Bavarian state authorities. NHR@FAU hardware is partially funded by DFG under project 440719683. The authors are thankful to Dr. Vikas Gulani and Dr. Xiaoqing Wang for insightful discussions. The authors thank ChatGPT for the revision of the Introduction section.

DATA AVAILABILITY STATEMENT

In the spirit of open science and reproducible research, source codes of this work are available in <https://github.com/ZhengguoTan/DeepDWI>. The presented 0.7 mm DWI raw k -space data is available in <https://doi.org/10.5281/zenodo.10781347> and <https://doi.org/10.5281/zenodo.13864504>.

ORCID

- Zhengguo Tan 0000-0002-4322-5109
- Patrick A Liebig 0000-0001-7342-3715
- Annika Hofmann 0009-0002-5723-1769
- Frederik B Laun 0000-0002-9269-5609
- Florian Knoll 0000-0001-5357-8656

REFERENCES

1. Jones Derek K. *Diffusion MRI: Theory, methods, and applications*. Oxford University Press; 2010.
2. Mansfield Peter. Multi-planar image formation using NMR spin echoes. *J Phys C*. 1977;10:55-58.
3. Butts Kim, Riederer Stephen J, Ehman Richard L, Thompson Richard M, Jack Clifford R. Interleaved echo planar imaging on a standard MRI system. *Magn. Reson. Med.*. 1993;31:67-72.
4. Liu Chunlei, Bammer Roland, Kim Dong-hyun, Moseley Michael E. Self-navigated interleaved spiral (SNAILS): Application to high-resolution diffusion tensor imaging. *Magn. Reson. Med.*. 2004;52:1388-1396.
5. Chen Nan-Kuei, Guidon Arnaud, Chang Hing-Chiu, Song Allen W. A robust multi-shot scan strategy for high-resolution diffusion weighted MRI enabled by

- 01 multiplexed sensitivity-encoding (MUSE). *NeuroImage*.
02 2013;72:41-47.
- 03 6. Porter David A, Heidemann Robin M. High resolution
04 diffusion-weighted imaging using readout-segmented
05 echo-planar imaging, parallel imaging and a two-
06 dimensional navigator-based reacquisition. *Magn. Reson.*
07 *Med.*. 2009;62:468-475.
- 08 7. Heidemann Robin M, Porter David A, Anwander Alfred,
09 et al. Diffusion imaging in humans at 7 T using readout-
10 segmented EPI and GRAPPA. *Magn. Reson. Med.*.
11 2010;64:9-14.
- 12 8. Andersson Jesper L R, Skare Stefan, Ashburner John.
13 How to correct susceptibility distortions in spin-echo
14 echo-planar images: application to diffusion tensor imag-
15 ing. *NeuroImage*. 2003;20:870-888.
- 16 9. Setsompop Kawin, Gagoski Borjan A, Polimeni
17 Jonathan R, Witzel Thomas, Wedeen Van J, Wald
18 Lawrence L. Blipped-controlled aliasing in parallel
19 imaging for simultaneous multislice echo planar imag-
20 ing with reduced *g*-factor penalty. *Magn. Reson. Med.*.
21 2012;67:1210-1224.
- 22 10. Pruessmann Klaas P, Weiger Markus, Scheidegger
23 Markus B, Boesiger Peter. SENSE: Sensitivity encoding
24 for fast MRI. *Magn. Reson. Med.*. 1999;42:952-962.
- 25 11. Griswold Mark A, Jakob Peter M, Heidemann Robin M,
26 et al. Generalized autocalibrating partially parallel acqui-
27 sitions (GRAPPA). *Magn. Reson. Med.*. 2002;47:1202-
28 1210.
- 29 12. Bammer Roland, Keeling Stephen L, Augustin Michael,
30 et al. Improved diffusion-weighted single-shot echo-planar
31 imaging (EPI) in stroke using sensitivity encoding
32 (SENSE). *Magn. Reson. Med.*. 2001;46:548-554.
- 33 13. Lustig Michael, Donoho David, Pauly John M. Sparse
34 MRI: The application of compressed sensing for rapid
35 MR imaging. *Magn. Reson. Med.*. 2007;58:1182-1195.
- 36 14. Mani Merry, Jacob Mathews, Kelley Douglas, Magnotta
37 Vincent. Multi-shot sensitivity-encoded diffusion data
38 recovery using structured low-rank matrix completion
39 (MUSSELS). *Magn. Reson. Med.*. 2017;78:494-507.
- 40 15. Hu Yuxin, Wang Xiaole, Tian Qiyuan, et al. Multi-shot
41 diffusion-weighted MRI reconstruction with magnitude-
42 based spatial-angular locally low-rank regularization
43 (SPA-LLR). *Magn. Reson. Med.*. 2020;83:1596-1607.
- 44 16. Veraart Jelle, Novikov Dmitry S, Christiaens Daan, Ades-
45 aron Benjamin, Sijbers Jan, Fieremans Els. Denoising of
46 diffusion MRI using random matrix theory. *NeuroImage*.
47 2016;142:394-406.
- 48 17. Tian Qiyuan, Li Ziyu, Fan Qiuyun, et al. SDnDTI: Self-
49 supervised deep learning-based denoising for diffusion
50 tensor MRI. *NeuroImage*. 2023;253:119033.
- 51 18. Setsompop Kawin, Fan Qiuyun, Stockmann Jason, et al.
52 High-resolution in vivo diffusion imaging of the human
53 brain with generalized slice dithered enhanced resolution:
Simultaneous multislice (gSlider-SMS). *Magn. Reson.*
Med.. 2018;79:141-151.
- 54 19. Dai Erpeng, Liu Simin, Guo Hua. High-resolution whole-
55 brain diffusion MRI at 3T using simultaneous multi-slab
56 (SMSlab) acquisition. *NeuroImage*. 2021;237:118099.
- 57 20. Dong Zijing, Reese Timothy G, Lee Hong-Hsi, et al. Romer-EPTI: Rotating-view motion-robust super-
58 resolution EPTI for SNR-efficient distortion-free in-vivo
59 mesoscale diffusion MRI and microstructure imaging.
Magn. Reson. Med.. 2024;.
- 60 21. Liang Zhi-Pei. Spatiotemporal imaging with partially
61 separable functions. In: :988-991; 2007.
- 62 22. Zhang Tao, Pauly John M, Levesque Ives R. Accelerated
63 parameter mapping with a locally low rank constraint.
Magn. Reson. Med.. 2015;73:655-661.
- 64 23. Dong Zijing, Wang Fuyixue, Reese Timothy G, Bilgic
65 Berkin, Setsompop Kawin. Echo planar time-resolved
66 imaging with subspace reconstruction and optimized spa-
tiotemporal encoding. *Magn. Reson. Med.*. 2020;84:2442-
2455.
- 67 24. Tan Zhengguo, Liebig Patrick Alexander, Heidemann
68 Robin Martin, Laun Frederik Bernd, Knoll Florian.
Accelerated diffusion-weighted magnetic resonance imag-
69 ing at 7 T: Joint reconstruction for shift-encoded
70 navigator-based interleaved echo planar imaging (JETS-
71 NAViEPI). *Imaging Neuroscience*. 2024;2:1-15.
- 72 25. Cho Jaejin, Jun Yohan, Wang Xiaoqing, Kobayashi
73 Caique, Bilgic Berkin. Improved Multi-shot Diffusion-
74 Weighted MRI with Zero-Shot Self-supervised Learning
75 Reconstruction. In: :457-466; 2023.
- 76 26. He Kaiming, Zhang Xiangyu, Ren Shaoqing, Sun Jian.
Deep residual learning for image recognition. In: :770-778;
2016.
- 77 27. Quan Yuhui, Chen Mingqin, Pang Tongyao, Ji Hui.
Self2Self With Dropout: Learning Self-Supervised
78 Denoising From Single Image. In: ; 2020.
- 79 28. Yaman Burhaneddin, Hosseini Seyed Amir Hossein,
Akçakaya Mehmet. Zero-shot self-supervised learning for
80 MRI reconstruction. In: ; 2022.
- 81 29. Mani Merry, Magnotta Vincent A, Jacob Mathews.
qModeL: A plug-and-play model-based reconstruction for
82 highly accelerated multi-shot diffusion MRI using learned
83 priors. *Magn. Reson. Med.*. 2021;86:835-851.
- 84 30. Hinton G. E., Salakhutdinov R. R.. Reducing the
85 dimensionality of data with neural networks. *Science*.
2006;313:504-507.
- 86 31. Boyd Stephen, Parikh Neal, Chu Eric, Peleato Borja,
Eckstein Jonathan. Distributed optimization and statisti-
87 cal learning via the alternating direction method of mul-
88 tipliers. *Foundations and Trends in Machine Learning*.
2010;3:1-122.
- 89 32. Wang Shanshan, Su Zhenghang, Ying Leslie, et al. Accel-
90 erating magnetic resonance imaging via deep learning. In:
91 :514-517; 2016.
- 92 33. Hammernik Kerstin, Klatzer Teresa, Kobler Erich, et
93 al. Learning a variational network for reconstruction of
94 accelerated MRI data. *Magn. Reson. Med.*. 2018;79:3055-
95 3071.
- 96 34. Aggarwal Hemant K, Mani Merry P, Jacob Mathews.
MoDL: Model-based deep learning architecture
97 for inverse problems. *IEEE Trans. Med. Imaging*.
2018;38:394-405.
- 98 35. Uecker Martin, Karaus Alexander, Frahm Jens. Inverse
99 reconstruction method for segmented multishot diffusion-
100 weighted MRI with multiple coils. *Magn. Reson. Med.*.
101 102 103
- 104 105 106

- 01 2009;62:1342-1348.
- 02 36. Merrem Andreas, Hofer Sabine, Hosseini Ali Seif Amir,
03 et al. Diffusion-weighted MRI of the prostate without
04 susceptibility artifacts: Undersampled multi-shot turbo-
05 STEAM with rotated radial trajectories. *NMR Biomed.*.
06 2019;32:e4074. 54
- 07 37. Kingma Diederik P, Ba Jimmy. ADAM: A Method for
08 Stochastic Optimization. In: ; 2015. 55
- 09 38. Yaman Burhaneddin, Hosseini Seyed Amir Hossein,
10 Moeller Steen, Ellermann Jutta, Uğurbil Kâmil,
11 Akçakaya Mehmet. Self-supervised learning of physics-
12 guided reconstruction neural networks without
13 fully sampled reference data. *Magn. Reson. Med.*.
14 2020;84:3172-3191. 61
- 15 39. Cordero-Grande Lucilio, Christiaens Daan, Hutter Jana,
16 Price Anthony N, Hajnal Jo V. Complex diffusion-
17 weighted image estimation via matrix recovery under
18 general noise models. *NeuroImage*. 2019;200:391-404. 66
- 19 40. Pietsch Maximilian, Christiaens Daan, Hajnal Joseph V,
20 Tournier J-Donald. dStripe: Slice artefact correction in
21 diffusion MRI via constrained neural network. *Medical
Image Analysis*. 2021;74:102255. 70
- 22 41. Zahneisen Benjamin, Aksoy Murat, Maclarens Julian,
23 Wuerslin Christian, Bammer Roland. Extended hybrid-
24 space SENSE for EPI: Off-resonance and eddy current
25 corrected joint interleaved blip-up/down reconstruction.
26 *NeuroImage*. 2017;153:97-108. 74
- 27 42. Tan Zhengguo, Unterberg-Buchwald Christina, Blumen-
28 thal Moritz, et al. Free-breathing liver fat, R_2^* and B_0
29 field mapping using multi-echo radial FLASH and reg-
30 ularized model-based reconstruction. *IEEE Transactions
31 on Medical Imaging*. 2023;42:1374-1387. 78
- 32 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106
-

MRM-24-25421: Responses to Editors and Reviewers

Editor's Comments to the Author:

Thanks for submitting to Magn Reson Med. Firstly, apologies for the delay in getting you an outcome – we struggled to identify willing reviewers. Finally, though, we found two expert referees. As you can see, although there is potential interest from them in the study there are substantial issues raised by both. As such I regret that we must return the current version of the manuscript to you. We would allow a future resubmission if you wished to undertake the considerable work outlined by the reviewers, but we would need to see a substantial improvement in reviewer enthusiasm and accompanying scores. Alternatively, you would be at liberty at this juncture to submit elsewhere.

Deputy Editor: Wang, Shanshan Comments to Author:

Dear Authors,

Thank you for your submission. The paper has been reviewed by two experts, with one recommending rejection and the other suggesting major revisions. The key concerns raised are as follows:

- * The baseline navigated method does not appear to work as expected.
- * There is no comparison with other state-of-the-art (SOTA) methods.
- * The results validation is primarily qualitative and subjective.

Given these issues, we are unable to consider the current version of your manuscript further unless you are willing to make substantial revisions to address these concerns. Please be aware that, should you choose to resubmit, the revised manuscript will typically be sent to the same reviewers.

Thank you again for your submission, and we appreciate your understanding.

We sincerely thank both editors for the constructive comments and resubmission invitation. We have thoroughly addressed the raised critiques in this revision:

- ★ The baseline navigated method did not show convincing results because the previous volunteer had abrupt and involuntary movements governed by underlying disease. We identified this with shot-to-shot phase maps in the response letter. We update the manuscript with Figures 5 and 6, which compare the navigated and selfgated methods with two reconstruction methods: MUSE and ADMM unrolling.
- ★ We choose to compare our proposed ADMM unrolling method with MUSE and LLR, because the former has been clinically adopted while the latter is widely used for multi-contrast compressed sensing reconstruciton. We did not compare to other self-supervised learning methods like SSDU, because SSDU still requires extra datasets for training. Both MUSE and LLR are competitive and representative methods to be compared with.
- ★ We have provided quantitative analysis in Figures 3 and 4, as well as DTI model fitted colored FA maps in Figure 8. The ablation study in Figure 3 was conducted with in-plane fully-sampled data, which we used as reference.

Please note that because this is a resubmission and the manuscript is treated as new, we don't provide the marked version.

Please also note that the goal of this work is to develop a selfgated self-supervised ADMM un-rolling reconstruction technique for high-resolution DWI. Collecting large-scale reference datasets to benchmark a wide spectrum of competing methods is beyond the scope of this work.

Reviewer 1

This manuscript presents a deep learning-based reconstruction method for high-resolution multi-shot DWI. Specifically, the authors employ ADMM unrolling with a network-learned regularization to enable multi-shot DWI reconstruction while mitigating artifacts caused by shot-to-shot phase variations. They report achieving 0.7 mm isotropic resolution DWI on a 7T scanner with good image quality. While the proposed approach is promising, I have several major concerns regarding the presentation and validation of the method.

- (1) *The current review of reconstruction methods in the Introduction feels somewhat mixed. Methods for motion-phase correction (e.g., MUSSELS) and k-q joint reconstruction (e.g., DAE) are discussed together, which might obscure their distinct purposes. I suggest that the authors introduce these methods separately to help readers better understand the core objectives of each category.*

Thank you for the suggestion. We have separated these methods in Introduction.

- (2) *Regarding the forward model A used in the proposed ADMM unrolling method (Eq. 2 and Algorithm 1), what was specifically used as the motion phase map Φ ? Was it derived from MUSE? The authors note: "However, it (MUSE) requires small undersampling factors per shot and fully sampled DWI acquisition assembling all shots. Alternatively, undersampled DWI acquisition can be enabled via the acquisition of navigators for shot phase estimation." This gives the impression that navigator-based phase correction should outperform MUSE. However, the results in Fig. 4 suggest otherwise, which seems contradictory.*

Thank you for the constructive comments. The motion phase map Φ can be estimated from either the imaging echo or the navigator echo, and is derived as following: (1) perform SMS-SENSE reconstruction on every shot, (2) extract and smooth phase from the reconstructed images. Yes, these are part of MUSE reconstruction steps. We do expect navigator-based phase correction outperforms MUSE, i.e., self-gated, as the navigator is less undersampled than the imaging echo.

- (3) *The navigator-based phase correction shown in Fig. 4 appears overly artifacted. The authors attribute this to increased scan time leading to greater sensitivity to inter-shot motion: "The main reason for these artifacts is that the acquisition of navigators increases the total scan time, resulting in higher sensitivity to accidental inter-shot motion." However, I find this explanation unconvincing. Since this is a retrospective experiment, the effective scan time for both navigated and self-gated acquisitions should be identical, meaning that the subject would experience the same motion in both cases. The inclusion of a navigator introduces a second echo, but since motion primarily affects the phase during diffusion encoding, the phase should remain static after diffusion encoding, allowing accurate phase correction using the navigator. It seems more likely that the artifacts arise from errors in navigator implementation. To clarify this, I recommend that the authors present the navigator phase data*

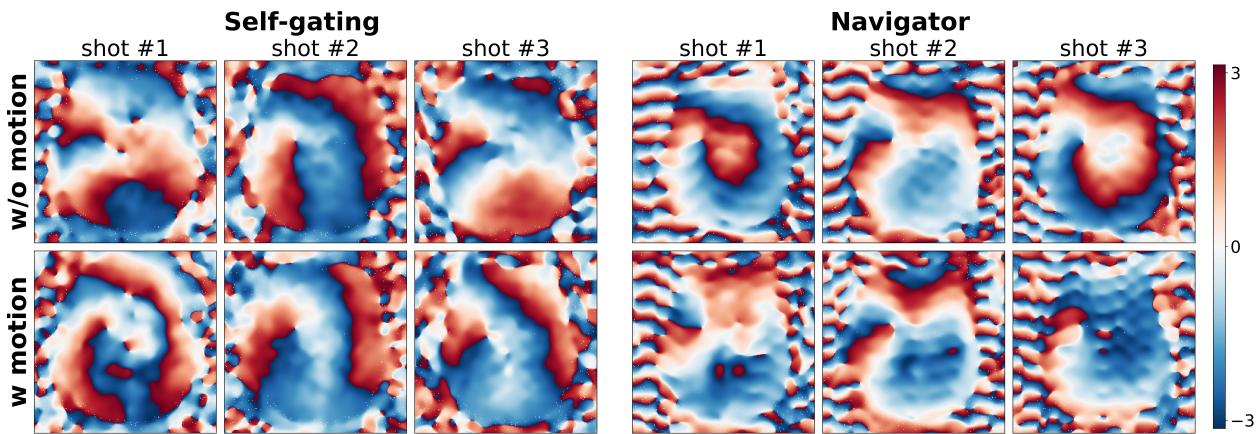


Figure 1: (Left Panel) Shot phases from self-gated reconstruction, i.e., reconstruct and extract phases from each shot of imaging echoes. (Right Panel) Shot phases from navigated reconstruction. (Top Row) A diffusion direction that does not show inter-shot motion. (Bottom Row) A diffusion direction that shows motion in the navigator from the 3rd shot, which appears as rapid phase jumps in the frontal brain region.

and compare it with the MUSE-estimated phase.

Thank you for the valuable insights. Here, Figure 1 shows that the navigator phase from the 3rd shot was indeed corrupted. After further review, we discovered that the volunteer presented here exhibited a very specific motion pattern, which was previously unknown to us. This motion is caused by an underlying disease that leads to abrupt and involuntary movements, which occur over extremely short durations. Consequently, the effect is only visible in isolated diffusion encoding where the navigator scans happen to be corrupted, while the rest show intact navigators and clean DWIs, as shown in Figure 2.

Based on these findings, we have decided to exclude this volunteer's data from our manuscript. We truly appreciate you bringing this to our attention. It led to an important investigation that we otherwise might have missed.

- (4) *Due to the failure of the navigator-based method, a reliable reference for comparison is lacking. Consequently, many of the presented results rely on subjective, qualitative assessments. This may weaken the overall impact of the manuscript. To strengthen the validation, I recommend acquiring a high-quality reference dataset and/or conducting simulations with known ground truth, which would enable more objective and quantitative evaluation.*

Thank you for the constructive suggestion. We provide a full-sampled reference data (listed as Protocol #1 in Table 1) and present the results in Figure 3 in the manuscript.

- (5) *The authors present only raw DWIs, which makes it difficult to evaluate the overall quality of the acquired diffusion datasets, particularly in the absence of a reference. I suggest performing additional diffusion analyses, such as DTI, to provide more convincing evidence of the proposed method's advantages over existing approaches (e.g., LLR).*

Thank you for the constructive suggestion. We perform DTI analysis with the 0.7 mm data

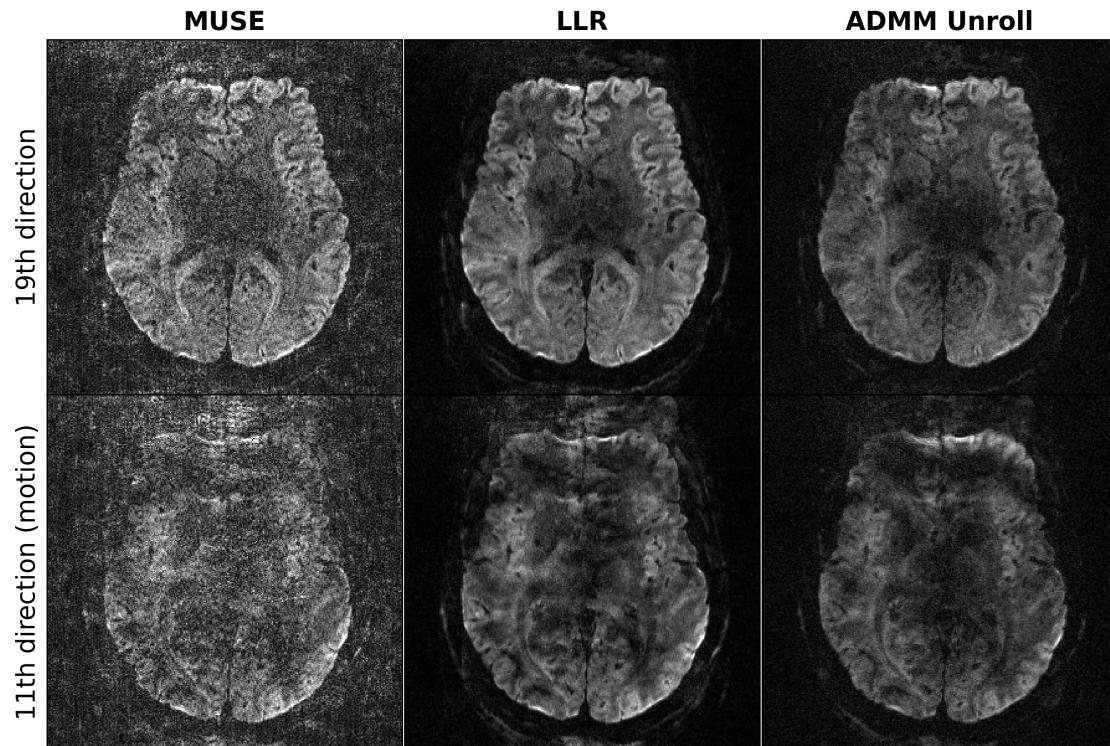
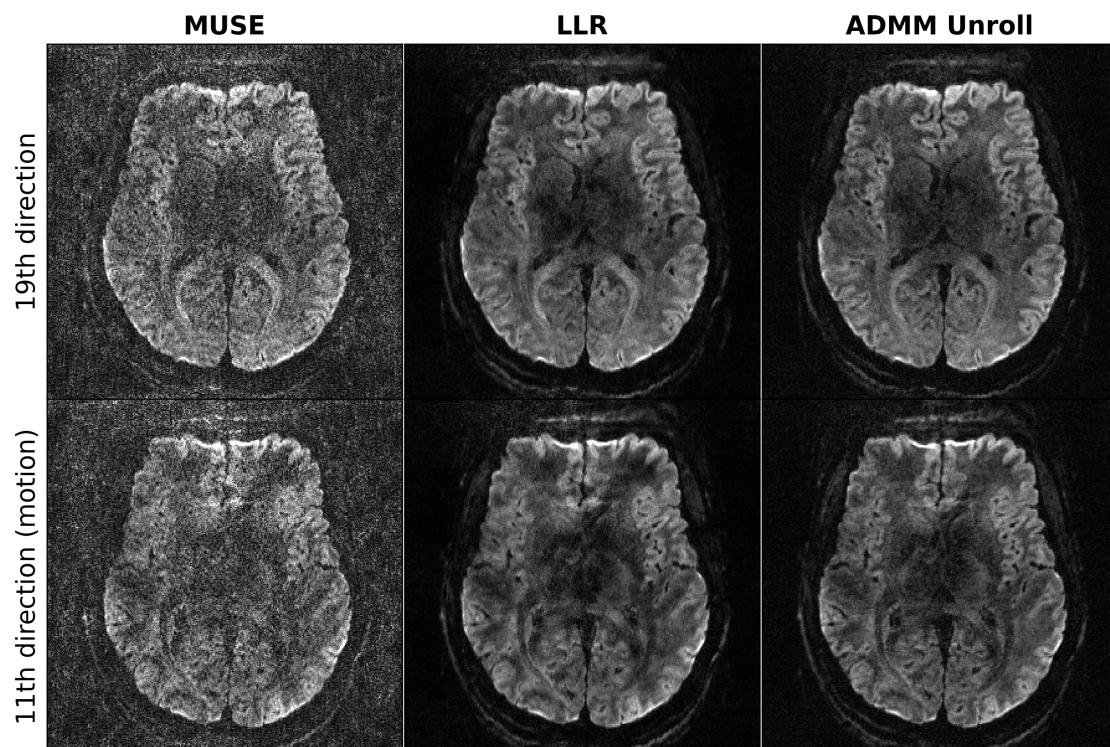
(A) Navigated reconstruction**(B) Selfgated reconstruction**

Figure 2: (A) Navigated and (B) selfgated reconstructions: MUSE, LLR, and ADMM Unroll from left to right. For both navigated and selfgated reconstructions, two diffusion directions are displayed. The first one (the 19th direction) shows clear DWI and no motion corruption, whereas the second (the 11th direction) shows artifacts in all reconstruction methods as the navigator data is corrupted by motion. The use of only imaging data, i.e., selfgated DWI reconstruction, shows reduced motion artifacts, as evident especially in the proposed ADMM Unroll reconstruction. Another advantage of selfgated reconstruction is the reduction in scan time.

and present the results in Figure 8 in the manuscript.

- (6) *While the authors compare slice-by-slice and single-slice training in Fig. 3, this comparison remains largely qualitative. The single-slice training result appears to exhibit diminished diffusion contrast, which was not reflected in the signal plot—likely because the selected voxel did not fall within a region of strong diffusion contrast. A more compelling evaluation could be plotting the signal variance across diffusion encoding directions to assess whether single-slice training introduces angular smoothing.*

Thank you for the valuable comment. Figure 4 in the manuscript presents both the difference and the signal variance in two regions of interest (mean and standard deviation) across diffusion directions between the two training strategies.

Reviewer 2

This paper proposes a self-supervised unrolled reconstruction algorithm for high-resolution and motion-robust diffusion-weighted imaging. Experimental results on a clinical 7T scanner demonstrate the effectiveness of the proposed method. Some issues remain to be addressed:

Method:

- (1) *The authors partitioned the sampling masks into 12 repetitions. Are the masks consistent across these repetitions? Each mask is subdivided into three disjoint subsets. What advantages does this approach offer compared to partitioning the masks into only two disjoint subsets?*

Thank you for the questions.

The masks are not consistent across repetitions. This is also reflected in the code: https://github.com/ZhengguoTan/DeepDWI/blob/main/examples/run_zssl.py#L45. For each repetition, the function ‘uniform_samp’ uses ‘torch.multinomial’ to sample a new mask (<https://github.com/ZhengguoTan/DeepDWI/blob/main/src/deepdwi/recons/zssl.py#L34>).

The zero-shot learning approach as is done in this work requires three disjoint subsets for training, training loss and validation loss (refer to Figure 2). In this way, a model can be trained and tested on a single dataset, without the need of large-scale datasets. In the case of only two disjoint subsets, you will need extra data for validation, as is done in SSDU.

- (2) *Dividing the sampling masks into three disjoint subsets reduces the amount of original effective data input for each part of the network. Does this approach allow the network to learn the overall data distribution effectively? Have the authors considered this potential limitation?*

Thank you for the insightful comment. As shown in Figure 9 in the manuscript, we train the model with 100 epoches and each epoch consists of 12 repetitions, both training and validation loss show convergent behavior. This approach allows the network to learn the overall data distribution.

- (3) *The authors proposed two distinct training strategies. I want to know whether the model was trained on a subset of data from a single volunteer and tested on another subset from the same volunteer.*

Yes, the model was trained on a subset of data (i.e., one multi-band slice k -space data) from a single volunteer and tested on another subset (i.e., the remaining multi-band slices) from the same volunteer.

- (4) *I would appreciate it if the authors could clarify the aspect in which motion robustness is manifested. Is it achieved through phase shift correction, or does it involve another technique?*

In our observation and from the comparison with other methods such as LLR, the motion robustness is achieved by the unrolled reconstruction utilizing spatial-diffusion convolutions.

Experiments:

- (1) *The experiment utilized data from only three volunteers, representing a very small sample size.*

Thank you for the comment. We put this as a limitation in the Discussion.

- (2) *Would it be possible to include deep learning-based self-supervised methods, such as SSDU, and multi-mask SSDU, as additional comparison methods?*

Thank you for the suggestion. As mentioned above, SSDU requires several datasets for training. Given the small sample size in this study, we are not able to compare our approach with SSDU type methods. However, we compare our approach with MUSE and LLR, which we believe are competitive and representative methods.

- (3) *In Figure 4, the authors assert that the unrolled ADMM offers an advantage in preserving clear tissue boundaries; however, the region marked by the red arrows in the figure is not clearly defined.*

Thank you for the insightful comment, which is inline with (3) from Reviewer 1. We provide detailed explanation above and update the manuscript with more compelling figures.

- (4) *The experiments in this paper are conducted using data from a single volunteer. Did the authors attempt to train on one part of the volunteer's data and test on another part from the same volunteer?*

Thank you for the questions. We clarify in the manuscript that this work recruited three volunteers. We tested both slice-by-slice training and single-slice training, both of which illustrate quantitatively similar results (refer to Figure 4). Yes, the single-slice training strategy is done by training on one part of the volunteer's data and testing on another part from the same volunteer.

- (5) *In Figure 7, could the authors clarify why the validation loss is lower than the training loss in the early stages of training?*

Thank you for the insightful comment. We explain in the Results section that "this is because more data is split into the training mask than the validation mask". In the function 'uniform_samp', the parameter 'rho' is used to control the split ratio. Please refer to <https://github.com/ZhengguoTan/DeepDWI/blob/main/src/deepdwi/recons/zssl.py#L14>.