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

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

***Abstract*— This work developed a self-gated zero-shot self-supervised learning (ZSSSL) reconstruction framework for navigator-free high-resolution diffusion-weighted imaging with undersampled multi-shot interleaved echo-planar imaging (iEPI) acquisition. ZSSSL belongs to algorithm unrolling with physics-guided data-consistency term and a learned regularization function. We unrolled the alternating direction method of multipliers (ADMM) with a residual neural network to explore the spatial-diffusion-dimension redundancy. First, we compared the proposed self-gated ZSSSL to conventional methods including parallel imaging as multiplexed sensitivity-encoding (MUSE), compressed sensing reconstruction with locally-low rank (LLR) regularization, and variational autoencoder (VAE) regularized reconstruction. ZSSSL supplied excellent reconstruction results in both 4-shot fully-sampled data and 2-shot undersampled data at 1.0 mm isotropic resolution. Second, self-gated ZSSSL was validated with both retrospectively and prospectively acquired data at 0.7 mm isotropic resolution. This approach outperformed both MUSE and LLR regularized reconstruction in terms of image sharpness and motion robustness. Although ZSSSL required up to 8 hours training time per slice, it is generalized to all other slices and its inference time was only 1 minute. Note that LLR required about 2 hours per slice. Overall, self-gated ZSSSL enables undersampled multi-shot iEPI acquisition without the need of navigators and offers sub-millimeter DWI at clinically feasible reconstruction time. The code is publicly available at: `https://github.com/ZhengguoTan/DeepDWI`.**

***Index Terms*— Diffusion-weighted imaging, Image reconstruction, Generative AI, Latent space, Self-supervised 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 R01 EB024532 and P41 EB017183. 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).

J. Glaser was with the Department of Medical Engineering, FAU, Erlangen, Germany. He is now with the Institute of Radiology, University Hospital Erlangen, FAU, Erlangen, Germany (e-mail: julius.glaser@fau.de).

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, which refers to 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, necessitates long acquisition, resulting in data vulnerable to subject motion and system imperfections, as well as high computational burden. DWI, in particular, poses challenges in the pursuit of high spatial, temporal, and angular resolution. DWI is typically acquired via the pulsed gradient spin echo 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 low spatial resolution. Plus, the interest in acquiring multiple diffusion directions for the improvement of angular resolution and for the probe to tissue microstructure increases the scan time.

Advances in parallel imaging [8]–[12] and compressed sensing [13]–[15] have enabled accelerated acquisition for HD-MRI. In particular, the low-rank model [16] has been a powerful tool in dimension reduction. Usually, 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 *prior* as regularization and faster inference than iterative reconstruction with hand-crafted regularization functions [26]. Algorithm unrolling has been introduced to accelerated MRI reconstruction and employed in various scenarios: supervised learning with fully sampled reference images [27], [28], and self-supervised learning with only undersampled data available for training [29], [30]. Noteworthy, it is rather difficult to acquire fully-sampled DWI for the training of a regularization functional. First, fully-sampled DWI requires a longer echo train in EPI, which not only elongates the scan time but also increases off-resonance-induced geometric distortion. Second, there exists a wide range of diffusion acquisition modes, thereby requiring a larger dataset than the two-dimensional imaging scenario [31]. Therefore, self-supervised learning is more appropriate 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 capability of DAE to learn DWI models, however, is open to questions. DAE is composed of sequential fully connected layers with nonlinear activation functions. This simple architecture may fail to learn complicated functions. DWI signal is such an example. The standard diffusion tensor model [34] consists of six tensor elements, and forms DWI signals based on the multiplication of exponential functions.

Contributions:

- We trained variational autoencoder (VAE) as regularization functional. VAE provides more accurate and robust representation of DWI signal.
- We unrolled ADMM to perform zero-shot self-supervised learning (ZSSSL). We incorporated self-gated shot-to-shot phase variation estimation into ZSSSL for deep diffusion-weighted imaging reconstruction.
- We enabled navigator-free high-resolution DWI with 21 diffusion encoding at $0.7\,\mathrm{mm}$ isotropic resolution and less than $10\,\mathrm{min}$ scan time.

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 ($\tilde{\mathbf{x}}$) to their corresponding k-space,

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

Here, the images $\tilde{\mathbf{x}}$ 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 ($\mathbf{P}$).

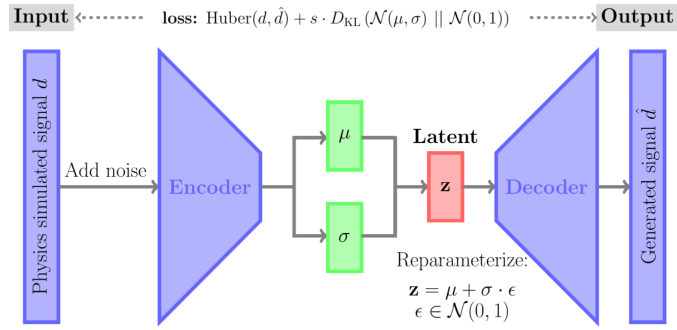

Fig. 1. The architecture of a variational autoencoder.

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 was 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 marginal 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 #2 at $0.7\,\mathrm{mm}$ isotropic resolution takes $16:27$ minutes with phase navigators, and can be reduced to about 10 minutes after the removal of phase navigators. However, as mentioned above, self gating poses a constraint on the achievable undersampling factor for parallel imaging and compressed sensing reconstruction methods (e.g., MUSE and joint reconstruction with LLR regularization).

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

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

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 $\mathcal{R}(\tilde{x})$ with the regularization strength λ. When using the Tikhonov regularization, i.e. $\mathcal{R}(\tilde{\mathbf{x}}) = \| \tilde{\mathbf{x}} \|_2^2$, Equation (2) can be solved via the conjugate gradient (CG) method. When using nonlinear regularization functions, e.g., 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. Variational Autoencoder (VAE)

Autoencoders comprise an encoder and a decoder, connected through a latent space. Conventional autoencoders have no regularization on the latent space. Consequently, the learned latent space lacks meaningful and structural representation.

To allow for dimension reduction while keeping the major part of the data structure, Kingma and Welling [40] proposed the variational autoencoder (VAE), as shown in Figure 1. In VAE, the encoder maps each diffusion-weighted signal into a Gaussian distribution ($\mathcal{N}(\mu, \sigma)$) within the latent space. The latent variable ($\mathbf{z}$) is sampled according to the encoded distribution. The decoder then maps the latent variable to the input space. The training of a VAE uses the Huber loss together with the Kullback-Leibler Divergence (KL-D). The Huber loss minimizes the difference between the input and the output, whereas KL-D minimizes the approximate posterior in latent space and the exact posterior (assumed to be Gaussian distribution).

C. Algorithm Unrolling for Deep Image Reconstruction

Algorithm unrolling has been an emerging technique in solving Equation (2) combining with deep neural networks. Algorithm unrolling consists of two ingredients. First, it learns a regularization function via deep neural networks. Second, it is constrained by the data-consistency term, i.e., the forward pass of the estimate $\mathcal{A}(\tilde{\mathbf{x}})$ must be close to the measured data $\mathbf{y}$. By mapping the operations used in iterative algorithms into networks, unrolled algorithms can be trained with data and achieve 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 [41]) 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 [42] as regularization.

D. Self-Supervised Learning for Image Reconstruction

It is difficult to acquire fully-sampled data for supervised learning in many MRI applications, e.g., dynamic imaging and diffusion-weighted imaging. To address this challenge, 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 [43].

III. METHODS

A. Data Acquisition

Table I lists two acquisition protocols implemented on a clinical 7 T MR system (MAGNETOM Terra, Siemens Healthineers, Erlangen, Germany) 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 $200\,\mathrm{T/m/s}$). The first protocol employed in-plane fully-sampled four-shot EPI and thus supplied ground truth data for the validation of our proposed methods. The second protocol implemented mesoscale $0.7\,\mathrm{mm}$ isotropic resolution based on NAViEPI or iEPI wiouth navigator with both in-plane and slice acceleration as 2, supplying a total acquisition time of $16:27$ and $9:57$ minutes, respectively. Three young healthy volunteers with written informed consent approved by the local ethics committee participated in this study.

TABLE I
NAVIEPI ACQUISITION PROTOCOLS

Protocol	#1	#2
Diffusion mode	MDDW	
Diffusion scheme	monopolar	
Diffusion direction	20	
b-value (s/mm^2)	1000	
b_0	1	
FOV (mm^2)	200	
Matrix size	200×200	286×286
In-plane resolution (mm^2)	1.0×1.0	0.7×0.7
Slice thickness (mm)	1.0	0.7
Slices	141	176
Navigator	No	Yes / No
Shots	4	3
TR (ms)	7700	15000
TE (ms)	67	58/98.3
ESP (ms)	1.02	1.17
Bandwidth (Hz/Pixel)	1086	972
Partial Fourier	6/8	5/8
Acceleration	1×3	2×2
Acquisition (min)	10 : 42	16 : 27 / 9 : 57

B. Image Reconstruction with a Learned VAE as regularization

To learn a VAE model as shown in Figure 1, we fed the VAE with diffusion-weighted signal simulated with the diffusion tensor model [34] and the employed b-value and diffusion-encoding vectors in data acquisition. The signal was augmented with added Gaussian noise and passed through the VAE network. The training loss function was defined as the sum of the Huber and and KL-D function between the simulated signal and the output of VAE.

In this work, both the encoder and the decoder in VAE consisted of 4 fully-connected linear layers. The last layer in the decoder used the sigmoid activation function, whereas the other layers used the ReLU activation function. The latent signal had 6 features, the same as the Gaussian distribution parameters μ and σ. The output features of each encoder layer gradually decreased from the number of features in simulated signal to the number of features in latent signal, whereas the input features of each decoder layer gradually increased from the number of features in latent signal to the number of features in simulated signal. After training, the VAE model was used as the regularization function in Equation (2), which was solved by ADMM in a plug-and-play manner.

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

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

ADMM updates the variables $\tilde{\mathbf{x}}$, $\mathbf{v}$, and $\mathbf{u}$ in an alternating scheme. It splits the unrolled reconstruction into three steps, as shown in Equation (3). First, the updating step for $\tilde{\mathbf{x}}$ 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 $\tilde{\mathbf{x}}$ and $\mathbf{u}$. Third, the variable $\mathbf{u}$ is updated by adding its current estimate to the difference between $\tilde{\mathbf{x}}$ and $\mathbf{v}$.

The index k in Equation (3) denotes the unrolling iteration, and $\mathcal{D}_\omega$ denotes the residual network (ResNet) [42] 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 DWIs ($\tilde{x}$) 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 $(N_Z, 2 \cdot N_{\text{diff}}, N_Y, N_X)$. 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.

D. Comparison of Regularization Techniques

In this work, based on the 4-shot fully-sampled iEPI data acquired by Protocol #1 in Table I, we compared the reconstruction performance utilizing four different regularization techniques, Tikhonov ℓ^2 regularization as in MUSE, LLR regularization, VAE 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. In other words, MUSE solves one DWI after another. On the contrary, all other regularized reconstructions fall into the joint reconstruction regime. First, the joint reconstruction jointly reconstructs all DWIs. Second, regularization terms that explore spatial-diffusion redundancy are imposed. For instance, LLR enforces low rankness of local spatial-diffusion matrices from DWIs, VAE learns a low-dimensional representation from the high-dimensional DWI signal, and ZSSSL learns a ResNet regularization function based on spatial-diffusion convolution kernels while enforcing data consistency during the unrolled training process.

E. Self-Gated ZSSSL

As mentioned in Section II-A, there exist two approaches for shot-to-shot phase variation estimation: self-gated and navigator-based. The self-gated approach as in MUSE [38] requires fully-sampled DWI acquisition and only marginal number of shots have been reported (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. NAViEPI renders shorter scan time than fully-sampled iEPI, but the use of phase navigator still elongates the acquisition, as listed in Table I. Therefore, one open question is whether it would be plausible to discard 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.

F. ZSSSL Model Generalization

We tested the ZSSSL model generalization in two aspects. First, given the 4-shot fully-sampled data acquired by Protocol #1 Table I, we trained ZSSSL with all 4 shots and then tested the trained model with retrospectively undersampled 2-shot data. Second, given the $0.7\,\text{mm}$ isotropic resolution DWI data with 88 multi-band slices acquired by Protocol #2 in Table I, we trained ZSSSL with only one slice and then performed the inference reconstruction on all other slices.

G. Computation

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

IV. Results

A. Comparison of Regularization Techniques

Figure 2 displays the regularized reconstruction results with the acquired 4-shot fully-sampled iEPI data as well as with the retrospectively 2-shot undersampled iEPI data.

First, compared to MUSE, all other regularized reconstructions (LLR, VAE, and ZSSSL) demonstrate denoising capabilities in the case of 4-shot fully-sampled iEPI. VAE shows more residual noise than LLR and ZSSSL, probably because VAE performs pixel-wise regularization, whereas LLR and ZSSSL extract local patches among DWIs in order to enforce spatial-diffusion low rankness and to perform spatial-diffusion convolution, respectively.

Second, when retrospectively undersampling the 4-shot iEPI data to 2 shots, the undersampling factor became 4×3 (4-fold in-plane undersampling and 3-fold slice undersampling). Both MUSE and VAE reconstructions show increased noise. In contrast, both LLR and ZSSSL reconstructions show better denoising performance. ZSSSL shows sharper and clearer delineation of brain tissues compared to LLR. This may be

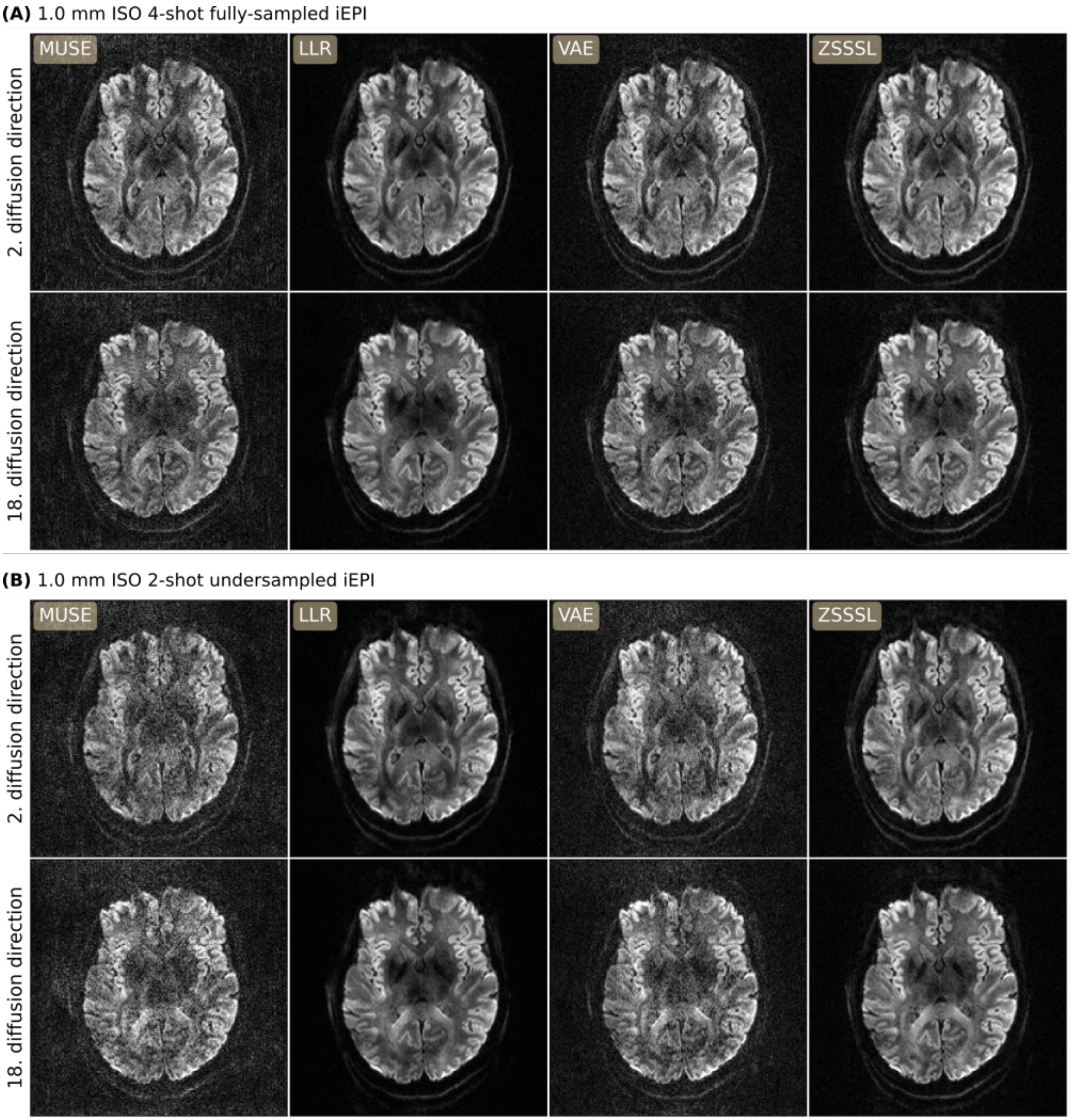

Fig. 2. **(A)** Comparison of different regularization functions on $1.0\,\mathrm{mm}$ isotropic resolution 4-shot fully-sampled iEPI DWI reconstruction: (1st column) MUSE with Tikhonov regularization, joint DWI reconstruction with (2nd column) LLR regularization, (3rd column) VAE regularization, and (4th column) ZSSSL. DWIs from the 2nd and the 18th diffusion-encoding direction were shown. **(B)** Comparison of the above regularization functions on $1.0\,\mathrm{mm}$ isotropic resolution retrospectively 2-shot undersampled iEPI DWI reconstruction.

due to the fact that the same LLR regularization strength was used for both the 4-shot and the 2-shot reconstruction. The empirically chosen regularization strength shows optimal for the 4-shot data, but results in slightly blurring artifacts for the 2-shot data. ZSSSL, on the contrary, learns the regularization strength during training and thus requires no empirical selection. Moreover, note that the LLR reconstruction took about $40\,\mathrm{min}$, whereas the training of ZSSSL lasted about $3\,\mathrm{h}$ but its inference took only $1\,\mathrm{min}$.

B. Retrospectively Self-Gated ZSSSL

Figure 3 compares LLR regularized and ZSSSL joint reconstruction. Both reconstruction methods were tested without and with the use of navigator data for shot phase estimation, respectively. Data were acquired by Protocol #2 in Table I. The 19th diffusion-encoding (without inter-shot motion) and the 11th diffusion-encoding (with inter-shot motion) reconstruction results were displayed.

As shown in Figure 3 (A), the LLR regularized joint reconstruction fails to recover diffusion-weighted images regardless of motion when performing self-gated shot phase estimation (i.e., without the use of navigator). This is because Protocol #2 employed high undersampling factor per shot 6×2, which resulted in downgraded shot phase estimation. This challenge can be mitigated with the use of navigator data for shot phase estimation. The undersampling factor of the navigator was 2×2, which renders reliable shot phase estimation in the absence of motion. This navigating approach, however, suffers from motion-induced aliasing artifacts, as shown in the bottom right panel of Figure 3 (A). This is because the inter-shot motion, when it occurs, affects only the imaging echo, but also the navigator echo.

Therefore, although the navigator approach allows for undersampled multi-shot acquisition with shot-to-shot phase variation estimated by navigator data, it prolongs the total acquisition time (as seen in Table I)) and is thus vulnerable to motion artifacts. Here, we aim to reduce the scan time via the removal of navigator acquisition and the use of ZSSSL for self-gated reconstruction. As shown in the left column of Figure 3 (B), given the self-gated shot phases (the navigator data, although acquired, was not used for shot phase estimation), ZSSSL enables motion-robust DWI reconstruction at high undersampling factor. When feeding the navigator phase into the forward model in Equation (1) for inverse reconstruction, ZSSSL also shows residual motion artifacts, as shown in the bottom right panel of Figure 3 (B). This again demonstrates that the occurrence of motion affects not only the imaging echo, but also the navigator.

Figure 4 shows diffusion-weighted images at the 11th diffusion encoding (with motion) in the coronal and the sagittal view, respectively. As mentioned in Section III-F, ZSSSL was trained using only one slice and then inferred on all other slices. The ZSSSL model generalizes well across slices. More importantly, navigated LLR reconstruction suffers from motion-induced stripping artifacts in both coronal and sagittal views [44], whereas the self-gated ZSSSL approach substantially removes such artifacts and supplies high-quality DWI without the need of navigator.

C. Prospectively Self-Gated ZSSSL

Figure 5 displays the reconstructed DWI of the 11th diffusion encoding using three different methods: MUSE, LLR, and ZSSSL. Data were acquired without navigator at $0.7\,\mathrm{mm}$ isotropic resolution. While both MUSE and LLR reconstructions suffer from undersampling- and motion-induced aliasing and striping artifacts, the proposed self-gated ZSSSL technique enables high-resolution DWI even without the need of navigators.

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. Self-gated ZSSSL achieved whole brain 21 direction diffusion encoding with the b-value of $1000\,\mathrm{s/mm^2}$ at $0.7\,\mathrm{mm}$ isotropic resolution and less than $10\,\mathrm{min}$ scan time.

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 and VAE regularization. Such an implementation assures fair comparison among different regularization methods.

While ZSSSL represents the algorithm unrolling approach to solve inverse problems, the VAE neural network was firstly trained with simulated data without the knowledge of the data consistency term in Equation (2) and then plugged in Equation (2) as the regularization function to be solved via ADMM. In this work, we observed that in the case of 2-shot undersampled data the VAE regularization performance was downgraded. First, VAE is trained by simulated dictionary data and uses fully-connected layers for every pixel. In other words, the implemented VAE architecture does not explore any spatial redundancy. On the contrary, LLR constructs local spatial-diffusion patches such as to enforce low rankness, and ZSSSL uses the 2D-convolution-based residual neural network in which the convolution window covers spatial-diffusion patches. Second, the strength of VAE lie in dimensionality reduction [21]. The data used to test regularization functions was acquired with 21 diffusion direction (Protocol #1 in Table I), which does not poses a large high-dimensional data. Therefore, it would be more valuable to apply VAE for nonlinear subspace representation of large high-dimensional data (e.g., multi-shell diffusion encoding with many directions).

The proposed self-gated ZSSSL approach is feasible for online reconstruction deployment. First, it requires much shorter acquisition time than the conventional MUSE approach with fully-sampled iEPI and our previous NAViEPI approach. Second, ZSSSL does not require large-scale fully-sampled data for training. Instead, the training of ZSSSL is scan specific. Plus, the trained ZSSSL model is applicable to different undersampling factors and to different slices. Third, the inference time of ZSSSL is much shorter than the LLR regularization approach.

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

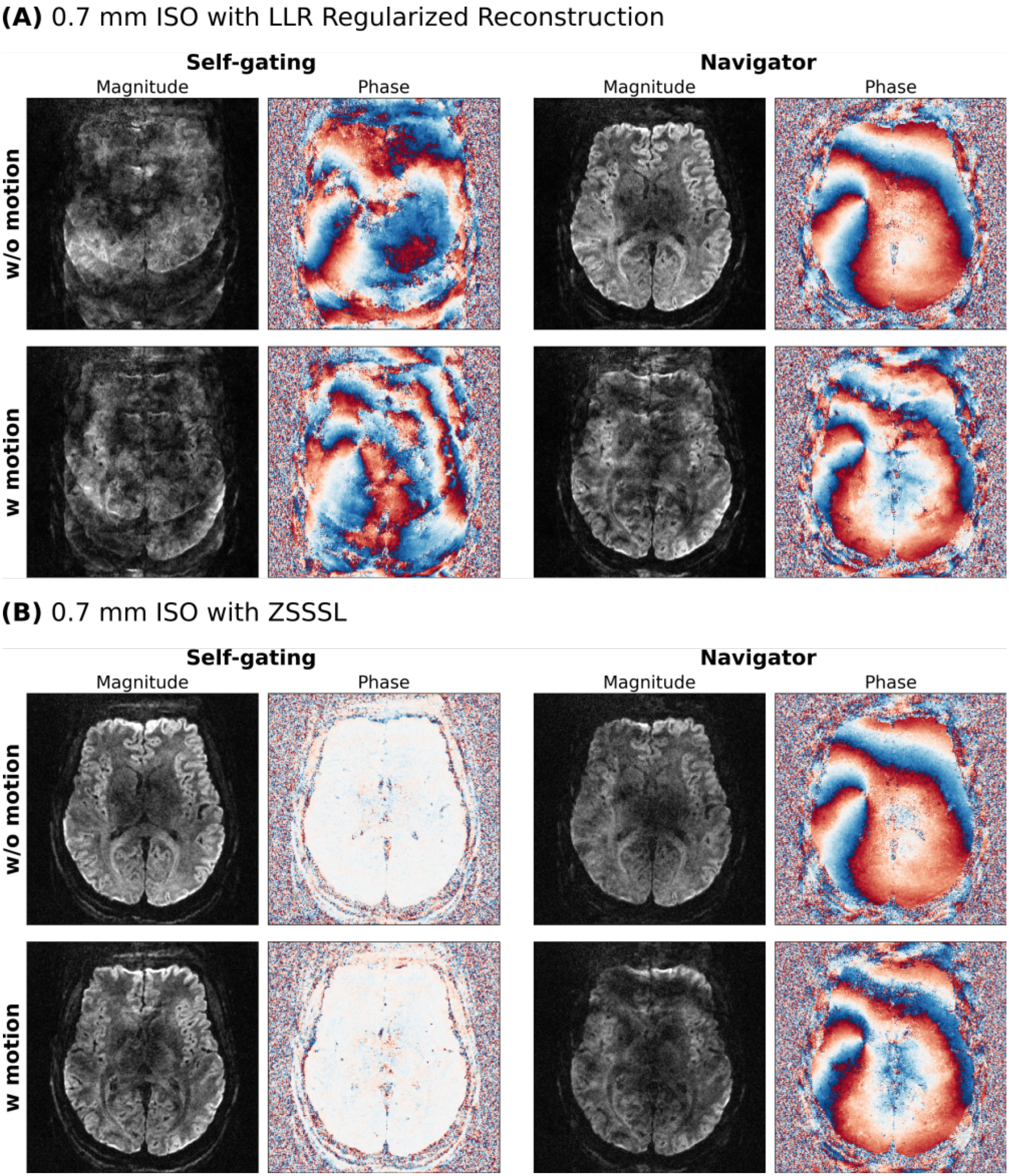

Fig. 3. $0.7\,\mathrm{mm}$ isotropic resolution DWI reconstruction results based on the NAViEPI data acquired with Protocol #2 in Table I. **(A)** LLR regularized reconstruction. **(B)** ZSSSL reconstruction. The displayed two columns from left to right are self-gated reconstruction without the use of navigator data and navigated reconstruction with the use of navigator data, respectively. In each column, the magnitude and the phase of the 19th-direction (without motion) and the 11th-direction (with motion) DWIs are displayed.

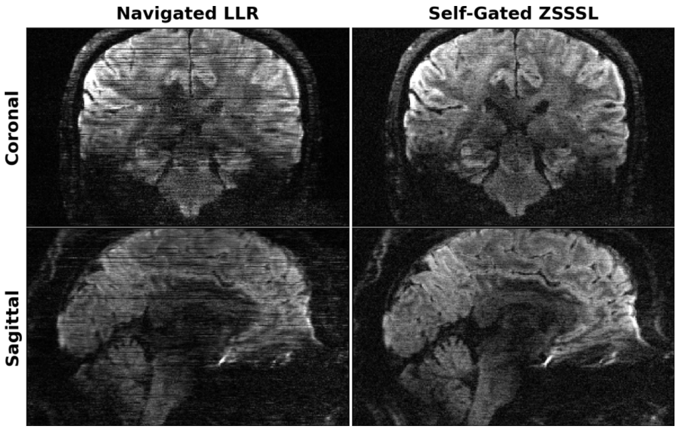

Fig. 4. $0.7\,\mathrm{mm}$ isotropic resolution DWI of the 11th diffusion encoding (with motion) in (top) the coronal and (bottom) the sagittal orientation, respectively. Displayed results are (left) the navigated LLR regularized and (right) the self-gated ZSSSL reconstruction.

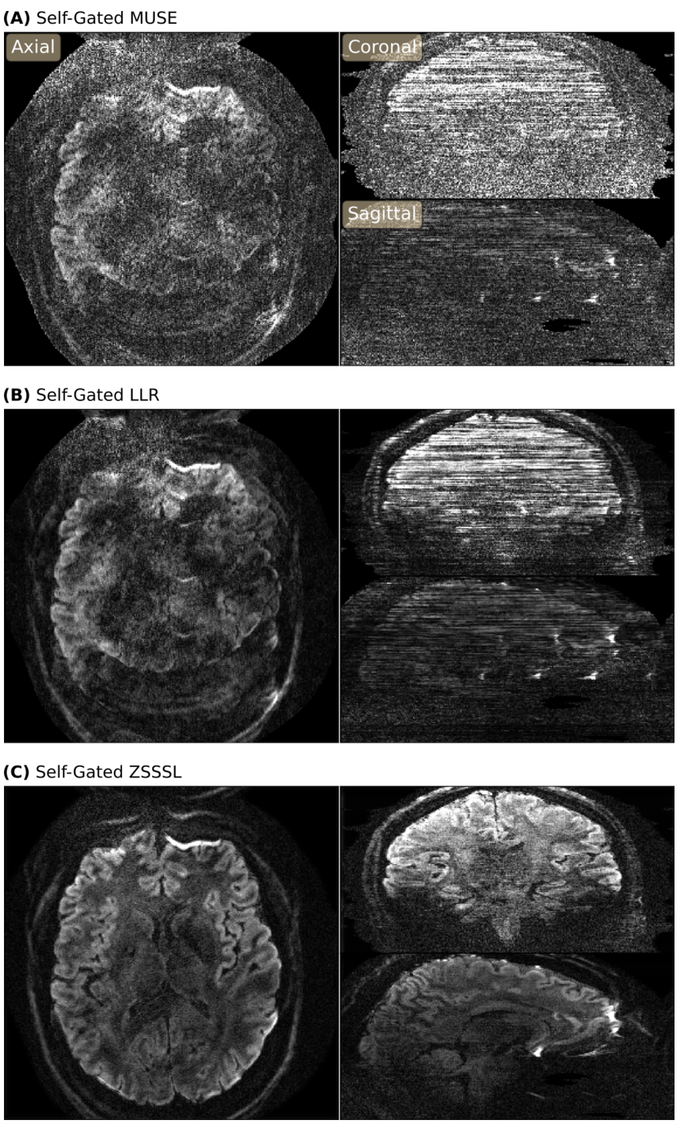

Fig. 5. $0.7\,\mathrm{mm}$ isotropic resolution DWI of the 11th diffusion encoding (with motion) without the acquisition of navigator. Displayed snapshots in the axial, the coronal and the sagittal views were reconstructed by **(A)** self-gated MUSE, **(B)** self-gated LLR, and **(C)** self-gated ZSSSL.

aliasing-free DWI in the case of higher acceleration factors (e.g. the $0.5 \times 0.5 \times 2.0\ \mathrm{mm}^3$ DWI data with an acceleration of 10×2 per shot). To overcome this problem, optimized trajectories with more densely-sampled k-space central region can be beneficial to estimate shot phase variation [45].

VI. Conclusion

In this work, we proposed a self-gated zero-shot self-supervised learning reconstruction framework for high-resolution 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] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," in *2nd International Conference on Learning Representations (ICLR'2)*, 2014.

[41] 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.

[42] 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.

[43] 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.

[44] 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.

[45] 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.