

# Memory-efficient Learning for High-Dimensional MRI Reconstruction

Ke Wang<sup>1</sup>, Michael Kellman<sup>2</sup>, Christopher M. Sandino<sup>3</sup>, Kevin Zhang<sup>1</sup>, Shreyas S. Vasanawala<sup>4</sup>, Jonathan I. Tamir<sup>5</sup>, Stella X. Yu<sup>6</sup>, and Michael Lustig<sup>1</sup>

<sup>1</sup> Electrical Engineering and Computer Sciences, UC Berkeley, Berkeley CA, USA

`mikilustig@berkeley.edu`

<sup>2</sup> Pharmaceutical Chemistry, UCSF, San Francisco CA, USA

<sup>3</sup> Electrical Engineering, Stanford University, Stanford CA, USA

<sup>4</sup> Radiology, Stanford University, Stanford CA, USA

<sup>5</sup> Electrical and Computer Engineering, UT Austin, Austin TX, USA

<sup>6</sup> International Computer Science Institute, UC Berkeley, Berkeley CA, USA

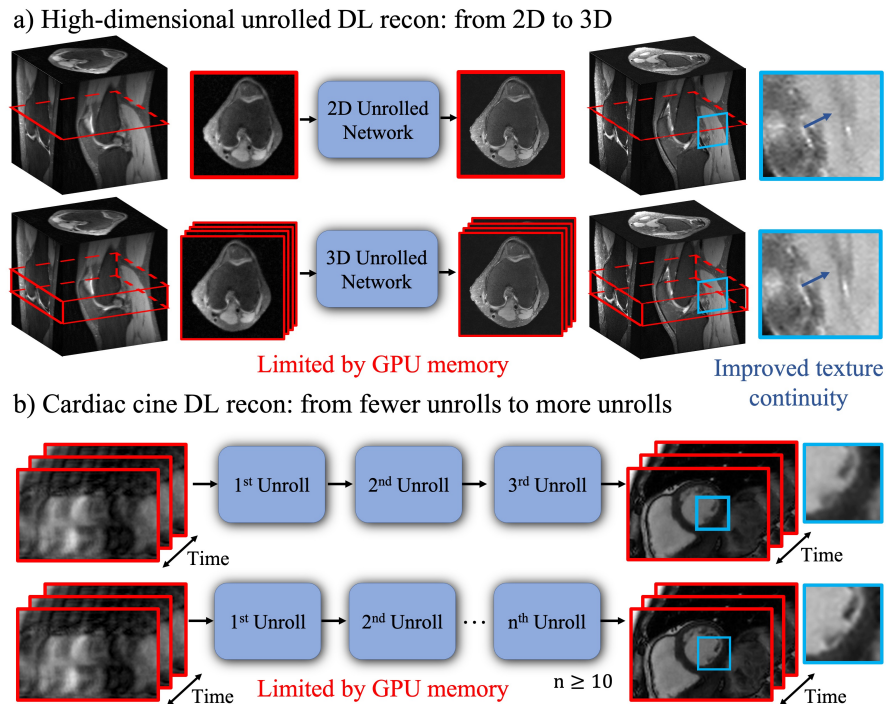
**Abstract.** Deep learning (DL) based unrolled reconstructions have shown state-of-the-art performance for under-sampled magnetic resonance imaging (MRI). Similar to compressed sensing, DL can leverage high-dimensional data (e.g. 3D, 2D+time, 3D+time) to further improve performance. However, network size and depth are currently limited by the GPU memory required for backpropagation. Here we use a memory-efficient learning (MEL) framework which favorably trades off storage with a manageable increase in computation during training. Using MEL with multi-dimensional data, we demonstrate improved image reconstruction performance for in-vivo 3D MRI and 2D+time cardiac cine MRI. MEL uses far less GPU memory while marginally increasing the training time, which enables new applications of DL to high-dimensional MRI.

**Keywords:** Magnetic Resonance Imaging (MRI) · Unrolled reconstruction · Memory-efficient learning.

## 1 Introduction

Deep learning-based unrolled reconstructions (Unrolled DL recons)[2,19,1,5,21,11] have shown great success at under-sampled MRI reconstruction, well beyond the capabilities of parallel imaging and compressed sensing (PICS)[12,4,16]. These methods are often formulated by unrolling the iterations of an image reconstruction optimization[5,1,21] and use a training set to learn an implicit regularization term represented by a deep neural network. It has been shown that increasing the number of unrolls improves upon finer spatial and temporal textures in the reconstruction[5,1,17]. Similar to compressed sensing and other low-dimensional representations, DL recons can take advantage of additional structure in very high-dimensional data (e.g. 3D, 2D+time, 3D+time) to further improve image quality. However, these large-scale DL recons are currently limited by GPU memory required for gradient-based optimization using backpropagation. Therefore,

most Unrolled DL recons focus on 2D applications or are limited to a small number of unrolls. In this work, we use our recently proposed memory-efficient learning (MEL) framework[9,23] to reduce the memory needed for backpropagation, which enables the training of Unrolled DL recons for 1) larger-scale 3D MRI; and 2) 2D+time cardiac cine MRI with a large number of unrolls (Figure 1). We evaluate the spatio-temporal complexity of our proposed method on the Model-based Deep Learning (MoDL) architecture [1] and train these high-dimensional DL recons on a single 12GB GPU. Our training uses far less memory while only marginally increasing the computation time. To demonstrate the advantages of high-dimensional reconstructions to image quality, we performed experiments on both retrospectively and prospectively under-sampled data for 3D MRI and cardiac cine MRI. Our in-vivo experiments indicate that by exploiting high-dimensional data redundancy, we can achieve better quantitative metrics and improved image quality with sharper edges for both 3D MRI and cardiac cine MRI.

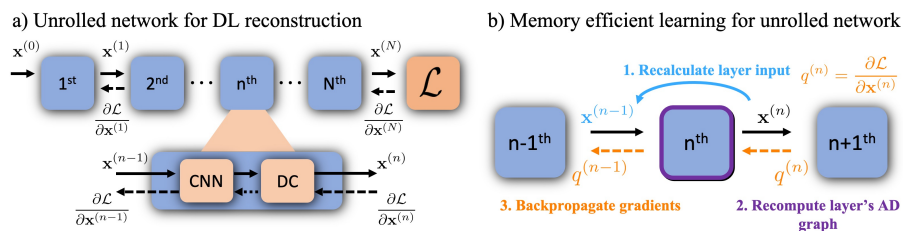


**Fig. 1.** GPU memory limitations for high-dimensional unrolled DL recons: a) Compared to a 2D unrolled network, the 3D unrolled network uses a 3D slab during training to leverage more redundancy, but is limited by GPU memory. b) Cardiac cine DL recons are often performed with a small number of unrolls due to memory limitations.

## 2 Methods

### 2.1 Memory-efficient learning

As shown in Figure 2 a), unrolled DL recons are often formulated by unrolling the iterations of an image reconstruction optimization[5,1]. Each unroll consists of two submodules: CNN based regularization layer and data consistency (DC) layer. In conventional backpropagation, the gradient must be computed for the entire computational graph, and intermediate variables from all  $N$  unrolls need to be stored at a significant memory cost. By leveraging MEL, we can process the full graph as a series of smaller sequential graphs. As shown in Figure 2 b), first, we forward propagate the network to get the output  $\mathbf{x}^{(N)}$  without computing the gradients. Then, we rely on the invertibility of each layer (required) to recompute each smaller auto-differentiation (AD) graph from the network’s output in reverse order. MEL only requires a single layer to be stored in memory at a time, which reduces the required memory by a factor of  $N$ . Notably, the required additional computation to invert each layer only marginally increases the backpropagation runtime.



**Fig. 2.** a) In conventional DL recon training, gradients of all layers are evaluated as a single computational graph, requiring significant GPU memory. b) In MEL, we sequentially evaluate each layer by: i) Recalculate the layer’s input  $\mathbf{x}^{(n-1)}$ , from the known output  $\mathbf{x}^{(n)}$ . ii) Reform the AD graph for that layer. iii) Backpropagate gradients  $q^{(n-1)}$  through the layer’s AD graph.

### 2.2 Memory-efficient learning for MoDL

Here, we use a widely used Unrolled DL Recon framework: MoDL[1]. We formulate the reconstruction of  $\hat{\mathbf{x}}$  as an optimization problem and solve it as below:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \mu \|\mathbf{x} - R_w(\mathbf{x})\|_2^2, \quad (1)$$

where  $\mathbf{A}$  is the system encoding matrix,  $\mathbf{y}$  denotes the k-space measurements and  $R_w$  is a learned CNN-based denoiser. For multi-channel MRI reconstruction,  $\mathbf{A}$  can be formulated as  $\mathbf{A} = \mathbf{P}\mathbf{F}\mathbf{S}$ , where  $\mathbf{S}$  represent the multi-channel sensitivity

maps,  $\mathbf{F}$  denotes Fourier Transform and  $\mathbf{P}$  is the undersampling mask used for selecting the acquired data. MoDL solves the minimization problem by an alternating procedure:

$$\mathbf{z}_n = R_w(\mathbf{x}_n) \quad (2)$$

$$\begin{aligned} \mathbf{x}_{n+1} &= \arg \min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \mu \|\mathbf{x} - \mathbf{z}_n\|_2^2, \\ &= (\mathbf{A}^H \mathbf{A} + \mu \mathbf{I})^{-1} (\mathbf{A}^H \mathbf{y} + \mu \mathbf{z}_n) \end{aligned} \quad (3)$$

which represents the CNN-based regularization layer and DC layer respectively. In this formulation, the DC layer is solved using Conjugate Gradient (CG)[20], which is unrolled for a finite number of iterations. For all the experiments, we used an invertible residual convolutional neural network (RCNN) introduced in [3,14,6], whose architecture is composed of a 5-layer CNN with 64 channels per layer. Detailed network architecture is shown in Figure S1. The residual CNN is inverted using the fixed-point algorithm as described in [9], while the DC layer is inverted through:

$$\mathbf{z}_n = \frac{1}{\mu} ((\mathbf{A}^H \mathbf{A} + \mu \mathbf{I}) \mathbf{x}_{n+1} - \mathbf{A}^H \mathbf{y}). \quad (4)$$

### 2.3 Training and evaluation of memory-efficient learning

With IRB approval and informed consent/assent, we trained and evaluated MEL on both retrospective and prospective 3D knee and 2D+time cardiac cine MRI. We conducted 3D MoDL experiments with and without MEL on 20 fully-sampled 3D knee datasets (320 slices each) from mridata.org[18]. 16 cases were used for training, 2 cases were used for validation and other 2 for testing. Around 5000 3D slabs with size  $21 \times 256 \times 320$  were used for training the reconstruction networks. All data were acquired on a 3T GE Discovery MR 750 with an 8-channel HD knee coil. An 8x Poisson Disk sampling pattern was used to retrospectively undersample the fully sampled k-space. Scan parameters included a matrix size of  $320 \times 256 \times 320$ , and TE/TR of 25ms/1550ms. In order to further demonstrate the feasibility of our 3D reconstruction with MEL on realistic prospectively undersampled scans, we reconstructed  $8 \times$  prospectively undersampled 3D FSE knee scans (available at mridata.org) with the model trained on retrospectively undersampled knee data. Scanning parameters includes: Volume size:  $320 \times 288 \times 236$ , TR/TE = 1400/20.46ms, Flip Angle:  $90^\circ$ , FOV:  $160 \text{ mm} \times 160 \text{ mm} \times 141.6 \text{ mm}$ .

For the cardiac cine MRI, fully-sampled bSSFP cardiac cine datasets were acquired from 15 volunteers at different cardiac views and slice locations on 1.5T and 3.0T GE scanners using a 32-channel cardiac coil. All data were coil compressed[24] to 8 virtual coils. Twelve of the datasets (around 190 slices) were used for training, 2 for validation, and one for testing. k-Space data were retrospectively under-sampled using a variable-density k-t sampling pattern to simulate 14-fold acceleration with 25% partial echo. We also conducted experiments on a prospectively under-sampled scan (R=12) which was acquired from a pediatric patient within a single breath-hold on a 1.5T scanner.

We compared the spatio-temporal complexity (GPU memory, training time) with and without MEL. In order to show the benefits of high-dimensional DL recons, we compared the reconstruction results of PICS, 2D and 3D MoDL with MEL for 3D MRI, and 2D+time MoDL with 4 unrolls and 10 unrolls for cardiac cine MRI. For both 2D MoDL and 3D MoDL with MEL, we used 5 unrolls, 10 CG steps and Residual CNN as the regularization layer. A baseline PICS reconstruction was performed using BART[22]. Sensitivity maps were computed using BART[22] and SigPy[13]. Common image quality metrics such as Peak Signal to Noise Ratio (pSNR), Structural Similarity (SSIM) [8] and Fréchet Inception Distance (FID)[7] were reported. FID is a widely used measure of perceptual similarity between two sets of images. All the experiments were implemented in Pytorch [15] and used Nvidia Titan XP (12GB) and Titan V CEO (32GB) GPUs. Networks were trained end-to-end using a per-pixel  $l_1$  loss and optimized using Adam [10] with a learning rate of  $1 \times 10^{-4}$ .

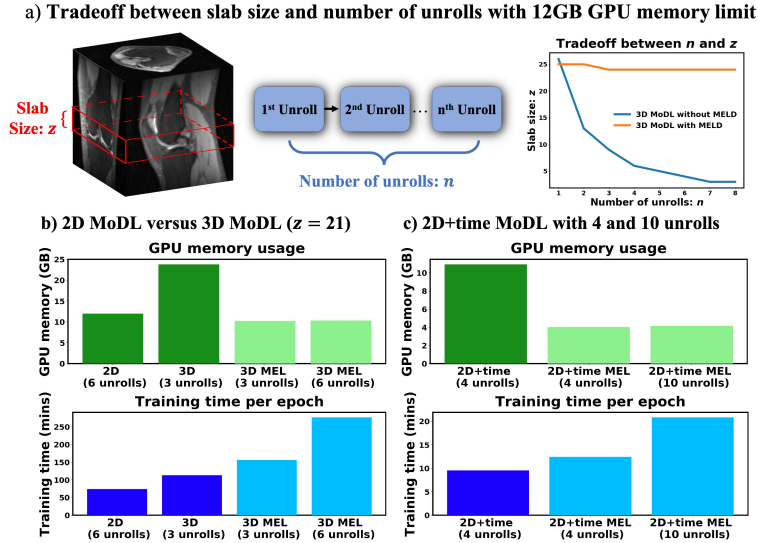
### 3 Results

We first evaluate the spatio-temporal complexity of MoDL with and without MEL (Figure 3). Without MEL, for a 12GB GPU memory limit, the maximum slab size decreases rapidly as the number of unrolls increases, which limits the performance of a 3D reconstruction. In contrast, using MEL, the maximum slab size is roughly constant. Figure 3 b) and c) show the comparisons from two different perspectives: 1)GPU memory usage; 2)Training time per epoch. Results indicate that for both 3D and 2D+time MoDL, MEL uses significantly less GPU memory than conventional backpropagation while marginally increasing training time. Notably, both MoDL with and without MEL have the same inference time.

Figure 4 shows a comparison of different methods for 3D reconstruction. Instead of learning from only 2D axial view slices (Figure 1 a), 3D MoDL with MEL captures the image features from all three dimensions. Zoomed-in details indicate that 3D MoDL with MEL is able to provide more faithful contrast with more continuous and realistic textures as well as higher pSNR over other methods. Figure 5 demonstrates that MEL enables the training of 2D+time MoDL with a large number of unrolls (10 unrolls), which outperforms MoDL with 4 unrolls with respect to image quality and y-t motion profile. With MEL, MoDL with 10 unrolls resolves the papillary muscles (yellow arrows) better than MoDL with 4 unrolls. Also, the y-t profile of MoDL with 10 unrolls depicts motion in a more natural way while MoDL with 4 unrolls suffers from blurring. Meanwhile, using 10 unrolls over 4 unrolls yields an improvement of 0.6dB in validation pSNR.

Table 1 shows the quantitative metric comparisons (pSNR, SSIM and FID) between different methods on both 3D MRI and cardiac cine MRI reconstructions. The results indicate that both 3D MoDL with MEL and 2D+time MoDL with MEL outperform other methods with respect to pSNR, SSIM and FID.

Figure 6 a) and Figure S2 show the reconstruction results on two representative prospectively undersampled 3D FSE knee scan. Note that in this scenario,

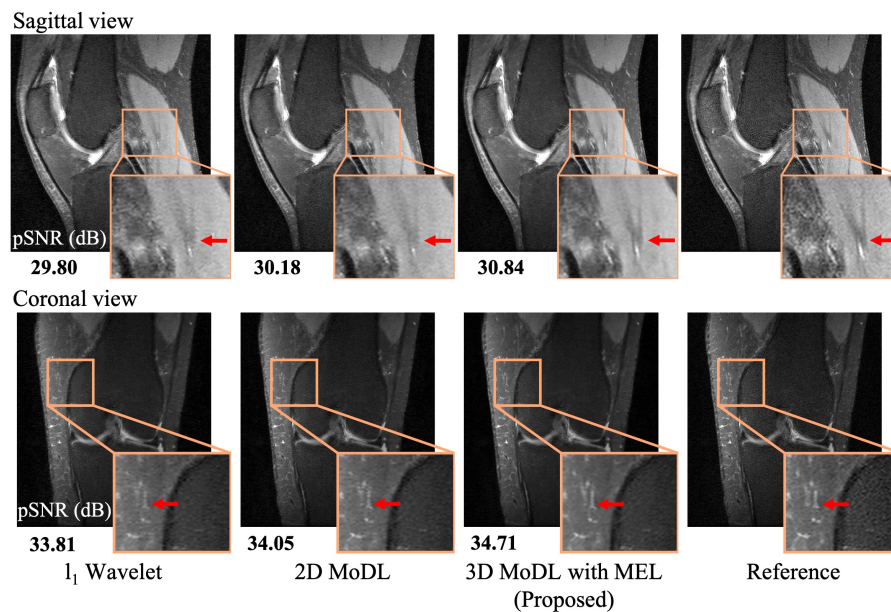


**Fig. 3.** Spatio-temporal complexity of MoDL with and without MEL. a) Tradeoff between 3D slab size  $z$  and a number of unrolls  $n$  with a 12GB GPU memory limitation. b) and c) show the memory and time comparisons for MoDL with and without MEL.

metric	method	3D MRI	2D cardiac cine MRI
pSNR (dB)	PICS	31.01±1.97	24.69±2.74
	2D MoDL	31.44±2.07	-
	3D MoDL with MEL	<b>32.11±2.05</b>	-
	2D+time MoDL: 4 unrolls	-	26.87±2.98
	2D+time MoDL with MEL: 10 unrolls	-	<b>27.42±3.21</b>
SSIM	PICS	0.816±0.046	0.824±0.071
	2D MoDL	0.821±0.044	-
	3D MoDL with MEL	<b>0.830±0.038</b>	-
	2D+time MoDL: 4 unrolls	-	0.870±0.042
	2D+time MoDL with MEL: 10 unrolls	-	<b>0.888±0.042</b>
FID	PICS	46.71	39.40
	2D MoDL	43.58	-
	3D MoDL with MEL	<b>41.48</b>	-
	2D+time MoDL: 4 unrolls	-	36.93
	2D+time MoDL with MEL: 10 unrolls	-	<b>31.64</b>

**Table 1.** Quantitative metrics (pSNR, SSIM and FID) of different methods on 3D MRI and cardiac cine MRI reconstructions (mean ± standard deviation of pSNR and SSIM).

there is no fully-sampled ground truth. Despite there exists some difference between the training and testing (e.g., matrix size, scanning parameters), 3D MoDL

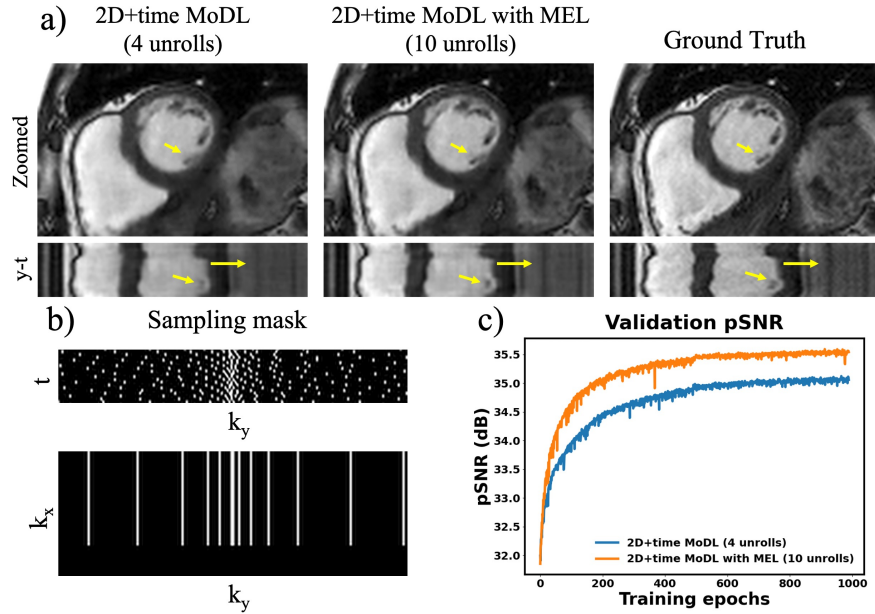


**Fig. 4.** A representative comparison of different methods (PICS, 2D MoDL, 3D MoDL with MEL) on 3D knee reconstruction (Sagittal view and Coronal view are shown). pSNRs are shown under each reconstructed image.

with MEL is still able to resolve more detailed texture and sharper edges over traditional PICS and learning-based 2D MoDL. Figure 6 b) and Video S1 shows the reconstruction on a representative prospective undersampled cardiac cine scan. We can clearly see that enabled by MEL, 2D+time MoDL with 10 unrolls can better depicts the finer details as well as more natural motion profile.

## 4 Conclusions

In this work, we show that MEL enables learning for high-dimensional MR reconstructions on a single 12GB GPU, which is not possible with standard back-propagation methods. We demonstrate MEL on two representative large-scale MR reconstruction problems: 3D volumetric MRI, 2D cardiac cine MRI with a relatively large number of unrolls. By leveraging the high-dimensional image redundancy and a large number of unrolls, we were able to get improved quantitative metrics and reconstruct finer details, sharper edges, and more continuous textures with higher overall image quality for both 3D and 2D cardiac cine MRI. Furthermore, 3D MoDL reconstruction results from prospectively undersampled k-space show that the proposed method is robust to the scanning parameters and could be potentially deployed in clinical systems. Overall, MEL brings a practical tool for training the large-scale high-dimensional MRI reconstructions



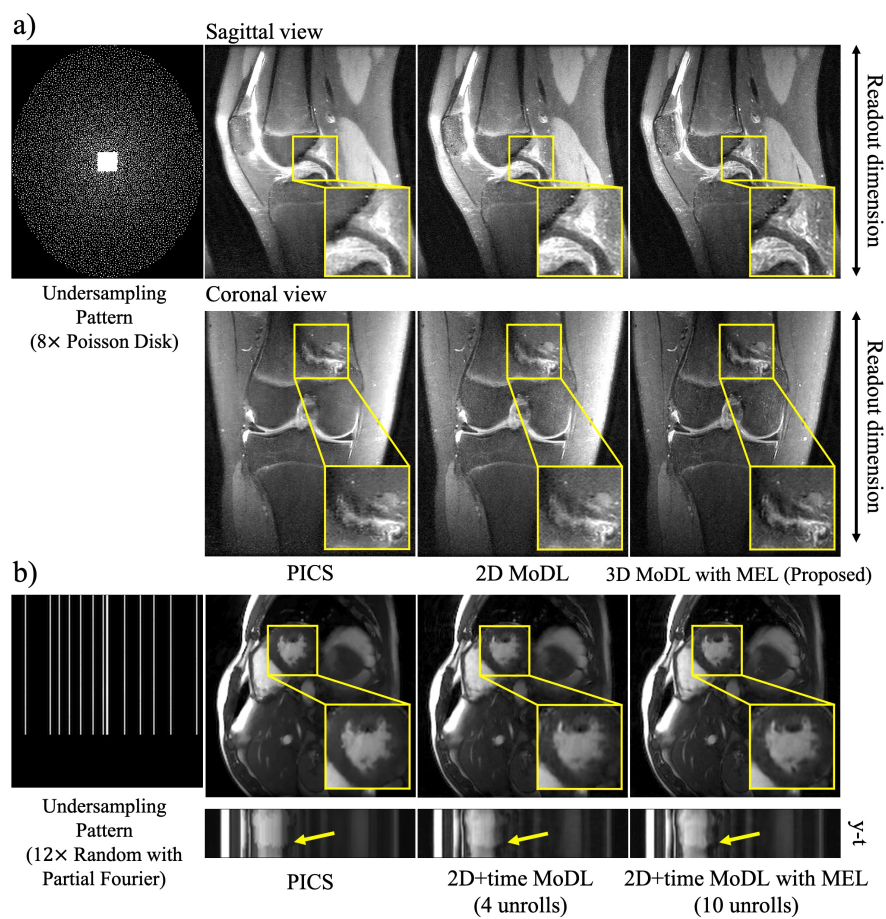
**Fig. 5.** a) Short-axis view cardiac cine reconstruction of a healthy volunteer on a 1.5T scanner. k-Space data was retrospectively undersampled to simulate 14-fold acceleration with 25% partial echo (shown in b) and reconstructed by: 2D+time MoDL with 4 unrolls, 2D+time MoDL with MEL and 10 unrolls. c) Validation pSNR of MoDL with 4 unrolls and MoDL with 10 unrolls.

with much less GPU memory and is able to achieve improved reconstructed image quality.

## 5 Acknowledgements

The authors would like to thank Dr. Gopal Nataraj for his helpful discussions and paper editing. We also acknowledge support from NIH R01EB009690, NIH R01HL136965, NIH R01EB026136 and GE Healthcare.





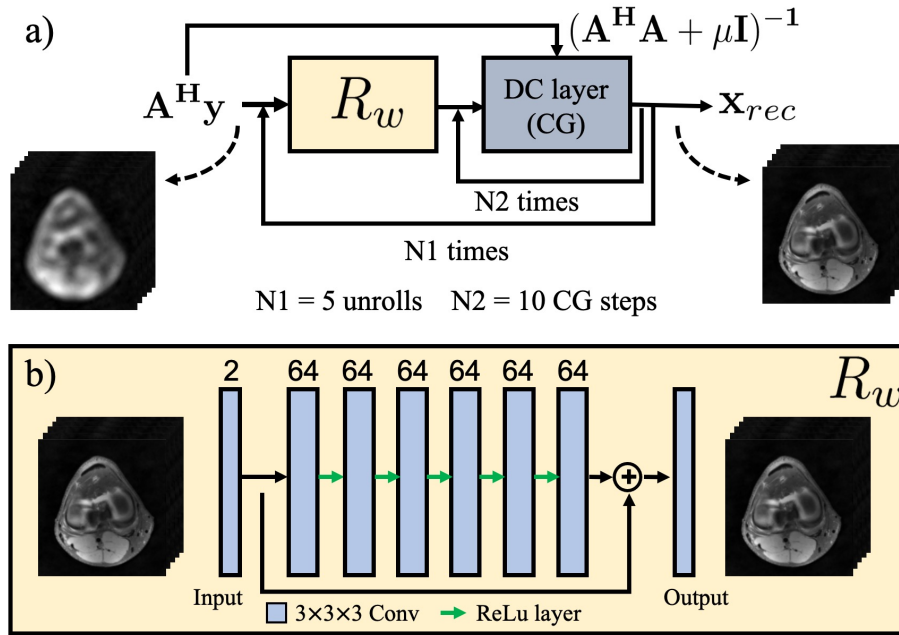
**Fig. 6.** a) Representative reconstruction results on a prospectively undersampled 3D FSE knee scan using different methods (PICS, 2D MoDL and 3D MoDL with MEL). b) Representative reconstruction results on a prospectively undersampled cardiac cine dataset.  $y$ - $t$  motion profiles are shown along with the reconstructed images.

## References

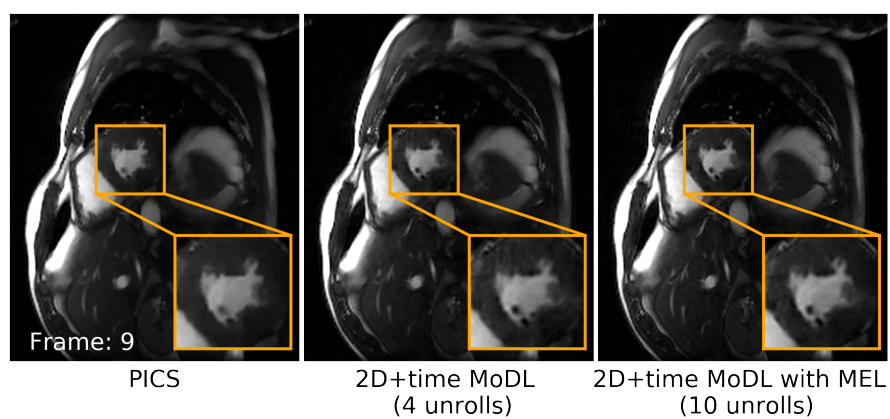
1. Aggarwal, H.K., Mani, M.P., Jacob, M.: Modl: Model-based deep learning architecture for inverse problems. *IEEE transactions on medical imaging* **38**(2), 394–405 (2018)
2. Diamond, S., Sitzmann, V., Heide, F., Wetzstein, G.: Unrolled optimization with deep priors. arXiv preprint arXiv:1705.08041 (2017)
3. Gomez, A.N., Ren, M., Urtasun, R., Grosse, R.B.: The reversible residual network: Backpropagation without storing activations. arXiv preprint arXiv:1707.04585 (2017)
4. Griswold, M.A., Jakob, P.M., Heidemann, R.M., Nittka, M., Jellus, V., Wang, J., Kiefer, B., Haase, A.: Generalized autocalibrating partially parallel acquisitions (grappa). *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* **47**(6), 1202–1210 (2002)
5. Hammernik, K., Klatzer, T., Kobler, E., Recht, M.P., Sodickson, D.K., Pock, T., Knoll, F.: Learning a variational network for reconstruction of accelerated mri data. *Magnetic resonance in medicine* **79**(6), 3055–3071 (2018)
6. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 770–778 (2016)
7. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: Gans trained by a two time-scale update rule converge to a local nash equilibrium. arXiv preprint arXiv:1706.08500 (2017)
8. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 1125–1134 (2017)
9. Kellman, M., Zhang, K., Markley, E., Tamir, J., Bostan, E., Lustig, M., Waller, L.: Memory-efficient learning for large-scale computational imaging. *IEEE Transactions on Computational Imaging* **6**, 1403–1414 (2020)
10. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
11. Küstner, T., Fuin, N., Hammernik, K., Bustin, A., Qi, H., Hajhosseiny, R., Masci, P.G., Neji, R., Rueckert, D., Botnar, R.M., et al.: Cinenet: deep learning-based 3d cardiac cine mri reconstruction with multi-coil complex-valued 4d spatio-temporal convolutions. *Scientific reports* **10**(1), 1–13 (2020)
12. Lustig, M., Donoho, D., Pauly, J.M.: Sparse mri: The application of compressed sensing for rapid mr imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* **58**(6), 1182–1195 (2007)
13. Ong, F., Lustig, M.: Sigpy: a python package for high performance iterative reconstruction. In: *Proc. ISMRM* (2019)
14. Ovadia, Y., Fertig, E., Ren, J., Nado, Z., Sculley, D., Nowozin, S., Dillon, J.V., Lakshminarayanan, B., Snoek, J.: Can you trust your model’s uncertainty? evaluating predictive uncertainty under dataset shift. arXiv preprint arXiv:1906.02530 (2019)
15. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S.: Pytorch: An imperative style, high-performance deep learning library. In: *Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E.,*

- Garnett, R. (eds.) *Advances in Neural Information Processing Systems* 32, pp. 8024–8035. Curran Associates, Inc. (2019), <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>
16. Pruessmann, K.P., Weiger, M., Scheidegger, M.B., Boesiger, P.: Sense: sensitivity encoding for fast mri. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* **42**(5), 952–962 (1999)
  17. Sandino, C.M., Lai, P., Vasanawala, S.S., Cheng, J.Y.: Accelerating cardiac cine mri using a deep learning-based espirit reconstruction. *Magnetic Resonance in Medicine* **85**(1), 152–167 (2021)
  18. Sawyer, A.M., Lustig, M., Alley, M., Uecker, P., Virtue, P., Lai, P., Vasanawala, S.: Creation of fully sampled mr data repository for compressed sensing of the knee. *Citeseer* (2013)
  19. Schlemper, J., Caballero, J., Hajnal, J.V., Price, A.N., Rueckert, D.: A deep cascade of convolutional neural networks for dynamic mr image reconstruction. *IEEE transactions on Medical Imaging* **37**(2), 491–503 (2017)
  20. Shewchuk, J.R., et al.: An introduction to the conjugate gradient method without the agonizing pain (1994)
  21. Tamir, J.I., Yu, S.X., Lustig, M.: Unsupervised deep basis pursuit: Learning inverse problems without ground-truth data. *arXiv preprint arXiv:1910.13110* (2019)
  22. Uecker, M., Ong, F., Tamir, J.I., Bahri, D., Virtue, P., Cheng, J.Y., Zhang, T., Lustig, M.: Berkeley advanced reconstruction toolbox. In: *Proc. Intl. Soc. Mag. Reson. Med.* No. 2486 in 23 (2015)
  23. Zhang, K., Kellman, M., Tamir, J.I., Lustig, M., Waller, L.: Memory-efficient learning for unrolled 3d mri reconstructions. In: *ISMRM Workshop on Data Sampling and Image Reconstruction* (2020)
  24. Zhang, T., Pauly, J.M., Vasanawala, S.S., Lustig, M.: Coil compression for accelerated imaging with cartesian sampling. *Magnetic resonance in medicine* **69**(2), 571–582 (2013)

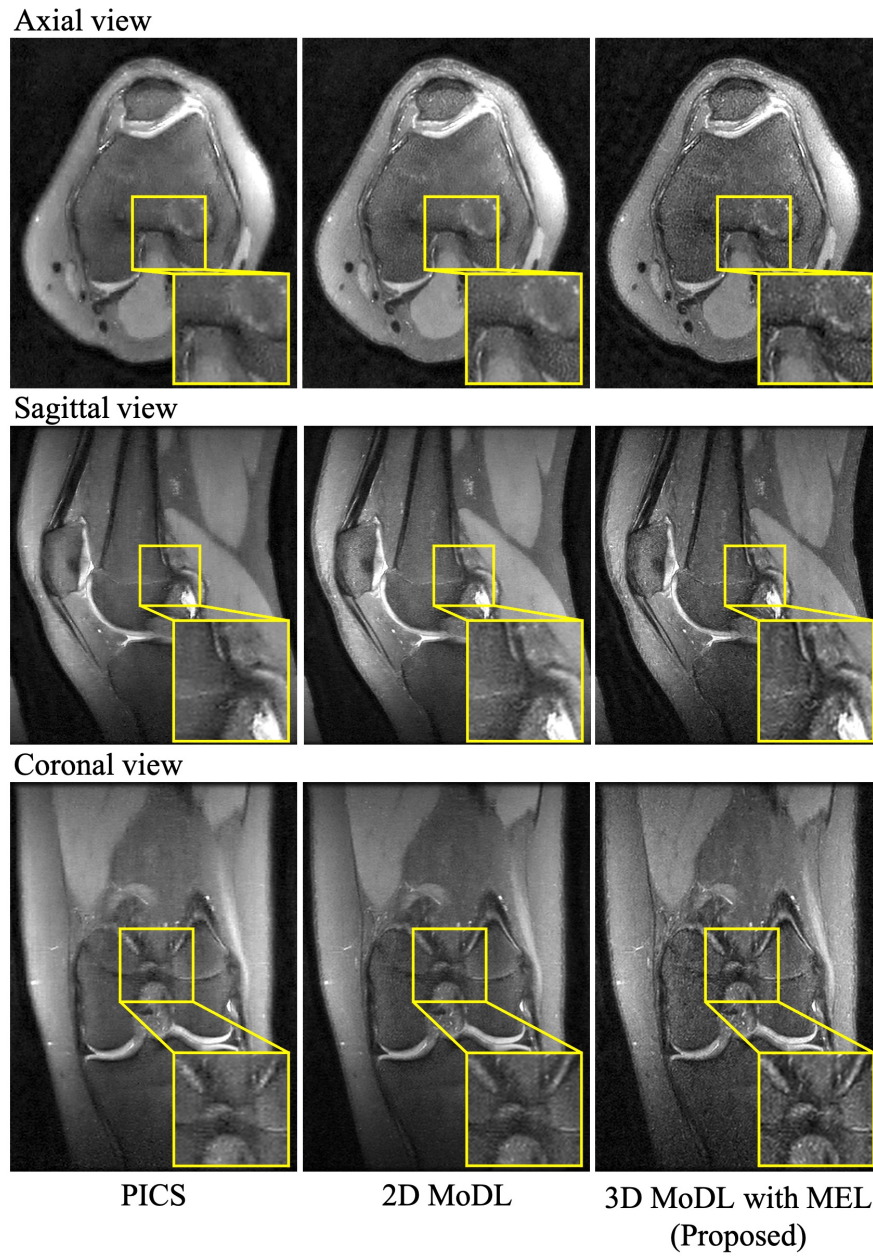
## 6 Supplementary Material



**Fig. S1.** Architectures of 3D MoDL framework and CNN based regularization layer. a) Zero-filled reconstruction is passed through  $N1$  unrolls consisting of a CNN based denoiser  $R_w$  and  $N2$  CG update steps. b) The CNN uses a Residual CNN architecture with convolutional blocks ( $3 \times 3 \times 3$  kernel, 64 channel) and ReLU activations, which has been proved to be invertible.



**Video. S1.** A dynamic video for a representative reconstruction results on a prospectively undersampled cardiac cine scan (See attached MP4 file).



**Fig. S2.** A representative reconstruction results on prospectively undersampled 3D FSE knee scan with the model trained on retrospectively undersampled knee data using different methods (PICS, 2D MoDL and 3D MoDL with MEL). Representative slices from 3 different views (Axial, Sagittal, Coronal) are shown here.