Greedy Learning for Large-Scale Neural MRI Reconstruction

Batu Ozturkler[1], Arda Sahiner[1], Tolga Ergen[1], Arjun D Desai[1], Shreyas Vasanawala[2], John M Pauly[1], Morteza Mardani[1], Mert Pilanci[1]

[1]*Electrical Engineering, Stanford University, Stanford, CA, United States*
[2]*Radiology, Stanford University, Stanford, CA, United States*

Introduction

- Model-based networks showed state-of-the-art performance for undersampled MRI reconstruction [1].
- However, training these networks end-to-end requires prohibitively intensive **memory** and **computation time, limiting their applicability** for high-dimensional imaging [2].
- First proposed for image classification [3], **greedy learning** splits the end-to-end network into decoupled network modules and performs gradient updates on each module independently which reduces memory footprint**.**
- We propose greedy learning for MRI reconstruction**,** which requires **6x less GPU memory during training,** and **achieves the same generalization performance as backpropagation.**

Background

- The forward model of undersampled MRI with parallel imaging and compressed sensing can be modeled as:

$$\boxed{y = UFSx + \epsilon}$$

where y is the observed measurements, x is the real image, S are sensitivity maps, F is the Fourier transform, U is the undersampling mask, and ϵ is additive noise.

- Then, undersampled MRI reconstruction can be formulated as:

$$\boxed{\hat{x} = \arg\min_{x} \frac{1}{2}\|Ax - y\|_2^2 + \lambda R(x)}$$

where $A = UFS$ is the forward model, R is a regularization function, and λ is the regularization strength.

- This optimization problem can be solved using **proximal gradient descent**:

$$x^{(i+)} = \mathbf{DC}(x^{(i)}) = x^{(i-1)} - 2tA^H(Ax^{(i)} - y)$$

$$x^{(i+1)} = f_{\theta_i}(x^{(i+)})$$

where parameters θ_i, $i \in 1,..,N$ of network f with N unrolled iterations are updated with backpropagation.

Methods

Greedy Learning: Split network with N unrolled iterations to M modules and update each set of parameters θ_i independently (Fig. 1).

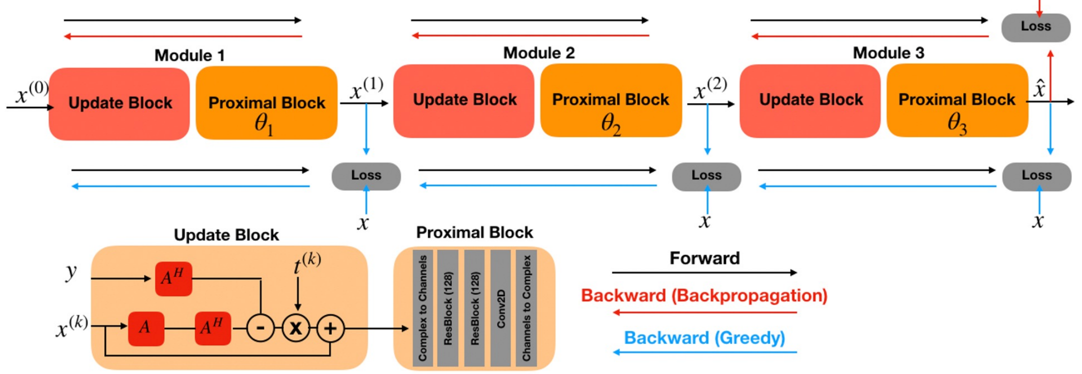

Figure 1: (a) Backpropagation. Forward pass is computed through all modules, and gradient updates are performed for all modules in the same backward pass. **(b) Greedy Learning**. At the end of each proximal block, loss is computed, and a local gradient update is performed on the current module. **(c)** Update block consists of data-consistency with measurements, and the proximal block consists of residual blocks.

Results

- **Dataset:** Fully-sampled 3D fast-spin echo (FSE) multi-coil knee MRI dataset available in mridata.org [4].
- **Setup:** Compare backpropagation with greedy learning where $M = \{4, 8\}$, and acceleration factor $R = \{12, 16\}$.
- Greedy learning achieves on par image quality (Fig. 2) and reconstruction performance with backpropagation in PSNR, SSIM, nRMSE with less memory (Table 1).

R	Method	SSIM	nRMSE	PSNR (dB)	Memory (MB)
12x	Backpropagation	0.896 (0.006)	0.127 (0.007)	40.06 (0.33)	10016
	Greedy ($M = 8$)	**0.919 (0.001)**	**0.124 (0.008)**	**40.24 (0.36)**	**1679**
	Greedy ($M = 4$)	0.903 (0.004)	0.126 (0.007)	40.12 (0.33)	2603
16x	Backpropagation	0.887 (0.007)	0.134 (0.008)	39.61 (0.33)	10016
	Greedy ($M = 8$)	**0.912 (0.001)**	**0.131 (0.008)**	**39.79 (0.36)**	**1679**
	Greedy ($M = 4$)	0.888 (0.007)	0.134 (0.008)	39.60 (0.32)	2603

Table 1: Comparison of test performance, maximum GPU memory during training for backpropagation and greedy learning with $M = 4$, and $M = 8$. Metrics are reported as mean (standard deviation).

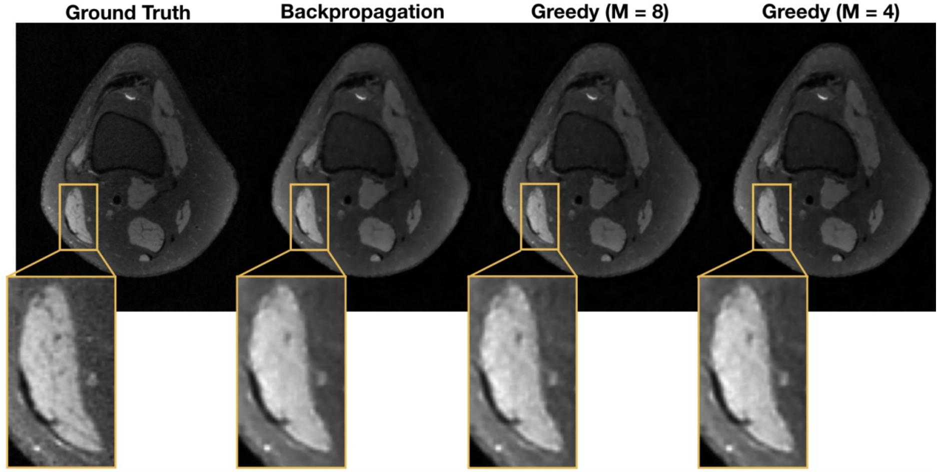

Figure 2: Reconstruction of a representative test knee slice with backpropagation, greedy learning with $M = 8$, and $M = 4$.

- On a single GPU, greedy learning and backpropagation have similar computation time. When independent modules in greedy learning are split across 2 GPUs, backward time is reduced which speeds up training. (Fig. 3)

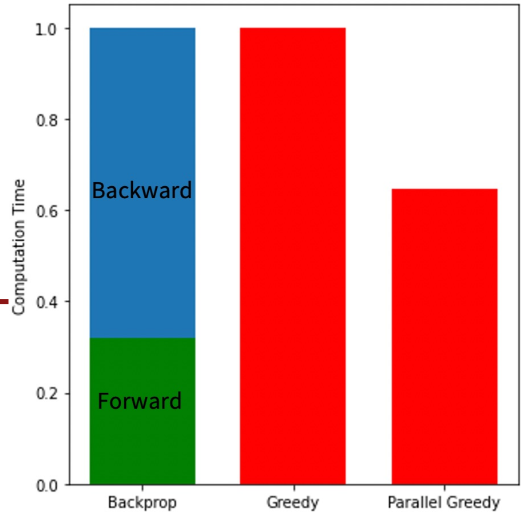

Figure 3: Computation time for greedy learning and backpropagation.

Conclusion

- Greedy learning for MRI reconstruction reduces memory footprint while preserving generalization performance, and can be applied to larger dimensional problems such as cardiac cine MRI.

References

[1] 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, Feb 2019.
[2] Morteza Mardani, Qingyun Sun, Shreyas Vasawanala, Vardan Papyan, Hatef Monajemi, John M. Pauly, and David L. Donoho. Neural proximal gradient descent for compressive imaging. In Proc. Neural Information Processing Systems (NeurIPS), 2018.
[3] Eugene Belilovsky, Michael Eickenberg, and Edouard Oyallon. Decoupled greedy learning of cnns, 2020.
[4] F Ong, S Amin, S Vasanawala, and M Lustig. Mridata. org: An open archive for sharing mri raw data. In Proc. Intl. Soc. Mag. Reson. Med, volume 26, 2018.