SELF-SUPERVISED PHYSICS-BASED DEEP LEARNING MRI RECONSTRUCTION WITHOUT FULLY-SAMPLED DATA

Burhaneddin Yaman[⋆†], *Seyed Amir Hossein Hosseini*[⋆†], *Steen Moeller*[†], *Jutta Ellermann*[†], *Kâmil Uğurbil*[†] *and Mehmet Akçakaya*[⋆†]

[⋆] Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN, USA
[†] Center for Magnetic Resonance Research, University of Minnesota, Minneapolis, MN, USA

ABSTRACT

Deep learning (DL) has emerged as a tool for improving accelerated MRI reconstruction. A common strategy among DL methods is the physics-based approach, where a regularized iterative algorithm alternating between data consistency and a regularizer is unrolled for a finite number of iterations. This unrolled network is then trained end-to-end in a supervised manner, using fully-sampled data as ground truth for the network output. However, in a number of scenarios, it is difficult to obtain fully-sampled datasets, due to physiological constraints such as organ motion or physical constraints such as signal decay. In this work, we tackle this issue and propose a self-supervised learning strategy that enables physics-based DL reconstruction without fully-sampled data. Our approach is to divide the acquired sub-sampled points for each scan into training and validation subsets. During training, data consistency is enforced over the training subset, while the validation subset is used to define the loss function. Results show that the proposed self-supervised learning method successfully reconstructs images without fully-sampled data, performing similarly to the supervised approach that is trained with fully-sampled references. This has implications for physics-based inverse problem approaches for other settings, where fully-sampled data is not available or possible to acquire.

Index Terms— Self-supervised learning, accelerated imaging, parallel imaging, compressed sensing, deep learning, neural networks, supervised learning

1. INTRODUCTION

Long acquisition times remain a limitation for MRI. In most clinical protocols, a form of accelerated imaging is utilized to improve spatio-temporal resolution. In these methods, data is sub-sampled, and then reconstructed using additional information. Parallel imaging [1, 2] which utilizes redundancies between receiver coils, and compressed sensing [3, 4] which uses compressibility of images are two common approaches.

Recently, deep learning (DL) has gained interest as a means of improving accelerated MRI reconstruction [5–12]. Several approaches have been proposed, including learning a mapping from zero-filled images to artifact-free images [10], learning interpolation rules in k-space [11, 12], and a physics-based approach that utilizes the known forward model during reconstruction [6–9]. The latter approach considers reconstruction as an inverse problem, including a data consistency term that involves the forward operator and a regularization term that is learned from training data. Such methods typically unroll an iterative algorithm for a predetermined number of iterations to solve this inverse problem [13]. Specifics of these networks vary [14]. For instance, in [7], data consistency used a gradient step and regularizer was a variational network, whereas in [8] conjugate gradient was used in data consistency and a residual network as regularizer.

To the best of our knowledge, most physics-based DL-MRI reconstructions to date use supervised learning, utilizing fully-sampled data as reference for training the network. However, in several practical settings, it is not possible to acquire fully-sampled data due to physiological constraints, e.g. high-resolution dynamic MRI, or due to systemic limitations, e.g. high-resolution single-shot diffusion MRI. Furthermore, accelerated imaging is often used to improve resolution feasible, which in turn makes it infeasible to acquire fully-sampled data at higher resolution due to scan time constraints.

In this work, we propose a self-supervised learning approach that uses only the acquired sub-sampled k-space data to train physics-based DL-MRI reconstruction without fully-sampled reference data. Succinctly, our approach divides the acquired k-space points into two sets. The k-space points in the first set are used for data consistency in the network, while the k-space points in the second set are used to define the loss function. Results on knee MRI [15] show that our self-supervised training performs similar to conventional supervised training for the same network structure, while outperforming parallel imaging and compressed sensing.

2. MATERIALS AND METHODS

2.1. Physics-Based Deep Learning MRI Reconstruction

Let $\mathbf{y}_\Omega$ be the acquired data in k-space, where Ω denotes the sub-sampling pattern of acquired locations, and $\mathbf{x}$ be the im-

age to be recovered. The forward model is given as

$$\mathbf{y}_\Omega = \mathbf{E}_\Omega \mathbf{x} + \mathbf{n}, \tag{1}$$

where $\mathbf{E}_\Omega : \mathbb{C}^{M\times N} \rightarrow \mathbb{C}^P$ is the forward encoding operator, including a partial Fourier matrix and the sensitivities of the receiver coil array [1], and $\mathbf{n} \in \mathbb{C}^P$ is the measurement noise. Equation (1) is generally ill-posed and thus commonly solved using a regularized least squares problem [3, 16] as follows

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

where first term enforces data consistency with acquired measurements, and $\mathcal{R}(\cdot)$ is a regularizer. There are several strategies to solve this optimization problem [17], where methods alternate between enforcing data consistency with acquired data $\mathbf{y}_\Omega$ and a proximal operation involving $\mathcal{R}(\cdot)$. For instance, using variable-splitting and quadratic relaxation [17], we have

$$\mathbf{z}^{(i-1)} = \arg\min_{\mathbf{z}} \mu\|\mathbf{x}^{(i-1)} - \mathbf{z}\|_2^2 + \mathcal{R}(\mathbf{z}) \tag{3a}$$

$$\mathbf{x}^{(i)} = \arg\min_{\mathbf{x}} \|\mathbf{y}_\Omega - \mathbf{E}_\Omega \mathbf{x}\|_2^2 + \mu\|\mathbf{x} - \mathbf{z}^{(i-1)}\|_2^2 \tag{3b}$$

where $\mathbf{z}^{(i)}$ is an intermediate variable and $\mathbf{x}^{(i)}$ is the desired image at iteration i. This algorithm can be unrolled for a fixed number of iterations [13], as depicted in Figure 1.

Physics-based DL-MRI methods train these types of unrolled algorithms end-to-end using fully-sampled training datasets [14]. The sub-problem (3a) is implemented by means of a neural network, while the data-consistency sub-problem (3b) is solved via

$$\mathbf{x}^{(i)} = (\mathbf{E}_\Omega^H \mathbf{E}_\Omega + \mu\mathbf{I})^{-1}(\mathbf{E}_\Omega^H \mathbf{y}_\Omega + \mu\mathbf{z}^{(i-1)}), \tag{4}$$

where $\mathbf{I}$ is the identity matrix and $(\cdot)^H$ is the Hermitian operator. This can be solved via conjugate gradient to avoid matrix inversion [8]. The unrolled network is then trained end-to-end, either by allowing different parameters for each iteration [7] or by sharing all trainable parameters across iterations [8].

2.2. Supervised Unrolled Network Training

In the supervised setting, SENSE-1 images generated from fully-sampled data are often utilized as ground truth for training. Let $\mathbf{x}_{\text{ref}}^i$ denote the ground truth image for subject i. Let $f(\mathbf{y}_\Omega^i, \mathbf{E}_\Omega^i; \boldsymbol{\theta})$ denote the output of the unrolled network for sub-sampled k-space data $\mathbf{y}_\Omega^i$, and encoding matrix $\mathbf{E}_\Omega^i$ of subject i, where the network is parametersized by $\boldsymbol{\theta}$. These parameters are learned using

$$\min_{\boldsymbol{\theta}} \frac{1}{N}\sum_{i=1}^{N} \mathcal{L}(\mathbf{x}_{\text{ref}}^i,\ f(\mathbf{y}_\Omega^i, \mathbf{E}_\Omega^i; \boldsymbol{\theta})), \tag{5}$$

where N is the number of fully-sampled datasets in the training database, and $\mathcal{L}(\cdot,\cdot)$ is a loss function between the network output image and the reference image. Common choices for $\mathcal{L}(\cdot,\cdot)$ include ℓ_2 norm, ℓ_1 norm, mixed norms and perception-based loss [7, 18–22].

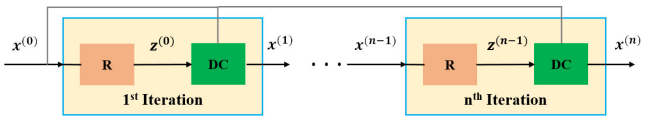

Fig. 1: The unrolled neural network architecture for solving Equation (2), where each step consists of a regularization (R) and a data consistency (DC) unit.

2.3. Proposed Self-Supervised Network Training

As discussed in Section 1, there are several scenarios where fully-sampled k-space data cannot be acquired. We propose to tackle this challenge by dividing the acquired sub-sampled data indices, Ω into two sets, Θ and Λ as

$$\Omega = \Theta \cup \Lambda, \tag{6}$$

where Θ denotes a set of k-space locations that is used within the network during training, and Λ denotes a set of k-space locations used in the loss function. In particular, in the absence of reference fully-sampled datasets, we minimize a loss function of the form

$$\min_{\boldsymbol{\theta}} \frac{1}{N}\sum_{i=1}^{N} \mathcal{L}\Big(\mathbf{y}_\Lambda^i,\ \mathbf{E}_\Lambda^i\big(f(\mathbf{y}_\Theta^i, \mathbf{E}_\Theta^i; \boldsymbol{\theta})\big)\Big). \tag{7}$$

We note that the loss function is defined between the network output image and a vector of k-space points, in contrast to the supervised case, which traditionally has a loss function defined over the image domain. We hypothesize that by calculating the loss only on Λ instead of the whole acquired data Ω, the network will be better-suited to avoid over-fitting issues and generalize to future data.

2.4. Implementation Details

Training was performed end-to-end by unrolling the algorithm in (3a)-(3b) for 10 iterations, where each iteration consisted of regularization and data consistency units. The regularization convolutional neural network (CNN) employed a ResNet structure, consisting of a layer of input and output convolution layers and 15 residual blocks (RB) with skip connections that facilitate the information flow during training [23]. Each RB comprised two convolutional layers, where the first layer is followed by a rectified linear unit (ReLU) and the second layer is followed by a constant multiplication layer [23]. All layers had a kernel size of 3×3, 64 channels. The data consistency unit used a conjugate gradient approach, which itself was unrolled for 10 iterations [8]. Coil sensitivity maps in the encoding matrices were generated using ESPIRiT [15]. The network had a total of 592,129 trainable parameters.

A normalized ℓ_1 - ℓ_2 loss, defined as

$$\mathcal{L}(\mathbf{u}, \mathbf{v}) = \frac{||\mathbf{u} - \mathbf{v}||_2}{||\mathbf{u}||_2} + \frac{||\mathbf{u} - \mathbf{v}||_1}{||\mathbf{u}||_1} \tag{8}$$

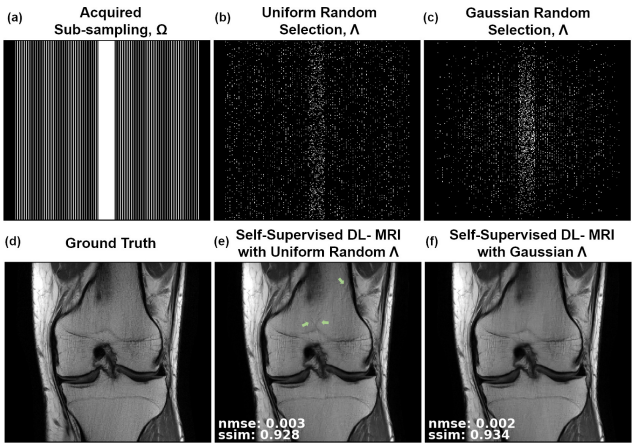

Fig. 2: a) The acquired sub-sampling pattern Ω; b) Uniform random and c) Gaussian random selection for subset Λ over which loss is defined; d) Reference fully-sampled data; e) Self-supervised DL-MRI reconstruction with Λ as in (b); f) with Λ as in (c). Green arrow marks residual artifacts in uniform random selection of Λ. These are suppressed further with Gaussian random selection.

was used for both the supervised and proposed self-supervised training. For the self-supervised setting, Θ was chosen as Ω/Λ. The networks were trained using an Adam optimizer with a learning rate of 10^{-3} by minimizing the respective loss function with a batch size of 1 over 100 epochs.

2.5. Imaging Experiments

Coronal proton density weighted knee MRI data from the New York University (NYU) fastMRI initiative database [15] was used for training and testing. Training data consisted of 300 slices from 10 patients. Each raw k-space data was of size $320 \times 368 \times 15$ where the first two dimensions are the matrix sizes and the last dimension is the number of coils. Testing was performed on 380 slices collected from 10 new subjects.

The fully sampled raw data were undersampled retrospectively using a uniform sub-sampling pattern provided by NYU with an acceleration rate of 4, and 24 lines as autocalibrated signal (ACS). The first set of experiments were designed to evaluate the choice of Λ on the proposed self-supervised training. Since Λ is a retrospectively selected subset of Ω during reconstruction, its choice is not constrained by physical limitations, such as gradient switching. Thus, Λ can be selected among all possible k-space locations in Ω. In particular, a uniform random selection of Λ, as well as a variable-density selection based on Gaussian weighting were investigated. Subsequently, the ratio $\rho = |\Lambda|/|\Omega|$ was varied among $\{0.05, 0.1, 0.2, 0.3, 0.4\}$, where $|\cdot|$ defines the cardinality of the index set.

Following the study on the choice of Λ, the proposed self-supervised learning method was compared with the conventional supervised learning approach, where the same network structure described in Section 2.4 were used for both approaches. Furthermore, the methods were compared to conjugate gradient SENSE (CG-SENSE) [24], as well as a conventional compressed sensing approach that uses a total generalized variation (TGV) [25] term for regularization in Equation (2). Experimental results were quantitatively evaluated using normalized mean square error (NMSE) and structural similarity index (SSIM).

3. RESULTS

Figure 2 shows the self-supervised network training with $\rho = 0.1$ for uniformly random and variable-density Gaussian selection of $\Lambda \subset \Omega$. The green arrows show that visible residual artifacts remained in the reconstruction using uniform random selection of Λ. These artifacts were further suppressed by selecting a variable-density Gaussian subset as Λ. The corresponding quantitative evaluations, depicted in the figure, confirm these observations.

The effect of varying $\rho \in \{0.05, 0.1, 0.2, 0.3, 0.4\}$ is shown in Figure 3. The residual artifacts, marked by green arrows, decrease with increasing ratio ρ and disappear for $\rho \in \{0.3, 0.4\}$. Quantitative results indicate that these two values have similar performance, with the latter showing slightly more visual improvement. Thus $\rho = 0.4$ was used for the rest of the study.

Figure 4 displays the comparison among the proposed self-supervised, supervised, TGV and CG-SENSE methods. The green arrows on CG-SENSE and TGV results show visible residual artifacts. The supervised and the proposed self-supervised DL-MRI reconstruction approaches eliminate these artifacts, while performing closely with each other. The SSIM and NMSE values for this slice further confirms the visual assessments. Figure 5a and b summarizes the mean and standard deviation of the NMSE and SSIM metrics over the 380 test slices using the proposed self-supervised DL-MRI, conventional supervised DL-MRI and CG-SENSE approaches. The two DL-MRI methods yield similar SSIM

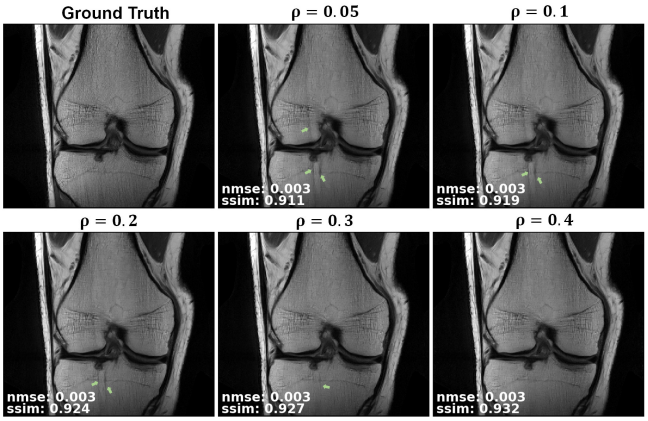

Fig. 3: A test slice reconstructed using different ratios of $\rho = |\Lambda|/|\Omega|$, where Λ is only used in the loss and Ω/Λ is only used in the data consistency unit of the network.

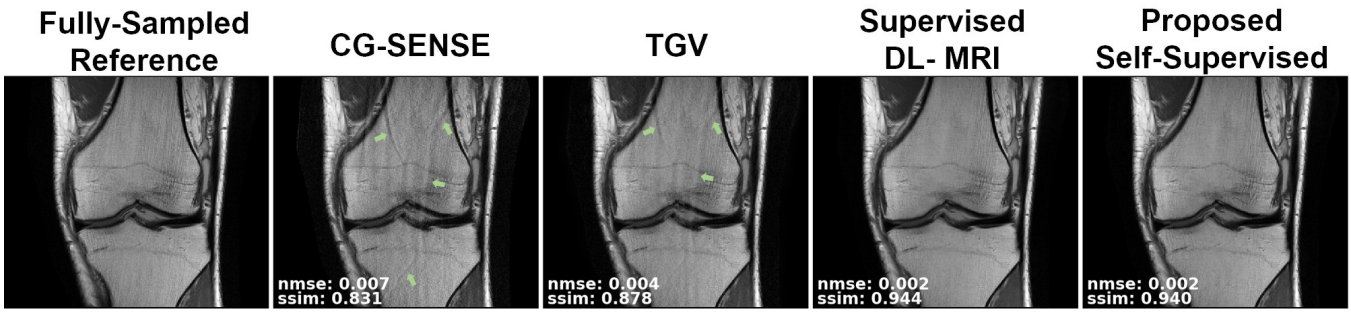

Fig. 4: A slice from a proton-density knee dataset, depicting the reference fully-sampled image, and reconstructions using CG-SENSE, TGV, supervised DL-MRI trained on fully-sampled data and the proposed self-supervised DL-MRI trained on only sub-sampled data. The proposed training strategy leads to a DL-MRI reconstruction that removes artifacts successfully and perform similar to the supervised approach, while visibly outperforming CG-SENSE and TGV approaches.

and NMSE values, even though the self-supervised approach does not use any fully-sampled datasets during training. Both methods outperform CG-SENSE.

4. DISCUSSION

We have proposed a self-supervised training method for physics-based DL-MRI reconstruction in the absence of fully-sampled data. The set of sub-sampled data indices Ω was divided into two sets Θ and Λ, where the former was used during data consistency in the unrolled network, and the latter was utilized in the loss function. Results on fastMRI dataset indicate that our approach is successful in training a reconstruction algorithm that removes aliasing artifacts, achieving comparable performance to the conventional supervised learning approach that has access to fully-sampled data, while outperforming traditional compressed sensing and parallel imaging. The same neural network architecture was used for the self-supervised and supervised training. While this is not the focus of this study, other choices of neural networks are possible for further improvement. The effect of different choices of Λ was also studied, where a variable-density selection with sufficient cardinality was favored.

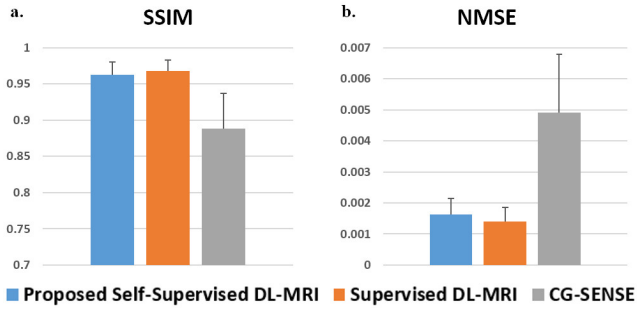

Fig. 5: a) and b) shows the SSIM and NMSE values for the 380 slices tested. Supervised DL-MRI approach performs slightly better than the proposed self-supervised DL-MRI approach, while both methods readily outperform CG-SENSE quantitatively.

In many scenarios, acquisition of fully-sampled data is challenging due to physiological and physical constraints. The lack of ground truth data hinders the utility of the supervised learning approaches in these scenarios. The proposed self-supervised approach relies only on available sub-sampled measurements. While we have focused on MRI reconstruction, the proposed approach naturally extends to other linear inverse problems, and has potential applications in other imaging modalities.

5. CONCLUSION

The proposed self-supervised learning strategy allows training of physics-based DL-MRI reconstruction without requiring fully-sampled data, while performing similar to conventional supervised learning approaches.

6. ACKNOWLEDGMENTS

This work was partially supported by NIH R00HL111410, NIH P41EB027061, NIH U01EB025144, NSF CAREER CCF-1651825. Knee MRI data were obtained from the NYU fastMRI initiative database [15]. NYU fastMRI database was acquired with the relevant institutional review board approvals as detailed in [15]. NYU fastMRI investigators provided data but did not participate in analysis or writing of this report. A listing of NYU fastMRI investigators, subject to updates, can be found at `fastmri.med.nyu.edu`.

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. A. Griswold, P. M. Jakob, et al., "Generalized autocalibrating partially parallel acquisitions (GRAPPA)," *Magn Reson Med*, vol. 47, pp. 1202–1210, 2002.

[3] 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.

[4] H. Jung, K. Sung, K. S. Nayak, E. Y. Kim, and J. C. Ye, "k-t FOCUSS: a general compressed sensing framework for high resolution dynamic MRI," *Magn Reson Med*, vol. 61, pp. 103–116, 2009.

[5] S. Wang, Z. Su, et al., "Accelerating magnetic resonance imaging via deep learning," in *Proc IEEE ISBI*, April 2016, pp. 514–517.

[6] Y. Yang J. Sun, H. Li, and Z. Xu, "Deep ADMM-Net for compressive sensing MRI," in *Advances in neural information processing systems*, 2016, pp. 10–18.

[7] 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.

[8] 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.

[9] M. Mardani, H. Monajemi, et al., "Recurrent generative adversarial networks for proximal learning and automated compressive image recovery," *arXiv preprint arXiv:1711.10046*, 2017.

[10] D. Lee, J. Yoo, S. Tak, and J. C. Ye, "Deep residual learning for accelerated MRI using magnitude and phase networks," *IEEE Trans Biomed Eng*, vol. 65, pp. 1985–1995, 2018.

[11] M. Akçakaya, S. Moeller, S. Weingärtner, and K. Uğurbil, "Scan-specific robust artificial-neural-networks for k-space interpolation (RAKI) reconstruction: Database-free deep learning for fast imaging," *Magn Reson Med*, vol. 81, pp. 439–453, 2019.

[12] Y. Han, L. Sunwoo, and J. C. Ye, "k-Space Deep Learning for Accelerated MRI," *IEEE Trans Med Imaging*, Jul 2019.

[13] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *Proc. Int. Conf. Mach. Learn*. Omnipress, 2010, pp. 399–406.

[14] D. Liang, J. Cheng, Z. Ke, and L. Ying, "Deep MRI Reconstruction," *IEEE Sig Proc Mag*, vol. 37, pp. 141–151, 2020.

[15] J. Zbontar, F. Knoll, et al., "fastMRI: An open dataset and benchmarks for accelerated MRI," *arXiv preprint arXiv:1811.08839*, 2018.

[16] S. G. Lingala, Y. Hu, E. DiBella, and M. Jacob, "Accelerated dynamic MRI exploiting sparsity and low-rank structure: k-t SLR," *IEEE Trans Med Imaging*, vol. 30, no. 5, pp. 1042–1054, May 2011.

[17] J.A. Fessler, "Optimization methods for MR image reconstruction," *IEEE Sig Proc Mag*, vol. 37, pp. 33–40, 2020.

[18] K. Hammernik, F. Knoll, D. K. Sodickson, and T. Pock, "L2 or not L2: impact of loss function design for deep learning MRI reconstruction," in *ISMRM 25th Annual Meeting*, 2017, p. 0687.

[19] J. Schlemper, J. Caballero, J. V. Hajnal, A. N. Price, and D. Rueckert, "A Deep Cascade of Convolutional Neural Networks for Dynamic MR Image Reconstruction," *IEEE Trans Med Imaging*, vol. 37, pp. 491–503, 2018.

[20] M. Mardani, E. Gong, et al., "Deep Generative Adversarial Neural Networks for Compressive Sensing MRI," *IEEE Trans Med Imaging*, vol. 38, pp. 167–179, 2019.

[21] C. Qin, J. Schlemper, et al., "Convolutional Recurrent Neural Networks for Dynamic MR Image Reconstruction," *IEEE Trans Med Imaging*, vol. 38, pp. 280–290, 2019.

[22] F. Knoll, K. Hammernik, et al., "Deep learning methods for parallel magnetic resonance image reconstruction," *IEEE Sig Proc Mag*, vol. 37, pp. 128–140, 2020.

[23] R. Timofte, E. Agustsson, L. Van Gool, M. H. Yang, and L. Zhang, "Ntire 2017 challenge on single image super-resolution: Methods and results," in *Proc IEEE CVPR*, 2017.

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

[25] F. Knoll, K. Bredies, T. Pock, and R. Stollberger, "Second order total generalized variation (TGV) for MRI," *Magn Reson Med*, vol. 65, no. 2, pp. 480–491, 2011.