Deep subspace learning for efficient reconstruction of spatiotemporal imaging data

Christopher M. Sandino[1], **Frank Ong**[1], **Siddharth S. Iyer**[1],
Adam M. Bush[2], **Shreyas S. Vasanawala**[2]
Department of Electrical Engineering[1] and Radiology[2]
Stanford University
Stanford, CA 94025
{sandino,franko,ssi,adambush,vasanawala}@stanford.edu

Abstract

Model-based deep learning approaches, such as unrolled neural networks, have been shown to be effective tools for efficiently solving inverse problems. However, the memory costs of training unrolled networks remain high, especially when the target data is high-resolution and high-dimensional. This often requires trade-offs in either network depth to reduce model size, or data resolution to reduce data size. To address this, we propose DL-Subspace - a novel unrolled network architecture which reduces memory usage by solving for a compact, low-dimensional representation of the target instead of the target itself. DL-Subspace is applied to accelerated magnetic resonance image reconstruction, demonstrating up to 4× higher memory efficiency and 4× faster inference speed while maintaining similar image quality as conventional unrolled networks.

1 Introduction

Model-based deep learning approaches have become effective tools for efficiently solving inverse problems that arise in a variety of computational imaging applications, including magnetic resonance imaging (MRI) [1, 2, 3, 4, 5]. Unrolled optimization is one such approach where a conventional optimization algorithm is unrolled to a fixed number of iterations, and its regularization steps are replaced with neural networks whose weights are learned by training the unrolled network end-to-end in a supervised fashion [6]. However, the memory costs of training unrolled networks remain high, especially when the target data is high-resolution and/or high-dimensional. This often requires trade-offs in either the number of unrolled iterations to reduce model size, or the training data resolution to reduce the input data size.

A novel unrolled network, known as DL-Subspace, is proposed to curb memory requirements by solving for a compact low-dimensional representation of the target instead of the target itself. We apply this technique to accelerated MRI reconstruction of high-resolution spatio-temporal data, and demonstrate 2-4× higher memory efficiency and 3-4× faster inference speed while maintaining similar image quality metrics as state-of-the-art unrolled neural network architectures.

2 Preliminaries

2.1 Dynamic MRI Reconstruction

Let $X \in \mathbb{C}^{N \times T}$ denote a time-series of MR images where N is the image size and T is the number of time frames. The goal of dynamic MRI reconstruction is to resolve X from a set of undersampled measurements $Y \in \mathbb{C}^{MC \times T}$ where M is the number of measurements per time frame, and C is the

NeurIPS 2021Workshop on Deep Learning and Inverse Problems, Online.

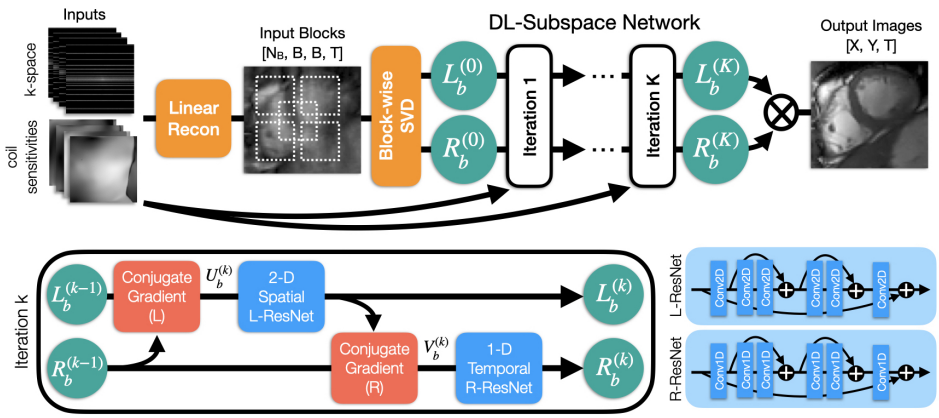

Figure 1: Linearly reconstructed images are converted into blocks and decomposed using the singular value decomposition ($X_b = U\Sigma V^H$) to initialize the block-wise basis functions as $L_b^{(0)} = U\Sigma^{1/2}$ and $R_b^{(0)} = V\Sigma^{1/2}$. These complex-valued basis functions are iteratively processed by the DL-Subspace network by alternating conjugate gradient (CG) and ResNet [7] updates. The real and imaginary part of the basis functions are treated as real-valued channels within each ResNet.

number of receivers in a phased coil array. These measurements, also known as k-space, are related to the images by a system of linear equations:

$$Y = \mathcal{A}(X) + W \tag{1}$$

where $\mathcal{A} : \mathbb{C}^{N\times T} \rightarrow \mathbb{C}^{MC\times T}$ is the MRI system matrix comprised of coil sensitivity maps, Fourier transform, and k-space sampling mask. The system noise $W \in \mathbb{C}^{MC\times T}$ is modelled by additive complex-valued white Gaussian noise. Image reconstruction is commonly performed by solving an inverse problem wherein X is estimated given explicit knowledge of Y and $\mathcal{A}$.

2.2 Subspace Model

However, solving for X can be computationally expensive, especially when X has high spatiotemporal resolution. In the past, subspace models have been used to re-formulate the dynamic MRI reconstruction problem to instead solve for the subspace of X instead of X directly [8, 9, 10, 11, 12]. Here, we consider a subspace model that exploits locally low-rank structure on image blocks of size $B \times B \times T$ [9]. Concretely, the time-series images can be split up into N_B overlapping blocks $X_b \in \mathbb{C}^{B^2\times T}$ where each block is denoted by index b. Each block can then be decomposed into a product of two matrices:

$$X_b = L_b R_b^H \tag{2}$$

where $L_b \in \mathbb{C}^{B^2\times K}$ and $R_b \in \mathbb{C}^{T\times K}$ contain K spatial and temporal basis functions respectively. The data is said to live in a low-dimensional subspace if X_b can be compactly represented using relatively few basis functions ($K << T$). The original time-series data X can be re-formed by multiplying L_b and R_b and summing over all of the blocks:

$$X = \sum_{b=1}^{N_B} \mathcal{M}_b\left(L_b R_b^H\right) \tag{3}$$

where the operator $\mathcal{M}_b : \mathbb{C}^{B^2\times T} \rightarrow \mathbb{C}^{N\times T}$ is used to embed each input block back into its original position in the full image. The new system of equations can be defined as:

$$Y = \mathcal{A}\left(\sum_{b=1}^{N_B} \mathcal{M}_b(L_b R_b^H)\right) + W. \tag{4}$$

In subspace reconstruction, a bi-linear inverse problem is solved to estimate L_b and R_b given explicit knowledge of Y and $\mathcal{A}$. Since this problem is ill-posed, it is common to enforce sparse structure on R_b such as in dictionary learning approaches [13, 14].

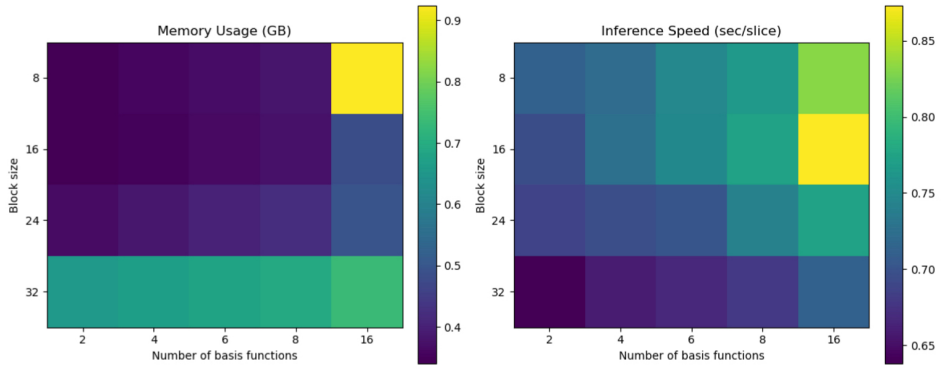

Figure 2: Memory usage (GB) and inference speed (sec/slice) benchmarks are evaluated for DL-Subspace with varying block size (B) and number of basis functions (K). For reference, memory usage during inference for DL-ESPIRiT and MoDL is 1.6 GB. Average inference speed for DL-ESPIRiT and MoDL are 2.496 and 2.800 sec/slice respectively.

3 Methods

3.1 Deep subspace learning

The dictionary learning approach is generalized by instead learning a prior on L_b and R_b in a data-driven manner by posing the following bi-linear inverse problem:

$$\underset{L_b,R_b}{\text{minimize}} \left\| Y - \mathcal{A}\left(\sum_{b=1}^{N_B} \mathcal{M}_b(L_b R_b^{(i)H})\right)\right\|_F^2 + \mu_l \sum_{b=1}^{N_B} ||\mathcal{D}_l(L_b) - L_b||_F^2 + \mu_r \sum_{b=1}^{N_B} ||\mathcal{D}_r(R_b) - R_b||_F^2 \tag{5}$$

where $\mathcal{D}_l$ and $\mathcal{D}_r$ are 2-D and 1-D convolutional neural networks (CNNs) that denoise the spatial and temporal basis functions respectively. Inspired by an unrolled alternating minimization approach by Arvinte et al. [15], this problem is broken down into simpler convex sub-problems by first minimizing with respect to L_b, then minimizing with respect to R_b, and repeating this process until convergence. As shown in Section A.2, each sub-problem can be solved using MoDL [3] to form the following update rule:

$$U_b^{(i+1)} = \arg\min_{L_b} \left\| Y - \mathcal{A}\left(\sum_{b=1}^{N_B} \mathcal{M}_b(L_b R_b^{(i)H})\right)\right\|_F^2 + \mu_l \sum_{b=1}^{N_B} ||L_b^{(i)} - L_b||_F^2 \tag{6}$$

$$L_b^{(i+1)} = \mathcal{D}_l\big(U_b^{(i+1)}\big) \tag{7}$$

$$V_b^{(i+1)} = \arg\min_{R_b} \left\| Y - \mathcal{A}\left(\sum_{b=1}^{N_B} \mathcal{M}_b(L_b^{(i+1)} R_b^{H})\right)\right\|_F^2 + \mu_r \sum_{b=1}^{N_B} ||R_b^{(i)} - R_b||_F^2 \tag{8}$$

$$R_b^{(i+1)} = \mathcal{D}_r\big(V_b^{(i+1)}\big). \tag{9}$$

This update rule is unrolled to fixed number of iterations and used to form the deep learning-based subspace (DL-Subspace) reconstruction network architecture shown in Fig. 1. The DL-Subspace network is trained end-to-end by minimizing the complex-valued l_1 loss:

$$\mathcal{L} = \left\| X - \sum_{b=1}^{N_B} \mathcal{M}_b(L_b^{(I)} R_b^{(I)H})\right\|_1 \tag{10}$$

where $L_b^{(I)}$ and $R_b^{(I)}$ are the final iterates of the network, and X is the fully-sampled ground truth.

3.2 Experiments

Training Data: Fully-sampled, multi-slice 2D cardiac cine MRI datasets were collected from 21 healthy volunteers and 1 pediatric patient under IRB approval. For training, 17 volunteer datasets

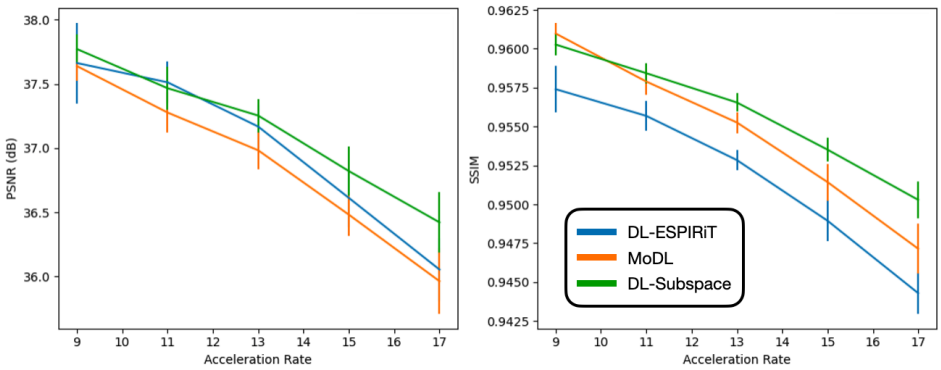

Figure 3: DL-ESPIRiT, MoDL, and DL-Subspace (proposed) networks are evaluated by reconstructing the test datasets with with different acceleration rates. Each reconstruction is evaluated in the image domain against the fully-sampled ground truth images with respect to PSNR and SSIM. All three methods demonstrate comparable metrics. DL-Subspace outperforms MoDL and DL-ESPIRiT by a small margin for higher acceleration rates.

are split slice-by-slice to create 294 unique training examples, each with 20-25 temporal frames. The input to the network is formed by retrospectively excluding k-space measurements using a pseudo-random, variable-density undersampling pattern [16]. This process simulates a rapid MRI acquisition that is accelerated by factors of 10-15 compared to the nominal fully-sampled acquisition. Two volunteer datasets are used for validation, and the remaining datasets for testing.

Evaluation: For testing, we compare DL-Subspace against unrolled proximal gradient descent (DL-ESPIRiT) [5] and unrolled half-quadratic splitting (MoDL) [3] networks. As shown in Table 1, the hyperparameters for each network are set such that they have the same number of learnable parameters. All three networks are trained using the same dataset with a batch size of 1 and Adam optimizer for 1000 epochs [17]. Once trained, reconstructions for each network are evaluated by computing standard image quality metrics such as peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) with respect to fully-sampled ground truth images.

Benchmarks: Memory usage and average inference speed are benchmarked for DL-ESPIRiT, MoDL, and DL-Subspace network with various hyperparameter configurations. Memory usage is measured by internal profiling functions provided by PyTorch. Average inference times are computed by averaging over 16 reconstructions of MRI data with matrix size of 200×180 and 20 temporal frames.

4 Results

DL-Subspace networks achieve 2-4$\times$ higher memory efficiency and 3-4$\times$ faster average inference speed than both DL-ESPIRiT and MoDL. As shown in Fig. 2, the exact gain in memory efficiency and inference speed depends on network hyperparameters such as block size (B) and the number of basis functions (K). Based on preliminary hyperparameter tuning experiments shown in Fig. 4, the DL-Subspace network with $B = 16$ and $K = 12$ is chosen for comparison against the other networks. As shown in Fig. 3, DL-Subspace demonstrates better image quality metrics on average, especially on data with high scan time accelerations. Example reconstructions shown in Figs. 5 & 6 depict similar image quality across all three networks, except DL-Subspace demonstrates slightly better noise performance upon visual inspection.

5 Conclusion

A novel deep learning-based subspace reconstruction framework known as DL Subspace is proposed, which uses a subspace model to curb the memory requirements of training unrolled neural networks for high-dimensional image reconstruction. As a result, both memory efficiency and average inference speed are improved by up to a factor of 4$\times$ without compromising image quality compared to state-of-the-art unrolled networks.

References

[1] Kerstin Hammernik, Teresa Klatzer, Erich Kobler, Michael P. Recht, Daniel K. Sodickson, Thomas Pock, and Florian Knoll. Learning a variational network for reconstruction of accelerated MRI data. *Magnetic Resonance in Medicine*, 79(6):3055–3071, 2018.

[2] Jo Schlemper, Jose Caballero, Joseph V. Hajnal, Anthony Price, and Daniel Rueckert. A Deep Cascade of Convolutional Neural Networks for Dynamic MR Image Reconstruction. *IEEE Transactions on Medical Imaging*, 37(2):491–503, 2018.

[3] Hemant K. Aggarwal, Merry P. Mani, and Mathews Jacob. MoDL: Model-Based Deep Learning Architecture for Inverse Problems. *IEEE Transactions on Medical Imaging*, 38(2):394–405, February 2019.

[4] Sampurna Biswas, Hemant K. Aggarwal, and Mathews Jacob. Dynamic MRI using model-based deep learning and SToRM priors: MoDL-SToRM. *Magnetic Resonance in Medicine*, 82(1):485–494, 2019.

[5] Christopher M. Sandino, Peng Lai, Shreyas S. Vasanawala, and Joseph Y. Cheng. Accelerating cardiac cine MRI using a deep learning-based ESPIRiT reconstruction. *Magnetic Resonance in Medicine*, 85(1):152–167, January 2021.

[6] Karol Gregor and Yann LeCun. Learning Fast Approximations of Sparse Coding. In *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ICML'10, pages 399–406, USA, 2010. Omnipress.

[7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity Mappings in Deep Residual Networks. *arXiv:1603.05027 [cs]*, March 2016.

[8] Zhi Pei Liang. Spatiotemporal imaging with partially separable functions. In *Proceedings of the 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 988–991, 2007.

[9] Joshua D. Trzasko and Armando Manduca. Local versus global low-rank promotion in dynamic MRI series reconstruction. In *Proceedings of the 19th Annual Meeting of the International Society of Magnetic Resonance in Medicine*, page 4371, Montreal, Quebec, Canada, 2011.

[10] Sajan Goud Lingala and Mathews Jacob. Blind compressive sensing dynamic MRI. *IEEE Transactions on Medical Imaging*, 32(6):1132–45, 2013.

[11] Anthony G. Christodoulou, Jaime L. Shaw, Christopher Nguyen, Qi Yang, Yibin Xie, Nan Wang, and Debiao Li. Magnetic resonance multitasking for motion-resolved quantitative cardiovascular imaging. *Nature Biomedical Engineering*, 2(4):215–226, April 2018.

[12] Frank Ong, Xucheng Zhu, Joseph Y. Cheng, Kevin M. Johnson, Peder E. Z. Larson, Shreyas S. Vasanawala, and Michael Lustig. Extreme MRI: Large-scale volumetric dynamic imaging from continuous non-gated acquisitions. *Magnetic Resonance in Medicine*, 84(4):1763–1780, October 2020.

[13] Saiprasad Ravishankar and Yoram Bresler. MR Image Reconstruction From Highly Undersampled k-Space Data by Dictionary Learning. *IEEE Transactions on Medical Imaging*, 30(5):1028–1041, May 2011.

[14] Jose Caballero, Anthony N. Price, Daniel Rueckert, and Joseph V. Hajnal. Dictionary learning and time sparsity for dynamic MR data reconstruction. *IEEE Transactions on Medical Imaging*, 33(4):256–263, 2014.

[15] Marius Arvinte, Sriram Vishwanath, Ahmed H. Tewfik, and Jonathan I. Tamir. Deep J-Sense: Accelerated MRI reconstruction via unrolled alternating optimization. In *Proceedings of the 24th International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 350–360, Strasbourg, France, 2021.

[16] Evan J. Zucker, Christopher M. Sandino, Aya Kino, Peng Lai, and Shreyas S. Vasanawala. Free-breathing Accelerated Cardiac MRI Using Deep Learning: Validation in Children and Young Adults. *Radiology*, page 202624, June 2021.

[17] Diederik P Kingma and Jimmy Lei Ba. Adam: A method for stochastic gradient descent. In *Proceedings of the 3rd International Conference on Learning Representations*, San Diego, CA, United States, 2015.

[18] Donald Geman and Chengda Yang. Nonlinear image recovery with half-quadratic regularization. *IEEE Transactions on Image Processing*, 4(7):932–946, July 1995.

A Appendix

A.1 Model-based Deep Learning (MoDL)

Unrolled neural networks have demonstrated state-of-the-art performance in reconstruction of accelerated dynamic MRI data [2, 4, 5]. Model-based deep learning reconstruction (MoDL) is one such approach that performs reconstruction by solving the following optimization problem [3]:

$$\hat{X} = \arg\min_{X} ||Y - \mathcal{A}(X)||_F^2 + \mu ||\mathcal{D}(X) - X||_F^2 \tag{11}$$

where $\mathcal{D}$ is a 3D CNN which acts as a spatiotemporal denoiser on X, and μ is the regularization strength. The problem in Eq. 11 is solved using the half quadratic splitting (HQS) algorithm which alternates between two steps [18]. In the first step, $\mathcal{D}(X)$ is treated as a constant, and the optimization problem in Eq. 11 is solved using a conjugate gradient (CG) solver. In the second step, the CNN is applied to denoise the output of the first step. The MoDL update rule is defined as follows:

$$Z^{(i+1)} = \arg\min_{X} ||Y - \mathcal{A}(X)||_F^2 + \mu ||X^{(i)} - X||_F^2 \tag{12}$$

$$X^{(i+1)} = \mathcal{D}(Z^{(i+1)}) \tag{13}$$

Both of the HQS and CG iterations are unrolled to have a fixed number of iterations, and trained end-to-end in a supervised fashion to learn the weights of $\mathcal{D}$ and the optimal value for μ from a training dataset of historical exam data.

A.2 Applying MoDL to Subspace Reconstruction

In subspace reconstruction, an inverse problem is solved to estimate the low-dimensional subspace of X instead of X directly. As mentioned in Section 3.1, this inverse problem can be written in the following form:

$$\underset{L_b, R_b}{\text{minimize}} \left|\left| Y - \mathcal{A}\left(\sum_{b=1}^{N_B} \mathcal{M}_b (L_b R_b^{(i)H}) \right) \right|\right|_F^2 + \mu_l \sum_{b=1}^{N_B} ||\mathcal{D}_l(L_b) - L_b||_F^2 + \mu_r \sum_{b=1}^{N_B} ||\mathcal{D}_r(R_b) - R_b||_F^2 \tag{14}$$

where $\mathcal{D}_l$ and $\mathcal{D}_r$ are simpler 2-D and 1-D CNNs that denoise the spatial and temporal basis functions respectively. The regularization strengths μ_l and μ_r are parameters that must be learned along with the CNN weights. This non-convex problem is broken down into simpler convex sub-problems by applying alternating minimization:

$$L_b^{(i+1)} = \arg\min_{L_b} \left|\left| Y - \mathcal{A}\left(\sum_{b=1}^{N_B} \mathcal{M}_b (L_b R_b^{(i)H}) \right) \right|\right|_F^2 + \mu_l \sum_{b=1}^{N_B} ||\mathcal{D}_l(L_b) - L_b||_F^2 \tag{15}$$

$$R_b^{(i+1)} = \arg\min_{R_b} \left|\left| Y - \mathcal{A}\left(\sum_{b=1}^{N_B} \mathcal{M}_b (L_b^{(i+1)} R_b^{H}) \right) \right|\right|_2^2 + \mu_r \sum_{b=1}^{N_B} ||\mathcal{D}_r(R_b) - R_b||_F^2. \tag{16}$$

Both sub-problems take the form of the problem solved by MoDL in Eqn. 11. Therefore, the MoDL update rule can be used to solve Eqns. 15 & 16 which forms the following 4-step update rule:

$$U_b^{(i+1)} = \arg\min_{L_b} \left|\left| Y - \mathcal{A}\left(\sum_{b=1}^{N_B} \mathcal{M}_b (L_b R_b^{(i)H}) \right) \right|\right|_F^2 + \mu_l \sum_{b=1}^{N_B} ||L_b^{(i)} - L_b||_F^2 \tag{17}$$

$$L_b^{(i+1)} = \mathcal{D}_l\big(U_b^{(i+1)}\big) \tag{18}$$

$$V_b^{(i+1)} = \arg\min_{R_b} \left|\left| Y - \mathcal{A}\left(\sum_{b=1}^{N_B} \mathcal{M}_b (L_b^{(i+1)} R_b^{H}) \right) \right|\right|_F^2 + \mu_r \sum_{b=1}^{N_B} ||R_b^{(i)} - R_b||_F^2 \tag{19}$$

$$R_b^{(i+1)} = \mathcal{D}_r\big(V_b^{(i+1)}\big). \tag{20}$$

A.3 How to choose the number of basis functions

The number of spatial and basis functions (K) is a network hyperparameter that needs to be tuned empirically for each imaging application. In this work, B fixed to 16, and K is chosen by training multiple networks with $K = 4, 8, 12$ and finding the network with the lowest average validation loss. As shown by the plots in Fig. 4, the DL-Subspace network with $K = 12$ basis functions performs the best on average. Better performance may be achieved by training with more basis functions, however, there are diminishing returns to adding more basis functions. Additionally, the representations L_b and R_b become less compact as the number of basis functions is increased, leading to decreased memory efficiency. Trade-offs between memory efficiency and reconstruction performance and will be the subject of future work.

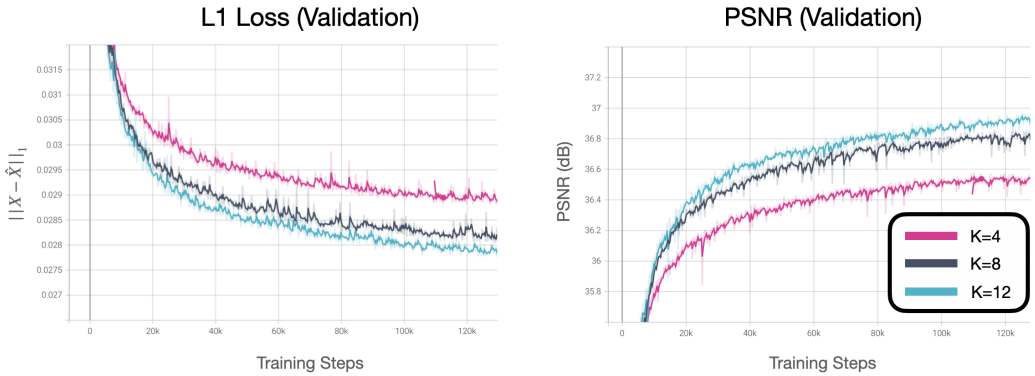

Figure 4: The average L_1 loss and peak signal-to-noise ratio (PSNR) is evaluated over the validation set for DL-Subspace networks with varying numbers of basis functions (K).

A.4 Network Hyperparameters

	# Unrolls	# CG Unrolls	# Learnable Parameters	Sub-Network Architecture	Kernel Size	# Input Features	# Hidden Features
DL-ESPIRiT	10	N/A	8.5 million	2D+Time ResNet-5	3 x 3 x 3	2	88
MoDL	10	10	8.5 million	2D+Time ResNet-5	3 x 3 x 3	2	88
DL-Subspace	10	10	8.6 million	2D ResNet-5	3 x 3	24	128
				1D ResNet-5	3 x 1	24	128

Table 1: Table containing hyperparameters for DL-ESPIRiT, MoDL, and DL-Subspace (proposed). Note that the number of hidden features for DL-ESPIRiT and MoDL networks is modified such that the total number of learnable parameters for all networks remain roughly the same. To increase model expressivity, the ResNet weights are not shared between unrolled iterations.

A.5 Reconstruction examples on volunteer data

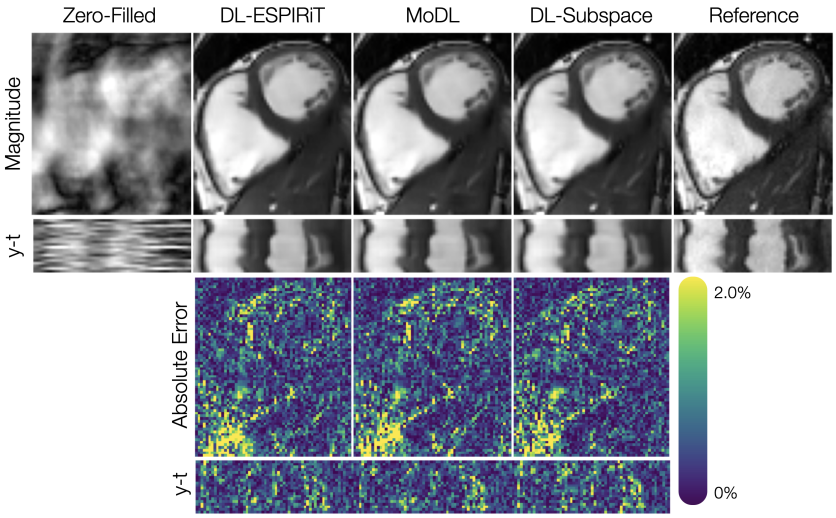

Figure 5: With IRB approval, one fully-sampled 2D cardiac cine MRI scan is performed on a healthy volunteer. The data is retrospectively undersampled to simulate an accelerated scan that is faster than the fully-sampled scan by a factor of 14. The undersampled data is reconstructed using (from left-to-right): zero-filling, DL-ESPIRiT, MoDL, and DL-Subspace. Magnitude images and y-t plots with corresponding error images and error y-t plots are shown here. No significant difference in image quality is observed between the deep learning-based methods.

A.6 Reconstruction examples on patient data

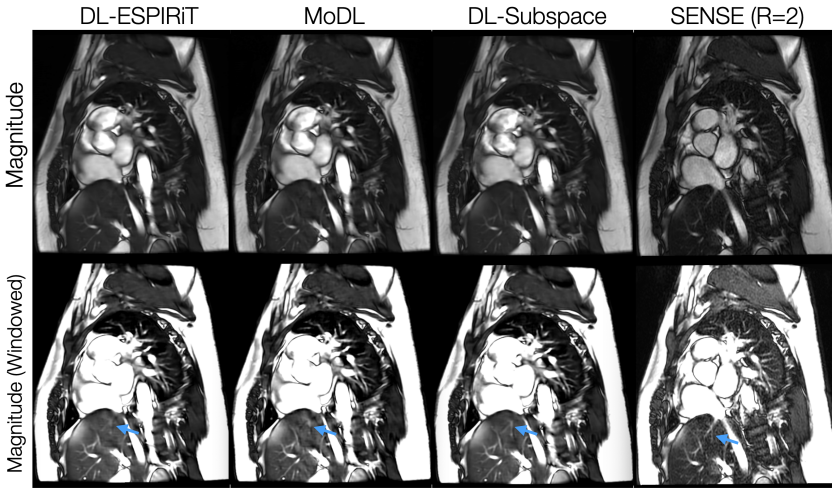

Figure 6: With IRB approval, two prospectively undersampled 2D cardiac cine MRI scans are performed in a pediatric patient. The first acquisition is uniformly undersampled (R=2) and reconstructed using SENSE (reference). The second acquisition is prospectively undersampled by a variable-density k_y-t undersampling pattern (R=12) and reconstructed using DL-ESPIRiT, MoDL, and DL-Subspace. The top and bottom rows show the same images, except the bottom row is windowed to better appreciate noise in the four reconstructions. The DL-Subspace images are observed to be the least noisy due to the denoising effect of reconstructing the target in a compressed subspace. This is best appreciated in the liver and hepatic vessels (blue arrows).