Memory-efficient Learning for Large-scale Computational Imaging

Michael Kellman*, Jon Tamir, Emrah Bostan, Michael Lustig & Laura Waller

Department of Electrical Engineering and Computer Sciences, University of California, Berkeley

*kellman@berkeley.edu

Overview

! Learning for large-scale computational imaging systems with standard backpropagation is **limited due to the memory capacity** of modern graphical processing units.

+ We propose a method that enables learning for large-scale computational imaging systems using **constant memory**.

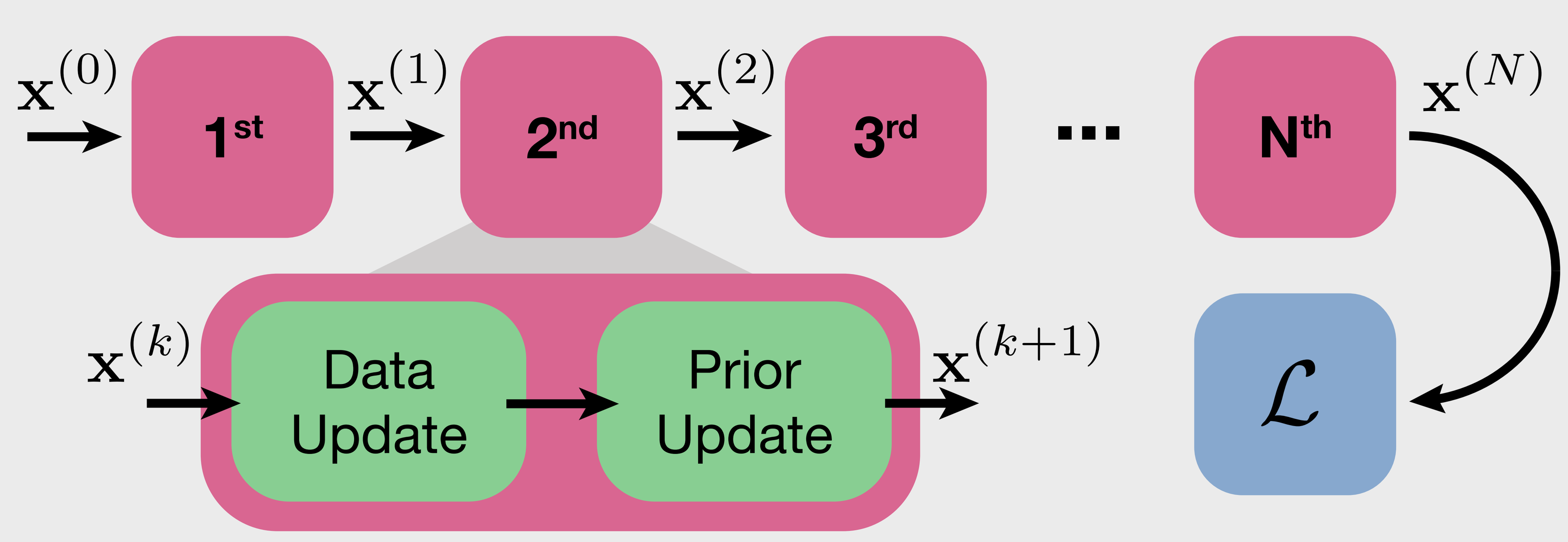
Introduction

Conventional Image Reconstruction:

$$\mathbf{x}^\star = \arg\min_{\mathbf{x}} \sum_{k=1}^{K} \|\mathbf{y}_k - \mathcal{A}_k(\mathbf{x})\|_2^2 + \mathcal{P}(\mathbf{x})$$

Data Consistency Term; Prior Term; Measurements; System Model

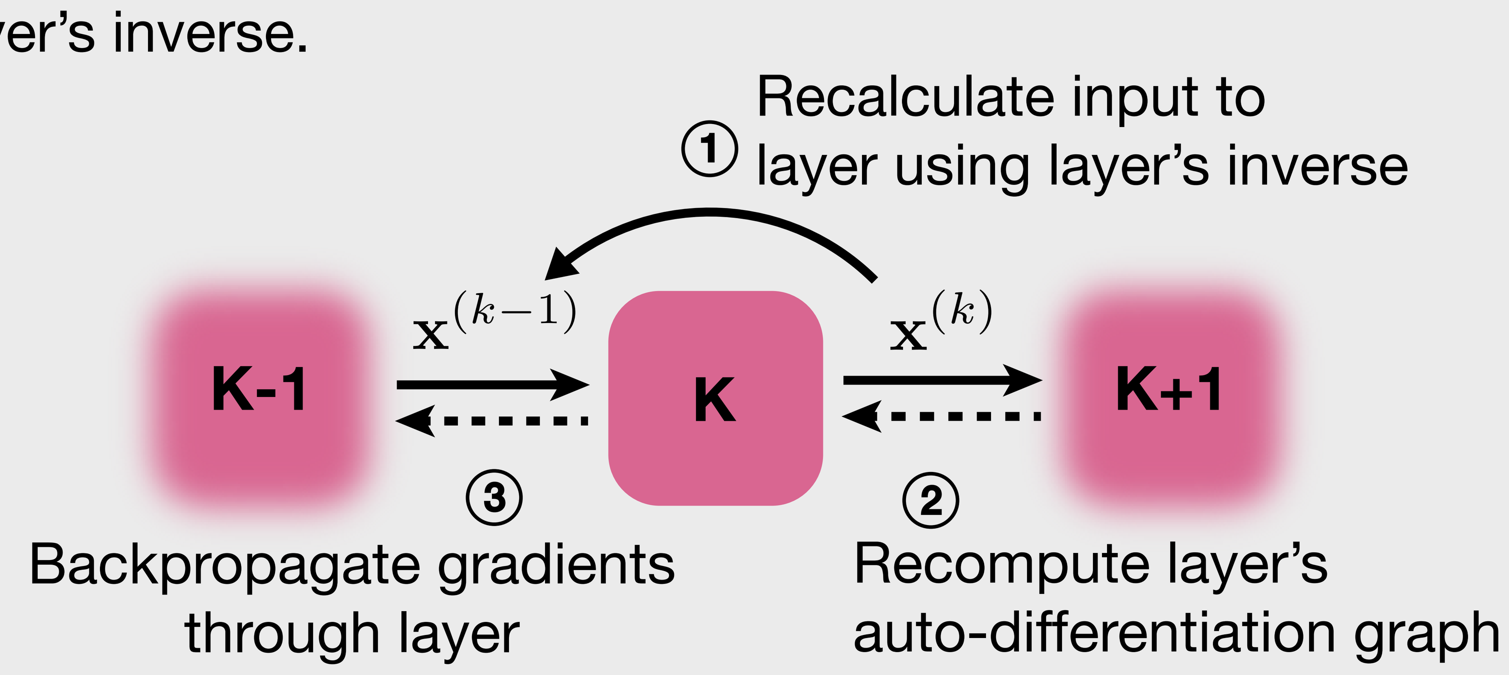
The system's image prior and experimental design can be **learned** to improve overall performance. Unroll the iterations of the reconstruction's optimizer to form the layers of a network [1]. Optimizers often alternate between applying data and prior updates to form the network:



Gradients are computed for the network using automatic differentiation [2]. For large-scale computational imaging systems, this will **exceed the GPU's memory capacity.**

Memory-efficient Learning

Rather than storing the whole graph for auto-differentiation, smaller graphs can be formed for each layer one at a time in reverse order by recalculating intermediate variables using each layer's inverse.



Gradient Layer

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \alpha\nabla_{\mathbf{x}} f(\mathbf{x}^{(k)})$$

$$\mathbf{x}^{(k-1)} \qquad \mathbf{x}^{(k)}$$

$$\mathbf{x}^{(k)} = \mathbf{x}^{(k+1)} + \alpha\nabla_{\mathbf{x}} f(\mathbf{x}^{(k)})$$

solved using fixed point method

Proximal Layer

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \nabla_{\mathbf{x}} p(\mathbf{x}^{(k+1)})$$

$$\mathbf{x}^{(k-1)} \qquad \mathbf{x}^{(k)}$$

$$\mathbf{x}^{(k)} = \mathbf{x}^{(k+1)} - \nabla_{\mathbf{x}} p(\mathbf{x}^{(k+1)})$$

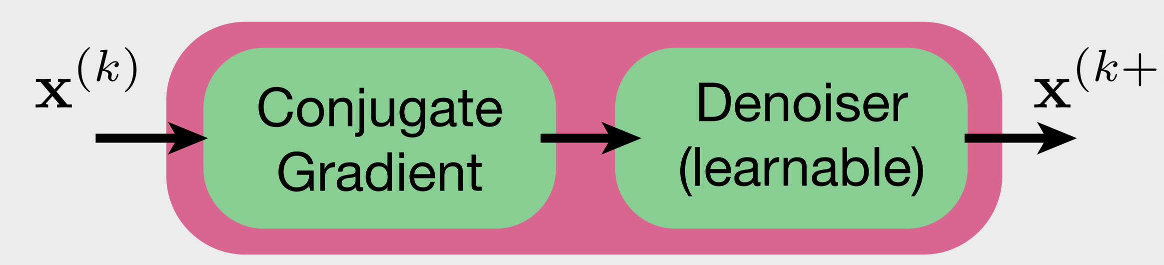
closed form solution

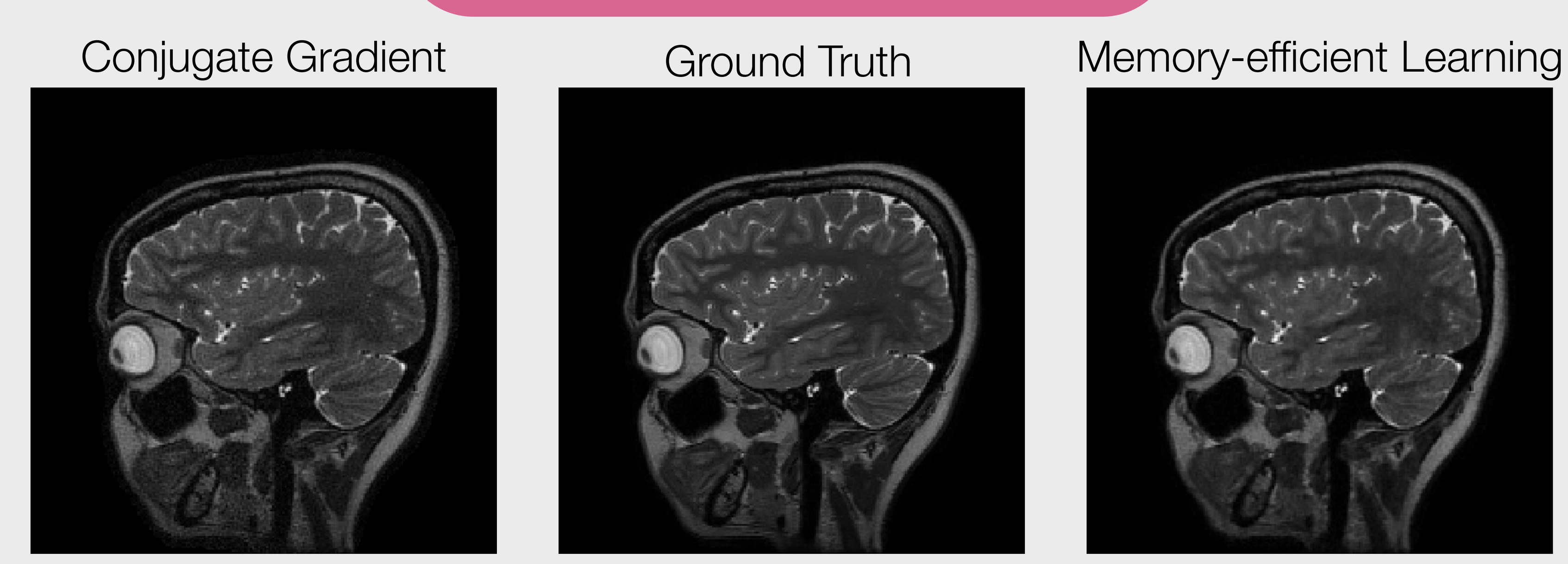
Results: Multi-channel MRI

+ **Multi-channel MRI** reduces scan time by undersampling and relying on image priors for reconstruction.

! **Learning image priors** for 3D multi-channel MRI [3] is **restricted by memory available on GPU.**

MR image reconstruction [3]:



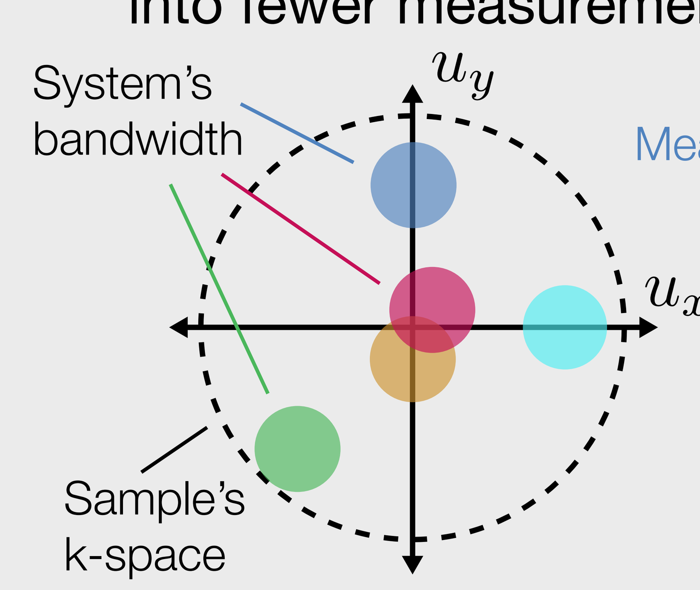


+ Our method enables learning at practical scales for this system, **ordinarily requiring >40GB using only 10GB**.

Results: Super Resolution Microscopy

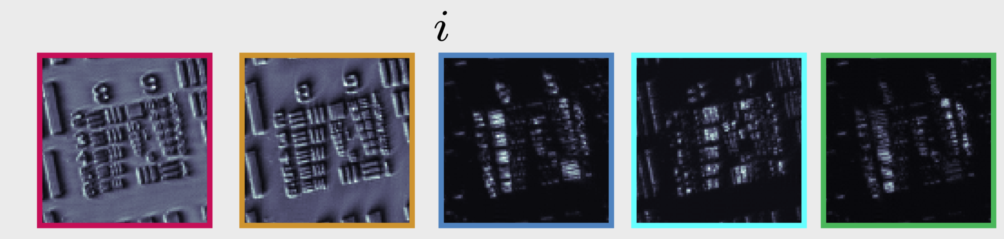
! **Fourier Ptychography** performs super resolution on an LED array microscope, however, has **poor time resolution.**

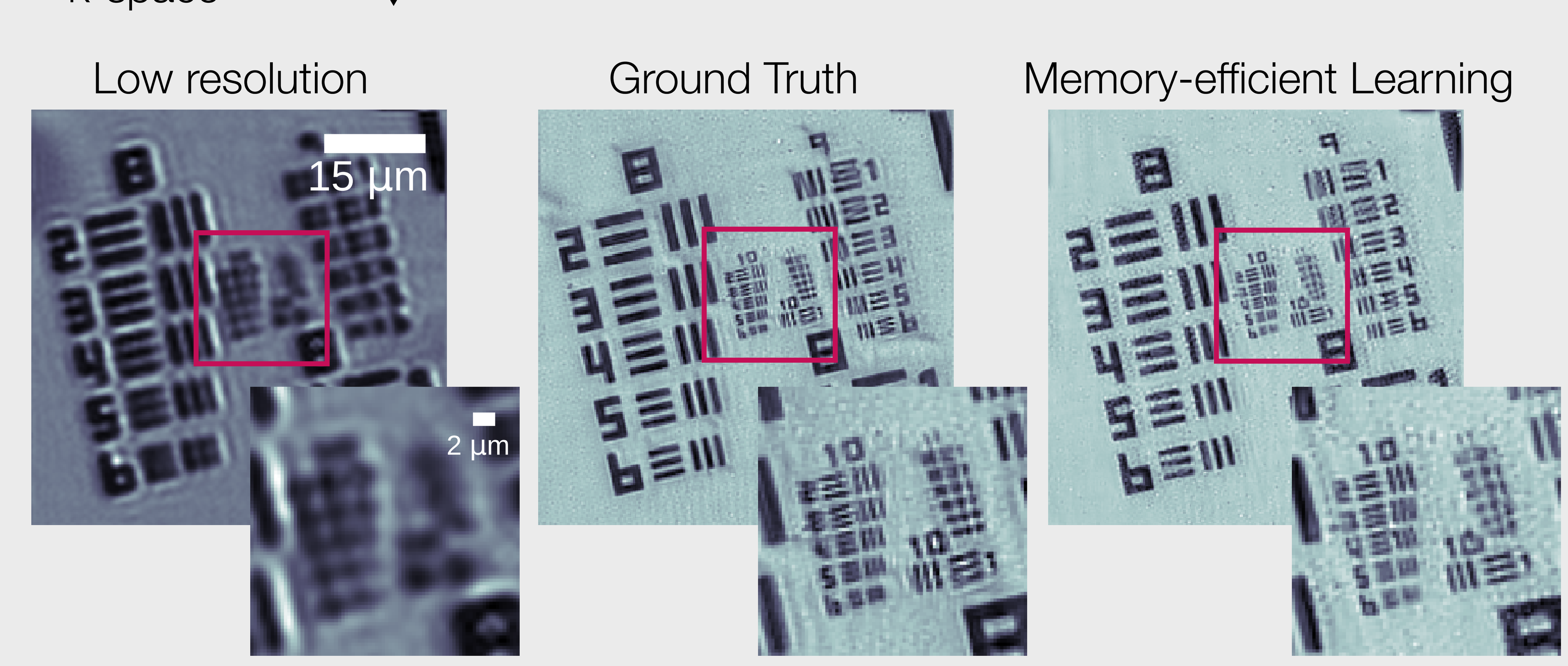
+ **Learn** the illumination design to **compress the information** into fewer measurements [4].



$$\mathbf{y_m} = \sum_i c_i |\mathbf{A}_i \mathbf{x}|^2$$

Measurement; LED Brightness (learnable); System Model





+ Our method enables learning at practical scales for this system, **ordinarily requiring 500GB using only 3GB**.

References

[1] K. Gregor and Y. LeCun. "Learning fast approximations of sparse coding." International Conference on Machine Learning. 2010.

[2] A. Griewank and A. Walther, "Evaluating derivatives: principles and techniques of algorithmic differentiation," SIAM, 2008.

[3] H. K. Aggarwal, M. P. Mani and M. Jacob, "Modl: Model-based deep learning architecture for inverse problems," IEEE transactions on medical imaging, 2018.

[4] M. R. Kellman, E. Bostan, N. Repina and L. Waller, "Physics-based Learned Design: Optimized Coded-Illumination for Quantitative Phase Imaging," IEEE Transactions on Computational Imaging, 2019.

M. Kellman is supported in part by the National Science Foundation (NSF) under grant DGE 1106400.

BERKELEY CENTER FOR COMPUTATIONAL IMAGING

BAIR
BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

Read more here!