



Research Article

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

Submission ID 1c7e01f3-c972-444a-927a-1d56301f6f8b

Submission Version Revision 1

PDF Generation 17 Nov 2025 10:02:58 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
MRM-AMA-clean.pdf	Supplementary Material for Review	6.0 MB	Page 17
supp.pdf	Supplementary Material for Review	2.0 MB	Page 30
responses.pdf	Author Response	85.7 KB	Page 36

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 **self-gated** self-supervised DeepDWI learning for multi-shot echo planar imaging with diffusion 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: 3900

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.

01 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}.

02 In the area of parallel imaging and compressed sensing, low rankness has been powerful prior knowledge for regularized iterative reconstruction. Mani et al.¹⁴ developed Multi-Shot Sensitivity Encoded Diffusion Data Recovery using Structured Low Rank Matrix Completion (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 magnitude-based spatial-angular locally low-rank regularization (SPA-LLR), which employs locally-low rank (LLR) regularization¹⁸ in joint k - q -space reconstruction. Tan et al.¹⁹ extended LLR to accelerated multi-band multi-shot multi-shell reconstruction. Dong et al.²⁰ proposed to better enforce the spatial-diffusion matrix low rankness by stacking shuffled patches together. However, low rank type iterative methods require long reconstruction time, due to the computation of singular value decomposition on many local patches.

03 Recently, algorithm unrolling emerges as an interpretable and efficient deep learning technique for signal and image processing²¹. In the context of iterative image reconstruction, algorithm unrolling naturally inherits domain knowledge, i.e., the physics-based forward modelling and the data consistency term. Therefore, algorithm unrolling does not need to learn that domain knowledge from intensive training data. Furthermore, algorithm unrolling can replace the regularization term in compressed sensing with deep neural networks, acting as an implicit regularizer in inverse problems. Traditional iterative image reconstruction represents the regularization term as one linear transformation followed by one nonlinear thresholding function. For instance, the low rankness regularization is done by singular value soft thresholding (nonlinear) on Casorati matrices. In contrast, deep neural networks are constructed by multiple layers comprising distinct linear transformations and nonlinear activation functions. Therefore, algorithm unrolling offers high representation power. There exist pioneering works that leverage algorithm unrolling for accelerated parallel imaging reconstruction^{22,23,24,25}.

04 Algorithm unrolling has been introduced to DWI image reconstruction. Mani et al.²⁶ proposed to learn a denoising autoencoder (DAE) model²⁷ from a dictionary simulated from the ball-and-stick model. 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.

05 Although promising, this method is specific to diffusion tensor models and may not generalize well to DWI image reconstruction. Alternatively, self-supervised learning that requires neither large-scale dictionaries nor high-quality fully-sampled reference data has been explored. Based on the zero-shot learning concept²⁸, Cho et al.²⁹ proposed to learn a k -space regularization function for multi-shot DWI reconstruction. 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. Further, its effectiveness for high-resolution DWI is yet to be explored.

06 To achieve submillimeter isotropic DWI at clinically feasible scan and inference 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. Our approach incorporates complementary k - q -space sampling and jointly reconstructs multiple slices and all diffusion-weighted images. Moreover, building upon the foundation of noise2noise^{30,31,28}, we train the ADMM unrolling model using a single multi-band slice, thereby enabling self-gated joint reconstruction with significantly reduced training time, and without the need for large-scale dictionaries or extra training data. 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

07 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

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.

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.

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 work¹⁹ 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

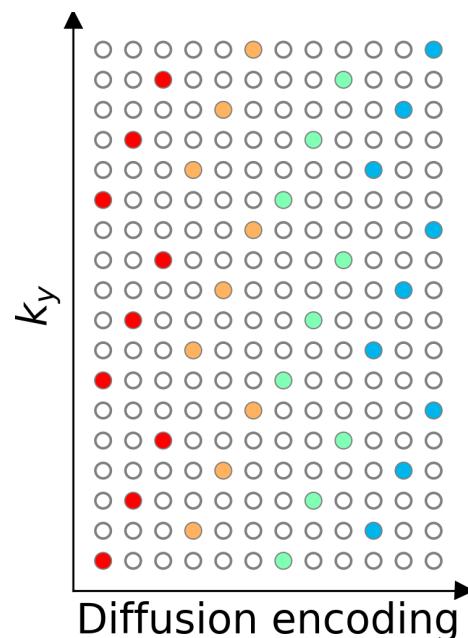


FIGURE 1 Three-shot DWI with diffusion shift encoding. This work employs three-shot per diffusion encoding and each shot has an in-plane undersampling factor of 6. Every three columns assemble one diffusion encoding and thus are colored the same. The starting k_y line is shifted between adjacent diffusion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

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^{4,32,5,33} 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). 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) regularization¹⁹ 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 MoDL²⁵, we employed the ADMM unrolling to solve the self-supervised learning reconstruction in Equation (2). The update rule of ADMM unrolling reads

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

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

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

```

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 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^{31,28} for scan-specific parallel imaging reconstruction. In contrast, Self-supervised learning via data undersampling (SSDU)³⁵ splits the sampling mask

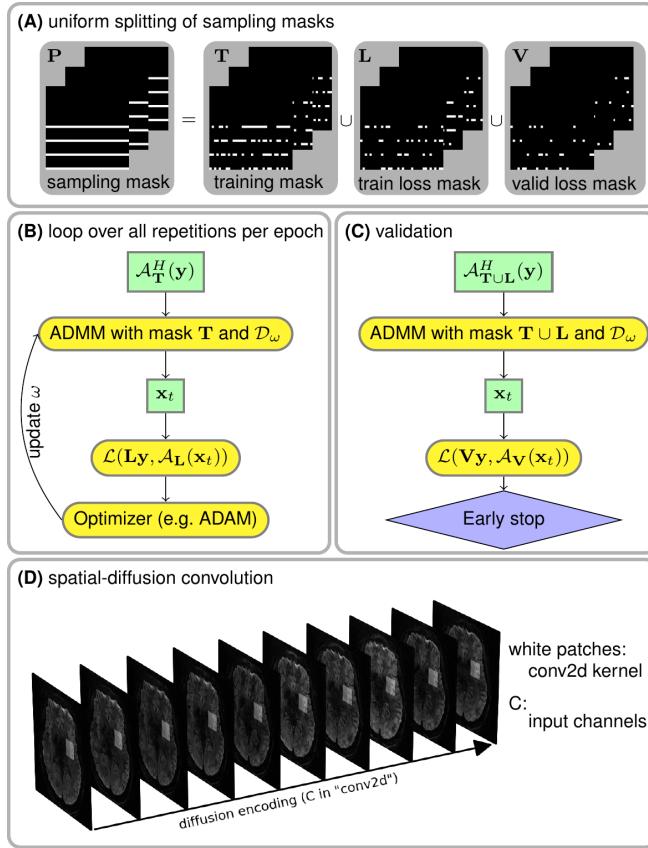


FIGURE 2 Illustration of the key components in ADMM unrolling. **(A)** The sampling mask P in Equation (1) was uniformly split into three disjoint sets: the training mask \mathbf{T} used for the data consistency term during training, the train loss mask \mathbf{L} used for the loss function calculation during training, and the validation loss mask \mathbf{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 ADAM³⁴ during training, but remain fixed during the validation step. **(D)** A stack of diffusion-weighted images is input into ResNet during ADMM unrolling.

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 ResNet³⁶ parameterized by ω (refer to Figure S1 in Supplementary Information). 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. 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 | Retrospective 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.

In addition, an ablation study that replaces the ResNet with Identity is provided in Figure S2 in Supplementary Information. Through this comparison, we observe effective denoising and sharp diffusion-weighted contrast with the ResNet-embedded ADMM unrolling.

3.2 | Model Generalizability

3.2.1 | Cross Slices

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 self-gated 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.

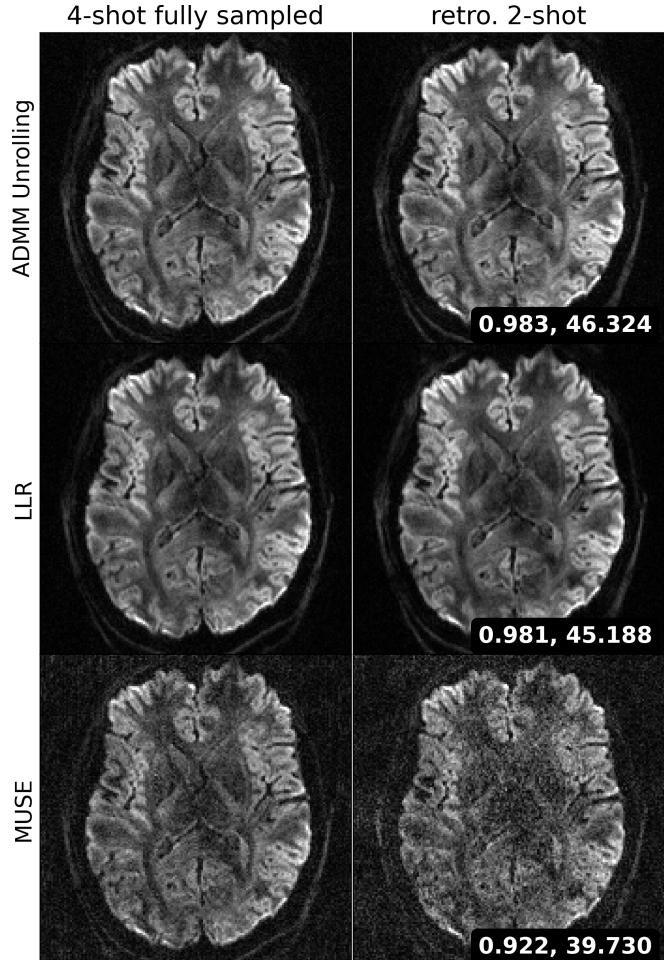


FIGURE 3 Retrospective study with the fully-sampled reference data acquired by Protocol #1. The first column displays one diffusion-weighted image from the 4-shot fully-sampled data reconstructed via (from top to bottom) the proposed self-supervised ADMM unrolling, LLR, and MUSE. The second column displays 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.

3.2.2 | Cross Subjects

We also evaluate the cross-subject generalizability. Figure S3 in Supplementary Information displays the results of a model trained on one subject but inferred on a different subject. The model performs well and is applicable to different subjects with the same acquisition parameters.

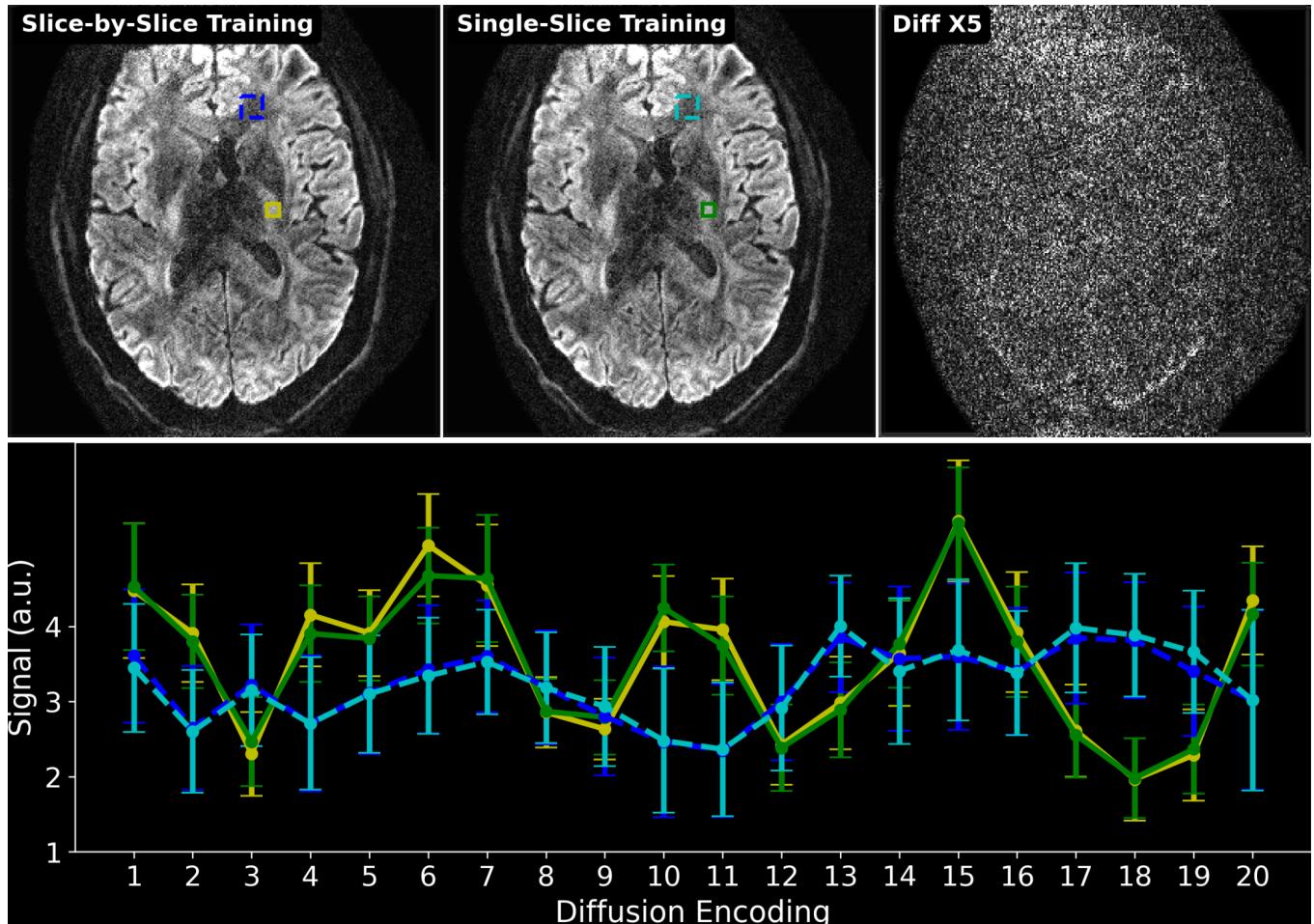


FIGURE 4 Comparison of two training strategies: (1) slice-by-slice training, where every slice is trained and tested individually; (2) single-slice training, where the unrolled ADMM model is trained on only one slice and tested on all remaining slices. The top-right image shows the absolute difference 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.

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 diffusion-weighted images. On the contrary, self-gated

reconstructions show only subtle phase, because shot-to-shot phase variations were estimated from the first echoes themselves. The reduced phase variation in the self-gated 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 denoiser³⁷, the self-gated ADMM unrolling reconstruction achieves strong denoising and resolves clear tissue details.

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

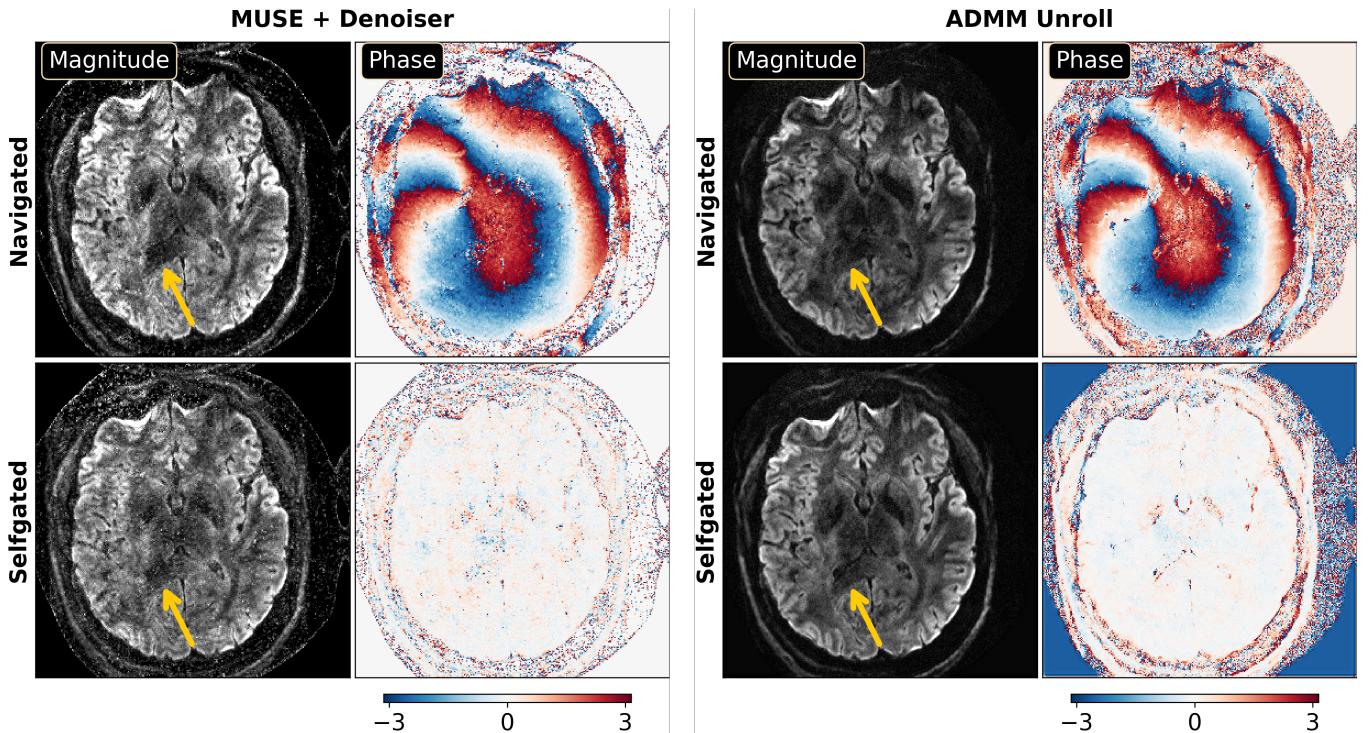


FIGURE 5 Validation of the proposed self-gated ADMM unrolling reconstruction with data acquired by Protocol #2 in Table 1. Both MUSE and ADMM unrolling were performed with navigated and self-gated shot-to-shot phase maps, respectively. Compared to MUSE, the ADMM unrolled reconstruction excels in denoising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

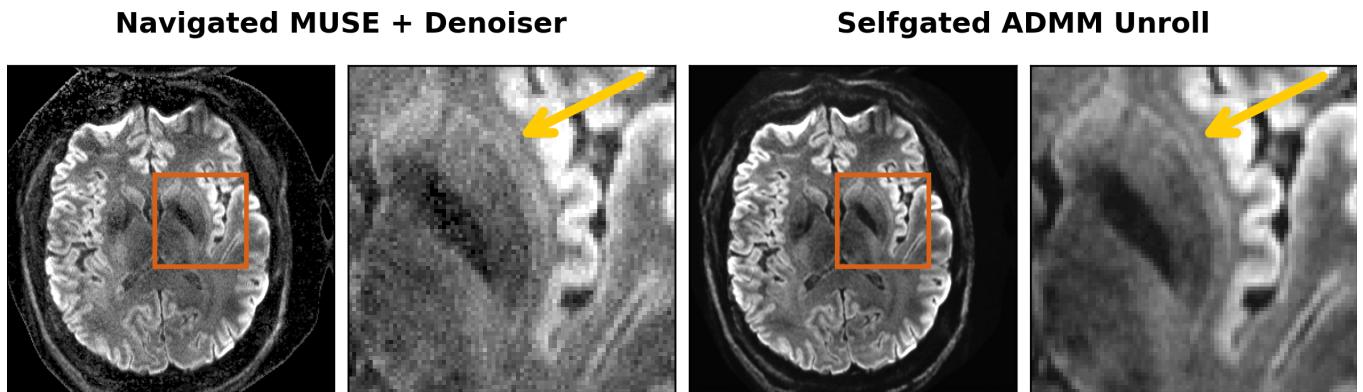


FIGURE 6 0.7 mm isotropic resolution DWI with the proposed self-gated ADMM unrolling enables the visualization of the tiny structure claustrum, whereas the MUSE reconstruction shows only blurred appearance. Displayed images are the mean diffusion-weighted image from 20 directions and its zoomed-in region.

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 artifacts³⁸, 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.

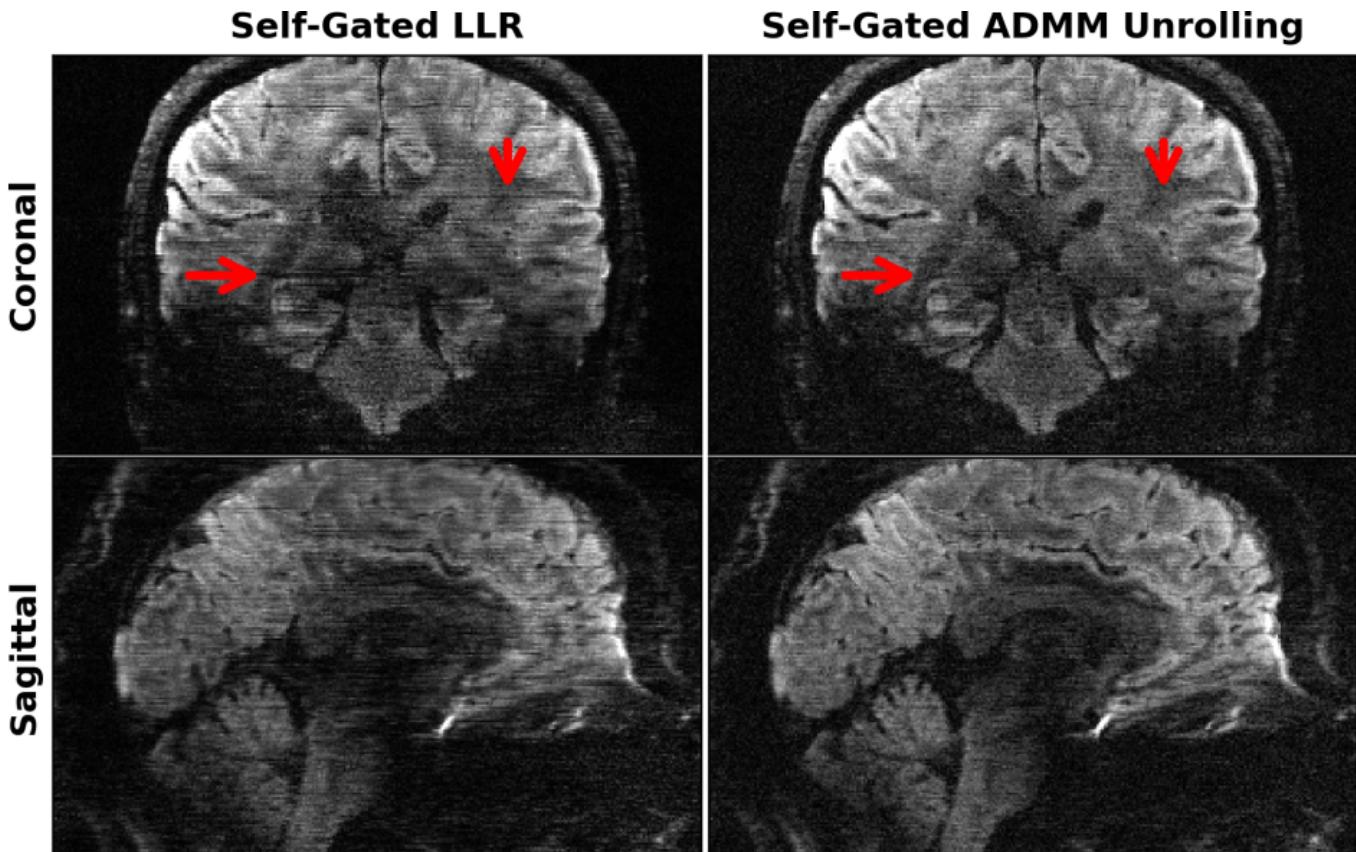


FIGURE 7 Single-direction diffusion-weighted images at 0.7 mm isotropic resolution as reconstructed by retrospectively self-gated (left) LLR and (right) ADMM unrolling in (top) the coronal and (bottom) the sagittal views, respectively. The same diffusion direction as in Figure 5 is chosen for display. 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).

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 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-gated 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 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 1000 s/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

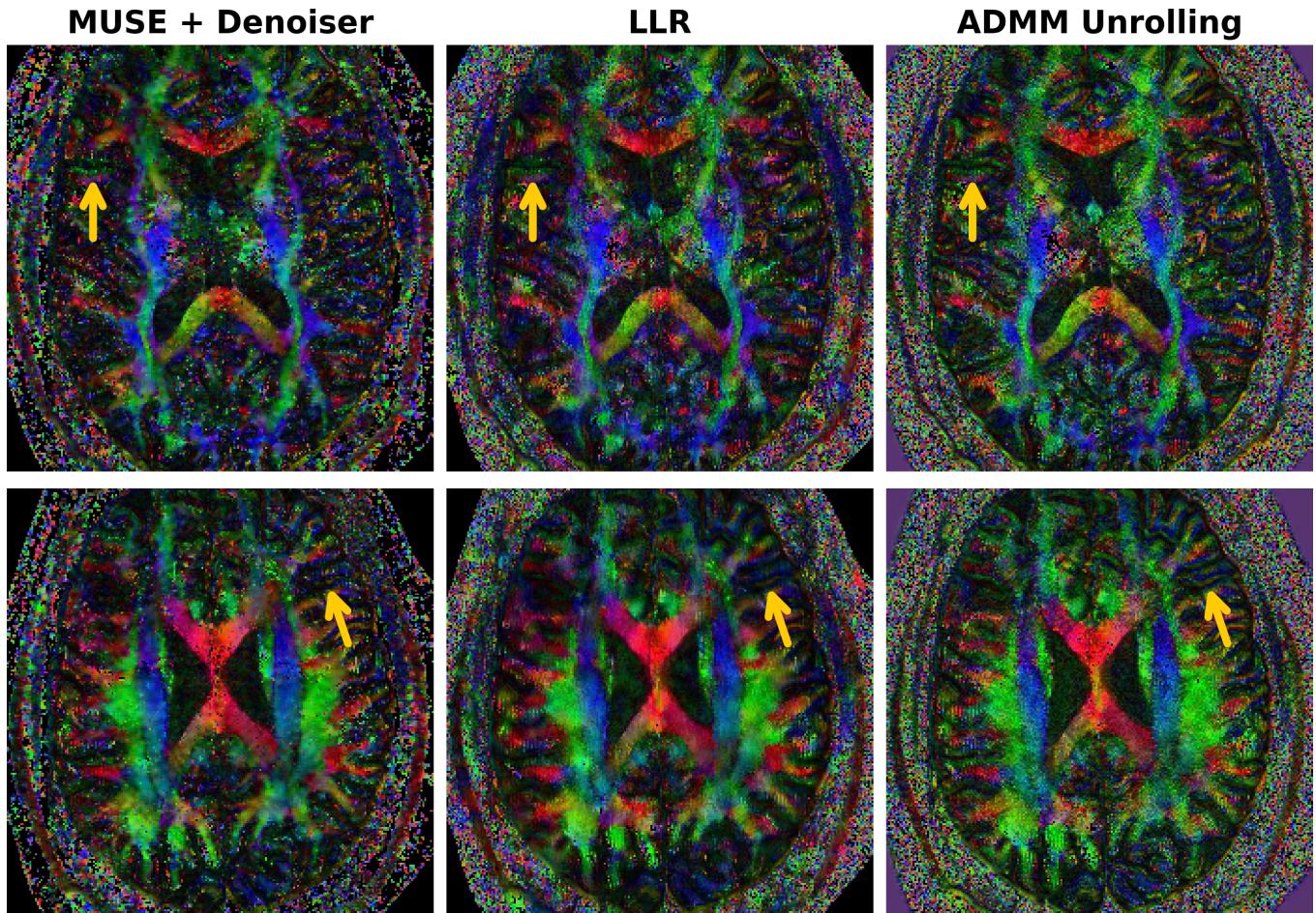


FIGURE 8 Diffusion tensor imaging (DTI) model derived colored fractional anisotropy (cFA) maps based on the diffusion-weighted images as reconstructed by MUSE, LLR, and ADMM unrolling, respectively. Two slices are displayed in the top and the bottom row, respectively. Even with limited scan time (about 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.

for single image recovery³¹ and parallel imaging²⁸, our approach trains an unrolled reconstruction network with only one dataset utilizing the concept of data splitting^{31,35,28}. 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 (refer to Table S1 in Supplementary Information), 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. 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

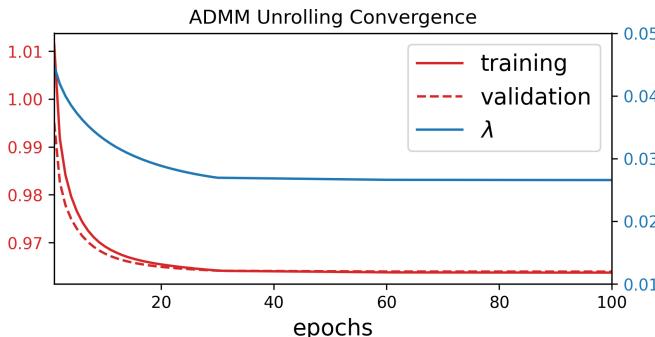


FIGURE 9 Convergence analysis along the ADMM unrolling training and validation epochs. Displayed curves are (red solid) the training loss, (red dashed) the validation loss, and (blue solid) the learned regularization strength λ , respectively. All parameters converge sufficiently and show no over-fitting.

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 encoding³⁹. 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 echoes⁴⁰. This approach does not require the 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 self-gated 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 SSDU³⁵. 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.

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. D. K. Jones. *Diffusion MRI: Theory, methods, and applications*. Oxford University Press; 2010.
2. P. Mansfield. Multi-planar image formation using NMR spin echoes. *J Phys C*. 1977;10:55-58.
3. K. Butts, S. J. Riederer, R. L. Ehman, R. M. Thompson, C. R. Jack. Interleaved echo planar imaging on a standard MRI system. *Magn. Reson. Med.*. 1993;31:67-72.
4. C. Liu, R. Bammer, D.-h. Kim, M. E. Moseley. Self-navigated interleaved spiral (SNAILS): Application to high-resolution diffusion tensor imaging. *Magn. Reson. Med.*. 2004;52:1388-1396.
5. N.-K. Chen, A. Guidon, H.-C. Chang, A. W. Song. A robust multi-shot scan strategy for high-resolution diffusion weighted MRI enabled by multiplexed sensitivity-encoding (MUSE). *NeuroImage*. 2013;72:41-47.
6. D. A. Porter, R. M. Heidemann. High resolution diffusion-weighted imaging using readout-segmented echo-planar imaging, parallel imaging and a two-dimensional navigator-based reacquisition. *Magn. Reson. Med.*. 2009;62:468-475.
7. R. M. Heidemann, D. A. Porter, A. Anwander, et al. Diffusion imaging in humans at 7 T using readout-segmented EPI and GRAPPA. *Magn. Reson. Med.*. 2010;64:9-14.
8. J. L. R. Andersson, S. Skare, J. Ashburner. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *NeuroImage*. 2003;20:870-888.
9. K. Setsompop, B. A. Gagoski, J. R. Polimeni, T. Witzel, V. J. Wedeen, L. L. Wald. Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced *g*-factor penalty. *Magn. Reson. Med.*. 2012;67:1210-1224.
10. K. P. Pruessmann, M. Weiger, M. B. Scheidegger, P. Boesiger. SENSE: Sensitivity encoding for fast MRI. *Magn. Reson. Med.*. 1999;42:952-962.
11. M. A. Griswold, P. M. Jakob, R. M. Heidemann, et al. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magn. Reson. Med.*. 2002;47:1202-1210.
12. R. Bammer, S. L. Keeling, M. Augustin, et al. Improved diffusion-weighted single-shot echo-planar imaging (EPI) in stroke using sensitivity encoding (SENSE). *Magn. Reson. Med.*. 2001;46:548-554.
13. M. Lustig, D. Donoho, J. M. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magn. Reson. Med.*. 2007;58:1182-1195.
14. M. Mani, M. Jacob, D. Kelley, V. Magnotta. Multi-shot sensitivity-encoded diffusion data recovery using structured low-rank matrix completion (MUSSELS). *Magn. Reson. Med.*. 2017;78:494-507.
15. Y. Hu, X. Wang, Q. Tian, et al. Multi-shot diffusion-weighted MRI reconstruction with magnitude-based spatial-angular locally low-rank regularization (SPALLR). *Magn. Reson. Med.*. 2020;83:1596-1607.
16. J. Veraart, D. S. Novikov, D. Christiaens, B. Ades-aron, J. Sijbers, E. Fieremans. Denoising of diffusion MRI using random matrix theory. *NeuroImage*. 2016;142:394-406.
17. Q. Tian, Z. Li, Q. Fan, et al. SDnDTI: Self-supervised deep learning-based denoising for diffusion tensor MRI. *NeuroImage*. 2023;253:119033.
18. T. Zhang, J. M. Pauly, I. R. Levesque. Accelerated parameter mapping with a locally low rank constraint. *Magn. Reson. Med.*. 2015;73:655-661.
19. Z. Tan, P. A. Liebig, R. M. Heidemann, F. B. Laun, F. Knoll. Accelerated diffusion-weighted magnetic resonance imaging at 7 T: Joint reconstruction for shift-encoded navigator-based interleaved echo planar imaging (JETS-NAVIEPI). *Imaging Neuroscience*. 2024;2:1-15.
20. Y. Dong, X. Ye, C. Li, M. J. P. van Osch, P. Börnert. Navigator-free multi-shot diffusion MRI via non-local low-rank reconstruction. *Magn. Reson. Med.*. 2025;94:1592-1603.
21. V. Monga, Y. Li, Y. C. Eldar. Algorithm Unrolling: Interpretable, Efficient Deep Learning for Signal and Image Processing. *IEEE Signal Processing Magazine*. 2021;38:18-44.
22. S. Wang, Z. Su, L. Ying, et al. Accelerating magnetic resonance imaging via deep learning. In: *13th IEEE International Symposium on Biomedical Imaging (ISBI'13)*:514-517; 2016.
23. K. Hammerkirk, T. Klatzer, E. Kobler, et al. Learning a variational network for reconstruction of accelerated MRI data. *Magn. Reson. Med.*. 2018;79:3055-3071.
24. Y. Yang, J. Sun, H. Li, Z. Xu. ADMM-CSNet: A Deep Learning Approach for Image Compressive Sensing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2018;42:521-538.
25. H. K. Aggarwal, M. P. Mani, M. Jacob. MoDL: Model-based deep learning architecture for inverse problems. *IEEE Trans. Med. Imaging*. 2018;38:394-405.
26. M. Mani, V. A. Magnotta, M. Jacob. qModeL: A plug-and-play model-based reconstruction for highly accelerated multi-shot diffusion MRI using learned priors. *Magn. Reson. Med.*. 2021;86:835-851.
27. G. E. Hinton, R. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*. 2006;313:504-507.
28. B. Yaman, S. A. H. Hosseini, M. Akçakaya. Zero-shot self-supervised learning for MRI reconstruction. In: *10th International Conference on Learning Representations (ICLR'10)*; 2022.
29. J. Cho, Y. Jun, X. Wang, C. Kobayashi, B. Bilgic. Improved Multi-shot Diffusion-Weighted MRI with Zero-Shot Self-supervised Learning Reconstruction. In: *26th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'26)*:457-466; 2023.
30. J. Lehtinen, J. Munkberg, J. Hasselgren, et al. Noise2Noise: Learning Imaging Restoration without Clean Data. In: *Proceedings of the 35th International Conference on Machine Learning (ICML'35)*:2965-2974; 2018.
31. Y. Quan, M. Chen, T. Pang, H. Ji. Self2Self With Dropout: Learning Self-Supervised Denoising From Single Image. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2020)*:1890-1898; 2020.

- 01 32. M. Uecker, A. Karaus, J. Frahm. Inverse reconstruction
02 method for segmented multishot diffusion-weighted MRI
03 with multiple coils. *Magn. Reson. Med.*. 2009;62:1342-
04 1348. 54
- 05 33. A. Merrem, S. Hofer, A. S. A. Hosseini, et al. Diffusion-
06 weighted MRI of the prostate without susceptibility
07 artifacts: Undersampled multi-shot turbo-STEAM with
08 rotated radial trajectories. *NMR Biomed.*. 2019;32:e4074.
09 55
- 10 34. D. P. Kingma, J. Ba. ADAM: A Method for Stochas-
11 tic Optimization. In: *3rd International Conference on*
12 *Learning Representations (ICLR'3)*; 2015.
13 56
- 14 35. B. Yaman, S. A. H. Hosseini, S. Moeller, J. Eller-
15 mann, K. Uğurbil, M. Akçakaya. Self-supervised learning
16 of physics-guided reconstruction neural networks with-
17 out fully sampled reference data. *Magn. Reson. Med.*.
18 2020;84:3172-3191. 57
- 19 36. K. He, X. Zhang, S. Ren, J. Sun. Deep residual learning
20 for image recognition. In: *IEEE Conference on Com-*
21 *puter Vision and Pattern Recognition (CVPR'16)*:770-
22 778; 2016. 58
- 23 37. L. Cordero-Grande, D. Christiaens, J. Hutter, A. N.
24 Price, J. V. Hajnal. Complex diffusion-weighted image
25 estimation via matrix recovery under general noise mod-
26 els. *NeuroImage*. 2019;200:391-404. 59
- 27 38. M. Pietsch, D. Christiaens, J. V. Hajnal, J.-D. Tournier.
28 dStripe: Slice artefact correction in diffusion MRI via
29 constrained neural network. *Medical Image Analysis*.
30 2021;74:102255. 60
- 31 39. B. Zahneisen, M. Aksoy, J. Maclarens, C. Wuerslin,
32 R. Bammer. Extended hybrid-space SENSE for EPI: Off-
33 resonance and eddy current corrected joint interleaved
34 blip-up/down reconstruction. *NeuroImage*. 2017;153:97-
108. 61
- 35 40. Z. Tan, C. Unterberg-Buchwald, M. Blumenthal, et al.
36 Free-breathing liver fat, R_2^* and B_0 field mapping using
37 multi-echo radial FLASH and regularized model-based
38 reconstruction. *IEEE Transactions on Medical Imaging*.
39 2023;42:1374-1387. 62
- 40 41 42 43 44 45 46 47 48 49 50 51 52 53
-
- 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106

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 self-gated self-supervised DeepDWI learning for multi-shot echo planar imaging with diffusion 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: 3900

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.

01 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}.

02 In the area of parallel imaging and compressed sensing, low rankness has been powerful prior knowledge for regularized iterative reconstruction. Mani et al.¹⁴ developed Multi-Shot Sensitivity Encoded Diffusion Data Recovery using Structured Low Rank Matrix Completion (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 magnitude-based spatial-angular locally low-rank regularization (SPA-LLR), which employs locally-low rank (LLR) regularization¹⁸ in joint k - q -space reconstruction. Tan et al.¹⁹ extended LLR to accelerated multi-band multi-shot multi-shell reconstruction. Dong et al.²⁰ proposed to better enforce the spatial-diffusion matrix low rankness by stacking shuffled patches together. However, low rank type iterative methods require long reconstruction time, due to the computation of singular value decomposition on many local patches.

03 Recently, algorithm unrolling emerges as an interpretable and efficient deep learning technique for signal and image processing²¹. In the context of iterative image reconstruction, algorithm unrolling naturally inherits domain knowledge, i.e., the physics-based forward modelling and the data consistency term. Therefore, algorithm unrolling does not need to learn that domain knowledge from intensive training data. Furthermore, algorithm unrolling can replace the regularization term in compressed sensing with deep neural networks, acting as an implicit regularizer in inverse problems. Traditional iterative image reconstruction represents the regularization term as one linear transformation followed by one nonlinear thresholding function. For instance, the low rankness regularization is done by singular value soft thresholding (nonlinear) on Casorati matrices. In contrast, deep neural networks are constructed by multiple layers comprising distinct linear transformations and nonlinear activation functions. Therefore, algorithm unrolling offers high representation power. There exist pioneering works that leverage algorithm unrolling for accelerated parallel imaging reconstruction^{22,23,24,25}.

04 Algorithm unrolling has been introduced to DWI image reconstruction. Mani et al.²⁶ proposed to learn a denoising autoencoder (DAE) model²⁷ from a dictionary simulated from the ball-and-stick model. 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.

05 Although promising, this method is specific to diffusion tensor models and may not generalize well to DWI image reconstruction. Alternatively, self-supervised learning that requires neither large-scale dictionaries nor high-quality fully-sampled reference data has been explored. Based on the zero-shot learning concept²⁸, Cho et al.²⁹ proposed to learn a k -space regularization function for multi-shot DWI reconstruction. 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. Further, its effectiveness for high-resolution DWI is yet to be explored.

06 To achieve submillimeter isotropic DWI at clinically feasible scan and inference 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. Our approach incorporates complementary k - q -space sampling and jointly reconstructs multiple slices and all diffusion-weighted images. Moreover, building upon the foundation of noise2noise^{30,31,28}, we train the ADMM unrolling model using a single multi-band slice, thereby enabling self-gated joint reconstruction with significantly reduced training time, and without the need for large-scale dictionaries or extra training data. 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

07 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

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.

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.

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 work¹⁹ 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

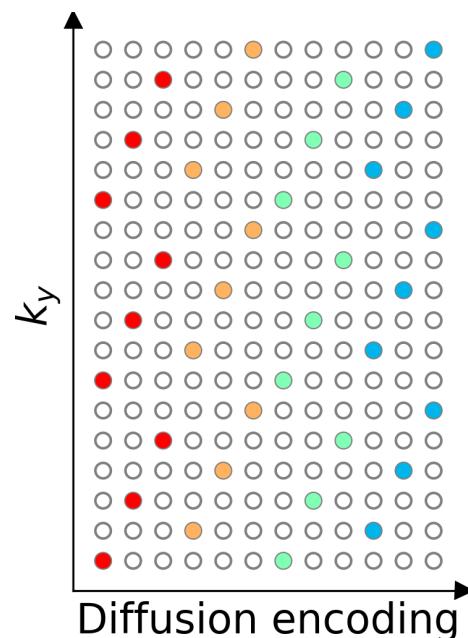


FIGURE 1 Three-shot DWI with diffusion shift encoding. This work employs three-shot per diffusion encoding and each shot has an in-plane undersampling factor of 6. Every three columns assemble one diffusion encoding and thus are colored the same. The starting k_y line is shifted between adjacent diffusion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

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^{4,32,5,33} 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). 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) regularization¹⁹ 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 MoDL²⁵, we employed the ADMM unrolling to solve the self-supervised learning reconstruction in Equation (2). The update rule of ADMM unrolling reads

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

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

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

```

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 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^{31,28} for scan-specific parallel imaging reconstruction. In contrast, Self-supervised learning via data undersampling (SSDU)³⁵ splits the sampling mask

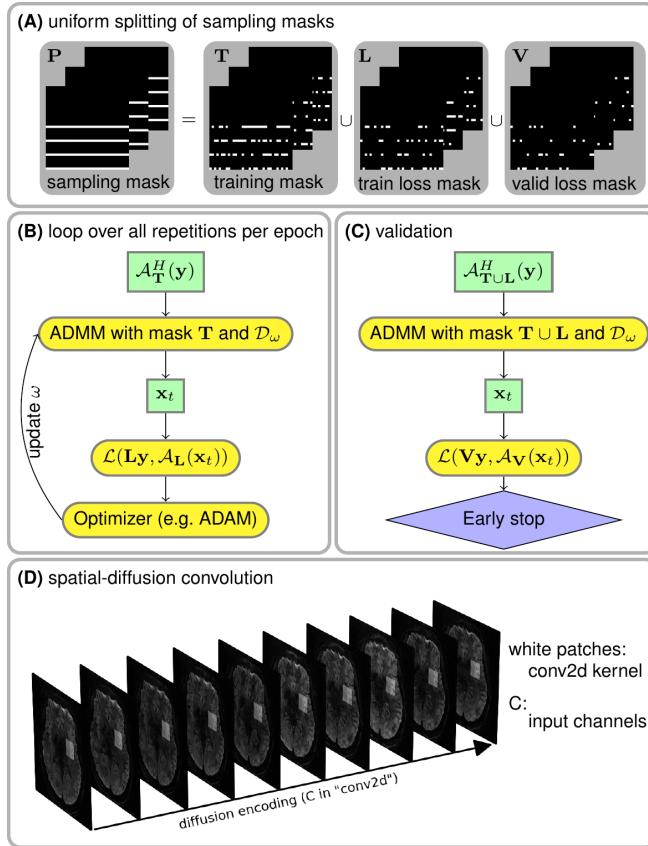


FIGURE 2 Illustration of the key components in ADMM unrolling. **(A)** The sampling mask P in Equation (1) was uniformly split into three disjoint sets: the training mask \mathbf{T} used for the data consistency term during training, the train loss mask \mathbf{L} used for the loss function calculation during training, and the validation loss mask \mathbf{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 ADAM³⁴ during training, but remain fixed during the validation step. **(D)** A stack of diffusion-weighted images is input into ResNet during ADMM unrolling.

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 ResNet³⁶ parameterized by ω (refer to Figure S1 in Supplementary Information). 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. 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 | Retrospective 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.

In addition, an ablation study that replaces the ResNet with Identity is provided in Figure S2 in Supplementary Information. Through this comparison, we observe effective denoising and sharp diffusion-weighted contrast with the ResNet-embedded ADMM unrolling.

3.2 | Model Generalizability

3.2.1 | Cross Slices

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 self-gated 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.

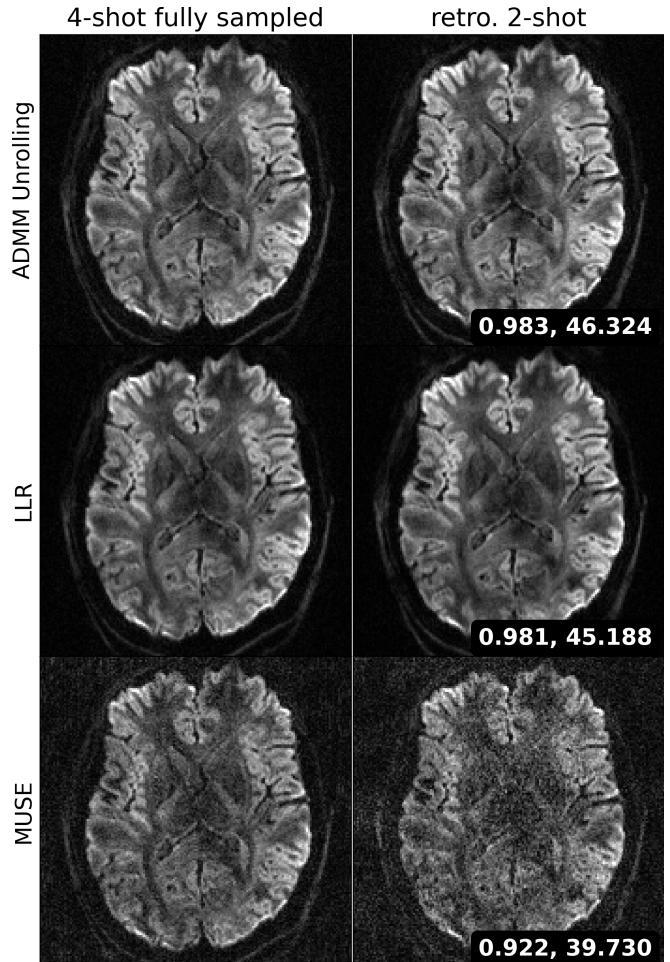


FIGURE 3 Retrospective study with the fully-sampled reference data acquired by Protocol #1. The first column displays one diffusion-weighted image from the 4-shot fully-sampled data reconstructed via (from top to bottom) the proposed self-supervised ADMM unrolling, LLR, and MUSE. The second column displays 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.

3.2.2 | Cross Subjects

We also evaluate the cross-subject generalizability. Figure S3 in Supplementary Information displays the results of a model trained on one subject but inferred on a different subject. The model performs well and is applicable to different subjects with the same acquisition parameters.

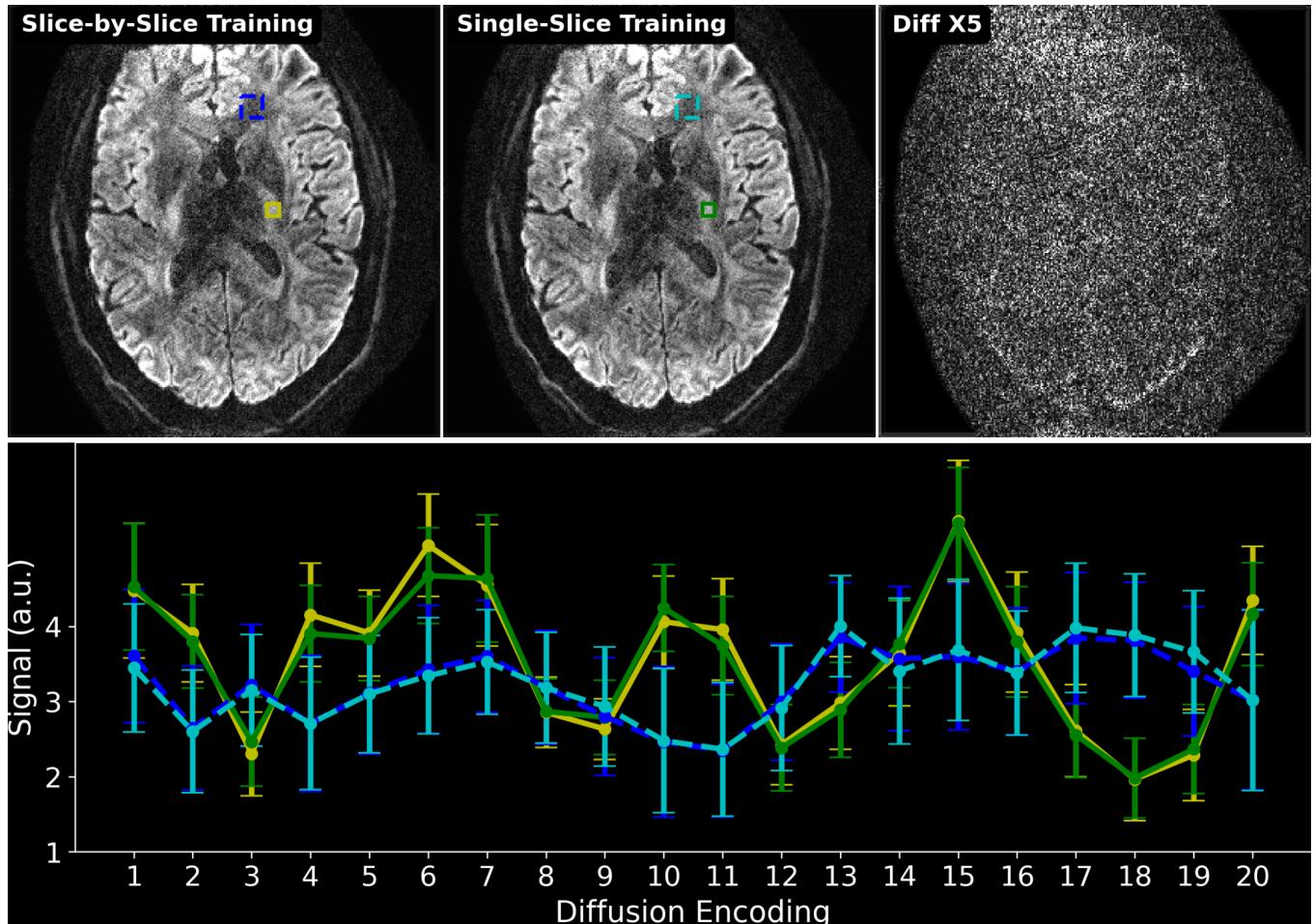


FIGURE 4 Comparison of two training strategies: (1) slice-by-slice training, where every slice is trained and tested individually; (2) single-slice training, where the unrolled ADMM model is trained on only one slice and tested on all remaining slices. The top-right image shows the absolute difference 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.

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 diffusion-weighted images. On the contrary, self-gated

reconstructions show only subtle phase, because shot-to-shot phase variations were estimated from the first echoes themselves. The reduced phase variation in the self-gated 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 denoiser³⁷, the self-gated ADMM unrolling reconstruction achieves strong denoising and resolves clear tissue details.

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

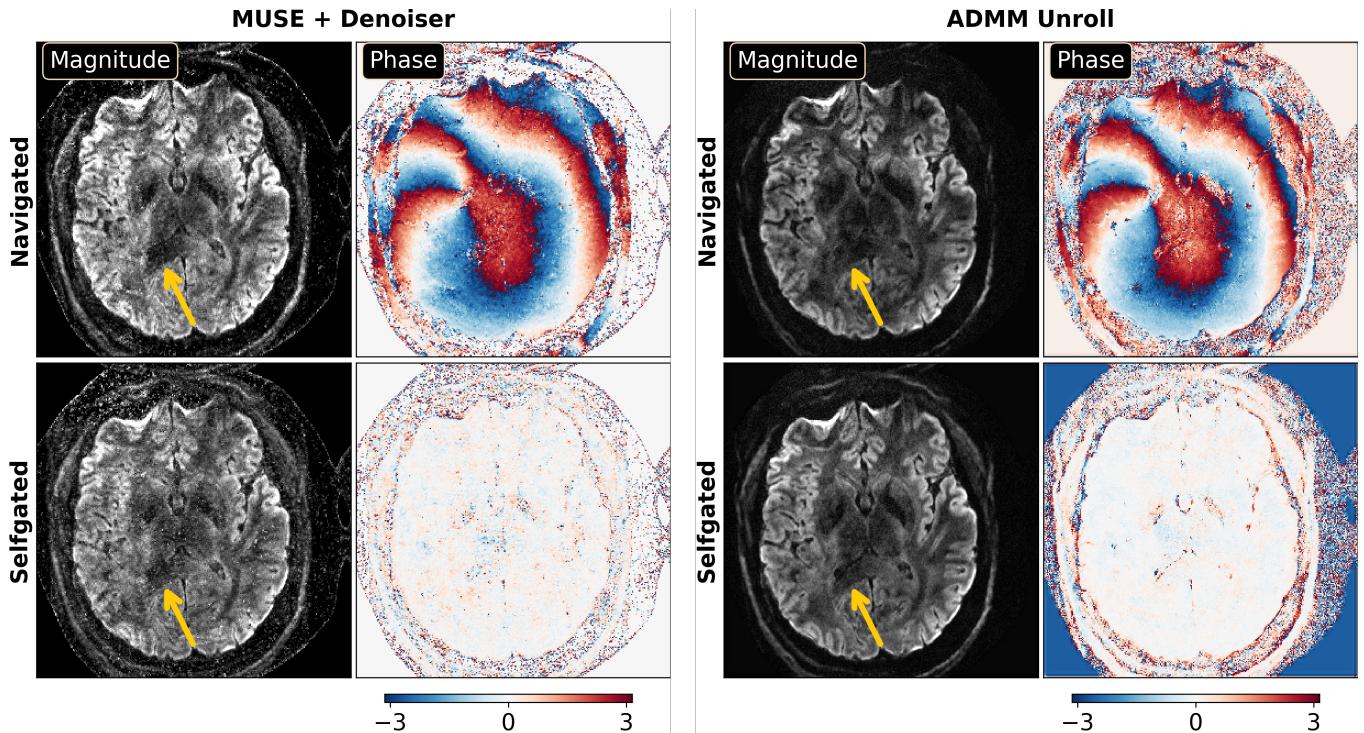
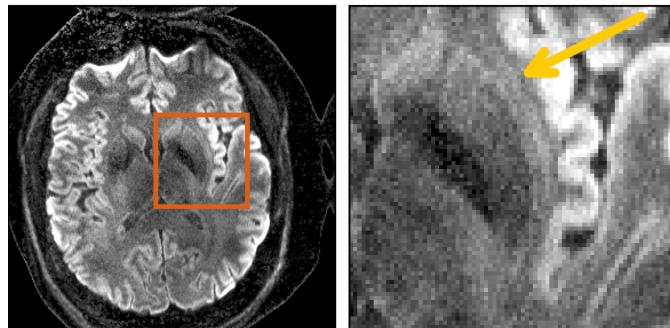


FIGURE 5 Validation of the proposed self-gated ADMM unrolling reconstruction with data acquired by Protocol #2 in Table 1. Both MUSE and ADMM unrolling were performed with navigated and self-gated shot-to-shot phase maps, respectively. Compared to MUSE, the ADMM unrolled reconstruction excels in denoising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

Navigated MUSE + Denoiser



Selfgated ADMM Unroll

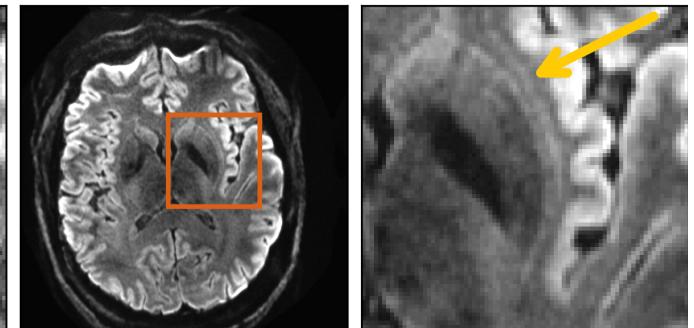


FIGURE 6 0.7 mm isotropic resolution DWI with the proposed self-gated ADMM unrolling enables the visualization of the tiny structure claustrum, whereas the MUSE reconstruction shows only blurred appearance. Displayed images are the mean diffusion-weighted image from 20 directions and its zoomed-in region.

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 artifacts³⁸, 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.

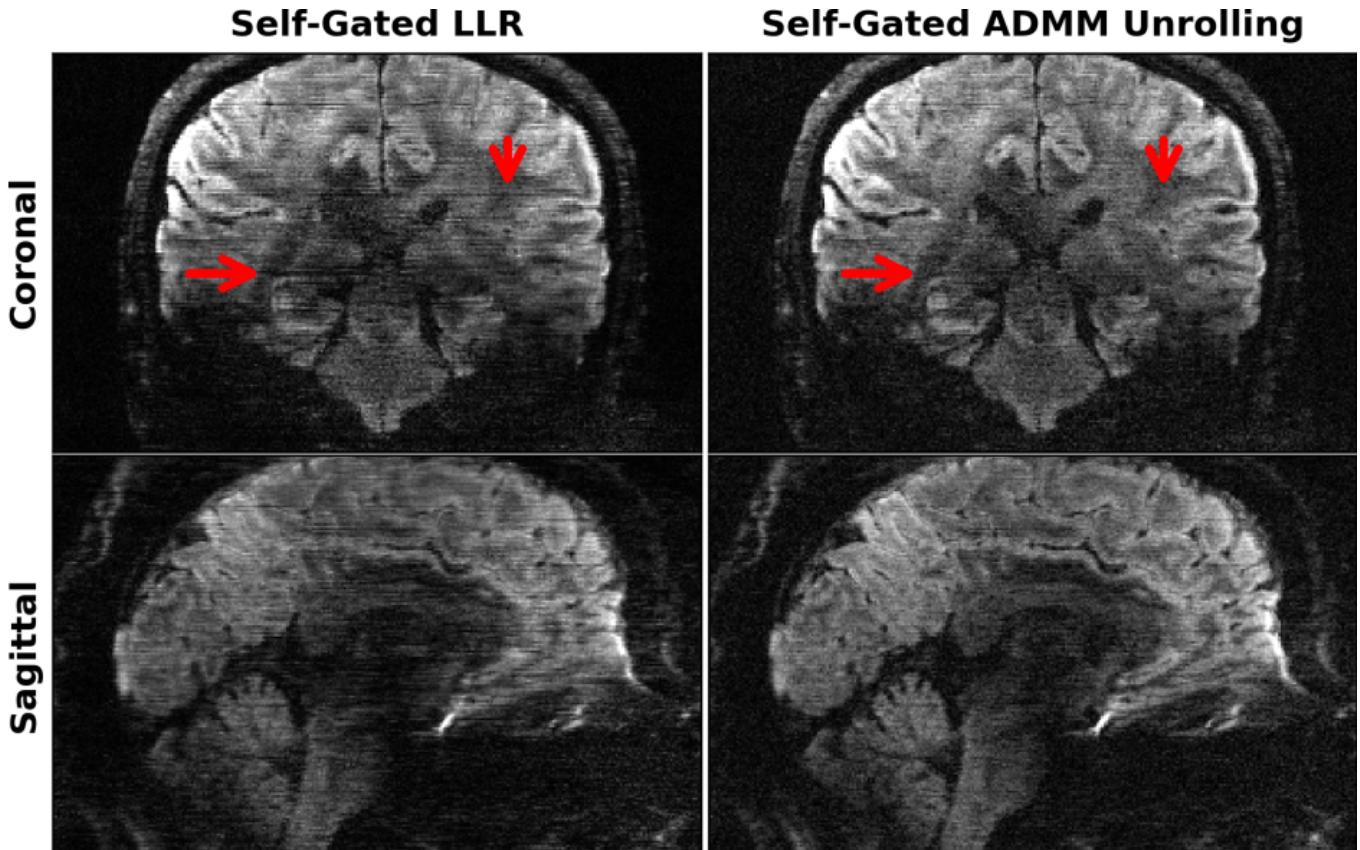


FIGURE 7 Single-direction diffusion-weighted images at 0.7 mm isotropic resolution as reconstructed by retrospectively self-gated (left) LLR and (right) ADMM unrolling in (top) the coronal and (bottom) the sagittal views, respectively. The same diffusion direction as in Figure 5 is chosen for display. 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).

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 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-gated 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 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 1000 s/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

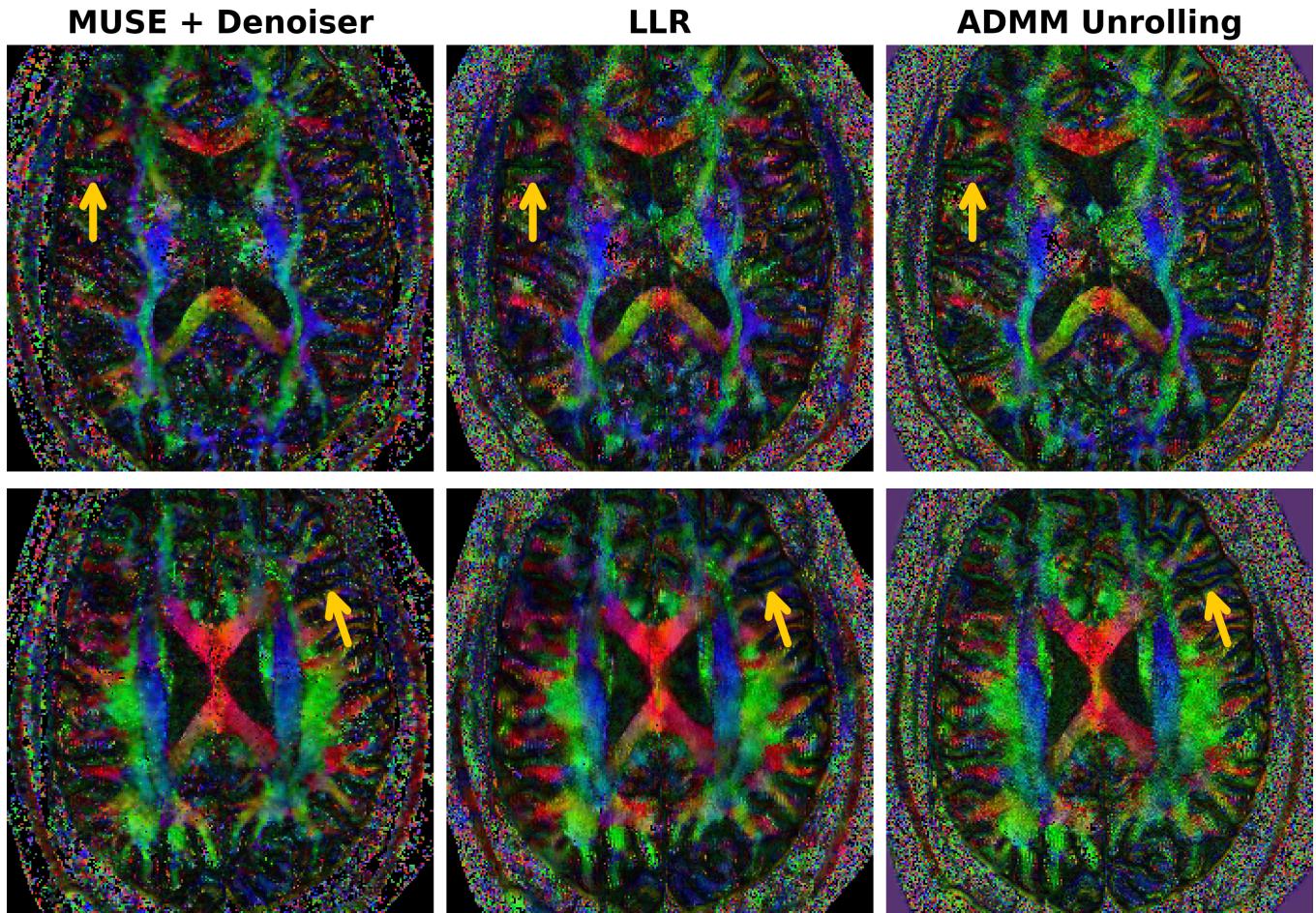


FIGURE 8 Diffusion tensor imaging (DTI) model derived colored fractional anisotropy (cFA) maps based on the diffusion-weighted images as reconstructed by MUSE, LLR, and ADMM unrolling, respectively. Two slices are displayed in the top and the bottom row, respectively. Even with limited scan time (about 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.

for single image recovery³¹ and parallel imaging²⁸, our approach trains an unrolled reconstruction network with only one dataset utilizing the concept of data splitting^{31,35,28}. 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 (refer to Table S1 in Supplementary Information), 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. 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

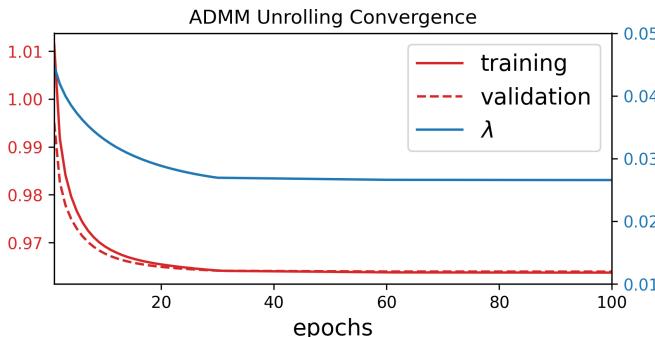


FIGURE 9 Convergence analysis along the ADMM unrolling training and validation epochs. Displayed curves are (red solid) the training loss, (red dashed) the validation loss, and (blue solid) the learned regularization strength λ , respectively. All parameters converge sufficiently and show no over-fitting.

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 encoding³⁹. 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 echoes⁴⁰. This approach does not require the 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 self-gated 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 SSDU³⁵. 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.

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. D. K. Jones. *Diffusion MRI: Theory, methods, and applications*. Oxford University Press; 2010.
2. P. Mansfield. Multi-planar image formation using NMR spin echoes. *J Phys C*. 1977;10:55-58.
3. K. Butts, S. J. Riederer, R. L. Ehman, R. M. Thompson, C. R. Jack. Interleaved echo planar imaging on a standard MRI system. *Magn. Reson. Med.*. 1993;31:67-72.
4. C. Liu, R. Bammer, D.-h. Kim, M. E. Moseley. Self-navigated interleaved spiral (SNAILS): Application to high-resolution diffusion tensor imaging. *Magn. Reson. Med.*. 2004;52:1388-1396.
5. N.-K. Chen, A. Guidon, H.-C. Chang, A. W. Song. A robust multi-shot scan strategy for high-resolution diffusion weighted MRI enabled by multiplexed sensitivity-encoding (MUSE). *NeuroImage*. 2013;72:41-47.
6. D. A. Porter, R. M. Heidemann. High resolution diffusion-weighted imaging using readout-segmented echo-planar imaging, parallel imaging and a two-dimensional navigator-based reacquisition. *Magn. Reson. Med.*. 2009;62:468-475.
7. R. M. Heidemann, D. A. Porter, A. Anwander, et al. Diffusion imaging in humans at 7 T using readout-segmented EPI and GRAPPA. *Magn. Reson. Med.*. 2010;64:9-14.
8. J. L. R. Andersson, S. Skare, J. Ashburner. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *NeuroImage*. 2003;20:870-888.
9. K. Setsompop, B. A. Gagoski, J. R. Polimeni, T. Witzel, V. J. Wedeen, L. L. Wald. Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced *g*-factor penalty. *Magn. Reson. Med.*. 2012;67:1210-1224.
10. K. P. Pruessmann, M. Weiger, M. B. Scheidegger, P. Boesiger. SENSE: Sensitivity encoding for fast MRI. *Magn. Reson. Med.*. 1999;42:952-962.
11. M. A. Griswold, P. M. Jakob, R. M. Heidemann, et al. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magn. Reson. Med.*. 2002;47:1202-1210.
12. R. Bammer, S. L. Keeling, M. Augustin, et al. Improved diffusion-weighted single-shot echo-planar imaging (EPI) in stroke using sensitivity encoding (SENSE). *Magn. Reson. Med.*. 2001;46:548-554.
13. M. Lustig, D. Donoho, J. M. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magn. Reson. Med.*. 2007;58:1182-1195.
14. M. Mani, M. Jacob, D. Kelley, V. Magnotta. Multi-shot sensitivity-encoded diffusion data recovery using structured low-rank matrix completion (MUSSELS). *Magn. Reson. Med.*. 2017;78:494-507.
15. Y. Hu, X. Wang, Q. Tian, et al. Multi-shot diffusion-weighted MRI reconstruction with magnitude-based spatial-angular locally low-rank regularization (SPALLR). *Magn. Reson. Med.*. 2020;83:1596-1607.
16. J. Veraart, D. S. Novikov, D. Christiaens, B. Ades-aron, J. Sijbers, E. Fieremans. Denoising of diffusion MRI using random matrix theory. *NeuroImage*. 2016;142:394-406.
17. Q. Tian, Z. Li, Q. Fan, et al. SDnDTI: Self-supervised deep learning-based denoising for diffusion tensor MRI. *NeuroImage*. 2023;253:119033.
18. T. Zhang, J. M. Pauly, I. R. Levesque. Accelerated parameter mapping with a locally low rank constraint. *Magn. Reson. Med.*. 2015;73:655-661.
19. Z. Tan, P. A. Liebig, R. M. Heidemann, F. B. Laun, F. Knoll. Accelerated diffusion-weighted magnetic resonance imaging at 7 T: Joint reconstruction for shift-encoded navigator-based interleaved echo planar imaging (JETS-NAViEPI). *Imaging Neuroscience*. 2024;2:1-15.
20. Y. Dong, X. Ye, C. Li, M. J. P. van Osch, P. Börnert. Navigator-free multi-shot diffusion MRI via non-local low-rank reconstruction. *Magn. Reson. Med.*. 2025;94:1592-1603.
21. V. Monga, Y. Li, Y. C. Eldar. Algorithm Unrolling: Interpretable, Efficient Deep Learning for Signal and Image Processing. *IEEE Signal Processing Magazine*. 2021;38:18-44.
22. S. Wang, Z. Su, L. Ying, et al. Accelerating magnetic resonance imaging via deep learning. In: *13th IEEE International Symposium on Biomedical Imaging (ISBI'13)*:514-517; 2016.
23. K. Hammerkirk, T. Klatzer, E. Kobler, et al. Learning a variational network for reconstruction of accelerated MRI data. *Magn. Reson. Med.*. 2018;79:3055-3071.
24. Y. Yang, J. Sun, H. Li, Z. Xu. ADMM-CSNet: A Deep Learning Approach for Image Compressive Sensing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2018;42:521-538.
25. H. K. Aggarwal, M. P. Mani, M. Jacob. MoDL: Model-based deep learning architecture for inverse problems. *IEEE Trans. Med. Imaging*. 2018;38:394-405.
26. M. Mani, V. A. Magnotta, M. Jacob. qModeL: A plug-and-play model-based reconstruction for highly accelerated multi-shot diffusion MRI using learned priors. *Magn. Reson. Med.*. 2021;86:835-851.
27. G. E. Hinton, R. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*. 2006;313:504-507.
28. B. Yaman, S. A. H. Hosseini, M. Akçakaya. Zero-shot self-supervised learning for MRI reconstruction. In: *10th International Conference on Learning Representations (ICLR'10)*; 2022.
29. J. Cho, Y. Jun, X. Wang, C. Kobayashi, B. Bilgic. Improved Multi-shot Diffusion-Weighted MRI with Zero-Shot Self-supervised Learning Reconstruction. In: *26th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'26)*:457-466; 2023.
30. J. Lehtinen, J. Munkberg, J. Hasselgren, et al. Noise2Noise: Learning Imaging Restoration without Clean Data. In: *Proceedings of the 35th International Conference on Machine Learning (ICML'35)*:2965-2974; 2018.
31. Y. Quan, M. Chen, T. Pang, H. Ji. Self2Self With Dropout: Learning Self-Supervised Denoising From Single Image. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2020)*:1890-1898; 2020.

- 01 32. M. Uecker, A. Karaus, J. Frahm. Inverse reconstruction
02 method for segmented multishot diffusion-weighted MRI
03 with multiple coils. *Magn. Reson. Med.*. 2009;62:1342-
04 1348. 54
- 05 33. A. Merrem, S. Hofer, A. S. A. Hosseini, et al. Diffusion-
06 weighted MRI of the prostate without susceptibility
07 artifacts: Undersampled multi-shot turbo-STEAM with
08 rotated radial trajectories. *NMR Biomed.*. 2019;32:e4074.
09 55
- 10 34. D. P. Kingma, J. Ba. ADAM: A Method for Stochas-
11 tic Optimization. In: *3rd International Conference on*
12 *Learning Representations (ICLR'3)*; 2015.
13 56
- 14 35. B. Yaman, S. A. H. Hosseini, S. Moeller, J. Eller-
15 mann, K. Uğurbil, M. Akçakaya. Self-supervised learning
16 of physics-guided reconstruction neural networks with-
17 out fully sampled reference data. *Magn. Reson. Med.*.
18 2020;84:3172-3191. 57
- 19 36. K. He, X. Zhang, S. Ren, J. Sun. Deep residual learning
20 for image recognition. In: *IEEE Conference on Com-*
21 *puter Vision and Pattern Recognition (CVPR'16)*:770-
22 778; 2016. 58
- 23 37. L. Cordero-Grande, D. Christiaens, J. Hutter, A. N.
24 Price, J. V. Hajnal. Complex diffusion-weighted image
25 estimation via matrix recovery under general noise mod-
26 els. *NeuroImage*. 2019;200:391-404. 59
- 27 38. M. Pietsch, D. Christiaens, J. V. Hajnal, J.-D. Tournier.
28 dStripe: Slice artefact correction in diffusion MRI via
29 constrained neural network. *Medical Image Analysis*.
30 2021;74:102255. 60
- 31 39. B. Zahneisen, M. Aksoy, J. Maclarens, C. Wuerslin,
32 R. Bammer. Extended hybrid-space SENSE for EPI: Off-
33 resonance and eddy current corrected joint interleaved
34 blip-up/down reconstruction. *NeuroImage*. 2017;153:97-
35 108. 61
- 36 40. Z. Tan, C. Unterberg-Buchwald, M. Blumenthal, et al.
37 Free-breathing liver fat, R_2^* and B_0 field mapping using
38 multi-echo radial FLASH and regularized model-based
39 reconstruction. *IEEE Transactions on Medical Imaging*.
40 2023;42:1374-1387. 62
- 41 63
- 42 64
- 43 65
- 44 66
- 45 67
- 46 68
- 47 69
- 48 70
- 49 71
- 50 72
- 51 73
- 52 74
- 53 75
- 54 76
- 55 77
- 56 78
- 57 79
- 58 80
- 59 81
- 60 82
- 61 83
- 62 84
- 63 85
- 64 86
- 65 87
- 66 88
- 67 89
- 68 90
- 69 91
- 70 92
- 71 93
- 72 94
- 73 95
- 74 96
- 75 97
- 76 98
- 77 99
- 78 100
- 79 101
- 80 102
- 81 103
- 82 104
- 83 105
- 84 106

Supplementary Information

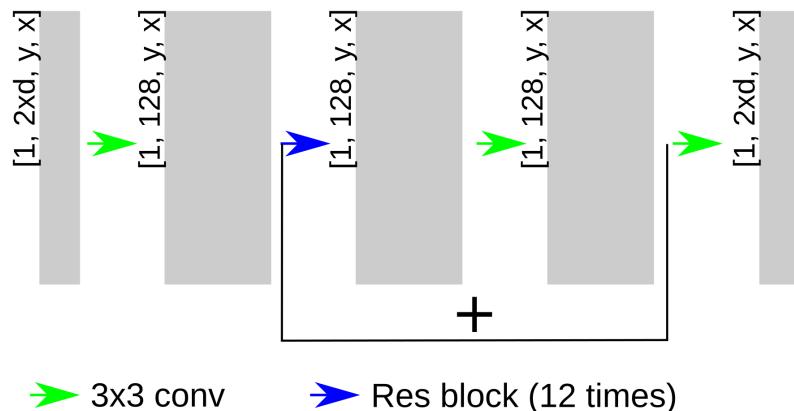
High-Resolution Diffusion-Weighted Imaging with Self-Gated Self-Supervised Unrolled Reconstruction

Zhengguo Tan, Patrick A. Liebig, Annika Hofmann, Frederik B. Laun, Florian Knoll

1	Residual Neural Network (ResNet)	2
2	Ablation Study	3
3	Generalizability: Cross Subjects	4
4	Reconstruction Time	6

1 Residual Neural Network (ResNet)

a) ResNet



b) Res Block

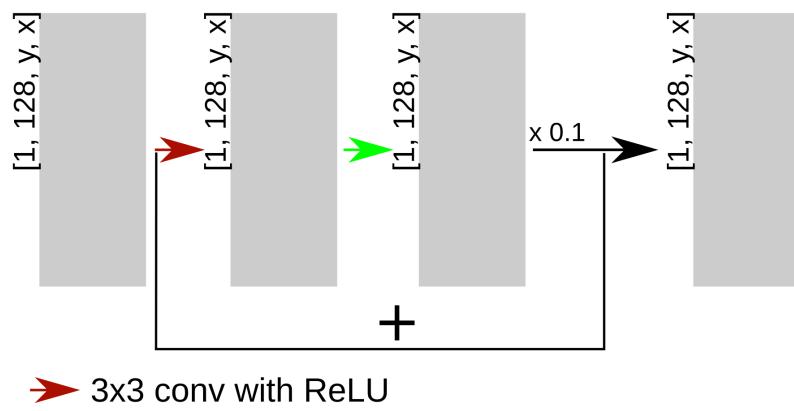


Figure S1: The architecture of ResNet. x and y denotes the shape of diffusion-weighted images, d denotes the number of diffusion encodings, and 2 indicates the real and the imaginary part.

In this study, the acquisition base resolution is 286, i.e., $x = y = 286$. The ResNet architecture in Figure S1 consists of 3,786,539 trainable parameters.

2 Ablation Study

In machine learning, ablation means the removal of a component of an artificial intelligence (AI) system. Here, we replace the ResNet with an Identity module (`torch.nn.Identity()`), which returns its input without any modification or computation, and has no trainable parameter. As a result, only the regularization parameter λ remains trainable during ADMM unrolling.

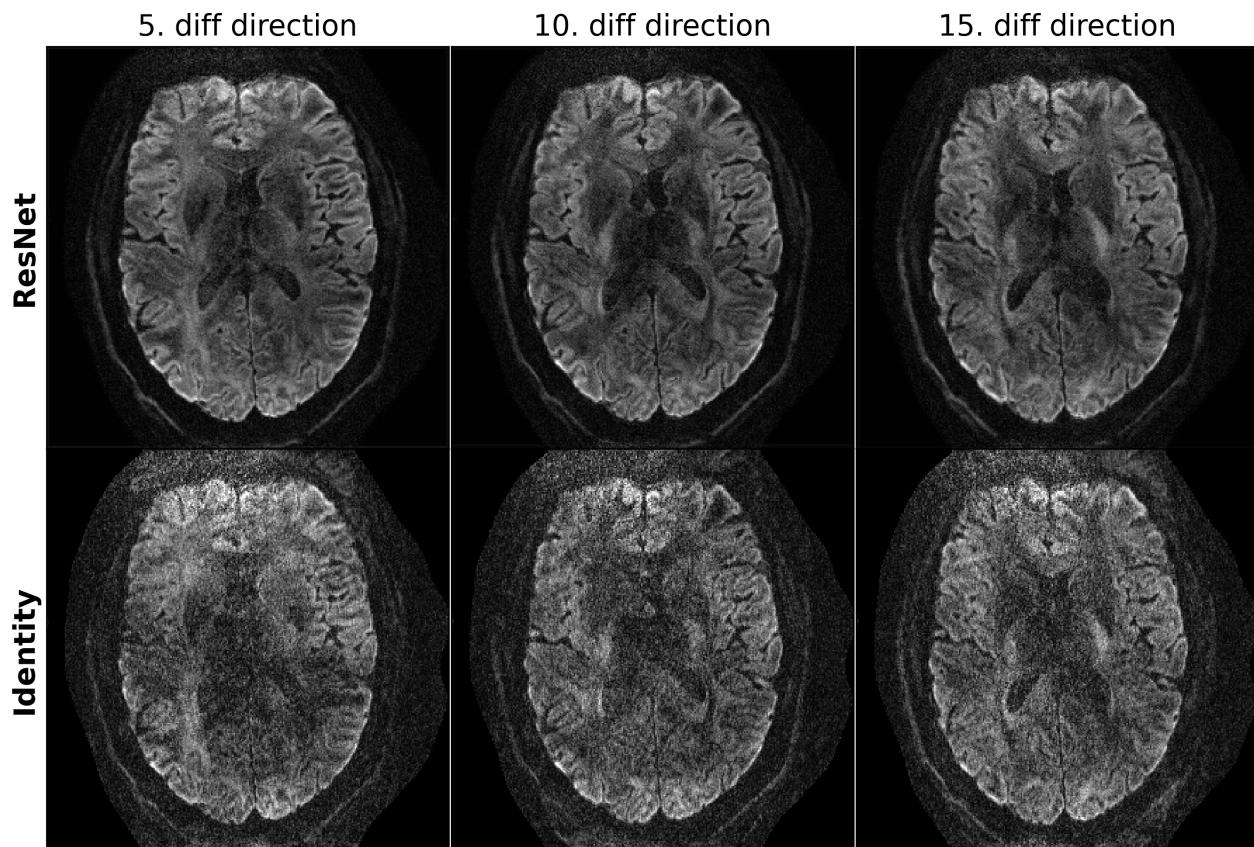


Figure S2: Ablation study. (1st Row) Reconstruction results of the proposed ADMM unrolling with the ResNet architecture. (2nd Row) Reconstruction results of ADMM unrolling without ResNet, but the Identity operator as the regularization.

Figure S2 displays the reconstruction results of the designed ablation study. This ablation study demonstrates the important role of the learned ResNet as a regularizer in removing noise while retaining sharp diffusion-weighted contrasts.

3 Generalizability: Cross Subjects

While Figure 4 in the main text shows the cross-slice generalizability of the proposed ADMM unrolling method, here we try to investigate whether the method generalizes among subjects. Specifically, we use the checkpoint from Figure S2 for the reconstruction of the data in Figure 4.

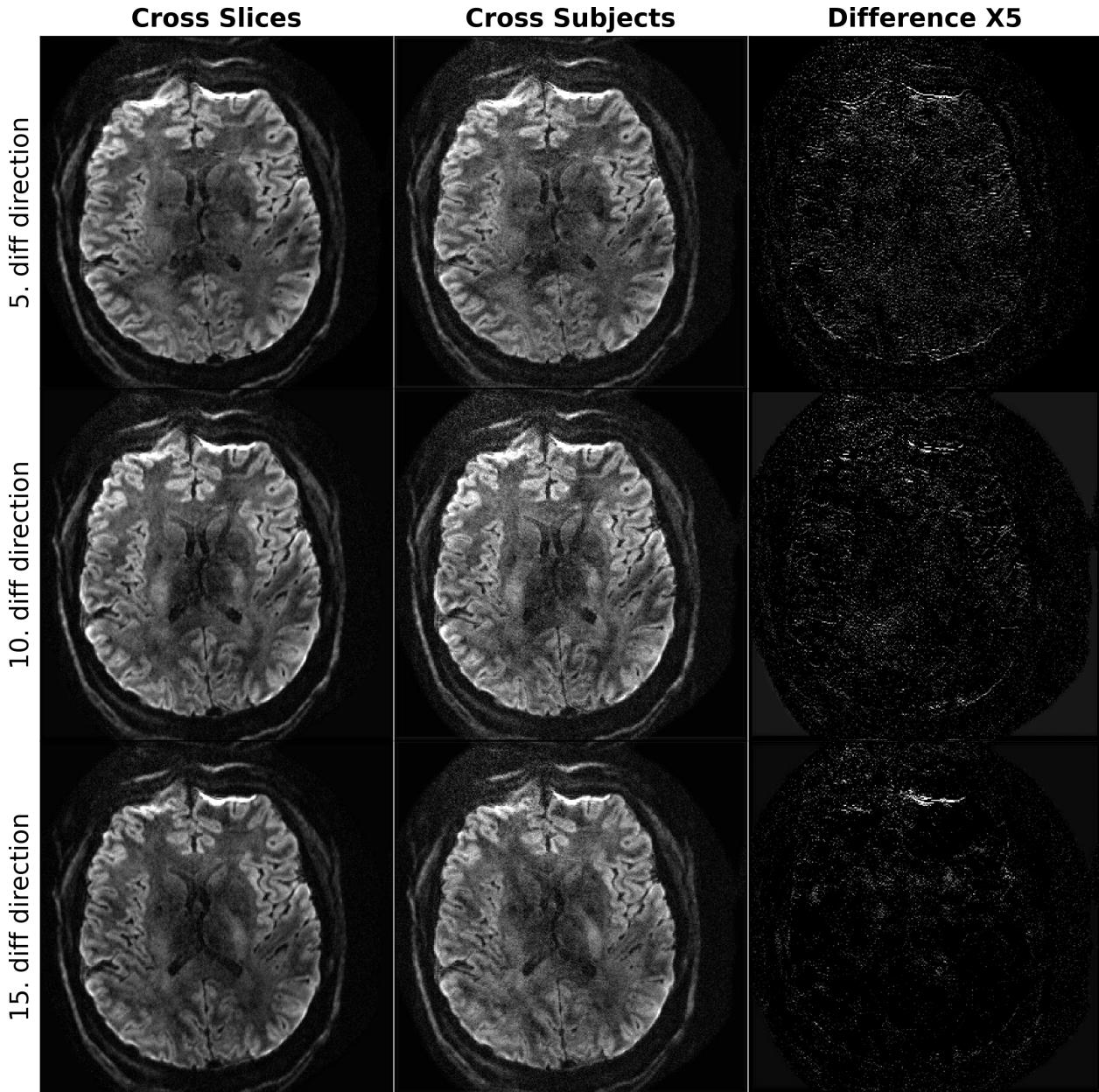


Figure S3: Cross-subject generalizability. (1st Column) Training and inference with the same subject, but only one slice is used for training. (2nd Column) Training and inference with difference subjects. Here, training is performed with the data in Figure S2, whereas inference is performed on a different subject. (3rd Column) Subtraction of diffusion-weighted images in the 1st column from those in the 2nd column.

Figure S3 shows the results of the cross-subject generalizability study. The difference images show only subtle differences between the two reconstruction results: cross-slice and cross-subject inferences. The proposed ADMM unrolling method generalizes well cross slices and subjects. We

foresee that ADMM unrolling with the self-supervised training can serve as a generalized model for diffusion-weighted image reconstruction and for fast inference.

4 Reconstruction Time

Table S1: Comparison of reconstruction times per multi-band slice. All reconstructions were performed on GPU A100 with 80GB memory.

Method	Training time (h)	Inference time (min)
MUSE	-	02:08
LLR	-	54:00
ADMM Unroll	04:27	01:50

MRM-25-26288: Responses to Editors and Reviewers**Editor's Comments to the Author:**

Thanks for resubmitting to Magn Reson Med. Your new manuscript went back to the original reviewers who both agree that it has been improved. However, there are still some remaining issues and corrections that are requested. We invite a further minor revision that seeks to address the remaining points that have been made. You may also wish to take this opportunity to ensure that your reference list is up to date and captures any recent relevant papers.

Deputy Editor: Wang, Shanshan Comments to Author:

The paper has been reviewed by two experts, who still have some concerns. Please properly address them. Many thanks.

We sincerely thank both editors for the encouraging comments. We have carefully addressed the comments:

- We have rewritten part of Introduction to better motivate the proposed ADMM unrolling for high-resolution DWI.
- We have updated our reference list, including the recent papers on noise2noise, self2self, and zero-shot self-supervised learning.
- We provide a document on Supplementary Information, which includes:
 - (1) The architecture of the ResNet;
 - (2) The results of an ablation study;
 - (3) The experiment that tests the trained model on a different subject to demonstrate the generalizability of the proposed self-supervised learning method;
 - (4) Reconstruction times.

Reviewer 1

This manuscript's application of an unrolled ADMM network with a zero-shot approach for DWI reconstruction is an interesting contribution, particularly the model's demonstrated generalizability across slices. However, there are several aspects that would benefit from further consideration and revision.

Regarding the manuscript's academic writing and presentation, there are several opportunities for enhancement.

- (1) *In the introduction, the presentation of related works could be strengthened by more clearly articulating the specific limitations of existing unrolled and self-supervised methods, which would better motivate the proposed solution.*

Thank you for the suggestion. We have rewritten the introduction.

- (2) *In the Method section, a couple of refinements could enhance clarity. The flowchart for Algorithm 1 is somewhat lengthy; streamlining it to highlight only the most essential steps would make it more accessible. Additionally, the self-supervised learning strategy should be supplemented with rigorous mathematical formulas to formally define the process.*

Thank you for the suggestion. After careful consideration, we would like to keep the pseudo code and Figure 2. We think this combination best explains how the data splitting is integrated with the training of ADMM unrolling.

- (3) *Finally, to substantiate the claim of faster reconstruction, the manuscript should include a direct, quantitative comparison of reconstruction times against competing methods.*

Please find the comparison in Table S1 in Supplementary Information.

There a few points require clarification in the manuscript.

- (4) *In Section 2.3, could you provide more specific details regarding the ADMM unrolling implementation? For example, information on the architecture of the ResNet used, such as the number of layers, would be very helpful.*

Thank you for the question. Please find the architecture of the ResNet in Figure S1 in Supplementary Information. We also provide the source code on ResNet here: <https://github.com/ZhengguoTan/DeepDWI/blob/main/src/deepdwi/models/resnet.py>.

- (5) *Figure 9 suggests that λ from Equation (3) is a learnable parameter. Is the penalty parameter ρ in the same equation also treated as learnable, or is it a fixed value?*

Thank you for the question. In the current implementation, the penalty parameter ρ in ADMM is kept as a fixed value, 0.05.

- (6) *In the ablation study described in Section 3.1, you compare the reconstruction results of MUSE, LLR, and ADMM unrolling for both 4-shot and 2-shot acquisitions. Could you clarify*

how this comparison constitutes an ablation study? An ablation study typically analyzes the impact of removing specific components of the proposed model, whereas this seems to be a performance comparison across different methods under varying conditions.

Thank you for the constructive comments.

- We agree that the result in Figure 3 (Section 3.1) is not an ablation study, but a retrospective study. We correct the wording in the manuscript.
- We provide an ablation study in Figure S2 in Supplementary Information. Here, we replace the ResNet with an Identity module. In other words, we don't embed ResNet into ADMM unrolling and compare the results without and with ResNet.

Reviewer 2

Thank you to the authors for addressing some of my previous concerns. However, the current version still does not meet the requirements for publication, and several issues remain to be resolved:

- (1) A major limitation of the proposed method is that it requires training a separate model for each subject, which significantly restricts its applicability. I suggest that the authors provide detailed training and testing times. Alternatively, they could evaluate whether a model trained on one subject's data can be generalized to other subjects.

Thank you for the insightful comments.

- Please find detailed training and testing times in Table S1 in Supplementary Information.
- We evaluate the model trained on one subject but inferred on a different subject. Please find details in Figure S3 in Supplementary Information.

- (2) In Table 3, it appears that the fully sampled data are different across methods. What is the rationale for this design choice? Why are the fully sampled datasets not identical?

Sorry, we don't have Table 3. If you meant Figure 3, we clarify that both columns in Figure 3 come from the same scan. In other words, we acquired 4-shot fully-sampled data as reference. Reconstruction results of the 4-shot data are provided as the first column in Figure 3. Then we retrospectively undersampled the data to have only 2 shots, i.e., 2-fold undersampling. Reconstruction results of the retro. 2-shot data are provided as the second column in Figure 3.

- (3) In Figure 9, is the λ parameter the same as the λ used in Algorithm 1? Their initial values seem to be inconsistent.

Thank you for the comment. The parameter λ in Figure 9 and Algorithm 1 is the same and is initialized as 0.05. Please note that the λ curve in Figure 9 is colored blue, and its corresponding y-axis is on the right side of Figure 9 (with the same color blue).

- (4) On page 9, line 50, the authors state: "However, we also observed that the self-gated ... (e.g., the $0.5 \times 0.5 \times 2.0 \text{ mm}^3$ DWI data with an acceleration of 15×2 per shot)." How was the $0.5 \times 0.5 \times 2.0 \text{ mm}^3$ dataset acquired? The manuscript does not explain.

Thank you for the insightful comment. We removed the sentence on $0.5 \times 0.5 \times 2.0 \text{ mm}^3$, which is not relevant to this manuscript.

- (5) For retrospective experiments, I recommend including reference data for comparison.

Thank you for the suggestion. We would like to clarify that in the retrospective experiment, we use the 4-shot fully-sampled data as reference.

- (6) Please check the consistency of abbreviations in the manuscript, such as “selfgated” vs. “self-gated,” and standardize them accordingly.

Done. We now use consistently “self-gated”.