Self Supervised Learning Methods for Imaging

Part 3: Learning from multiple operators

Mike Davies, University of Edinburgh

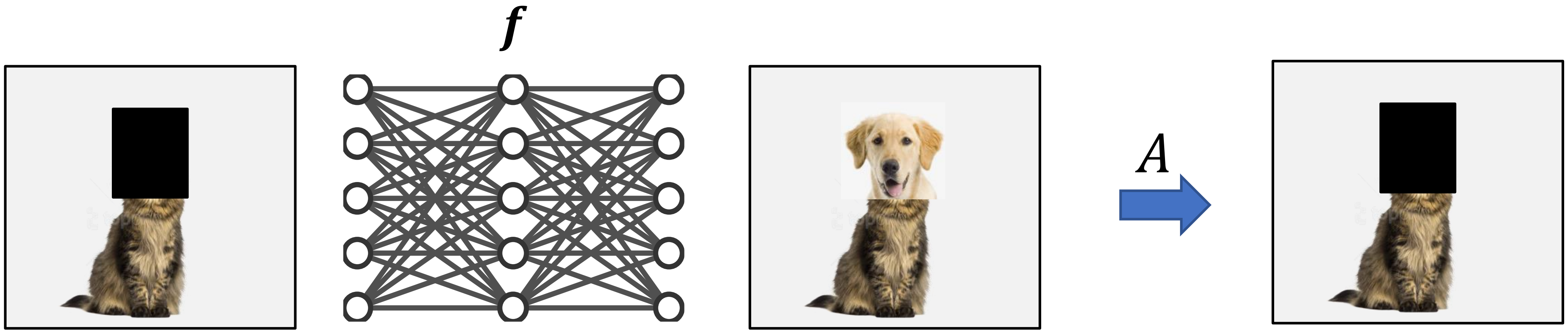
Julián Tachella, CNRS, École Normale Supérieure de Lyon

Learning Approach

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

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

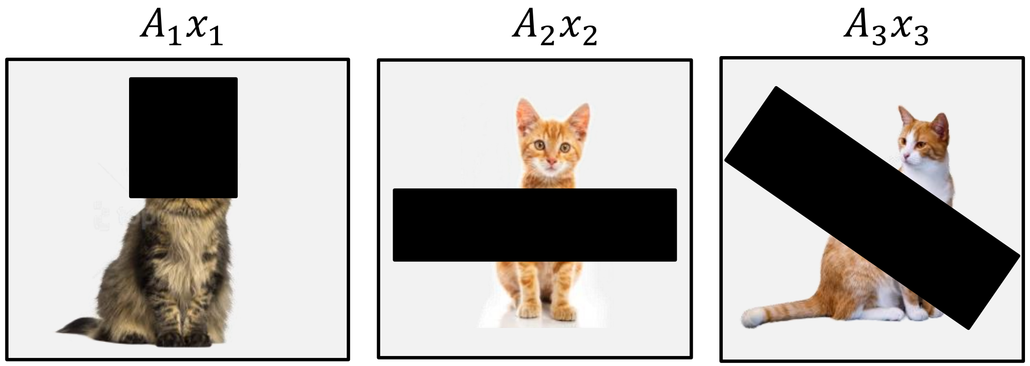
Proposition: Any reconstruction function $f(\boldsymbol{y}) = A^{\dagger}\boldsymbol{y} + g(\boldsymbol{y})$ is measurement consistent where $g: \mathbb{R}^m \mapsto \mathcal{N}_A$ is any function whose image belongs to the nullspace of A.

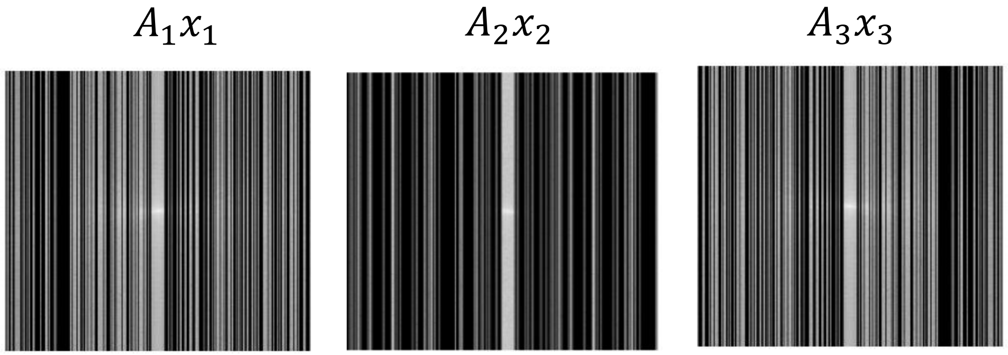


Learning from Measurements

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

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





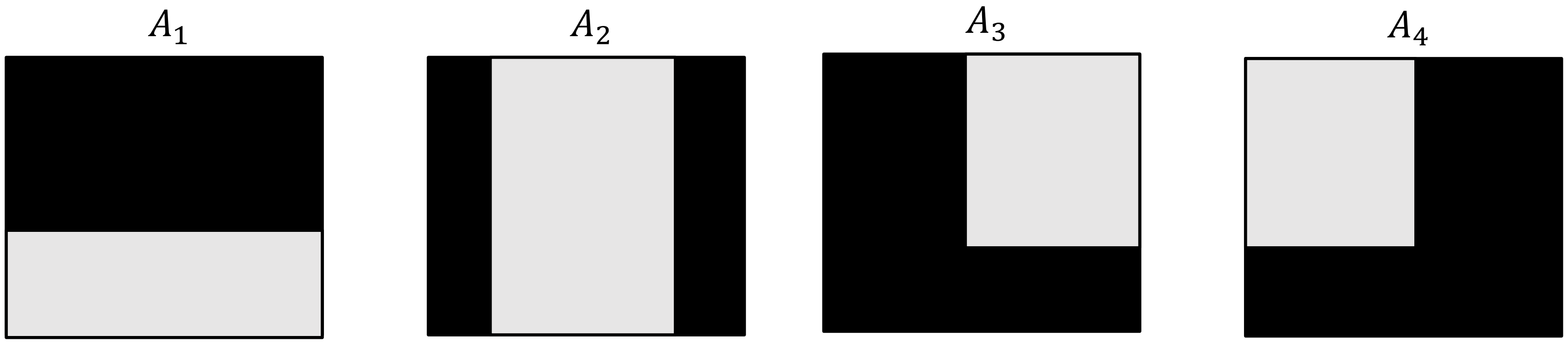
Necessary Condition

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

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

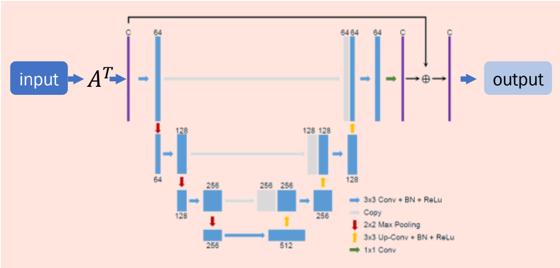
and thus, if $m \geq n/G$.

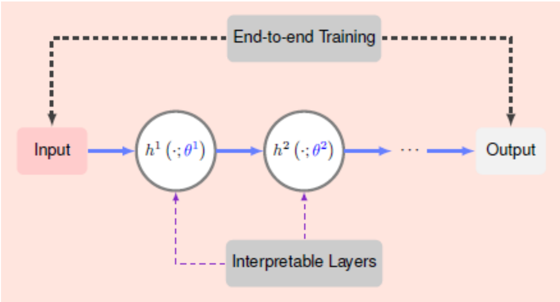


Learning Approach

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:

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

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:

$$\boldsymbol{y}_a = A_a \boldsymbol{x} + \boldsymbol{\epsilon}_a$$

and

$$\boldsymbol{y}_b = A_b \boldsymbol{x} + \boldsymbol{\epsilon}_b$$

with $A_a, A_b \sim P_A$, $\boldsymbol{\epsilon}_a, \boldsymbol{\epsilon}_b \sim P_\epsilon$, i.i.d**.**

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

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

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

- 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.

Multiplicative Noisier2Noise

Assumption: [Moran et al., 2020] considered a Bernoulli multiplicative noise model which can also be viewed as noiseless subsampling

$$\boldsymbol{y} = \mathrm{diag}(\boldsymbol{n})\boldsymbol{x}$$

with $n_i \in \{0,1\}$ and $\mathbb{P}(n_i = 0) = p$,

Bernoulli-Noisier2Noise: synthesise:

$$\boldsymbol{z} = \mathrm{diag}(\boldsymbol{m})\boldsymbol{y} = \mathrm{diag}(\boldsymbol{m})\mathrm{diag}(\boldsymbol{n})\boldsymbol{x}$$

with $m_i \in \{0,1\}$ and $\mathbb{P}(m_i = 0) = q$,

Minimise: $\mathcal{L}_{\mathrm{BNr2N}}(\boldsymbol{y}, f) = \mathbb{E}_{\boldsymbol{x},\boldsymbol{m},\boldsymbol{n}}\left\|\boldsymbol{y} - f\big(\boldsymbol{z}, \mathrm{diag}(\boldsymbol{m})\big)\right\|_2^2$ $\longrightarrow$ MMSE estimator: $f\big(\boldsymbol{z}, \mathrm{diag}(\boldsymbol{m})\big) \approx \mathbb{E}\{\boldsymbol{y}|\boldsymbol{z},\boldsymbol{m}\}$

Then

$$\mathbb{E}\{\boldsymbol{x}|\boldsymbol{z},\boldsymbol{m}\} = (1-k)^{-1}(\mathbb{E}\{\boldsymbol{y}|\boldsymbol{z},\boldsymbol{m}\} - k\boldsymbol{z})$$

with $k = \frac{p}{p+q-pq}$

Measurement Splitting

What happens if we do not have pairs, i.e. $\boldsymbol{y}_g = A_g \boldsymbol{x} + \boldsymbol{\epsilon}$

Self-Supervised Learning via Data Undersampling (SSDU) [Yaman et al., 2019]:

randomly split $\boldsymbol{y} = \begin{bmatrix} \boldsymbol{y}_a \\ \boldsymbol{y}_b \end{bmatrix}$ and $A_g = \begin{bmatrix} A_{g,a} \\ A_{g,b} \end{bmatrix}$ at each sample

$$\mathcal{L}_{\text{SSDU}}(\boldsymbol{y}, f) = \mathbb{E}_{g,a,b} || \, \boldsymbol{y}_b - A_{g,b} \, f\left(\boldsymbol{y}_a, A_{g,a}\right) ||^2$$

- The trivial solution $f(\boldsymbol{y}, A) = A^{\dagger}\boldsymbol{y}$ is not a minimizer
- Choice of splitting is important!
    - $\boldsymbol{A}_{g,a}$ should keep most measurements
- Variable density Noisier2Noise interpretation [Millard & Chiew, 2023]
    - In $\boldsymbol{\epsilon} = \boldsymbol{0}$, setting, with measurement consistent f, we get: $f\left(\boldsymbol{y}_a, A_{g,a}\right) = \mathbb{E}\{\boldsymbol{x} | \boldsymbol{y}_a, A_{g,a}\}$
- Does not denoise $\boldsymbol{y}_a$ but can be modified to do so [Millard & Chiew, 2024]

Measurement Splitting Revisited

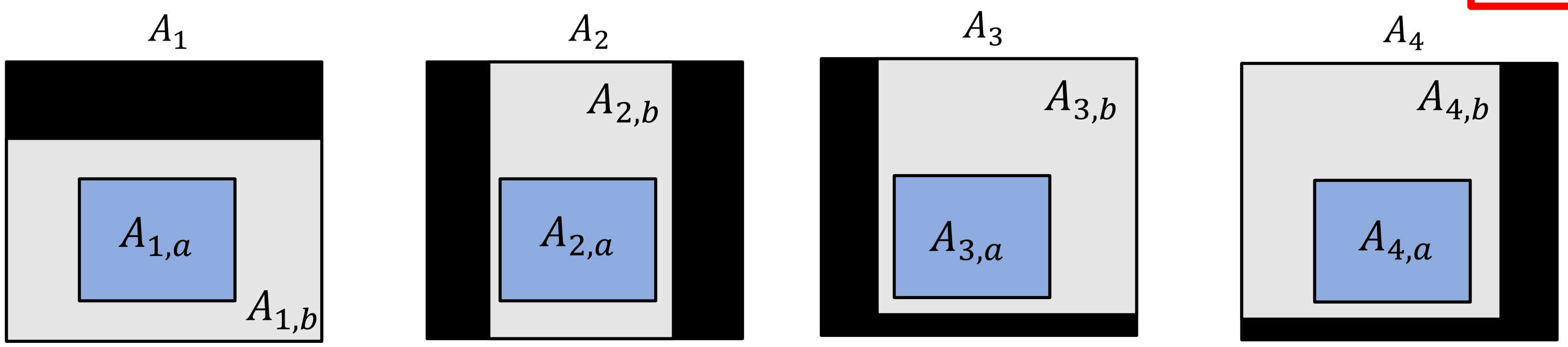
Same (noiseless) setting: $\boldsymbol{y}_g = A_g\boldsymbol{x}$ and randomly split $\boldsymbol{y} = \begin{bmatrix}\boldsymbol{y}_a \\ \boldsymbol{y}_b\end{bmatrix}$ and $A_g = \begin{bmatrix}A_{g,a} \\ A_{g,b}\end{bmatrix}$ at each sample.

Proposition [Daras et al., 2023]: if $\mathbb{E}_{A_g|A_{g,a}}\, A_g^\top A_g$ has full rank, then measurement splitting loss:

$$\mathcal{L}_{MS}(\boldsymbol{y}, f) = \mathbb{E}_{g,a,b}\left\| \boldsymbol{y} - A_g f(\boldsymbol{y}_a, A_{g,a})\right\|^2$$

has as minimizer $f(\boldsymbol{y}_a, A_{g,a}) = \mathbb{E}\{\boldsymbol{x}|\boldsymbol{y}_a, A_{g,a}\}$ as the supervised loss in expectation.

Unlike SSDU this includes **all** of $\boldsymbol{y}$. It enforces data consistency on $\boldsymbol{y}_a$



Warning: This condition is sufficient, but **not necessary** in general (more on this later)

Measurement Splitting Revisited

The MS loss has same minimiser as supervised loss that uses less measurements [Daras et al. 2023]:

Write $r(\boldsymbol{y}_a, A_{g,a}) = f(\boldsymbol{y}_a, A_{g,a}) - \mathbb{E}\{\boldsymbol{x}|\ \boldsymbol{y}_a, A_{g,a}\}$ as deviation from MMSE loss.

$$\mathcal{L}_{\text{MS}}(\boldsymbol{y}, f) = \mathbb{E}\left\|(\boldsymbol{y} - A_g\mathbb{E}\{\boldsymbol{x}|\ \boldsymbol{y}_a, A_{g,a}\}) - A_g\, r(\boldsymbol{y}_a, A_{g,a})\right\|^2$$

$$= \underbrace{\mathbb{E}\left\|\boldsymbol{y} - A_g\mathbb{E}\{\boldsymbol{x}|\ \boldsymbol{y}_a, A_{g,a}\}\right\|^2}_{①} + \underbrace{\mathbb{E}\left\|A_g\, r(\boldsymbol{y}_a, A_{g,a})\right\|^2}_{②} - \underbrace{\mathbb{E}\left[(\boldsymbol{y} - A_g\mathbb{E}\{\boldsymbol{x}|\ \boldsymbol{y}_a, A_{g,a}\})^\top A_g\, r(\boldsymbol{y}_a, A_{g,a})\right]}_{③}$$

① = irreducible error (const.); ② = deviation from optimal (= 0 if $\mathbb{E}_{A_g|A_{g,a}} A_g^\top A_g$ full rank) ③ = correlation term = 0

Final thought: if there is known Gaussian noise, then we can simply replace $\mathcal{L}_{\text{MS}}(\boldsymbol{y}, f)$ by its SURE equivalent or Noisier2Noise…

Using All Measurements

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

$$\mathcal{L}_{MOI}(\boldsymbol{y}, f) = \left|\left|\boldsymbol{y} - A_g f\left(\boldsymbol{y}, A_g\right)\right|\right|^2 + \sum_s ||f(A_s\hat{\boldsymbol{x}}, A_s) - \hat{\boldsymbol{x}}||^2 \quad \text{with } \hat{\boldsymbol{x}} = f\left(\boldsymbol{y}, A_g\right)$$

Can be replaced by SURE, R2R, etc.

Enforces $f\left(A_g\boldsymbol{x}, A_g\right) \approx f(A_s\boldsymbol{x}, A_s)$

- 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 $\mathrm{rank}\left(\mathbb{E}_{A_g|A_{g,a}} A_g^{\top} A_g\right) = n$ constraint

Inpainting

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

Measurements y | Signal x | Supervised | AmbientGAN [Bora et al, 2018] | MOI

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

References

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

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