Unsupervised Machine Imaging

when is data-driven knowledge discovery really possible?

Deep learning, image analysis, inverse problems, and optimization (DIPOpt) Workshop
November 27th - 30th, 2023

Mike Davies
University of Edinburgh
Joint work with Julián Tachella and Dongdong Chen

Dedicated to: Professor Nick Kingsbury who died earlier this year

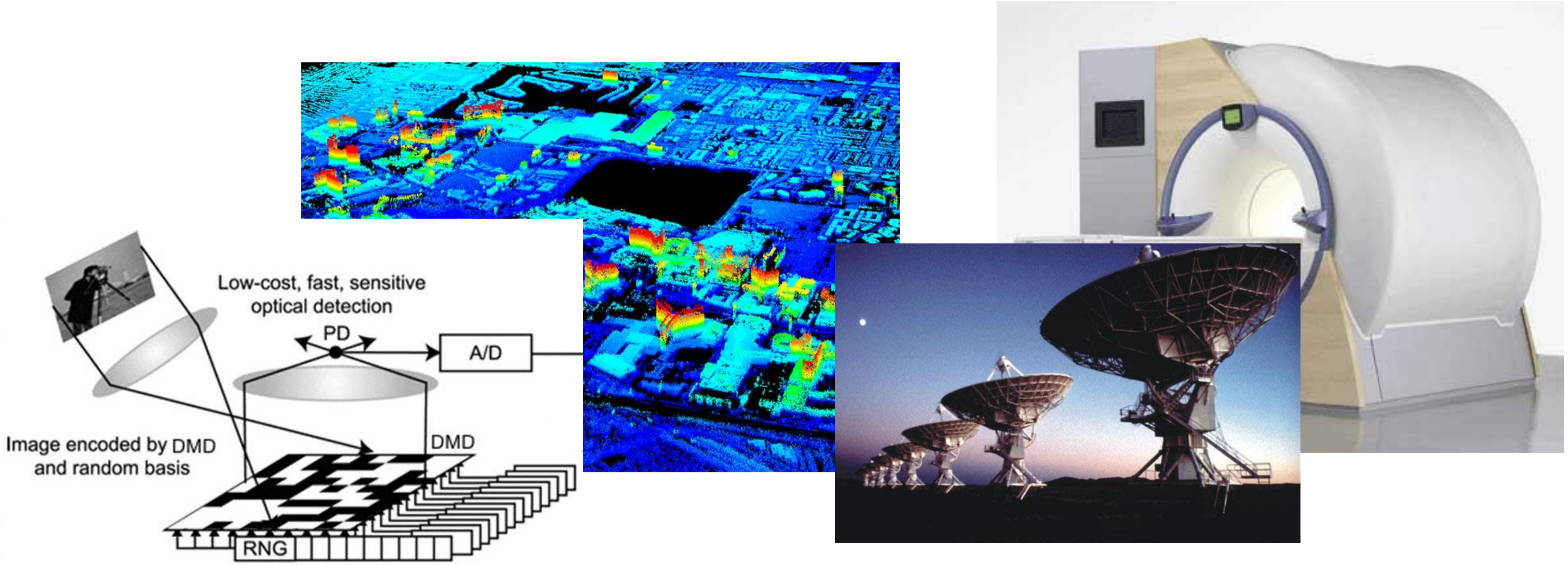
An early pioneer of sparse signal and image processing, rotation invariant wavelets, etc.

Computational & Compressed Imaging

Imaging and Sensing Challenge

Insufficient Measurements

Complete measurements can be costly, time consuming and sometimes just impossible!



The Inverse problem

Goal: estimate signal x from y

signal $\in \mathbb{R}^n$

$$\text{measurements } \in \mathbb{R}^m \rightarrow y = Ax + \epsilon \leftarrow \text{noise/error}$$

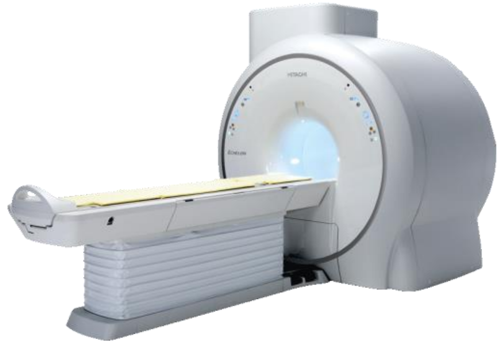
Physics

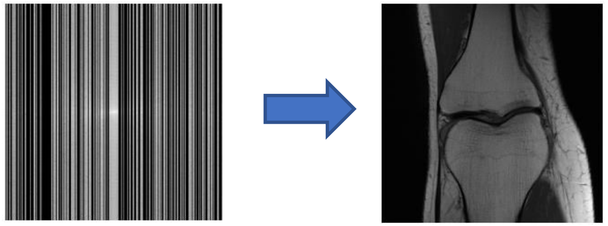
- We are particularly interested in the underdetermined case: $n > m$
- To invert, we need to restrict x to some low dimensional set, $\mathcal{X} \subset \mathbb{R}^n$

Examples

Magnetic resonance imaging

- A = subset of Fourier modes ($k - \mathrm{space}$) of 2D/3D images

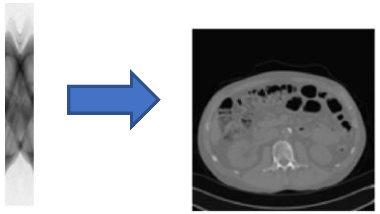




y x

Computed tomography

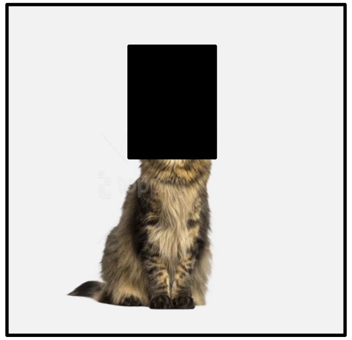
- A = 2D projections (sinograms) of 2D/3D images

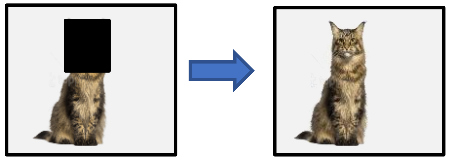


y x

Image inpainting

- A = diagonal matrix with 1's and 0s.

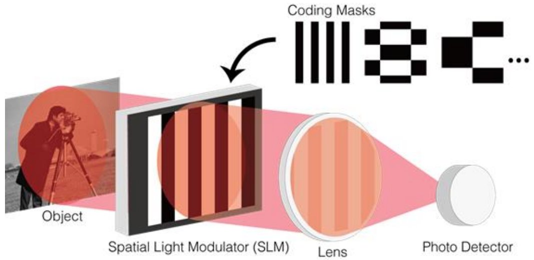


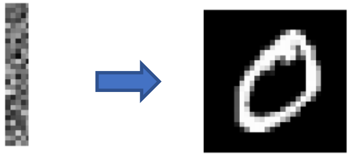


y x

Single-pixel camera

- A = binary matrix with iid Bernoulli entries





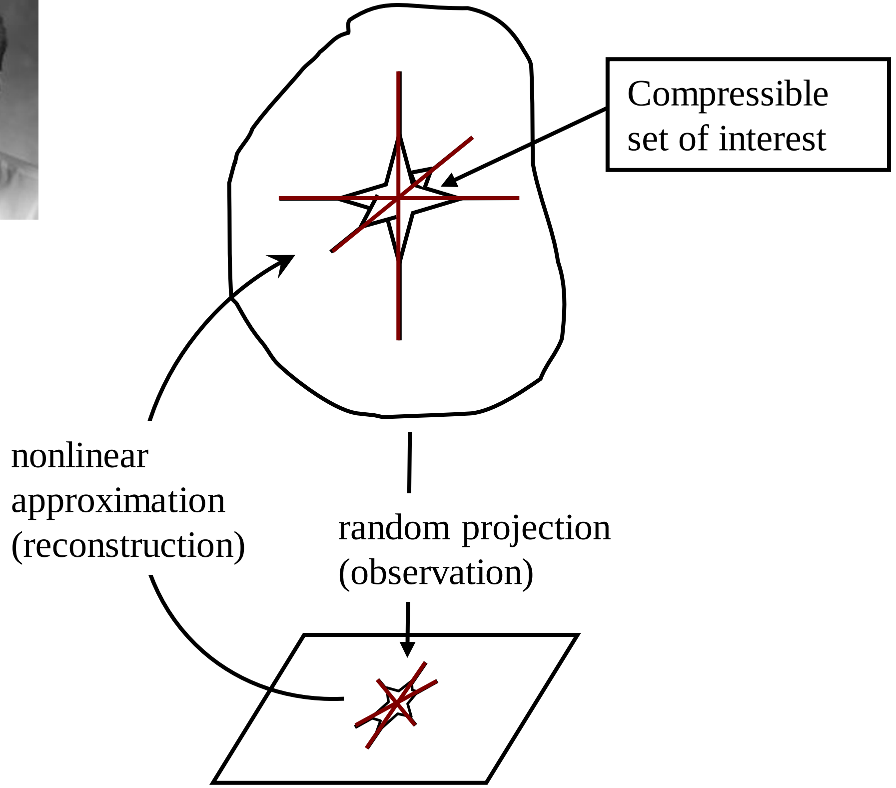
y x

Compressed sensing

This is the principle behind Compressed Sensing

E. Candès, J. Romberg, and T. Tao, **"Robust Ucertainty principles: Exact signal reconstruction from highly incomplete frequency information,"** IEEE Trans. Information Theory, 2006

D. Donoho, **"Compressed sensing,"** IEEE Trans. Information Theory, 2006



Compressed Sensing assumes a hand crafted low dimensional signal models, e.g. approximately k-sparse, low rank,...

Then using approximately

$$m \geq \mathcal{O}(k \ldots)$$

random measurements have little or no information loss. Signal reconstruction by a nonlinear mapping, e.g. L_1 min., OMP, CoSaMP, IHT, etc...

Machine Imaging

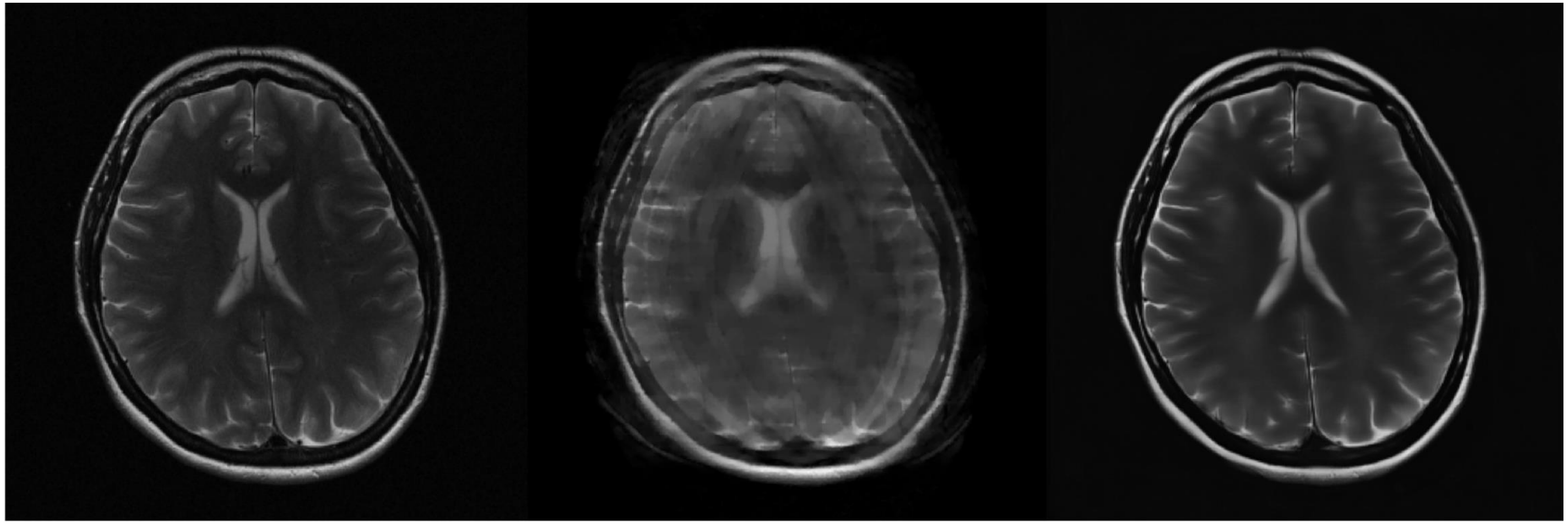
Machine Imaging

- State-of-the-art reconstructions
- Once trained, $\boldsymbol{f}$ is easy to evaluate

fastMRI

Accelerating MR Imaging with AI

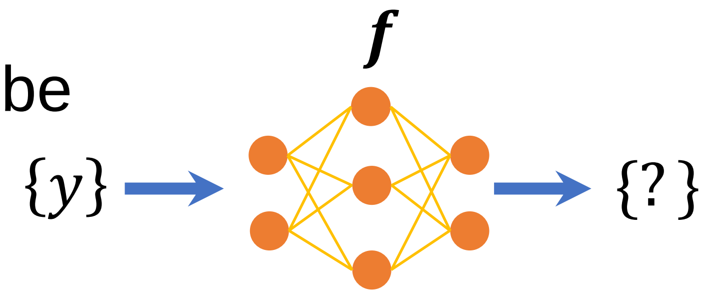
Ground-truth | Total variation (28.2 dB) | Deep network (34.5 dB)



x8 accelerated MRI [Zbontar et al., 2019]

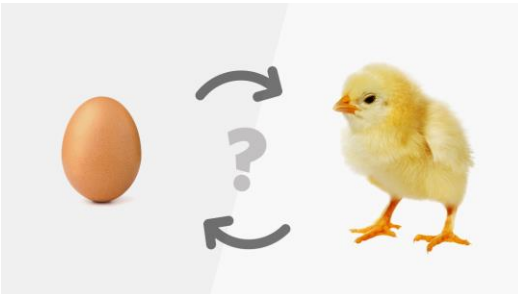
Learning approach

Main disadvantage: Obtaining training signals x_i can be expensive or impossible.



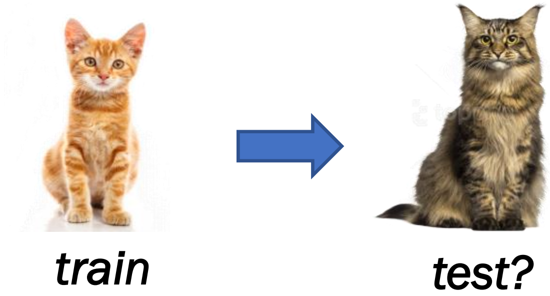
- Medical and scientific imaging

- Only solves inverse problems which we already know what to expect



- Risk of training with signals from a different distribution

- Raises the question: can AI be used for data-driven knowledge discovery in imaging?

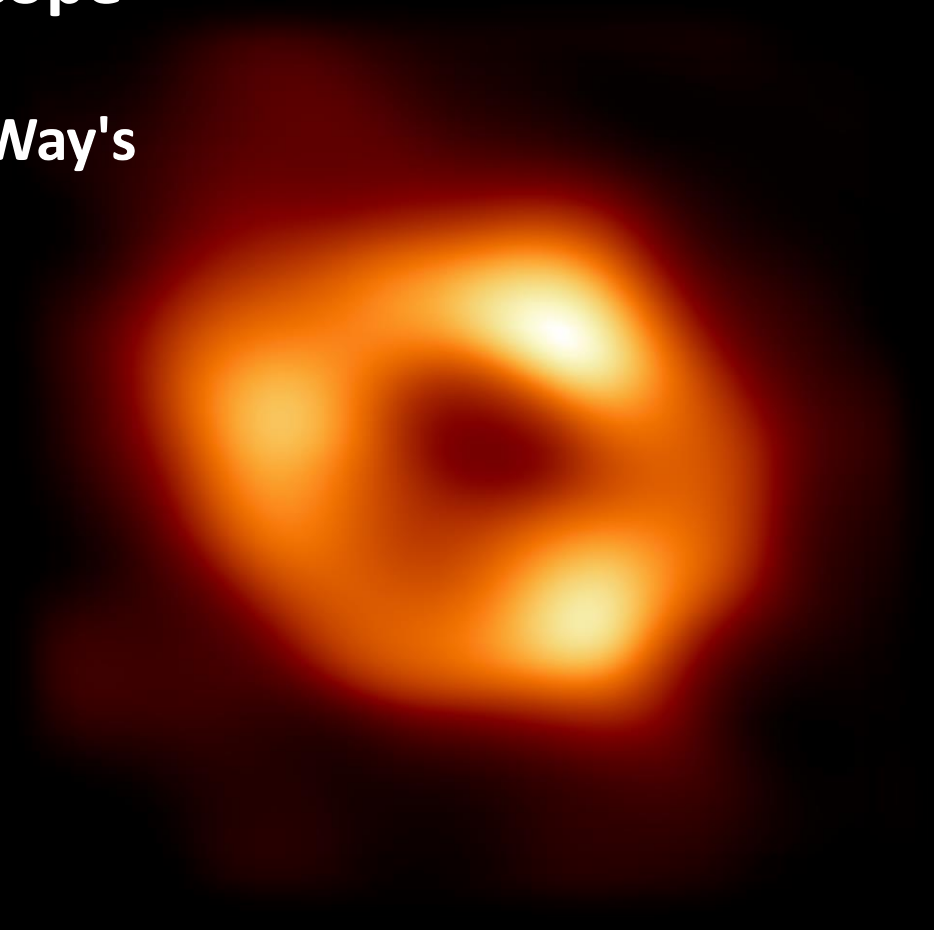


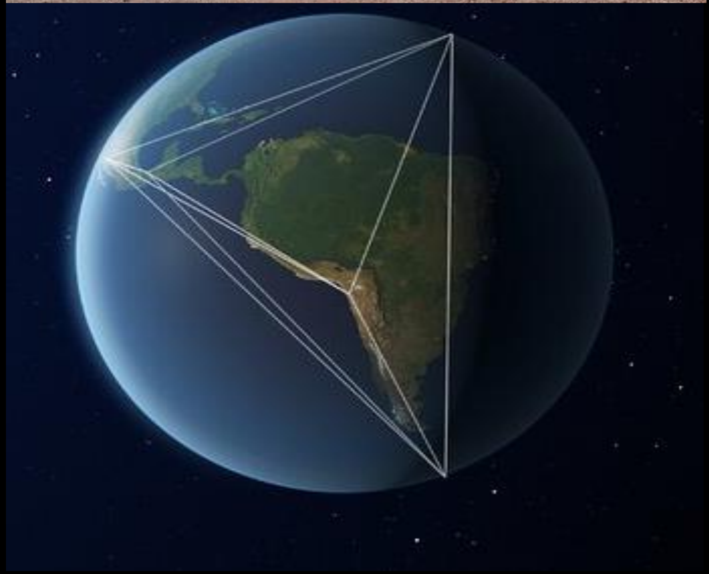
The Challenge

AI for Knowledge Discovery?

The Event Horizon Telescope

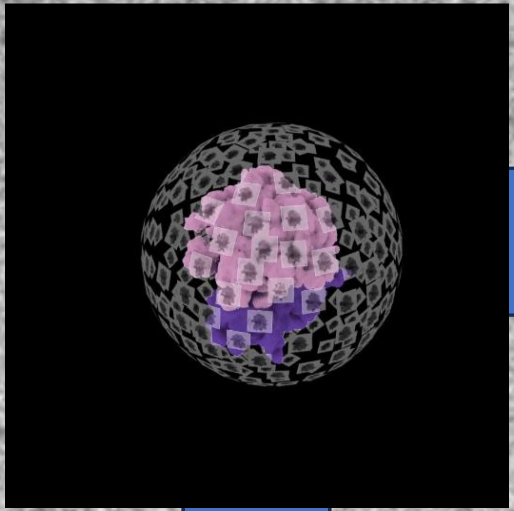
The First Image of the Milky Way's Black Hole: Sagittarius A*



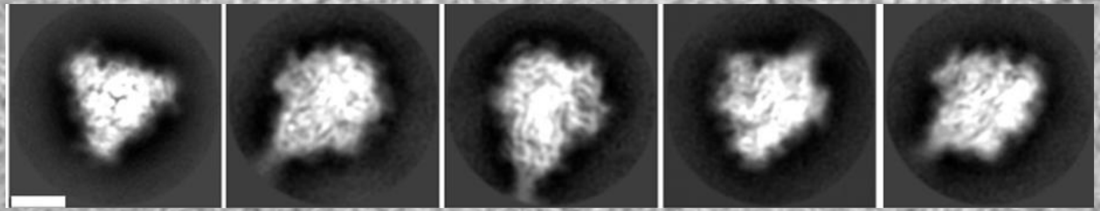


AI for Knowledge Discovery?

Cryo-electron microscopy: here are tomographic slices of SARS-CoV-2 virus particles

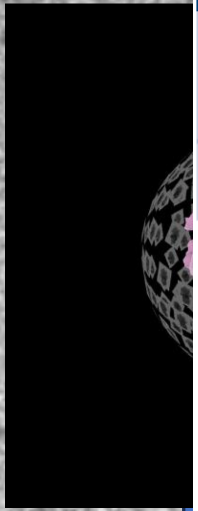


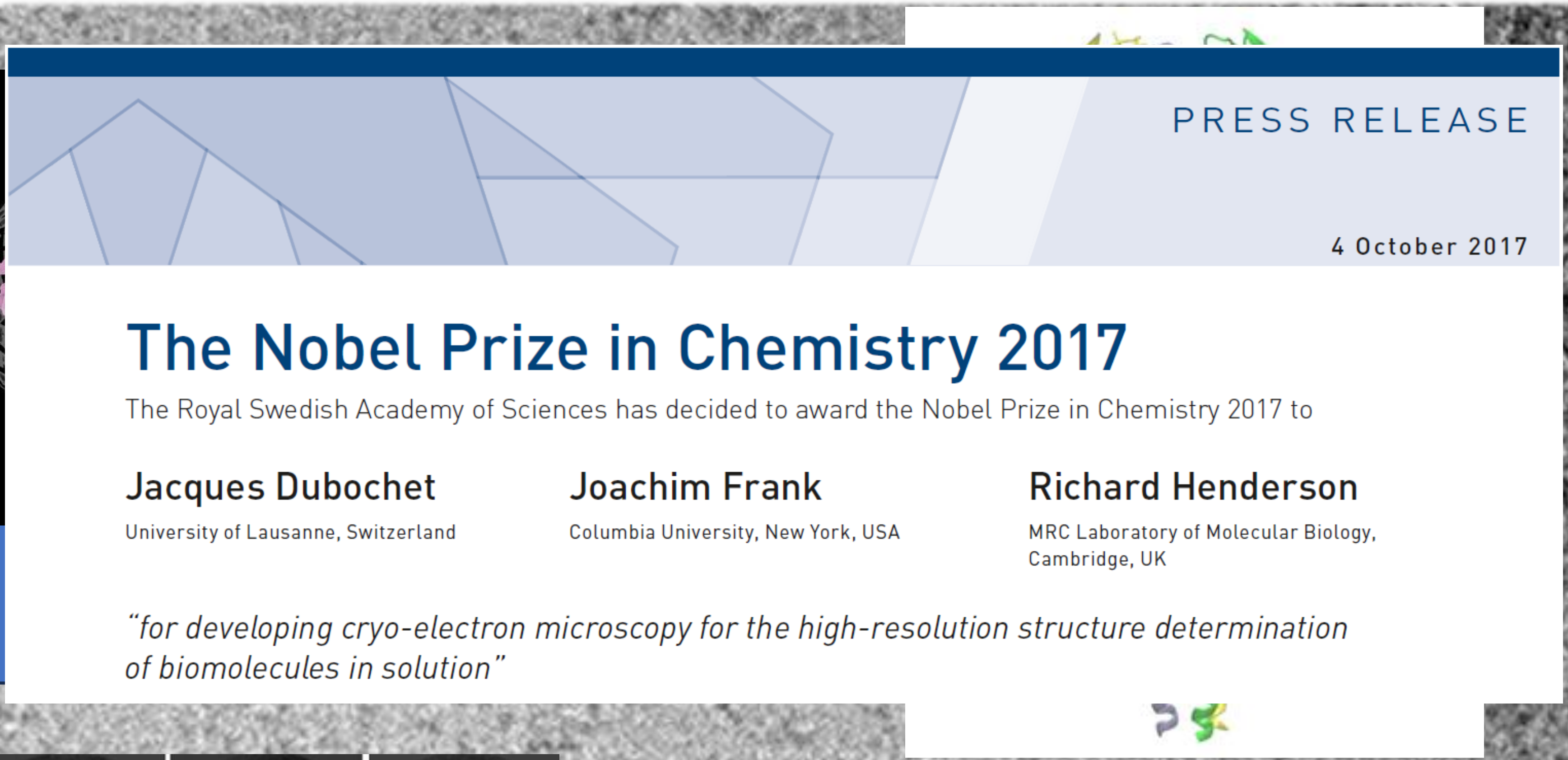
Recover



AI for Knowledge Discovery?

Cryo-electron microscopy: here are tomographic slices of SARS-CoV-2 virus particles





PRESS RELEASE

4 October 2017

The Nobel Prize in Chemistry 2017

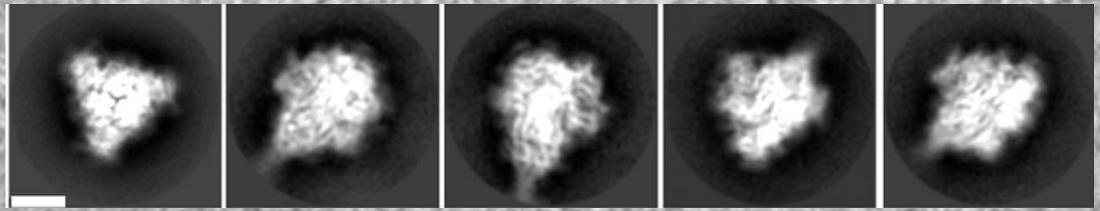
The Royal Swedish Academy of Sciences has decided to award the Nobel Prize in Chemistry 2017 to

Jacques Dubochet
University of Lausanne, Switzerland

Joachim Frank
Columbia University, New York, USA

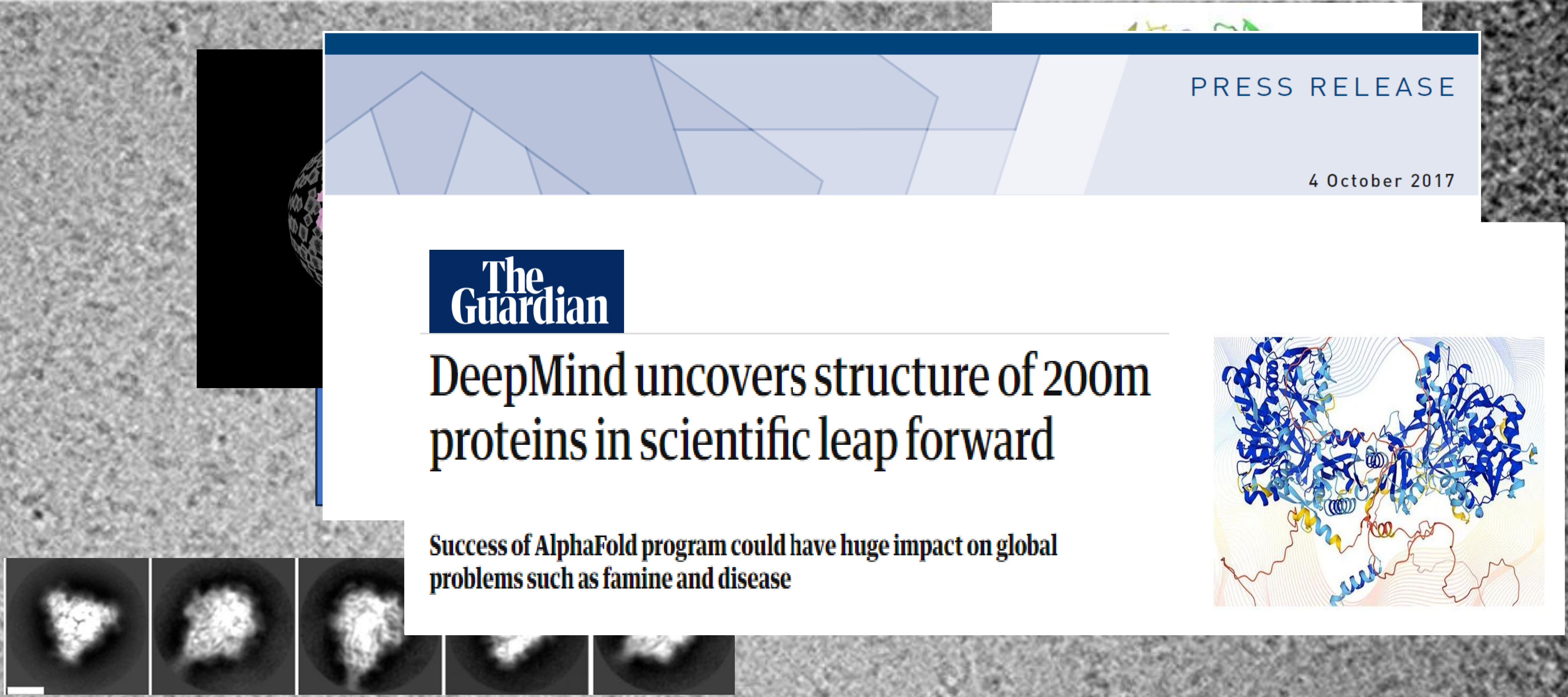
Richard Henderson
MRC Laboratory of Molecular Biology, Cambridge, UK

"for developing cryo-electron microscopy for the high-resolution structure determination of biomolecules in solution"



AI for Knowledge Discovery?

Cryo-electron microscopy: here are tomographic slices of SARS-CoV-2 virus particles



The Inverse problem (again)

Goal: estimate signal x from y

$$y = Ax + \epsilon$$

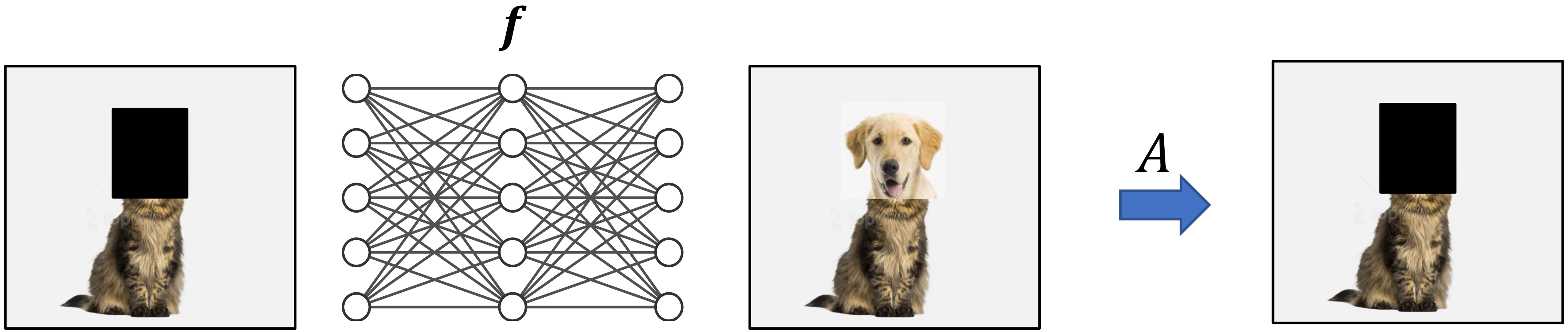
- We are interested in the underdetermined case: $n > m$
- To invert, we need to restrict x to some low dimensional set, $\mathcal{X} \subset \mathbb{R}^n$
- However, $\mathcal{X}$ is unknown and we need to learn it

A Simple Null(-space) Result

Can we learn from only the measurements, $\{y_i\}$?

(Can we minimise the measurement consistency $\sum_i \|y_i - Af(y_i)\|^2$?)

Proposition: Any reconstruction function $f(y) = A^{\dagger}y + g(y)$ is measurement consistent for any $g: \mathbb{R}^m \mapsto \mathcal{N}_A$ whose image belongs to the nullspace of A.



Possible Solutions

How to learn from $\{y_i\}$?

(1) *We either need more measurement information, or*

(2) *We need more prior information*

Multi-Operator & Equivariant Imaging

Multiple Measurement Operators

more measurement information…

Multiple Operators: what if we observe the images through a set of multiple different (known) operators? Can we learn from $\{y_i, A_{g_i}\}$ pairs where:

$$y_i = A_{g_i} x_i, \qquad g_i \in G$$

Example: accelerated MRI with different k-space (sub-)sampling.

Cramér-Wold Theorem: A Borel probability measure on $\mathbb{R}^n$ is uniquely determined by its one-dimensional projections.

Argument used in Ambient GAN [Bora et al. 2018] but in most settings we will not have access to such a large set of projections. Can we use fewer?

Symmetry

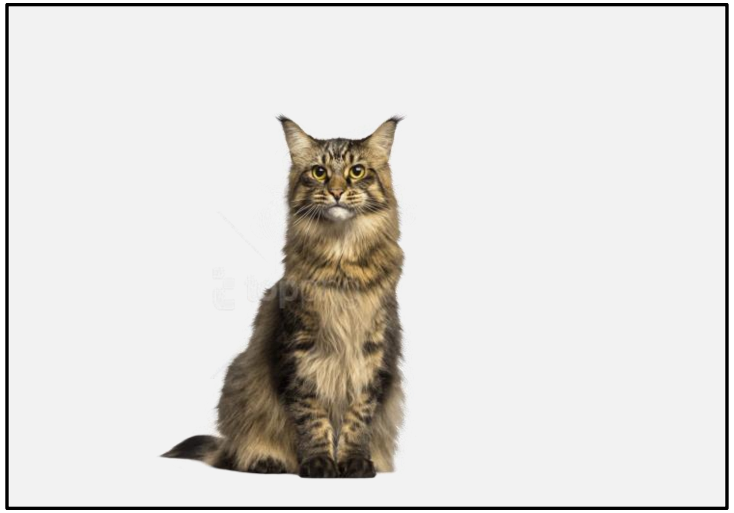
more prior information…

Idea: Most natural image and signal distributions are invariant to certain groups of transformations:

$$\forall x \in \mathcal{X},\ \forall g \in G,\ x' = T_g x \in \mathcal{X}$$

where T_g is the group action for $g \in G$

Example: natural images are invariant to…
G = group of translational shifts
G = group of 2D rotations



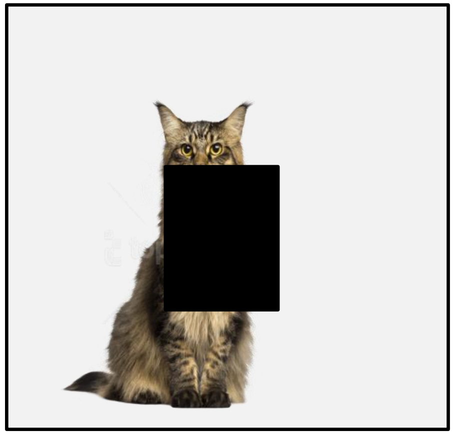
Exploiting Invariance

These two views are somewhat related…

Consider

$$y = Ax = \underbrace{AT_g}_{A_g}\,\overbrace{T_g^{-1}x}^{x'} = A_g x'$$

- Implicitly invariance gives access to multiple operators A_g,
- Each operator *may* have a different nullspace



Model identification

When can we *uniquely* identify the set of signals $\mathcal{X} \subset \mathbb{R}^n$ from the observed measurement sets $\{\mathcal{Y}_g = A_g\mathcal{X}\}$?

Necessary conditions

Proposition [Tachella, Chen and D. ’23]: Identifying $\mathcal{X}$ from multiple sets $\{\mathcal{Y}_g = A_g\mathcal{X}\}$ possible only if

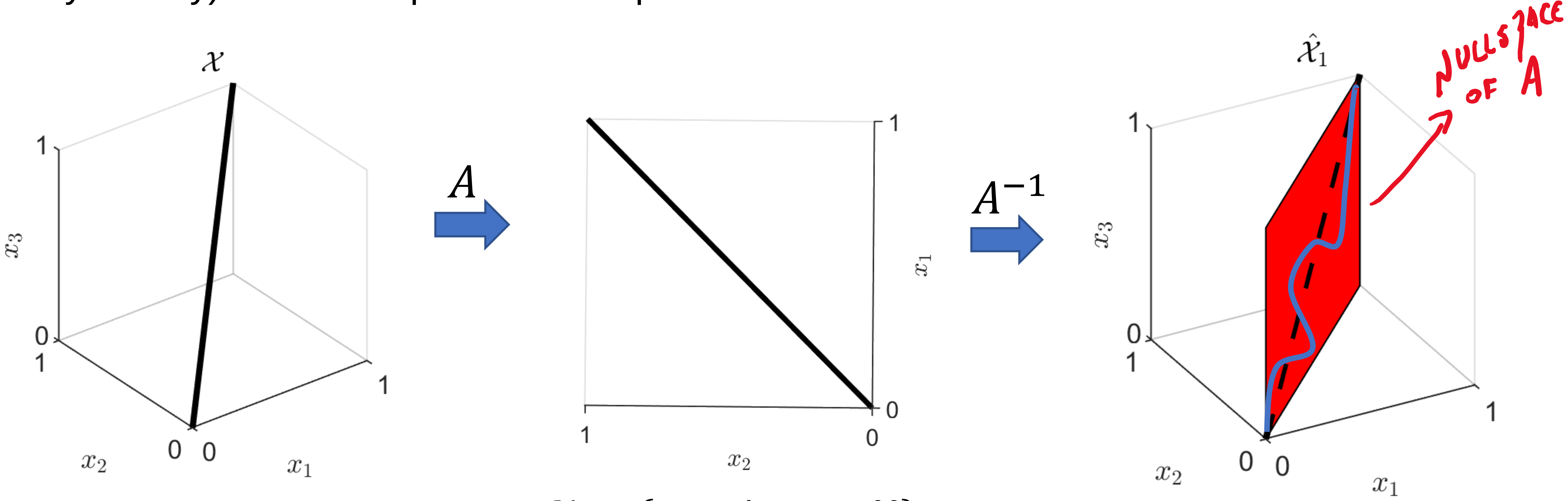
$$\operatorname{rank}\left(\begin{bmatrix} A_1 \\ \vdots \\ A_{|G|} \end{bmatrix}\right) = n,$$

and thus, only if: $m \geq \frac{n}{|G|}$ or, for invariance, $\mathrm{m} \geq \max \frac{c_j}{s_j}$ where s_j and c_j are dimension and multiplicity of the irreducible representations of T_g.

Corollary [Tachella, Chen and D. ’23]: For invariance, identifying $\mathcal{X}$ requires that A is **not equivariant:** $AT_g \neq \tilde{T}_g A$

Geometric intuition

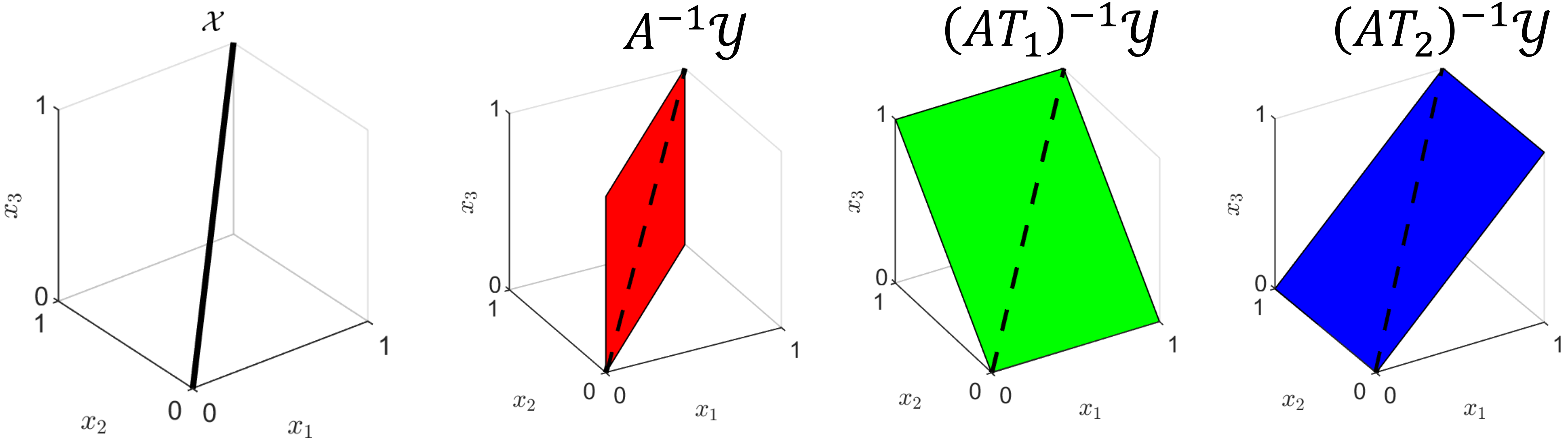
Toy example ($n = 3, m = 2$): Signal set is $\mathcal{X} = \mathrm{span}([1,1,1]^{\mathrm{T}})$ (note the permutation symmetry). Forward operator A keeps first 2 coordinates.



$$\mathcal{Y} := \{y = Ax, x \in \mathcal{X}\}$$

Geometric intuition

Toy example ($n = 3, m = 2$): Signal set is $\mathcal{X} = \mathrm{span}[1,1,1]^{\mathrm{T}}$. Forward operator A keeps first 2 coordinates. **Now with explicit permutation symmetry**



$$\mathcal{X} = \bigcap_{g\in G} A_g^{-1}\mathcal{Y}$$

Sufficient conditions

Additional assumption: The signal set is **low-dimensional** (c.f. Compressed Sensing)

The signal model has a box-counting dimension $k < n$

Examples: Sparse dictionaries, union of subspaces, manifold models, etc.

Theorem [Tachella, Chen and D. '22] (**multi-operator imaging**):
A k-dimensional model $\mathcal{X} \subset \mathbb{R}^n$ is identifiable from the sets $\mathcal{Y}_g := A_g\mathcal{X},\ g \in G$ for *almost every set of operators* $A_1, \ldots, A_{|G|} \in \mathbb{R}^{n\times m}$ if

$$m > k + \frac{n}{|G|}$$

Sufficient conditions

Additional assumption: The signal set is **low-dimensional** (c.f. Compressed Sensing)

The signal model has a box-counting dimension $k < n$

Examples: Sparse dictionaries, union of subspaces, manifold models, etc.

Theorem [Tachella, Chen and D. '22] (**equivariant imaging**): Let G be a compact cyclic group. A k-dimensional model $\mathcal{X}$ is identifiable from the sets $\mathcal{Y}_g := AT_g\mathcal{X},\ g \in G$ for *almost every* $A \in \mathbb{R}^{n\times m}$ if

$$m > 2k + \max c_j + 1 \geq 2k + \frac{n}{|G|} + 1$$

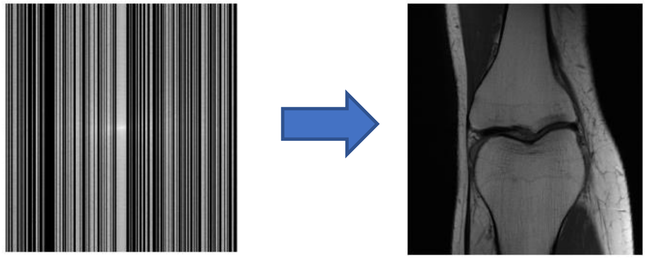
where c_j is the multiplicity of the representation and $\max c_j \geq n/|G|$.

Consequences

Magnetic Resonance Imaging

- A = subset of Fourier modes
- Equivariant to translations
- Not equivariant to rotations, which have $\max c_j \approx \sqrt{n}$

$$m > 2k + \sqrt{n} + 1$$

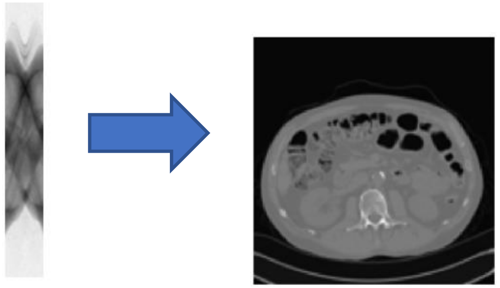


y x

Computed Tomography (CT)

- A = 1D projections (sinograms)
- Equivariant to translations
- Not equivariant to rotations, which have $\max c_j \approx \sqrt{n}$

$$m > 2k + \sqrt{n} + 1$$

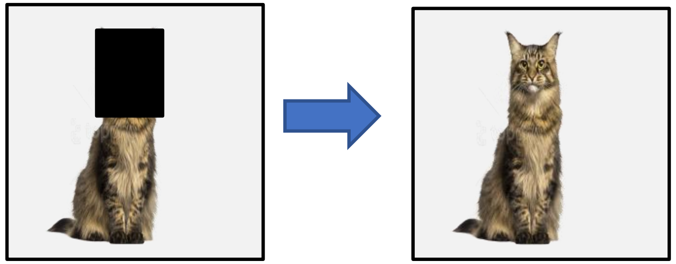


y x

Image inpainting

- A = diagonal matrix with 1’s and 0s.
- Not equivariant to translations, which have $\max c_j \approx 1$

$$m > 2k + 2$$



y x

Equivariant Imaging:
the Algorithm

Equivariant Imaging

How can we enforce invariance in practice?

Idea: we would like $f \circ A$ to be G-equivariant, i.e. $f\left(AT_g x\right) = T_g f(Ax)$

Unsupervised training loss

$$\underset{f}{\operatorname{argmin}} \ \mathcal{L}_{MC}(f) \ + \mathcal{L}_{EI}(f)$$

- $\mathcal{L}_{MC}(f) = \sum_i \|y_i - Af(y_i)\|^2$ measurement consistency
- $\mathcal{L}_{EI}(f) = \sum_{i,g} \left\|f\left(AT_g f(y_i)\right) - T_g f(y_i)\right\|^2$ enforces equivariance of $f \circ A$

Network-agnostic: applicable to any existing deep model!

Equivariant Imaging

How can we enforce invariance in practice?

Idea: we would like $f \circ A$ to be G-equivariant, i.e. $f\left(AT_g x\right) = T_g f(Ax)$

Unsupervised training loss

$$\underset{f}{\operatorname{argmin}}\ \mathcal{L}_{SURE}(f) + \mathcal{L}_{EI}(f)$$

- $\mathcal{L}_{MC}(f) = \sum_i \|y_i - Af(y_i)\|^2$ measurement consistency
- $\mathcal{L}_{EI}(f) = \sum_{i,g} \left\| f\left(AT_g f(y_i)\right) - T_g f(y_i) \right\|^2$ enforces equivariance of $f \circ A$
- For noisy measurements replace $\mathcal{L}_{MC}(f)$ with a SURE loss function: $\mathcal{L}_{\text{SURE}}(f)$

Network-agnostic: applicable to any existing deep model!

Handling noise via SURE

Gaussian noise $y \sim \mathcal{N}(u, I\sigma^2)$

$$\mathcal{L}_{SURE}(f) = \sum_i \|y_i - Af(y_i)\|^2 - \sigma^2 m + 2\sigma^2 \mathrm{div}(A \circ f)(y_i)$$

where $\mathrm{div}(h(x)) = \sum_j \frac{\delta h_j}{\delta x_j}$ is approximated with a Monte Carlo estimate which only requires evaluations of h [Ramani, 2008]

Theorem [Stein, 1981] Under mild differentiability conditions on the function $A \circ f$, the following holds

$$\mathbb{E}_{y,u}\{\mathcal{L}_{MC}(f)\} = \mathbb{E}_y\{\mathcal{L}_{SURE}(f)\}$$

Can also be extended to **Poisson noise**, and mixed **Poisson-Gaussian noise**

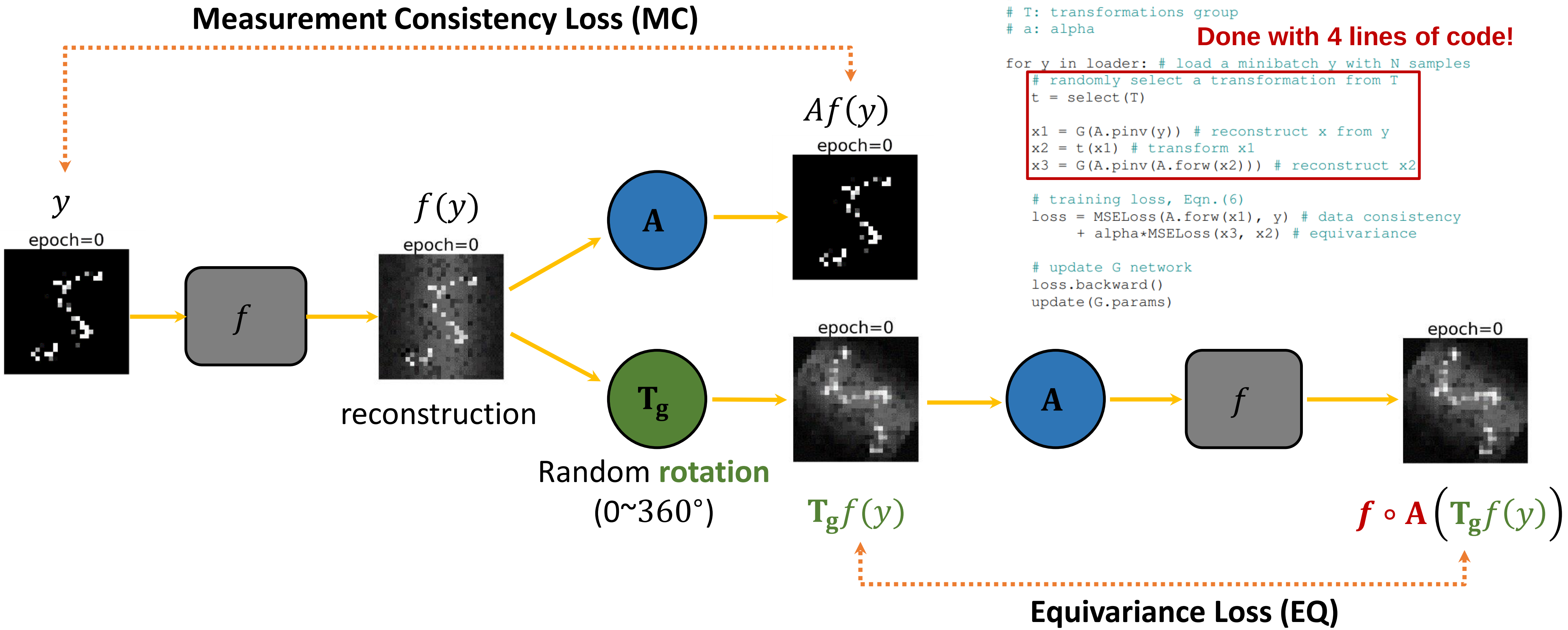
Equivariant Imaging

Example: Imaging Inpainting (MNIST)

Algorithm 1 Pseudocode of EI in a PyTorch-like style.

```
# A.forw, A.pinv: forward and pseudo inverse operators
# G: neural network
# T: transformations group
# a: alpha

for y in loader: # load a minibatch y with N samples
    # randomly select a transformation from T
    t = select(T)

    x1 = G(A.pinv(y)) # reconstruct x from y
    x2 = t(x1) # transform x1
    x3 = G(A.pinv(A.forw(x2))) # reconstruct x2

    # training loss, Eqn.(6)
    loss = MSELoss(A.forw(x1), y) # data consistency
         + alpha*MSELoss(x3, x2) # equivariance

    # update G network
    loss.backward()
    update(G.params)
```

Done with 4 lines of code!



"Equivariant Imaging: Learning Beyond the Range Space", Chen, Tachella, and Davies, *ICCV*, 2021.

Equivariant Imaging

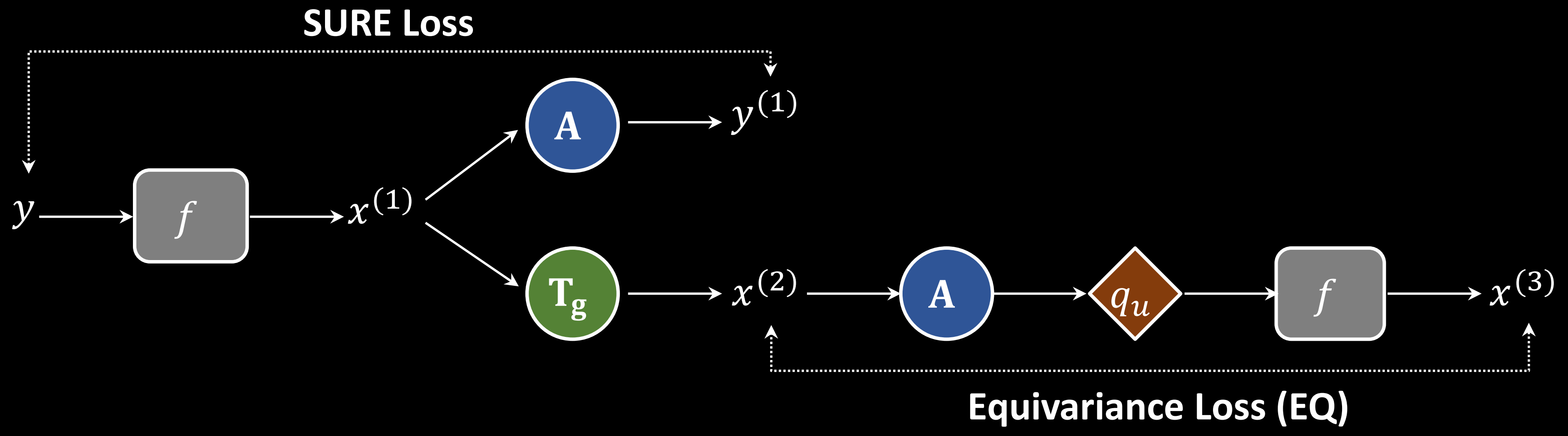
Example: Imaging Inpainting (MNIST)

Algorithm 1 Pseudocode of EI in a PyTorch-like style.

```
# A.forw, A.pinv: forward and pseudo inverse operators
# G: neural network
```

Robust Equivariant Imaging

REI also possible:



Equivariance Loss (EQ)

“Equivariant Imaging: Learning Beyond the Range Space”, Chen, Tachella, and Davies, *ICCV*, 2021.

Experiments

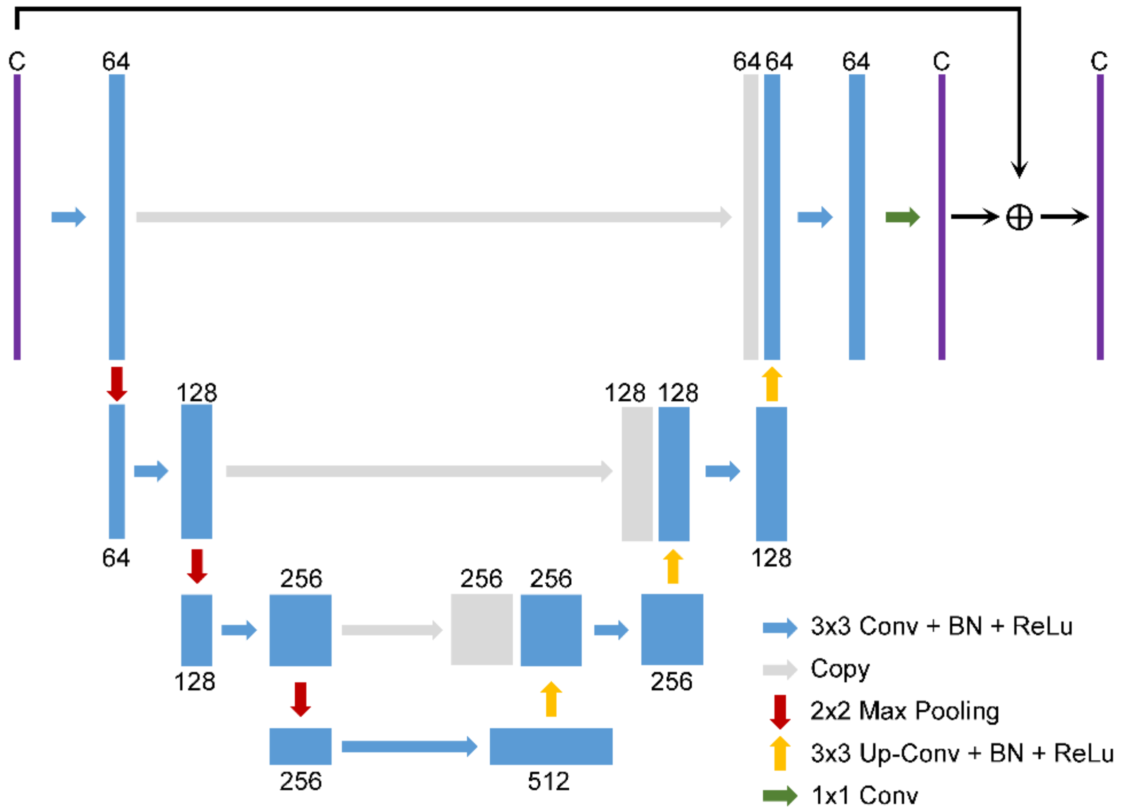
Experiments

Network

- $f = g_\theta \circ A^\dagger$ where g_θ is a U-net CNN

Comparison

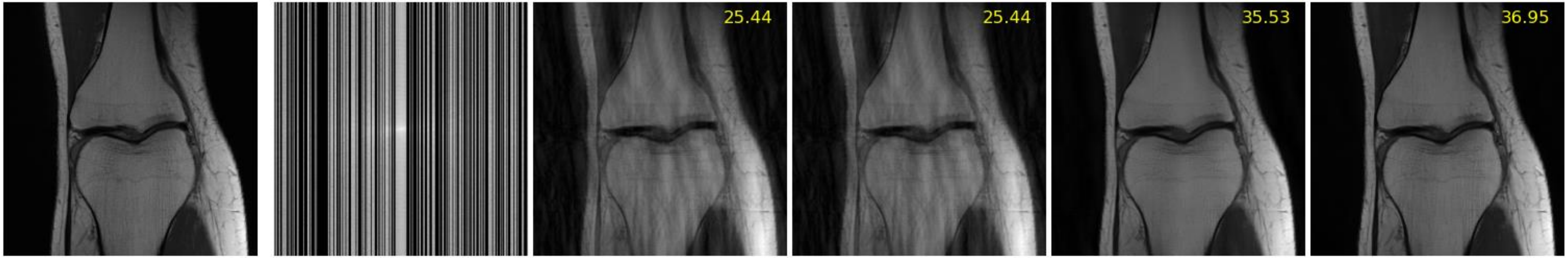
- Pseudo-inverse $A^\dagger y_i$ (no training)
- Meas. consistency $Af(y_i) = y_i$
- Fully supervised loss: $f(y_i) = x_i$
- Equivariant imaging (unsupervised)
  $Af(y_i) = y_i$ and equivariant $f \circ A$



Magnetic resonance imaging

- Operator A is a subset of Fourier measurements (x2 downsampling)
- Dataset is approximately rotation invariant

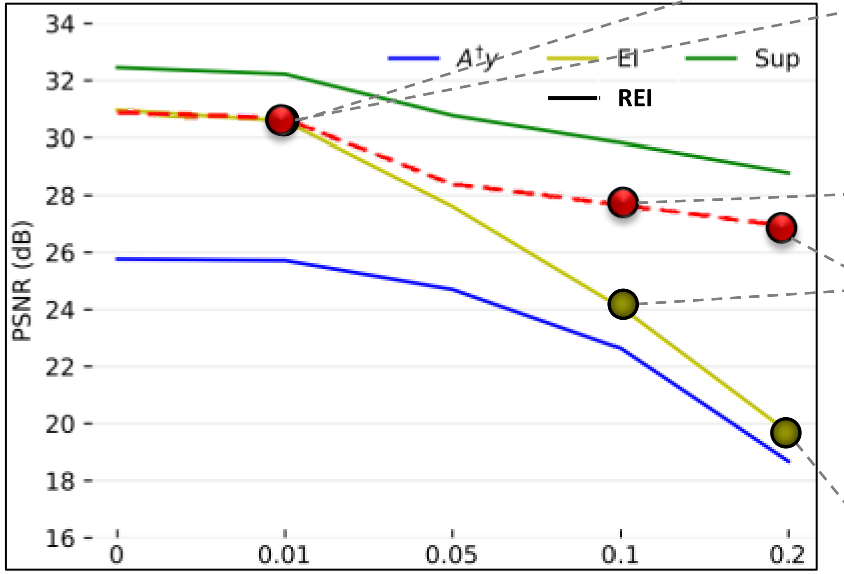
Signal x | Measurements y | $A^{\dagger}y$ | Meas. consistency | Equivariant imaging | Fully supervised

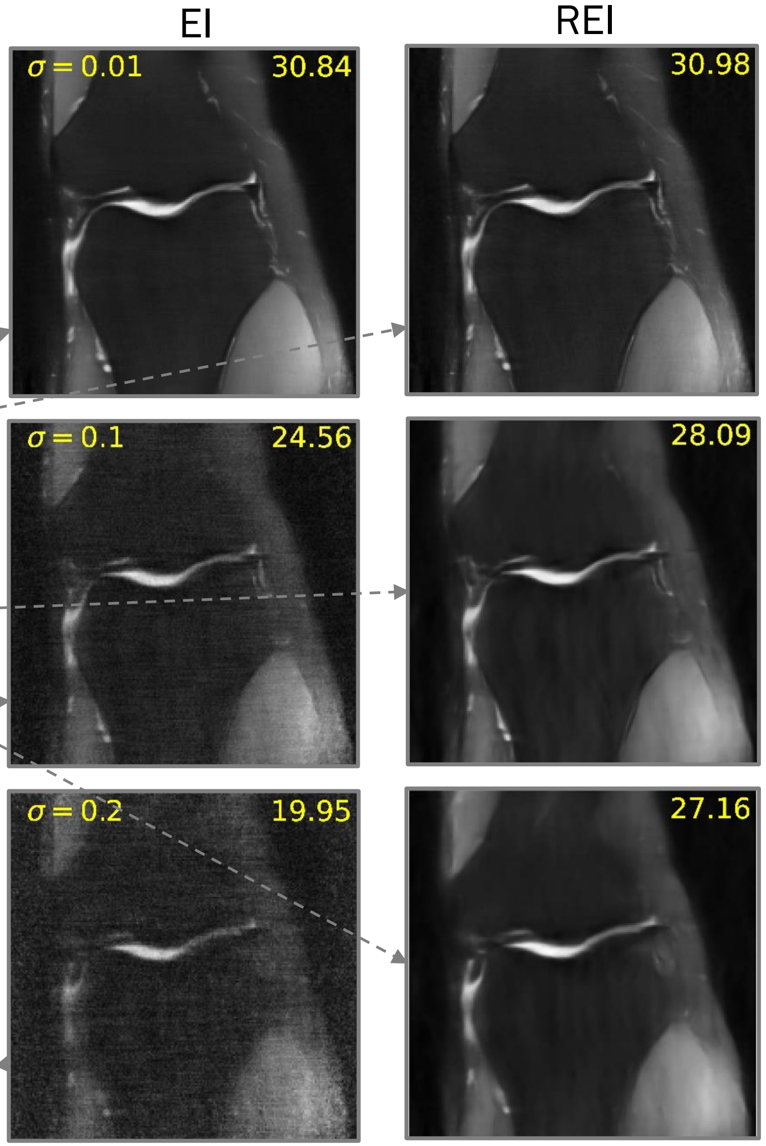


Magnetic resonance imaging

However, EI degrades with measurement noise! (4x undersamping)

MRI with different noise levels:

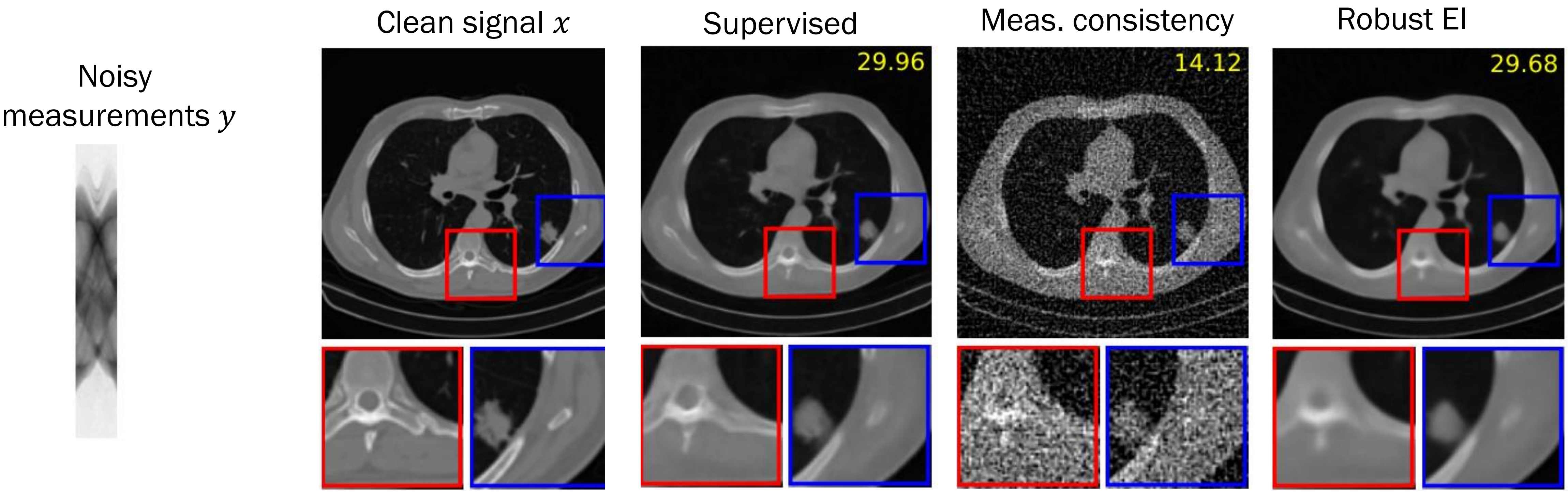




σ: Gaussian noise standard deviation $y \sim \mathcal{N}(u, \sigma^2 I)$ with $u = \mathbf{A}(x)$.

Low Dose Computed tomography

- Operator A is (non-linear variant) sparse radon transform (50 views)
- Mixed Poisson-Gaussian noise
- Dataset is approximately rotation invariant



Single-pixel camera

Measurement only training is avoids issues of distribution shift

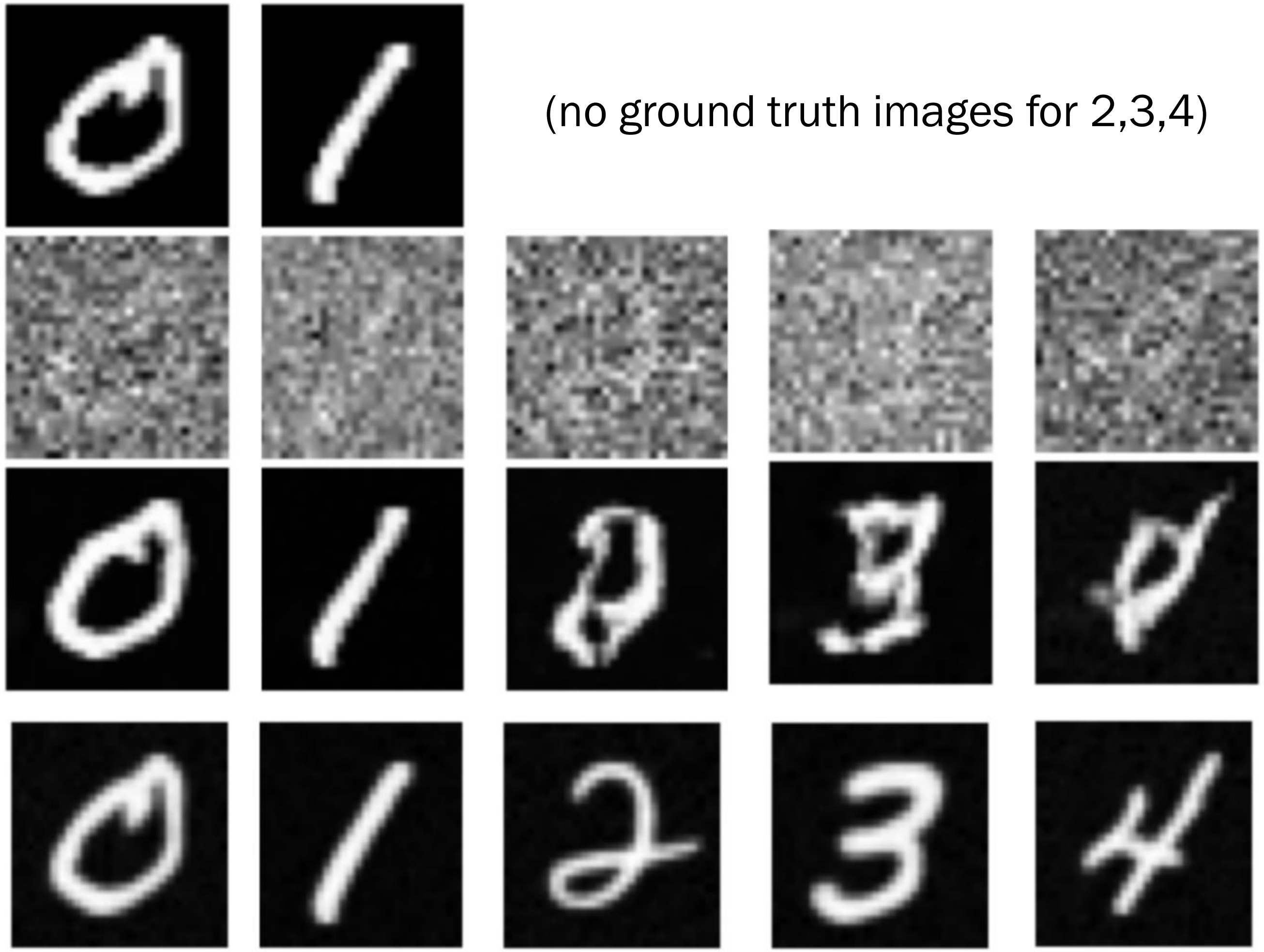
Training Images

(no ground truth images for 2,3,4)

Training Measurements

Supervised Reconstruction

Unsupervised Reconstruction (EI)



Conclusions

Novel unsupervised learning framework for imaging

Theory: Necessary & sufficient conditions for learning

- Number of measurements
- Exploiting either multiple measurements or equivariance
- Interplay between forward operator & data invariance

Practice: deep learning approach

- Unsupervised loss which can be applied to any model
- Can be made robust to noise through SURE loss

Outlook

Ongoing/future work

- Extending the theory
- Nonlinear inverse problems
- Semi-group actions
- Approximate low-dimensional models
- Sample complexity for learning
- New applications…
- Links to other Self-Supervised Learning Techniques…

Papers

[1] "Equivariant Imaging: Learning Beyond the Range Space", Chen, Tachella and Davies, *ICCV 2021*

[2] "Robust Equivariant Imaging: a fully unsupervised framework for learning to image from noisy and partial measurements", Chen, Tachella and Davies, *CVPR 2022*

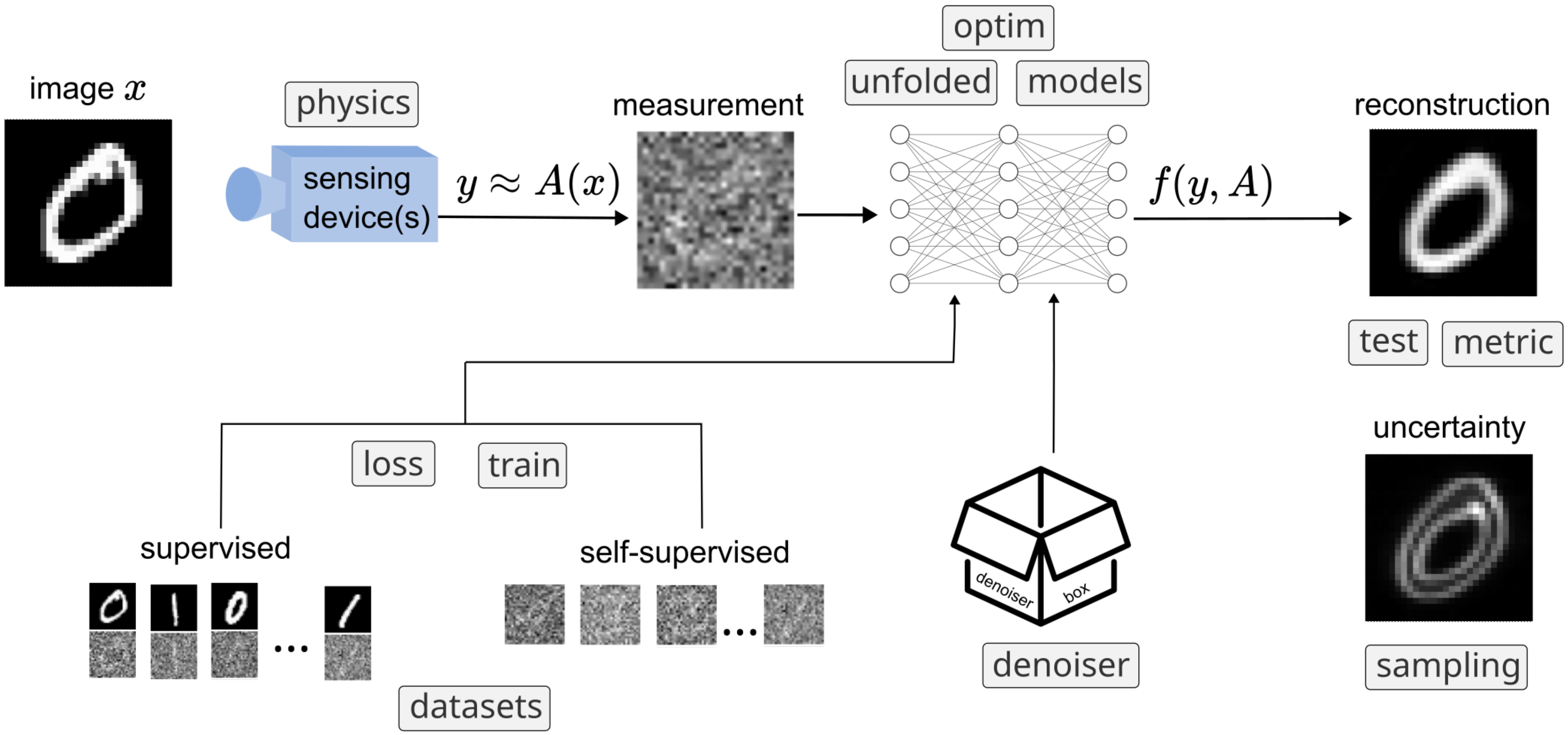
[3] "Unsupervised Learning From Incomplete Measurements for Inverse Problems", Tachella, Chen and Davies, *NeurIPS 2022*.

[4] "Sensing Theorems for Unsupervised Learning in Inverse Problems", Tachella, Chen and Davies, *JMLR 2023*.

[5] "Imaging with Equivariant Deep Learning", Chen, Davies, Ehrhardt, Schonlieb, Sherry and Tachella, IEEE SPM *2023*.

[6] "Self-Supervised Learning for Image Super-Resolution and Deblurring", Scanvic, Abry, Davies and Tachella, Submitted, 2023.

New PyTorch Library



Questions?