High-Resolution Motion-Robust Diffusion-Weighted Imaging with Self-Gated Zero-Shot Self-Supervised Reconstruction

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

***Abstract*— This work introduces a self-gated zero-shot self-supervised learning (ZSSSL) reconstruction framework for navigator-free, high-resolution diffusion-weighted imaging using undersampled multi-shot interleaved echo-planar imaging (iEPI) acquisition. ZSSSL belongs to the algorithm unrolling technique, with a physics-guided data-consistency term and a learned regularization function. We unroll the alternating direction method of multipliers (ADMM) with a 2D residual neural network to leverage redundancy in the spatial-diffusion dimension. First, the proposed self-gated ZSSSL method is validated with both retrospectively and prospectively acquired data at 0.7 mm isotropic resolution. This approach outperforms both multiplexed sensitivity-encoding (MUSE) and compressed sensing with locally-low rank (LLR) regularization in terms of image sharpness, tissue continuity and motion robustness. Second, the implemented ADMM unrolling converges sufficiently in 100 epochs, and enables the learning of the regularization strength. Third, while ZSSSL requires up to eight hours training time per slice, it generalizes well to all other slices, with an inference time of just one minute. In comparison, LLR requires about an hour per slice. Overall, self-gated ZSSSL enables undersampled multi-shot iEPI acquisition without the need of navigators, providing sub-millimeter DWI at clinically feasible reconstruction times. The source code is publicly available at: https://github.com/ZhengguoTan/DeepDWI.**

***Index Terms*— Diffusion weighted imaging, Magnetic resonance imaging, Image reconstruction, End-to-end learning in medical imaging, Machine learning**

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. In addition, scientific support and HPC resources were 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. *(Corresponding Author: Zhengguo Tan)*

Z. Tan was with the Department Artificial Intelligence in Biomedical Engineering (AIBE), FAU, Erlangen, Germany. He is now with the Michigan Institute for Imaging Technology and Translation (MIITT), Department of Radiology, University of Michigan, Ann Arbor, MI 48109 USA (e-mail: zgtan@med.umich.edu).

P. A. Liebig is with Siemens Healthcare GmbH, Erlangen, Germany (e-mail: patrick.liebig@siemens-healthineers.com).

A. Hofmann is with the Department AIBE, FAU, Erlangen, Germany (e-mail: annika.ah.hofmann@fau.de).

F. B. Laun is with the Institute of Radiology, University Hospital Erlangen, FAU, Erlangen, Germany (e-mail: Frederik.Laun@uk-erlangen.de).

F. Knoll is with the Department AIBE, FAU, Erlangen, Germany (e-mail: florian.knoll@fau.de).

I. Introduction

HIGH-dimensional magnetic resonance imaging (HD-MRI) has been a flourishing field, focused on the acquisition, reconstruction and analysis of multi-dimensional multi-contrast-weighted MRI data. Examples of HD-MRI include but are not limited to magnetic resonance spectroscopic imaging (MRSI) [1], diffusion-weighted imaging (DWI) [2], [3], and quantitative parameter mapping [4], [5]. Conventional HD-MRI, however, requires long acquisition times, making the data susceptible to subject motion and system imperfections, and imposing high computational burden. DWI, in particular, poses challenges in the pursuit of high spatial, temporal, and angular resolution. DWI is typically acquired using the pulsed gradient spin echo diffusion-weighted sequence [6] followed by fast echo-planar imaging (EPI) readouts [7]. However, the use of long echo trains in EPI results in geometric distortion artifacts and reduced spatial resolution. Additionally, acquiring multiple diffusion directions to enhance angular resolution and to better probe tissue microstructure further extends the scan time.

Advances in parallel imaging [8]–[12] and compressed sensing [13]–[15] have enabled accelerated acquisition for HD-MRI. Notably, the low-rank model [16] has been a powerful tool in dimension reduction. Typically, singular value decomposition (SVD) is used to learn a truncated temporal basis function from a large-scale physics-informed dictionary [17]–[19]. The temporal basis function is then integrated with the MRI forward model, i.e. the sensitivity encoding operator [11], for joint reconstruction of the corresponding spatial basis images. In addition, low-rank regularization can be employed in the joint reconstruction [20].

Beyond the low-rank technique, advanced neural networks, e.g. autoencoder [21], have been explored for HD-MRI reconstruction and proven to supply more accurate representations of high-dimensional data than SVD. Lam et al. [22] and Mani et al. [23] proposed to first learn a denoising autoencoder (DAE) model from a physics-informed simulated dictionary and then incorporate the learned DAE model as a regularizer in the alternating direction method of multipliers (ADMM) [24] unrolling reconstruction. Pioneered by Gregor and LeCun [25], algorithm unrolling enables the use of learned deep *priors* as regularization and offers faster inference compared to iterative reconstruction methods that rely on hand-crafted regularization functions [26]. Algorithm unrolling has been applied to accelerated MRI reconstruction in various scenarios: including but not limited to supervised learning with fully sampled reference images [27], [28], and self-supervised learning with only undersampled data available for training [29], [30]. Notably, acquiring fully sampled DWI for training a regularization function is quite challenging. First, fully-sampled DWI requires longer echo times in EPI, which not only elongates the scan times but also reduces the signal-to-noise ratio (SNR). Second, the variety of diffusion acquisition modes necessitates a larger dataset compared to two-dimensional imaging scenarios [31]. As a result, self-supervised learning is better suited for DWI reconstruction.

Deep neural networks are capable of learning not only regularization functions, but also MR-physics forward operators. Liu et al. [32] proposed the reference-free T_1 parameter maps extraction (RELAX) self-supervised deep learning reconstruction, which learns the mapping from T_1 parameter maps to undersampled multi-coil multi-contrast k-space data. Arefeen et al. [33] proposed to replace the conventional SVD-based linear subspace modeling [17] by the latent decoder model within DAE for improved T_2-weighted image reconstruction. The ability of DAE to learn DWI models is somewhat uncertain. DAE is composed of sequential fully connected layers with nonlinear activation functions, which may struggle with complex functions like those required for DWI signals. On the other hand, the standard diffusion tensor model [34] consists of six tensor elements, so the diffusion-weighted signal dictionary based on the diffusion tensor model usually consists of tens of millions atoms.

Contributions:

- We unrolled ADMM to perform zero-shot self-supervised learning (ZSSSL) and incorporated self-gated shot-to-shot phase variation estimation into ZSSSL for deep diffusion-weighted imaging reconstruction.
- We demonstrated that the trained ZSSSL model from one single slice can be applied to all other slices. This significantly reduces the training time.
- We achieved navigator-free high-resolution DWI with 21 diffusion-encoding directions at $0.7\,\mathrm{mm}$ isotropic resolution, and a scan time of under 10 minutes.

II. Related Work

A. Multi-Band Multi-Shot DWI Acquisition & Modeling

Our previous work [35] demonstrated the joint k-q-slice forward operator for multi-band multi-shot navigator-based interleaved EPI (NAViEPI) DWI acquisition. This operator can be understood as an extended sensitivity encoding (SENSE) operator [11], which maps the multi-slice multi-diffusion-weighted images ($\mathbf{x}$) to their corresponding k-space,

$$\mathcal{A}(\mathbf{x}) = \mathbf{P\Sigma\Theta FS\Phi x} \tag{1}$$

Here, the images $\mathbf{x}$ are point-wise multiplied with the pre-computed shot-to-shot phase variation maps ($\mathbf{\Phi}$) and coil sensitivity maps ($\mathbf{S}$). The output images are then converted to k-space via two-dimensional fast Fourier transform ($\mathbf{F}$), point-wise multiplied with the multi-band phases ($\mathbf{\Theta}$), summed along the slice dimension ($\mathbf{\Sigma}$), 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 [36]–[39] realized 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. Self-gated shot phase estimation does not require the acquisition of phase navigator data. However, it requires small undersampling factors per shot and fully-sampled DWI acquisition assembling all shots. Alternatively, undersampled DWI acquisition can be enabled via the acquisition of navigators for shot phase estimation [35]. This approach allows for mesoscale-resolution DWI at $7\,\mathrm{T}$, but still needs long scan time. As listed in Table I, the total acquisition of Protocol #3 at $0.7\,\mathrm{mm}$ isotropic resolution takes $16:27$ minutes with phase navigators. This scan time can be reduced to approximately 10 minutes be removing the phase navigators (Protocol #2 in Table I).

With the operator $\mathcal{A}$, the joint reconstruction is expressed as,

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

where $\mathbf{y}$ is the measured k-space data. The first term in Equation (2) presents data consistency, 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) regularzation [35] or neural networks with nonlinear activation functions, ADMM was employed in this work to solve for Equation (2).

B. Algorithm Unrolling for Deep Image Reconstruction

Algorithm unrolling has been an emerging technique in solving inverse problems with learnable deep neural networks. Algorithm unrolling consists of two ingredients. First, it uses deep neural networks to learn regularization function. Second, it is constrained by the data-consistency term. In other words, the forward pass of the estimate $\mathcal{A}(\mathbf{x})$ must be close to the measured data $\mathbf{y}$. By mapping the operations used in iterative algorithms onto networks, unrolled algorithms can be trained with data, leading to much faster inference than conventional iterative algorithms [26]. Further, recent developments have shown that the operations used in compressed sensing MRI, i.e., sparsifying transformation and soft thresholding, can be learned via neural networks. For instance, Hammernik et al. [27] proposed to unroll the gradient descent algorithm with a learned neural network (e.g. U-net [40]) as the regularization function. Aggarwal et al. [28] proposed the model-based deep learning architecture for inverse problems (MoDL) to unroll the alternating minimization algorithm with a learned residual denoising network [41] as regularization.

C. Self-Supervised Learning for Image Reconstruction

In many MRI applications, such as dynamic imaging and diffusion-weighted imaging, acquiring fully-sampled data for

TABLE I
NAVIEPI ACQUISITION PROTOCOLS

Protocol	#1	#2
Diffusion mode	MDDW	
Diffusion scheme	monopolar	
Diffusion direction	20	
b-value ($\mathrm{s/mm^2}$)	1000	
b_0	1	
FOV ($\mathrm{mm^2}$)	200	
Matrix size	$286 \times 286 \times 176$	
Voxel ($\mathrm{mm^3}$)	$0.7 \times 0.7 \times 0.7$	
Shots	3	
Acceleration	2×2	
Partial Fourier	$5/8$	
Bandwidth ($\mathrm{Hz/Pixel}$)	972	
ESP (ms)	1.17	
Navigator	Yes	No
TE (ms)	$58/98.3$	58
TR (ms)	15000	8900
Acquisition (min)	$16:27$	$9:57$

supervised learning can be challenging. To tackle this issue, Yaman et al. [29] proposed self-supervised learning via data undersampling (SSDU), which learns the regularization function in Equation (2) by splitting available undersampled data into two disjoint sets, one of which is used in the data consistency term and another used for the computation in the training loss function. The training of SSDU requires large undersampled data sets. To close the domain gap between training and test data, Yaman et al. [30] proposed scan-specific zero-shot self-supervised learning (ZSSSL), which splits a single data set into three disjoint sets for (a) the data consistency term, (b) the loss calculation during training, and (c) validation, respectively. Recently, ZSSSL has been adopted for multi-contrast image reconstruction [42].

III. METHODS

A. Data Acquisition

Table I lists three 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 $80\,\mathrm{mT/m}$ and a peak slew rate $200\,\mathrm{T/m/s}$). Protocols #1 and #2 realized mesoscale DWI with $0.7\,\mathrm{mm}$ isotropic resolution. Two-fold acceleration is employed in both in-plane and slice directions. Every DWI data is acquired by three shots in an interleaved manner, which resultes in 6×2-fold acceleration per shot. Noteworthy, the total scan time can be reduced to about 10 minutes (Protocol #2) when switching off navigator acquisition. Three young healthy volunteers with written informed consent approved by the local ethics committee participated in this study.

B. Image Reconstruction via ADMM Unrolling and Zero-Shot Self-Supervised Learning

Instead of the two-step alternating minimization unrolling scheme as used in MoDL [28], we employed the ADMM

Algorithm 1 ADMM Unrolling for ZSSSL

Initialization:
  split sampling mask $\mathbf{P}$ into 12 repetitions, each of which consists of three disjoint sets $\mathbf{T}$, $\mathbf{L}$, and $\mathbf{V}$
  $p \leftarrow 0$ and $N_{\text{epoch}} \leftarrow 100$
  $\mathcal{D}_\omega$ set as ResNet
  $\rho \leftarrow 0.05$ and $\lambda \leftarrow 0.05$
  $\text{Loss}_{\text{valid}} \leftarrow \text{inf}$ and $\text{trace} \leftarrow 0$
function ADMM(mask)
  $\mathcal{A}_{\text{mask}} \leftarrow$ set the mask in the forward operator $\mathcal{A}$
  $\mathbf{x}^{(0)} \leftarrow \mathcal{A}^H_{\text{mask}}(\mathbf{y})$
  $\mathbf{v}^{(0)} \leftarrow \mathbf{x}^{(0)}$ and $\mathbf{u}^{(0)} \leftarrow \mathbf{0}$
  $k \leftarrow 0$ and $N_{\text{unroll}} \leftarrow 8$
  while $k < N_{\text{unroll}}$ **do**
    $\mathbf{x}^{(k+1)} \leftarrow$ conjugate gradient with 6 iterations
    $\mathbf{v}^{(k+1)} \leftarrow (\lambda/\rho) \cdot \mathcal{D}_\omega(\mathbf{x}^{(k+1)} + \mathbf{u}^{(k)})$
    $\mathbf{u}^{(k+1)} \leftarrow \mathbf{u}^{(k)} + \mathbf{x}^{(k+1)} - \mathbf{v}^{(k+1)}$
    $k \leftarrow k + 1$
  end while
  return $\mathbf{x}^{(k+1)}$
end function
Training:
while $p < N_{\text{epoch}}$ or $\text{trace} \leq 12$ **do**
  $\mathbf{x}_t \leftarrow \text{ADMM}(\mathbf{T})$
  $\text{Loss}_{\text{train}} \leftarrow \mathcal{L}(\mathbf{Ly}, \mathcal{A}_{\mathbf{L}}(\mathbf{x}_t))$
  update ω via ADAM
  Validation:
  $\mathbf{x}_t \leftarrow \text{ADMM}(\mathbf{T} \cup \mathbf{L})$
  $\text{Loss}_{\text{temp}} \leftarrow \mathcal{L}(\mathbf{Vy}, \mathcal{A}_{\mathbf{V}}(\mathbf{x}_t))$
  if $\text{Loss}_{\text{temp}} \leq \text{Loss}_{\text{valid}}$ **then**
    $\text{Loss}_{\text{valid}} \leftarrow \text{Loss}_{\text{temp}}$
    $\text{trace} \leftarrow 0$
  else
    $\text{trace} \leftarrow \text{trace} + 1$
  end if
end while

unrolling to solve the self-supervised learning reconstruction in Equation (2). The update rule of ADMM unrolling reads

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

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 $\mathcal{D}_\omega$ 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 $\mathbf{v}$.

As shown in Figure 1, the data sampling mask $\mathbf{P}$ in ZSSSL [30] is split into three disjoint sets, the training mask $\mathbf{T}$ for the data consistency term, the training loss mask $\mathbf{L}$ for the loss function calculation, and the validation loss mask $\mathbf{V}$. Each

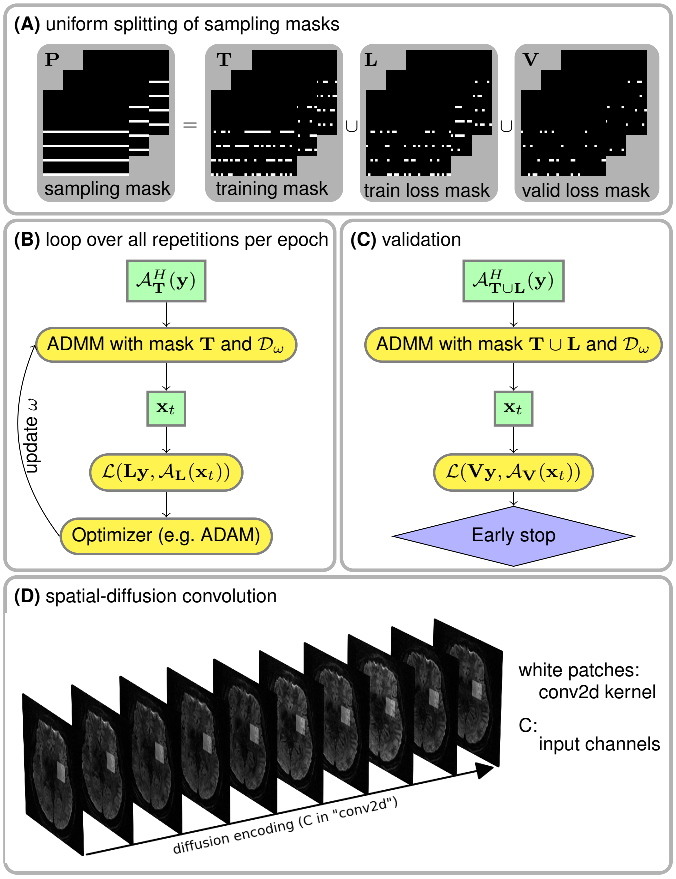

Fig. 1. Illustration of the key components in ZSSSL. **(A)** The sampling mask $\boldsymbol{P}$ in Equation (1) was uniformly split into three disjoint sets: the training mask $\mathbf{T}$ used for the data consistency term during ZSSSL training, the train loss mask $\mathbf{L}$ used for the loss function calculation during ZSSSL training, and the validation loss mask $\mathbf{V}$ used for the loss function calculation during ZSSSL validation. **(B)** and **(C)** show the flowchart for the training and the validation of ZSSSL, respectively. Note that the ResNet parameters $\boldsymbol{\omega}$ are updated during training, but remain fixed during the validation step. **(D)** A stack of DWIs is input into ResNet during ADMM unrolling.

set consists of 12 repetitions constructed via random uniform sampling of the data mask $\mathbf{P}$. In each training epoch, every repetition is looped through in order to update the ResNet parameters ω. 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 index k in Equation (3) denotes the unrolling iteration, and $\mathcal{D}_\omega$ denotes the residual network (ResNet) [41] parameterized by ω. In this work, 2D convolution was employed to construct the ResNet. 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 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 (x) to be reconstructed has 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 DWIs, N_X and N_Y are the width and the height of DWIs, N_Z is the number of slices (same as the multi-band factor), and N_{diff} is the number of diffusion encodings. To train a ResNet based on 2D convolution, the DWIs were reshaped and permuted as $(N_Z, 2\cdot N_{\text{diff}}, N_Y, N_X)$, as illustrated in Figure 1 (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.

C. 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 ZSSSL 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, all other regularized reconstructions fall into the joint reconstruction regime. They jointly reconstruct all DWIs and imposes regularization terms that explore spatial-diffusion redundancy. For example, LLR enforces low rankness of local spatial-diffusion matrices from DWIs, whereas ZSSSL learns a ResNet regularization function based on spatial-diffusion convolution kernels while enforcing data consistency during the unrolled training process.

D. Self-Gated ZSSSL

As discussed in Section II-A, there are two approaches for estimating shot-to-shot phase variation: self-gated and navigator-based. The self-gated approach, as used in MUSE [38], requires fully-sampled DWI acquisition and has typically reported only a small number of shots (up to 4). The previously proposed NAViEPI approach enabled high-resolution DWI with the use of undersampled iEPI and shot-to-shot phase navigator acquisition. While NAViEPI results in shorter scan time than fully-sampled iEPI, the use of phase navigator still elongates the acquisition, as listed in Table I. Therefore, a key question is whether it is feasible to discard the shot-to-shot phase navigator while keeping undersampled iEPI acquisition. In this work, we investigated the feasibility of ZSSSL in self-gated scan for $0.7\,\text{mm}$ isotropic resolution DWI.

E. ZSSSL Model Generalizability

Volumetric whole brain DWI acquisition consists of many multi-band slices, and the training of algorithm unrolling models on every slice 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 ZSSSL model with only one multi-band slice data, and then tested the model on all remaining multi-band slices. We dubbed this approach as "single-slice training". Second, we trained and tested every multi-band slice individually, which was dubbed 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 training.

F. Computation

All reconstructions were in this work done on a single A100 SXM4/NVLink GPU with $80\,\mathrm{GB}$ memory (NVIDIA, Santa Clara, CA, USA). Computing infrastructure was provided by the Erlangen National High Performance Computing Center.

IV. Results

A. Retrospectively Self-Gated ZSSSL

Figure 2 demonstrates the efficacy of the self-gated ZSSSL reconstruction by comparing to the navigated and self-gated reconstructions on the first volunteer. Data were acquired by the NAViEPI sequence, as listed in Protocol #1 in Table I. The single-direction DWIs with accidental motion were displayed.

The selected DWIs showed residual aliasing-like and severe motion-blurring artifacts in the navigated reconstructions, including both LLR and ZSSSL. The main reason of these artifacts is that the acquisition of navigators increased the total scan time, resulting in higher sensitivity to accidental inter-shot motion. Admittedly, navigators are valuable in the case of ultra high spatial resolution using many shots, e.g. 3-fold in-plane undersampling and 5-shot acquisition for the in-plane resolution of 0.5 mm [35]. In this experiment, however, the utilization of 3 shots yielded 6×2-fold acceleration per shot (refer to Protocol #1). Such an acceleration rate proved achievable in the self-gated approach. Both LLR and ZSSSL reconstructions supplied geometrically correct DWIs without noticeable aliasing artifacts. This in turn indicated that motion corrupted the navigator data in this measurement. Further, self-gated ZSSSL exhibits much clearer tissue delineation in reconstructed DWIs, as indicated by red arrows in the zoomed-in views in Figure 2, whereas self-gated LLR suffers from slightly blurry tissue boundaries and ambiguous signals.

Figure 3 shows coronal- and saggital-view diffusion-weighted images with the same diffusion encoding as in Figure 2. As mentioned in Section III-E, the ZSSSL model was trained using only one slice and then inferred on the remaining slices. The ZSSSL model generalized well across slices. The inference of every slice took only about one minute, whereas the LLR reconstruction took about 48 minutes per slice. More importantly, the self-gated LLR reconstruction exhibited residual motion-induced stripping artifacts (refer to red arrows in Figure 3)) [43], whereas the self-gated ZSSSL approach substantially removed these artifacts and supplied high-quality DWI without the need of navigators. Both reconstructions showed B_1 field inhomogeneities in the cerebellum region and residual spatial distortion in the frontal brain region. These artifacts, however, are beyond the scope of this work.

B. Prospectively Self-Gated ZSSSL

Figure 4 compares the reconstruction results using the prospectively acquired iEPI data without navigators of the second volunteer (refer to Protocol #2 in Table I). The snapshot single diffusion-direction DWIs as well as mean DWIs at three orthogonal views were displayed.

The MUSE reconstruction suffered from strong noise at such mesoscale voxel size. The mean DWIs of MUSE improved the visibility of brain tissues, but the overall image quality is not sufficient. The LLR regularized reconstruction largely reduced noise, but still exhibited suspicious dark signal in the axial view, which appeared as striping artifacts in the coronal and sagittal views (refer to the red arrows in Figure 4). These artifacts can be reduced by incorporating adaptive noise estimation in LLR regularization [44]. On the other hand, the LLR regularized reconstruction showed residual noise in the cerebellum region (refer to the blue arrows in Figure 4), which could be caused by the B_1 excitation field inhomogeneity at $7\,\mathrm{T}$. The above-mentioned striping artifacts were nearly gone in the ZSSSL reconstruction. Moreover, the DWI from ZSSSL in the axial view showed clearer diffusion contrasts and thus better tissue delineation and continuity in the coronal and the sagittal views. Further, as indicated by the blue arrows in Figure 4, the DWI in the sagittal view showed more homogeneous signal distribution and reduced noise surrounding the cerebellum.

Figure 5 displays the training and validation loss as well as the learned regularization strength along epochs for the results shown in Figure 4. It can be seen that 100 epochs were sufficient for the convergence of ADMM unrolling. The ZSSSL model converged well along epochs, and did not show any over-fitting behavior. In addition, the regularization strength converged to the value of about 0.027.

Figure 6 showed the reconstructed DWIs at four different diffusion directions based on the iEPI data acquired from the third volunteer. In this experiment, the volunteer was instructed to keep still during scan. Again, the proposed self-gated ZSSSL reconstruction with spatial-diffusion convolution illustrated superior tissue structure delineation and diffusion contrasts to the LLR regularized reconstruction. Similar to the cerebellum region, the frontal brain region suffers from higher noise in the LLR reconstruction. In contrast, the ZSSSL approach generally illustrated more homogeneous signal and noise distribution across the field-of-view. LLR builds upon one single linear transformation (singular-value decomposition, SVD) and one nonlinear soft thresholding, whereas the ResNet in ZSSSL builds upon multiple convolutions and non-linear activation functions. Therefore, deep neural networks enable more in-depth exploration of key features in the high-dimensional data.

C. Model Generalizability

Based on the data shown in Figure 6, Figure 7 demonstrated the generalizability of the proposed ADMM unrolling approach. Single-direction DWIs from both the slice-by-slice training and the single-slice training strategies are displayed. The difference image between these two DWIs shows no residual structural information, but mainly noise. Further, we plotted the mean and standard deviation within the selected region-of-interest along all diffusion encodings. The plotted curves show quantitatively similar values between the two reconstructions.

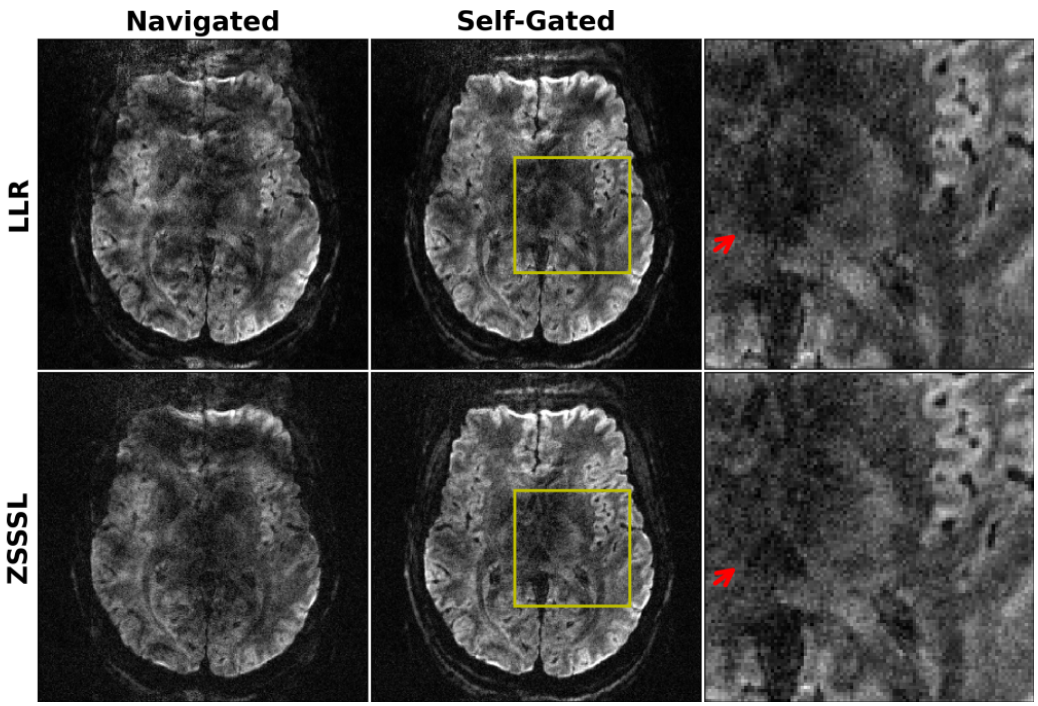

Fig. 2. Comparison of (top) LLR regularized and (bottom) ZSSSL reconstruction on 0.7 mm isotropic resolution DWI acquired by Protocol #1 with shot phase estimated from (left) navigators and (middle) imaging echoes, respectively. Zoomed views of the yellow boxes from the self-gated reconstruction are displayed in the right column. The use of navigators prolongs the total scan time, and thus increases the sensitivity to motion, as shown in the single-direction diffusion-weighted image reconstructed with navigated shot phase. The retrospectively self-gated reconstruction discards navigators, and renders sharper diffusion-weighted images. Compared to LLR, ZSSSL is advantageous in resolving clearer tissue boundaries in diffusion-weighted images, as indicated by red arrows.

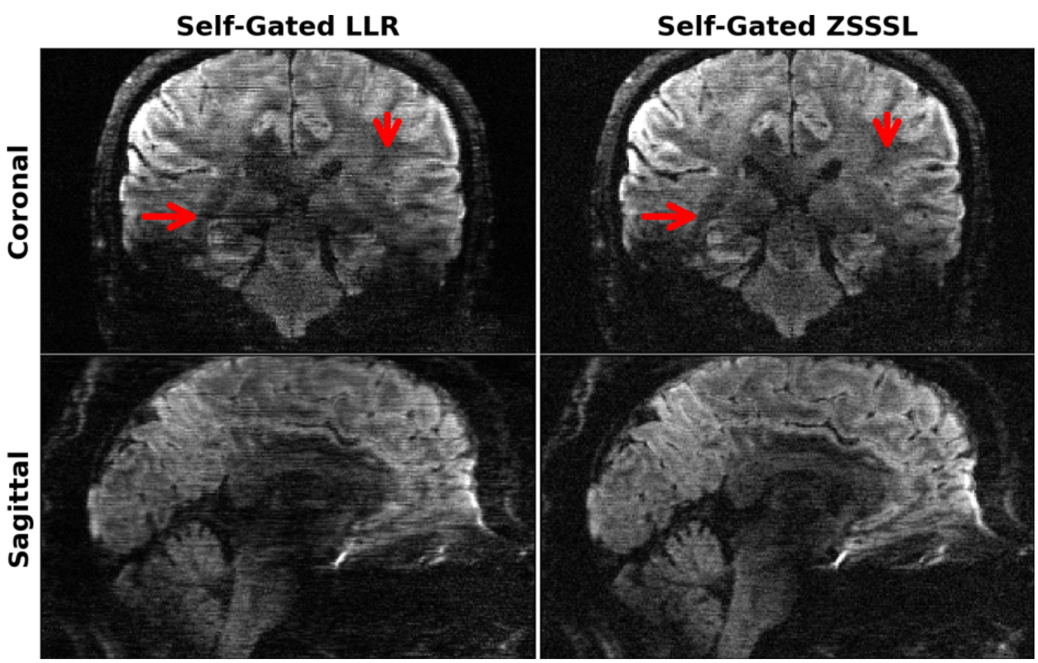

Fig. 3. Single-direction diffusion-weighted images at 0.7 mm isotropic resolution as reconstructed by retrospectively self-gated (left) LLR and (right) ZSSSL in (top) the coronal and (bottom) the sagittal views. The same diffusion direction as in Figure 2 is chosen for display. ZSSSL reduces phase ambiguities in the shot-combined reconstruction , thereby rendering clearer tissue delineation and reduced stripping artifacts (as indicated by the red arrows).

V. Discussion

This work reported a novel self-gated zero-shot self-supervised learning approach for multi-shot undersampled iEPI acquisition and high-resolution DWI reconstruction. The self-gated ZSSSL achieved whole brain diffusion encoding in 21 directions with a b-value of $1000\,\mathrm{s/mm^2}$ at $0.7\,\mathrm{mm}$ isotropic resolution, all within a scan time of less than 10 minutes. Technically, this work unrolled ADMM to perform ZSSSL training and testing. Likewise, ADMM was employed to solve the inverse problem in Equation (2) with LLR regularization. This approach assures fair comparison among different regularization methods.

The proposed self-gated ZSSSL approach is well-suited for online reconstruction deployment. Firstly, it requires much shorter acquisition time than the conventional MUSE approach with fully-sampled iEPI and our previous NAViEPI method. Secondly, ZSSSL does not require large-scale fully-sampled data for training. Instead, the training of ZSSSL is scan specific. Last but not the least, the trained ZSSSL model is applicable to different undersampling factors and to different slices. Fourth, the inference time of ZSSSL is much shorter compared to the LLR regularization approach.

We observed that stripping-type motion artifacts occurred more frequently with sub-millimeter isotropic resolution DWI. In addition, sub-millimeter isotropic voxel resulted in higher noise in DWI. This makes sense, as scans with reduced slice

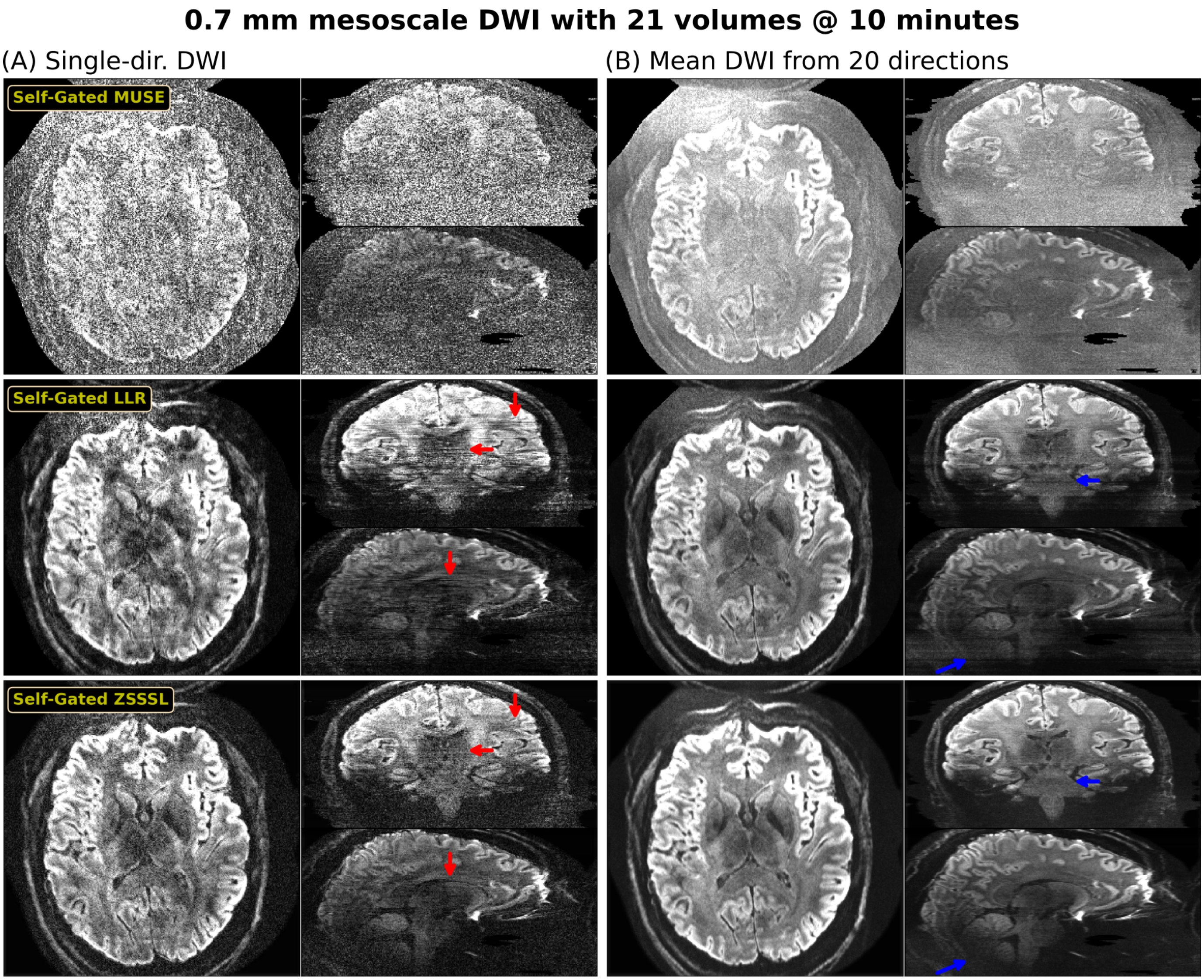

Fig. 4. 0.7 mm isotropic mesoscale DWI with 21 volumes at about 10 minutes without the acquisition of navigators. (A) Single-direction diffusion-weighted images and (B) Mean diffusion-weighted images of 20 diffusion directions at three orthogonal orientations are displayed. Diffusion-weighted images were reconstructed by (top) MUSE, (middle) LLR, and (bottom) ZSSSL. MUSE suffers from severe noise artifacts at such small voxel size. LLR is able to clean up most of the noise, but is still hampered by signal void artifacts in the axial view, which appears as stripping artifacts in coronal and sagittal views (as indicated by red arrows). ZSSSL significantly reduces both noise and signal voids. In the mean diffusion-weighted images, LLR shows amplified noise in the cerebellum region, whereas ZSSSL yields more homogeneous signal distributions.

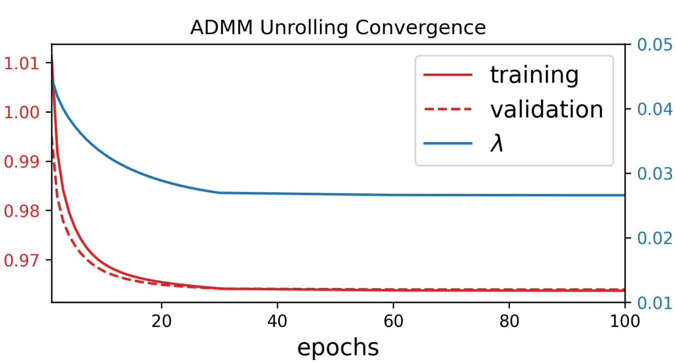

Fig. 5. Convergence analysis along the ADMM unrolling training and validation epochs for the ZSSSL results in Figure 4. Displayed curves are (red solid) the training loss, (red dashed) the validation loss, and (blue solid) the learned regularization strength $\boldsymbol{\lambda}$, respectively.

thickness are more susceptible to shot-to-shot phase variations. To enable sub-millimeter mesoscale DWI, Setsompop et al. [45] proposed the gSlider technique with slice phase–dither encoding, which excites one slab multiple times with complementary slice encoding schemes. gSlider has been proven effective in alleviating motion sensitivity, because the thicker slab (in comparison to the thin single slice) reduced inter-slice motion. Meanwhile, Hadamard encoding of the slices within a slab gained SNR in the linear inverse reconstruction. However, it has been reported that gSlider has stricter requirements on B_0 and B_1 field homogeneity and shows residual slab boundary artifacts [46]. In contrast, the proposed self-gated ZSSSL method requires no such advanced slab encoding, while achieves sub-millimeter resolution at a clinical feasible reconstruction time. Thus, the proposed method can be useful for the probe to high-resolution brain micro-structures in the

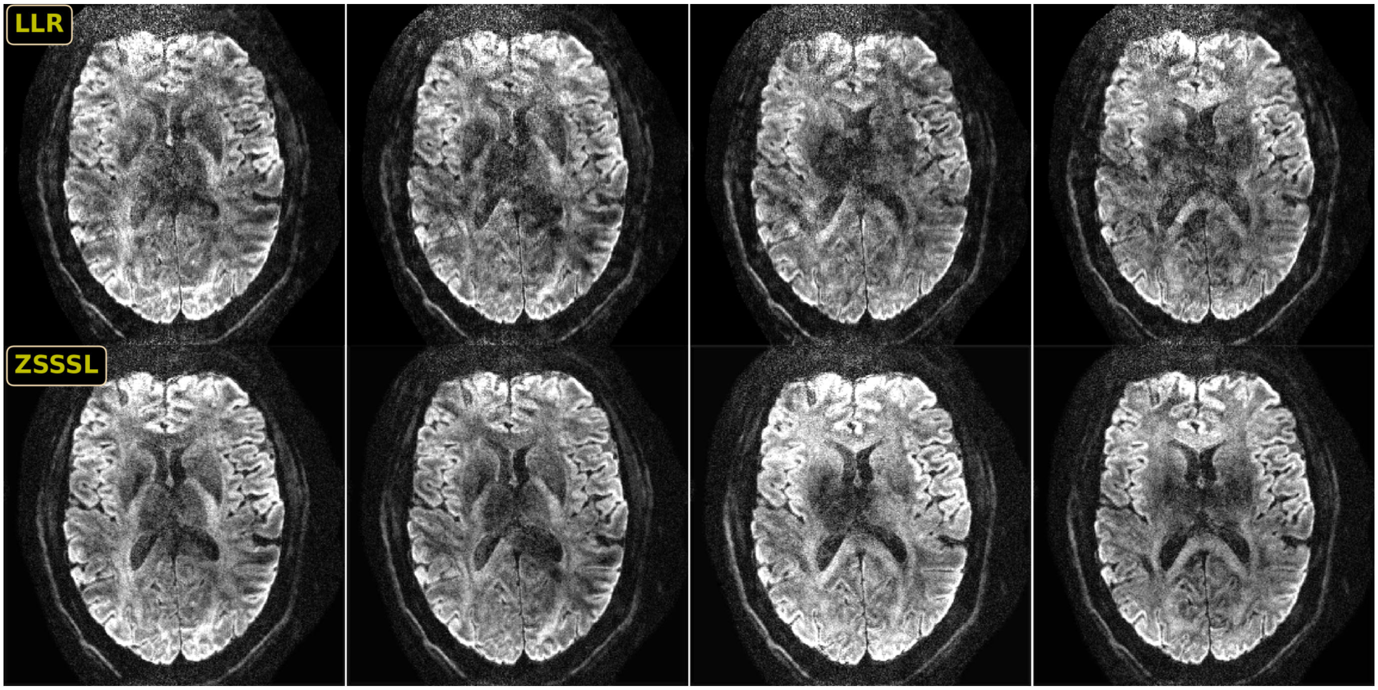

Fig. 6. Prospectively self-gated DWI reconstruction results at 0.7 mm isotropic resolution. Displayed images are one axial slice at four different diffusion-encoding directions. ZSSSL enables much cleaner delineations of diffusion contrasts than LLR.

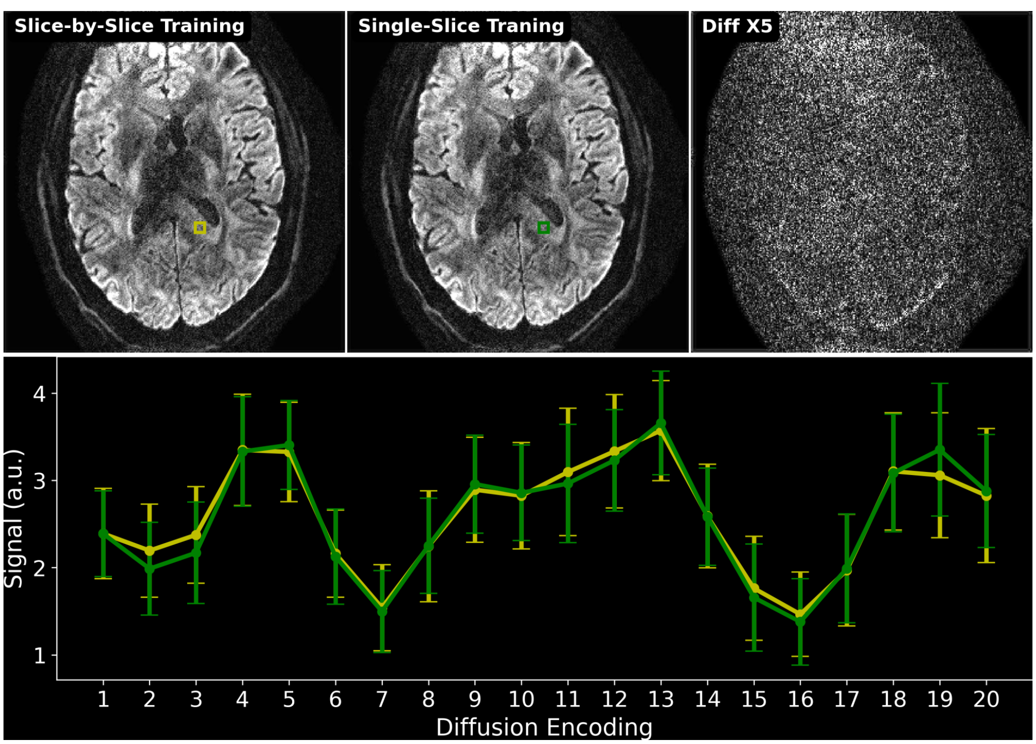

Fig. 7. Comparison of two training strategies: (1) slice-by-slice training, where every slice is trained and tested individually; (2) single-slice training, where the ZSSSL model is trained on only one slice and tested on all remaining slices. The top-right image shows the absolute difference between the reconstructed DWIs at the 10th diffusion direction between (1) and (2). The bottom panel plots the mean and standard deviation of the signal within yellow and green rectangles in the slice-by-slice training and the single-slice training, respectively. No major qualitative or quantitative difference can be see between the two training strategies.

human connectome project [47].

This work demonstrated the capability of self-gated ZSSSL in reconstructing $0.7\,\mathrm{mm}$ isotropic resolution 3-shot iEPI DWI with (6×2)-fold acceleration per shot. However, we also observed that the self-gated approach failed to recover aliasing-free DWI in the case of higher acceleration factors (e.g. the $0.5\times0.5\times2.0\,\mathrm{mm}^3$ DWI data with an acceleration of 15×2 per shot). To address this issue, acquiring shot-to-shot phase navigators helps with the shot-combined DWI reconstruction [35]. Alternatively, employing optimized trajectories with a more densely-sampled k-space central region could help better estimate shot phase variations [36], [48].

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 encoding [50]. This approach can potentially be combined with the model-based reconstruction for joint reconstructions of DWIs and B_0 field maps [51].

VI. Conclusion

In this work, we proposed a self-gated zero-shot self-supervised learning reconstruction framework based on ADMM unrolling for high-resolution and motion-robust DWI.

References

[1] T. R. Brown, B. M. Kincaid, and K. Uğurbil, "NMR chemical shift imaging in three dimensions," *Proc. Natl. Acad. Sci. USA*, vol. 79, pp. 3532–3526, 1982.

[2] D. Le Bihan, E. Breton, D. Lallemand, P. Grenier, E. Cabanis, and M. Laval-Jeantet, "MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders," *Radiology*, vol. 161, pp. 401–407, 1986.

[3] K.-D. Merboldt, W. Hanicke, and J. Frahm, "Self-diffusion NMR imaging using stimulated echoes," *J. Magn. Reson.*, vol. 64, pp. 479–486, 1985.

[4] M. Doneva, P. Börnert, H. Eggers, C. Stehning, J. Sénégas, and A. Mertins, "Compressed sensing for magnetic resonance parameter mapping," *Magn. Reson. Med.*, vol. 64, pp. 1114–1120, 2010.

[5] D. Ma, V. Gulani, N. Seiberlich, K. Liu, J. L. Sunshine, J. L. Duerk, and M. A. Griswold, "Magnetic resonance fingerprinting," *Nature*, vol. 495, pp. 187–192, 2013.

[6] E. O. Stejskal and J. Tanner, "Spin diffusion measurements: Spin echoes in the presence of time-dependent field gradient," *J. Chem. Phys.*, vol. 42, pp. 288–292, 1965.

[7] P. Mansfield, "Multi-planar image formation using NMR spin echoes," *J Phys C*, vol. 10, pp. 55–58, 1977.

[8] P. B. Roemer, W. A. Edelstein, C. E. Hayes, S. P. Souza, and O. M. Mueller, "The NMR phased array," *Magn. Reson. Med.*, vol. 16, pp. 192–225, 1990.

[9] D. K. Sodickson and W. J. Manning, "Simultaneous acquisition of spatial harmonics (SMASH): Fast imaging with radiofrequency coil arrays," *Magn. Reson. Med.*, vol. 38, pp. 591–603, 1997.

[10] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger, "SENSE: Sensitivity encoding for fast MRI," *Magn. Reson. Med.*, vol. 42, pp. 952–962, 1999.

[11] K. P. Pruessmann, M. Weiger, P. Börnert, and P. Boesiger, "Adcances in sensitivity encoding with arbitrary k-space trajectories," *Magn. Reson. Med.*, vol. 46, pp. 638–651, 2001.

[12] M. A. Griswold, P. M. Jakob, R. M. Heidemann, M. Nittka, V. Jellus, J. Wang, B. Kiefer, and A. Haase, "Generalized autocalibrating partially parallel acquisitions (GRAPPA)," *Magn. Reson. Med.*, vol. 47, pp. 1202–1210, 2002.

[13] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Reson. Med.*, vol. 58, pp. 1182–1195, 2007.

[14] K. T. Block, M. Uecker, and J. Frahm, "Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint," *Magn. Reson. Med.*, vol. 57, pp. 1186–1098, 2007.

[15] Z.-P. Liang, "Spatiotemporal imaging with partially separable functions," in *4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'4)*, 2007, pp. 988–991.

[16] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM. J. Optim.*, vol. 20, pp. 1956–1982, 2010.

[17] C. Huang, C. G. Graff, E. W. Clarkson, A. Bilgin, and M. I. Altbach, "T_2 mapping from highly undersampled data by reconstruction of principal component coefficient maps using compressed sensing," *Magn. Reson. Med.*, vol. 67, pp. 1355–1366, 2012.

[18] F. Lam and Z.-P. Liang, "A subspace approach to high-resolution spectroscopic imaging," *Magn. Reson. Med.*, vol. 71, pp. 1349–1357, 2014.

[19] D. F. McGivney, E. Pierre, D. Ma, Y. Jiang, H. Saybasili, V. Gulani, and M. A. Griswold, "SVD compression for magnetic resonance fingerprinting in the time domain," *IEEE Trans. Med. Imaging*, vol. 33, pp. 2311–2322, 2014.

[20] J. I. Tamir, M. Uecker, W. Chen, P. Lai, M. T. Alley, S. S. Vasanawala, and M. Lustig, "T_2 shuffling: Sharp, multicontrast, volumetric fast spin-echo imaging," *Magn. Reson. Med.*, vol. 77, pp. 180–195, 2017.

[21] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, pp. 504–507, 2006.

[22] F. Lam, Y. Li, and X. Peng, "Constrained magnetic resonance spectroscopic imaging by learning nonlinear low-dimensional models," *IEEE Trans. Med. Imaging*, vol. 39, pp. 545–555, 2019.

[23] M. Mani, V. A. Magnotta, and M. Jacob, "qModeL: A plug-and-play model-based reconstruction for highly accelerated multi-shot diffusion MRI using learned priors," *Magn. Reson. Med.*, vol. 86, pp. 835–851, 2021.

[24] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, pp. 1–122, 2010.

[25] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *27th International Conference on Machine Learning (ICML'27)*, 2010, pp. 399–406.

[26] V. Monga, Y. Li, and Y. C. Eldar, "Algorithm Unrolling: Interpretable, Efficient Deep Learning for Signal and Image Processing," *IEEE Signal Processing Magazine*, vol. 38, pp. 18–44, 2021.

[27] K. Hammernik, T. Klatzer, E. Kobler, M. P. Recht, D. K. Sodickson, T. Pock, and F. Knoll, "Learning a variational network for reconstruction of accelerated MRI data," *Magn. Reson. Med.*, vol. 79, pp. 3055–3071, 2018.

[28] H. K. Aggarwal, M. P. Mani, and M. Jacob, "MoDL: Model-based deep learning architecture for inverse problems," *IEEE Trans. Med. Imaging*, vol. 38, pp. 394–405, 2018.

[29] B. Yaman, S. A. H. Hosseini, S. Moeller, J. Ellermann, K. Uğurbil, and M. Akçakaya, "Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data," *Magn. Reson. Med.*, vol. 84, pp. 3172–3191, 2020.

[30] B. Yaman, S. A. H. Hosseini, and M. Akçakaya, "Zero-shot self-supervised learning for MRI reconstruction," in *10th International Conference on Learning Representations (ICLR'10)*, 2022.

[31] F. Knoll, J. Zbontar, A. Sriram, M. J. Muckley, M. Bruno, A. Defazio, M. Parente, K. J. Geras, J. Katsnelson, H. Chandarana, Z. Zhang, M. Drozdzalv, A. Romero, M. Rabbat, P. Vincent, J. Pinkerton, D. Wang, N. Yakubova, E. Owens, C. L. Zitnick, M. P. Recht, D. K. Sodickson, and Y. W. Lui, "fastMRI: A Publicly Available Raw k-Space and DICOM Dataset of Knee Images for Accelerated MR Image Reconstruction Using Machine Learning," *Radiology: Artificial Intelligence*, vol. 2, p. e190007, 2020.

[32] F. Liu, R. Kijowski, G. E. Fakhri, and L. Feng, "Magnetic resonance parameter mapping using model-guided self-supervised deep learning," *Magn. Reson. Med.*, vol. 85, pp. 3211–3226, 2021.

[33] Y. Arefeen, J. Xu, M. Zhang, Z. Dong, F. Wang, J. White, B. Bilgic, and E. Adalsteinsson, "Latent signal models: Learning compact representations of signal evolution for improved time-resolved, multi-contrast MRI," *Magn. Reson. Med.*, vol. 90, pp. 483–501, 2023.

[34] P. J. Basser, J. Mattiello, and D. Le Bihan, "MR diffusion tensor spectroscopy and imaging," *Biophys. J.*, vol. 66, pp. 259–267, 1994.

[35] Z. Tan, P. A. Liebig, R. M. Heidemann, F. B. Laun, and 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*, vol. 2, pp. 1–15, 2024.

[36] C. Liu, R. Bammer, D.-h. Kim, and M. E. Moseley, "Self-navigated interleaved spiral (SNAILS): Application to high-resolution diffusion tensor imaging," *Magn. Reson. Med.*, vol. 52, pp. 1388–1396, 2004.

[37] M. Uecker, A. Karaus, and J. Frahm, "Inverse reconstruction method for segmented multishot diffusion-weighted MRI with multiple coils," *Magn. Reson. Med.*, vol. 62, pp. 1342–1348, 2009.

[38] N.-K. Chen, A. Guidon, H.-C. Chang, and A. W. Song, "A robust multi-shot scan strategy for high-resolution diffusion weighted MRI enabled by multiplexed sensitivity-encoding (MUSE)," *NeuroImage*, vol. 72, pp. 41–47, 2013.

[39] A. Merrem, S. Hofer, A. S. A. Hosseini, D. Voit, K.-D. Merboldt, Z. Tan, and J. Frahm, "Diffusion-weighted MRI of the prostate without susceptibility artifacts: Undersampled multi-shot turbo-STEAM with rotated radial trajectories," *NMR Biomed.*, vol. 32, p. e4074, 2019.

[40] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *18th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'18)*, 2015, pp. 234–241.

[41] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'16)*, 2016, pp. 770–778.

[42] A. Heydari, A. Ahmadi, T. H. Kim, and B. Bilgic, "Joint MAPLE: Accelerated joint T_1 and T_2^* mapping with scan-specific self-supervised networks," *Magn. Reson. Med.*, 2024.

[43] W.-T. Chang, K. M. Huynh, P.-T. Yap, and W. Lin, "Navigator-Free Submillimeter Diffusion MRI using Multishot-encoded Simultaneous Multi-slice (MUSIUM) Imaging," in *Proceedings of the 29th Annual Meeting of ISMRM*, 2021, p. 1322.

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

[45] K. Setsompop, Q. Fan, J. Stockmann, B. Bilgic, S. Huang, S. F. Cauley, A. Nummenmaa, F. Wang, Y. Rathi, T. Witzel, and L. L. Wald, "High-resolution in vivo diffusion imaging of the human brain with generalized slice dithered enhanced resolution: Simultaneous multislice (gSlider-SMS)," *Magn. Reson. Med.*, vol. 79, pp. 141–151, 2018.

[46] E. Dai, S. Liu, and H. Guo, "High-resolution whole-brain diffusion MRI at 3T using simultaneous multi-slab (SMSlab) acquisition," *NeuroImage*, vol. 237, p. 118099, 2021.

[47] S. Y. Huang, T. Witzel, B. Keil, A. Scholz, M. Davids, P. Dietz, E. Rummert, R. Ramb, J. E. Kirsch, A. Yendiki, Q. Fan, Q. Tian, G. Ramos-Llordén, H.-H. Lee, A. Nummenmaa, B. Bilgic, K. Setsompop, F. Wang, A. V. Avram, M. Komlosh, D. Benjamini, K. N. Magdoom, S. Pathak, W. Schneider, D. S. Novikov, E. Fieremans, S. Tounekti, C. Mekkaoui, J. Augustinack, D. Berger, A. Shapson-Coe, J. Lichtman, P. J. Basser, L. L. Wald, and B. R. Rosen, "Connectome 2.0: Developing the next-generation ultra-high gradient strength human MRI scanner for bridging studies of the micro-, meso- and macro-connectome," *NeuroImage*, vol. 243, p. 118530, 2021.

[48] E. Dai, P. K. Lee, Z. Dong, F. Fu, K. Setsompop, and J. A. McNab, "Distortion-free diffusion imaging using self-navigated Cartesian echo-planar time resolved acquisition and joint magnitude and phase constrained reconstruction," *IEEE Trans Med Imaging*, vol. 41, pp. 63–74, 2022.

[49] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," in *2nd International Conference on Learning Representations (ICLR'2)*, 2014.

[50] B. Zahneisen, M. Aksoy, J. Maclaren, C. Wuerslin, and R. Bammer, "Extended hybrid-space SENSE for EPI: Off-resonance and eddy current corrected joint interleaved blip-up/down reconstruction," *NeuroImage*, vol. 153, pp. 97–108, 2017.

[51] Z. Tan, C. Unterberg-Buchwald, M. Blumenthal, N. Scholand, P. Schaten, C. Holme, X. Wang, D. Raddatz, and M. Uecker, "Free-breathing liver fat, R_2^* and B_0 field mapping using multi-echo radial FLASH and regularized model-based reconstruction," *IEEE Transactions on Medical Imaging*, vol. 42, pp. 1374–1387, 2022.