Self Supervised Learning Methods for Imaging

Part 3: Learning from multiple operators

Mike Davies, University of Edinburgh

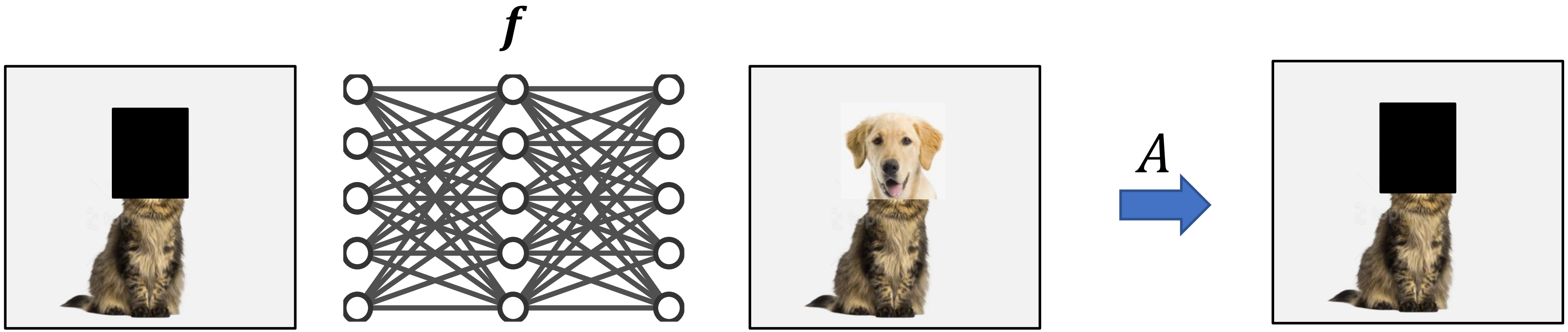
Julián Tachella, CNRS, École Normale Supérieure de Lyon

Learning Approach

Learning from **incomplete noiseless** measurements $\boldsymbol{y}$**?**

$$\underset{f}{\operatorname{argmin}} \sum_i \|\boldsymbol{y}_i - Af(\boldsymbol{y}_i)\|^2$$

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

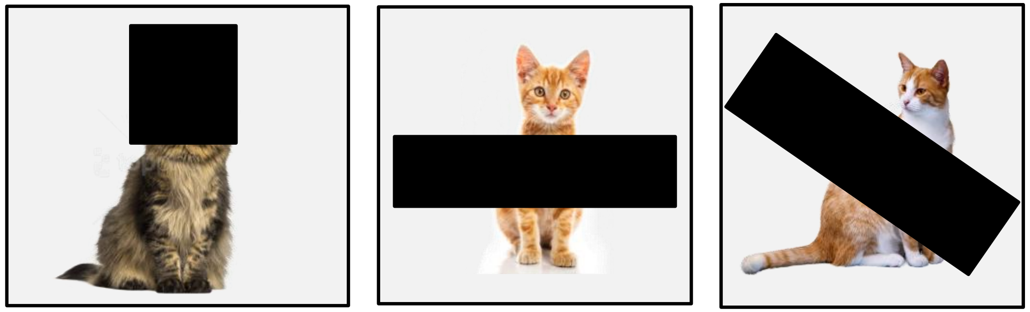


Learning from Measurements

How to learn from only $\boldsymbol{y}$?

- Access multiple operators $\boldsymbol{y}_i = A_{g_i}\boldsymbol{x}_i$ with $g \in \{1, \dots, G\}$
- Each A_g with different nullspace
- Offers the possibility for learning using multiple measurement operators

A_1x_1 A_2x_2 A_3x_3



A_1x_1 A_2x_2 A_3x_3

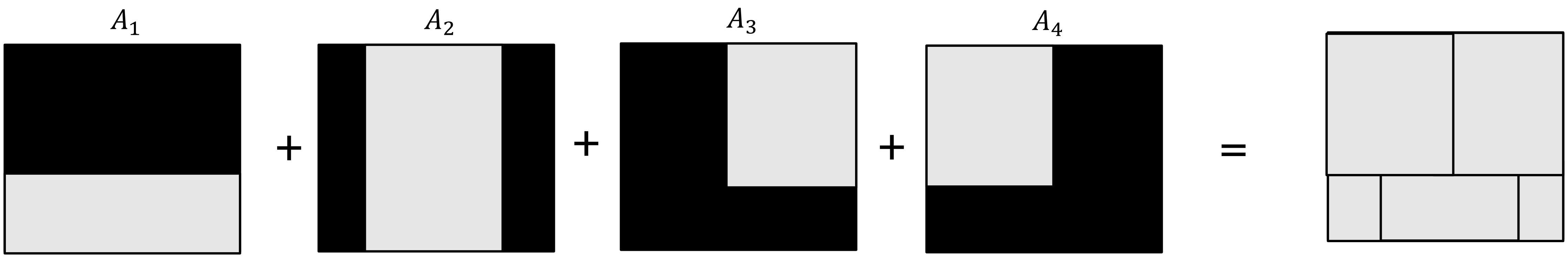
Necessary Condition

Intuition: we need that the operators $A_1, A_2, \ldots A_G$ cover the whole ambient space [T., 2022].

Proposition: Learning reconstruction mapping f from observed measurements possible only if

$$\mathrm{rank}\left(\mathbb{E}_g\, A_g^\top A_g\right) = n$$

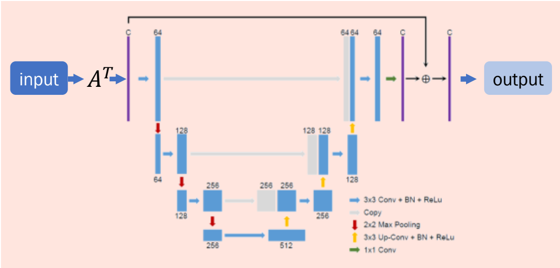
and thus, if $m \geq n/G$.

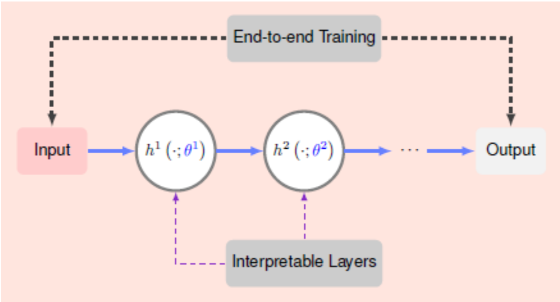


Learning Approach

We will consider networks $\hat{\boldsymbol{x}} = f(\boldsymbol{y}, A)$, where f is also a function of measurement operator e.g.,

- Filtered back projection $f(\boldsymbol{y}, A) = f(A^{\dagger}\boldsymbol{y})$
- Unrolled networks…





- Naïve **measurement consistency** loss:

$$\mathcal{L}_{\mathrm{MC}}(\boldsymbol{y}, f) = \mathbb{E}_{\boldsymbol{y},g}\, ||\boldsymbol{y} - A_g\, f(\boldsymbol{y}, A_g)||^2$$

Without noise, a minimizer is the trivial solution $f(\boldsymbol{y}, A) = A^{\dagger}\boldsymbol{y}$

Noise2Noise Revisited

Artifact2Artifact [Liu et al., 2020]

Assumption: Observe **independent** subsampled pairs:
with $A_a, A_b \sim p_A$ and $\boldsymbol{\epsilon}_a, \boldsymbol{\epsilon}_b \sim p_{\epsilon}$, iid

$$\begin{cases} \boldsymbol{y}_a = A_a \boldsymbol{x} + \boldsymbol{\epsilon}_a \\ \boldsymbol{y}_b = A_b \boldsymbol{x} + \boldsymbol{\epsilon}_b \end{cases}$$

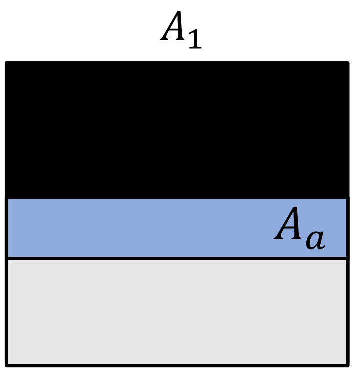
$$\mathcal{L}_{\mathrm{A2A}}(\boldsymbol{y}, f) = \mathbb{E}_{a,b} || \, \boldsymbol{y}_b - A_b \, f(\boldsymbol{y}_a, A_a) ||^2 = \mathbb{E}_a \| \boldsymbol{x} - f(\boldsymbol{y}_a, A_a) \|^2_{\mathbb{E}_b \{A_b^\top A_b\}}$$

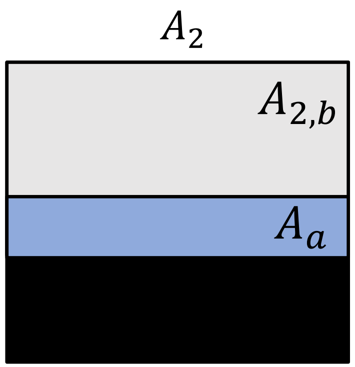
- If $\mathrm{rank}\left(\mathbb{E}_b \, A_b^\top A_b\right) = n$, then $\mathcal{L}_{\mathrm{A2A}}$ is equivalent to a weighted supervised loss with the same minimum

- If A_a, A_b are subsampling of ortho basis, reweighted version possible [Gan et al., 2021]
- The independence assumption is hard to meet in practice.

Measurement Splitting Revisited

Inpainting examples with $G = 2$ operators





$$\begin{aligned} f_1^*(\boldsymbol{y}) &= \operatorname{argmin} \mathbb{E}_{\boldsymbol{y}_1, \boldsymbol{y}_a} ||\boldsymbol{y}_1 - A_1 f(\boldsymbol{y}_a, A_a)||_2^2 \\ &= \mathbb{E}\{A_1^\dagger A_1 \boldsymbol{x} | \boldsymbol{y}_a, A_a\} \end{aligned}$$

$$\begin{aligned} f_2^*(\boldsymbol{y}) &= \operatorname{argmin} \mathbb{E}_{\boldsymbol{y}_1, \boldsymbol{y}_a} ||\boldsymbol{y}_{2,b} - A_{2,b} f(\boldsymbol{y}_a, A_a)||_2^2 \\ &= \mathbb{E}\{A_{2,b}^\dagger A_{2,b} \boldsymbol{x} | \boldsymbol{y}_a, A_a\} \end{aligned}$$

$$\begin{aligned} f^*(\boldsymbol{y}) &= f_1^*(\boldsymbol{y}) + f_2^*(\boldsymbol{y}) \\ &= A_1^\dagger A_1 \mathbb{E}\{\boldsymbol{x} | \boldsymbol{y}_a, A_a\} + A_{2,b}^\dagger A_{2,b} \mathbb{E}\{\boldsymbol{x} | \boldsymbol{y}_a, A_a\} \\ &= \mathbb{E}\{\boldsymbol{x} | \boldsymbol{y}_a, A_a\} \end{aligned}$$

Measurement Splitting Revisited

In general: $\boldsymbol{y}_g = A_g \boldsymbol{x}$ and sample two (possibly overlapping) subsets $S_a, S_b \sim p(S_a, S_b | A_g)$

$$\begin{cases} \boldsymbol{y}_a = S_a \boldsymbol{y}_g, & A_a = S_a A_g \\ \boldsymbol{y}_b = S_b \boldsymbol{y}_g, & A_b = S_b A_g \end{cases}$$

$$\mathcal{L}_{\mathrm{SSDU}}(\boldsymbol{y}, f) = \mathbb{E}_{a,b} || \, \boldsymbol{y}_b - A_b \, f(\boldsymbol{y}_a, A_a) ||^2 = \mathbb{E}_a \mathbb{E}_{a|b} || \, \boldsymbol{x} - \, f(\boldsymbol{y}_a, A_a) ||^2_{\mathbb{E}_b\{A_b^\top A_b | A_a\}}$$

If $\mathbb{E}_b\{A_b^\top A_b | A_a\}$ has full rank, it has same minimizer as supervised loss, $f^*(\boldsymbol{y}_a, A_a) = \mathbb{E}\{\boldsymbol{x} | \boldsymbol{y}_a, A_a\}$

- If $\boldsymbol{y}_g$ has noise, replace measurement consistency by SURE, R2R, cross-validation, etc.
- **Warning:** this condition is sufficient, but **not necessary** in general (more on this later)

Using All Measurements

Can we use all measurements?
Multi Operator Imaging (MOI) [Tachella et al., 2022]

$$\mathcal{L}_{\mathrm{MOI}}(\boldsymbol{y}, f) = \underbrace{\left|\left|\boldsymbol{y} - A_g f\left(\boldsymbol{y}, A_g\right)\right|\right|^2}_{\text{Can be replaced by SURE, R2R, etc.}} + \underbrace{\mathbb{E}_s \left|\left| f(A_s \hat{\boldsymbol{x}}, A_s) - \hat{\boldsymbol{x}}\right|\right|^2}_{\text{Enforces } f\left(A_g \boldsymbol{x}, A_g\right) \approx f(A_s \boldsymbol{x}, A_s)} \quad \text{with } \hat{\boldsymbol{x}} = f\left(\boldsymbol{y}, A_g\right)$$

- The trivial solution $f(\boldsymbol{y}, A) = A^{\dagger}\boldsymbol{y}$ is not generally a minimizer
- Motivated by identifiability theory of low dimensional set (more later)
- Does not have the $\mathrm{rank}\left(\mathbb{E}\, A_b^{\top} A_b |\, A_a\right) = n$ constraint

Magnetic Resonance Imaging

- Unrolled network
- FastMRI dataset (single coil)
- A_g are subsets of Fourier measurements (x4 downsampling)

Measurements y | Signal x | Supervised | Measurement Splitting [Yaman et al., 2019] | MOI

Inpainting

- U-Net network
- CelebA dataset
- A_g are inpainting masks

Measurements y | Signal $\boldsymbol{x}$ | Supervised | AmbientGAN [Bora et al, 2018] | MOI

References

The full reference list for this tutorial can be found here:

https://tachella.github.io/projects/selfsuptutorial/