# Improved Multi-shot Diffusion-Weighted MRI with Zero-Shot Self-supervised Learning Reconstruction

Jaejin Cho[1,2]([✉]) , Yohan Jun[1,2] , Xiaoqing Wang[1,2] ,
Caique Kobayashi[1,2,3,4] , and Berkin Bilgic[1,2,5]

[1] Athinoula A. Martinos Center for Biomedical Imaging, Charlestown, MA 02129, USA

[2] Harvard Medical School, Boston, MA 02115, USA
jcho18@mgh.harvard.edu

[3] University of São Paulo, São Paulo, State of São Paulo 05508-060, Brazil

[4] Technische Universität München, 80333 München, Bavaria, Germany

[5] Harvard-MIT Health Sciences and Technology, Cambridge, MA 02139, USA

Abstract. Diffusion MRI is commonly performed using echo-planar imaging (EPI) due to its rapid acquisition time. However, the resolution of diffusion-weighted images is often limited by magnetic field inhomogeneity-related artifacts and blurring induced by T_2- and T_2^*-relaxation effects. To address these limitations, multi-shot EPI (msEPI) combined with parallel imaging techniques is frequently employed. Nevertheless, reconstructing msEPI can be challenging due to phase variation between multiple shots. In this study, we introduce a novel msEPI reconstruction approach called zero-MIRID (zero-shot self-supervised learning of Multi-shot Image Reconstruction for Improved Diffusion MRI). This method jointly reconstructs msEPI data by incorporating deep learning-based image regularization techniques. The network incorporates CNN denoisers in both k- and image-spaces, while leveraging virtual coils to enhance image reconstruction conditioning. By employing a self-supervised learning technique and dividing sampled data into three groups, the proposed approach achieves superior results compared to the state-of-the-art parallel imaging method, as demonstrated in an in-vivo experiment.

Keywords: Self-supervised learning · Multi-shot echo planar imaging · diffusion MRI

1 Introduction

Magnetic resonance imaging (MRI) is widely used for diagnosis and treatment monitoring as it provides structural and physiological information related to dis-

ease progression. Diffusion MRI (dMRI) measures molecular diffusion in biological tissues and provides microscopic details of tissue architecture, as molecules interact with many different obstacles while diffusing throughout tissues [16]. However, dMRI requires repeated acquisitions with different diffusion directions. Echo-planar imaging (EPI), which enables fast encoding per imaging slice, is commonly used for dMRI due to its fast acquisition time. However, single-shot (ss-) EPI is susceptible to severe susceptibility-induced geometric distortion and T_2- and T_2^*-induced voxel blurring. These artifacts worsen at higher in-plane resolutions as the time required to acquire each line of k-space increases approximately linearly.

Multi-shot (ms-) acquisition is an effective approach to mitigate EPI-related artifacts, which segments k-space into multiple portions covered across multiple repetition times (TRs) to reduce the effective echo spacing. However, potential shot-to-shot phase variations across multiple EPI shots can introduce additional artifacts. Recent algorithms, such as low-rank prior methods like low-rank modeling of local k-space neighborhoods (LORAKS) [7,8,14,15,17,18], and multi-shot sensitivity encoded diffusion data recovery using structured low-rank matrix completion (MUSSELS) [19], have successfully addressed this challenge by jointly reconstructing msEPI images through a low-rank constraint applied across the EPI shots.

In recent years, deep learning has emerged as a promising approach for image reconstruction, offering potential solutions to the challenges of existing techniques, including long reconstruction times, residual artifacts at high acceleration factors, and over-smoothing [6,9,10]. One notable development is model-based deep learning (MoDL), which leverages an unrolled convolutional neural network (CNN) and a parallel imaging (PI) forward model to denoise and unalias undersampled data [1]. MoDL has also been applied to multi-shot diffusion-weighted echo-planar imaging, known as MoDL-MUSSELS, effectively replacing MUSSELS and significantly reducing reconstruction times while achieving comparable results to state-of-the-art methods [2]. MoDL-MUSSELS includes CNN denoisers in both image- and k-spaces, as recent work has demonstrated that utilizing both domains has yielded improvement in performance based on metrics such as PSNR and SSIM [6]. However, it is worth noting that existing deep learning networks for dMRI have typically been trained in a supervised manner, which requires a significant amount of ground truth images that are not easily acquired in EPI acquisitions.

In contrast, self-supervised learning [3,24,25] does not rely on external training data and can be used in denoising, reconstruction, quantitative mapping, and other applications. Recent advancements in zero-shot self-supervised learning (ZS-SSL) have demonstrated successful scan-specific network training without any external database [24]. This approach has shown comparable or superior results to supervised networks. However, in dMRI, where the same volume is repeatedly acquired while changing diffusion directions, ZS-SSL typically requires training separate networks for different directions, which can be impractical.

The virtual coil (VC) approach is a highly effective technique for enhancing the performance of parallel MRI [5], particularly in the case of EPI that utilizes partial Fourier acquisition. VC generates virtual coils by incorporating conjugate symmetric k-space signals from actual coils. This integration provides supplementary information for missing data points in k-space, further being useful when combined with partial Fourier acquisition. Conceptually, the utilization of VC consistently ensures an image quality equivalent to or exceeds that of the image reconstructed without VC.

In this study, we propose a novel msEPI reconstruction method called zero-MIRID (zero-shot self-supervised learning of Multi-shot Image Reconstruction for Improved Diffusion MRI). Our method jointly reconstructs msEPI data by incorporating zero-shot self-supervised learning-based image reconstruction. Our key contributions are as follows:

– We jointly reconstruct multiple-shot images using self-supervised learning.
– We train one network for all diffusion directions, which accelerates training speed and improves performance.
– We used network denoisers in both k- and image-space and employed the VC [5] to improve the conditioning of the reconstruction.
– In the in-vivo experiment, the proposed method demonstrates more robust images and better diffusion metrics than the state-of-art PI technique for dMRI.
– To our best knowledge, this study proposes the first self-supervised learning reconstruction for dMRI.

Overall, our zero-MIRID method offers a promising approach to enhance msEPI reconstruction in dMRI, providing improved image quality and diffusion metrics through the integration of self-supervised learning techniques.

2   Method

2.1   PI Techniques for dMRI

For msEPI data, SENSE is commonly used for image reconstruction. SENSE individually reconstructs each shot's data using the spatial variation of the coil sensitivity profile. The mth shot image in the dth diffusion direction, $x_{d,m}$, can be reconstructed as follow.

$$x_{d,m} = \operatorname*{argmin}_{x_{d,m}} \|\mathcal{F}_m \mathbf{C} x_{d,m} - b_{d,m}\|_2^2 \tag{1}$$

where $\mathcal{F}_m$ is the undersampled Fourier transform for the mth shot, C is the coil sensitivity map, and $b_{d,m}$ is the acquired k-space data of dth direction and m^{th} shot.

On the other hand, MUSSELS and LORAKS jointly reconstruct multiple-shot images using the low-rank property among msEPI data. The images in the dth diffusion direction can be reconstructed using LORAKS as follows.

$$x_d = \underset{x_d}{\text{argmin}} \sum_{m=0}^{M} \|\mathcal{F}_m \mathbf{C} x_{d,m} - b_{d,m}\|_2^2 + \lambda \mathcal{J}(\mathcal{F}x_d) \tag{2}$$

where $\mathcal{J}$ is the LORAKS regularization. In this work, we utilized S-LORAKS, which employs phase information and k-space symmetry [14,15].

2.2 Network Design
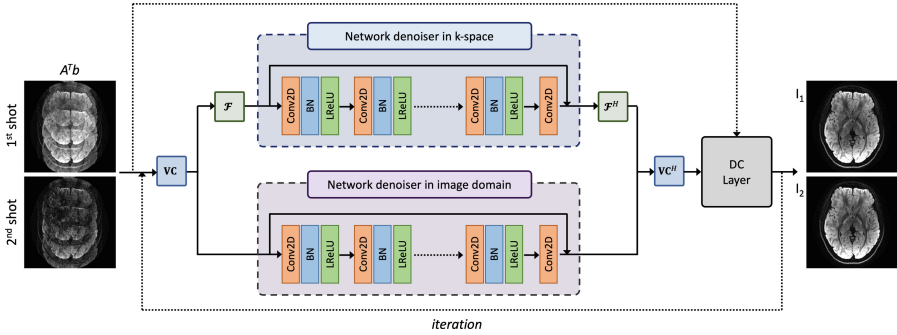
Fig. 1. The proposed image reconstruction diagram of zero-MIRID. The virtual coil (VC) layer was used to efficiently reconstruct the data accelerated by partial Fourier. The network denoisers in both the k-space and image domain were used. The DC layer enforces the consistency between the acquired data and the reconstructed images.

Figure 1 shows the proposed network diagram of zero-MIRID. The input of the network is $A_m^T b_d$, where $A_m = \mathcal{F}_m C$. The network consists of two CNNs in the k- and image-spaces. The VC was added and removed before and after the denoising CNNs, respectively. The images in the dth diffusion direction can be jointly reconstructed using zero-MIRID as follows.

$$\begin{aligned} x_d =& \underset{x_d}{\text{argmin}} \sum_{m=0}^{M} \|\mathcal{F}_m \mathbf{C} x_{d,m} - b_{d,m}\|_2^2 \\ &+ \lambda_1 \left\|\mathcal{V}_C^H N_i \mathcal{V}_C x_d\right\|_2^2 + \lambda_2 \left\|\mathcal{V}_C^H \mathcal{F}^H N_k \mathcal{F} \mathcal{V}_C x_d\right\|_2^2 \end{aligned} \tag{3}$$

where $\mathcal{V}_C$ is the VC operator, and N_i and N_k are denoising CNNs in the image- and k-space, respectively. We define $Nx = x - Dx$, where D is the CNN network, and modified the alternating minimization-based solution in [2] to get the solutions of equation (3), as follows.

$$\begin{aligned} x_{n+1} =& (A^H A + \lambda_1 I + \lambda_2 I)(A^H b + \lambda_1 \eta_n + \lambda_2 \zeta_n) \\ \zeta_{n+1} =& \mathcal{V}_C^H \mathcal{F}^H D_k \mathcal{F} \mathcal{V}_C x_{n+1} \\ \eta_{n+1} =& \mathcal{V}_C^H D_i \mathcal{V}_C x_{n+1} \end{aligned} \tag{4}$$

where n is the optimization step (iteration) number, η and ζ is the network denoising terms in k- and image-space, and $A = \mathcal{F}C$.

2.3   Zero-Shot Self-supervised Learning
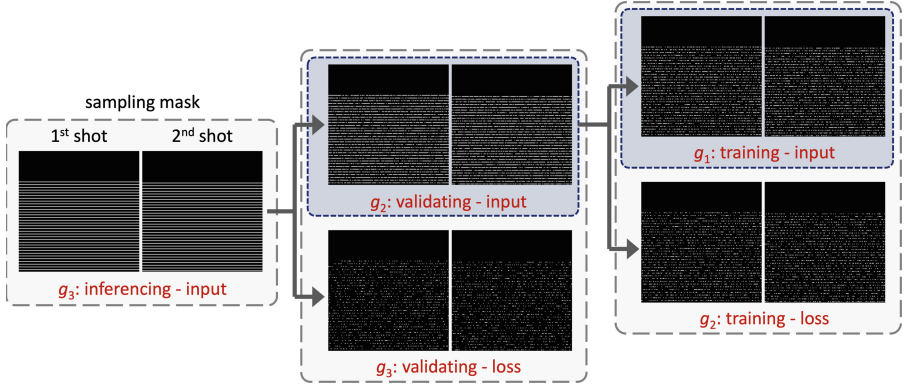
Fig. 2. The masks used for training, validation, and inference phases. The sampling mask was split into three different masks.

As proposed in the recent ZS-SSL study [24], we split the sampling mask into three different groups, as shown in Fig. 2, where g_3 is the entire sampling mask and $g_3 \supset g_2 \supset g_1$. In the training phase, g_1 was used for network input, while g_2 was used to calculate training losses. In the validating phase, g_2 was used for network input, while g_3 was used to calculate validating losses. In the inferencing phase, g_3 was used for network inputs. The loss in the dth direction in the training phase can be described as follows.

$$\mathcal{L}(g_2 \cdot b_d, \ g_2 \cdot Af(g_1 \cdot b_d; \theta)) \tag{5}$$

where $\mathcal{L}$ is the loss function, f is the zero-MIRID reconstruction, and θ is the trainable network parameters. Similarly, the loss in the dth direction in the validating phase can be described as follows.

$$\mathcal{L}(g_3 \cdot b_d, \ g_3 \cdot Af(g_2 \cdot b_d; \theta)) \tag{6}$$

In this study, we used the normalized root mean square error (NRMSE) and normalized mean absolute error (NMAE) as the loss functions.

2.4   Experiment Details

In-vivo experiments were conducted on a 3T Siemens Prisma system with a 32-channel head coil. For dMRI, we acquired the diffusion-weighted data in 32

different directions using 2-shot EPI, with each shot accelerated by 5-fold (R=5) and employing 75% Partial Fourier, resulting in 15% coverage of the k-space in each shot relative to a fully-sampled readout. Imaging parameters are; field of view (FOV)=224 $\times$ 224 $\times$ 128 mm^3, voxel size =1 $\times$ 1 $\times$ 4 mm^3, TR=3.5 s, and effective echo time (TE) =59 ms.

SENSE and S-LORAKS reconstructions were performed with MATLAB R2022a using Intel Xeon 6248R and 512 GB RAM. All neural network implementations were conducted with Python, using the Keras library in Tensorflow 2.4.1. NVIDIA Quadro RTX 8000 (RAM: 48 GB) was used to train, validate, and test the network. The denoising CNNs consist of 16 layers of which the depth is 46. For the 16 layer-CNN, we employed a filter size of 3×3. The depth of our network is 46, resulting in a total of 583,114 trainable parameters. The DC layer takes ten conjugate gradient steps, and the reconstruction block iterates ten times, where the MoDL paper [1] has demonstrated the saturated performance. For training the model, the Adam optimizer is used with a learning rate of 1e-3. Leaky ReLU was used as the activation function. For every diffusion direction, one g_2 and 50 cases of g_1 were generated. The ratio of the number of k-space points of $g_3:g_2:g_1 = 1.00:0.80:0.48$. We trained a single network for 32 diffusion directions and used that network to reconstruct all directions. For comparison, we trained two separate networks for the individual reconstruction for each shot (zero-SIRID, single-shot image reconstruction). We used the FSL toolbox for diffusion analysis [13,22,23]. To estimate multiple fiber orientations, we used the Bayesian Estimation of Diffusion Parameters Obtained using Sampling Techniques (BEDPOSTX) [4,11,12].

Example data and code can be found in the following link:
https://github.com/jaejin-cho/miccai2023

3   Results

Figure 3 the reconstructed diffusion-weighted images at 5-fold acceleration per shot in the selected diffusion directions. The reference images were obtained from 5-shot EPI data that covers complementary k-space lines to each other with the S-LORAKS constraint. While SENSE shows severe noise amplification and remaining folding artifacts, zero-SIRID was able to partially mitigate the noise amplification. S-LORAKS jointly reconstructed two shots, considerably reduced noise, and improved the signal-to-noise ratio (SNR). Nonetheless, in the selected diffusion directions, folding artifacts were amplified, and the center of the image shows a dropped signal (pointed by yellow arrows). In contrast, zero-MIRID demonstrated robust image reconstruction even with a high reduction factor per shot. The NRMSE and NAE across the diffusion direction are provided in the supplementary material, demonstrating notable reductions in NRMSE and NMAE when the proposed method is compared to S-LORAKS.

Figure 4 presents the average diffusion-weighted image (DWI), fractional anisotropy (FA) map, and 2nd crossing fiber image calculated from the reconstructed images. S-LORAKS and zero-MIRID produced high-fidelity average
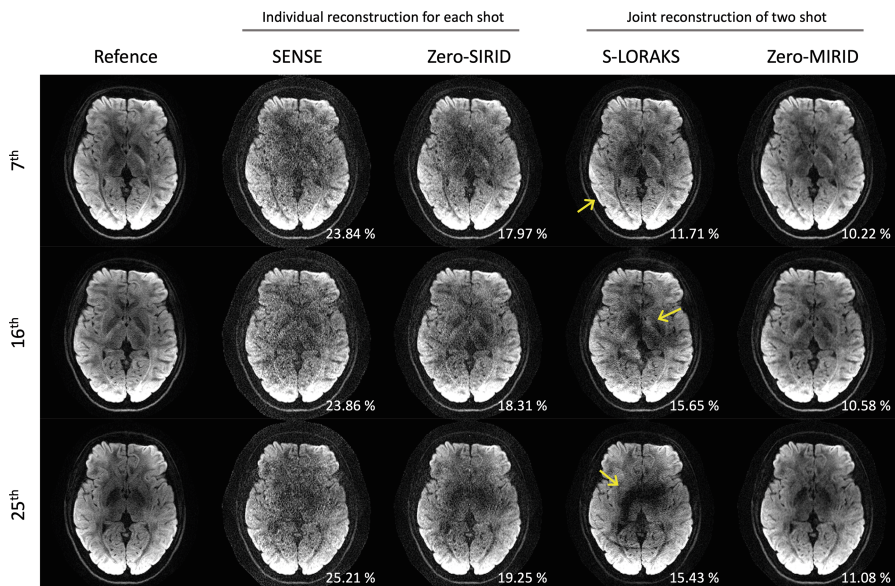
Fig. 3. The reconstructed diffusion-weighted images at R=5 per shot. Selected diffusion directions were shown. Reference images were obtained from 5-shot EPI data with S-LORAKS reconstruction. SENSE and zero-SIRID individually reconstruct each shot image, whereas S-LORAKS and zero-MIRID jointly reconstruct two shot images. NRMSE was shown at the bottom of each image.
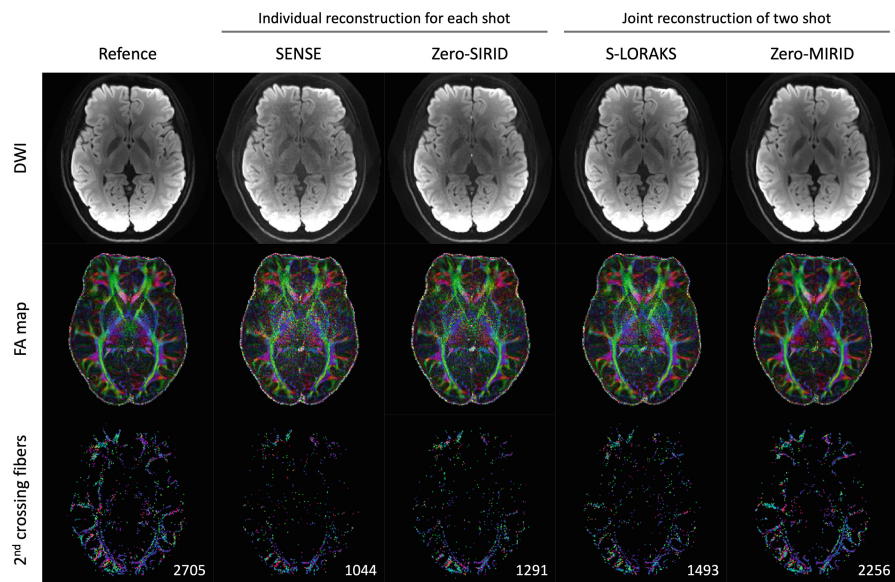
Fig. 4. Average DWI, FA map, and 2nd crossing fiber image from the reconstructed images in Fig. 3. The number of 2nd crossing fibers was shown at the bottom of each column.

DWIs, whereas SENSE and zero-SIRID show remaining artifacts. SENSE, zero-SIRID, and S-LORAKS show amplified noise in the center of the FA maps, whereas zero-MIRID effectively mitigated the noise. Furthermore, zero-MIRID well preserved the number of 2nd crossing fibers, often considered a crucial factor in evaluating successful dMRI acquisition [4,12].

4   Discussion and Conclusion

In this study, we proposed an improved image reconstruction method for msEPI and dMRI in a self-supervised deep learning manner. In-vivo experiment demonstrates the proposed method outperformed S-LORAKS, the state-of-art PI method for dMRI.

Acquiring reference images of msEPI can be challenging because each shot is typically highly accelerated and shot-to-shot phase variation prevents jointly reconstructing multiple shots efficiently. Advanced PI techniques that jointly reconstruct many EPI shots can improve the PI condition and provide high-fidelity images, but using a PI method may induce bias to that particular method. Therefore, supervised learning might not be an ideal solution for msEPI. On the other hand, self-supervised learning, which does not require reference images, could be a more suitable approach for msEPI. Due to the difficulty in obtaining reliable ground truth data, conventional quantitative metrics such as SSIM and NRMSE may be less reliable for evaluation. In dMRI, FA maps and 2nd crossing fibers could be used for obtaining more suitable metrics.

We trained a single network for all diffusion directions, which improved performance and reduced training time (please see the supplementary material). NRMSE and NMAE were reduced from 14.69% to 13.61% and from 15.73% to 14.41%, respectively. The training time for the proposed network was 22:30 min per diffusion direction/slice (on GPU). This is expected to be reduced by transfer learning. Inference took approximately 1 s per direction/slice, and 2-shot LORAKS took approximately 20 s per direction/slice (on CPU). Since the images are highly similar across diffusion directions, training on the entire diffusion direction has a similar effect to increasing the size of the training database, thereby enhancing network training. Moreover, using a single network for all directions reduces training time compared to training separate networks for each direction, from 40:01 min to 22:30 min per diffusion direction and slice.

As a future work, the simultaneous multi-slice (SMS) technique [21], which is often used for further acceleration, can be easily incorporated into the current network (please see the preliminary images in the supplementary material). At R_{sms}=5 × 2-fold acceleration, NRMSE and NMAE were significantly reduced compared with SENSE, from 22.91% to 9.07% and from 26.09% to 11.12%, respectively. g-Slider could be a good application as well [20], because RF-encoded images also have highly similar image features.

References

1. Aggarwal, H.K., Mani, M.P., Jacob, M.: MoDL: model-based deep learning architecture for inverse problems. IEEE Trans. Med. Imaging **38**(2), 394–405 (2019)
2. Aggarwal, H.K., Mani, M.P., Jacob, M.: MoDL-MUSSELS: model-based deep learning for multishot Sensitivity-Encoded diffusion MRI. IEEE Trans. Med. Imaging **39**(4), 1268–1277 (2020)
3. Akçakaya, M., Moeller, S., Weingärtner, S., Uğurbil, K.: Scan-specific robust artificial-neural-networks for k-space interpolation (RAKI) reconstruction: database-free deep learning for fast imaging. Magn. Reson. Med. **81**(1), 439–453 (2019)
4. Behrens, T.E., Berg, H.J., Jbabdi, S., Rushworth, M.F., Woolrich, M.W.: Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? neuroimage **34**(1), 144–155 (2007)
5. Blaimer, M., Gutberlet, M., Kellman, P., Breuer, F.A., Köstler, H., Griswold, M.A.: Virtual coil concept for improved parallel MRI employing conjugate symmetric signals. Mag. Resonance Med.: Off. J. Int. Society Mag. Resonance Med. **61**(1), 93–102 (2009)
6. Eo, T., Jun, Y., Kim, T., Jang, J., Lee, H.J., Hwang, D.: Kiki-net: cross-domain convolutional neural networks for reconstructing undersampled magnetic resonance images. Magn. Reson. Med. **80**(5), 2188–2201 (2018)
7. Haldar, J.P.: Low-rank modeling of local k-space neighborhoods (LORAKS) for constrained MRI. IEEE Trans. Med. Imaging **33**(3), 668–681 (2014)
8. Haldar, J.P., Zhuo, J.: P-LORAKS: Low-rank modeling of local k-space neighborhoods with parallel imaging data. Magn. Reson. Med. **75**(4), 1499–1514 (2016)
9. Hammernik, K., et al.: Learning a variational network for reconstruction of accelerated mri data. Magn. Reson. Med. **79**(6), 3055–3071 (2018)
10. Han, Y., Sunwoo, L., Ye, J.C.: k-space deep learning for accelerated MRI. IEEE Trans. Med. Imaging **39**(2), 377–386 (2019)
11. Hernändez, M., et al.: Accelerating fibre orientation estimation from diffusion weighted magnetic resonance imaging using Gpus. PLoS ONE **8**(4), e61892 (2013)
12. Jbabdi, S., Sotiropoulos, S.N., Savio, A.M., Graña, M., Behrens, T.E.: Model-based analysis of multishell diffusion MR data for tractography: how to get over fitting problems. Magn. Reson. Med. **68**(6), 1846–1855 (2012)
13. Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M.: Fsl. Neuroimage **62**(2), 782–790 (2012)
14. Kim, T.H., Setsompop, K., Haldar, J.P.: Loraks makes better sense: phase-constrained partial fourier sense reconstruction without phase calibration. Magn. Reson. Med. **77**(3), 1021–1035 (2017)
15. Kim, T., Haldar, J.: Loraks software version 2.0: Faster implementation and enhanced capabilities. University of Southern California, Los Angeles, CA, Tech. Rep. USC-SIPI-443 (2018)
16. Le Bihan, D., Iima, M.: Diffusion magnetic resonance imaging: what water tells us about biological tissues. PLoS Biol. **13**(7), e1002203 (2015)
17. Lobos, R.A., et al.: Robust autocalibrated structured low-rank epi ghost correction. Magn. Reson. Med. **85**(6), 3403–3419 (2021)
18. Lobos, R.A., Kim, T.H., Hoge, W.S., Haldar, J.P.: Navigator-free epi ghost correction with structured low-rank matrix models: New theory and methods. IEEE Trans. Med. Imaging **37**(11), 2390–2402 (2018)

19. Mani, M., Jacob, M., Kelley, D., Magnotta, V.: Multi-shot sensitivity-encoded diffusion data recovery using structured low-rank matrix completion (MUSSELS). Magn. Reson. Med. **78**(2), 494–507 (2017)
20. Setsompop, K., et al.: High-resolution in vivo diffusion imaging of the human brain with generalized slice dithered enhanced resolution: Simultaneous multislice (g s lider-sms). Magn. Reson. Med. **79**(1), 141–151 (2018)
21. Setsompop, K., Gagoski, B.A., Polimeni, J.R., Witzel, T., Wedeen, V.J., Wald, L.L.: Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced g-factor penalty. Magn. Reson. Med. **67**(5), 1210–1224 (2012)
22. Smith, S.M., et al.: Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage **23**, S208–S219 (2004)
23. Woolrich, M.W., et al.: Bayesian analysis of neuroimaging data in FSL. Neuroimage **45**(1), S173–S186 (2009)
24. Yaman, B., Hosseini, S.A.H., Akcakaya, M.: Zero-Shot Self-Supervised learning for MRI reconstruction. In: International Conference on Learning Representations (2022)
25. Yaman, B., Hosseini, S.A.H., Moeller, S., Ellermann, J., Uğurbil, K., Akçakaya, M.: Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data. Magn. Reson. Med. **84**(6), 3172–3191 (2020)