Washington University in St. Louis

Computational Imaging: Reconciling **Models** and **Learning**

Ulugbek S. Kamilov
Computational Imaging Group (CIG)

Point your phone cam to connect on Twitter

IPAM UCLA (Los Angeles, CA) — 27 Jan 2020

@ukmlv • cigroup.wustl.edu • kamilov@wustl.edu

Deep learning has captured our collective imagination

Source: ??? (friend)

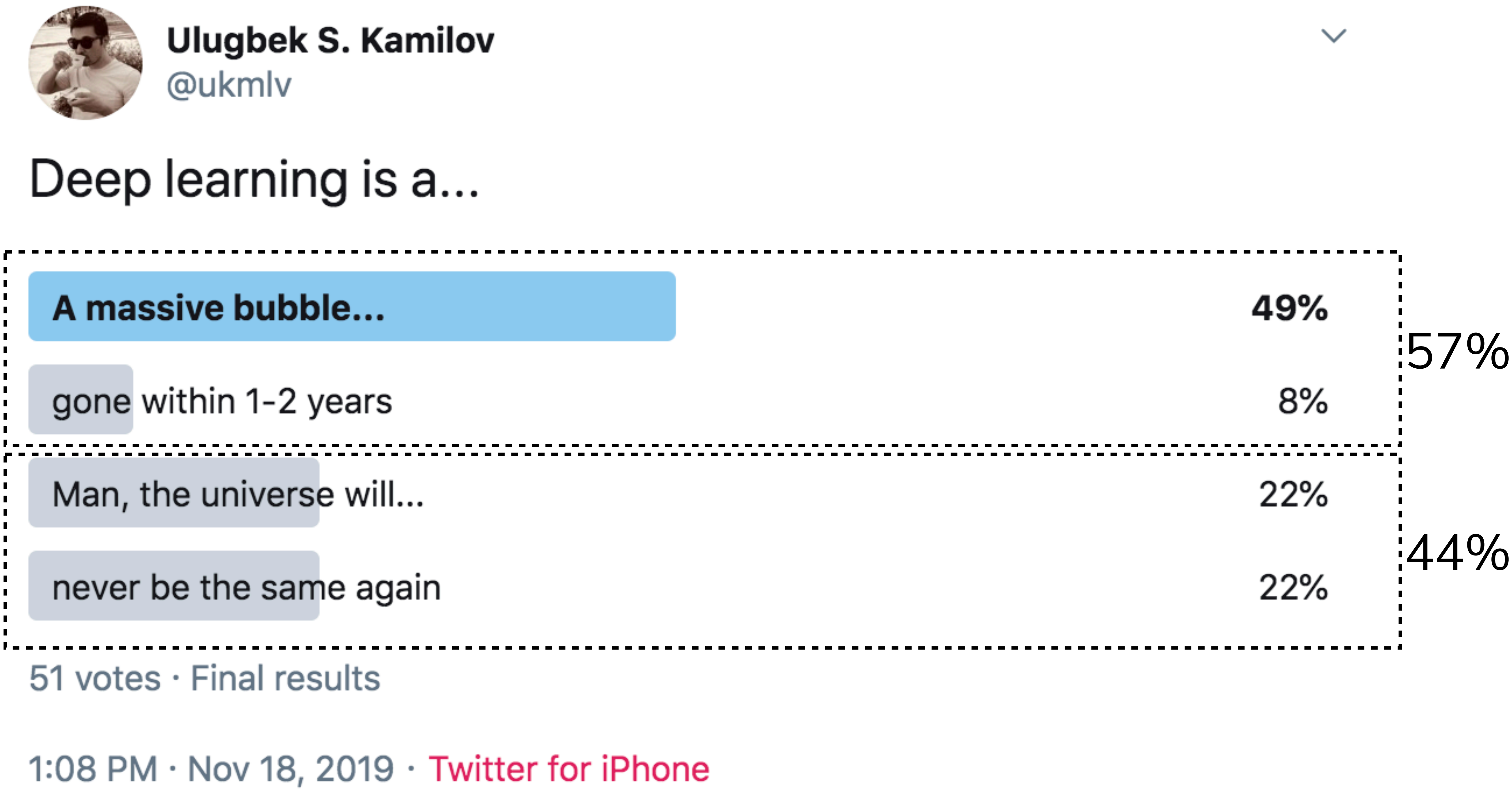
Ulugbek S. Kamilov
@ukmlv

Deep learning is a...

massive bubble…
gone within 1-2 years

universe will…
never be the same again

1:08 PM · Nov 18, 2019 · Twitter for iPhone

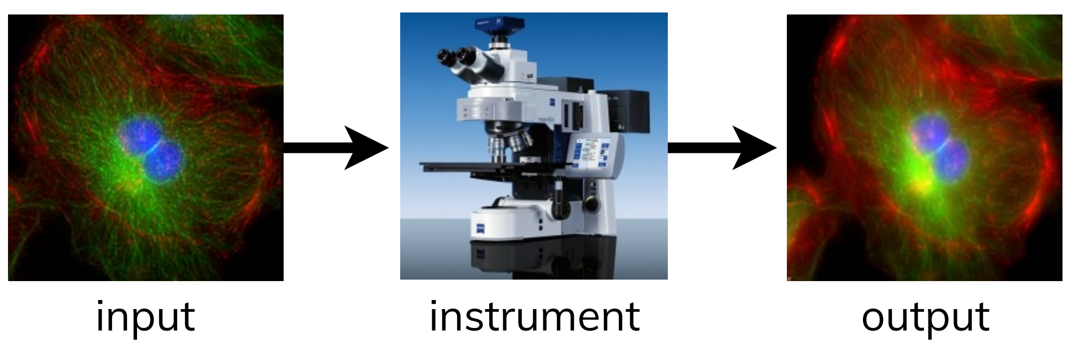


Ulugbek S. Kamilov
@ukmlv
Deep learning is a...
A massive bubble...
49%
gone within 1-2 years
8%
57%
Man, the universe will...
22%
never be the same again
22%
44%
51 votes · Final results
1:08 PM · Nov 18, 2019 · Twitter for iPhone

Imaging is going through a paradigm shift driven by machine learning

Imaging is going through a paradigm shift driven by machine learning

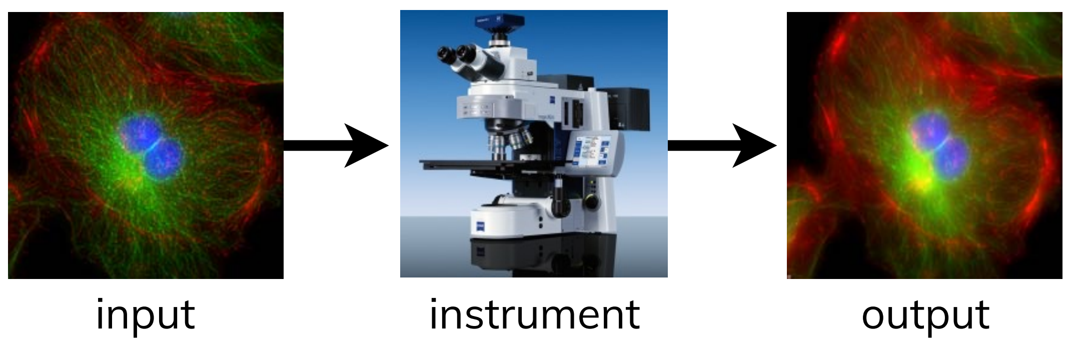
Past: Focus on hardware for image formation



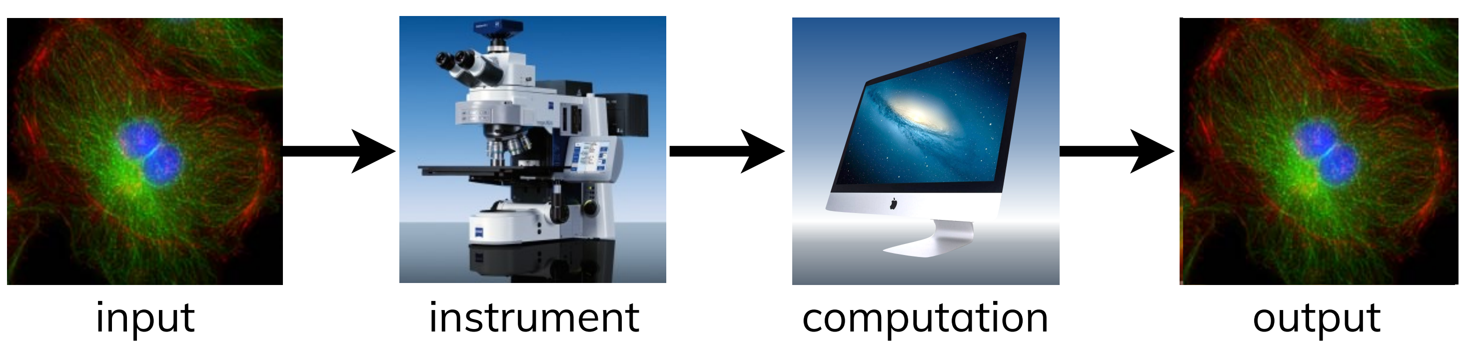
Washington University in St.Louis

Imaging is going through a paradigm shift driven by machine learning

Past: Focus on hardware for image formation



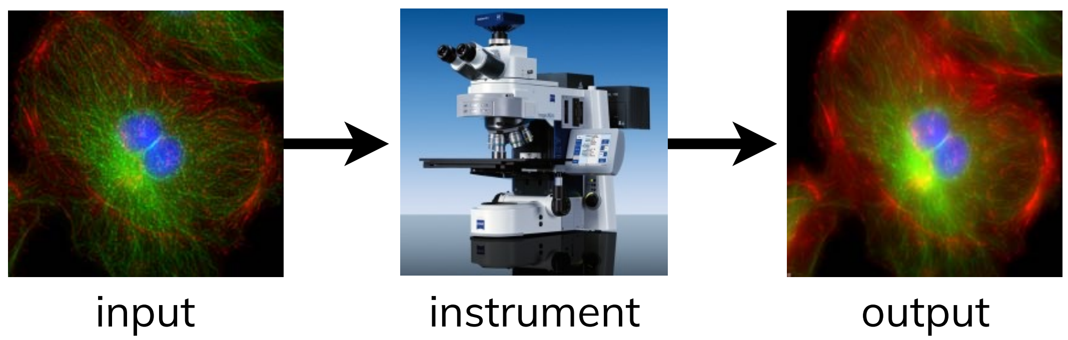
Present: Use digital signal processing for improved performance



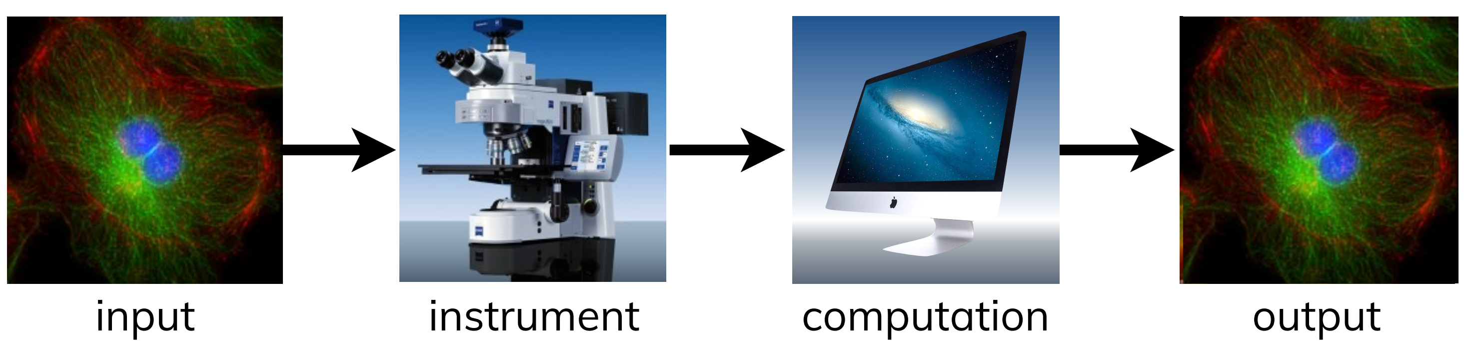
Washington University in St. Louis

Imaging is going through a paradigm shift driven by machine learning

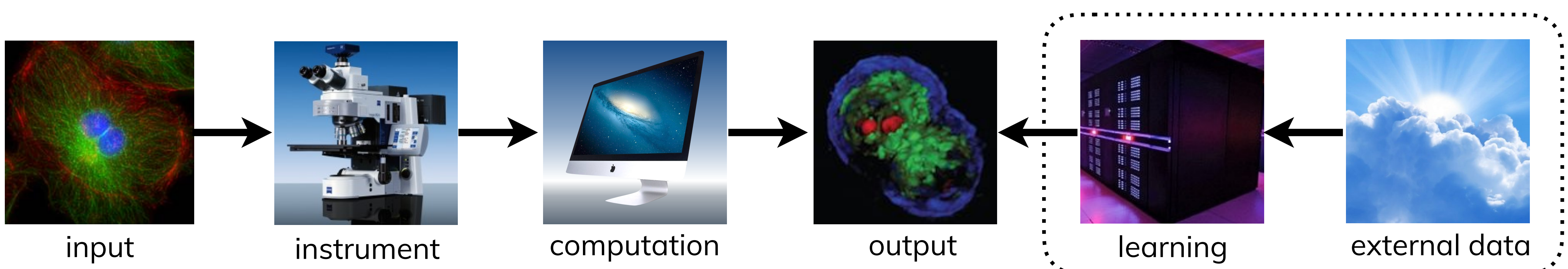
Past: Focus on hardware for image formation



Present: Use digital signal processing for improved performance

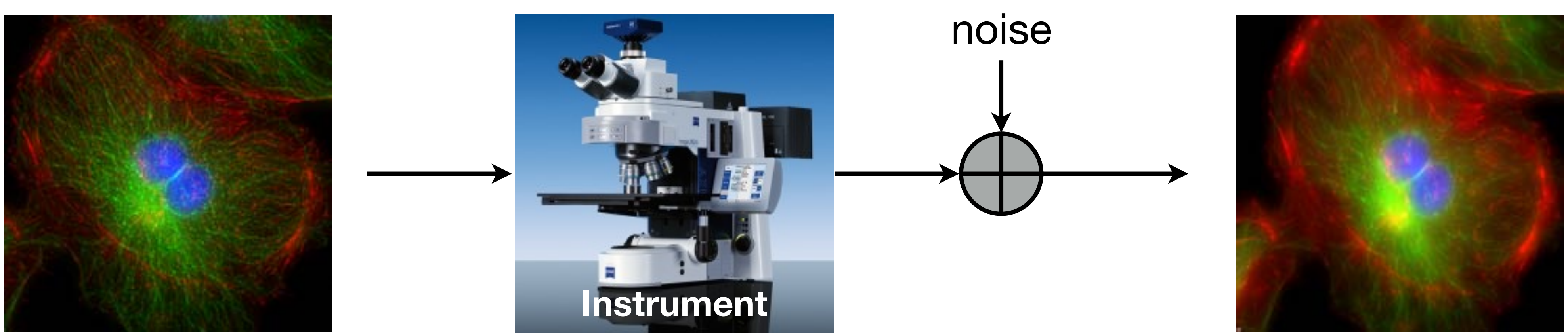


Future: Machine learning for retrieving ***hidden*** information



The vast majority of imaging problems can be formulated as inverse problems

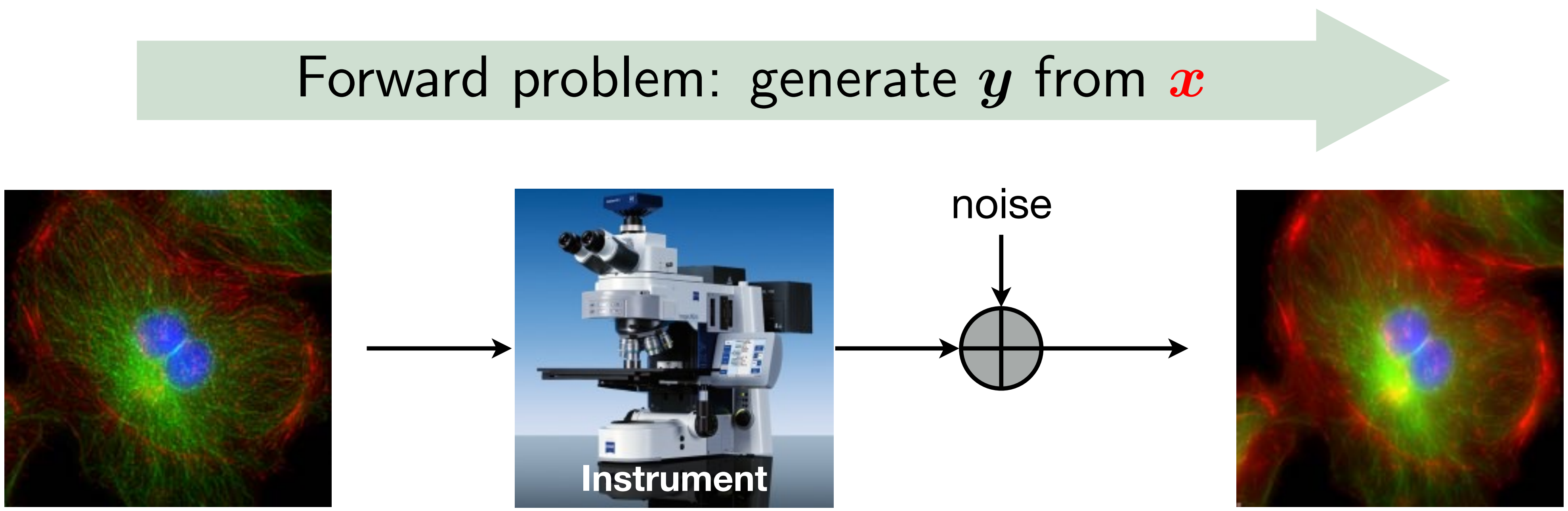
The vast majority of imaging problems can be formulated as inverse problems



$\boldsymbol{x}$ $\qquad$ $\mathbf{H}$ $\qquad$ $\boldsymbol{y} = \mathbf{H}(\boldsymbol{x}) + \boldsymbol{e}$

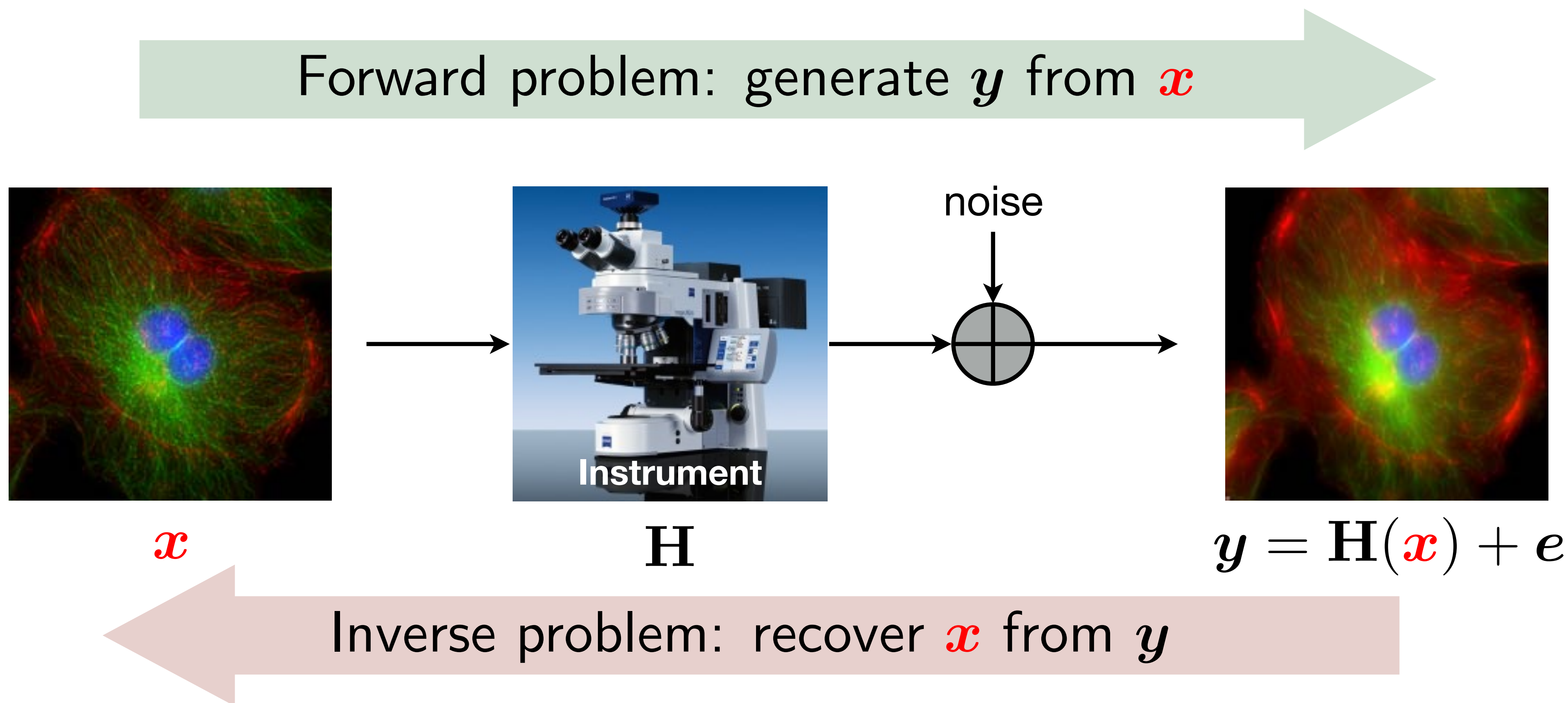
Source: Michael Unser

The vast majority of imaging problems can be formulated as inverse problems



$\boldsymbol{x}$ $\qquad$ $\mathbf{H}$ $\qquad$ $\boldsymbol{y} = \mathbf{H}(\boldsymbol{x}) + \boldsymbol{e}$

The vast majority of imaging problems can be formulated as inverse problems



Source: Michael Unser

The vast majority of imaging problems can be formulated as inverse problems

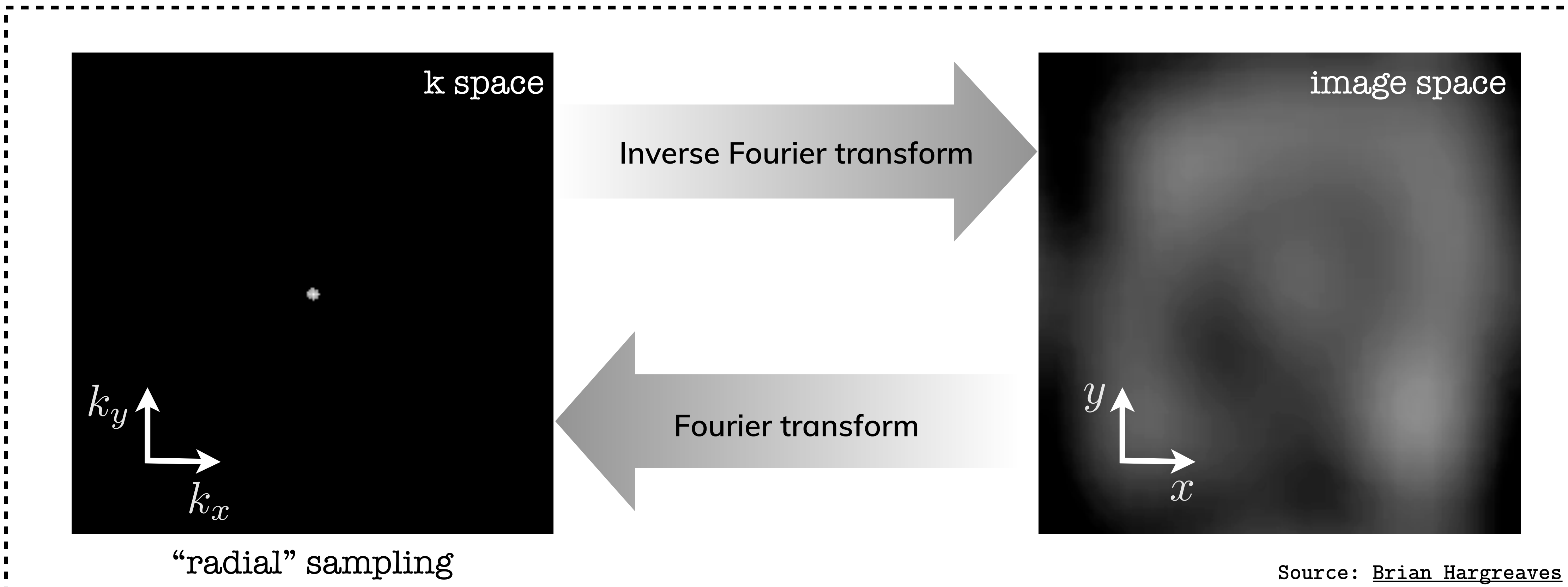
Imaging Problem	Rediation	Forward Model	Variations
2D or 3D tomography	coherent x-ray	$\boldsymbol{y}_i = \mathbf{R}_{\boldsymbol{\theta}_i}\boldsymbol{x}$	parallel, cone beam
3D deconvolution microscopy	fluorescence	$\boldsymbol{y} = \mathbf{H}\boldsymbol{x}$	brightfield, confocal, light sheet
structured illumination microscopy (SIM)	fluorescence	$\boldsymbol{y}_i = \mathbf{H}\mathbf{W}_i\boldsymbol{x}$	full 3D reconstruction, non-sinusoidal patters
positron emission tomography (PET)	gamma rays	$\boldsymbol{y}_i = \mathbf{H}_{\boldsymbol{\theta}_i}\boldsymbol{x}$	list mode with time-of-flight
magnetic resonance imaging (MRI)	radio frequency	$\boldsymbol{y} = \mathbf{S}\mathbf{F}\boldsymbol{x}$	uniform or nonuniform sampling in k-space
Cardiac MRI (parallel, nonuniform)	radio frequency	$\boldsymbol{y}_{t,i} = \mathbf{S}_t\mathbf{F}\mathbf{W}_i\boldsymbol{x}$	gated or nongated, retrospective registration
optical diffraction tomography (ODT)	coherent light	$\boldsymbol{y}_i = \mathbf{W}_i\mathbf{F}\boldsymbol{x}$	with holography or gating interferometry

Source: Michael Unser

Example #1: Magnetic resonance imaging (MRI) collects data in the spatial-frequency domain

Example #1: Magnetic resonance imaging (MRI) collects data in the spatial-frequency domain

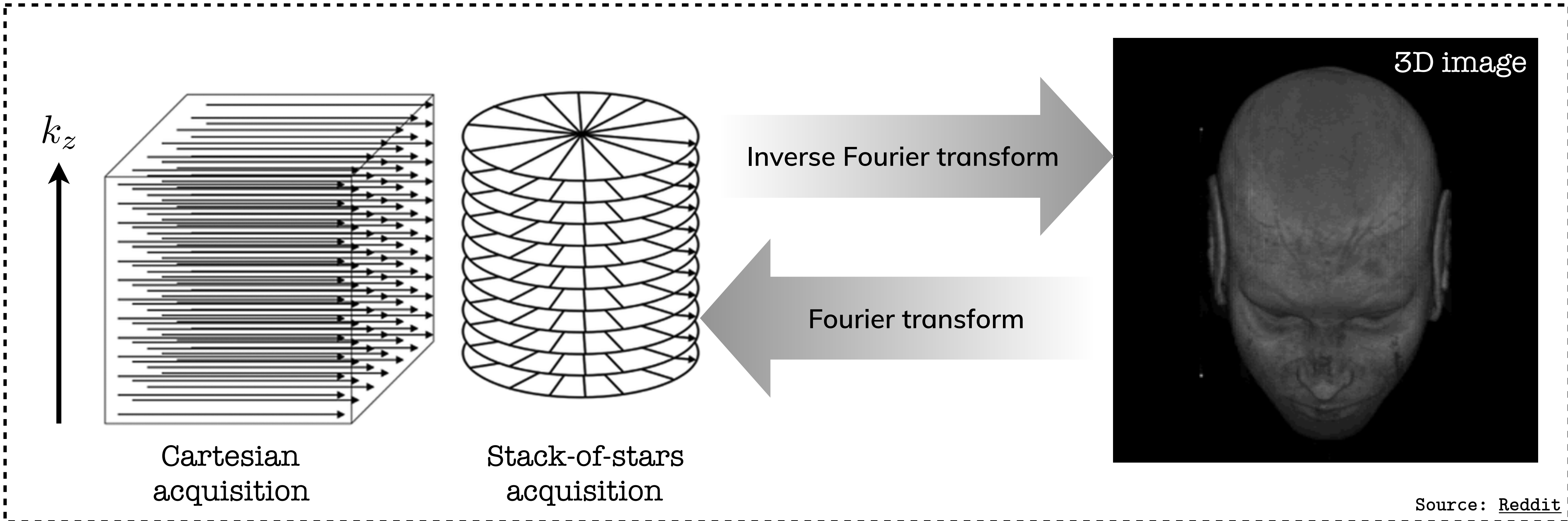
Example #1: Magnetic resonance imaging (MRI) collects data in the spatial-frequency domain



By sampling the entire k-space plane and taking the inverse Fourier transform, one can produce an image!

Eldeniz *et al.*, “CAPTURE: Consistently Acquired Projections for Tuned and Robust Estimation,” *Investigative Radiology*, vol. 53, no. 5, May 2018

Example #1: Magnetic resonance imaging (MRI) collects data in the spatial-frequency domain



Source: Reddit

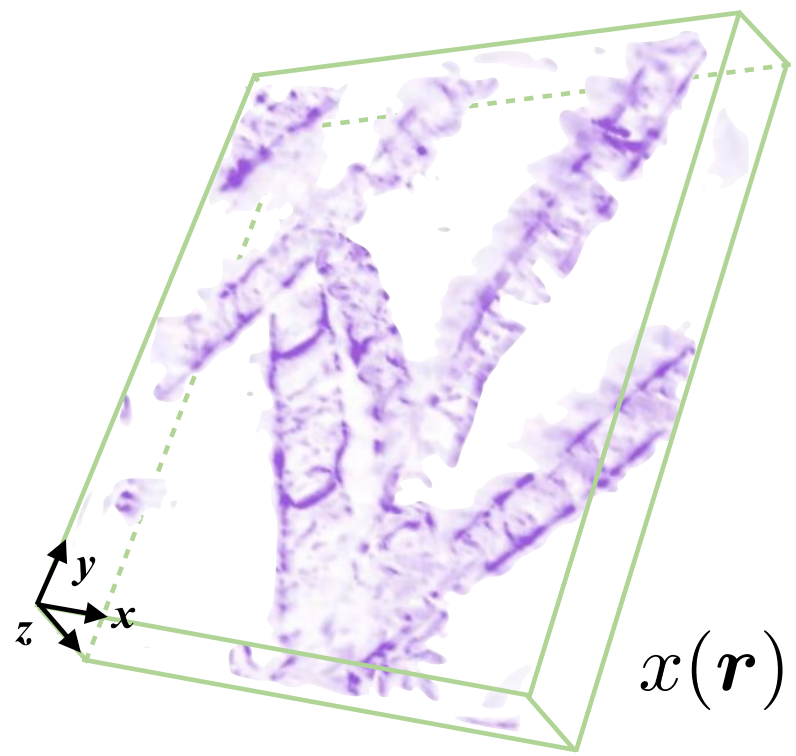
Volumetric images are formed by collecting the k-space data across multiple slices

Ichikawa *et al.*, "Utility of Stack-of-stars Acquisition for Hepatobiliary Phase Imaging without Breath-holding," *Magn Reson Med Sci*, May 2019

Example #2: Optical diffraction tomography (ODT) replaces x-rays with the visible light

Washington University in St. Louis

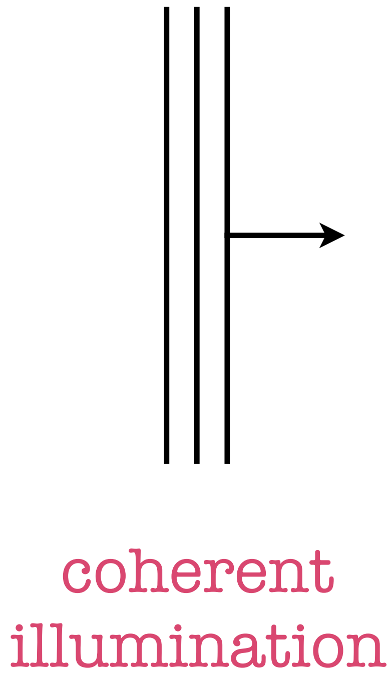
Example #2: Optical diffraction tomography (ODT) replaces x-rays with the visible light

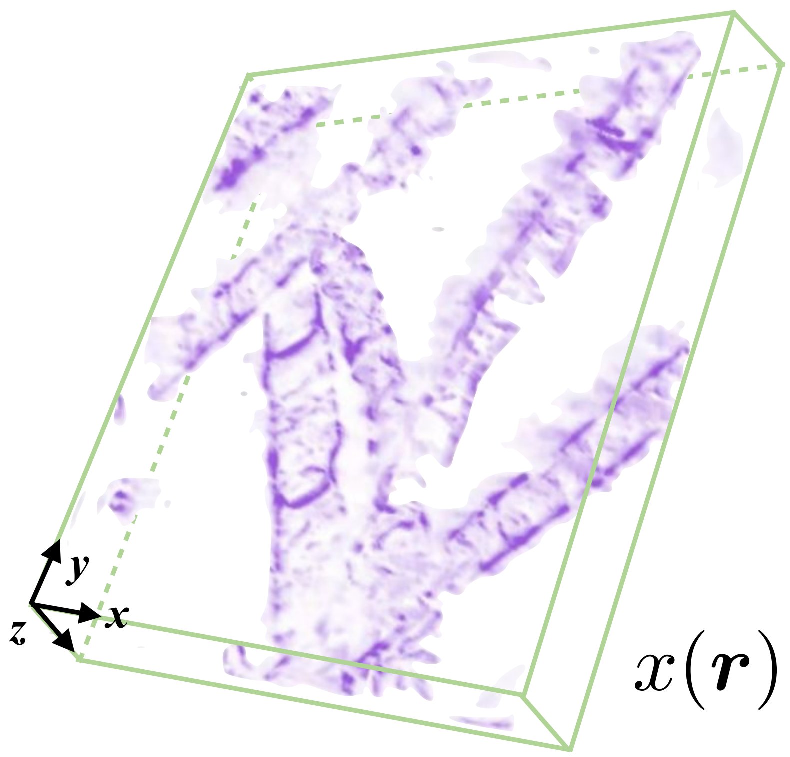


3D object of interest
(spirogyra algae cluster)

Wu *et al.*, "SIMBA: Scalable Inversion in Optical Tomography using Deep Denoising Priors," arXiv:1911.13241, 2019

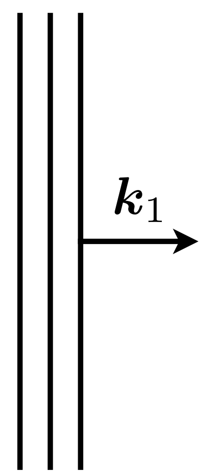
Example #2: Optical diffraction tomography (ODT) replaces x-rays with the visible light





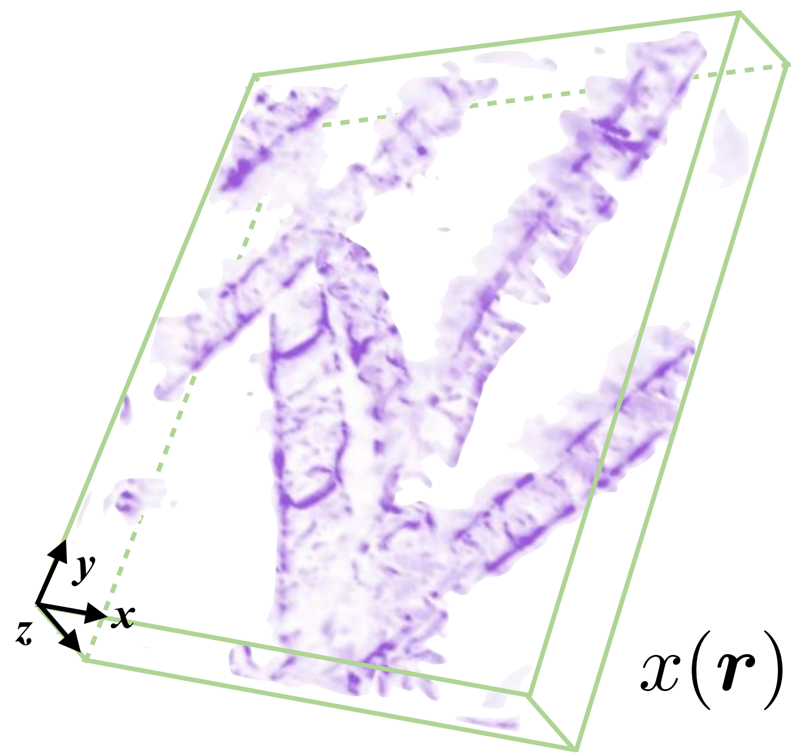
Wu *et al.*, “SIMBA: Scalable Inversion in Optical Tomography using Deep Denoising Priors,” arXiv:1911.13241, 2019

Example #2: Optical diffraction tomography (ODT) replaces x-rays with the visible light



coherent illumination

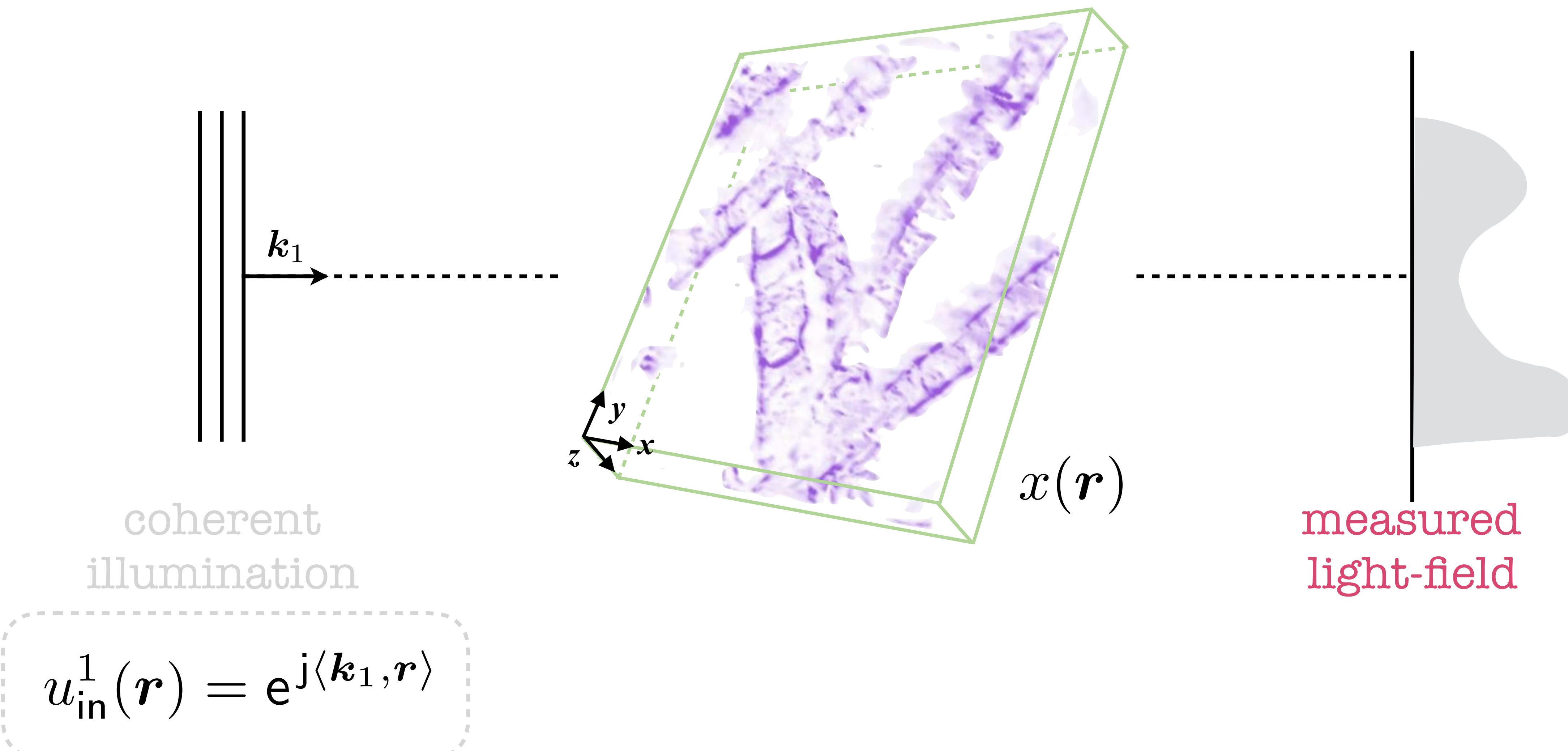
$$u^1_{\mathsf{in}}(\boldsymbol{r}) = \mathsf{e}^{\mathsf{j}\langle \boldsymbol{k}_1, \boldsymbol{r}\rangle}$$



$x(\boldsymbol{r})$

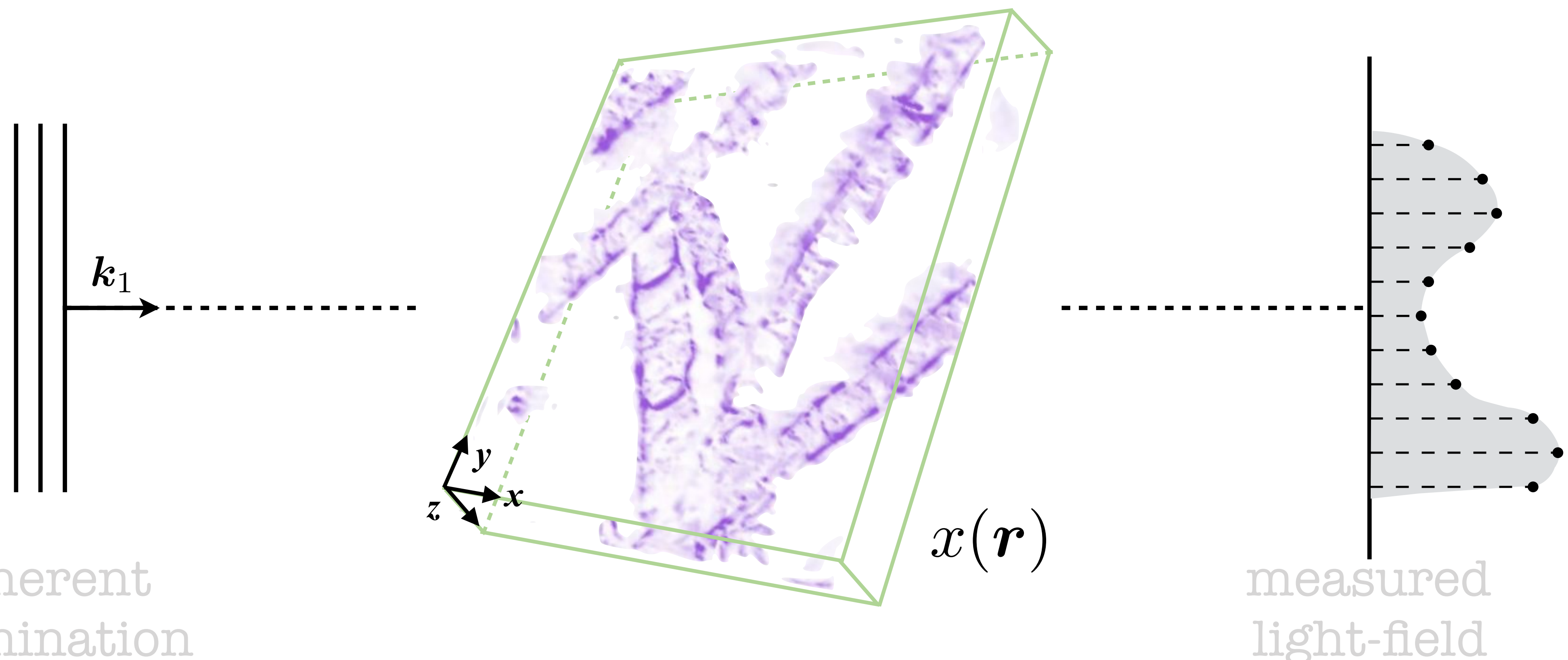
Wu *et al.*, "SIMBA: Scalable Inversion in Optical Tomography using Deep Denoising Priors," arXiv:1911.13241, 2019

Example #2: Optical diffraction tomography (ODT) replaces x-rays with the visible light



Wu *et al.*, "SIMBA: Scalable Inversion in Optical Tomography using Deep Denoising Priors," arXiv:1911.13241, 2019

Example #2: Optical diffraction tomography (ODT) replaces x-rays with the visible light

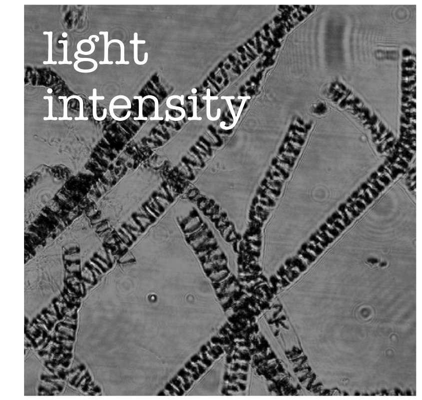


$$u_{\text{in}}^1(\boldsymbol{r}) = \mathrm{e}^{\mathrm{j}\langle \boldsymbol{k}_1, \boldsymbol{r} \rangle}$$

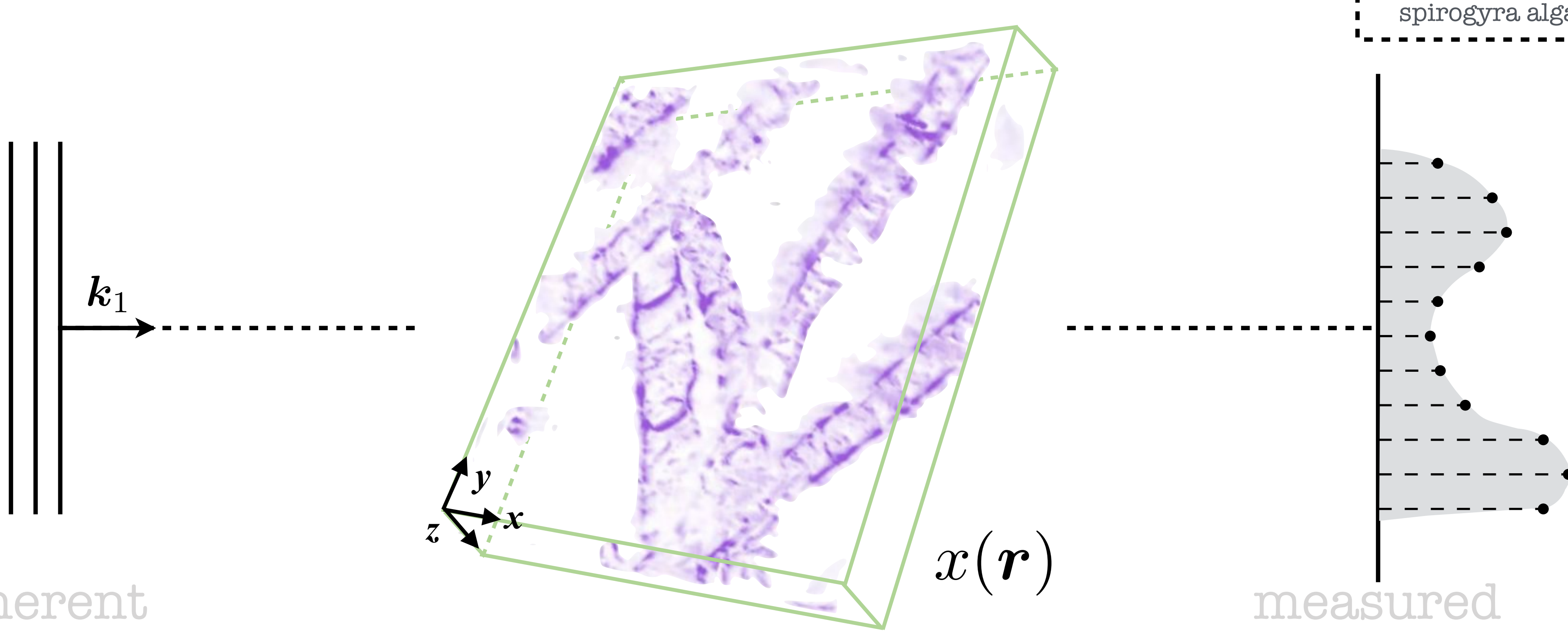
$\boldsymbol{y}_1$

Wu *et al.*, "SIMBA: Scalable Inversion in Optical Tomography using Deep Denoising Priors," arXiv:1911.13241, 2019

Example #2: Optical diffraction tomography (ODT) replaces x-rays with the visible light



spirogyra algae cluster



coherent illumination

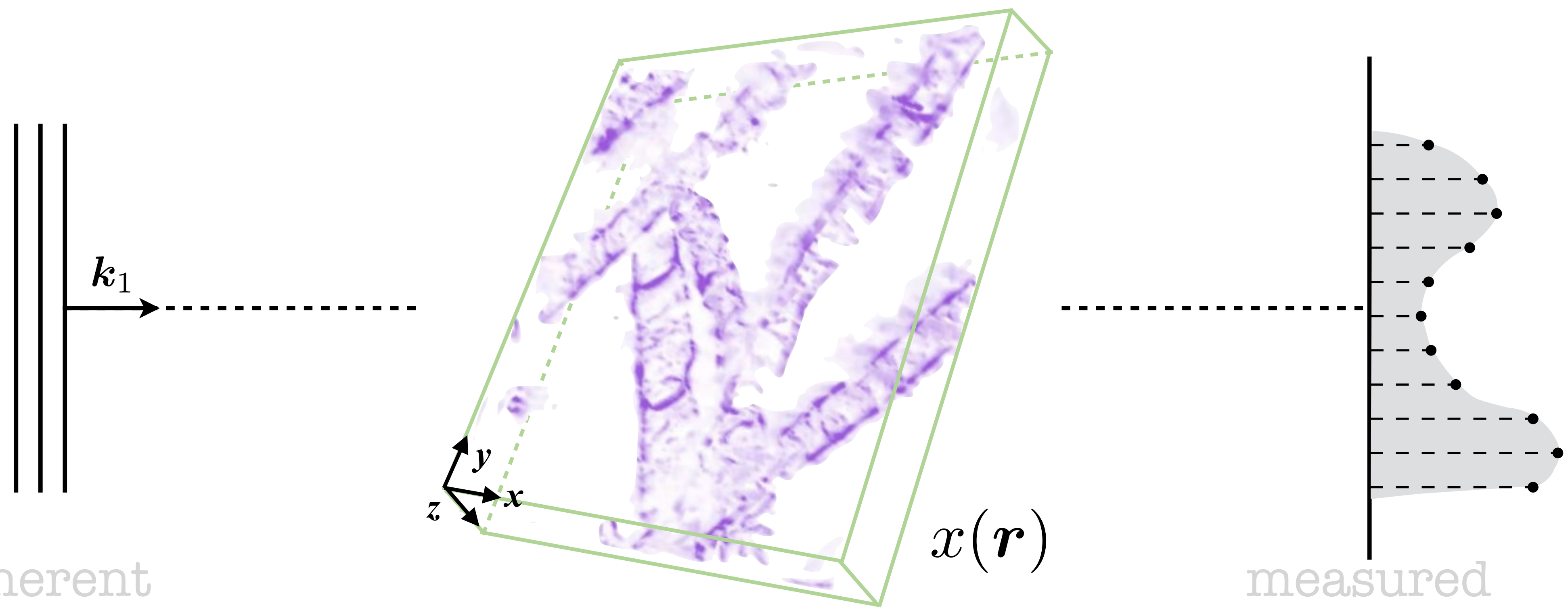
$$u_{\text{in}}^1(\boldsymbol{r}) = \mathrm{e}^{\mathrm{j}\langle \boldsymbol{k}_1, \boldsymbol{r} \rangle}$$

measured light-field

$$\boldsymbol{y}_1$$

Wu *et al.*, "SIMBA: Scalable Inversion in Optical Tomography using Deep Denoising Priors," arXiv:1911.13241, 2019

Example #2: Optical diffraction tomography (ODT) replaces x-rays with the visible light



coherent illumination

measured light-field

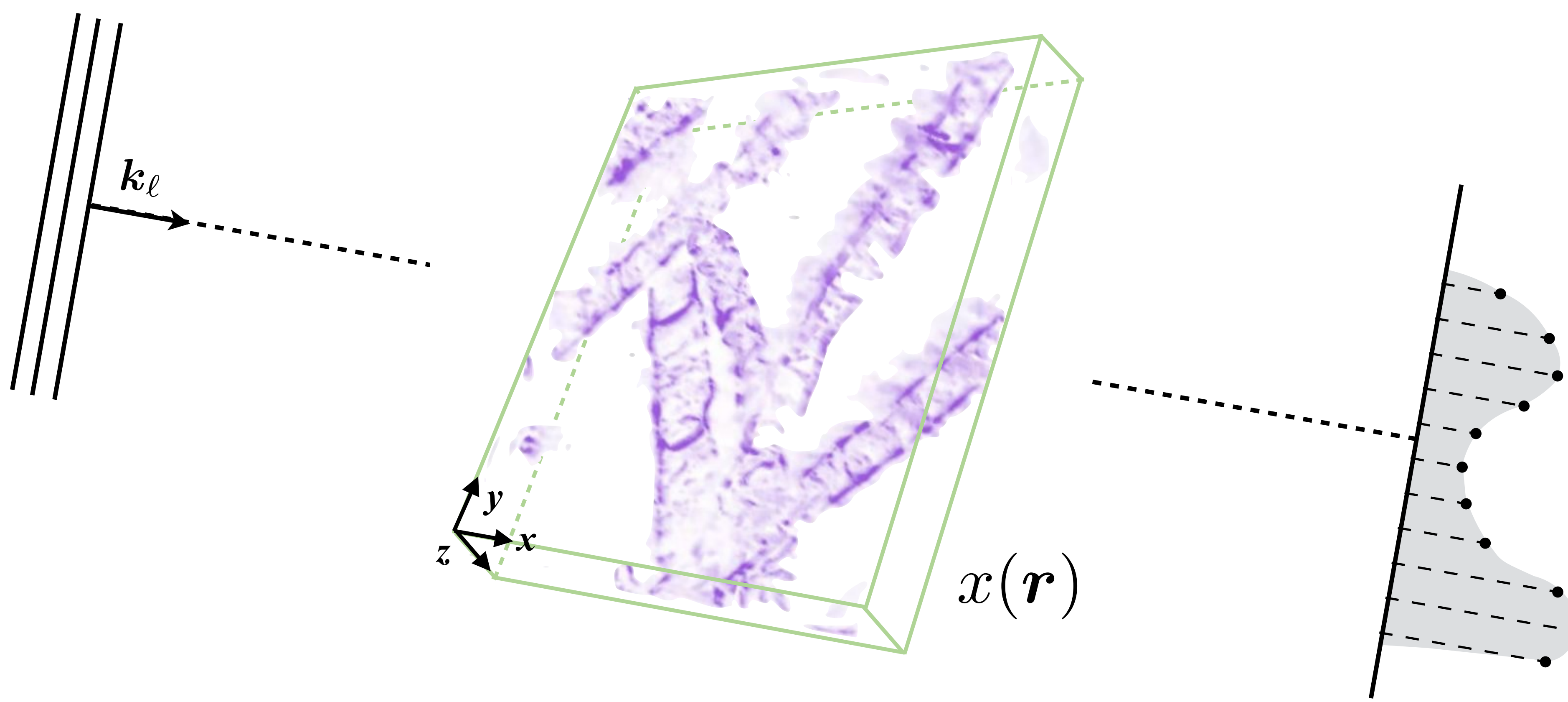
$$u_{\mathrm{in}}^{1}(\boldsymbol{r}) = \mathrm{e}^{\mathrm{j}\langle \boldsymbol{k}_1, \boldsymbol{r} \rangle}$$

$$\boldsymbol{y}_1$$

Wu *et al.*, "SIMBA: Scalable Inversion in Optical Tomography using Deep Denoising Priors," arXiv:1911.13241, 2019

Washington University in St. Louis

Example #2: Optical diffraction tomography (ODT) replaces x-rays with the visible light



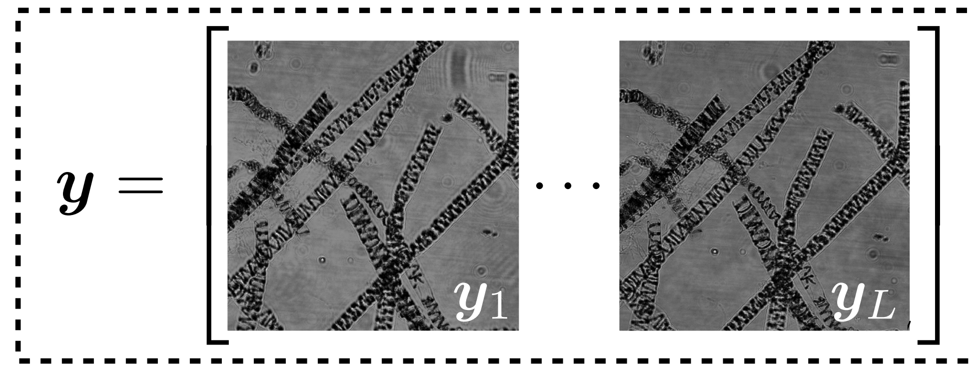
$$u_{\mathrm{in}}^{\ell}(\boldsymbol{r}) = \mathrm{e}^{\mathrm{j}\langle \boldsymbol{k}_\ell, \boldsymbol{r} \rangle}$$

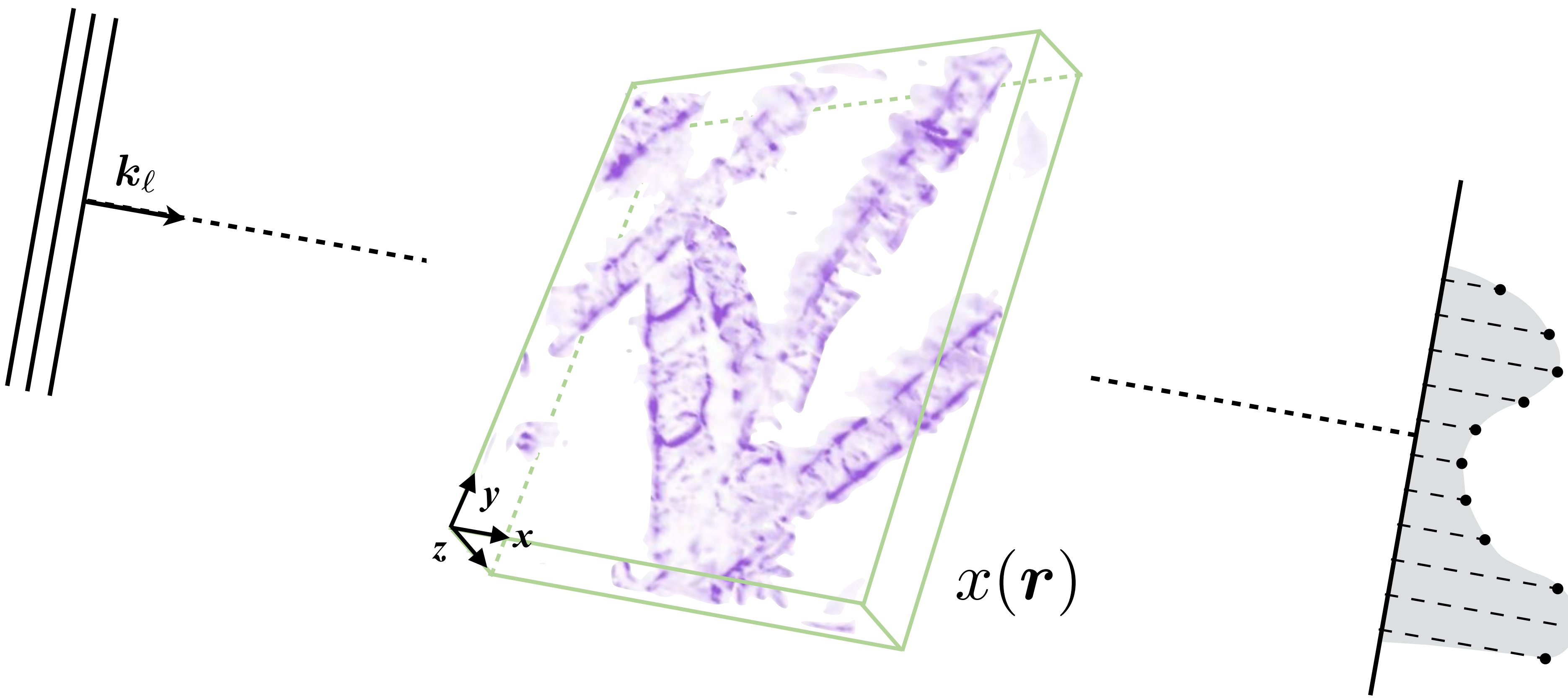
$$\boldsymbol{y}_\ell$$

Wu *et al.*, “SIMBA: Scalable Inversion in Optical Tomography using Deep Denoising Priors,” arXiv:1911.13241, 2019

Example #2: Optical diffraction tomography (ODT) replaces x-rays with the visible light

array of measurements





$$u_{\text{in}}^{\ell}(\boldsymbol{r}) = \mathrm{e}^{\mathrm{j}\langle \boldsymbol{k}_\ell, \boldsymbol{r}\rangle}$$

$$\boldsymbol{y}_\ell$$

Wu *et al.*, "SIMBA: Scalable Inversion in Optical Tomography using Deep Denoising Priors," arXiv:1911.13241, 2019

What are the typical challenges when solving imaging inverse problems?

What are the typical challenges when solving imaging inverse problems?

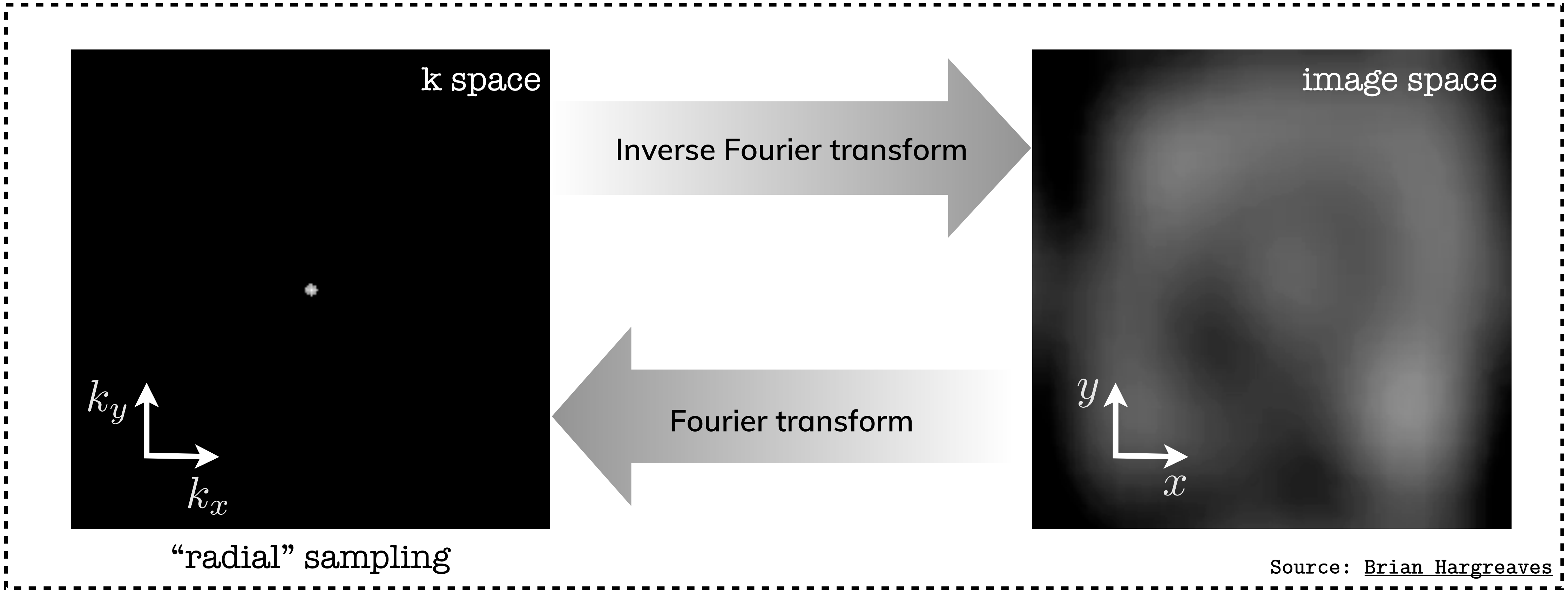
Slow acquisition:

Due to **sequential** and **indirect** acquisition of imaging data

What are the typical challenges when solving imaging inverse problems?

Slow acquisition:

Due to **sequential** and **indirect** acquisition of imaging data

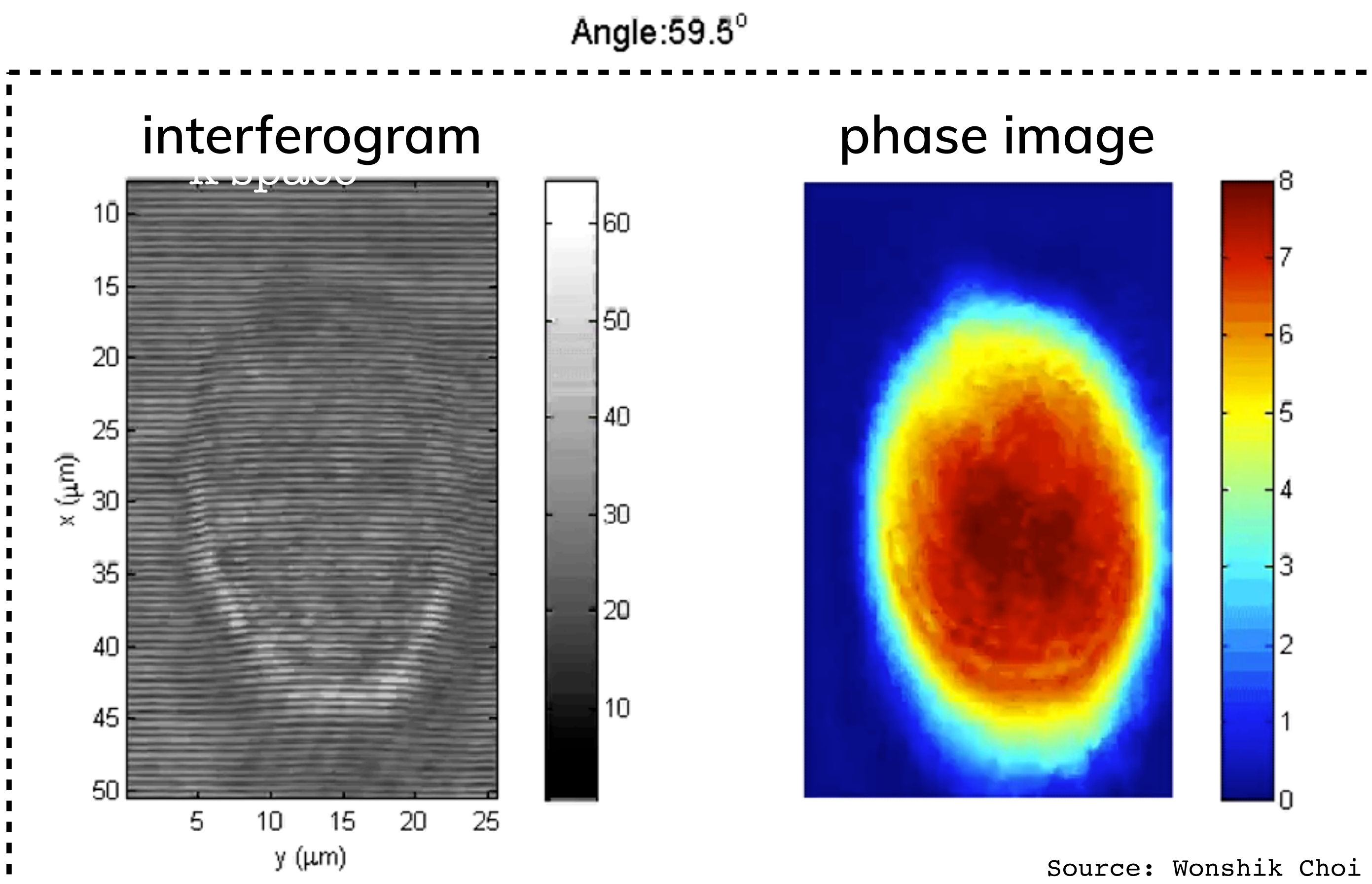


MRI: Need to sample a sufficient amount of k-space data to obtain a high-quality image

What are the typical challenges when solving imaging inverse problems?

Slow acquisition:

Due to **sequential** and **indirect** acquisition of imaging data



ODT: Need to sample a sufficient amount of projection data to obtain a high-quality image

Choi *et al.*, "Tomographic phase microscopy," *Nature Methods*, vol. 4, no. 9, pp. 717–719, September 2007

What are the typical challenges when solving imaging inverse problems?

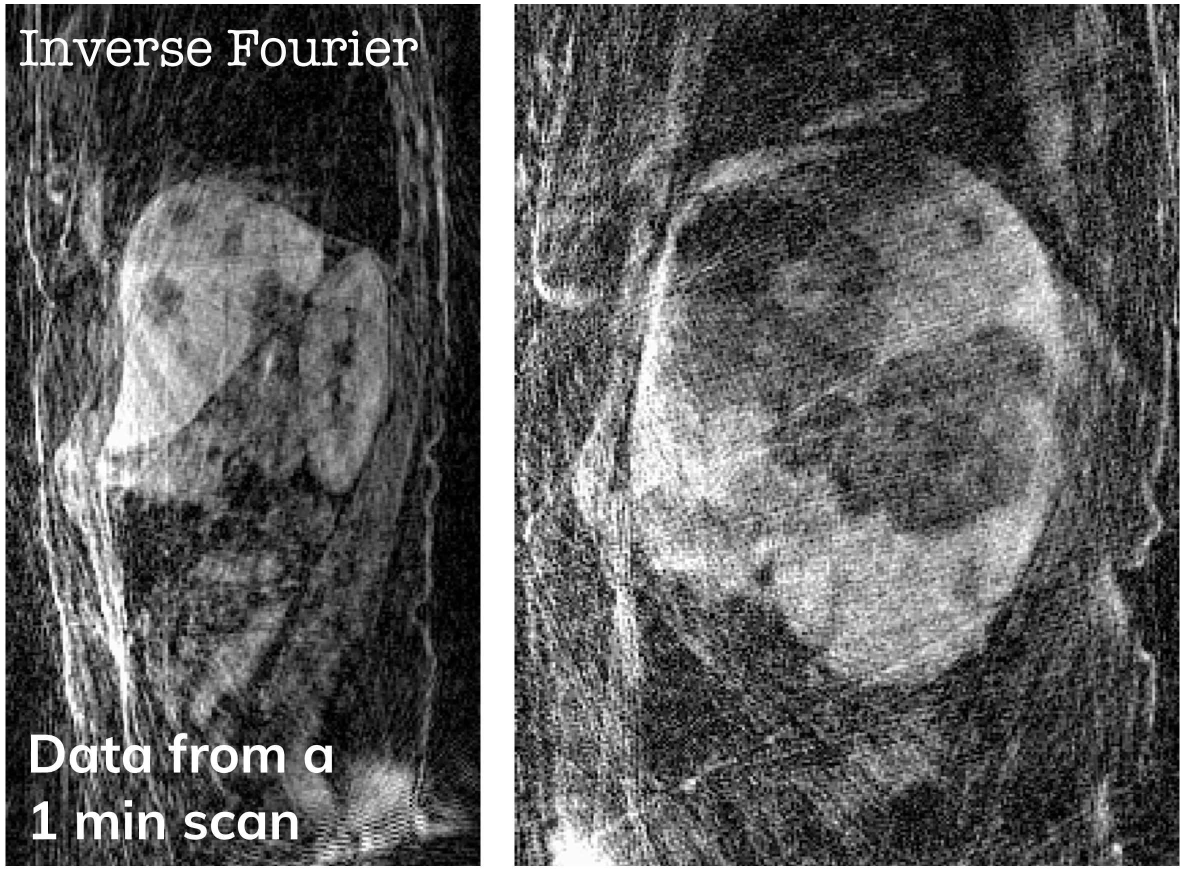
Slow acquisition:
Due to sequential and indirect acquisition of imaging data

Imaging artifacts:
Due to **subsampling**, **model mismatch**, and **noise**

What are the typical challenges when solving imaging inverse problems?

Slow acquisition:
Due to sequential and indirect acquisition of imaging data

Imaging artifacts:
Due to **subsampling**, **model mismatch**, and **noise**

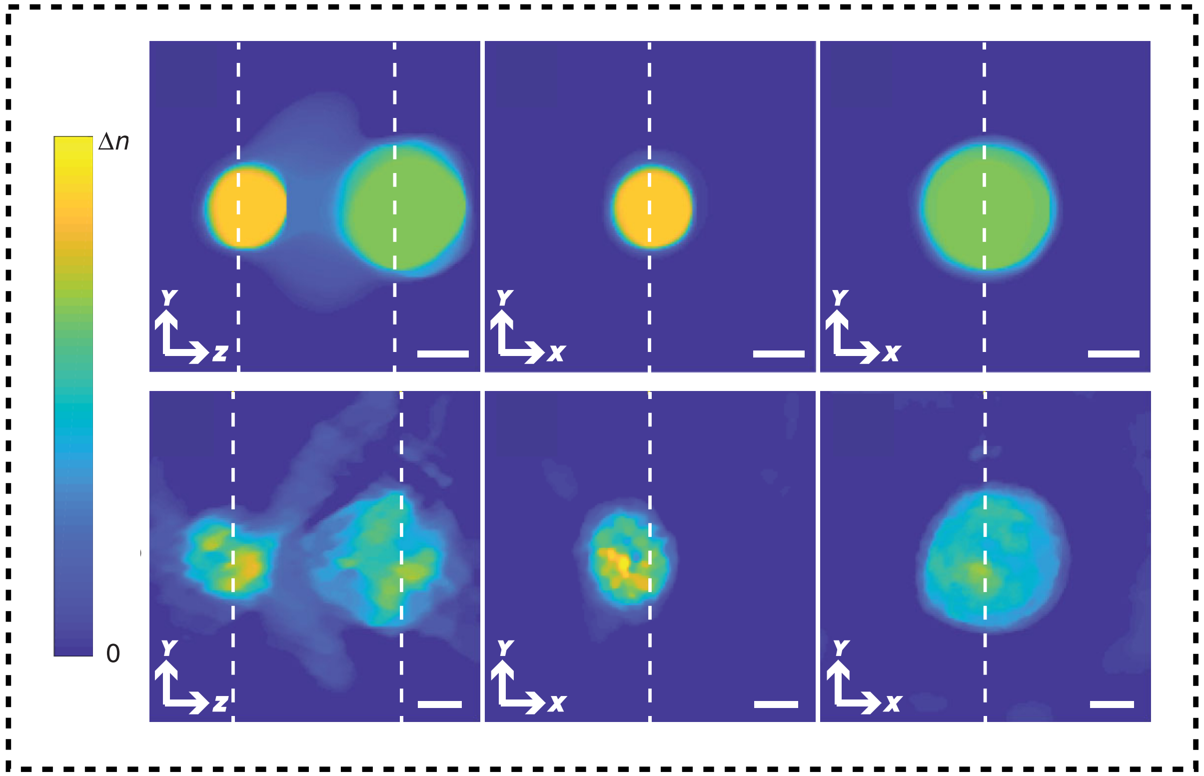


Example in free-breathing MRI: Streaking artifacts due to severe undersampling.

What are the typical challenges when solving imaging inverse problems?

Slow acquisition:
Due to sequential and indirect acquisition of imaging data

Imaging artifacts:
Due to **subsampling**, **model mismatch**, and **noise**



Examples in ODT:
Identical objects might end up looking different

Lim *et al.*, "Learning tomography assessed using Mie theory," *Physical Review Applied*, vol. 9, no. 034027, March 2018.

What are the typical challenges when solving imaging inverse problems?

Slow acquisition:
Due to sequential and indirect acquisition of imaging data

Imaging artifacts:
Due to subsampling, model mismatch, and noise

High computational and memory requirements:
Due to large volumes of data that need to be processed

What are the typical challenges when solving imaging inverse problems?

Slow acquisition:
Due to sequential and indirect acquisition of imaging data

Imaging artifacts:
Due to subsampling, model mismatch, and noise

High computational and memory requirements:
Due to large volumes of data that need to be processed

Source: Frank Ong

k-space data of 2 GB and
image of size 100 GB
(392 x 318 x 165 x 500 matrix)

What are the typical challenges when solving imaging inverse problems?

Slow acquisition:
Due to sequential and indirect acquisition of imaging data

Imaging artifacts:
Due to subsampling, model mismatch, and noise

High computational and memory requirements:
Due to large volumes of data that need to be processed

Goal: Overcome these limitations by bringing together physical models and learning as information sources

Washington University in St. Louis

Goal: Overcome these limitations by bringing together physical models and learning as information sources

IEEE TRANSACTIONS ON COMPUTATIONAL IMAGING, VOL. 5, NO. 3, SEPTEMBER 2019 395

An Online Plug-and-Play Algorithm for Regularized Image Reconstruction

Yu Sun, *Student Member, IEEE*, Brendt Wohlberg, *Senior Member, IEEE*, and Ulugbek S. Kamilov, *Member, IEEE*

PnP-SGD
IEEE TCI (2019)

Block Coordinate Regularization by Denoising

Yu Sun Washington University in St. Louis sun.yu@wustl.edu

Jiaming Liu Washington University in St. Louis jiaming.liu@wustl.edu

Ulugbek S. Kamilov Washington University in St. Louis kamilov@wustl.edu

BC-RED
NeurIPS (2019)

Online Regularization by Denoising with Applications to Phase Retrieval

Zihui Wu Yu Sun Jiaming Liu Ulugbek S. Kamilov
Washington University in St. Louis
{ray.wu, sun.yu, jiaming.liu, kamilov}@wustl.edu
https://cigroup.wustl.edu

SIMBA
ICCVW (2019)

RARE: Image Reconstruction using Deep Priors Learned without Ground Truth

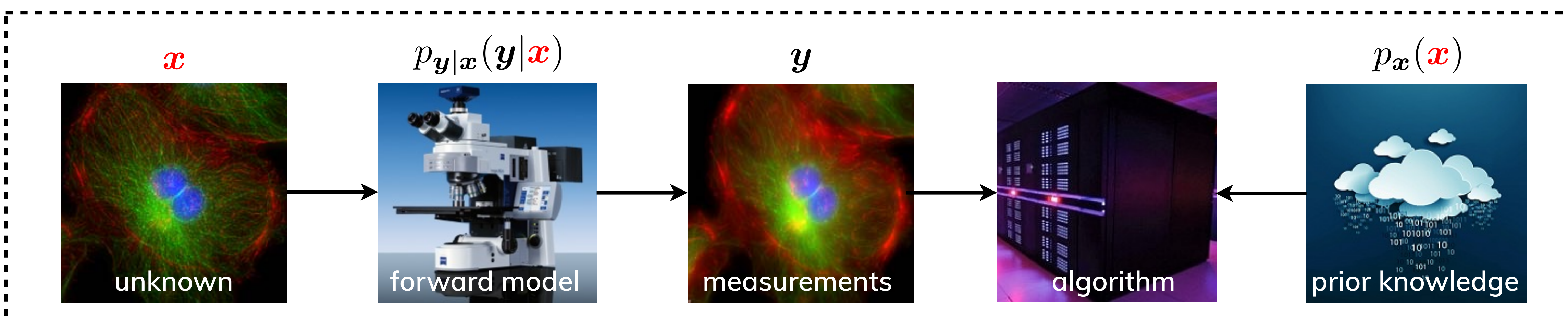
Jiaming Liu, *Student Member, IEEE,* Yu Sun, *Student Member, IEEE,* Cihat Eldeniz, Weijie Gan, Hongyu An, and Ulugbek S. Kamilov, *Member, IEEE*

RARE
arXiv:1912.05854 (2019)

Washington University in St. Louis

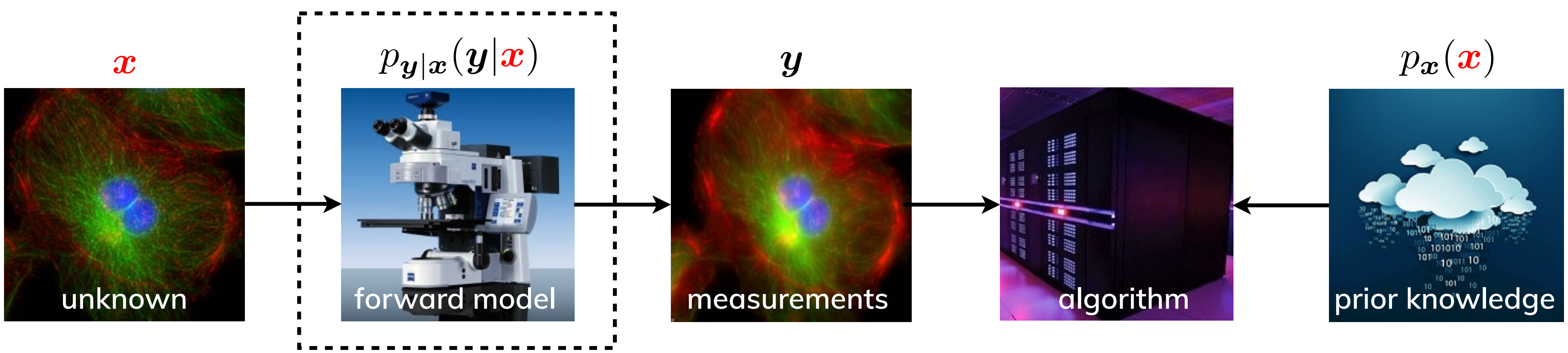
Goal: Overcome these limitations by bringing together physics and learning as information sources

imaging as a very high-dimensional **inference problem**



Goal: Overcome these limitations by bringing together physics and learning as information sources

Forward model:
describes the **physics of data acquisition**

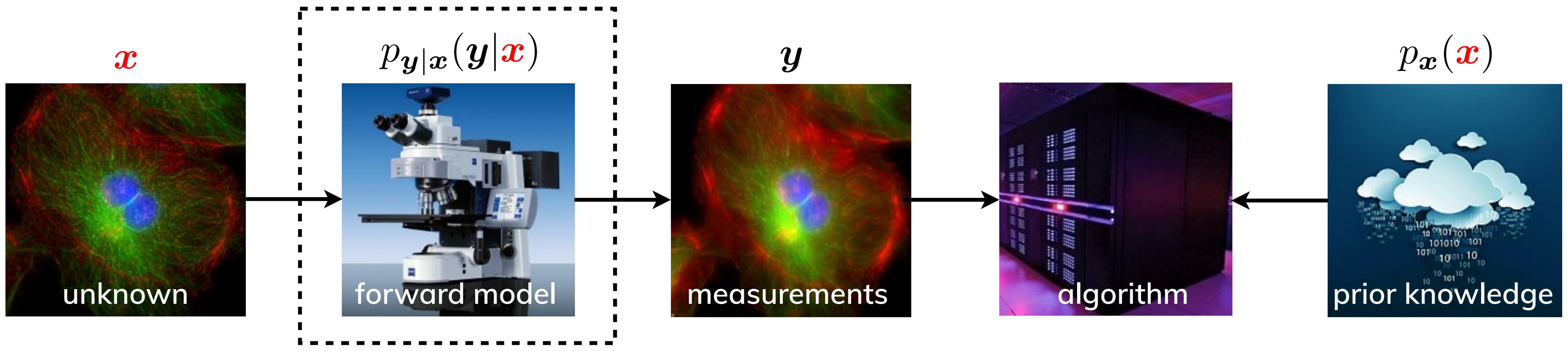


Washington University in St. Louis

Goal: Overcome these limitations by bringing together physics and learning as information sources

Forward model:
describes the physics of data acquisition

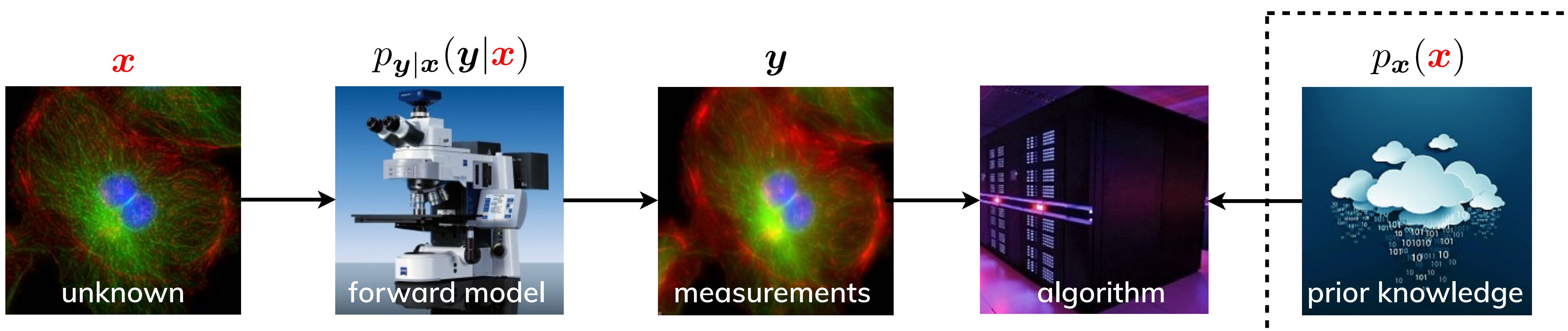
Need **advanced forward models** to:
- **correctly interpret the data** *(physics of data acquisition)*
- **account for known error sources** *(noise and model mismatch)*



Goal: Overcome these limitations by bringing together physics and learning as information sources

Forward model:
describes the physics of data acquisition

Imaging prior:
infuses domain-specific info about the unknown image



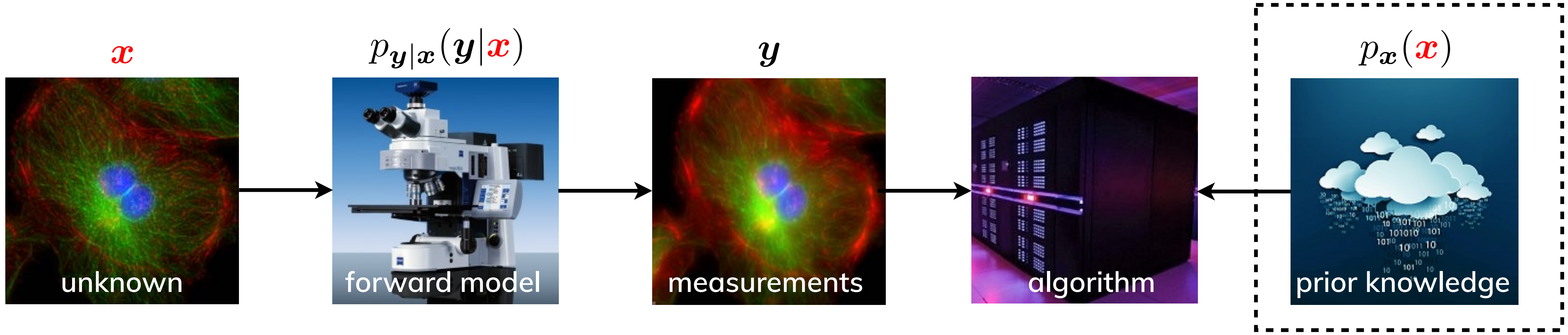
Goal: Overcome these limitations by bringing together physics and learning as information sources

Forward model:
describes the physics of data acquisition

Imaging prior:
infuses domain-specific info about the unknown image

Need **advanced priors** to mitigate:

- long acquisition times *(for in-vivo imaging)*
- incomplete measurements *(missing frequencies, phase loss)*
- noisy measurements *(sensor noise)*

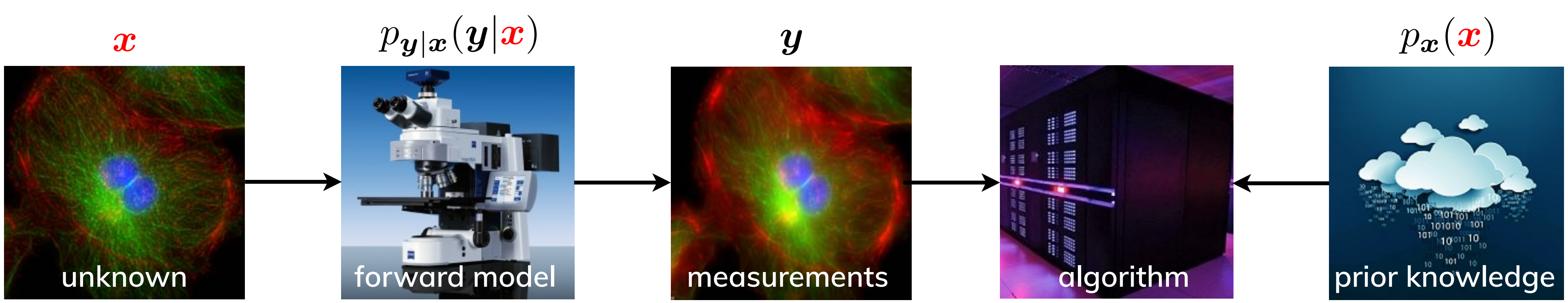


Goal: Overcome these limitations by bringing together physics and learning as information sources

Forward model:
describes the physics of data acquisition

Imaging prior:
infuses domain-specific info about the unknown image

Computational Imaging:
Our goal is to integrate forward models and learned CNN priors to enable fast and accurate biomedical imaging.



Today we will talk about

- Imaging as an inverse problem
  Infusing prior knowledge into image formation
- RARE: Regularization by Artifact Removal
  Using CNN priors learned without ground truth
- SIMBA: Scalable algorithms using CNN priors
  Enabling large-scale tomographic imaging

Today we will talk about

- **Imaging as an inverse problem**
  Infusing prior knowledge into image formation

- RARE: Regularization by Artifact Removal
  Using CNN priors learned without ground truth

- SIMBA: Scalable algorithms using CNN priors
  Enabling large-scale tomographic imaging

Image denoising is a fundamental inverse problem that highlights the importance of prior knowledge

Image denoising is a fundamental inverse problem that highlights the importance of prior knowledge

Additive white Gaussian noise (AWGN) model

$$\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{e}$$

noisy observation =
unknown desired + unknown undesired

Image denoising is a fundamental inverse problem that highlights the importance of prior knowledge

Additive white Gaussian noise (AWGN) model

$$\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{e}$$

Problem:
There are ∞ many possible solutions!

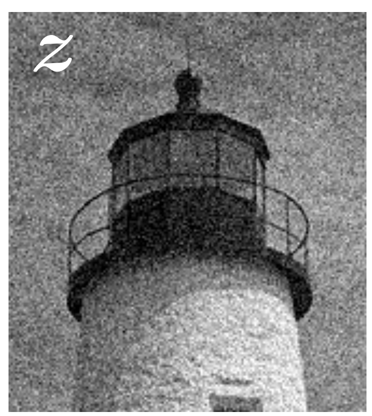
Image denoising is a fundamental inverse problem that highlights the importance of prior knowledge

Additive white Gaussian noise (AWGN) model

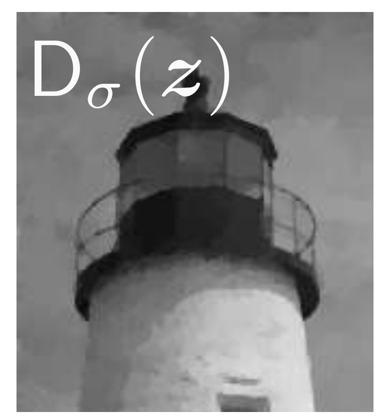
$$\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{e}$$

Image denoiser is a function for separating **signal** from **noise**

$$\mathsf{D}_\sigma : \text{more noisy image} \mapsto \text{less noisy image}$$



$\mapsto$



Source: Pascal Getreuer

Image denoising is a fundamental inverse problem that highlights the importance of prior knowledge

Additive white Gaussian noise (AWGN) model

$$\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{e}$$

Image denoiser is a function for separating signal from noise

$$\mathsf{D}_\sigma : \text{more noisy image} \mapsto \text{less noisy image}$$

Proximal operator is a special type of denoiser

$$\mathsf{prox}_{\tau h}(\boldsymbol{z}) := \mathop{\mathsf{arg\,min}}_{\boldsymbol{x}} \left\{ \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{z}\|_2^2 + \tau h(\boldsymbol{x}) \right\}$$

Source: Pascal Getreuer

Image denoising is a fundamental inverse problem that highlights the importance of prior knowledge

Additive white Gaussian noise (AWGN) model

$$\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{e}$$

Image denoiser is a function for separating signal from noise

$$\mathsf{D}_\sigma : \text{more noisy image} \mapsto \text{less noisy image}$$

Proximal operator is a special type of denoiser

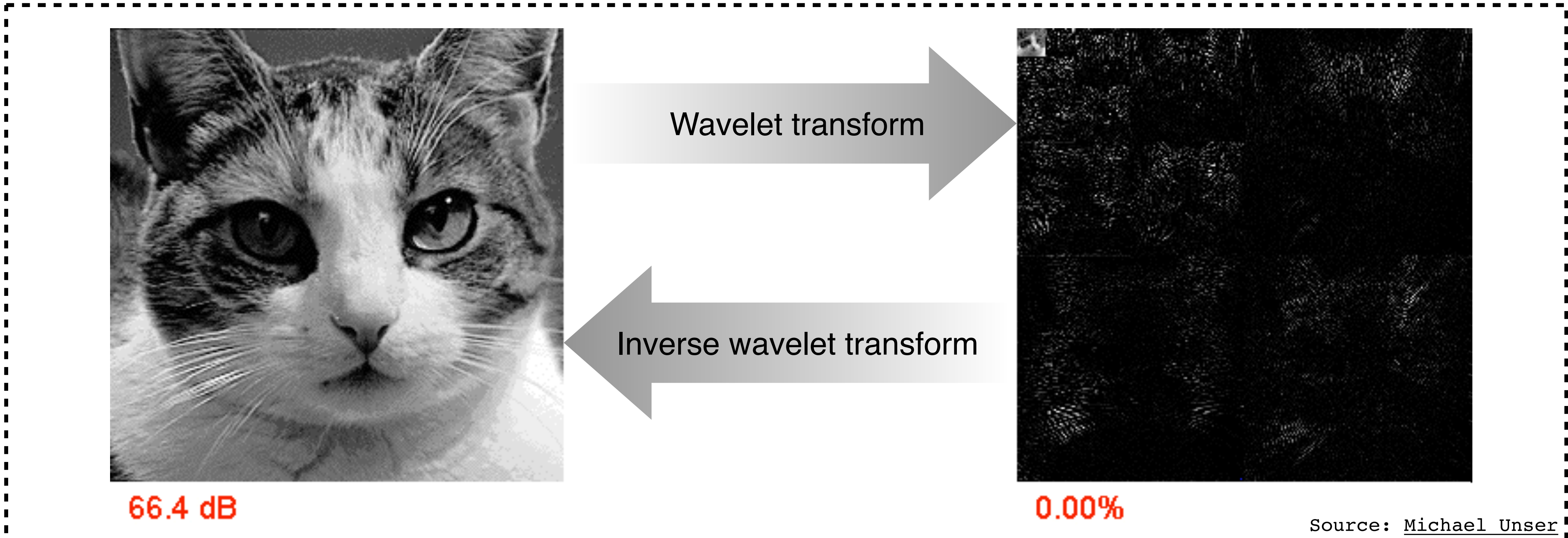
$$\mathsf{prox}_{\tau h}(\boldsymbol{z}) \;:=\; \mathop{\mathsf{arg\,min}}_{\boldsymbol{x}} \left\{ \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{z}\|_2^2 + \tau h(\boldsymbol{x}) \right\}$$

It turns out proximal operators are the bridge from image denoising to other imaging inverse problems!

Example: Wavelet-domain soft-thresholding uses the model that images are wavelet compressible

Example: Wavelet-domain soft-thresholding uses the model that images are wavelet compressible

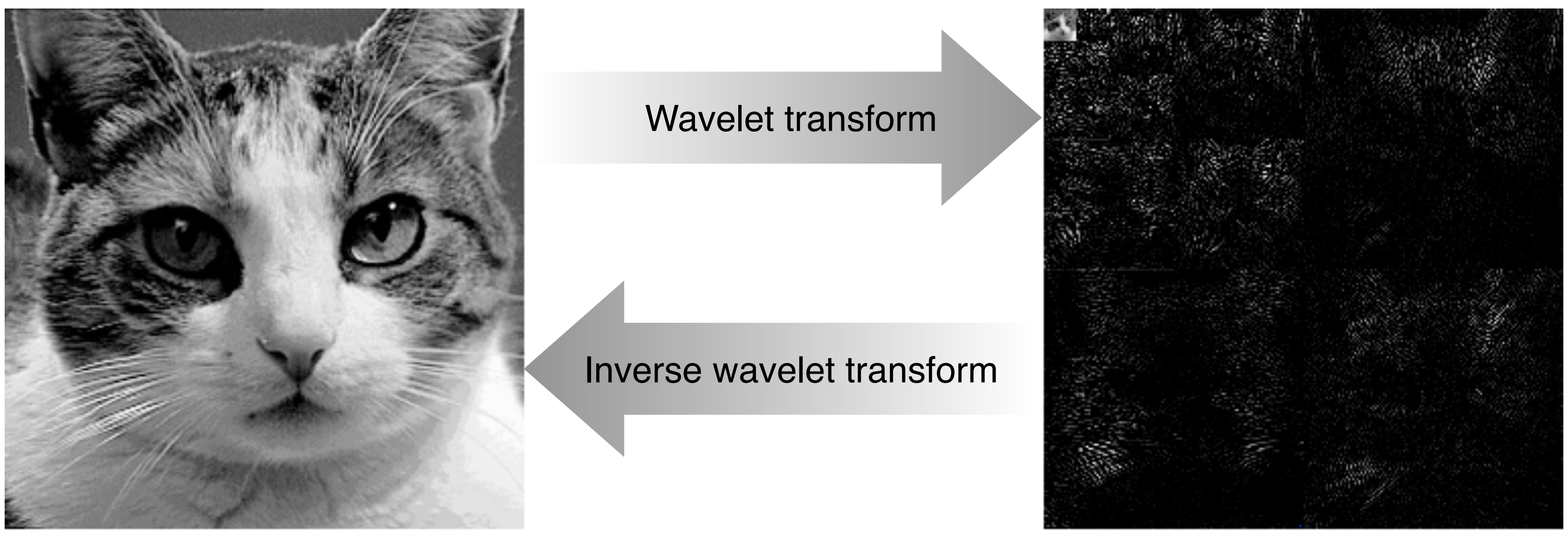
An image (but not noise) is **compressible** in the wavelet domain



Wavelet sparsity is behind the **JPEG-2000** standard

Example: Wavelet-domain soft-thresholding uses the model that images are wavelet compressible

An image (but not noise) is compressible in the wavelet domain



Source: Michael Unser

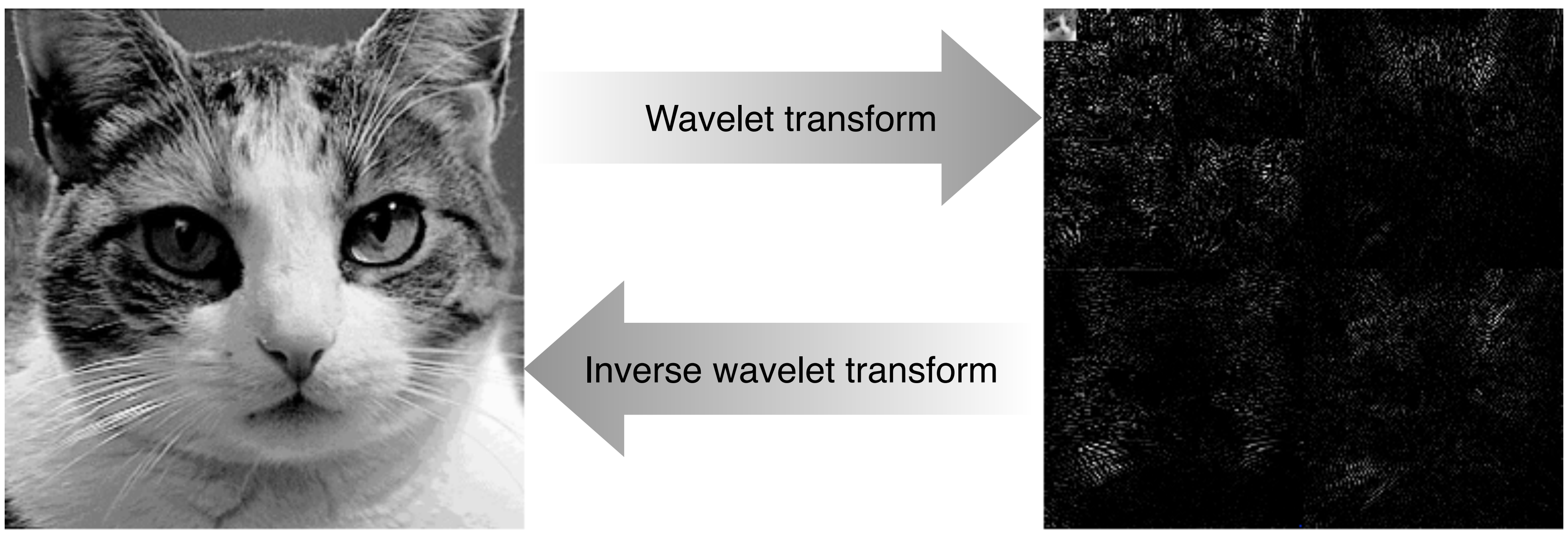
Idea: Denoise an image by finding a wavelet-sparse solution

$$\mathsf{D}_\tau(\boldsymbol{z}) = \underset{\boldsymbol{x}}{\arg\min} \left\{ \frac{1}{2} \|\boldsymbol{x} - \boldsymbol{z}\|_2^2 + \tau \|\mathbf{W}\boldsymbol{x}\|_0 \right\}$$

l_0-norm counts # of non-zeroes

Example: Wavelet-domain soft-thresholding uses the model that images are wavelet compressible

An image (but not noise) is compressible in the wavelet domain



Source: Michael Unser

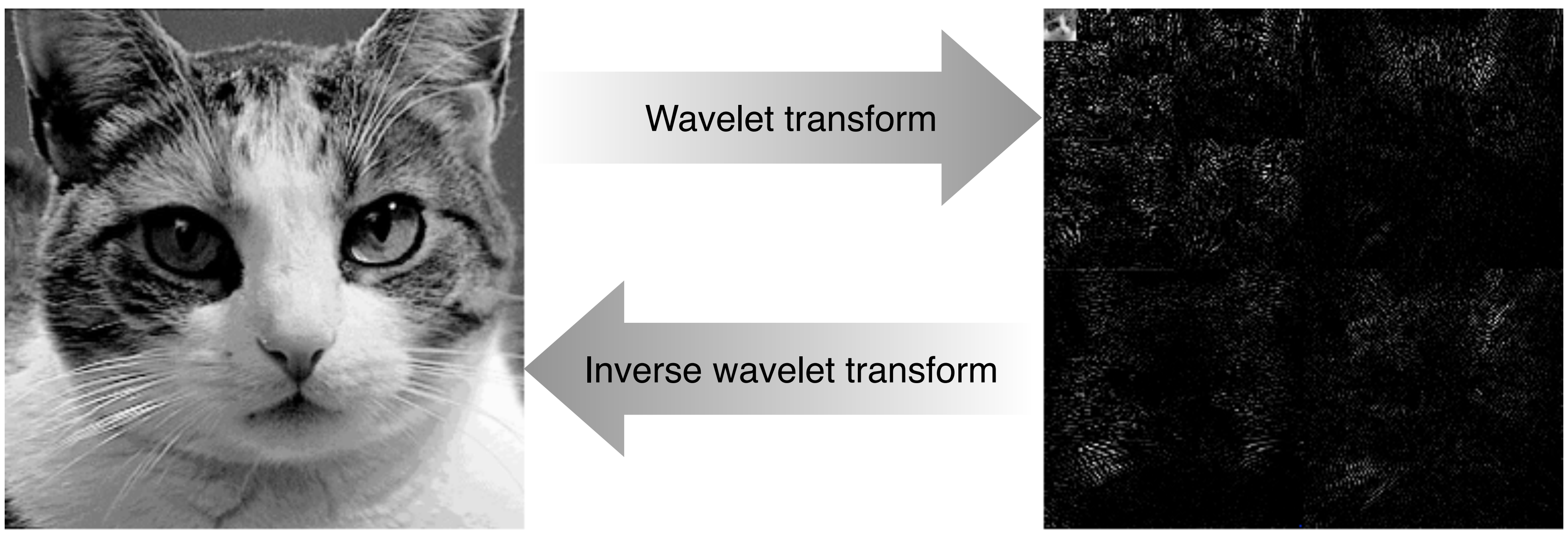
Idea: Denoise an image by finding a wavelet-sparse solution

$$\mathsf{D}_\tau(\boldsymbol{z}) = \underset{\boldsymbol{x}}{\arg\min} \left\{ \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{z}\|_2^2 + \tau \|\mathbf{W}\boldsymbol{x}\|_0 \right\}$$

Problem: **Nonconvex** and **nonsmooth**

Example: Wavelet-domain soft-thresholding uses the model that images are wavelet compressible

An image (but not noise) is compressible in the wavelet domain



Source: Michael Unser

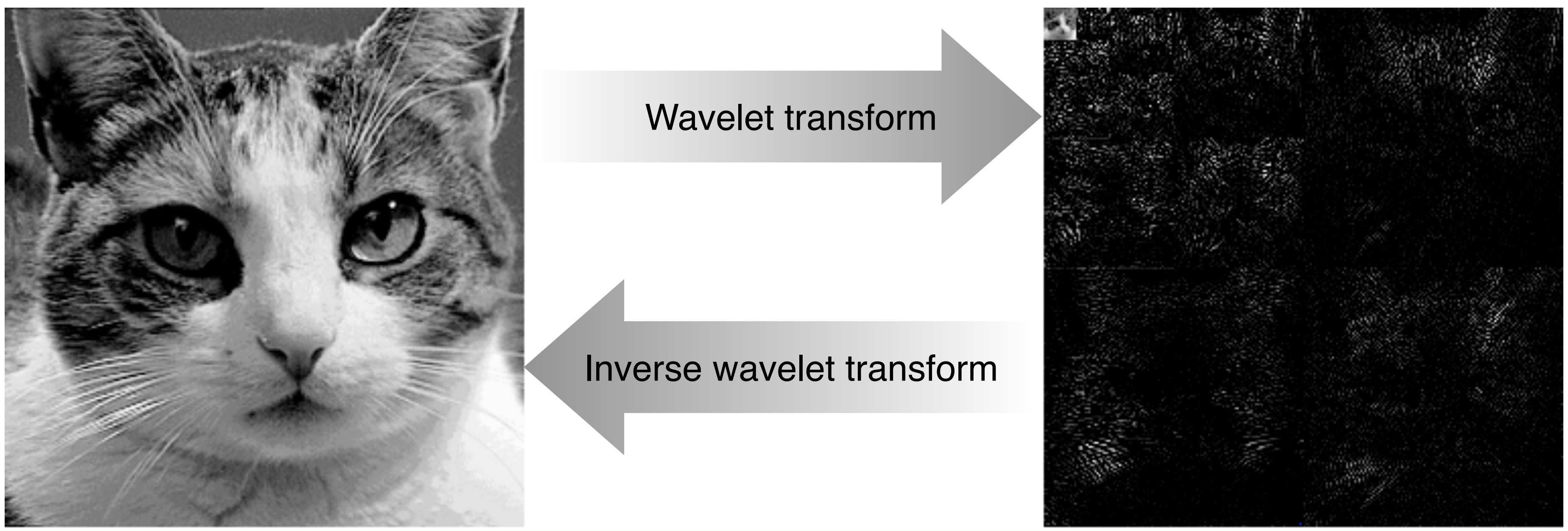
Idea: Denoise an image by finding a wavelet-sparse solution

$$\mathsf{D}_\tau(\boldsymbol{z}) = \mathop{\mathsf{arg\,min}}_{\boldsymbol{x}} \left\{ \frac{1}{2} \|\boldsymbol{x} - \boldsymbol{z}\|_2^2 + \tau \|\mathbf{W}\boldsymbol{x}\|_1 \right\}$$

Convex and **nonsmooth**, but still **sparsity promoting**

Example: Wavelet-domain soft-thresholding uses the model that images are wavelet compressible

An image (but not noise) is compressible in the wavelet domain

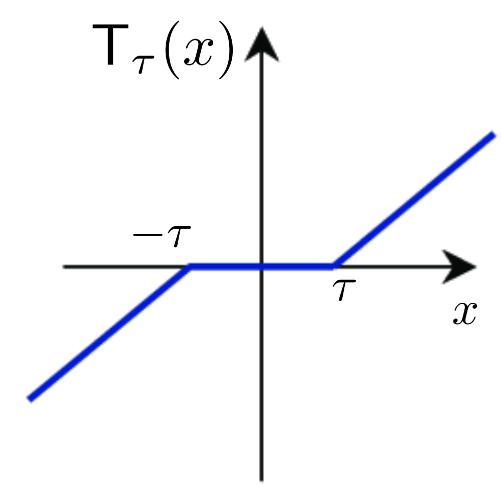


Source: Michael Unser

Idea: Denoise an image by finding a wavelet-sparse solution

$$\mathsf{D}_\tau(\boldsymbol{z}) = \underset{\boldsymbol{x}}{\arg\min} \left\{ \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{z}\|_2^2 + \tau \|\mathbf{W}\boldsymbol{x}\|_1 \right\}$$
$$= \mathbf{W}^\mathsf{T} \mathsf{T}_\tau(\mathbf{W}\boldsymbol{z})$$

Closed-form solution: **Wavelet thresholding**!



Example: Wavelet-domain soft-thresholding uses the model that images are wavelet compressible

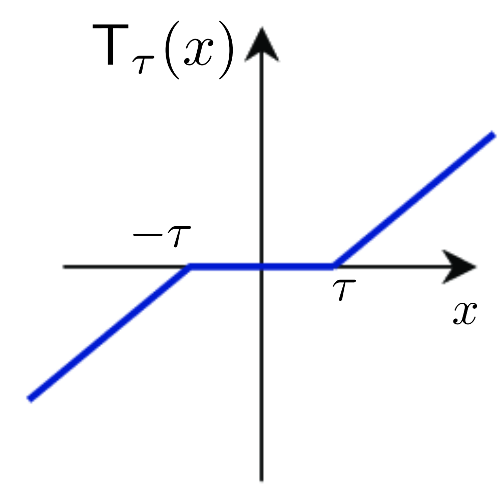
An image (but not noise) is compressible in the wavelet domain

Source: Pascal Getreuer

Intuition: Small wavelet coefficients correspond to noise!

Idea: Denoise an image by finding a wavelet-sparse solution

$$\mathsf{D}_\tau(\boldsymbol{z}) = \arg\min_{\boldsymbol{x}} \left\{ \frac{1}{2} \|\boldsymbol{x} - \boldsymbol{z}\|_2^2 + \tau \|\mathbf{W}\boldsymbol{x}\|_1 \right\}$$
$$= \mathbf{W}^\mathsf{T} \mathsf{T}_\tau(\mathbf{W}\boldsymbol{z})$$



Closed-form solution: **Wavelet thresholding**!

Regularized inversion provides a unified approach for integrating physics and prior knowledge

Regularized inversion provides a unified approach for integrating physics and prior knowledge

Image recovery is often an **ill-posed inverse problem**

$$\boldsymbol{y} = \mathbf{H}(\boldsymbol{x}) + \boldsymbol{e}$$

noise, but also **subsampling**, **physics**, **model uncertainties**, etc.

Regularized inversion provides a unified approach for integrating physics and prior knowledge

Image recovery is often an ill-posed inverse problem

$$\boldsymbol{y} = \mathbf{H}(\boldsymbol{x}) + \boldsymbol{e}$$

Formulation as a **regularized optimization** problem

$$\widehat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\mathsf{arg\,min}} \{f(\boldsymbol{x})\}$$

$$f(\boldsymbol{x}) := g(\boldsymbol{x}) + h(\boldsymbol{x})$$

data-fidelity + prior

Regularized inversion provides a unified approach for integrating physics and prior knowledge

Image recovery is often an ill-posed inverse problem

$$\boldsymbol{y} = \mathbf{H}(\boldsymbol{x}) + \boldsymbol{e}$$

Formulation as a regularized optimization problem

$$\widehat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\text{arg min}} \{f(\boldsymbol{x})\} \qquad f(\boldsymbol{x}) := g(\boldsymbol{x}) + h(\boldsymbol{x})$$

Example: **Linear inverse problems** (20th century theory)

$$\widehat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\text{arg min}} \left\{ \frac{1}{2}\|\boldsymbol{y} - \mathbf{H}\boldsymbol{x}\|_2^2 + \frac{\lambda}{2}\|\mathbf{D}\boldsymbol{x}\|_2^2 \right\}$$
$$= (\mathbf{H}^{\mathsf{H}}\mathbf{H} + \lambda\mathbf{D}^{\mathsf{H}}\mathbf{D})^{-1}\mathbf{H}^{\mathsf{H}}\boldsymbol{y} = \mathbf{R}_\lambda \cdot \boldsymbol{y}$$

Regularized inversion provides a unified approach for integrating physics and prior knowledge

Image recovery is often an ill-posed inverse problem

$$\boldsymbol{y} = \mathbf{H}(\boldsymbol{x}) + \boldsymbol{e}$$

Formulation as a regularized optimization problem

$$\widehat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\arg\min} \{f(\boldsymbol{x})\} \qquad f(\boldsymbol{x}) := g(\boldsymbol{x}) + h(\boldsymbol{x})$$

Example: **Linear inverse problems** (20th century theory)

$$\widehat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\arg\min} \left\{ \frac{1}{2}\|\boldsymbol{y} - \mathbf{H}\boldsymbol{x}\|_2^2 + \frac{\lambda}{2}\|\mathbf{D}\boldsymbol{x}\|_2^2 \right\}$$
$$= (\mathbf{H}^{\mathsf{H}}\mathbf{H} + \lambda\mathbf{D}^{\mathsf{H}}\mathbf{D})^{-1}\mathbf{H}^{\mathsf{H}}\boldsymbol{y} = \mathbf{R}_\lambda \cdot \boldsymbol{y}$$

Classical
"filtered" backprojection

Regularized inversion provides a unified approach for integrating physics and prior knowledge

Image recovery is often an ill-posed inverse problem

$$\boldsymbol{y} = \mathbf{H}(\boldsymbol{x}) + \boldsymbol{e}$$

Formulation as a regularized optimization problem

$$\widehat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\text{arg min}} \{f(\boldsymbol{x})\} \qquad f(\boldsymbol{x}) := g(\boldsymbol{x}) + h(\boldsymbol{x})$$

Example: **Linear inverse problems** (20th century theory)

$$\widehat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\text{arg min}} \left\{\frac{1}{2}\|\boldsymbol{y} - \mathbf{H}\boldsymbol{x}\|_2^2 + \frac{\lambda}{2}\|\mathbf{D}\boldsymbol{x}\|_2^2\right\}$$
$$= (\mathbf{H}^{\mathsf{H}}\mathbf{H} + \lambda\mathbf{D}^{\mathsf{H}}\mathbf{D})^{-1}\mathbf{H}^{\mathsf{H}}\boldsymbol{y} = \mathbf{R}_\lambda \cdot \boldsymbol{y}$$

Classical
“filtered” backprojection

Assumption:
Gaussians everywhere

Regularized inversion provides a unified approach for integrating physics and prior knowledge

Image recovery is often an ill-posed inverse problem

$$\boldsymbol{y} = \mathbf{H}(\boldsymbol{x}) + \boldsymbol{e}$$

Formulation as a regularized optimization problem

$$\widehat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\arg\min} \{f(\boldsymbol{x})\} \qquad f(\boldsymbol{x}) := g(\boldsymbol{x}) + h(\boldsymbol{x})$$

Example: **Maximum a posteriori probability (MAP)** estimator

$$g(\boldsymbol{x}) = -\log(p_{\boldsymbol{y}|\boldsymbol{x}}(\boldsymbol{y}|\boldsymbol{x})) \qquad h(\boldsymbol{x}) = -\log(p_{\boldsymbol{x}}(\boldsymbol{x}))$$

general likelihood term — general prior term

Regularized inversion provides a unified approach for integrating physics and prior knowledge

Image recovery is often an ill-posed inverse problem

$$\boldsymbol{y} = \mathbf{H}(\boldsymbol{x}) + \boldsymbol{e}$$

Formulation as a regularized optimization problem

$$\widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \{f(\boldsymbol{x})\} \qquad f(\boldsymbol{x}) := g(\boldsymbol{x}) + h(\boldsymbol{x})$$

Example: **AWGN** and **sparsity-promoting** prior

$$g(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{y} - \mathbf{H}(\boldsymbol{x})\|_2^2$$

smooth data-fidelity
(least-squares)

$$h(\boldsymbol{x}) = \lambda\|\mathbf{D}\boldsymbol{x}\|_1$$

nonsmooth regularizer
(sparsity promoting prior)

FISTA and ADMM use proximal operators for solving large-scale and nonsmooth problems

FISTA and ADMM use proximal operators for solving large-scale and nonsmooth problems

Goal: Minimize the following

$$f(\boldsymbol{x}) = g(\boldsymbol{x}) + h(\boldsymbol{x})$$

FISTA and ADMM use proximal operators for solving large-scale and nonsmooth problems

Fast iterative shrinkage/thresholding algorithm (FISTA) vs. alternating direction method of multipliers (ADMM)

$$\begin{aligned}
\boldsymbol{z}^k &\leftarrow \boldsymbol{s}^{k-1} - \gamma \nabla g(\boldsymbol{s}^{k-1}) \\
\boldsymbol{x}^k &\leftarrow \mathsf{prox}_{\gamma h}(\boldsymbol{z}^k) \\
\boldsymbol{s}^k &\leftarrow \boldsymbol{x}^k + ((q_{k-1} - 1)/q_k)(\boldsymbol{x}^k - \boldsymbol{x}^{k-1})
\end{aligned}$$

FISTA: grad data + prox prior

$$\begin{aligned}
\boldsymbol{z}^k &\leftarrow \mathsf{prox}_{\gamma g}(\boldsymbol{x}^{k-1} - \boldsymbol{s}^{k-1}) \\
\boldsymbol{x}^k &\leftarrow \mathsf{prox}_{\gamma h}(\boldsymbol{z}^k + \boldsymbol{s}^{k-1}) \\
\boldsymbol{s}^k &\leftarrow \boldsymbol{s}^{k-1} + (\boldsymbol{z}^k - \boldsymbol{x}^k)
\end{aligned}$$

ADMM: prox data + prox prior

$$f(\boldsymbol{x}) = g(\boldsymbol{x}) + h(\boldsymbol{x})$$

FISTA and ADMM use proximal operators for solving large-scale and nonsmooth problems

Fast iterative shrinkage/thresholding algorithm (FISTA) vs. alternating direction method of multipliers (ADMM)

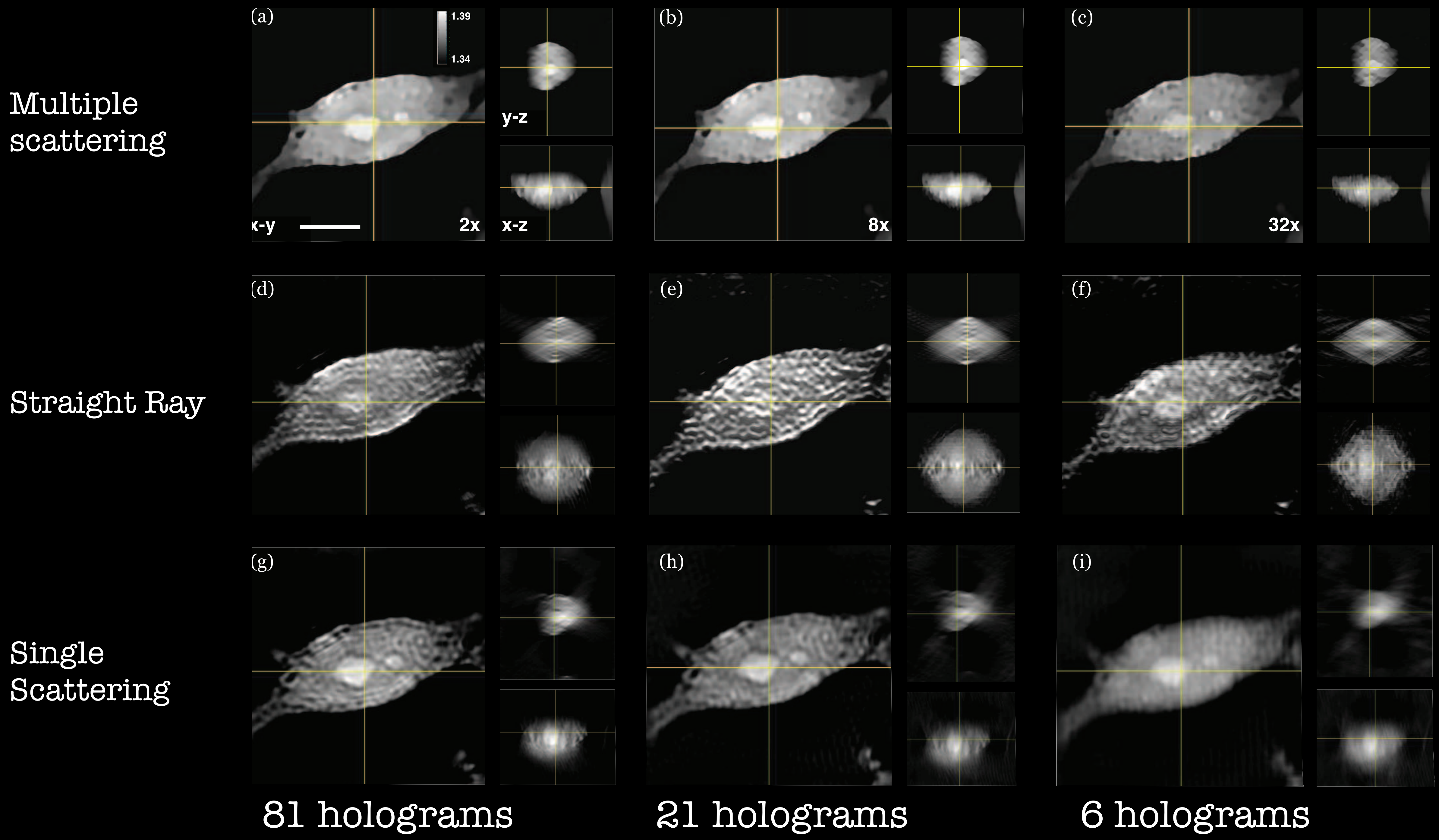
FISTA		ADMM
$\boldsymbol{z}^k \leftarrow \boldsymbol{s}^{k-1} - \gamma \nabla g(\boldsymbol{s}^{k-1})$	more data consistent	$\boldsymbol{z}^k \leftarrow \text{prox}_{\gamma g}(\boldsymbol{x}^{k-1} - \boldsymbol{s}^{k-1})$
$\boldsymbol{x}^k \leftarrow \text{prox}_{\gamma h}(\boldsymbol{z}^k)$	less noisy	$\boldsymbol{x}^k \leftarrow \text{prox}_{\gamma h}(\boldsymbol{z}^k + \boldsymbol{s}^{k-1})$
$\boldsymbol{s}^k \leftarrow \boldsymbol{x}^k + ((q_{k-1} - 1)/q_k)(\boldsymbol{x}^k - \boldsymbol{x}^{k-1})$		$\boldsymbol{s}^k \leftarrow \boldsymbol{s}^{k-1} + (\boldsymbol{z}^k - \boldsymbol{x}^k)$

Both FISTA and ADMM alternate between increasing data consistency and reducing noise

- $\mathsf{I} - \gamma \nabla g$: less data consistent $\mapsto$ more data consistent
- $\text{prox}_{\gamma g}$: less data consistent $\mapsto$ more data consistent
- $\text{prox}_{\gamma h}$: more noisy image $\mapsto$ less noisy image

$$f(\boldsymbol{x}) = g(\boldsymbol{x}) + h(\boldsymbol{x})$$

Application: Reduce the scanning time in optical diffraction tomography



Kamilov *et al.*, “Optical Tomographic Image Reconstruction Based on Beam Propagation and Sparse Regularization,” *IEEE Trans. Comp. Imag.*, 2016.

To summarize our discussion so far

- Imaging quality in ill-posed inverse problems can be significantly improved by using prior knowledge

- Sparsity-promoting priors are highly effective, but typically result in nonsmooth optimization problems

- FISTA and ADMM are the two most popular algorithms used for nonsmooth regularized inversion

- We will seek to achieve improvements by using priors specified by **deep neural nets for image restoration**

Today we will talk about

- Imaging as an inverse problem
  Infusing prior knowledge into image formation
- RARE: Regularization by Artifact Removal
  Using CNN priors learned without ground truth
- SIMBA: Scalable algorithms using CNN priors
  Enabling large-scale tomographic imaging

Is there a better way to represent images than using fixed sparsity-promoting transforms?

Is there a better way to represent images than using fixed sparsity-promoting transforms?

The interest in sparsity-driven imaging highlighted the importance of **structural priors for image formation**

Is there a better way to represent images than using fixed sparsity-promoting transforms?

The interest in sparsity-driven imaging highlighted the importance of structural priors for image formation

Do we know a more flexible, sophisticated, and data-adaptive tool for **characterizing imaging priors**?

Is there a better way to represent images than using fixed sparsity-promoting transforms?

The interest in sparsity-driven imaging highlighted the importance of structural priors for image formation

Do we know a more flexible, sophisticated, and data-adaptive tool for characterizing imaging priors?

Yes, **deep neural nets** provide a state-of-the-art tool for representing and enforcing sophisticated structural information

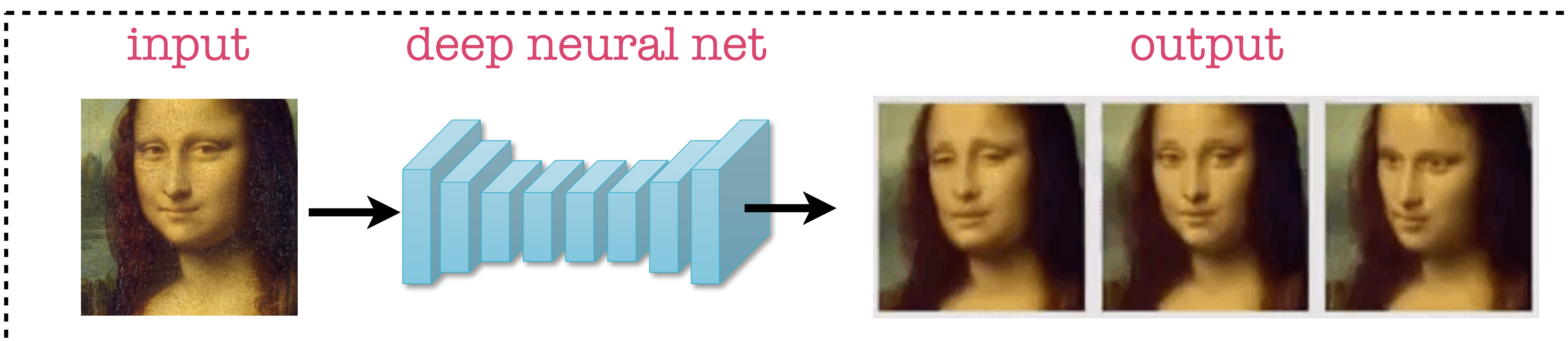
Washington University in St. Louis

Is there a better way to represent images than using fixed sparsity-promoting transforms?

The interest in sparsity-driven imaging highlighted the importance of structural priors for image formation

Do we know a more flexible, sophisticated, and data-adaptive tool for characterizing imaging priors?

Yes, **deep neural nets** provide a state-of-the-art tool for representing and enforcing sophisticated structural information



Zakharov *et al.*, "Few-Shot Adversarial Learning of Realistic Neural Talking Head Models," *Proc. ICCV*, October 2019

Is there a better way to represent images than using fixed sparsity-promoting transforms?

The interest in sparsity-driven imaging highlighted the importance of structural priors for image formation

Do we know a more flexible, sophisticated, and data-adaptive tool for characterizing imaging priors?

Yes, **deep neural nets** provide a state-of-the-art tool for representing and enforcing sophisticated structural information

- Wow, this result is excellent. Why?
- This is the power of deep neural net…

(ref: Charlie Bouman)

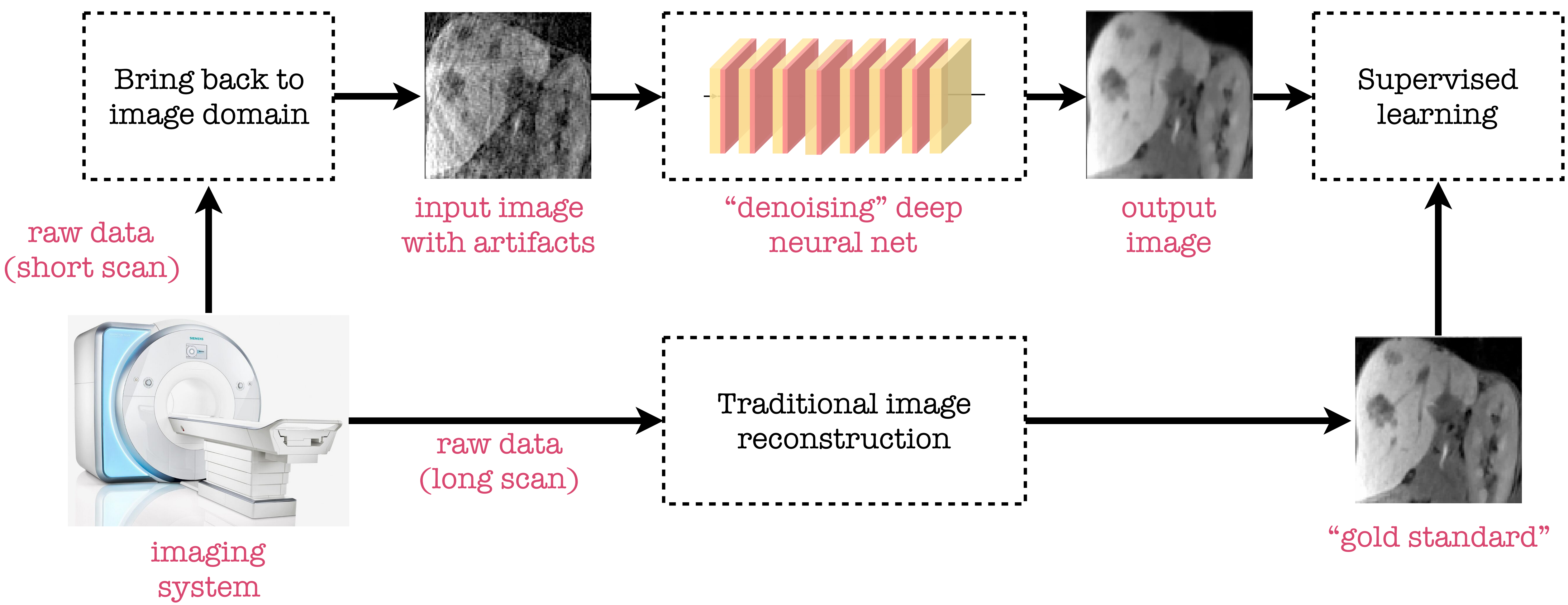
Simple recipe: Supervised learning from the measured data to the desired image

Jin *et al.*, "Deep Convolutional Neural Network for Inverse Problems in Imaging," *IEEE TIP*, 2017

Washington University in St. Louis

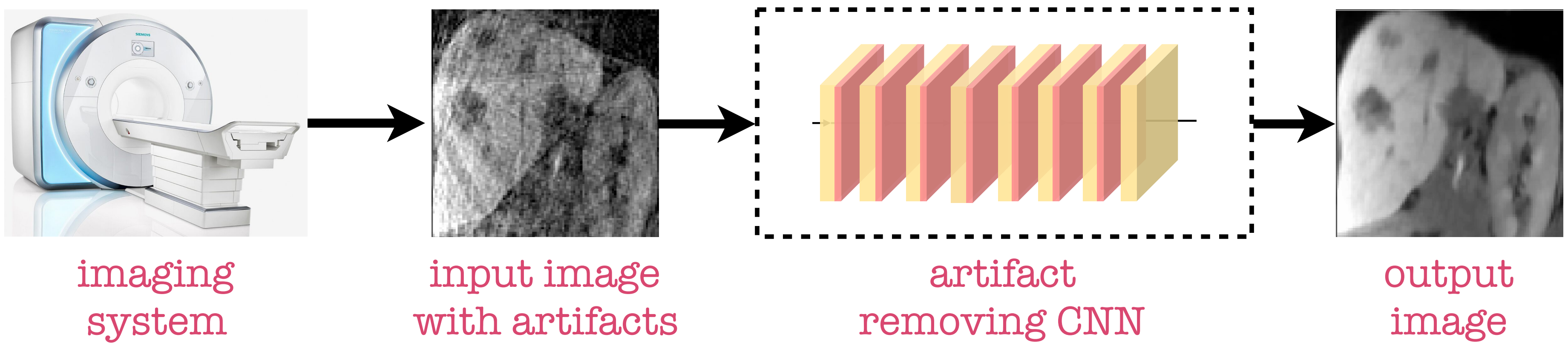
Simple recipe: Supervised learning from the measured data to the desired image

Example: Train a deep neural net to remove artifacts from an image



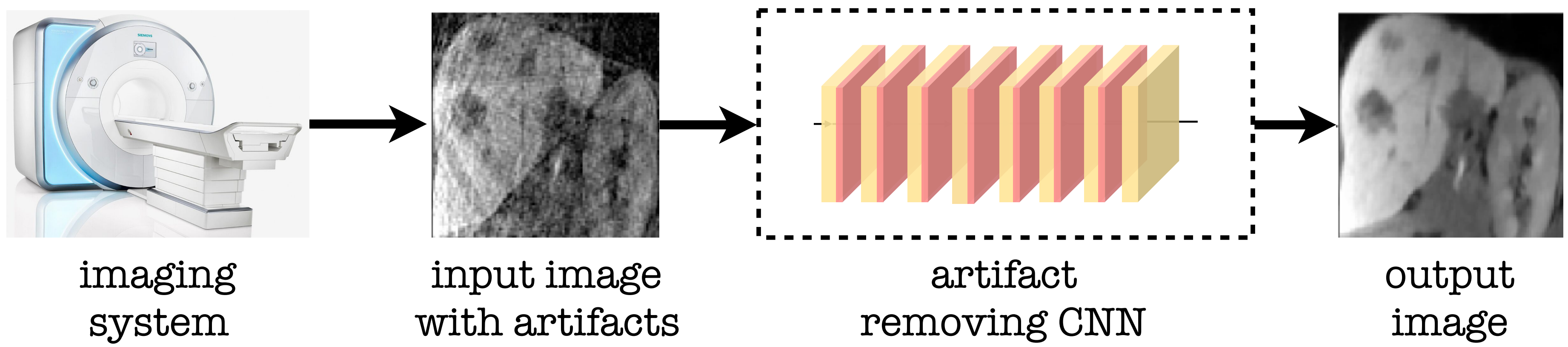
Simple recipe: Supervised learning from the measured data to the desired image

Example: Train a deep neural net to remove artifacts from an image



Simple recipe: Supervised learning from the measured data to the desired image

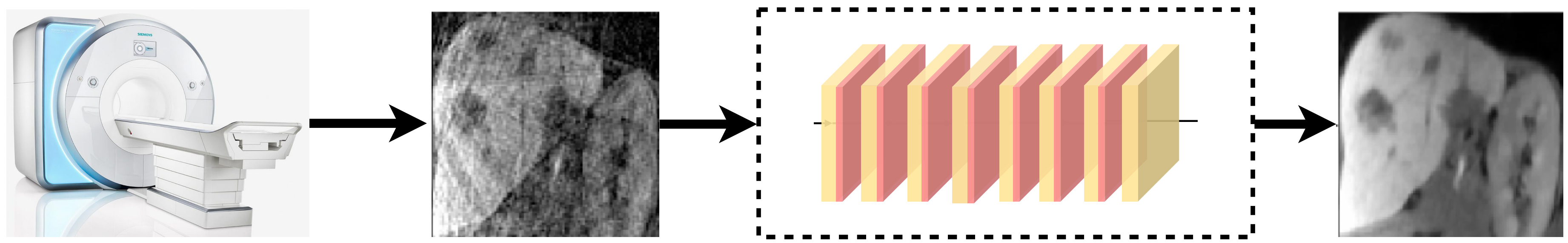
Example: Train a deep neural net to remove artifacts from an image



Question: What are some of the key benefits of this approach?

Simple recipe: Supervised learning from the measured data to the desired image

Example: Train a deep neural net to remove artifacts from an image



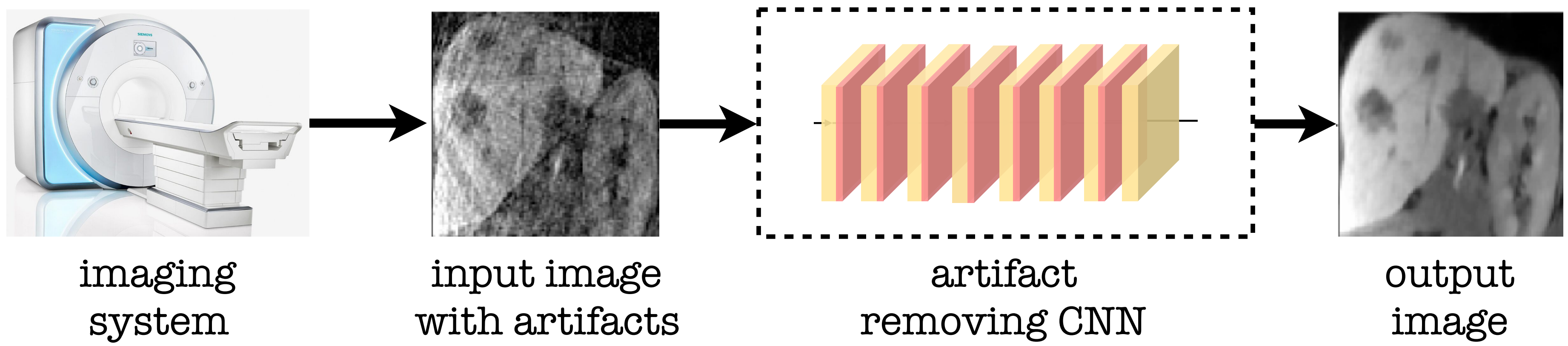
Question: What are some of the key benefits of this approach?

1) Very easy to implement and deploy

Use existing frameworks and architectures

Simple recipe: Supervised learning from the measured data to the desired image

Example: Train a deep neural net to remove artifacts from an image



Question: What are some of the key benefits of this approach?

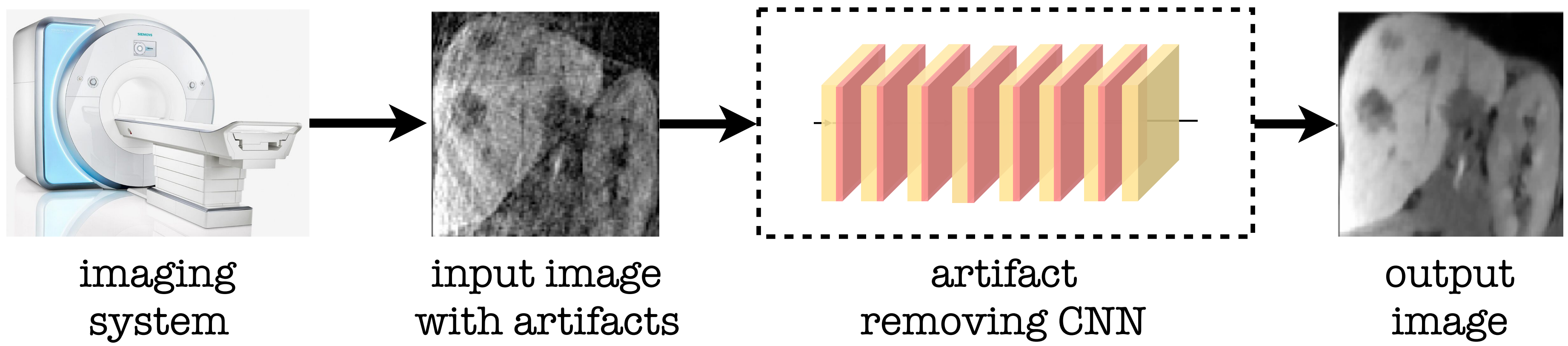
1) Very easy to implement and deploy

2) Extremely fast at test time (perfect for volumetric imaging!)

Takes seconds, while optimization takes hours

Simple recipe: Supervised learning from the measured data to the desired image

Example: Train a deep neural net to remove artifacts from an image



Question: What are some of the key benefits of this approach?

1) Very easy to implement and deploy

2) Extremely fast at test time (perfect for volumetric imaging!)

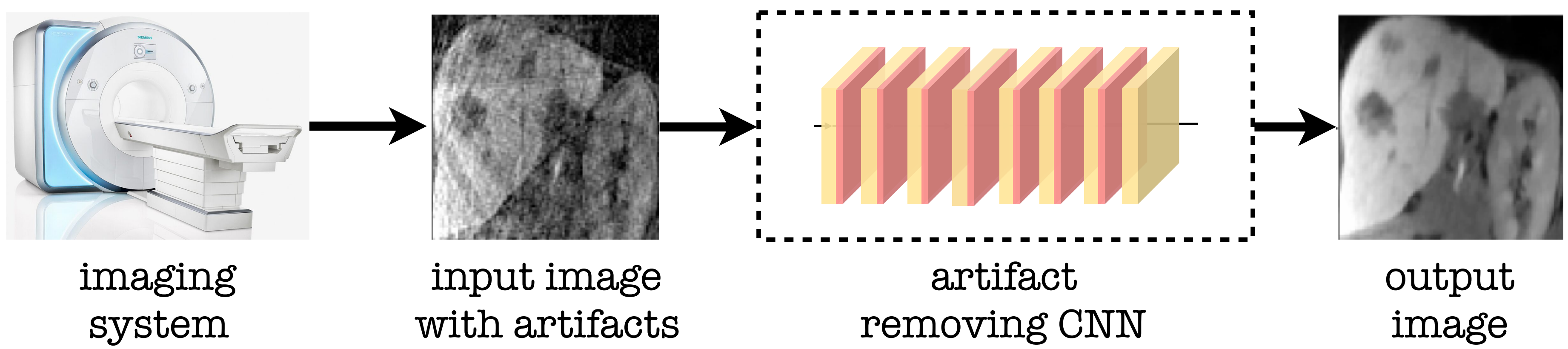
3) No need to explicitly models anything, just learn end-to-end

Everything is learned automatically from data

Washington University in St. Louis

Simple recipe: Supervised learning from the measured data to the desired image

Example: Train a deep neural net to remove artifacts from an image



Question: What are some of the key benefits of this approach?

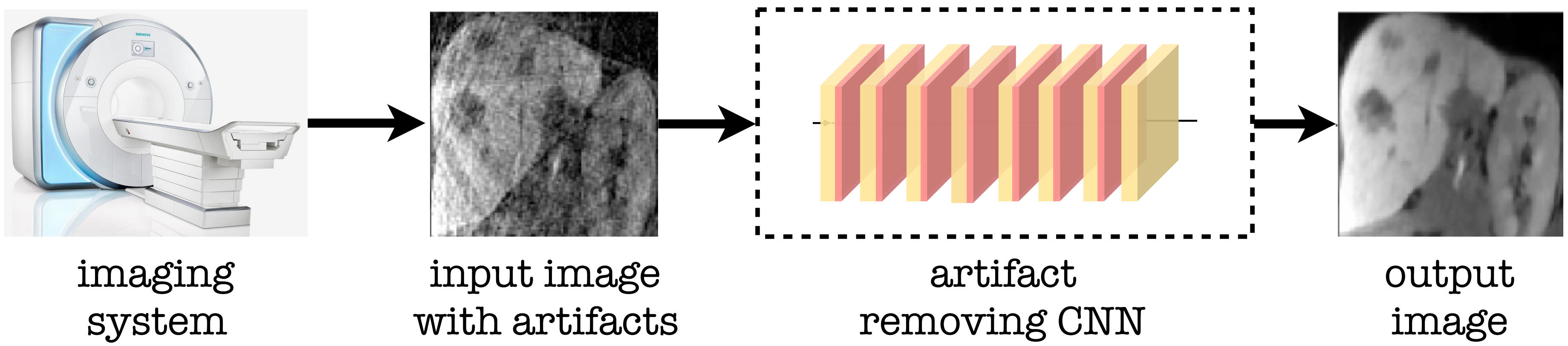
1) Very easy to implement and deploy

2) Extremely fast at test time (perfect for volumetric imaging!)

3) No need to explicitly models anything, just learn end-to-end

Simple recipe: Supervised learning from the measured data to the desired image

Example: Train a deep neural net to remove artifacts from an image

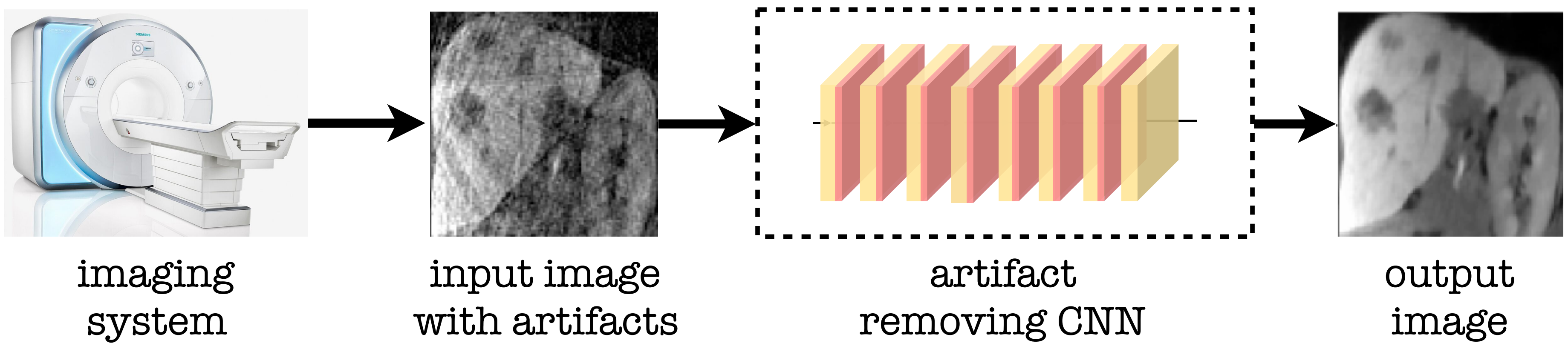


Question: What are some of the key limitations of this approach?

Washington University in St. Louis

Simple recipe: Supervised learning from the measured data to the desired image

Example: Train a deep neural net to remove artifacts from an image



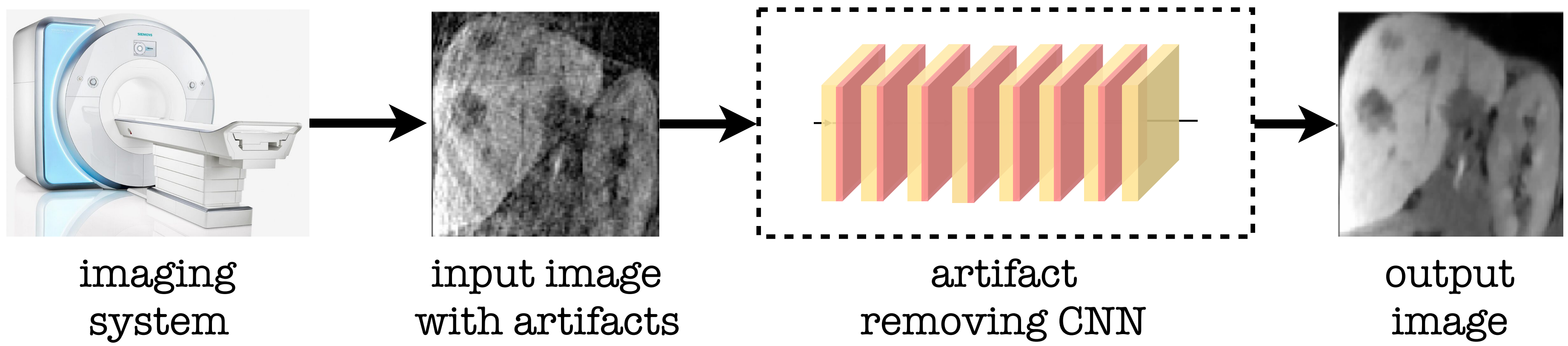
Question: What are some of the key limitations of this approach?

1) Need **ground truth** for training the network

Impossible to have for some applications

Simple recipe: Supervised learning from the measured data to the desired image

Example: Train a deep neural net to remove artifacts from an image



Question: What are some of the key limitations of this approach?

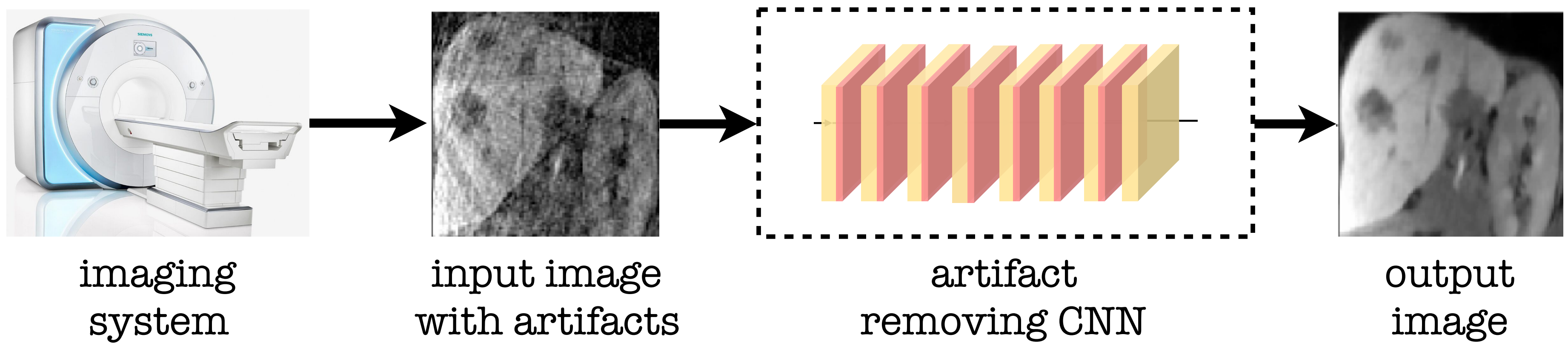
1) Need ground truth for training the network

2) **Consistency** to the measured data is not explicit

No explicit measure of deviation from the raw data

Simple recipe: Supervised learning from the measured data to the desired image

Example: Train a deep neural net to remove artifacts from an image



Question: What are some of the key limitations of this approach?

1) Need ground truth for training the network

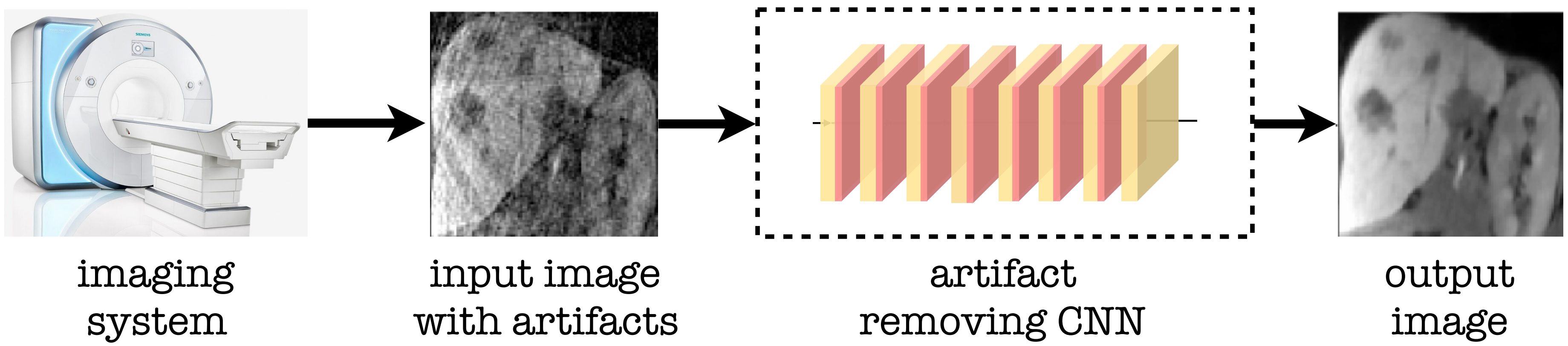
2) Consistency to the measured data is not explicit

3) Does not exploit known physical models and information

Why re-learn something we know? Also, limits generalization

Simple recipe: Supervised learning from the measured data to the desired image

Example: Train a deep neural net to remove artifacts from an image



Question: What are some of the key limitations of this approach?

1) Need ground truth for training the network

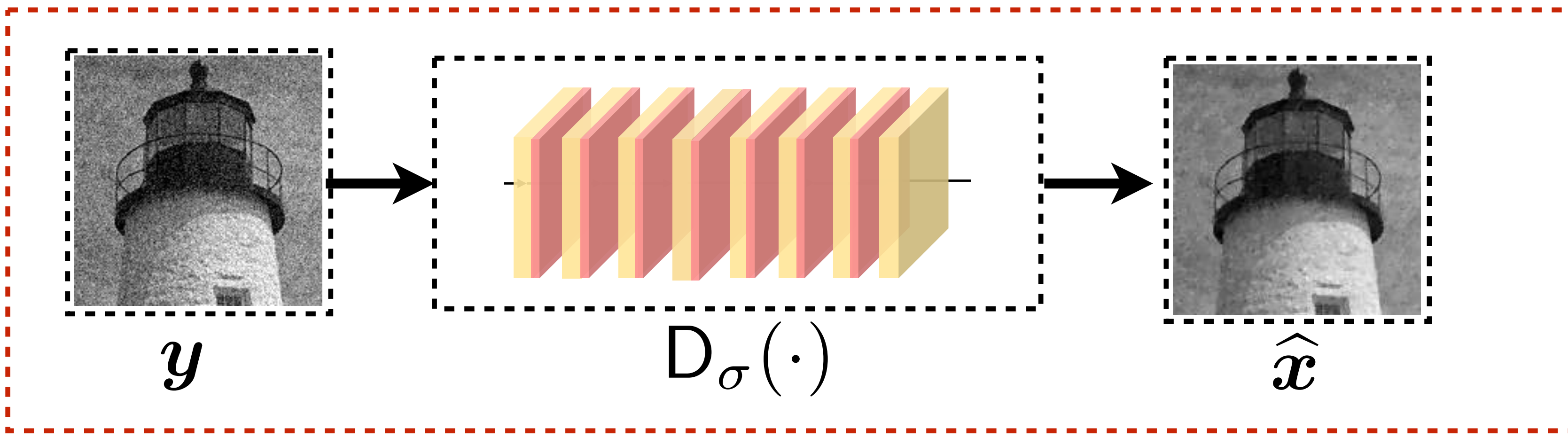
2) Consistency to the measured data is not explicit

3) Does not exploit known physical models and information

Plug-and-play priors (PnP) methods separate the forward model from the learned prior

Plug-and-play priors (PnP) methods separate the forward model from the learned prior

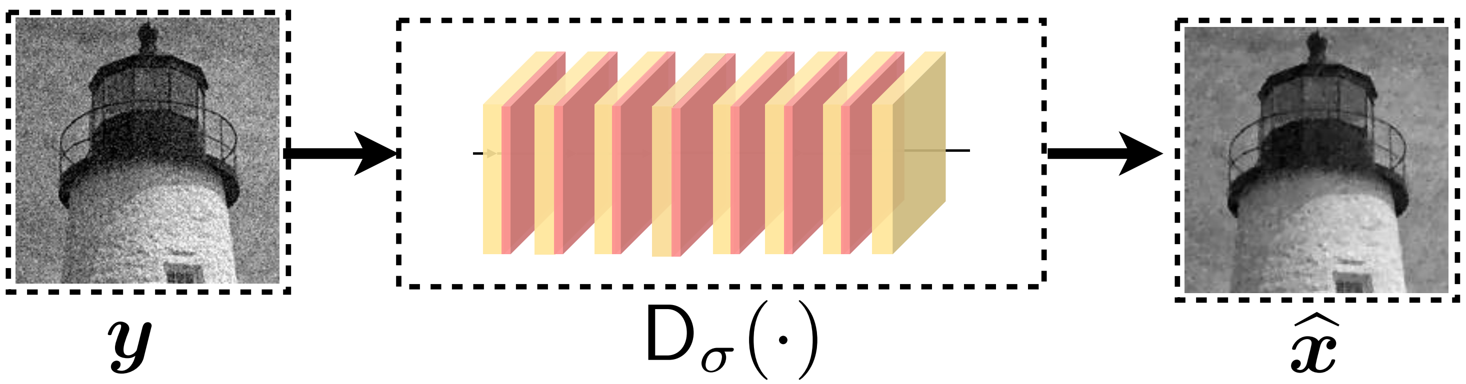
Idea: Treat a **denoising CNN** as a proximal operator



$\approx \mathsf{prox}_{\gamma h}(\boldsymbol{y})$

Plug-and-play priors (PnP) methods separate the forward model from the learned prior

Idea: Treat a denoising CNN as a proximal operator

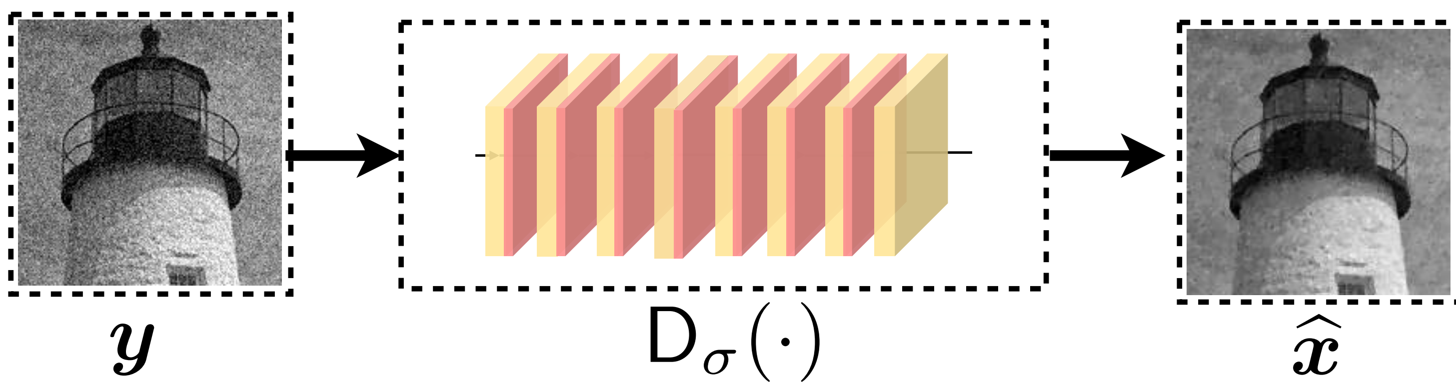


We define the following **"plug-and-play"** operators

- D_σ : more noisy image $\mapsto$ less noisy image
- $\mathsf{I} - \gamma \nabla g$: less data consistent $\mapsto$ more data consistent
- $\mathsf{prox}_{\gamma g}$: less data consistent $\mapsto$ more data consistent

Plug-and-play priors (PnP) methods separate the forward model from the learned prior

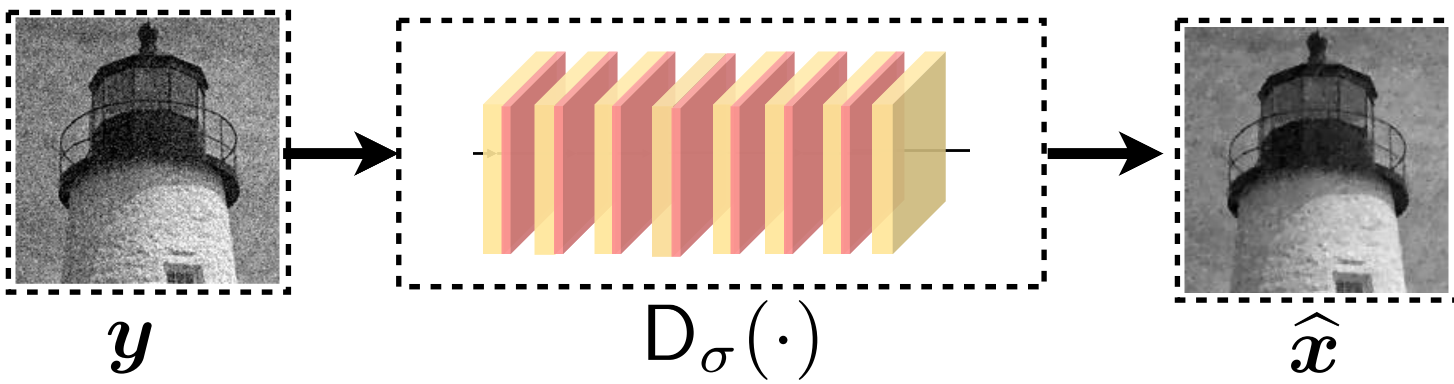
Idea: Treat a denoising CNN as a proximal operator



Combine operators into **Plug-and-Play Priors (PnP)** algorithms

Plug-and-play priors (PnP) methods separate the forward model from the learned prior

Idea: Treat a denoising CNN as a proximal operator



Combine operators into Plug-and-Play Priors (PnP) algorithms

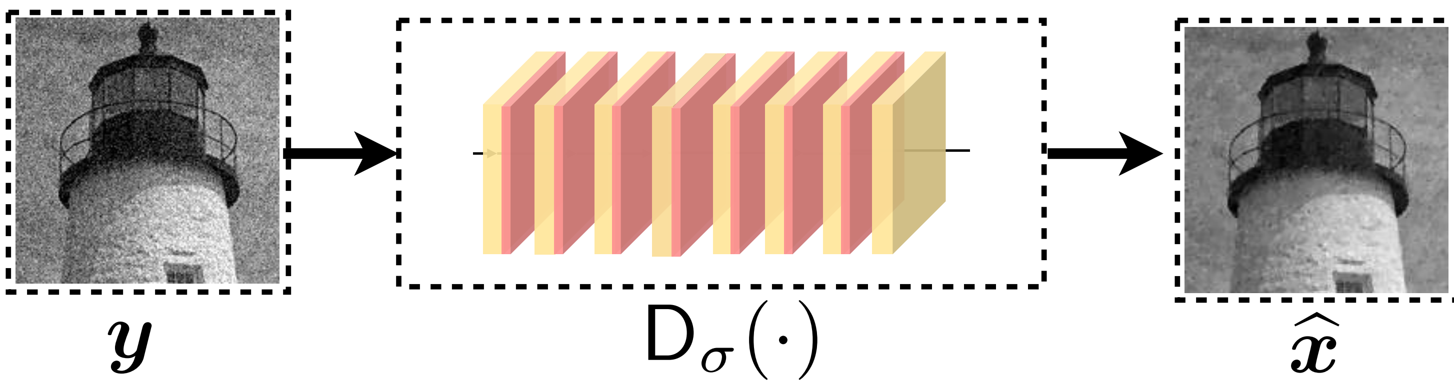
$$\begin{aligned}
\boldsymbol{z}^k &\leftarrow \mathsf{prox}_{\gamma g}(\boldsymbol{x}^{k-1} - \boldsymbol{s}^{k-1}) \\
\boldsymbol{x}^k &\leftarrow \mathsf{D}_\sigma(\boldsymbol{z}^k + \boldsymbol{s}^{k-1}) \\
\boldsymbol{s}^k &\leftarrow \boldsymbol{s}^{k-1} + (\boldsymbol{z}^k - \boldsymbol{x}^k)
\end{aligned}$$

PnP-ADMM

Venkatakrishnan *et al.*, "Plug-and-Play Priors for Model Based Reconstruction," *Proc. IEEE GlobalSIP*, pp. 945-948, December 2013.

Plug-and-play priors (PnP) methods separate the forward model from the learned prior

Idea: Treat a denoising CNN as a proximal operator



Combine operators into Plug-and-Play Priors (PnP) algorithms

$$\begin{aligned}
\boldsymbol{z}^k &\leftarrow \mathsf{prox}_{\gamma g}(\boldsymbol{x}^{k-1} - \boldsymbol{s}^{k-1}) \\
\boldsymbol{x}^k &\leftarrow \mathsf{D}_\sigma(\boldsymbol{z}^k + \boldsymbol{s}^{k-1}) \\
\boldsymbol{s}^k &\leftarrow \boldsymbol{s}^{k-1} + (\boldsymbol{z}^k - \boldsymbol{x}^k)
\end{aligned}$$

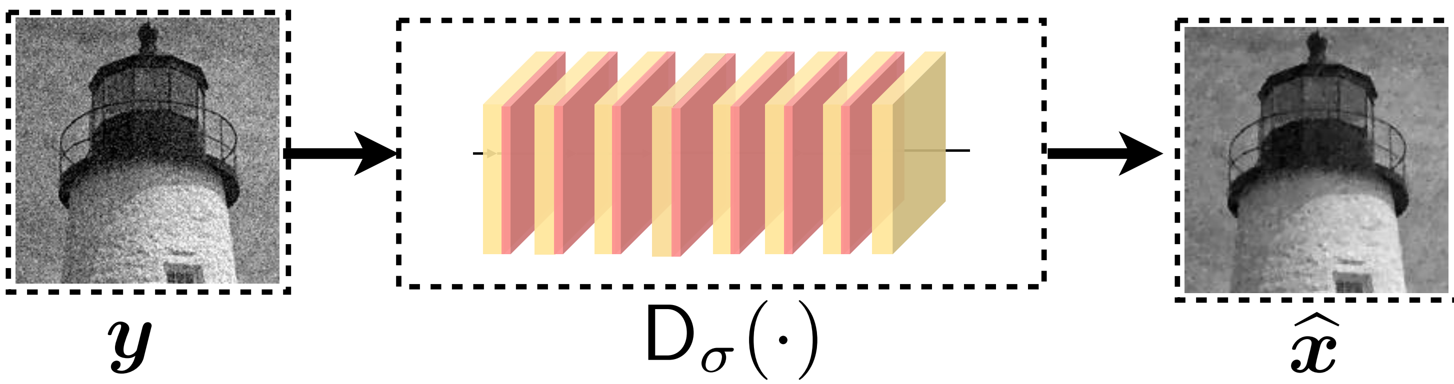
$$\begin{aligned}
\boldsymbol{z}^k &\leftarrow \boldsymbol{s}^{k-1} - \gamma \nabla g(\boldsymbol{s}^{k-1}) \\
\boldsymbol{x}^k &\leftarrow \mathsf{D}_\sigma(\boldsymbol{z}^k) \\
\boldsymbol{s}^k &\leftarrow \boldsymbol{x}^k + ((q_{k-1} - 1)/q_k)(\boldsymbol{x}^k - \boldsymbol{x}^{k-1})
\end{aligned}$$

PnP-FISTA

Kamilov *et al.*, “A Plug-and-Play Priors Approach for Solving Nonlinear Imaging Inverse Problems,” *IEEE Signal Process. Lett.*, 2017

Plug-and-play priors (PnP) methods separate the forward model from the learned prior

Idea: Treat a denoising CNN as a proximal operator



Combine operators into Plug-and-Play Priors (PnP) algorithms

$$\boldsymbol{z}^k \leftarrow \mathsf{prox}_{\gamma g}(\boldsymbol{x}^{k-1} - \boldsymbol{s}^{k-1})$$
$$\boldsymbol{x}^k \leftarrow \mathsf{D}_\sigma(\boldsymbol{z}^k + \boldsymbol{s}^{k-1})$$
$$\boldsymbol{s}^k \leftarrow \boldsymbol{s}^{k-1} + (\boldsymbol{z}^k - \boldsymbol{x}^k)$$

PnP-ADMM

$$\boldsymbol{z}^k \leftarrow \boldsymbol{s}^{k-1} - \gamma \nabla g(\boldsymbol{s}^{k-1})$$
$$\boldsymbol{x}^k \leftarrow \mathsf{D}_\sigma(\boldsymbol{z}^k)$$
$$\boldsymbol{s}^k \leftarrow \boldsymbol{x}^k + ((q_{k-1} - 1)/q_k)(\boldsymbol{x}^k - \boldsymbol{x}^{k-1})$$

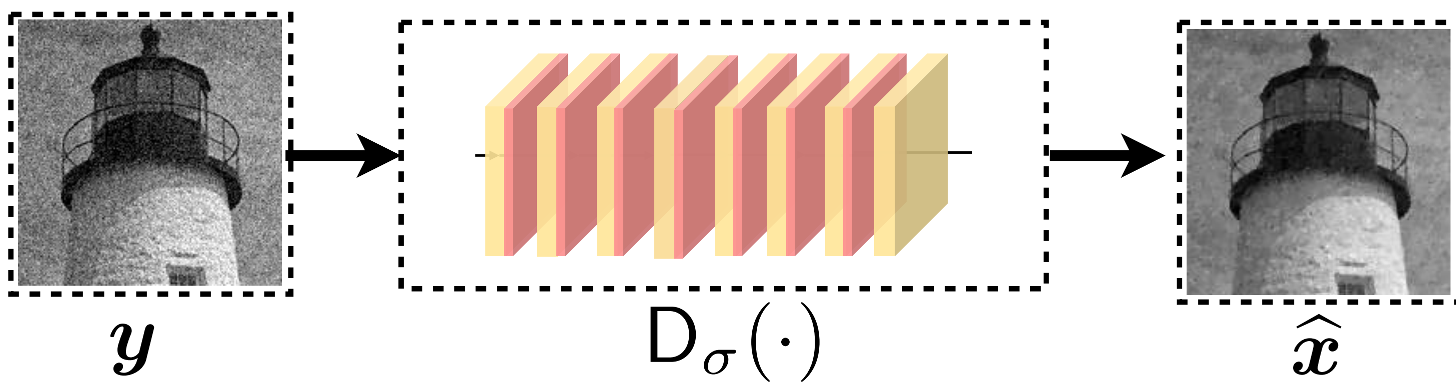
PnP-FISTA

Venkatakrishnan *et al.*, "Plug-and-Play Priors for Model Based Reconstruction," *Proc. IEEE GlobalSIP*, pp. 945-948, December 2013.

Kamilov *et al.*, "A Plug-and-Play Priors Approach for Solving Nonlinear Imaging Inverse Problems," *IEEE Signal Process. Lett.*, 2017

Plug-and-play priors (PnP) methods separate the forward model from the learned prior

Idea: Treat a denoising CNN as a proximal operator



Combine operators into Plug-and-Play Priors (PnP) algorithms

$$\boldsymbol{z}^k \leftarrow \mathsf{prox}_{\gamma g}(\boldsymbol{x}^{k-1} - \boldsymbol{s}^{k-1})$$
$$\boldsymbol{x}^k \leftarrow \mathsf{D}_\sigma(\boldsymbol{z}^k + \boldsymbol{s}^{k-1})$$
$$\boldsymbol{s}^k \leftarrow \boldsymbol{s}^{k-1} + (\boldsymbol{z}^k - \boldsymbol{x}^k)$$

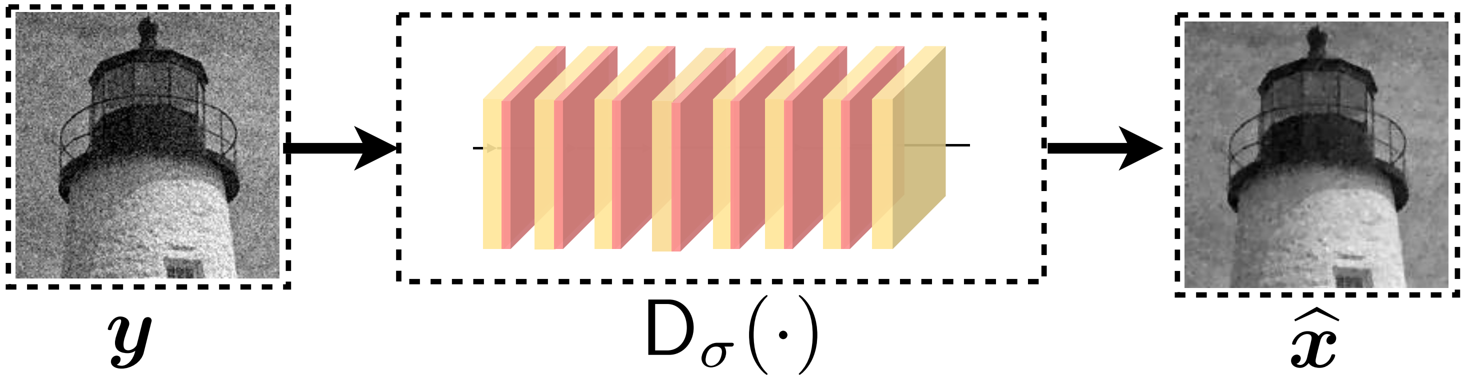
$$\boldsymbol{z}^k \leftarrow \boldsymbol{s}^{k-1} - \gamma \nabla g(\boldsymbol{s}^{k-1})$$ data consistent
$$\boldsymbol{x}^k \leftarrow \mathsf{D}_\sigma(\boldsymbol{z}^k)$$ less noisy
$$\boldsymbol{s}^k \leftarrow \boldsymbol{x}^k + ((q_{k-1} - 1)/q_k)(\boldsymbol{x}^k - \boldsymbol{x}^{k-1})$$

Venkatakrishnan *et al.*, “Plug-and-Play Priors for Model Based Reconstruction,” *Proc. IEEE GlobalSIP*, pp. 945-948, December 2013.

Kamilov *et al.*, “A Plug-and-Play Priors Approach for Solving Nonlinear Imaging Inverse Problems,” *IEEE Signal Process. Lett.*, 2017

Plug-and-play priors (PnP) methods separate the forward model from the learned prior

Idea: Treat a denoising CNN as a proximal operator



Combine operators into Plug-and-Play Priors (PnP) algorithms

$$\boldsymbol{z}^k \leftarrow \mathsf{prox}_{\gamma g}(\boldsymbol{x}^{k-1} - \boldsymbol{s}^{k-1})$$
$$\boldsymbol{x}^k \leftarrow \mathsf{D}_\sigma(\boldsymbol{z}^k + \boldsymbol{s}^{k-1})$$
$$\boldsymbol{s}^k \leftarrow \boldsymbol{s}^{k-1} + (\boldsymbol{z}^k - \boldsymbol{x}^k)$$

$$\boldsymbol{z}^k \leftarrow \boldsymbol{s}^{k-1} - \gamma \nabla g(\boldsymbol{s}^{k-1})$$ data consistent
$$\boldsymbol{x}^k \leftarrow \mathsf{D}_\sigma(\boldsymbol{z}^k)$$ less noisy
$$\boldsymbol{s}^k \leftarrow \boldsymbol{x}^k + ((q_{k-1} - 1)/q_k)(\boldsymbol{x}^k - \boldsymbol{x}^{k-1})$$

Venkatakrishnan *et al.*, "Plug-and-Play Priors for Model Based Reconstruction," *Proc. IEEE GlobalSIP*, pp. 945-948, December 2013.

Kamilov *et al.*, "A Plug-and-Play Priors Approach for Solving Nonlinear Imaging Inverse Problems," *IEEE Signal Process. Lett.*, 2017

RED is an alternative to PnP that seeks to have an explicit regularizer for a given denoiser

RED is an alternative to PnP that seeks to have an explicit regularizer for a given denoiser

Consider the following GM-RED algorithm

$$\boldsymbol{x}^t \leftarrow \boldsymbol{x}^{t-1} - \gamma \mathsf{G}(\boldsymbol{x}^{t-1})$$

"gradient" descent

$$\mathsf{G}(\boldsymbol{x}) := \nabla g(\boldsymbol{x}) + \tau(\boldsymbol{x} - \mathsf{D}_\sigma(\boldsymbol{x}))$$

data — prior

Romano *et al.*, "The Little Engine That Could: Regularization by Denoising," *SIAM J. Imaging Sci.*, vol. 10, no. 4, 2017

RED is an alternative to PnP that seeks to have an explicit regularizer for a given denoiser

Consider the following GM-RED algorithm

$$\boldsymbol{x}^t \leftarrow \boldsymbol{x}^{t-1} - \gamma \mathsf{G}(\boldsymbol{x}^{t-1}) \qquad \mathsf{G}(\boldsymbol{x}) := \nabla g(\boldsymbol{x}) + \tau(\boldsymbol{x} - \mathsf{D}_\sigma(\boldsymbol{x}))$$

For a locally homogeneous denoiser that has a symmetric Jacobian, GM-RED* solves the following problem

$$\min_{\boldsymbol{x}} \{\underbrace{g(\boldsymbol{x})}_{\text{data}} + \underbrace{h_{\mathsf{red}}(\boldsymbol{x})}_{\text{prior}}\} \qquad h_{\mathsf{red}}(\boldsymbol{x}) := \frac{\tau}{2}\boldsymbol{x}^{\mathsf{T}}(\boldsymbol{x} - \mathsf{D}_\sigma(\boldsymbol{x}))$$

Romano *et al.*, "The Little Engine That Could: Regularization by Denoising," *SIAM J. Imaging Sci.*, vol. 10, no. 4, 2017

Reehorst *et al.*, "Regularization by Denoising: Clarifications and New Interpretations," *IEEE TCI*, vol. 5, no. 1, 2019

RED is an alternative to PnP that seeks to have an explicit regularizer for a given denoiser

Consider the following GM-RED algorithm

$$\boldsymbol{x}^t \leftarrow \boldsymbol{x}^{t-1} - \gamma \mathsf{G}(\boldsymbol{x}^{t-1}) \qquad \mathsf{G}(\boldsymbol{x}) := \nabla g(\boldsymbol{x}) + \tau(\boldsymbol{x} - \mathsf{D}_\sigma(\boldsymbol{x}))$$

For a locally homogeneous denoiser that has a symmetric Jacobian, GM-RED* solves the following problem

$$\min_{\boldsymbol{x}} \{g(\boldsymbol{x}) + h_{\mathsf{red}}(\boldsymbol{x})\} \qquad h_{\mathsf{red}}(\boldsymbol{x}) := \frac{\tau}{2}\boldsymbol{x}^{\mathsf{T}}(\boldsymbol{x} - \mathsf{D}_\sigma(\boldsymbol{x}))$$

While this convex-optimization interpretation of RED is powerful, it also restricts the class of acceptable denoisers

Recall that one of the key conditions we assumed for the denoiser $f(\mathbf{x})$ is that it is homogeneous of degree 1,

$$f(c\mathbf{x}) = cf(\mathbf{x}). \tag{51}$$

This immediately notifies us that the $\rho(\mathbf{x})$ we wish to mimic in (50) must be 2-homogeneous

Appendix A: Can we mimic any denoiser?

Romano *et al.*, "The Little Engine That Could: Regularization by Denoising," *SIAM J. Imaging Sci.*, vol. 10, no. 4, 2017

Regularization by Artifact Removal (RARE) generalizes RED beyond AWGN denoisers

Regularization by Artifact Removal (RARE) generalizes RED beyond AWGN denoisers

Consider the gradient method (GM) variant of RARE

$$\boldsymbol{x}^t \leftarrow \boldsymbol{x}^{t-1} - \gamma \mathsf{G}(\boldsymbol{x}^{t-1})$$

"gradient" descent

$$\mathsf{G}(\boldsymbol{x}) := \nabla g(\boldsymbol{x}) + \tau(\boldsymbol{x} - \mathsf{R}_{\boldsymbol{\theta}}(\boldsymbol{x}))$$

data prior

Sun *et al.*, "Block Coordinate Regularization by Denoising," *Proc. NeurIPS*, Dec 2019

Liu *et al.*, "RARE: Image Reconstruction using Deep Priors Learned without Ground Truth," arXiv:1912.05854, 2019

Regularization by Artifact Removal (RARE) generalizes RED beyond AWGN denoisers

Consider the gradient method (GM) variant of RARE

$$\boldsymbol{x}^t \leftarrow \boldsymbol{x}^{t-1} - \gamma \mathsf{G}(\boldsymbol{x}^{t-1}) \qquad \mathsf{G}(\boldsymbol{x}) := \nabla g(\boldsymbol{x}) + \tau(\boldsymbol{x} - \mathsf{R}_{\boldsymbol{\theta}}(\boldsymbol{x}))$$

and the set of fixed points of G

$$\mathsf{zer}(\mathsf{G}) := \{\boldsymbol{x} : \mathsf{G}(\boldsymbol{x}) = \mathbf{0}\}$$

a set of all images where G is zero

Sun *et al.*, “Block Coordinate Regularization by Denoising,” *Proc. NeurIPS*, Dec 2019

Liu *et al.*, “RARE: Image Reconstruction using Deep Priors Learned without Ground Truth,” arXiv:1912.05854, 2019

Regularization by Artifact Removal (RARE) generalizes RED beyond AWGN denoisers

Consider the gradient method (GM) variant of RARE

$$\boldsymbol{x}^t \leftarrow \boldsymbol{x}^{t-1} - \gamma \mathsf{G}(\boldsymbol{x}^{t-1}) \qquad \mathsf{G}(\boldsymbol{x}) := \nabla g(\boldsymbol{x}) + \tau(\boldsymbol{x} - \mathsf{R}_{\boldsymbol{\theta}}(\boldsymbol{x}))$$

and the set of fixed points of G

$$\mathsf{zer}(\mathsf{G}) := \{\boldsymbol{x} : \mathsf{G}(\boldsymbol{x}) = \mathbf{0}\}$$

RARE leverages deep neural net priors without differentiating them, by also giving an explicit control over data fidelity

If $\boldsymbol{x}^* \in \mathsf{zer}(\nabla g) \cap \mathsf{fix}(\mathsf{R}_{\boldsymbol{\theta}})$, then $\boldsymbol{x}^* \in \mathsf{zer}(\mathsf{G})$

If $\mathsf{zer}(\nabla g) \cap \mathsf{fix}(\mathsf{R}_{\boldsymbol{\theta}}) = \varnothing$, then $\boldsymbol{x}^* \in \mathsf{zer}(\mathsf{G})$ is a trade-off controlled by $\tau > 0$

Sun *et al.*, “Block Coordinate Regularization by Denoising,” *Proc. NeurIPS*, Dec 2019

Liu *et al.*, “RARE: Image Reconstruction using Deep Priors Learned without Ground Truth,” arXiv:1912.05854, 2019

RARE might look heuristic, but it has a rigorous foundation in monotone operator theory

RARE might look heuristic, but it has a rigorous foundation in monotone operator theory

Theorem 1. Run RARE for $t \geq 1$ iterations under Assumptions 1-3 using a fixed step-size $0 < \gamma \leq 1/(L + 2\tau)$. Then, we have

$$\|\mathsf{G}(\boldsymbol{x}^{t-1})\|_2^2 \leq \frac{(L+2\tau)}{\gamma t} R_0^2$$

1 – RARE converges to zer(G) as O(1/t)

Sun *et al.*, "Block Coordinate Regularization by Denoising," *Proc. NeurIPS*, Dec 2019

Liu *et al.*, "RARE: Image Reconstruction using Deep Priors Learned without Ground Truth," arXiv:1912.05854, 2019

RARE might look heuristic, but it has a rigorous foundation in monotone operator theory

Theorem 1. Run RARE for $t \geq 1$ iterations under Assumptions 1-3 using a fixed step-size $0 < \gamma \leq 1/(L + 2\tau)$. Then, we have

$$\|\mathsf{G}(\boldsymbol{x}^{t-1})\|_2^2 \leq \frac{(L+2\tau)}{\gamma t} R_0^2$$

Theorem 2. Run RARE for $t \geq 1$ iterations with a proximal operator denoiser under Assumptions 1-4 using a fixed step-size $0 < \gamma \leq 1/(L + 2\tau)$. Then, we have

$$f(\boldsymbol{x}^t) - f^* \leq \frac{2}{\gamma t} R_0^2 + \frac{G_0^2}{2\tau},$$

where the function $f = g + h$ and f^* is its minimum.

2 – RARE is backward compatible with proximal optimization

Sun *et al.*, "Block Coordinate Regularization by Denoising," *Proc. NeurIPS*, Dec 2019

Liu *et al.*, "RARE: Image Reconstruction using Deep Priors Learned without Ground Truth," arXiv:1912.05854, 2019

RARE might look heuristic, but it has a rigorous foundation in monotone operator theory

Theorem 1. Run RARE for $t \geq 1$ iterations under Assumptions 1-3 using a fixed step-size $0 < \gamma \leq 1/(L + 2\tau)$. Then, we have

$$\|\mathsf{G}(\boldsymbol{x}^{t-1})\|_2^2 \leq \frac{(L + 2\tau)}{\gamma t} R_0^2$$

Theorem 2. Run RARE for $t \geq 1$ iterations with a proximal operator denoiser under Assumptions 1-4 using a fixed step-size $0 < \gamma \leq 1/(L + 2\tau)$. Then, we have

$$f(\boldsymbol{x}^t) - f^* \leq \frac{2}{\gamma t} R_0^2 + \frac{G_0^2}{2\tau},$$

where the function $f = g + h$ and f^* is its minimum.

Sun *et al.*, "Block Coordinate Regularization by Denoising," *Proc. NeurIPS*, Dec 2019

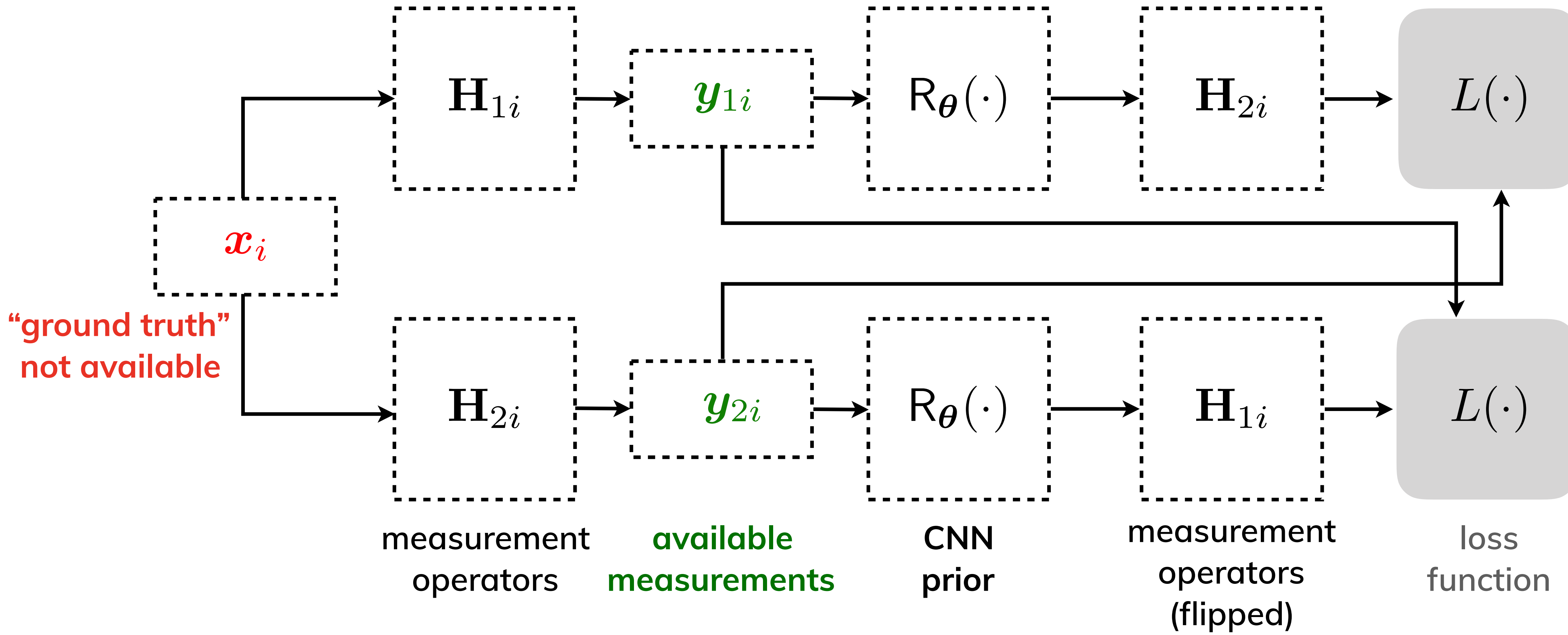
Liu *et al.*, "RARE: Image Reconstruction using Deep Priors Learned without Ground Truth," arXiv:1912.05854, 2019

RARE leads to a simple recipe for training CNN priors without ground truth

Lehtinen *et al.*, “Noise2Noise: Learning Image Restoration without Clean Data,” Proc. ICML, 2018

RARE leads to a simple recipe for training CNN priors without ground truth

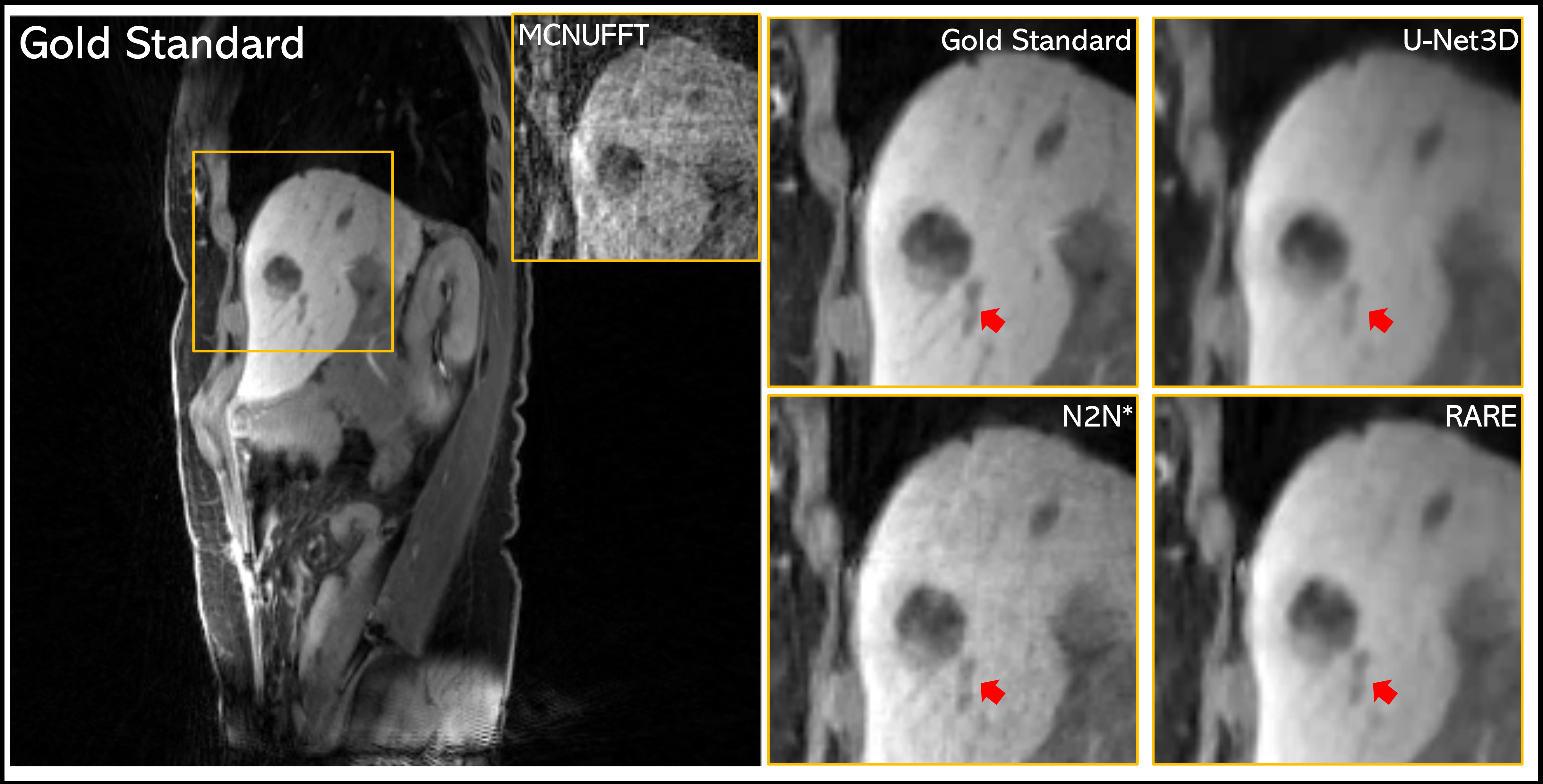
Consider several random and independent views for each object
(we consider training of artifact-removing CNNs)



Lehtinen *et al.*, "Noise2Noise: Learning Image Restoration without Clean Data," Proc. ICML, 2018

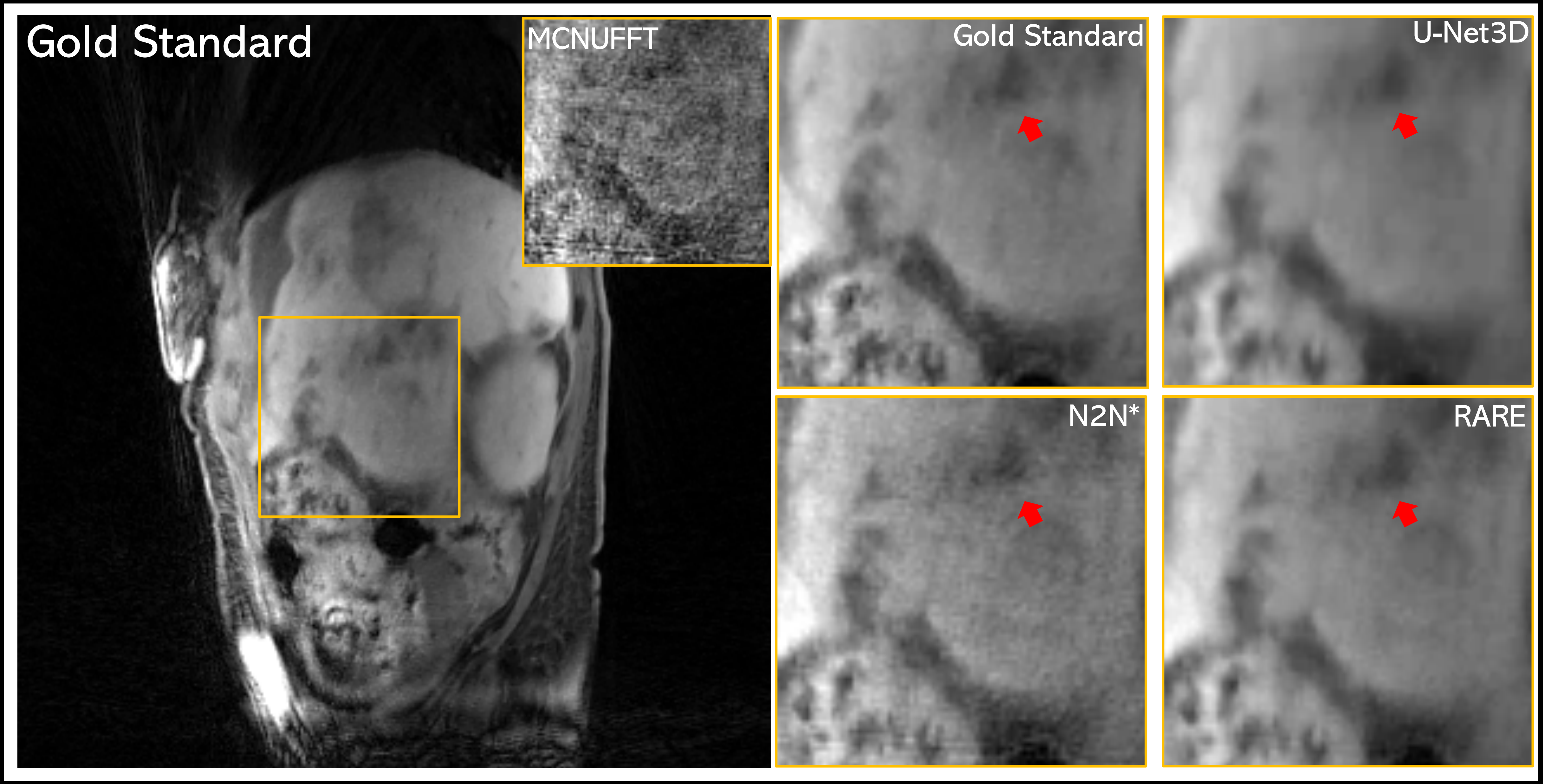
Liu *et al.*, "RARE: Image Reconstruction using Deep Priors Learned without Ground Truth," arXiv:1912.05854, 2019

Application: Reconstructing 10 motion phases from a 1 minute free-breathing MRI scan



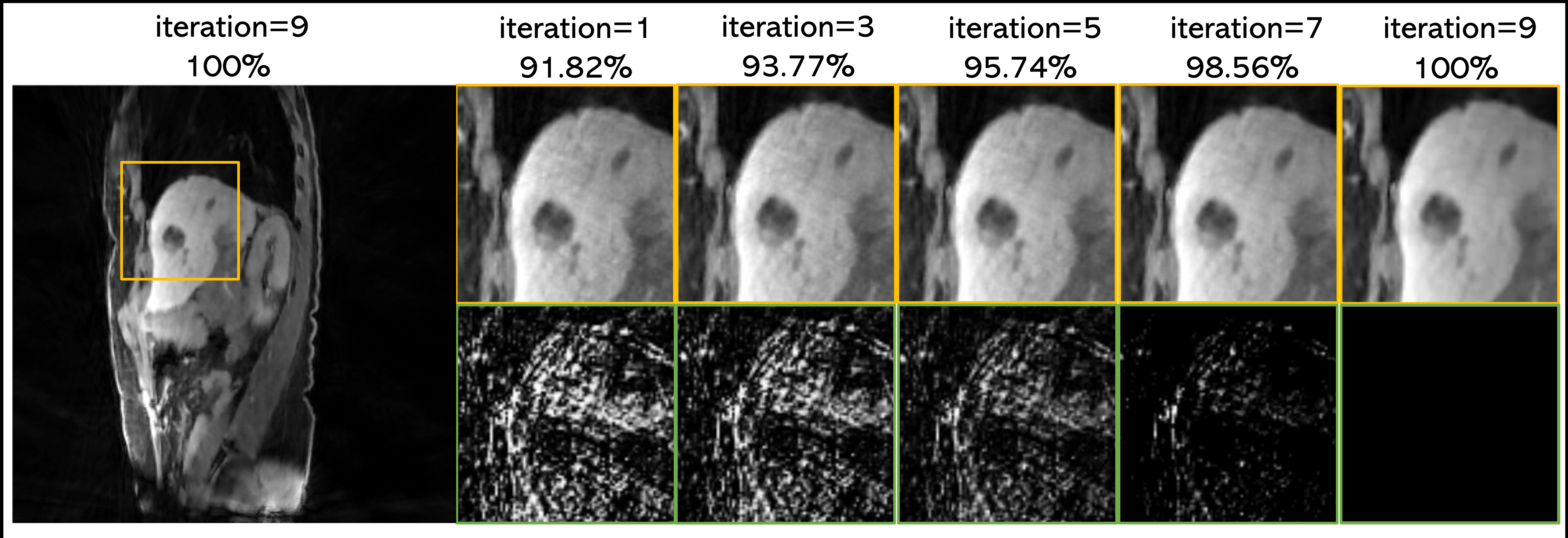
Liu *et al.*, "RARE: Image Reconstruction using Deep Priors Learned without Ground Truth," arXiv:1912.05854, 2019

Application: Reconstructing 10 motion phases from a 1 minute free-breathing MRI scan



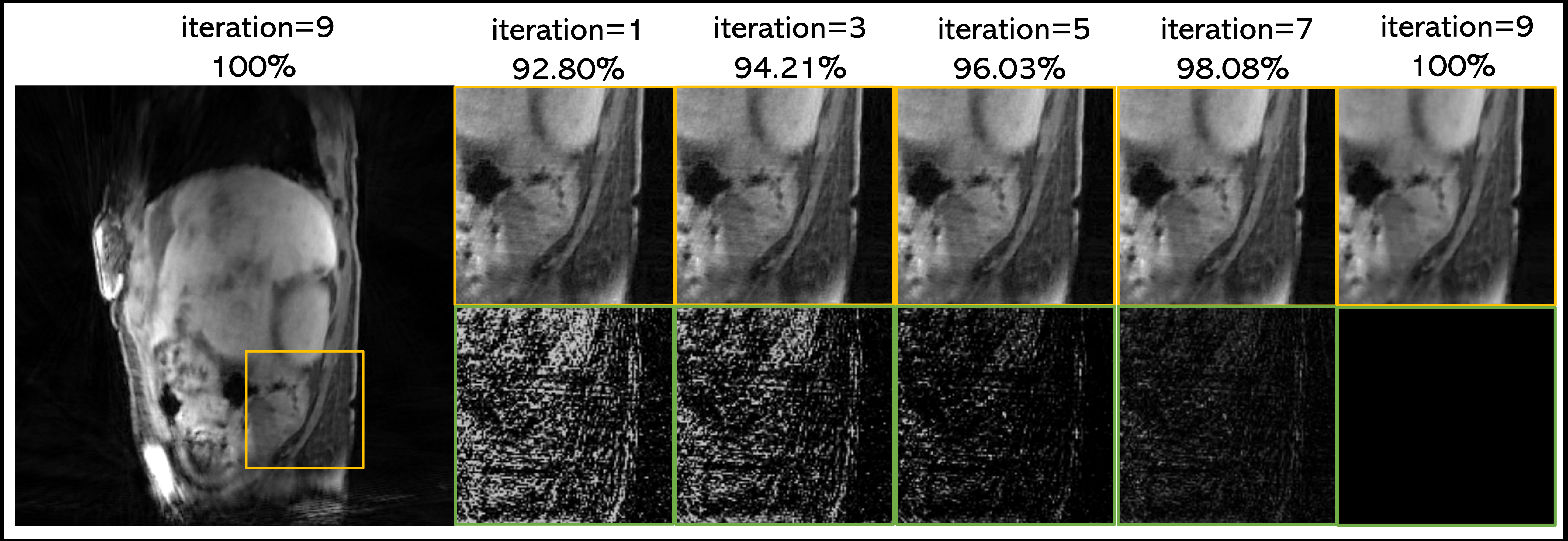
Liu *et al.*, “RARE: Image Reconstruction using Deep Priors Learned without Ground Truth,” arXiv:1912.05854, 2019

Application: Reconstructing 10 motion phases from a 1 minute free-breathing MRI scan



Liu *et al.*, "RARE: Image Reconstruction using Deep Priors Learned without Ground Truth," arXiv:1912.05854, 2019

Application: Reconstructing 10 motion phases from a 1 minute free-breathing MRI scan



Liu *et al.*, “RARE: Image Reconstruction using Deep Priors Learned without Ground Truth,” arXiv:1912.05854, 2019

To conclude

- We increasingly rely on implicit regularization using nonlinear operators, such as deep neural nets
- PnP and RED are algorithms that enable combining a learned CNN denoiser with the forward model
- RARE considers more general artifact removing CNN priors that are easier to train without ground truth
- For certain large-scale problems, it is beneficial to consider **streaming algorithms** that we will discuss next

Today we will talk about

- Imaging as an inverse problem
  Infusing prior knowledge into image formation
- RARE: Regularization by Artifact Removal
  Using CNN priors learned without ground truth
- SIMBA: Scalable algorithms using CNN priors
  Enabling large-scale tomographic imaging

For many imaging inverse problems, the traditional batch data processing is suboptimal

Washington University in St. Louis

For many imaging inverse problems, the traditional batch data processing is suboptimal

Consider the data-fidelity of the following form

$$g(\boldsymbol{x}) = \frac{1}{L}\sum_{\ell=1}^{L} g_\ell(\boldsymbol{x}) = \frac{1}{2L}\sum_{\ell=1}^{L} \|\boldsymbol{y} - \mathbf{H}_\ell(\boldsymbol{x})\|_2^2$$

Sum of L data-fidelity terms

For many imaging inverse problems, the traditional batch data processing is suboptimal

Consider the data-fidelity of the following form

$$g(\boldsymbol{x}) = \frac{1}{L}\sum_{\ell=1}^{L} g_\ell(\boldsymbol{x}) = \frac{1}{2L}\sum_{\ell=1}^{L} \|\boldsymbol{y} - \mathbf{H}_\ell(\boldsymbol{x})\|_2^2$$

Consider the complexity of computing
the **full-batch gradient** and the **mini-batch gradient**

$$\nabla g(\boldsymbol{x}) = \frac{1}{L}\sum_{\ell=1}^{L} \nabla g_\ell(\boldsymbol{x})$$

Uses all the measurements

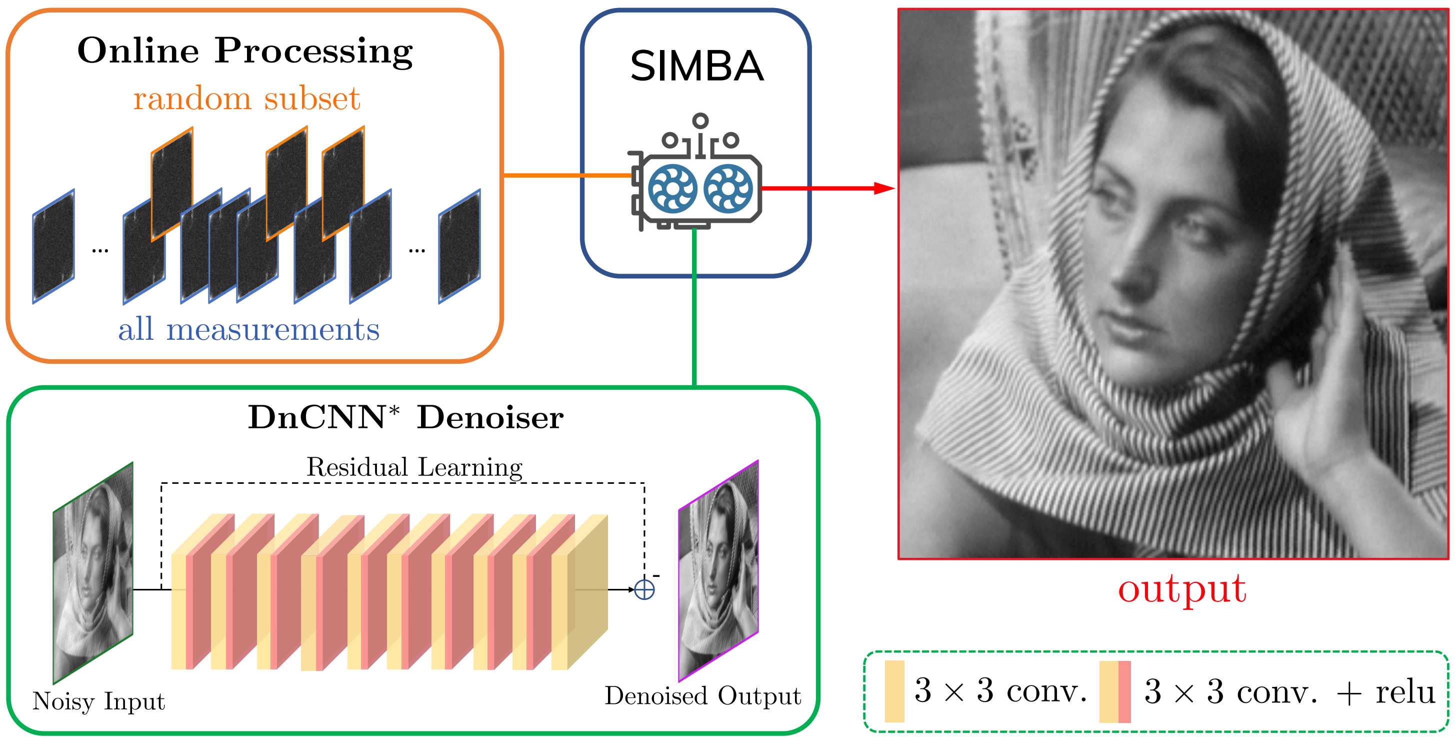
vs.

$$\hat{\nabla} g(\boldsymbol{x}) = \frac{1}{B}\sum_{b=1}^{B} \nabla g_{\ell_b}(\boldsymbol{x})$$

Uses a subset of measurements

SIMBA is a “streaming” variant of RARE suitable for large-scale imaging problems

SIMBA is a "streaming" variant of RARE suitable for large-scale imaging problems



SIMBA uses only B << L measurements per iteration

RARE

$$\nabla g(\boldsymbol{x}^{k-1}) \leftarrow \mathsf{FullGradient}(\boldsymbol{x}^{k-1})$$
$$\mathsf{G}(\boldsymbol{x}^{k-1}) \leftarrow \nabla g(\boldsymbol{x}^{k-1}) + \tau(\boldsymbol{x}^{k-1} - \mathsf{R}_{\boldsymbol{\theta}}(\boldsymbol{x}^{k-1}))$$
$$\boldsymbol{x}^{k} \leftarrow \boldsymbol{x}^{k-1} - \gamma \mathsf{G}(\boldsymbol{x}^{k-1})$$

SIMBA

$$\widehat{\nabla} g(\boldsymbol{x}^{k-1}) \leftarrow \mathsf{MinibatchGradient}(\boldsymbol{x}^{k-1})$$
$$\mathsf{G}(\boldsymbol{x}^{k-1}) \leftarrow \widehat{\nabla} g(\boldsymbol{x}^{k-1}) + \tau(\boldsymbol{x}^{k-1} - \mathsf{R}_{\boldsymbol{\theta}}(\boldsymbol{x}^{k-1}))$$
$$\boldsymbol{x}^{k} \leftarrow \boldsymbol{x}^{k-1} - \gamma \mathsf{G}(\boldsymbol{x}^{k-1})$$

Wu *et al.*, "Online Regularization by Denoising with Applications to Phase Retrieval," *Proc. ICCVW*, 2019

Wu *et al.*, "SIMBA: Scalable Inversion in Optical Tomography using Deep Denoising Priors," arXiv:1911.13241, 2019

SIMBA is a "streaming" variant of RARE suitable for large-scale imaging problems

Theorem 1. Run SIMBA for $t \geq 1$ iterations under Assumptions 1-3 using a fixed step-size $0 < \gamma \leq 1/(L + 2\tau)$ and a fixed minibatch size $B = t$. Then, we have

$$\mathbb{E}\left[\frac{1}{t}\sum_{k=1}^{t}\|\mathsf{G}(\boldsymbol{x}^{k-1})\|_2^2\right] \leq \frac{C}{\sqrt{t}},$$

where $C > 0$ is a constant.

Convergence similar to SGD

SIMBA uses only B << L measurements per iteration

RARE

$\nabla g(\boldsymbol{x}^{k-1}) \leftarrow \mathsf{FullGradient}(\boldsymbol{x}^{k-1})$

$\mathsf{G}(\boldsymbol{x}^{k-1}) \leftarrow \nabla g(\boldsymbol{x}^{k-1}) + \tau(\boldsymbol{x}^{k-1} - \mathsf{R}_{\boldsymbol{\theta}}(\boldsymbol{x}^{k-1}))$

$\boldsymbol{x}^k \leftarrow \boldsymbol{x}^{k-1} - \gamma\mathsf{G}(\boldsymbol{x}^{k-1})$

SIMBA

$\widehat{\nabla} g(\boldsymbol{x}^{k-1}) \leftarrow \mathsf{MinibatchGradient}(\boldsymbol{x}^{k-1})$

$\mathsf{G}(\boldsymbol{x}^{k-1}) \leftarrow \widehat{\nabla} g(\boldsymbol{x}^{k-1}) + \tau(\boldsymbol{x}^{k-1} - \mathsf{R}_{\boldsymbol{\theta}}(\boldsymbol{x}^{k-1}))$

$\boldsymbol{x}^k \leftarrow \boldsymbol{x}^{k-1} - \gamma\mathsf{G}(\boldsymbol{x}^{k-1})$

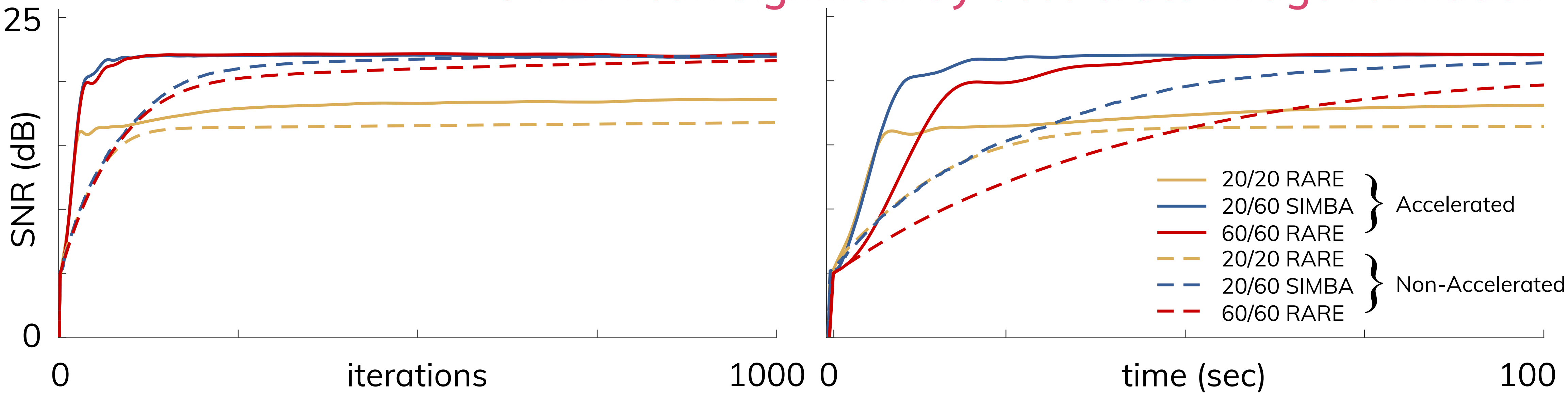
Wu *et al.*, "Online Regularization by Denoising with Applications to Phase Retrieval," *Proc. ICCVW*, 2019

Wu *et al.*, "SIMBA: Scalable Inversion in Optical Tomography using Deep Denoising Priors," arXiv:1911.13241, 2019

Washington University in St. Louis

SIMBA is a “streaming” variant of RARE suitable for large-scale imaging problems

SIMBA can significantly accelerate image formation



SIMBA uses only B << L measurements per iteration

RARE

$$\nabla g(\boldsymbol{x}^{k-1}) \leftarrow \mathsf{FullGradient}(\boldsymbol{x}^{k-1})$$

$$\mathsf{G}(\boldsymbol{x}^{k-1}) \leftarrow \nabla g(\boldsymbol{x}^{k-1}) + \tau(\boldsymbol{x}^{k-1} - \mathsf{R}_{\boldsymbol{\theta}}(\boldsymbol{x}^{k-1}))$$

$$\boldsymbol{x}^{k} \leftarrow \boldsymbol{x}^{k-1} - \gamma \mathsf{G}(\boldsymbol{x}^{k-1})$$

SIMBA

$$\widehat{\nabla} g(\boldsymbol{x}^{k-1}) \leftarrow \mathsf{MinibatchGradient}(\boldsymbol{x}^{k-1})$$

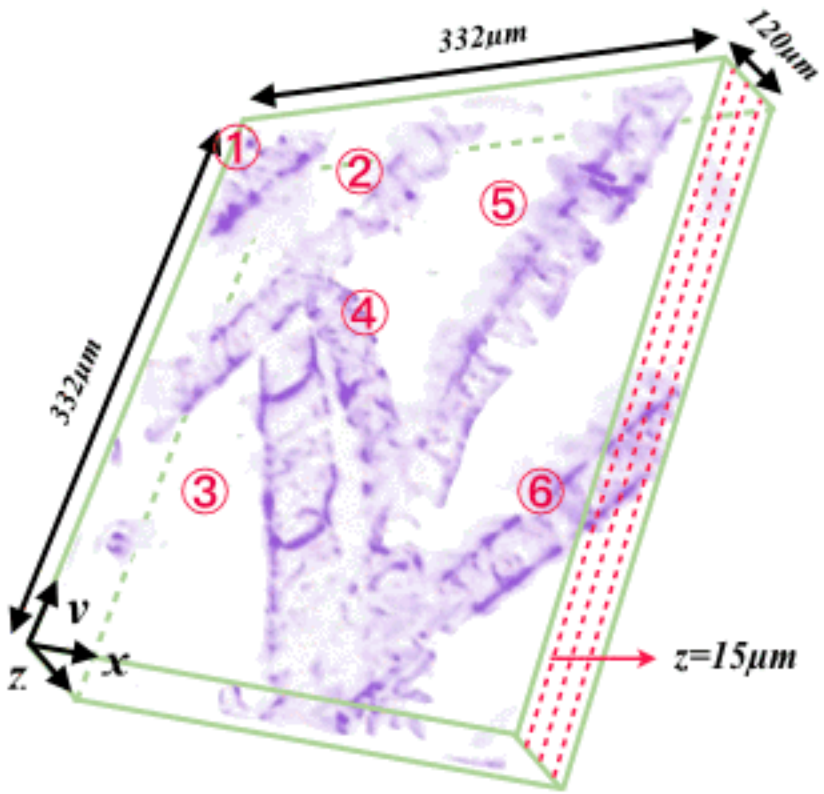
$$\mathsf{G}(\boldsymbol{x}^{k-1}) \leftarrow \widehat{\nabla} g(\boldsymbol{x}^{k-1}) + \tau(\boldsymbol{x}^{k-1} - \mathsf{R}_{\boldsymbol{\theta}}(\boldsymbol{x}^{k-1}))$$

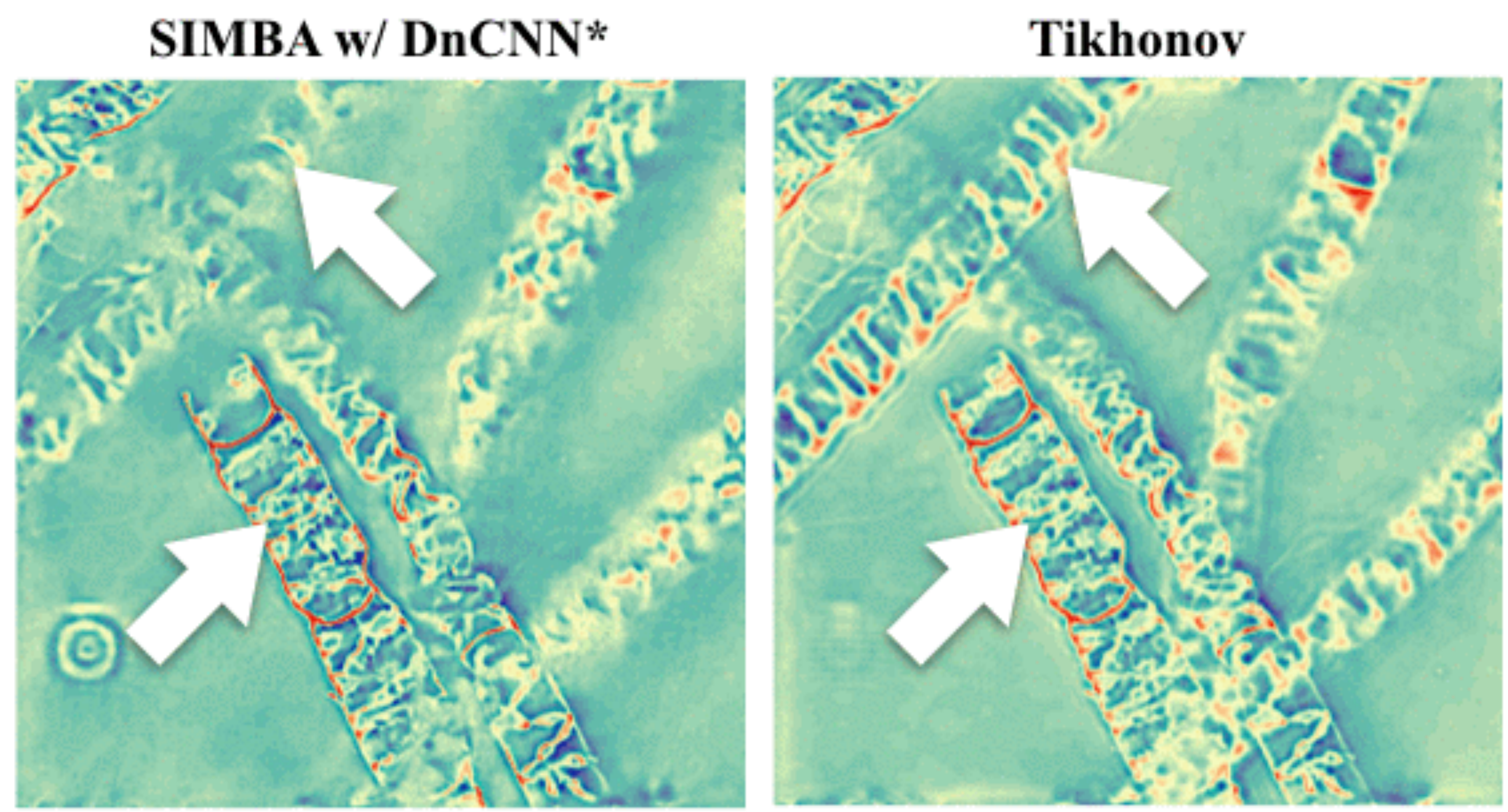
$$\boldsymbol{x}^{k} \leftarrow \boldsymbol{x}^{k-1} - \gamma \mathsf{G}(\boldsymbol{x}^{k-1})$$

Wu *et al.*, “Online Regularization by Denoising with Applications to Phase Retrieval,” *Proc. ICCVW*, 2019

Wu *et al.*, “SIMBA: Scalable Inversion in Optical Tomography using Deep Denoising Priors,” arXiv:1911.13241, 2019

SIMBA is a “streaming” variant of RARE suitable for large-scale imaging problems

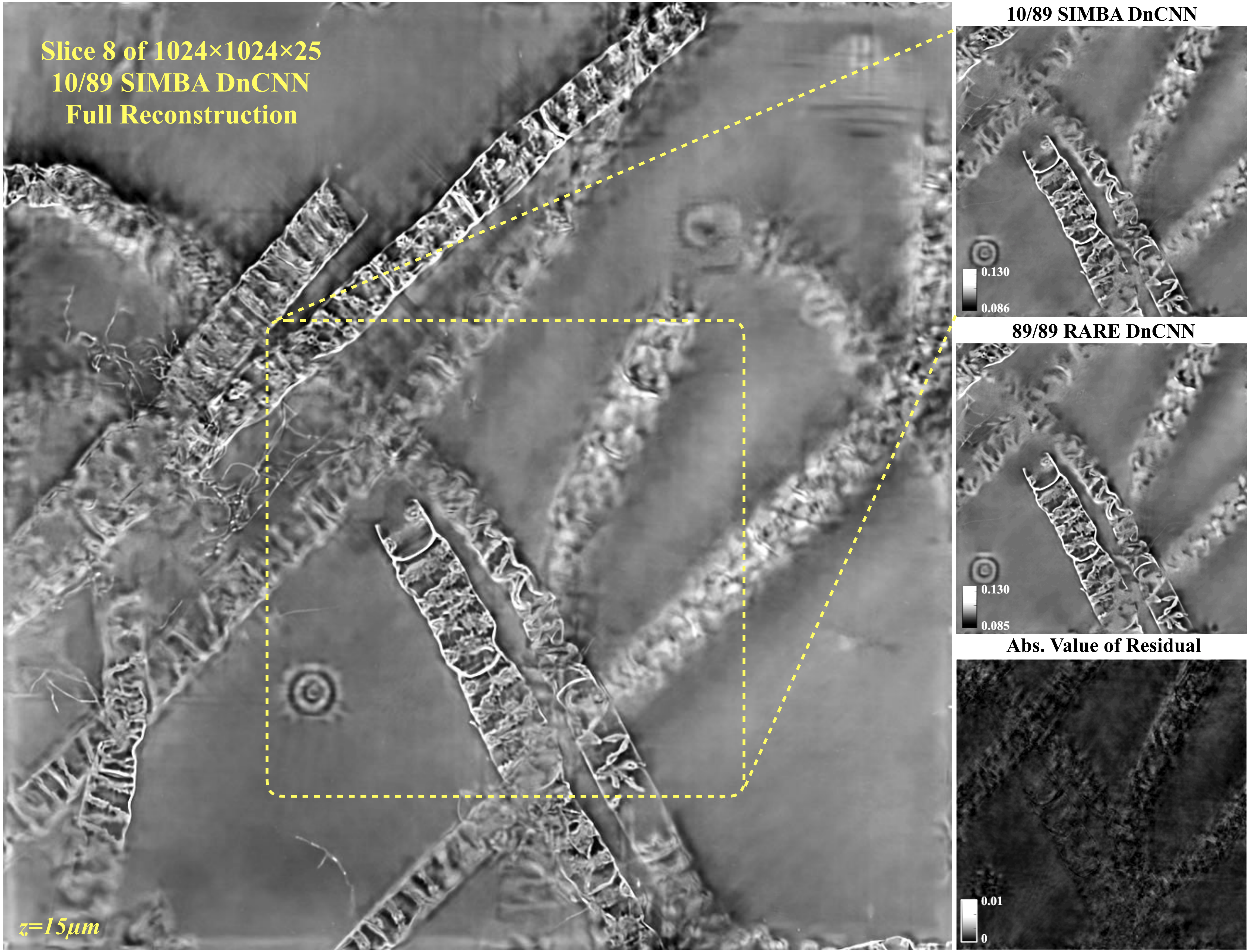




SIMBA leads to better 3D sectioning

Wu *et al.*, “Online Regularization by Denoising with Applications to Phase Retrieval,” *Proc. ICCVW*, 2019

Wu *et al.*, “SIMBA: Scalable Inversion in Optical Tomography using Deep Denoising Priors,” arXiv:1911.13241, 2019



SIMBA is as good as full RARE

To conclude

- We increasingly rely on implicit regularization using nonlinear operators, such as deep neural nets
- RARE is a theoretically sound algorithm that combines the artifact removing CNN prior with the forward model
- SIMBA is a minibatch variant of RARE beneficial for certain large-scale problems such as tomography

Computational Imaging Group (CIG) at Washington University in St. Louis (WashU)

Washington University in St. Louis

CONTACT INFO

Prof. Ulugbek Kamilov
Email: kamilov@wustl.edu
Personal Twitter: @ukmlv

Web: http://cigroup.wustl.edu
Group Twitter: @wustlcig

ACKNOWLEDGEMENTS

Yu Sun	Prof. Hongyu An
Jiaming Liu	Prof. Lei Tian
Ray Wu	Dr. Brendt Wohlberg
Weijie Gan	Dr. Hassan Mansour
Xiaojian Xu	Prof. Michael Unser
...	...

Support:

NSF

CCF-181391

Scan this to connect on Twitter

Acknowledgements

- This material is based upon work supported by the National Science Foundation under Grant No. 1813910. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

- Research reported in this publication was supported by the Washington University Institute of Clinical and Translational Sciences grant UL1TR002345 from the National Center for Advancing Translational Sciences (NCATS) of the National Institutes of Health (NIH). The content is solely the responsibility of the authors and does not necessarily represent the official view of the NIH.