

Computational Microscopy: Advanced Priors beyond Optimization

Ulugbek Kamilov
Computational Imaging Group (CIG)

ICERM (Providence, RI) — 21 Mar 2019

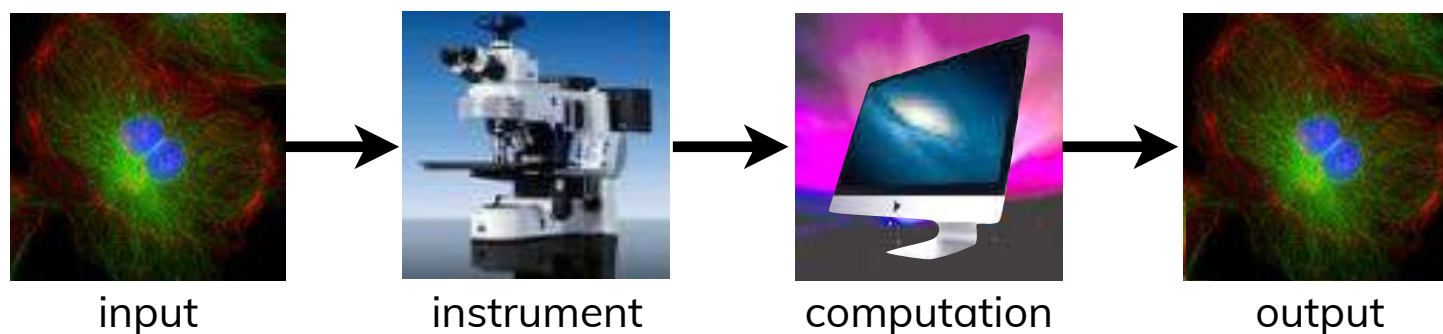
@wustlcig • cigroup.wustl.edu • kamilov@wustl.edu

Optical microscopy is going through a paradigm shift with computation at its core

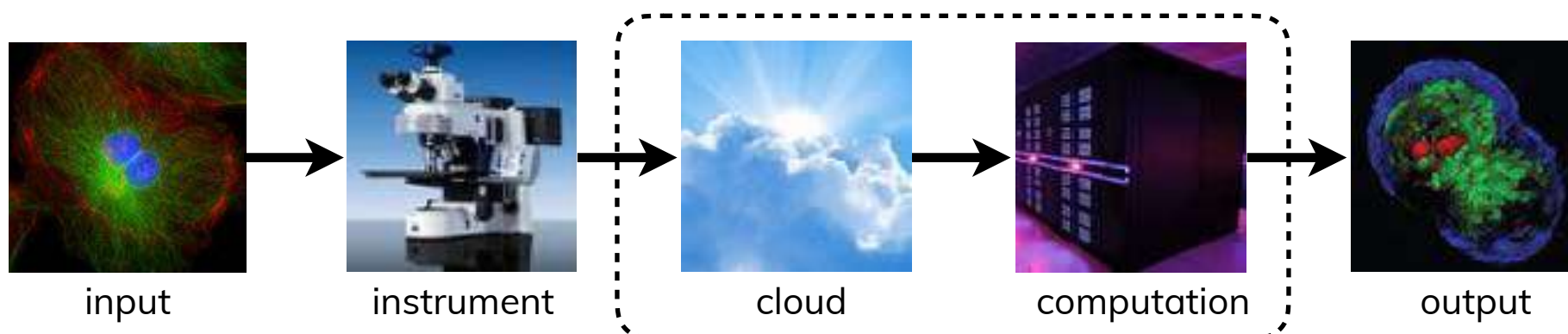
Past: Solely rely on optics for image formation



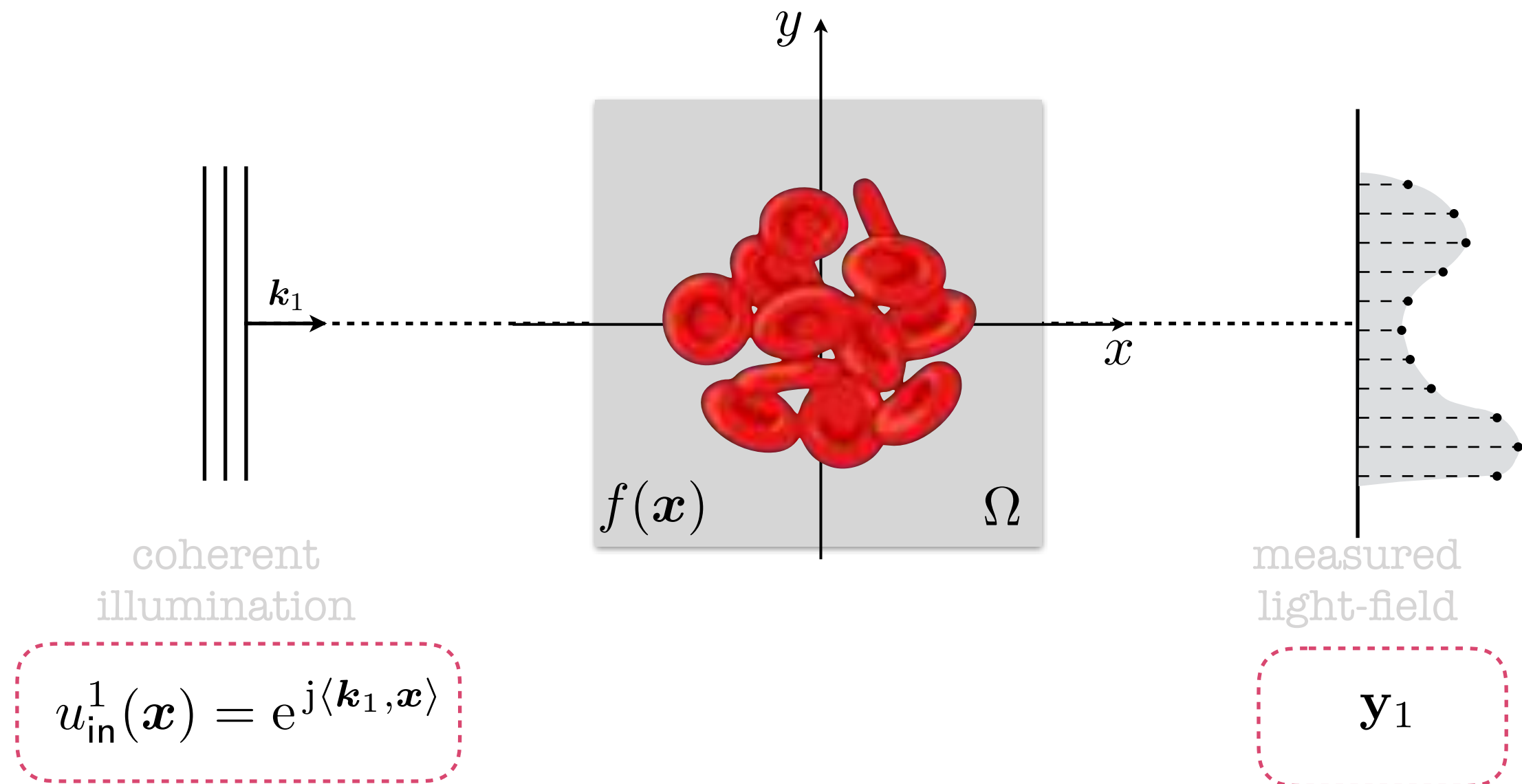
Present: Use signal processing for improved performance



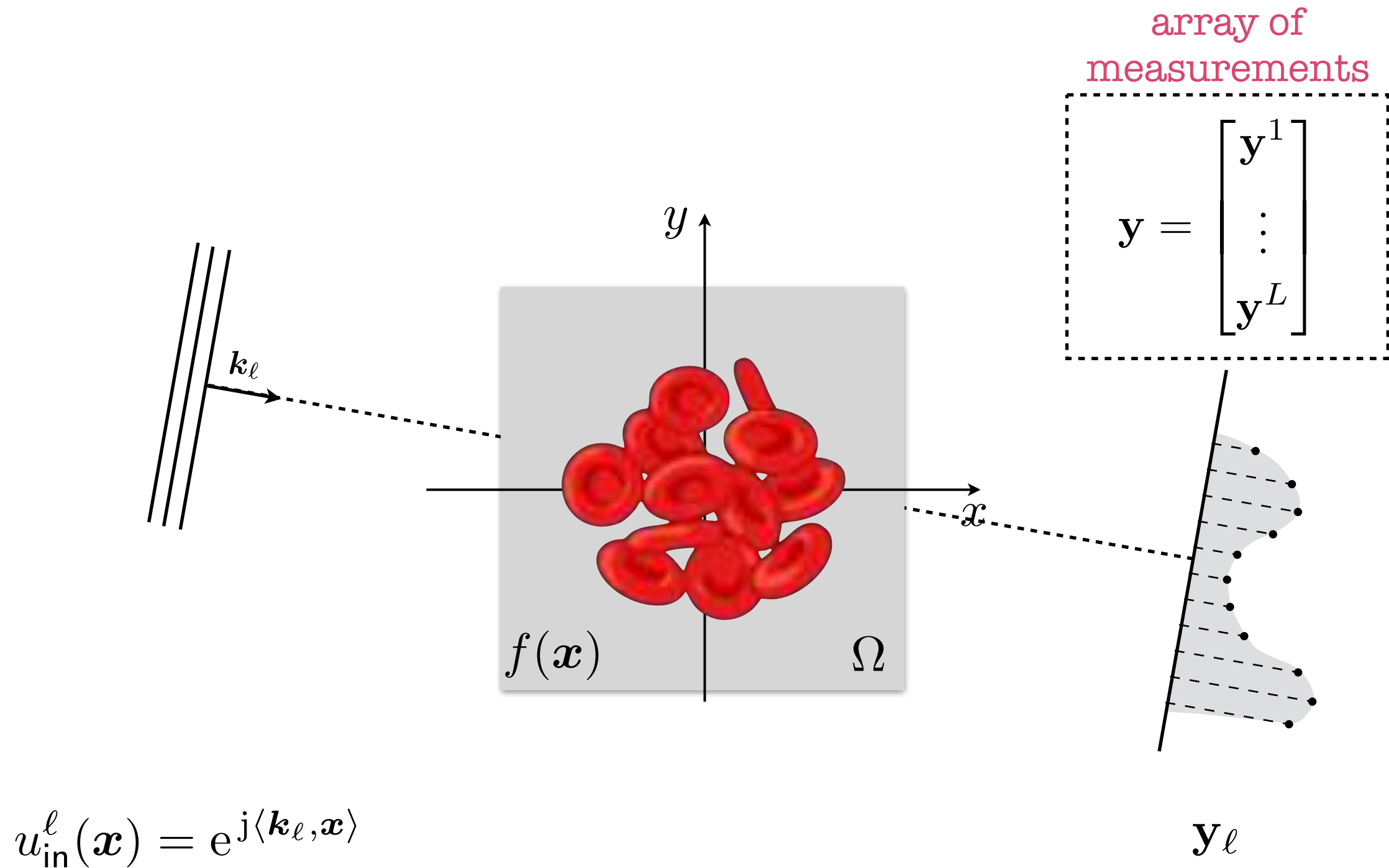
Future: Advanced inference for retrieving “hidden” information



Optical tomographic microscopy replaces x-rays with the visible light



Optical tomographic microscopy replaces x-rays with the visible light



Optical tomography is a powerful tool for live cell imaging

3D + Time:

Reveals internal cell structure across time

Quantitative and Label-free:

Relies on the refractive index as an intrinsic contrast

High-resolution and Non-ionizing:

Visible light spectrum (380-700 nm) is ideal for cell imaging

Optical tomography suffers from several critical limitations

Lengthy acquisition:

Needs 100s or 1000s of illuminations

Imaging artifacts:

Missing information and model mismatch

Sophisticated optics:

Holographic acquisition of phase limits applicability

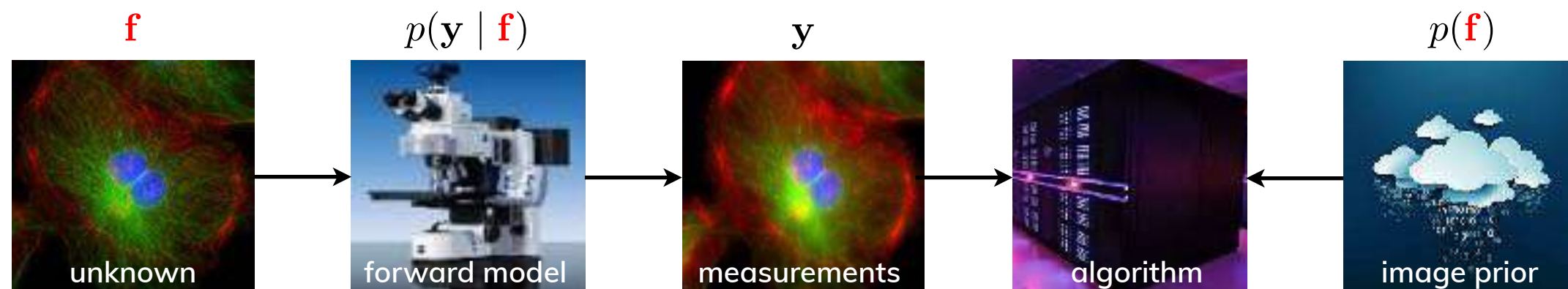
Goal: Overcome these limitations by leveraging advanced computational imaging

Forward model:
describes the physics of data acquisition

Image prior:
infuses domain-specific knowledge about the unknown image

Computational imaging to the rescue:

Can we use the very best computational tools to enable fast and accurate optical tomography?



Today we will talk about

- Accounting for nonlinearities in optical tomography
Going beyond linear inverse problems
- Fast online imaging using “plug-in” operators
Enforcing priors beyond traditional optimization
- Total variation for deep image prior (DIP-TV)
Using untrained CNNs as imaging priors

Today we will talk about

- **Accounting for nonlinearities in optical tomography**
Going beyond linear inverse problems
- Fast online imaging using “Plug-In” operators
Enforcing priors beyond traditional optimization
- Total variation for deep image prior (DIP-TV)
Using untrained CNNs as imaging priors

Absorption and scattering limit the ability of noninvasive imaging deep inside the tissue



light absorption and scattering
at different wavelengths

Absorption and scattering limit the ability of noninvasive imaging deep inside the tissue

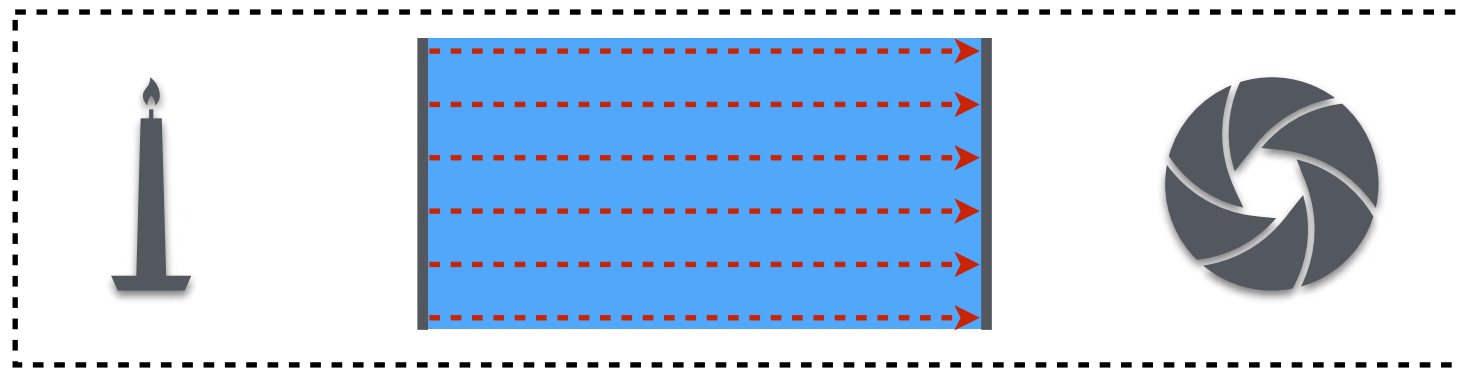
Scattering is the deflection of a propagating wave 'ray' from its original direction



homogeneous object

Absorption and scattering limit the ability of noninvasive imaging deep inside the tissue

Scattering is the deflection of a propagating wave 'ray' from its original direction

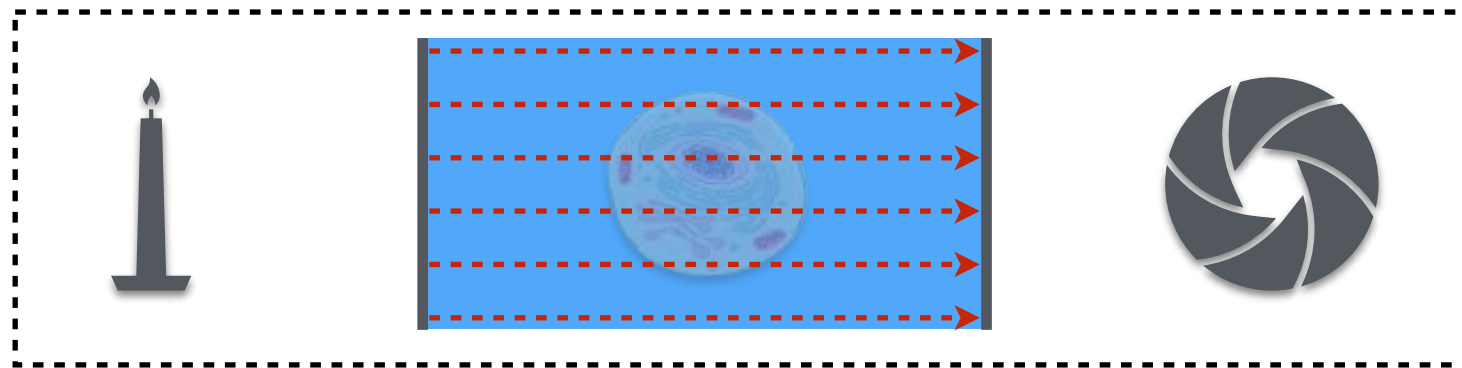


homogeneous object

No
scattering

Absorption and scattering limit the ability of noninvasive imaging deep inside the tissue

Scattering is the deflection of a propagating wave 'ray' from its original direction

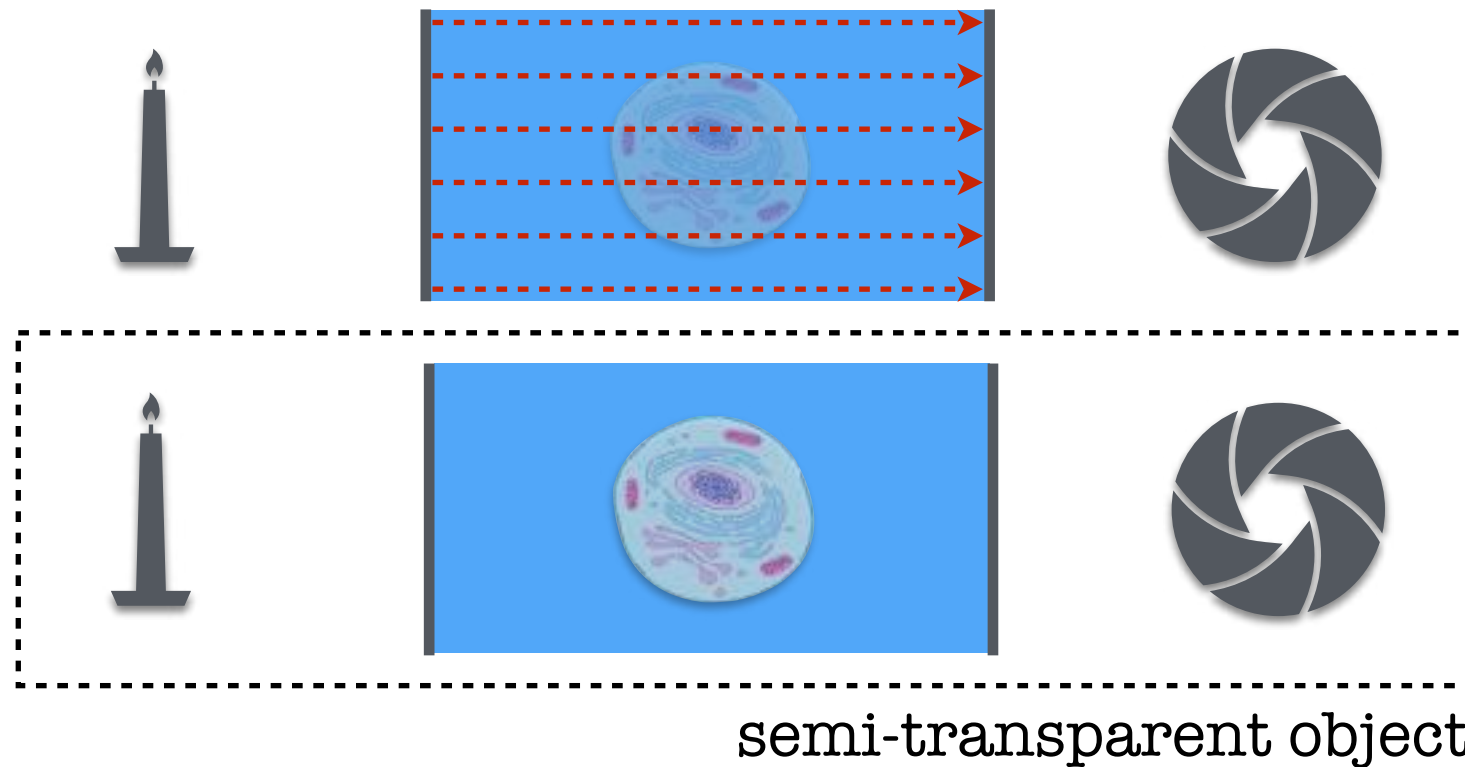


near-homogeneous object

No
scattering

Absorption and scattering limit the ability of noninvasive imaging deep inside the tissue

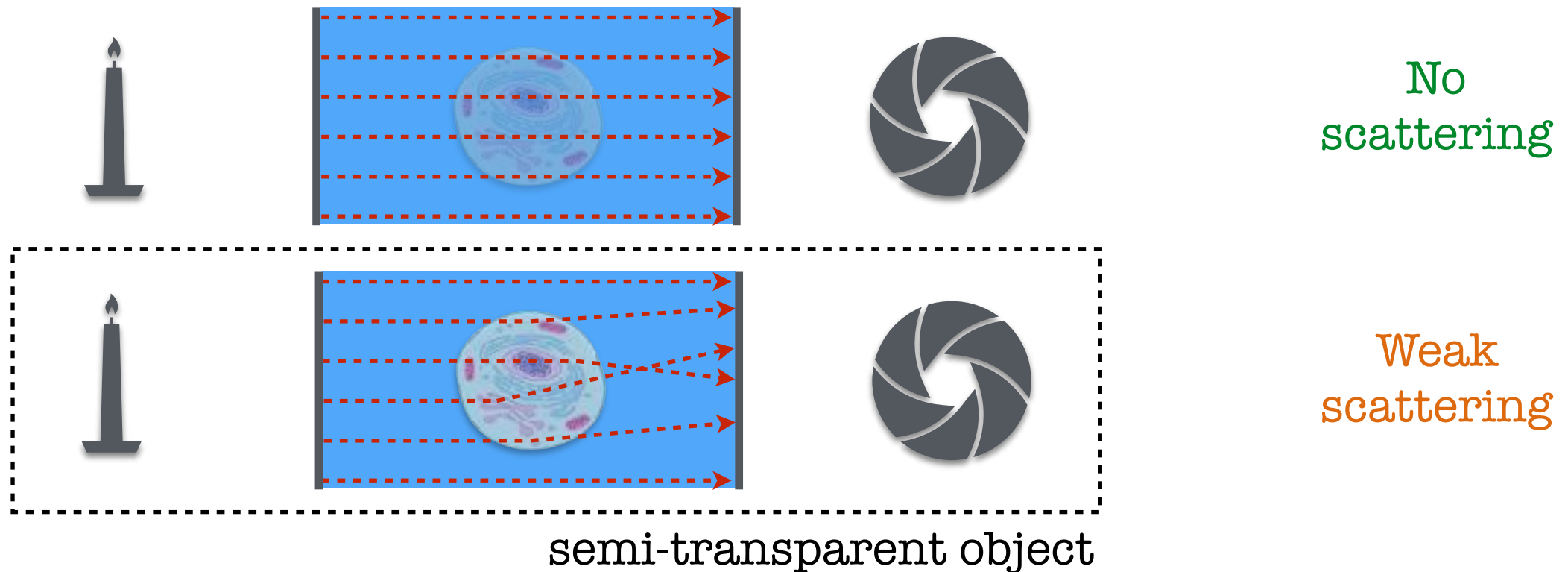
Scattering is the deflection of a propagating wave 'ray' from its original direction



No
scattering

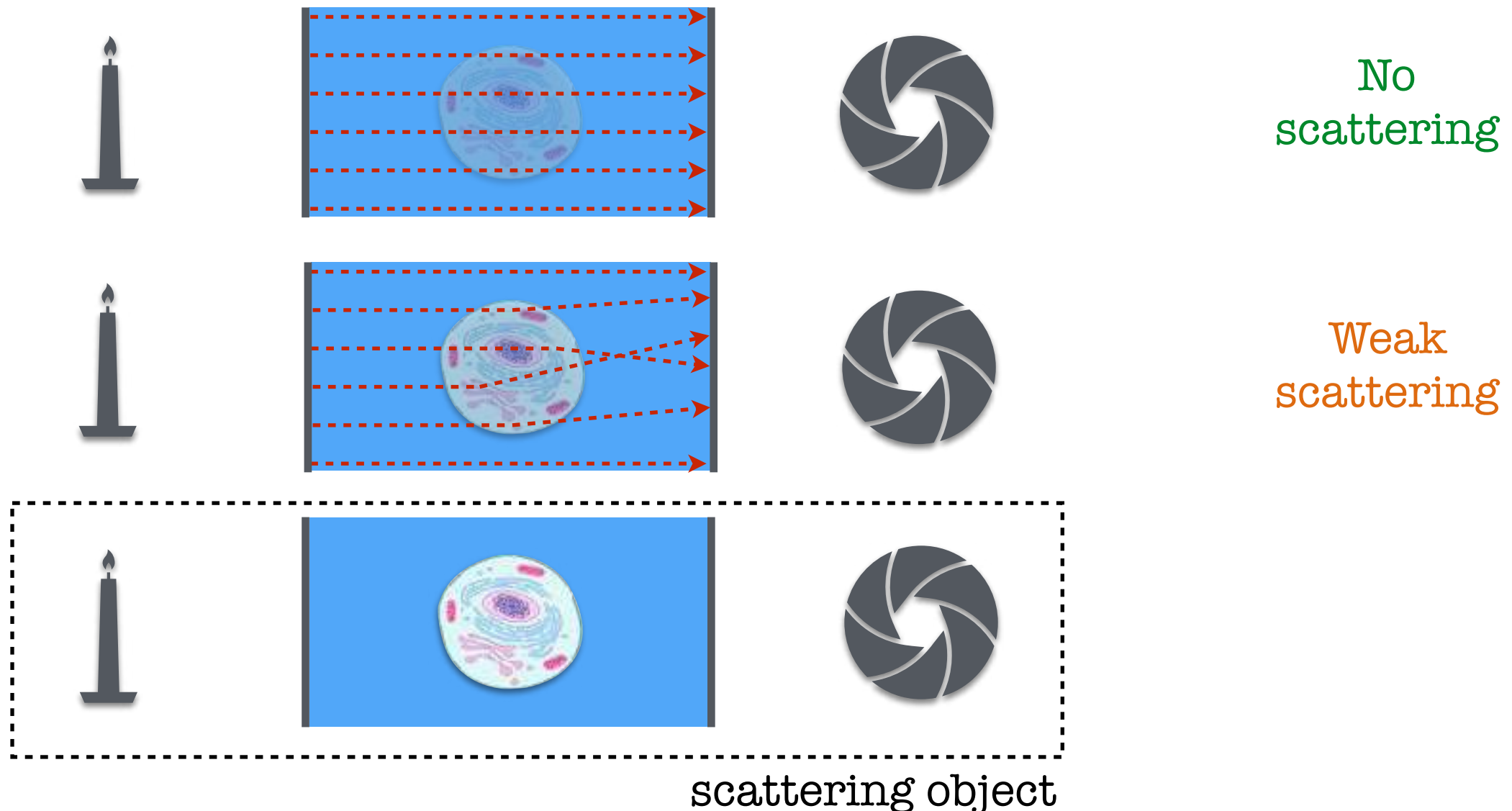
Absorption and scattering limit the ability of noninvasive imaging deep inside the tissue

Scattering is the deflection of a propagating wave 'ray' from its original direction



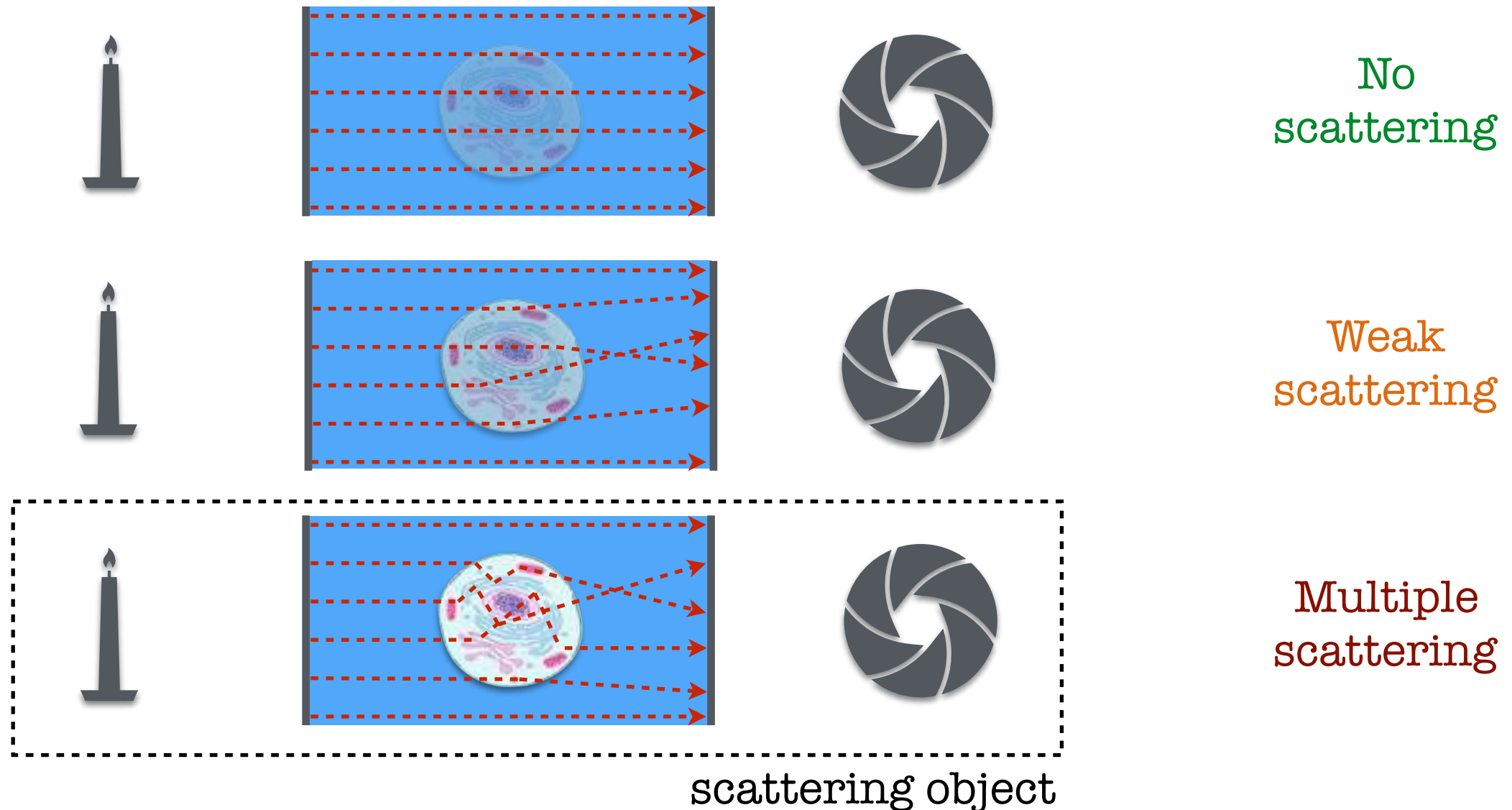
Absorption and scattering limit the ability of noninvasive imaging deep inside the tissue

Scattering is the deflection of a propagating wave 'ray' from its original direction



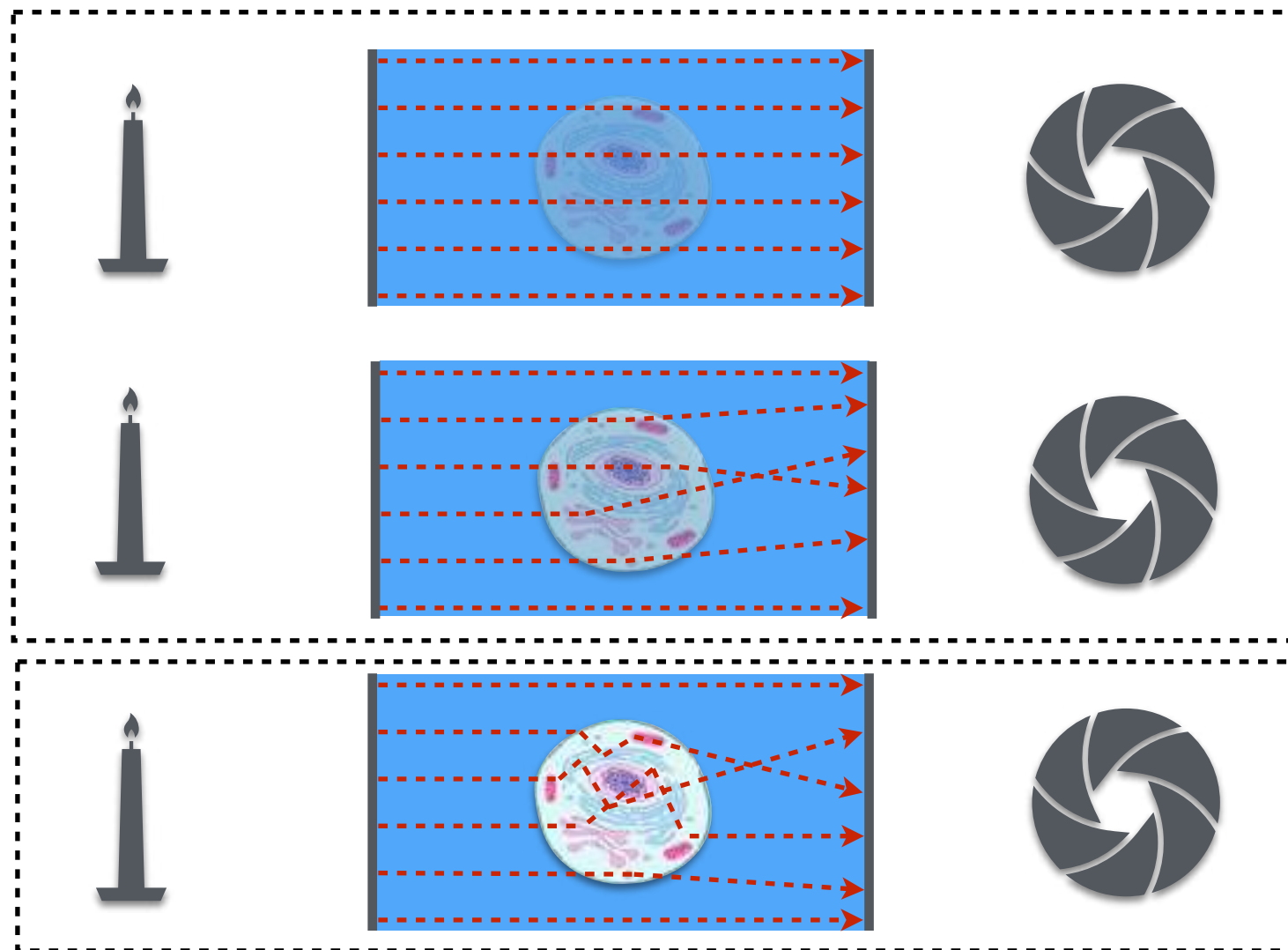
Absorption and scattering limit the ability of noninvasive imaging deep inside the tissue

Scattering is the deflection of a propagating wave 'ray' from its original direction



Absorption and scattering limit the ability of noninvasive imaging deep inside the tissue

Scattering is the deflection of a propagating wave 'ray' from its original direction



- linear measurements
- convex optimization
- fast algorithms

- complicated models
- nonconvex optimization
- hard to analyze

Absorption and scattering limit the ability of noninvasive imaging deep inside the tissue

Scattering is the deflection of a propagating wave 'ray' from its original direction

Scattering limits conventional imaging systems to superficial layers of an object



Optical tomography is traditionally simplified to a linear forward model

The Helmholtz equation for modeling object-light interactions

$$(\Delta + k_b^2 \mathbf{I})u_{\text{sc}}(\mathbf{x}) = f(\mathbf{x})u(\mathbf{x}) \quad u(\mathbf{x}) = u_{\text{in}}(\mathbf{x}) + u_{\text{sc}}(\mathbf{x})$$

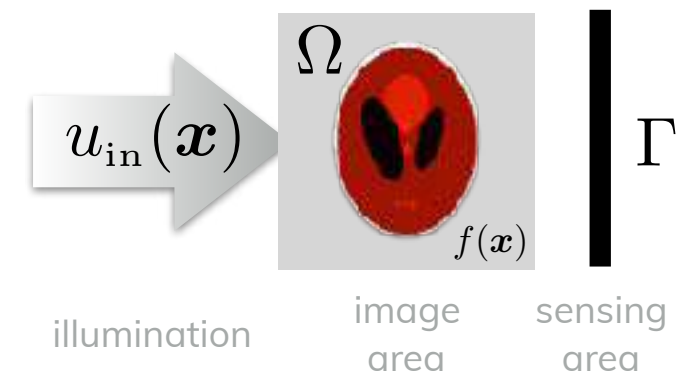
The Domain-integral formulation with the Green's function

$$u_{\text{sc}}(\mathbf{x}) = \int_{\Omega} g(\mathbf{x} - \mathbf{x}') f(\mathbf{x}') u(\mathbf{x}') d\mathbf{x} \quad g(\mathbf{x}) \triangleq \frac{e^{jk_b \|\mathbf{x}\|}}{4\pi \|\mathbf{x}\|}$$

The first Born approximation linearizes the model

$$u_{\text{sc}}(\mathbf{x}) \approx H_b\{f\}(\mathbf{x}) = \int_{\Omega} g(\mathbf{x} - \mathbf{x}') f(\mathbf{x}') u_{\text{in}}(\mathbf{x}') d\mathbf{x}$$

(by ignoring multiple scattering)



Linearized scattering model leads to the Fourier diffraction theorem

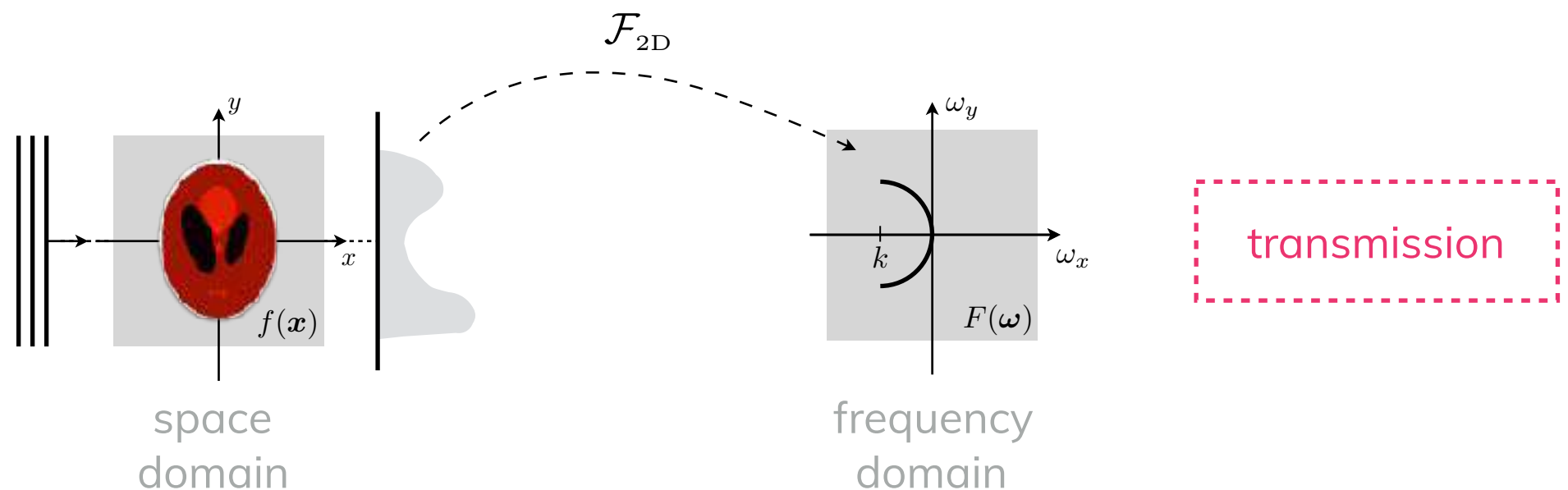
Assume a weakly scattering (i.e., quite transparent) object

$$|u_{\text{sc}}(\mathbf{x})| \ll |u_{\text{in}}(\mathbf{x})|$$

This leads to the Fourier diffraction theorem

$$\mathbf{y} = S_k \{ \mathcal{F}_{3D} \{ f(\mathbf{x}) \} \}$$

subsampling in
Fourier space



Linearized scattering model leads to the Fourier diffraction theorem

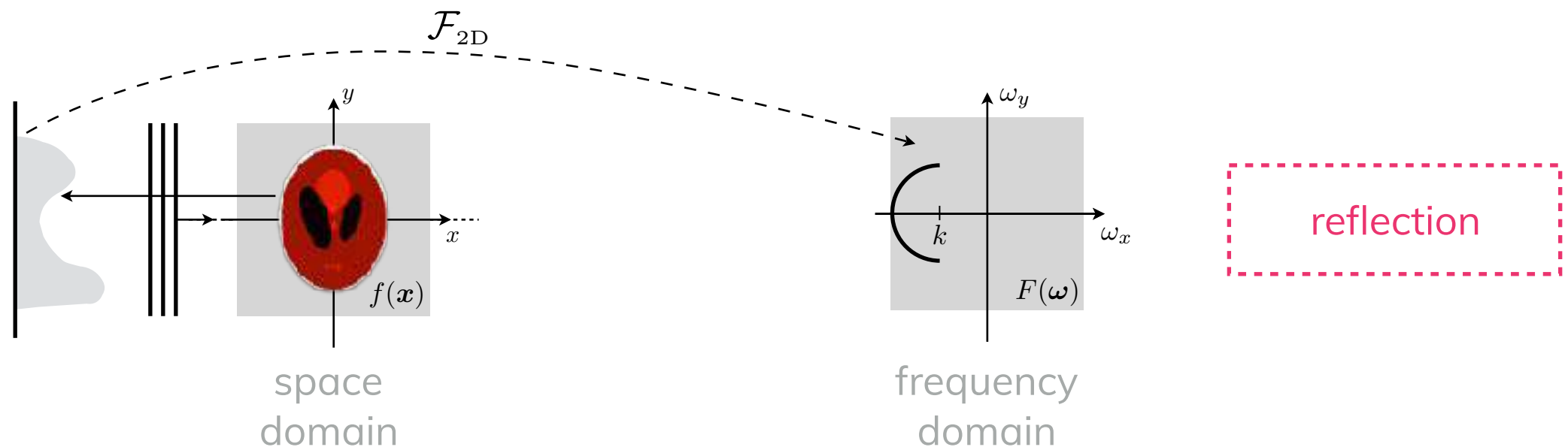
Assume a weakly scattering (i.e., quite transparent) object

$$|u_{\text{sc}}(\mathbf{x})| \ll |u_{\text{in}}(\mathbf{x})|$$

This leads to the Fourier diffraction theorem

$$\mathbf{y} = S_k \{ \mathcal{F}_{3D} \{ f(\mathbf{x}) \} \}$$

subsampling in
Fourier space



Linearized scattering model leads to the Fourier diffraction theorem

Assume a weakly scattering (i.e., quite transparent) object

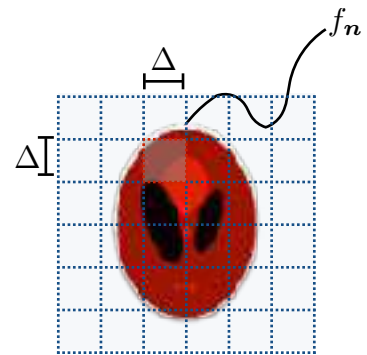
$$|u_{\text{sc}}(\mathbf{x})| \ll |u_{\text{in}}(\mathbf{x})|$$

This leads to the Fourier diffraction theorem

$$\mathbf{y} = S_k \{ \mathcal{F}_{3D} \{ f(\mathbf{x}) \} \}$$

Discretize by approximating the object with its samples

$$f(\mathbf{x}) \approx \sum_{\mathbf{n} \in \Omega} f_{\mathbf{n}} \delta(\mathbf{x} - \mathbf{n}\Delta) \quad f_{\mathbf{n}} = f(\mathbf{x})|_{\mathbf{x}=\mathbf{n}\Delta}$$



Linearized scattering model leads to the Fourier diffraction theorem

Assume a weakly scattering (i.e., quite transparent) object

$$|u_{\text{sc}}(\boldsymbol{x})| \ll |u_{\text{in}}(\boldsymbol{x})|$$

This leads to the Fourier diffraction theorem

$$\mathbf{y} = S_{\mathbf{k}} \{ \mathcal{F}_{3\text{D}} \{ f(\boldsymbol{x}) \} \}$$

Discretize by approximating the object with its samples

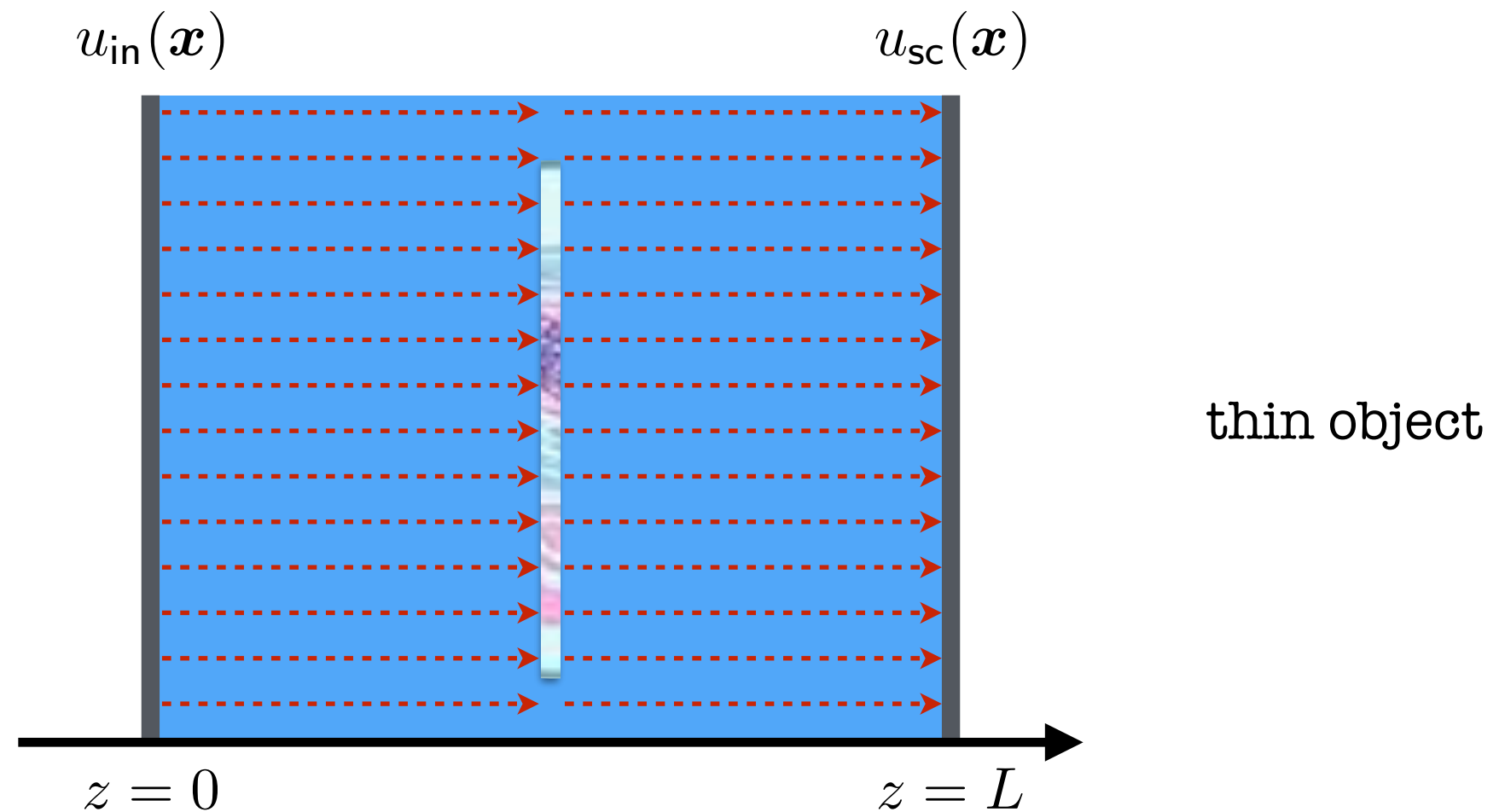
$$f(\boldsymbol{x}) \approx \sum_{\boldsymbol{n} \in \Omega} f_{\boldsymbol{n}} \delta(\boldsymbol{x} - \boldsymbol{n}\Delta)$$

Thus, we obtain a linear inverse problem:

$$\mathbf{y} = \mathbf{H}\mathbf{f} + \mathbf{e}$$

Beam propagation method (BPM)

efficiently models forward multiple scattering



$$u_1(\mathbf{x}) = (\phi_{L/2} * u_{\text{in}})(\mathbf{x})$$

convolution

$$u_2(\mathbf{x}) = u_1(\mathbf{x}) \cdot o_{L/2}(\mathbf{x})$$

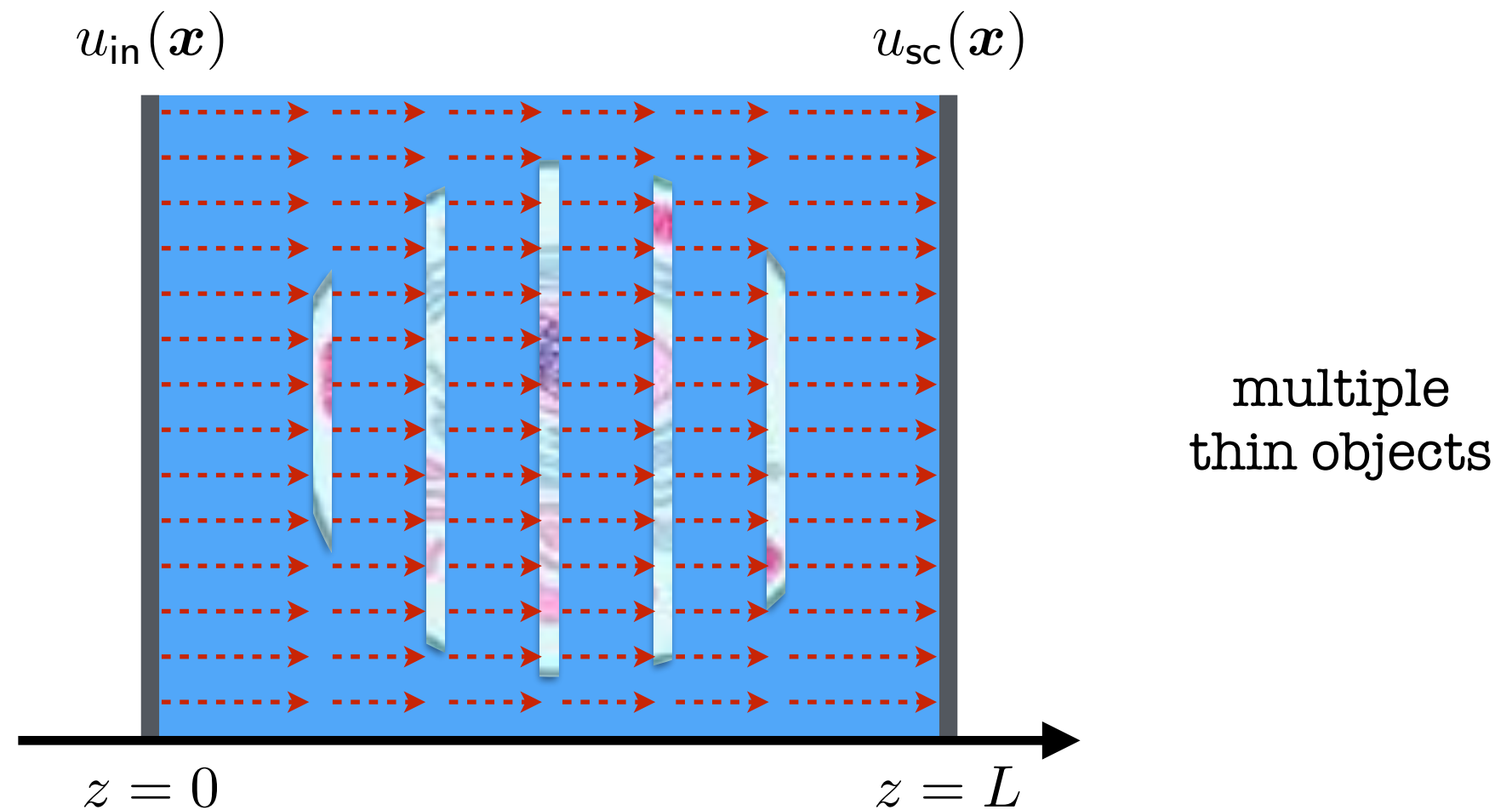
phase-shift

$$u_{\text{sc}}(\mathbf{x}) = (\phi_{L/2} * u_2)(\mathbf{x})$$

convolution

Beam propagation method (BPM)

efficiently models forward multiple scattering

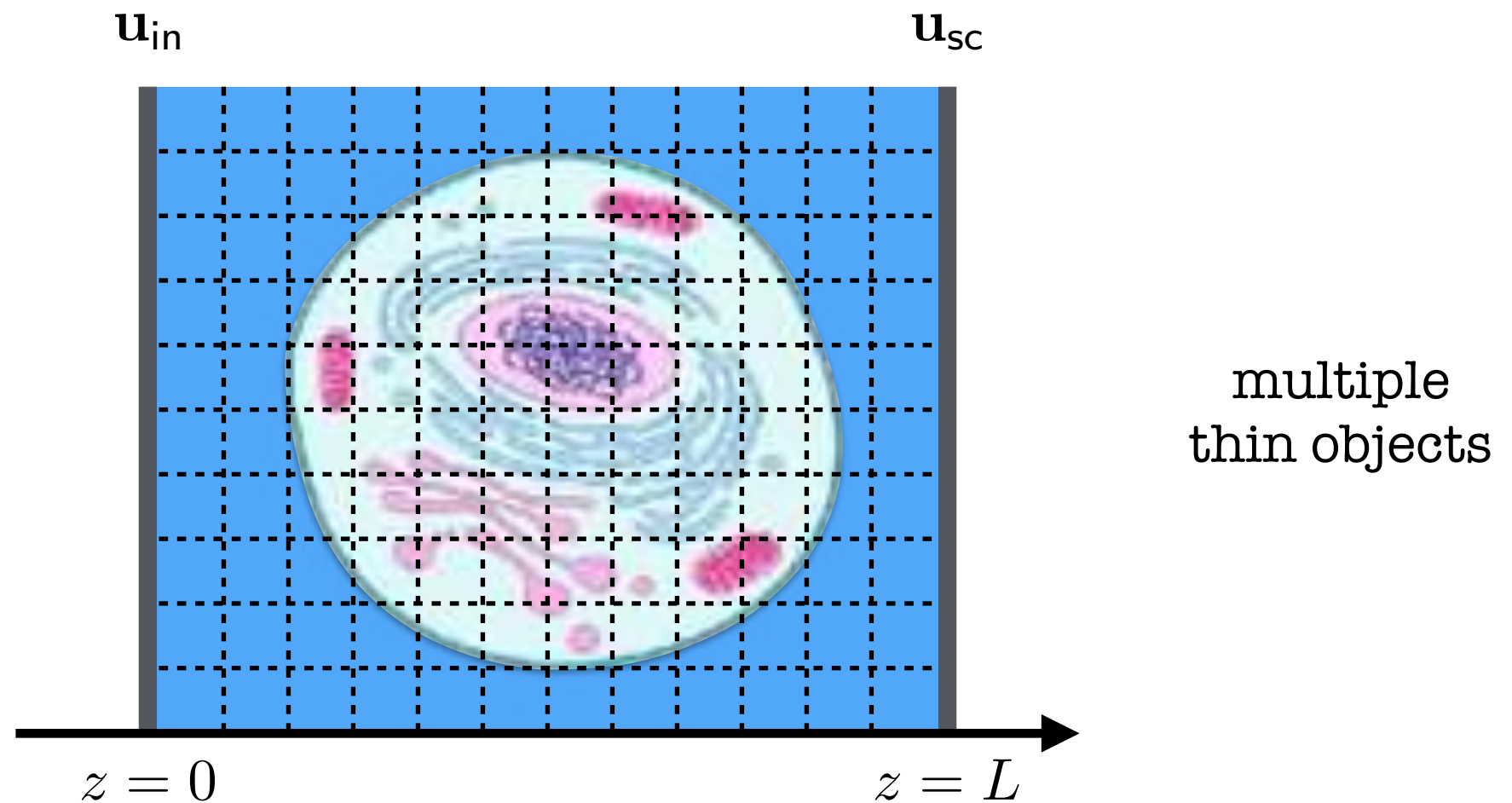


$$u_k(\mathbf{x}) = o_k(\mathbf{x}) \cdot (\phi_\Delta * u_{k-1})(\mathbf{x}) \quad k = 1, \dots, K$$

recursive structure

Beam propagation method (BPM)

efficiently models forward multiple scattering

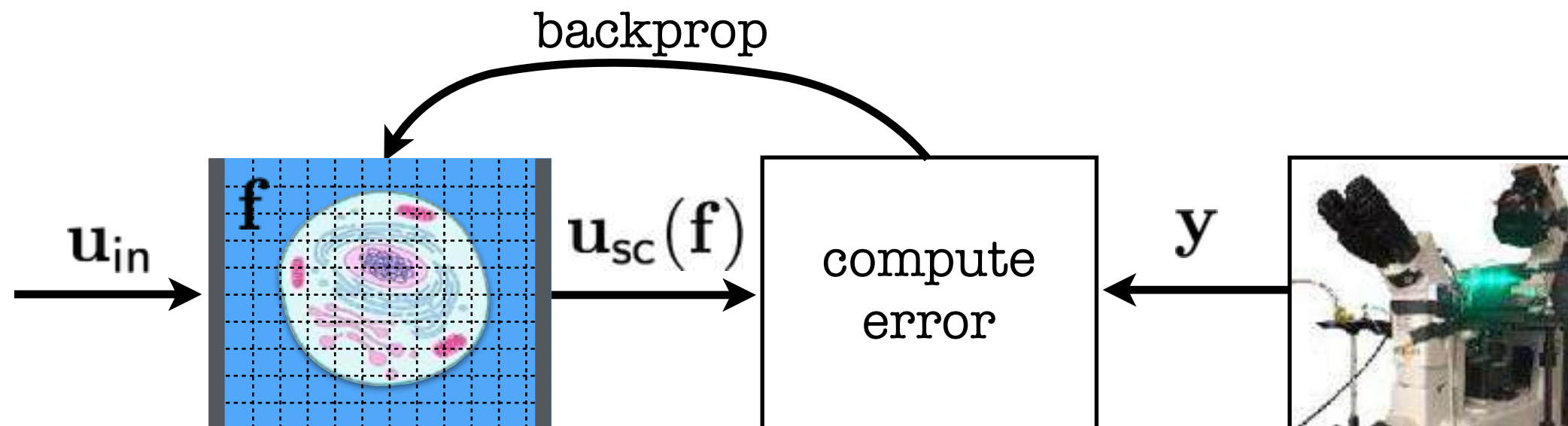


$$\mathbf{u}_k = \mathbf{o}_k \cdot (\phi * \mathbf{u}_{k-1}) \quad k = 1, \dots, K$$

recursive structure

Image formation under BPM is analogous to the training of convolutional neural nets (CNNs)

- 1) Initialize object
- 2) Illuminate object and measure the scattered field
- 3) Run forward BPM propagation
- 4) Run BPM error back-propagation to obtain the gradient
- 5) Update the image
- 6) Return object after convergence



FISTA and ADMM are two popular algorithms for large-scale and nonsmooth optimization

Consider a minimization problem

$$\min_{\mathbf{f}} \left\{ \mathcal{C}(\mathbf{f}) \triangleq \mathcal{D}(\mathbf{f}) + \mathcal{R}(\mathbf{f}) \right\} \quad \mathcal{D}(\mathbf{f}) \triangleq \frac{1}{2} \|\mathbf{y} - |\mathbf{u}_{\text{sc}}(\mathbf{f})|^2\|_{\ell_2}^2$$

Define the proximal operator for avoiding differentiating the regularizer

$$\text{prox}_{\lambda\mathcal{R}}(\mathbf{y}) \triangleq \arg \min_{\mathbf{f}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{f}\|_{\ell_2}^2 + \lambda\mathcal{R}(\mathbf{f}) \right\}$$

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

$$\begin{aligned} \mathbf{z}^k &\leftarrow \mathbf{s}^{k-1} - \gamma \nabla \mathcal{D}(\mathbf{s}^{k-1}) \\ \mathbf{f}^k &\leftarrow \text{prox}_{\gamma\mathcal{R}}(\mathbf{z}^k) \\ \mathbf{s}^k &\leftarrow \mathbf{f}^k + ((q_{k-1} - 1)/q_k)(\mathbf{f}^k - \mathbf{f}^{k-1}) \end{aligned}$$

ISTA: $q_k = 1$ /FISTA: specific q_k

$$\begin{aligned} \mathbf{z}^k &\leftarrow \text{prox}_{\gamma\mathcal{D}}(\mathbf{f}^{k-1} - \mathbf{s}^{k-1}) \\ \mathbf{f}^k &\leftarrow \text{prox}_{\gamma\mathcal{R}}(\mathbf{z}^k + \mathbf{s}^{k-1}) \\ \mathbf{s}^k &\leftarrow \mathbf{s}^{k-1} + (\mathbf{z}^k - \mathbf{f}^k) \end{aligned}$$

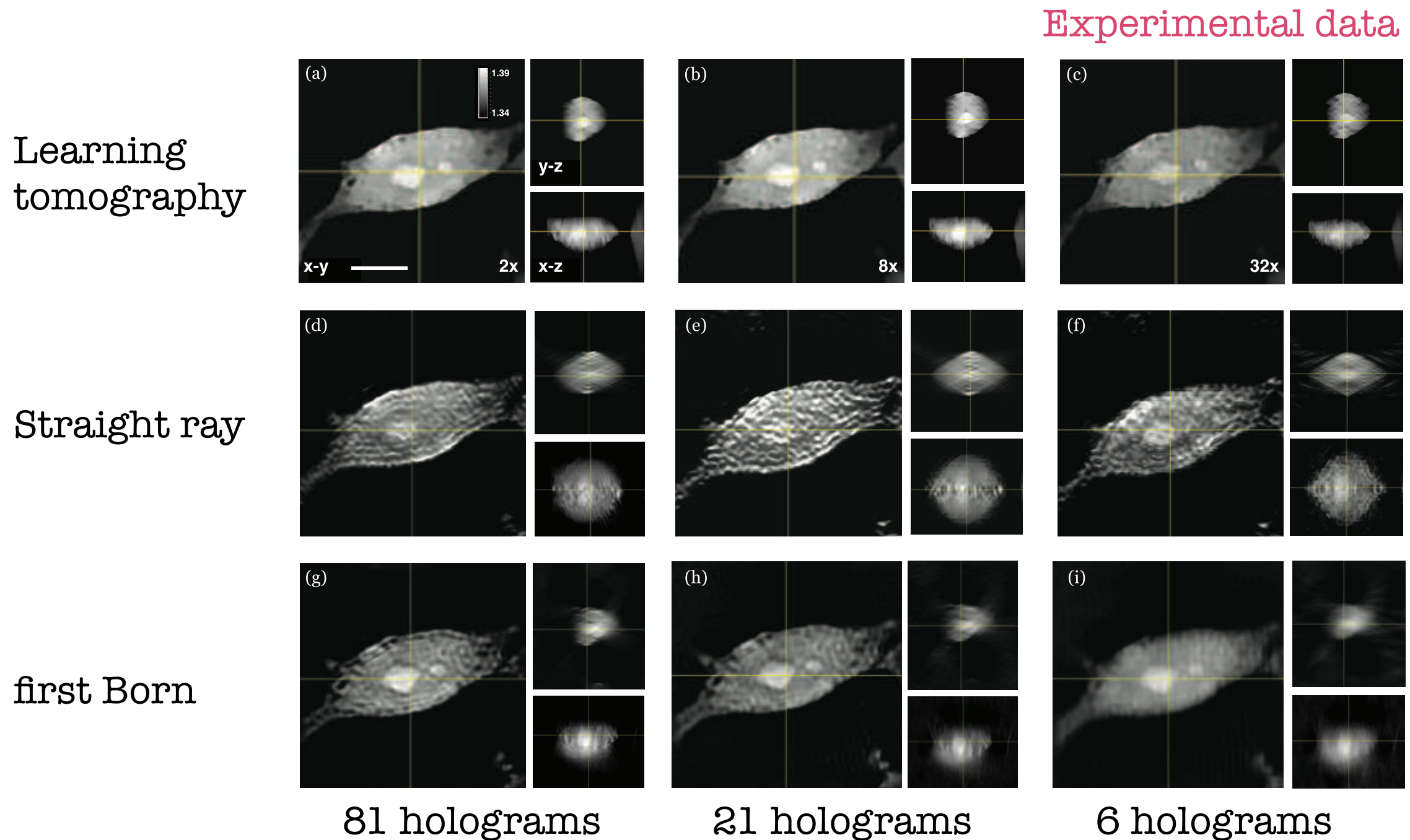
ADMM fast practical convergence

Our regularized BPM framework was extensively validated on 3D optical tomography

$$\min_{\mathbf{f}} \left\{ \frac{1}{2L} \sum_{\ell=1}^L \|\mathbf{y}_{\ell} - \mathbf{u}_{\text{sc}}^{\ell}(\mathbf{f})\|_{\ell_2}^2 + \lambda \sum_{n=1}^N \|[\mathbf{D}\mathbf{f}]_n\|_{\ell_2} \right\}$$

Fit to L illuminations + isotropic 3D-TV prior

Our regularized BPM framework was extensively validated on 3D optical tomography



Today we will talk about

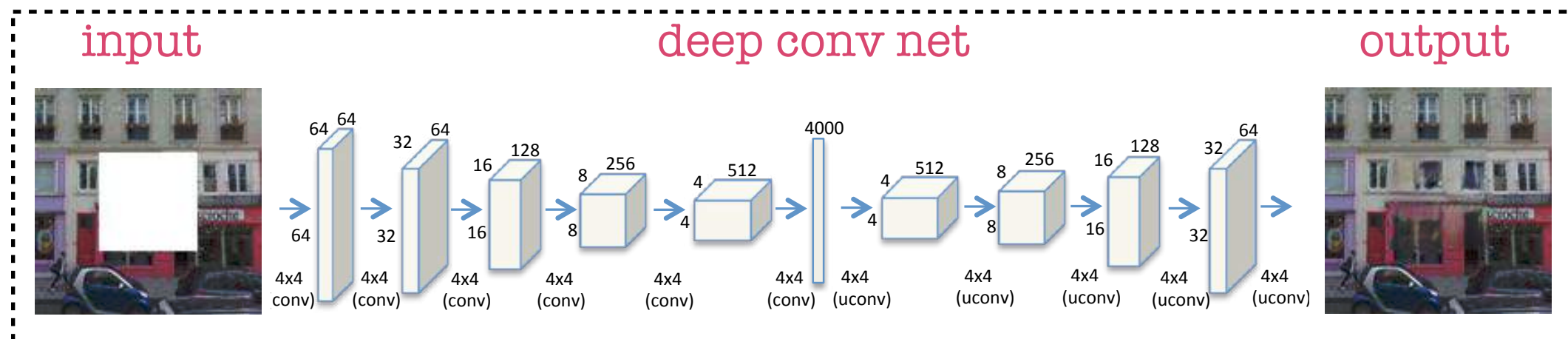
- Accounting for nonlinearities in optical tomography
Going beyond linear inverse problems
- **Fast online imaging using “plug-in” operators**
Enforcing priors beyond traditional optimization
- Total variation for deep image prior (DIP-TV)
Using untrained CNNs as imaging priors

Can we use semantic priors for improving image formation?

The recent interest in sparse recovery highlighted the importance of structural priors in image formation

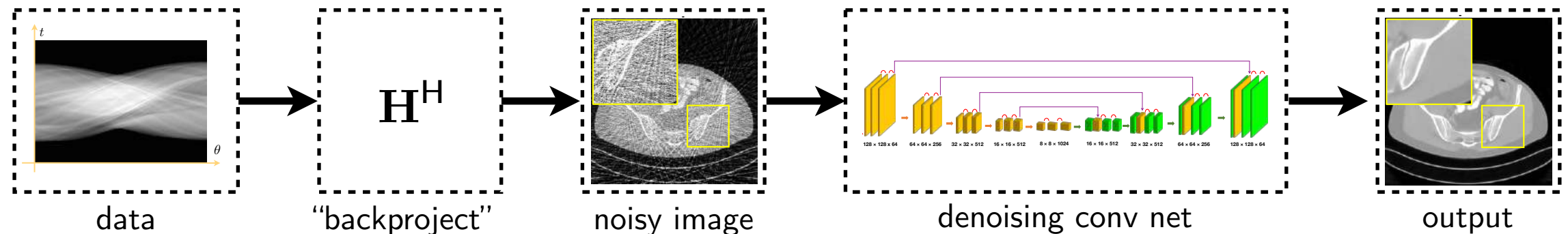
How can we create priors beyond simple constraints (for example: we know that we are looking at cells)?

Deep neural nets provide a powerful tool for representing and enforcing sophisticated image priors



A well established deep learning pipeline: first backproject then denoise with a conv net

Data processing pipeline

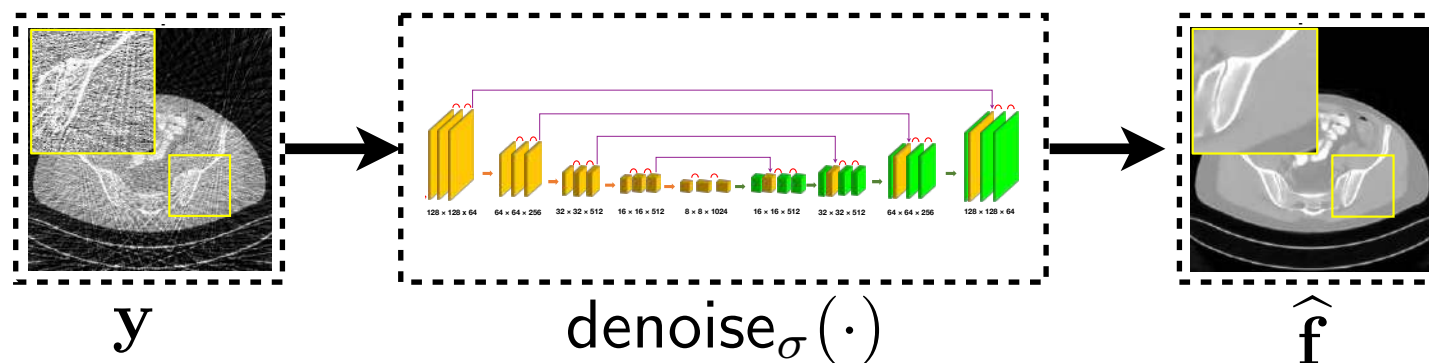


Question: What are some of the key limitations of this approach?

- 1) Implicit dependance of the conv net on the forward model
- 2) Consistency with the measured data is unclear
- 3) Needs a sufficiently good starting point to denoise

Treating a denoiser as a proximal operator allows to separate the prior from the forward model

Build a denoiser at various noise levels



Use the denoiser as a Plug-and-Play Prior (PnP)

$$\begin{aligned} \mathbf{z}^k &\leftarrow \text{prox}_{\gamma\mathcal{D}}(\mathbf{f}^{k-1} - \mathbf{s}^{k-1}) \\ \mathbf{f}^k &\leftarrow \text{denoise}_\sigma(\mathbf{z}^k + \mathbf{s}^{k-1}) \\ \mathbf{s}^k &\leftarrow \mathbf{s}^{k-1} + (\mathbf{z}^k - \mathbf{f}^k) \end{aligned}$$

PnP-ADMM

$$\begin{aligned} \mathbf{z}^k &\leftarrow \mathbf{s}^{k-1} - \gamma \nabla \mathcal{D}(\mathbf{s}^{k-1}) \\ \mathbf{f}^k &\leftarrow \text{denoise}_\sigma(\mathbf{z}^k) \\ \mathbf{s}^k &\leftarrow \mathbf{f}^k + ((q_{k-1} - 1)/q_k)(\mathbf{f}^k - \mathbf{f}^{k-1}) \end{aligned}$$

PnP-FISTA

Plug-and-play priors (PnP) approach has been shown to yield state-of-the-art results

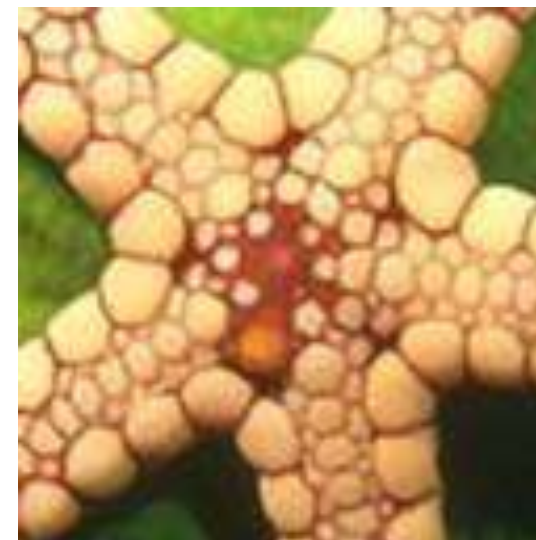
Method	Average PSNR (dB) over 10 images
TV	29.22
IDD-BM3D	30.92
ASDS-Reg	30.11
NCSR	31.09
PnP	31.33



(a) Ground Truth



(b) Input 20.83dB



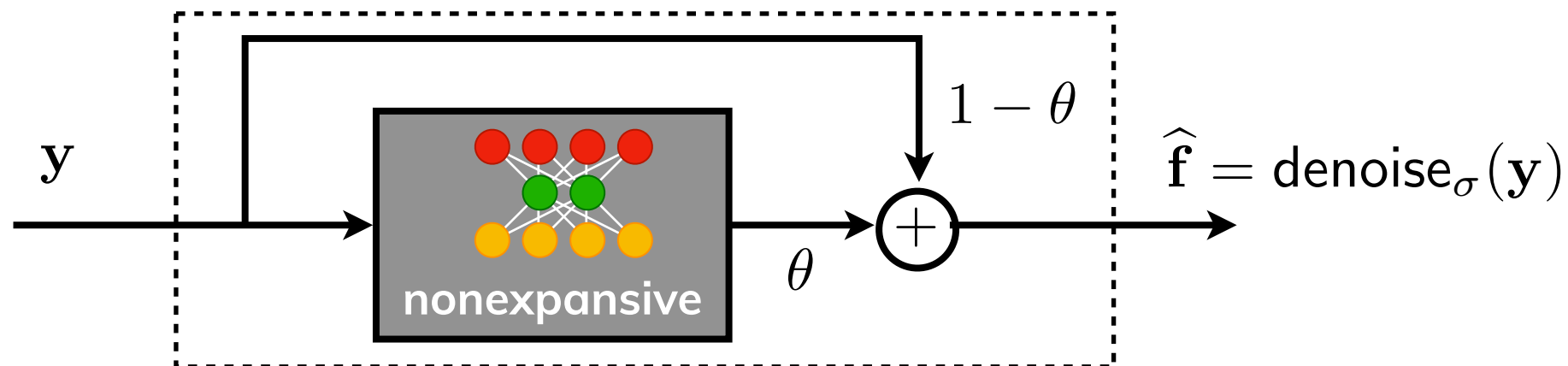
(d) NCSR 28.39dB



(e) P^3 -TNRD 28.43dB

We prove using monotone operator theory that PnP-ISTA converges for averaged denoisers

An averaged deep conv net is straightforward to build

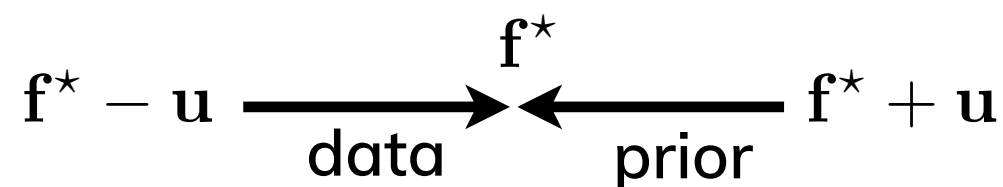


For a convex data-fidelity, the iterates of PnP-ISTA satisfy

$$\|\mathbf{f}^t - \mathbf{P}(\mathbf{f}^t)\|^2 \leq \frac{2}{t} \left(\frac{1 + \theta}{1 - \theta} \right) \|\mathbf{f}^0 - \mathbf{f}^*\|^2 \quad \mathbf{P}(\mathbf{f}) = \text{denoise}_\sigma(\mathbf{f} - \gamma \nabla \mathcal{D}(\mathbf{f}))$$

where the fixed point satisfies the **consensus equilibrium (CE)**

$$\begin{aligned} \mathbf{f}^* &= \text{prox}_{\gamma \mathcal{D}}(\mathbf{f}^* - \mathbf{u}) \\ \mathbf{f}^* &= \text{denoise}_\sigma(\mathbf{f}^* + \mathbf{u}) \end{aligned}$$



Our analysis extends recent results on the convergence of PnP schemes

[Sreehari *et al.*]: When $\mathcal{D}(\cdot)$ is convex and $\nabla \text{denoise}_\sigma(\cdot)$ is a symmetric matrix with eigenvalues in $[0, 1]$, then $\text{denoise}_\sigma(\cdot)$ is a proximal operator.

Denoiser is an implicit proximal operator

[Chan *et al.*]: When both $\nabla \mathcal{D}(\cdot)$ and $\text{denoise}_\sigma(\cdot)$ are bounded operators, PnP-ADMM with a quadratic parameter update scheme converges to a fixed point.

Unfortunately no convergence rate
PnP-ISTA can diverges for bounded operators!

PnP-SPGM accelerates image formation in optical tomography with many measurements

In reality, the data-fidelity has the following form

$$\mathcal{D}(\mathbf{f}) = \frac{1}{2L} \sum_{\ell=1}^L \|\mathbf{y}_{\ell} - \mathbf{u}_{\text{sc}}^{\ell}(\mathbf{f})\|_{\ell_2}^2 \Rightarrow \nabla \mathcal{D}(\mathbf{f}) = \frac{1}{L} \sum_{\ell=1}^L \left[\frac{\partial}{\partial \mathbf{f}} \mathbf{u}_{\text{sc}}^{\ell}(\mathbf{f}) \right] (\mathbf{u}_{\text{sc}}^{\ell}(\mathbf{f}) - \mathbf{y}_{\ell})$$

PnP-SPGM can accelerate image formation by reducing per-iteration cost (and also parallelizing the algorithm)

$$\hat{\nabla} \mathcal{D}(\mathbf{s}^{k-1}) \leftarrow \text{minibatchGradient}(\mathbf{s}^{k-1}, B)$$

$$\mathbf{z}^k \leftarrow \mathbf{s}^{k-1} - \gamma \hat{\nabla} \mathcal{D}(\mathbf{s}^{k-1})$$

$$\mathbf{f}^k \leftarrow \text{denoise}_{\sigma}(\mathbf{z}^k)$$

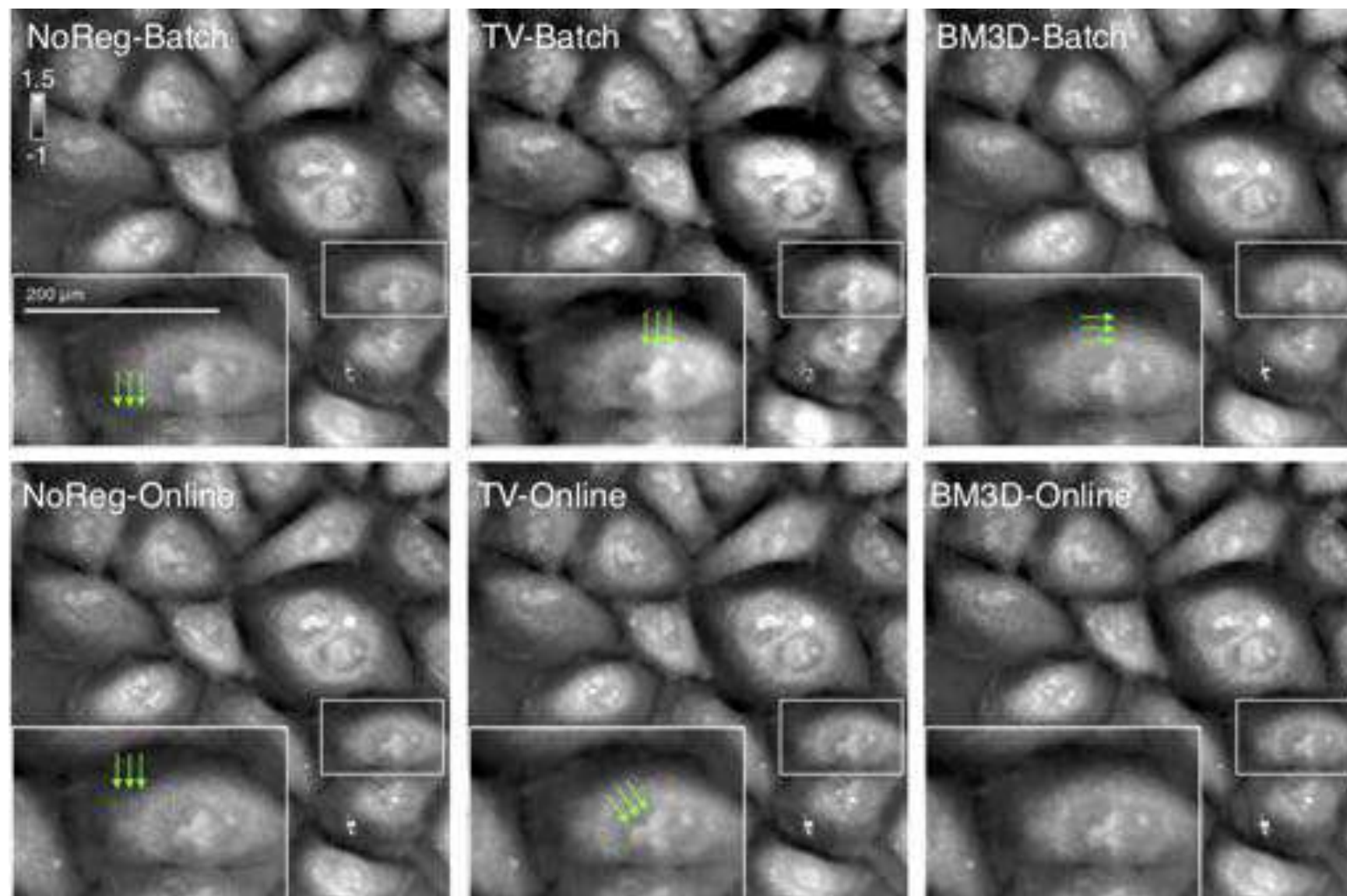
$$\mathbf{s}^k \leftarrow \mathbf{f}^k + ((q_{k-1} - 1)/q_k)(\mathbf{f}^k - \mathbf{f}^{k-1})$$

use only $B \ll L$
measurements per iteration

Converges to the same
solution as PnP-ISTA*

When the number of measurements is large, PnP-SPGM converges faster than batch algorithms

Experimental FPM data



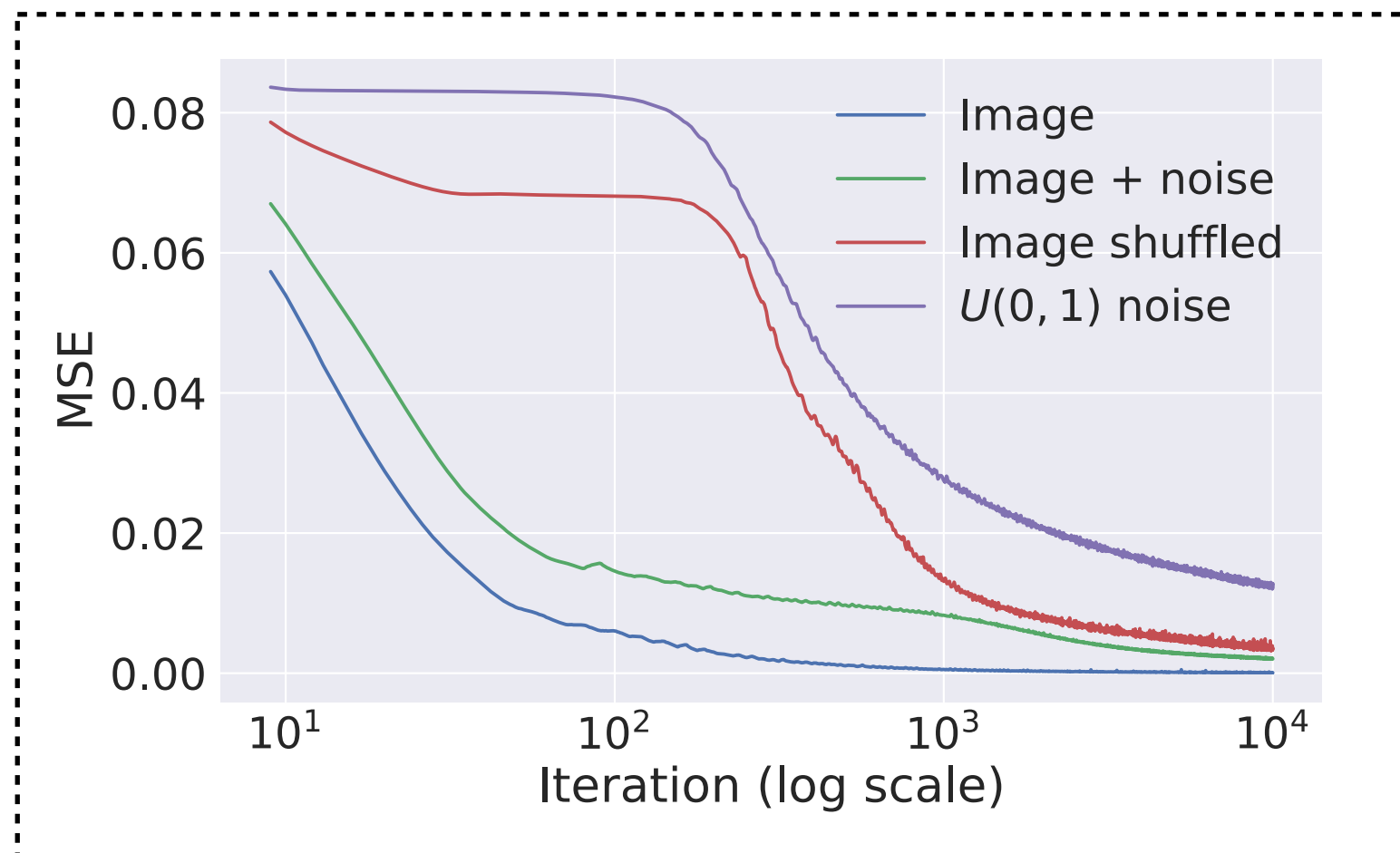
Using 60 (out of total 293) illuminations per iteration

Today we will talk about

- Accounting for nonlinearities in optical tomography
Going beyond linear inverse problems
- Fast online imaging using “Plug-In” operators
Enforcing priors beyond traditional optimization
- **Total variation for deep image prior (DIP-TV)**
Using untrained CNNs as imaging priors

Does the excellent performance of conv nets exclusively come from from learning?

A deep conv net fits more easily to natural images compared to noise

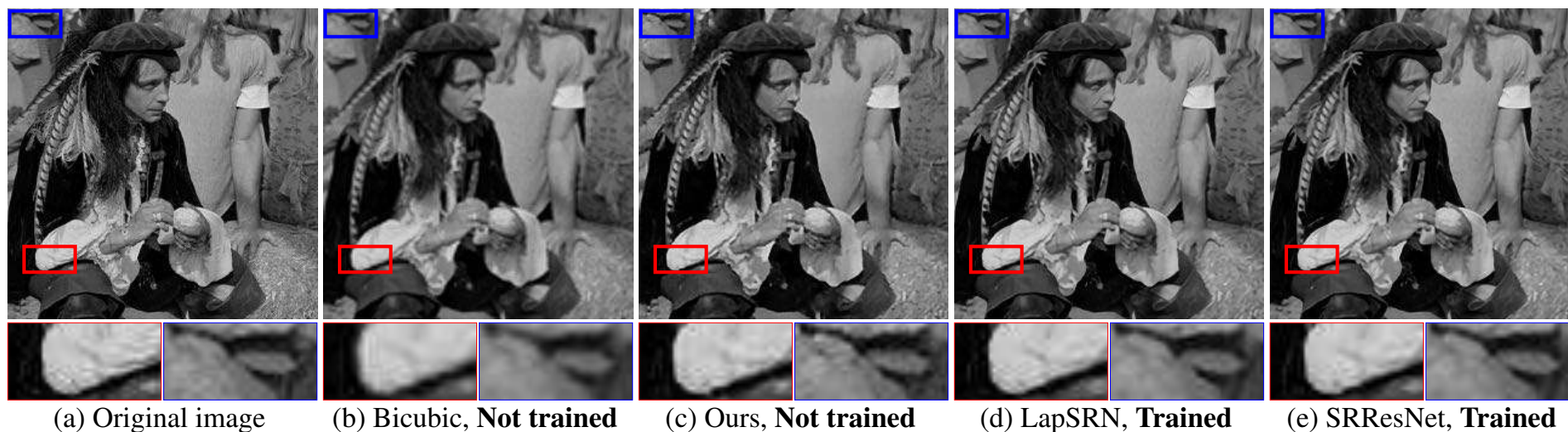


Does the excellent performance of conv nets exclusively come from from learning?

A deep conv net fits more easily to natural images compared to noise

This suggests that it can be used as a deep image prior (DIP) in an inverse problem

$$\hat{\mathbf{f}} = f_{\theta^*}(\mathbf{z}) \quad \theta^* = \arg \min_{\theta} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{H}f_{\theta}(\mathbf{z})\|_{\ell_2}^2 \right\}$$

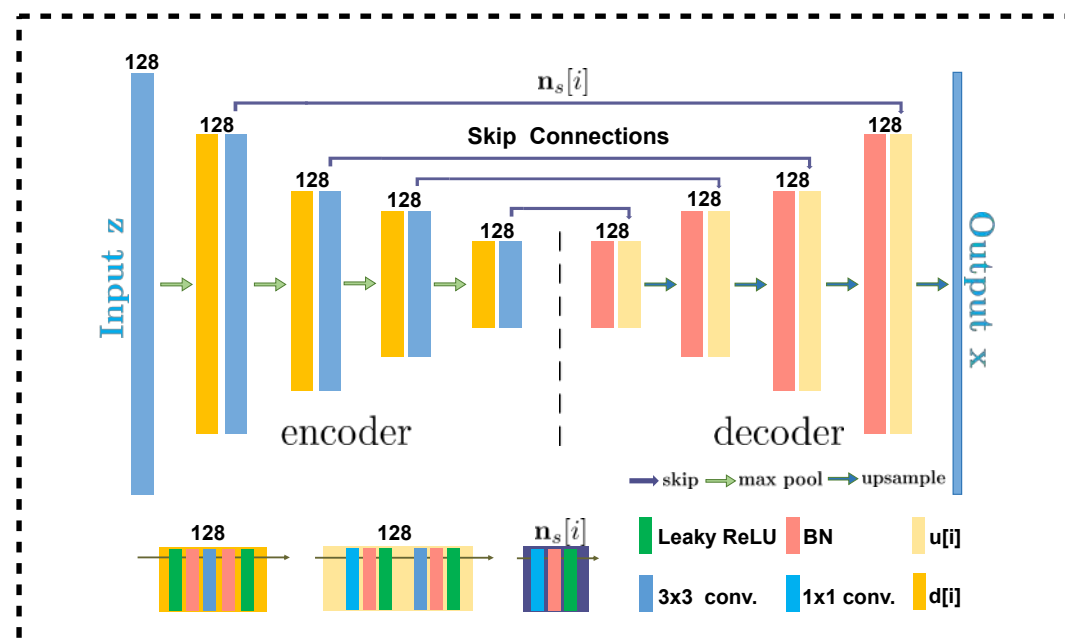


DIP can be conveniently combined with other priors to further stabilize and improve it

Can a combination of TV and DIP improve over both when they are used separately?

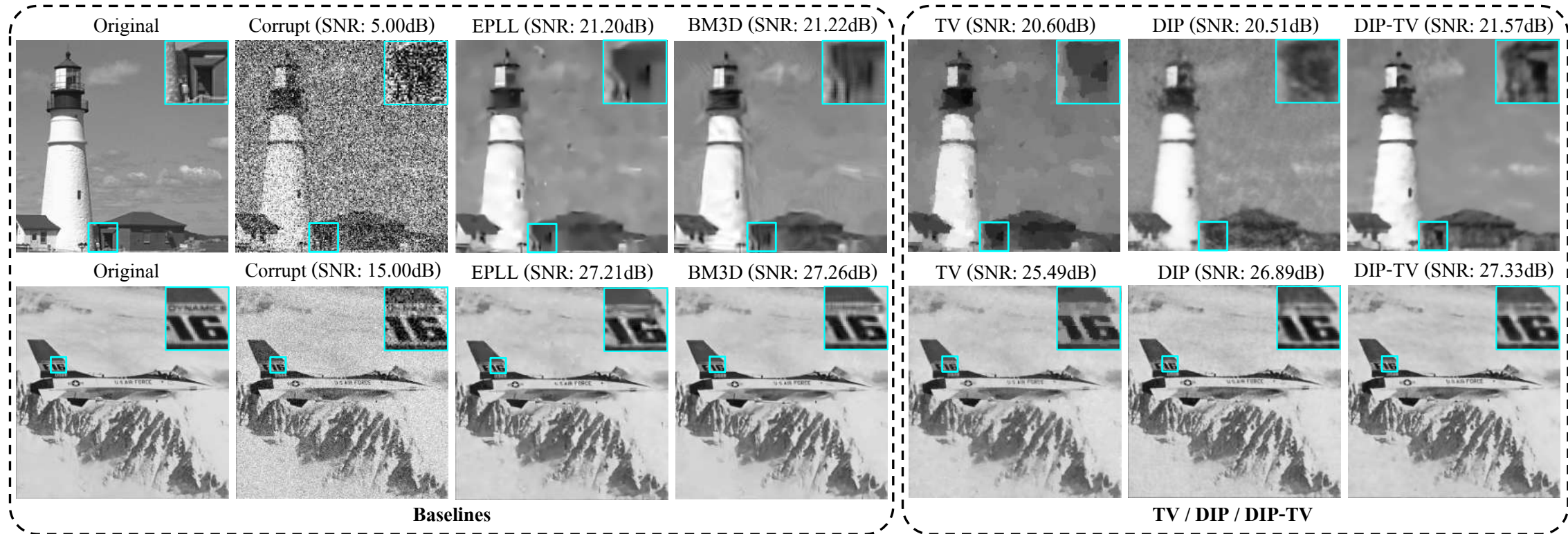
$$\hat{\mathbf{f}} = f_{\theta^*}(\mathbf{z}) \quad \theta^* = \arg \min_{\theta} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{H}f_{\theta}(\mathbf{z})\|_{\ell_2}^2 + \lambda \|\mathbf{D}f_{\theta}(\mathbf{z})\|_{\ell_1} \right\}$$

We adopt a simple **modified U-Net architecture** considered in the original DIP paper



DIP can be conveniently combined with other priors to further stabilize and improve it

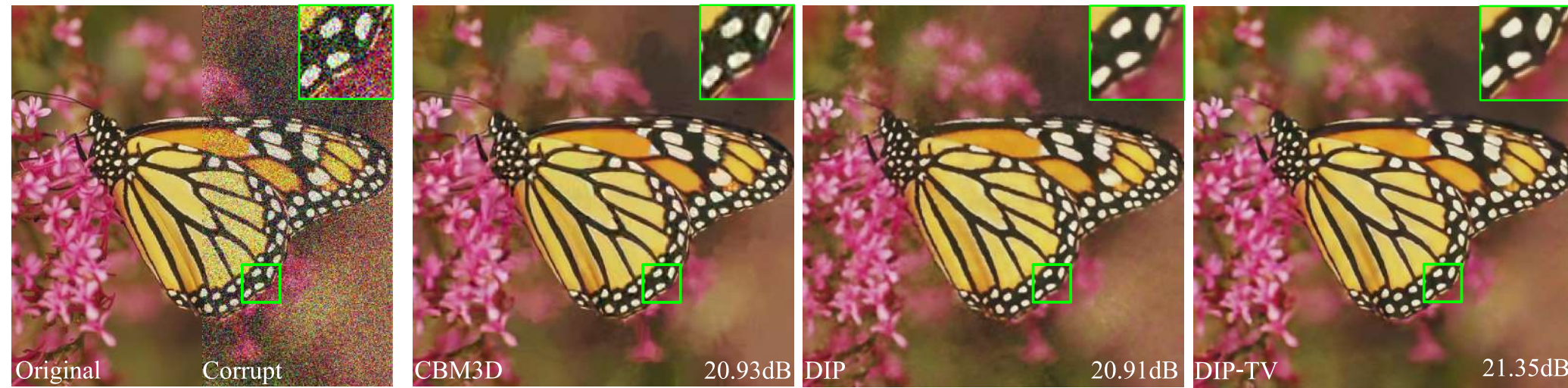
grayscale denoising



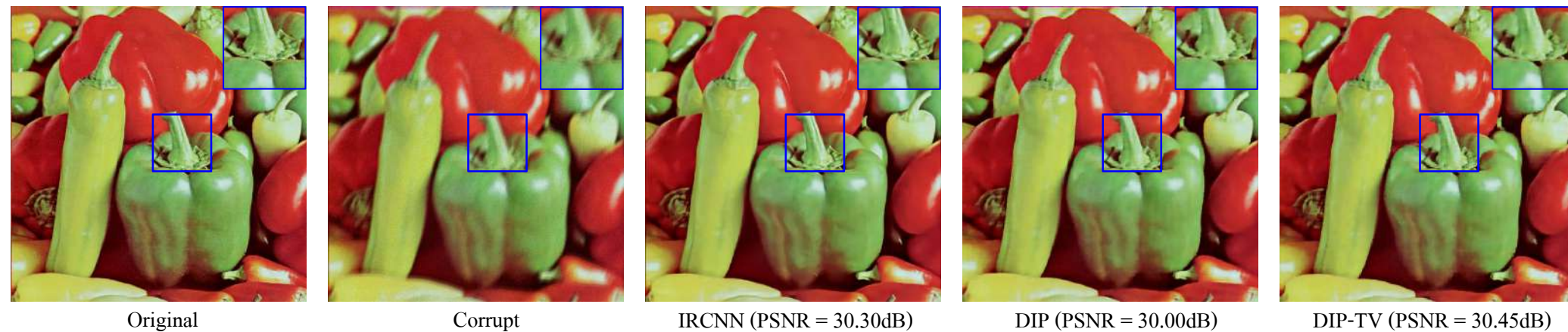
Images	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Input SNR = 5 dB / $\sigma = 76.26$														
EPLL	18.60	21.39	19.18	15.29	16.88	16.54	18.33	21.80	21.21	20.19	19.38	19.85	16.85	21.20
BM3D	18.72	22.22	18.81	15.31	16.86	16.50	18.30	21.87	21.55	20.25	19.52	20.35	17.33	21.22
TV	17.22	20.38	17.65	13.74	16.24	15.42	16.57	19.71	20.09	18.38	18.49	18.27	16.23	20.60
DIP	17.98	21.19	18.78	14.98	16.16	16.19	17.61	21.44	21.08	18.67	18.97	20.19	16.64	20.51
DIP-TV	18.84	22.41	19.56	15.52	16.99	16.79	18.48	22.26	21.61	19.10	19.55	20.52	17.80	21.57
Input SNR = 10 dB / $\sigma = 53.43$														
EPLL	21.21	24.21	21.96	17.81	19.42	19.65	20.88	24.59	23.68	21.20	21.79	22.98	19.65	23.91
BM3D	21.30	25.10	21.57	17.81	19.39	19.58	20.84	24.65	24.01	21.28	21.90	23.39	20.20	23.85
TV	19.76	22.82	20.39	16.34	18.45	18.04	18.91	22.62	22.15	20.34	20.56	20.80	18.85	22.83
DIP	20.76	24.32	21.55	17.81	18.82	19.14	20.21	24.43	23.24	21.01	21.22	23.46	19.90	22.99
DIP-TV	21.33	25.11	22.10	17.96	19.43	19.61	20.89	24.77	23.81	21.57	21.65	23.60	20.46	24.12

DIP can be conveniently combined with other priors to further stabilize and improve it

color denoising



color deblurring



To conclude

- ◉ Optical tomographic live-cell imaging could benefit from nonlinear forward models and advanced priors
- ◉ BPM is a simple, yet effective, nonlinear model that accounts for forward multiple scattering
- ◉ We increasingly rely on implicit regularization using nonlinear operators, such as deep neural nets
- ◉ Plug-In SPGM is a theoretically sound algorithm that can regularize at large-scales using nonlinear operators
- ◉ Deep conv nets can regularize with or without training, and can be combined with traditional regularizers

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

CONTACT INFO

Prof. Ulugbek Kamilov

Email: kamilov@wustl.edu

Personal Twitter: [@ukmlv](https://twitter.com/ukmlv)

Web: <http://cigroup.wustl.edu>

Group Twitter: [@wustlcig](https://twitter.com/wustlcig)



Support:



CCF-181391



Institute of Clinical and
Translational Sciences

WashU ICTS UL1TR002345



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.