Zero-Shot Physics-Guided Deep Learning for Subject-Specific MRI Reconstruction

***Abstract*— Physics-guided deep learning (PG-DL) has emerged as a powerful tool for accelerated MRI reconstruction, while often necessitating a database of fully-sampled measurements for training. Recent self-supervised and unsupervised learning approaches enable training without fully-sampled data. However, a database of undersampled measurements may not be available in many scenarios, especially for scans involving contrast or recently developed sequences, necessitating new methodology for subject-specific PG-DL reconstructions. A main challenge for developing subject-specific PG-DL methods is the large number of parameters, making it prone to overfitting. Moreover, database-trained models may not generalize well to unseen measurements that differ in terms of SNR, image contrast, sampling pattern, and anatomy. In this work, we propose a zero-shot self-supervised learning approach to perform subject-specific PG-DL reconstruction to tackle these issues. The proposed approach splits available measurements for each scan into three disjoint sets. Two of these sets are used to enforce data consistency and define loss during training, while the last set is used to establish an early stopping criterion. In the presence of models pre-trained on a database, we show that the proposed approach can be adapted as subject-specific fine-tuning via transfer learning to further improve reconstruction quality.**

I. INTRODUCTION

Lengthy acquisition times in MRI remain a challenge, necessitating the use of accelerated imaging techniques. Conventional acceleration methods, such as parallel imaging [1] or compressed sensing [2] are commonly used, but the acceleration rates they achieve may be limited due to noise amplification and/or residual aliasing artifacts in the reconstruction. Recently, deep learning (DL) based methods have emerged for accelerating MRI further [3], [4]. Among DL methods, physics-guided DL (PG-DL) approaches have received attention due to their robustness and improved reconstruction quality [3]–[6], while explicitly incorporating the physics of the encoding matrix to the neural network via a procedure known as algorithm unrolling [7].

Despite its effectiveness as an accelerated MRI reconstruction strategy, a number of challenges remain for DL methods. Most training paradigms require large databases of fully-sampled data, which may not be available. Transfer learning has been proposed to enable training with smaller fully-sampled datasets [8], while unsupervised training approaches [9]–[11] have been proposed to tackle challenges associated with not having fully-sampled data in some applications. In particular, self-supervised learning via data undersampling (SSDU) has shown that similar reconstructions to supervised PG-DL can be achieved while training only on a database of undersampled measurements [9].

Although the aforementioned unsupervised learning strategies enable training from undersampled data, they still require a database for training in order to learn the large number of parameters for the neural network. However, in some MRI applications involving time-varying physiological processes, dynamic information such as time courses of signal changes, contrast-related uptake or breathing patterns may differ substantially between subjects, making it difficult to generate high-quality databases of sufficient size for the aforementioned strategies. Furthermore, database training in general brings along concerns about generalization [12], [13]. Particularly, for MRI reconstruction, this may translate to training and test dataset mismatches on image contrast, sampling pattern, SNR, vendor, and anatomy. For instance, the fastMRI transfer track challenge shows that the performance of pretrained models degrades due to distribution shift or changes in acquisition parameters at inference time [14]. Moreover, bias due to datasets lacking examples of rare and/or subtle pathologies increases the risk of generalization failure [13], [15]. A recent work [16] proposed to fine-tune a pre-trained neural network for MRI reconstruction in a subject-specific manner using transfer learning. However, it required a pre-trained model and lacked an early stopping criterion to avoid overfitting [16], [17].

In this work, we develop a zero-shot self-supervised learning strategy to enable subject-specific training of PG-DL methods and tackle the challenges associated with database training. We propose a holdout self-supervision method, where the acquired data is split into at least three disjoint sets, which are respectively used only in the PG-DL neural network, to define the training loss, and to establish an early stopping strategy to avoid overfitting. In cases where a database-trained network may be available, we also propose a subject-specific fine-tuning approach, which combines the concept of transfer learning and our zero-shot self-supervised learning to achieve improved reconstruction quality and further reduce computational complexity.

II. MATERIALS AND METHODS

A. Problem Formulation and PG-DL Reconstruction

Let $\mathbf{y}_\Omega$ be the undersampled noisy data from a multi-coil MRI system, where Ω denotes the under-sampling pattern, and $\mathbf{E}_\Omega : \mathbb{C}^{M\times N} \rightarrow \mathbb{C}^P$ be the forward encoding operator that includes a partial Fourier matrix sampling the locations in Ω and coil sensitivites [1]. The inverse problem for accelerated MRI reconstruction is given as

$$\arg\min_{\mathbf{x}} \|\mathbf{y}_\Omega - \mathbf{E}_\Omega \mathbf{x}\|_2^2 + \mathcal{R}(\mathbf{x}), \tag{1}$$

where $\mathcal{R}(\cdot)$ is a regularizer. Using standard optimization methods [7], Eq. (1) can be decoupled into two sub-problems, one involving a regularization step and the other forming a data consistency (DC) step. In PG-DL methods, such iterative optimization algorithms are unrolled for a fixed number of iterations [7]. Neural networks are used to implicitly solve the regularization subproblem, while the DC sub-problem is solved by conventional linear approaches [4].

B. Supervised Learning for PG-DL Reconstruction

In supervised PG-DL, training is performed using fully-sampled data. Let $\mathbf{y}^i_{\text{ref}}$ be fully-sampled k-space for subject i and $f(\mathbf{y}^i_\Omega, \mathbf{E}^i_\Omega; \boldsymbol{\theta})$ be the output of the unrolled network for sub-sampled k-space $\mathbf{y}^i_\Omega$, where the network is parameterized by $\boldsymbol{\theta}$. End-to-end training minimizes [4], [9]

$$\min_{\boldsymbol{\theta}} \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}(\mathbf{y}^i_{\text{ref}},\ \mathbf{E}^i_{\text{full}} f(\mathbf{y}^i_\Omega, \mathbf{E}^i_\Omega; \boldsymbol{\theta})), \tag{2}$$

where N is the number of samples in the training database, $\mathbf{E}^i_{\text{full}}$ is the fully-sampled encoding operator that transform network output to k-space and $\mathcal{L}(\cdot,\cdot)$ is a loss function.

C. Self-Supervised Learning for PG-DL Reconstruction

Unlike supervised learning, SSDU performs training without fully-sampled data by only utilizing acquired measurements [9]. In SSDU, the acquired undersampled data indices, Ω are split into two disjoint sets Θ and Λ as $\Omega = \Theta \cup \Lambda$. Here, Θ is the set of k-space locations that is used within the DC units of the PG-DL network during training, while Λ is a set of k-space locations used in the loss function. End-to-end training is performed using a loss function of the form

$$\min_{\boldsymbol{\theta}} \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}\Big(\mathbf{y}^i_\Lambda,\ \mathbf{E}^i_\Lambda\big(f(\mathbf{y}^i_\Theta, \mathbf{E}^i_\Theta; \boldsymbol{\theta})\big)\Big). \tag{3}$$

Recently, this method was also extended to a multi-mask setting, which was shown to substantially improve the performance at the cost of increased training time [18].

D. Proposed Zero-Shot Learning for PG-DL Reconstruction

SSDU uses 2-way partitioning of acquired measurements for training and defining the loss. The use of this 2-way partitioning for subject-specific learning was explored in our previous work [16]. However, it was observed that training had to be stopped early to avoid overfitting. This is similar to other single-image learning strategies, such as deep image prior (DIP) [17]. In both cases, it is difficult to develop a stopping strategy with a 2-way partition. Thus, in this work, we propose a zero-shot self-supervised learning via data undersampling (ZS-SSDU) approach for subject-specific reconstruction with a well-defined early stopping criterion.

1) ZS-SSDU Formulation and Training : ZS-SSDU extends the partitioning in SSDU to a 3-way split, which is reminiscent of the machine learning framework of using a validation set in addition to testing and training sets for hyperparameter tuning or for regularization by early stopping. We define the following partition for Ω:

$$\Omega = \Theta \sqcup \Lambda \sqcup \Gamma, \tag{4}$$

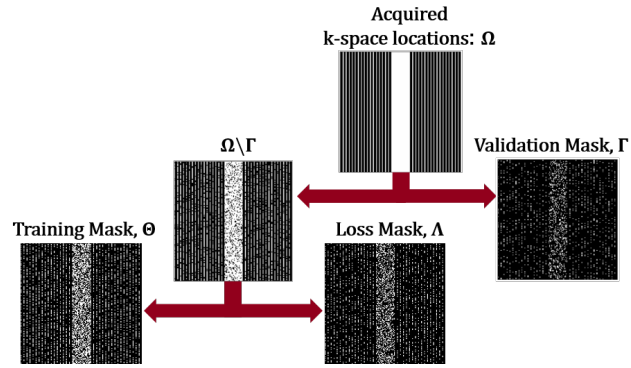

Fig. 1: ZS-SSDU splits acquired measurements into three disjoint sets, used for data consistency in the unrolled network, for defining loss, and for validation to establish an early stopping criterion.

where $\sqcup$ denotes a disjoint union, i.e. Θ, Λ and Γ are pairwise disjoint (Figure 1). Similar to Section II-C, Θ is used in the DC units of the unrolled network, and Λ is used to define the loss in k-space. The third partition Γ is a set of acquired k-space indices that are set aside for defining a k-space validation loss. The proposed zero-shot learning for MRI reconstruction uses two types of losses at each epoch during training. The training loss for ZS-SSDU can be written as:

$$\min_{\boldsymbol{\theta}} \mathcal{L}\Big(\mathbf{y}^i_\Lambda,\ \mathbf{E}^i_\Lambda\big(f(\mathbf{y}^i_\Theta, \mathbf{E}^i_\Theta; \boldsymbol{\theta})\big)\Big). \tag{5}$$

This is now supplemented by a new k-space validation loss, which tests the generalization performance of the trained network on the k-space validation partition Γ. For the lth epoch, where the learned network weights are specified by $\boldsymbol{\theta}^{(l)}$, this loss is given by:

$$\mathcal{C}^{(l)} = \mathcal{L}\Big(\mathbf{y}^i_\Gamma,\ \mathbf{E}^i_\Gamma\big(f(\mathbf{y}^i_{\Omega\setminus\Gamma}, \mathbf{E}^i_{\Omega\setminus\Gamma}; \boldsymbol{\theta}^{(l)})\big)\Big). \tag{6}$$

Note that in Eq. (6), the network output is calculated by applying the DC units on $\Omega\setminus\Gamma = \Theta \cup \Lambda$, i.e. all the acquired points outside of Γ to get a more accurate representation of its generalizability performance. The key idea is that while the training loss will decrease over epochs, the k-space validation loss will start increasing once overfitting starts to be observed. Thus, we monitor the loss in Eq. (6) during training to define an early stopping criterion to avoid overfitting. Let L represent the epoch in which the training needs to be stopped. Then at inference time, the network output is calculated as $f(\mathbf{y}_\Omega, \mathbf{E}_\Omega; \boldsymbol{\theta}^{(L)})$, i.e. all acquired points are used to calculate the final network output [9].

2) Multi-Mask Augmentation for ZS-SSDU Training : ZS-SSDU can be extended to multi-mask setting for improved performance [18], by fixing a k-space validation partition $\Gamma \subset \Omega$, and performing the multi-masking on $\Omega\setminus\Gamma$. Formally, $\Omega\setminus\Gamma$ is partitioned K times such that

$$\Omega\setminus\Gamma = \Theta_k \cup \Lambda_k, \qquad k \in \{1, \ldots, K\}, \tag{7}$$

where Λ_k, Θ_k and Γ are pairwise disjoint. This leads to the following training loss

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \frac{1}{K} \sum_{k=1}^{K} \mathcal{L}\Big(\mathbf{y}_{\Lambda_j},\ \mathbf{E}_{\Lambda_k}\big(f(\mathbf{y}_{\Theta_k}, \mathbf{E}_{\Theta_k}; \boldsymbol{\theta})\big)\Big)$$

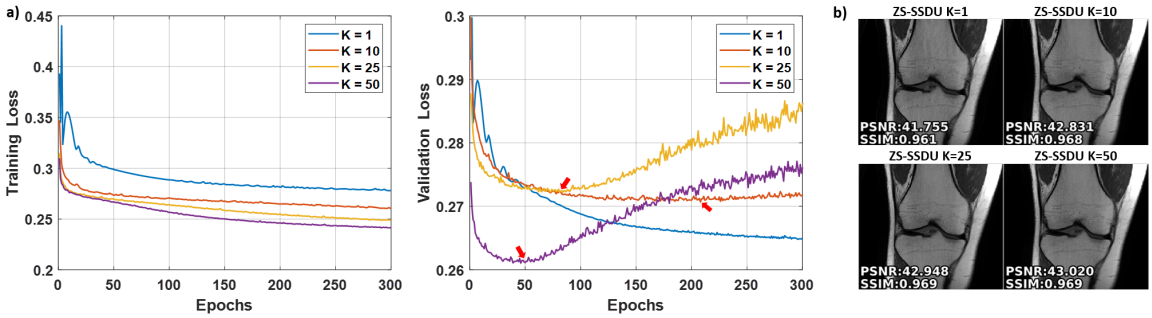

Fig. 2: a) Representative training and validation losses for ZS-SSDU with multiple $K \in \{1, 10, 25, 50\}$ masks on a coronal PD knee dataset using uniform undersampling at R=4. For $K > 1$ the validation loss forms an L-curve, whose breaking point (red arrows) dictates the early stopping criterion for training. b) Corresponding ZS-SSDU reconstruction results. At $K = 1$, without a clear stopping criterion, visible artifacts remain, highlighting the overfitting. For $K > 1$, ZS-SSDU shows good reconstruction quality without residual artifacts.

with the k-space validation loss

$$\mathcal{C}^{(l)}_{\text{m-mask}} = \mathcal{L}\Big(\mathbf{y}_\Gamma,\ \mathbf{E}_\Gamma\big(f(\mathbf{y}_{\Omega\setminus\Gamma}, \mathbf{E}_{\Omega\setminus\Gamma}; \boldsymbol{\theta}^{(l)})\big)\Big). \tag{8}$$

Note that this training strategy has a K-fold longer training time compared to that of Eq. 5.

3) ZS-SSDU Scan Specific Fine-Tuning via Transfer Learning: DL approaches are typically trained on large databases in a supervised manner. Similarly, for smaller datasets, a network pre-trained on a large database can be re-trained with supervision via transfer learning [8]. In both cases, the learned models are fixed during inference and used to reconstruct undersampled measurements that may have different acquisition parameters than the pre-trained model. Hence, the models may not generalize well during inference, leading to sub-optimal reconstructions [15]. In the presence of pre-trained models, proposed ZS-SSDU can be used to fine-tune this model using subject-specific transfer learning. This approach, referred to as ZS-SSDU-TL uses the network pre-trained on a different database as the starting point and optimizes the network parameters for the scan of interest using the objective function in Eq. 8. Furthermore, using pre-trained model weights as starting point leads to faster convergence and reduced computational time.

E. Knee MRI Datasets

Fully-sampled coronal proton density (Coronal PD) knee datasets [19] were obtained from the fastMRI database. Relevant imaging parameters: matrix size = 320×368, in-plane resolution = 0.49×0.44 mm^2. Fully-sampled knee datasets were retrospectively subsampled with an acceleration rate of 4 by keeping 24 lines of autocalibrated signal (ACS) from center of k-space using a uniform undersampling pattern.

F. Implementation Details

All PG-DL approaches were trained end-to-end using 10 unrolled iterations. Conjugate gradient method and a ResNet structure were employed in DC and regularizer units of the unrolled network, respectively [9]. Coil sensitivity maps were generated from central 24×24 ACS using ESPIRiT. End-to-end training was performed with a normalized ℓ_1-ℓ_2 loss (Adam optimizer, learning rate=$5 \cdot 10^{-4}$, batch size=1) [9].

The stopping criterion for the proposed ZS-SSDU was investigated on slices from the knee dataset. Γ was selected from the acquired measurements Ω using a uniformly random selection with $|\Gamma|/|\Omega| = 0.2$. The remaining acquired measurements $\Omega\setminus\Gamma$ were retrospectively split into disjoint 2-tuples multiple times based on uniformly random selection with the ratio $\rho = |\Lambda_k|/|\Omega\setminus\Gamma| = 0.4$ for $\forall k \in \{1, \ldots, K\}$. Network parameters for ZS-SSDU were initialized randomly with a normal distribution. The supervised PG-DL network was used for comparison purposes, as well as the starting point for ZS-SSDU-TL. CG-SENSE was also implemented for further comparisons.

III. RESULTS

Figure 2a shows representative subject-specific training and validation loss curves at R=4 of ZS-SSDU for $K \in \{1, 10, 25, 50\}$. As expected, training loss decreases with increasing epochs for all K. Validation loss for $K = 1$ decreases without showing a clear breaking point for stopping. For $K > 1$, validation loss forms an L-curve, and the breaking point of the L-curve is used as the stopping criterion. Figure 2b shows reconstructions corresponding to the K values using the proposed stopping criterion. For

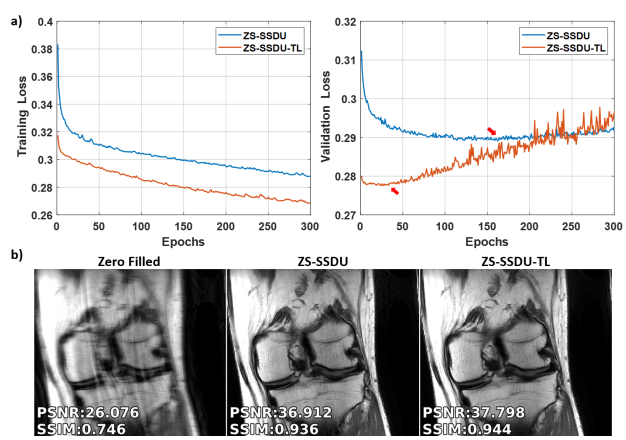

Fig. 3: a) Loss curves ZS-SSDU with/without TL for $K = 10$ on a representative slice from Coronal PD dataset. ZS-SSDU with TL converges faster compared to ZS-SSDU (red arrows). b) Reconstruction results corresponding to the loss curves. Both ZS-SSDU and ZS-SSDU-TL removes residual artifacts.

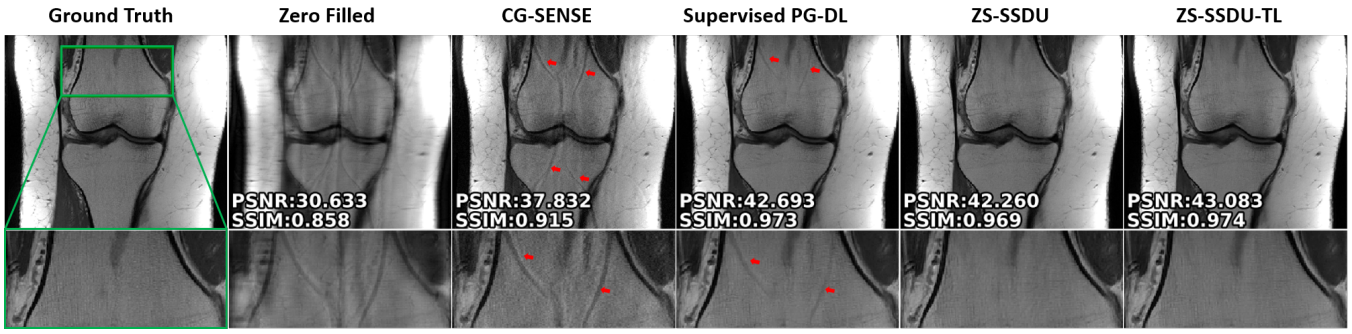

Fig. 4: Reconstruction results for a representative test slice using uniform undersampling at $R = 4$. CG-SENSE suffers from major artifacts. Supervised PG-DL improves reconstruction quality, but suffers from residual artifacts for this particular test slice. ZS-SSDU and ZS-SSDU-TL successfully reduce these artifacts, while the latter achieves the highest visual and quantitative reconstruction quality.

$K > 1$, good reconstruction quality is observed with no visible residual aliasing artifacts. For $K = 1$, without a clear breaking point for stopping criterion, reconstructions show lower visual and quantitative quality. $K = 10$ is used for the remainder of the study.

Figure 3a and b show loss curves and reconstruction results on a representative slice with and without transfer learning. As expected, ZS-SSDU-TL converges faster compared to ZS-SSDU, substantially reducing the total training time. Both ZS-SSDU and ZS-SSDU-TL remove residual artifacts, while the latter shows visually and quantitatively improved reconstruction performance.

Figure 4 shows reconstruction results for another test slice. CG-SENSE suffers from major residual artifacts. While supervised PG-DL further improves reconstruction performance, it still suffers from residual artifacts in this slice, marked by red arrows. Both ZS-SSDU and ZS-SSDU-TL achieves artifact-free reconstruction despite being trained on a single slice, while the latter expectedly achieves better quantitative metrics due to transfer learning.

IV. Conclusion

In this study, we proposed a zero-shot self-supervised PG-DL approach for subject-specific MRI reconstruction with a well-defined stopping criterion that avoids overfitting. ZS-SSDU was also used to fine-tune a pretrained model via transfer learning for improved reconstruction quality and reduced risk of generalization at a lower computational cost. Results on knee datasets showed that the proposed ZS-SSDU achieves comparable or better image quality than supervised PG-DL despite being trained on a single dataset.

References

[1] 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.

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

[3] K. Hammernik, T. Klatzer, et al., "Learning a variational network for reconstruction of accelerated MRI data," *Magn Reson Med*, vol. 79, pp. 3055–3071, 2018.

[4] F. Knoll, K. Hammernik, et al., "Deep-learning methods for parallel magnetic resonance imaging reconstruction," *IEEE Sig Proc Mag*, vol. 37, no. 1, pp. 128–140, 2020.

[5] 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, 2019.

[6] S. A. H. Hosseini, B. Yaman, S. Moeller, M. Hong, and M. Akçakaya, "Dense recurrent neural networks for accelerated MRI: History-cognizant unrolling of optimization algorithms," *IEEE J. Sel. Top. Signal Process.*, vol. 14, no. 6, pp. 1280–1291, 2020.

[7] V. Monga, Y. Li, and Y. Eldar, "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing," *arXiv preprint arXiv:1912.10557*, 2019.

[8] S. U. H. Dar, M. Özbey, A. B. Çatlı, and T. Çukur, "A transfer-learning approach for accelerated MRI using deep neural networks," *Magn Reson Med*, vol. 84, no. 2, pp. 663–685, 2020.

[9] B. Yaman, S. A. H. Hosseini, et al., "Self-supervised learning of physics-guided reconstruction neural networks without fully-sampled reference data," *Magn Reson Med*, vol. 84, pp. 3172–3191, Dec 2020.

[10] G. Oh, B. Sim, H. Chung, L. Sunwoo, and J. C. Ye, "Unpaired deep learning for accelerated MRI using optimal transport driven cycleGAN," *IEEE Trans Comp Imaging*, vol. 6, pp. 1285–1296, 2020.

[11] Z. Ke, J. Cheng, et al., "An unsupervised deep learning method for multi-coil cine MRI," *Physics in Medicine & Biology*, 2020.

[12] Y. C. Eldar, A. O. Hero III, et al., "Challenges and open problems in signal processing: Panel discussion summary from ICASSP 2017," *IEEE Sig Proc Magazine*, vol. 34, pp. 8–23, 2017.

[13] F. Knoll, T. Murrell, et al., "Advancing machine learning for MR image reconstruction with an open competition: Overview of the 2019 fastMRI challenge," *Magn Reson Med*, vol. 84, pp. 3054–3070, 2020.

[14] M. J. Muckley, B. Riemenschneider, et al., "Results of the 2020 fastmri challenge for machine learning mr image reconstruction," *IEEE Transactions on Medical Imaging*, 2021.

[15] F. Knoll, K. Hammernik, et al., "Assessment of the generalization of learned image reconstruction and the potential for transfer learning," *Magn Reson Med*, vol. 81, no. 1, pp. 116–128, 01 2019.

[16] S. A. H. Hosseini, B. Yaman, S. Moeller, and M. Akçakaya, "High-fidelity accelerated MRI reconstruction by scan-specific fine-tuning of physics-based neural networks," in *IEEE Engineering in Medicine Biology Society (EMBC)*, 2020, pp. 1481–1484.

[17] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *Proc. IEEE CVPR*, June 2018.

[18] B. Yaman, S. A. H. Hosseini, et al., "Ground-truth free multi-mask self-supervised physics-guided deep learning in highly accelerated mri," in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2021, pp. 1850–1854.

[19] F. Knoll, J. Zbontar, et al., "fastMRI: A publicly available raw k-space and DICOM dataset of knee images for accelerated MR image reconstruction using machine learning," *Radiol AI*, p. e190007, 2020.