Equivariant splitting: self-supervised learning from incomplete data

Victor Sechaud[*1], Jérémy Scanvic[*1,2], Quentin Barthélemy[2], Patrice Abry[1], Julián Tachella[1]

[1]CNRS, ENS de Lyon
[2]Prysm, Lyon, France

2025-11-07

[*]Equal contribution

Introduction

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\varepsilon}$$

$$\mathcal{X} \subset \mathbb{R}^n$$

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\varepsilon}$$

Introduction

$$\boldsymbol{y} = \underbrace{\boldsymbol{A}}_{\mathbb{R}^{m\times n}}\overbrace{\boldsymbol{x}}^{\mathcal{X} \subset \mathbb{R}^{n}} + \boldsymbol{\varepsilon}$$

Introduction

$$\mathcal{X} \subset \mathbb{R}^n$$

$$\mathcal{Y} \subseteq \mathbb{R}^m \qquad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\varepsilon}$$

$$\mathbb{R}^{m \times n}$$

Introduction

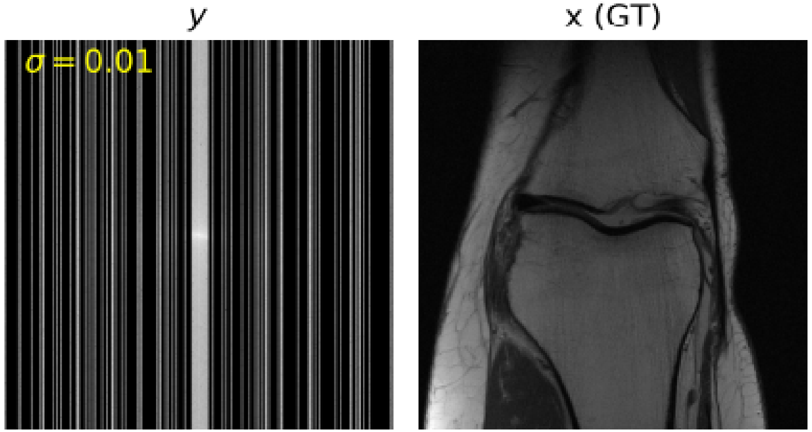
$$\mathcal{X} \subset \mathbb{R}^n$$

$$\mathcal{Y} \subseteq \mathbb{R}^m \quad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\varepsilon} \longrightarrow \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I}_n)$$

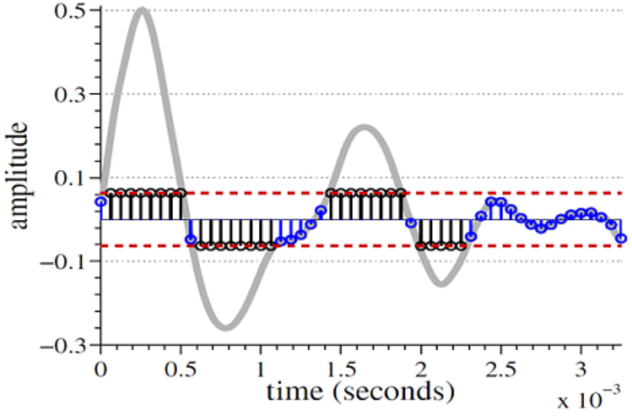
$$\mathbb{R}^{m \times n}$$

$$\mathcal{Y} \subseteq \mathbb{R}^m \quad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\varepsilon} \longrightarrow \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I}_n)$$

$$\mathcal{X} \subset \mathbb{R}^n \qquad \mathbb{R}^{m \times n}$$



Linear: MRI, inpainting, ...



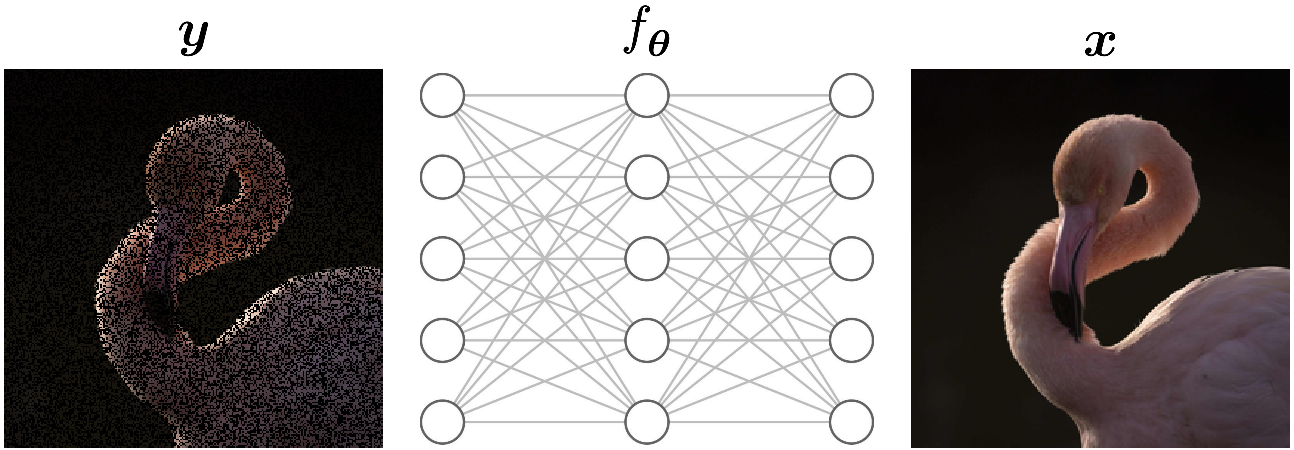
Non-linear: declipping, phase retrieval

Deep learning approches

- Deep learning is now the state of the art.

Objective

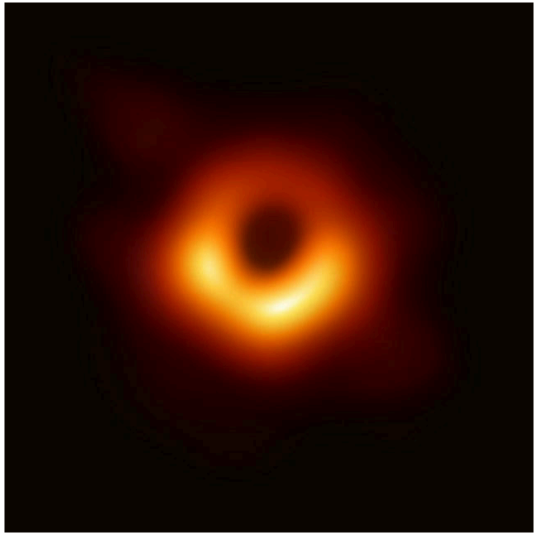
- **Learn** inverse forward operator of $\boldsymbol{A}$, $f_\theta(\boldsymbol{y}, \boldsymbol{A}) \approx \boldsymbol{x}$.
- **Using** a dataset $\{(\boldsymbol{x}_i, \boldsymbol{y}_i)\}_{i \in I}$ + a neural network.

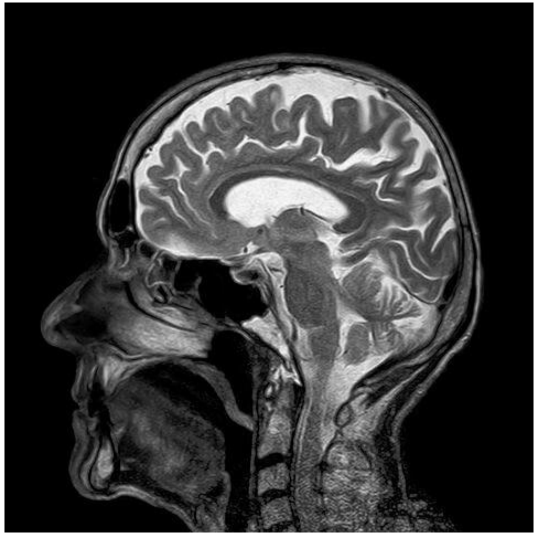


$$\underset{\theta}{\operatorname{argmin}} \sum_{i \in I} \left\| f_\theta(\boldsymbol{y}_i, \boldsymbol{A}) - \boldsymbol{x}_i \right\|^2$$

Limits of the method

- Training and testing datasets can be very different.
- We need a large set of $\{\boldsymbol{x}_i\}_{i\in I}$ (ground-truth).





Astronomical and medical imaging.

Self-supervised learning

- **Dataset:** $\cancel{(\boldsymbol{x}_i, \boldsymbol{y}_i)_{i \in I}}$ only $(\boldsymbol{y}_i)_{i \in I}$.
- What can we do with $\boldsymbol{y}_i$?

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \Rightarrow \boldsymbol{y} \approx \boldsymbol{A} f_{\boldsymbol{\theta}}(\boldsymbol{y}, \boldsymbol{A})$$

Self-supervised learning

- **Dataset:** $\cancel{(\boldsymbol{x}_i, \boldsymbol{y}_i)_{i \in I}}$ only $(\boldsymbol{y}_i)_{i \in I}$.
- What can we do with $\boldsymbol{y}_i$?

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \Rightarrow \boldsymbol{y} \approx \boldsymbol{A} f_{\boldsymbol{\theta}}(\boldsymbol{y}, \boldsymbol{A})$$

$$\underset{\boldsymbol{\theta}}{\operatorname{argmin}} \sum_{i \in I} \left\| \boldsymbol{y}_i - \boldsymbol{A} f_{\boldsymbol{\theta}}(\boldsymbol{y}_i, \boldsymbol{A}) \right\|^2$$

Self-supervised learning

- **Dataset:** $\cancel{(\boldsymbol{x}_i, \boldsymbol{y}_i)_{i \in I}}$ only $(\boldsymbol{y}_i)_{i \in I}$.
- What can we do with $\boldsymbol{y}_i$?

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \Rightarrow \boldsymbol{y} \approx \boldsymbol{A} f_{\boldsymbol{\theta}}(\boldsymbol{y}, \boldsymbol{A})$$

$$\operatorname*{argmin}_{\boldsymbol{\theta}} \sum_{i \in I} \|\boldsymbol{y}_i - \boldsymbol{A} f_{\boldsymbol{\theta}}(\boldsymbol{y}_i, \boldsymbol{A})\|^2$$

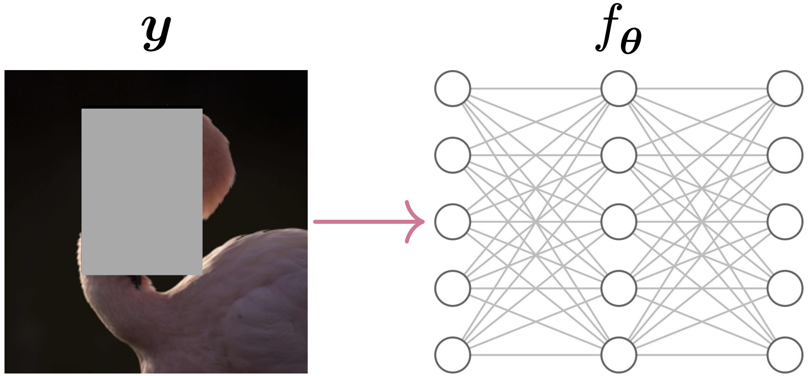
$\boldsymbol{y}$

Self-supervised learning

- **Dataset:** $\cancel{(\boldsymbol{x}_i, \boldsymbol{y}_i)_{i \in I}}$ only $(\boldsymbol{y}_i)_{i \in I}$.
- What can we do with $\boldsymbol{y}_i$?

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \Rightarrow \boldsymbol{y} \approx \boldsymbol{A} f_{\boldsymbol{\theta}}(\boldsymbol{y}, \boldsymbol{A})$$

$$\underset{\boldsymbol{\theta}}{\operatorname{argmin}} \sum_{i \in I} \|\boldsymbol{y}_i - \boldsymbol{A} f_{\boldsymbol{\theta}}(\boldsymbol{y}_i, \boldsymbol{A})\|^2$$

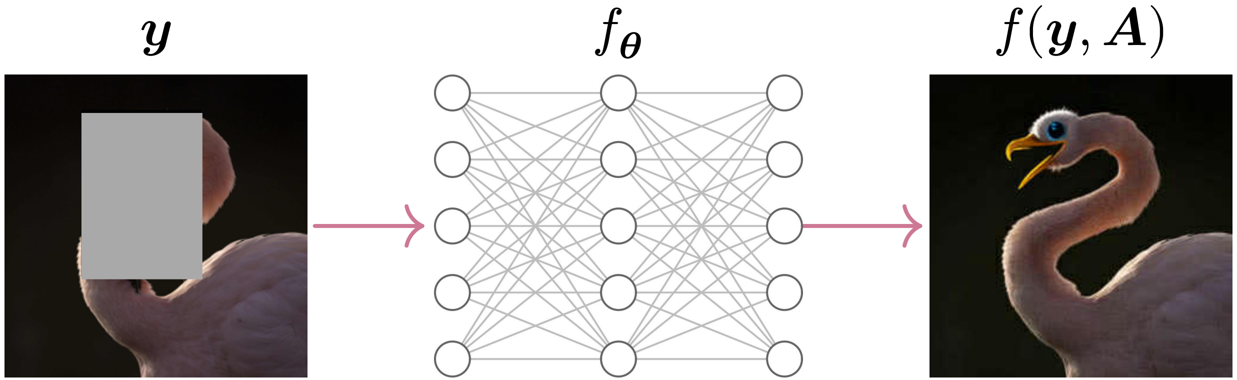


Self-supervised learning

- **Dataset:** $\cancel{(\boldsymbol{x}_i, \boldsymbol{y}_i)_{i \in I}}$ only $(\boldsymbol{y}_i)_{i \in I}$.
- What can we do with $\boldsymbol{y}_i$?

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \Rightarrow \boldsymbol{y} \approx \boldsymbol{A} f_{\boldsymbol{\theta}}(\boldsymbol{y}, \boldsymbol{A})$$

$$\underset{\boldsymbol{\theta}}{\operatorname{argmin}} \sum_{i \in I} \|\boldsymbol{y}_i - \boldsymbol{A} f_{\boldsymbol{\theta}}(\boldsymbol{y}_i, \boldsymbol{A})\|^2$$

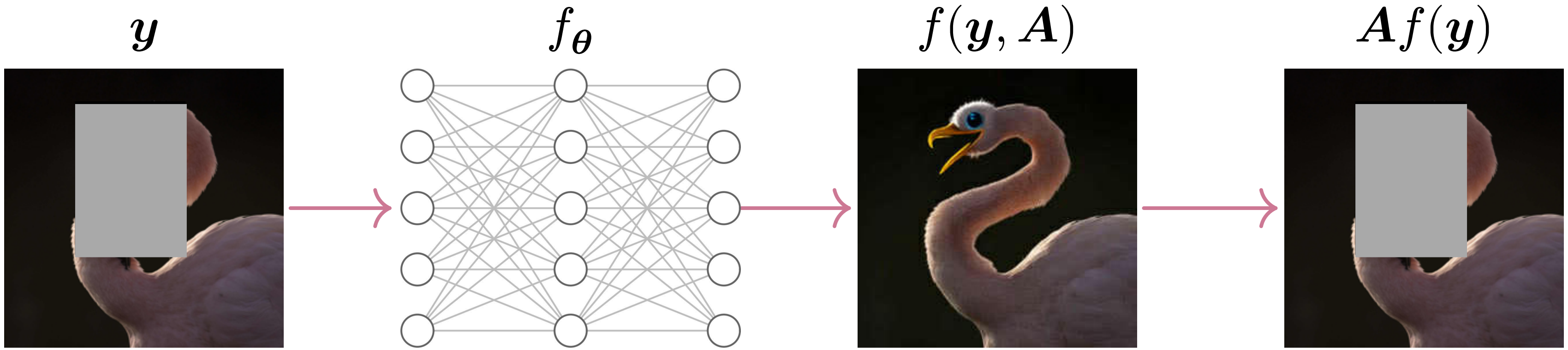


Self-supervised learning

- **Dataset:** $\cancel{(\boldsymbol{x}_i, \boldsymbol{y}_i)_{i\in I}}$ only $(\boldsymbol{y}_i)_{i\in I}$.
- What can we do with $\boldsymbol{y}_i$?

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \Rightarrow \boldsymbol{y} \approx \boldsymbol{A} f_{\boldsymbol{\theta}}(\boldsymbol{y}, \boldsymbol{A})$$

$$\underset{\boldsymbol{\theta}}{\operatorname{argmin}} \sum_{i\in I} \|\boldsymbol{y}_i - \boldsymbol{A} f_{\boldsymbol{\theta}}(\boldsymbol{y}_i, \boldsymbol{A})\|^2$$

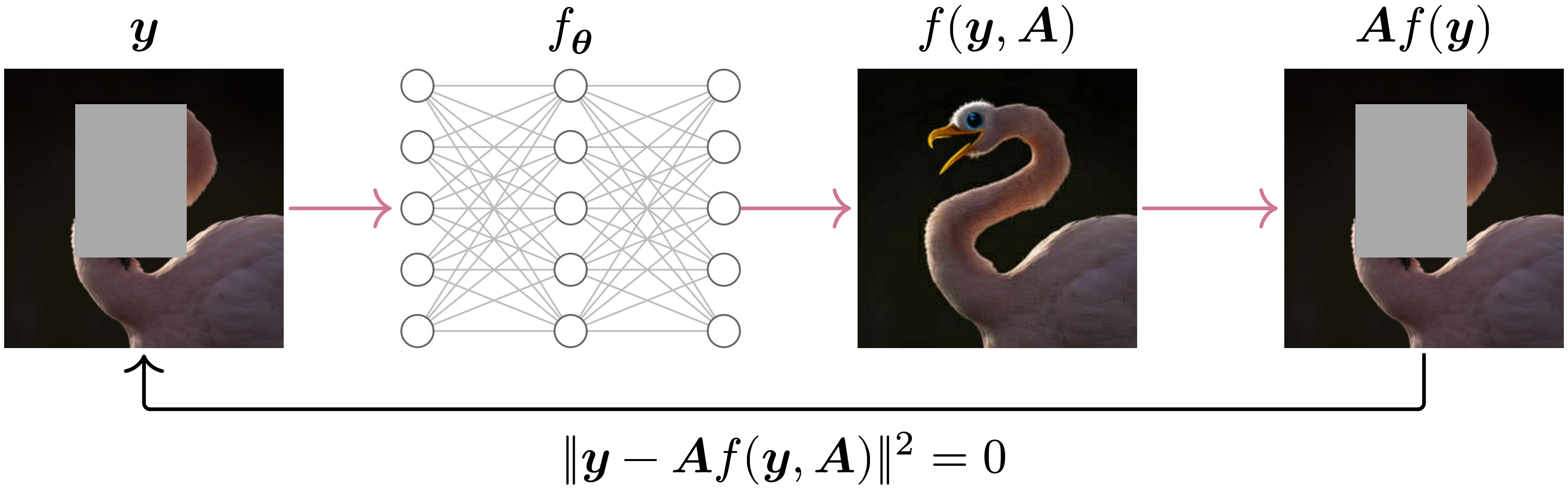


Self-supervised learning

- **Dataset:** $\cancel{(\boldsymbol{x}_i, \boldsymbol{y}_i)_{i \in I}}$ only $(\boldsymbol{y}_i)_{i \in I}$.
- What can we do with $\boldsymbol{y}_i$?

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \Rightarrow \boldsymbol{y} \approx \boldsymbol{A} f_{\boldsymbol{\theta}}(\boldsymbol{y}, \boldsymbol{A})$$

$$\underset{\boldsymbol{\theta}}{\operatorname{argmin}} \sum_{i \in I} \|\boldsymbol{y}_i - \boldsymbol{A} f_{\boldsymbol{\theta}}(\boldsymbol{y}_i, \boldsymbol{A})\|^2$$



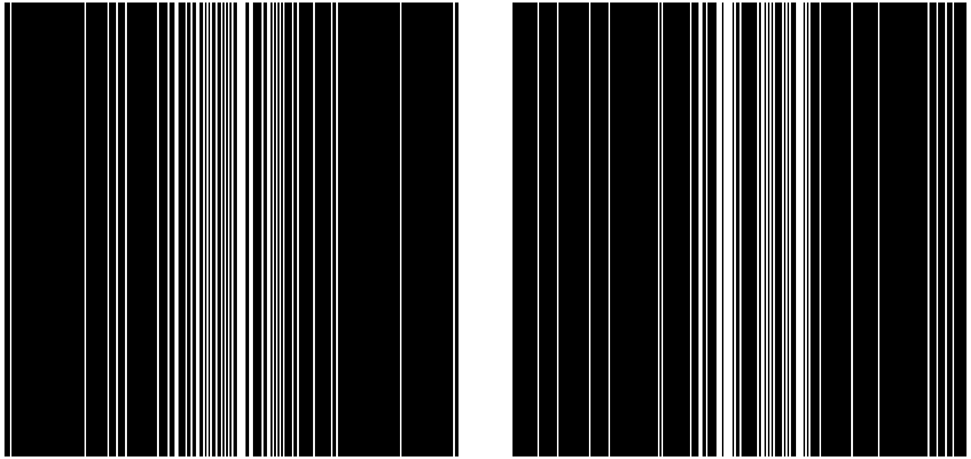
Multiple operators

- Leveraging **multiple** operators: $\boldsymbol{A} \sim p(\boldsymbol{A})$.
- **Differ** for each measurements $\boldsymbol{y} \sim p(\boldsymbol{y} \mid \boldsymbol{A}\boldsymbol{x})$.

Multiple operators

- Leveraging **multiple** operators: $\boldsymbol{A} \sim p(\boldsymbol{A})$.
- **Differ** for each measurements $\boldsymbol{y} \sim p(\boldsymbol{y} \mid \boldsymbol{A}\boldsymbol{x})$.

Inpainting



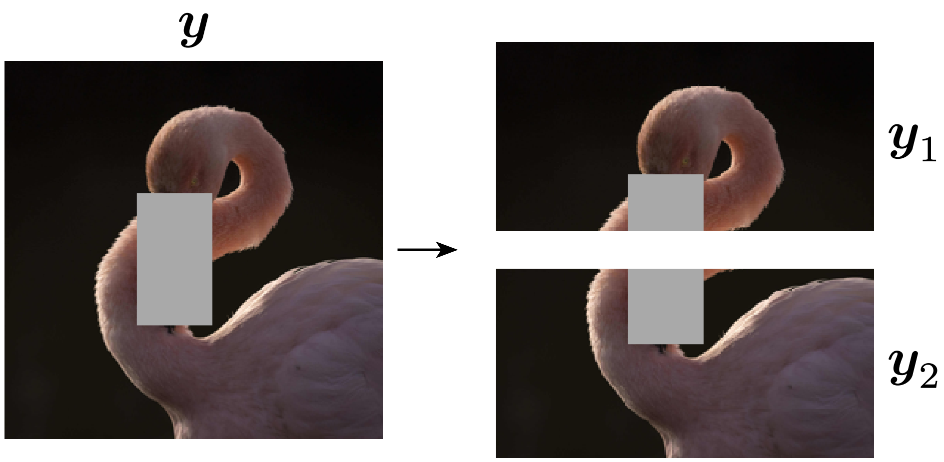
MRI

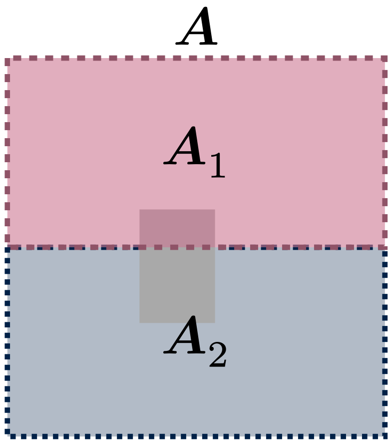
Measurements splitting

- Leveraging **multiple** operators: $\boldsymbol{y} \sim p(\boldsymbol{y} \mid \boldsymbol{A}\boldsymbol{x})$ with $\boldsymbol{A} \sim p(\boldsymbol{A})$.
- **Split** measurements $\boldsymbol{y} = [\boldsymbol{y}_1, \boldsymbol{y}_2]$ with **corresponding** operators $\boldsymbol{A} = [\boldsymbol{A}_1, \boldsymbol{A}_2]$.

$$f^* \in \operatorname*{argmin}_f \mathbb{E}_{\boldsymbol{y},\boldsymbol{A}}\{\mathcal{L}_{\text{SPLIT}}(\boldsymbol{y}, \boldsymbol{A}, f)\}$$

$$\text{with} \quad \mathcal{L}_{\text{SPLIT}}(\boldsymbol{y}, \boldsymbol{A}, f) = \mathbb{E}_{\boldsymbol{y}_1, \boldsymbol{A}_1 \mid \boldsymbol{y}, \boldsymbol{A}}\left[\|\boldsymbol{A} f(\boldsymbol{y}_1, \boldsymbol{A}_1) - \boldsymbol{y}\|^2\right]$$



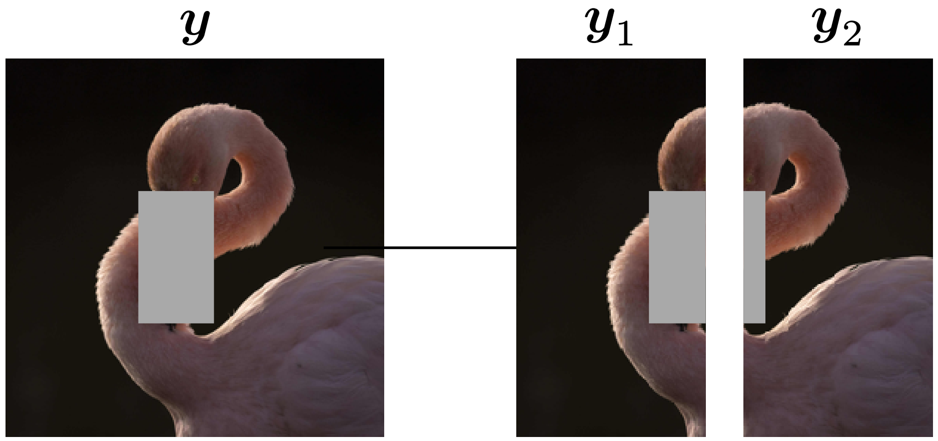


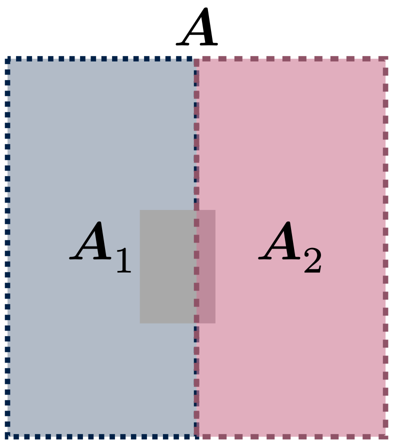
Measurements splitting

- Leveraging **multiple** operators: $\boldsymbol{y} \sim p(\boldsymbol{y} \mid \boldsymbol{A}\boldsymbol{x})$ with $\boldsymbol{A} \sim p(\boldsymbol{A})$.
- **Split** measurements $\boldsymbol{y} = [\boldsymbol{y}_1, \boldsymbol{y}_2]$ with **corresponding** operators $\boldsymbol{A} = [\boldsymbol{A}_1, \boldsymbol{A}_2]$.

$$f^* \in \operatorname*{argmin}_f \mathbb{E}_{\boldsymbol{y},\boldsymbol{A}} \{\mathcal{L}_{\mathrm{SPLIT}}(\boldsymbol{y}, \boldsymbol{A}, f)\}$$

$$\text{with} \quad \mathcal{L}_{\mathrm{SPLIT}}(\boldsymbol{y}, \boldsymbol{A}, f) = \mathbb{E}_{\boldsymbol{y}_1,\boldsymbol{A}_1 \mid \boldsymbol{y},\boldsymbol{A}} \left[\|\boldsymbol{A} f(\boldsymbol{y}_1, \boldsymbol{A}_1) - \boldsymbol{y}\|^2 \right]$$





A visual interpretation of $\boldsymbol{Q}_{\boldsymbol{A}_1}$

Solution with $\boldsymbol{Q}_{\boldsymbol{A}_1} \triangleq \mathbb{E}_{\boldsymbol{A} \mid \boldsymbol{A}_1}\{\boldsymbol{A}^\top \boldsymbol{A}\}$ and v any function in:

$$f^*(\boldsymbol{y}_1, \boldsymbol{A}_1) = \boldsymbol{Q}_{\boldsymbol{A}_1}^\dagger \boldsymbol{Q}_{\boldsymbol{A}_1} \underbrace{\mathbb{E}_{\boldsymbol{x} \mid \boldsymbol{y}_1, \boldsymbol{A}_1}\{\boldsymbol{x}\}}_{\text{MMSE estimator}} + \underbrace{\left(\boldsymbol{I} - \boldsymbol{Q}_{\boldsymbol{A}_1}^\dagger \boldsymbol{Q}_{\boldsymbol{A}_1}\right) v(\boldsymbol{y}_1)}_{\text{Error}}$$

A visual interpretation of $\boldsymbol{Q}_{\boldsymbol{A}_1}$

Solution with $\boldsymbol{Q}_{\boldsymbol{A}_1} \triangleq \mathbb{E}_{\boldsymbol{A} \mid \boldsymbol{A}_1}\{\boldsymbol{A}^\top \boldsymbol{A}\}$ and v any function in:

$$f^*(\boldsymbol{y}_1, \boldsymbol{A}_1) = \boldsymbol{Q}^\dagger_{\boldsymbol{A}_1} \boldsymbol{Q}_{\boldsymbol{A}_1} \underbrace{\mathbb{E}_{\boldsymbol{x} \mid \boldsymbol{y}_1, \boldsymbol{A}_1}\{\boldsymbol{x}\}}_{\text{MMSE estimator}} + \underbrace{\left(\boldsymbol{I} - \boldsymbol{Q}^\dagger_{\boldsymbol{A}_1} \boldsymbol{Q}_{\boldsymbol{A}_1}\right) v(\boldsymbol{y}_1)}_{\text{Error}}$$

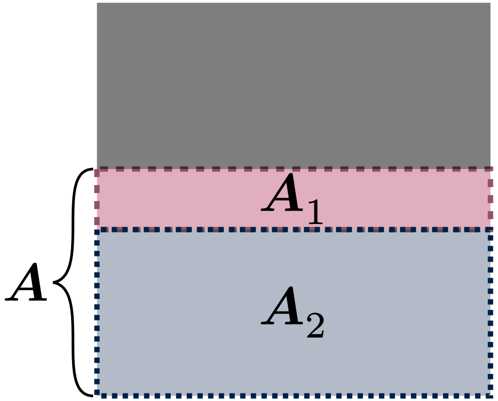
If $\boldsymbol{Q}_{\boldsymbol{A}_1}$ **invertible**: $\quad f^*(\boldsymbol{y}_1, \boldsymbol{A}_1) = \mathbb{E}_{\boldsymbol{x} \mid \boldsymbol{y}_1, \boldsymbol{A}_1}\{\boldsymbol{x}\}$

A visual interpretation of $Q_{\boldsymbol{A}_1}$

Solution with $Q_{\boldsymbol{A}_1} \triangleq \mathbb{E}_{\boldsymbol{A} \mid \boldsymbol{A}_1}\{\boldsymbol{A}^\top \boldsymbol{A}\}$ and v any function in:

$$f^*(\boldsymbol{y}_1, \boldsymbol{A}_1) = Q^\dagger_{\boldsymbol{A}_1} Q_{\boldsymbol{A}_1} \underbrace{\mathbb{E}_{\boldsymbol{x} \mid \boldsymbol{y}_1, \boldsymbol{A}_1}\{\boldsymbol{x}\}}_{\text{MMSE estimator}} + \underbrace{\left(\boldsymbol{I} - Q^\dagger_{\boldsymbol{A}_1} Q_{\boldsymbol{A}_1}\right) v(\boldsymbol{y}_1)}_{\text{Error}}$$

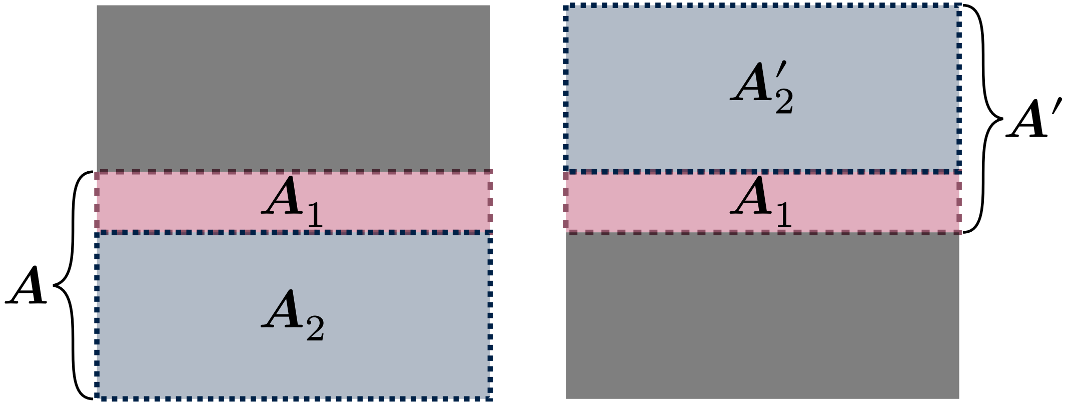
If $Q_{\boldsymbol{A}_1}$ **invertible**: $\quad f^*(\boldsymbol{y}_1, \boldsymbol{A}_1) = \mathbb{E}_{\boldsymbol{x} \mid \boldsymbol{y}_1, \boldsymbol{A}_1}\{\boldsymbol{x}\}$



Solution with $Q_{A_1} \triangleq \mathbb{E}_{A \mid A_1}\{A^\top A\}$ and v any function in:

$$f^*(y_1, A_1) = Q_{A_1}^\dagger Q_{A_1} \underbrace{\mathbb{E}_{x \mid y_1, A_1}\{x\}}_{\text{MMSE estimator}} + \underbrace{\left(I - Q_{A_1}^\dagger Q_{A_1}\right) v(y_1)}_{\text{Error}}$$

If Q_{A_1} **invertible**: $\quad f^*(y_1, A_1) = \mathbb{E}_{x \mid y_1, A_1}\{x\}$

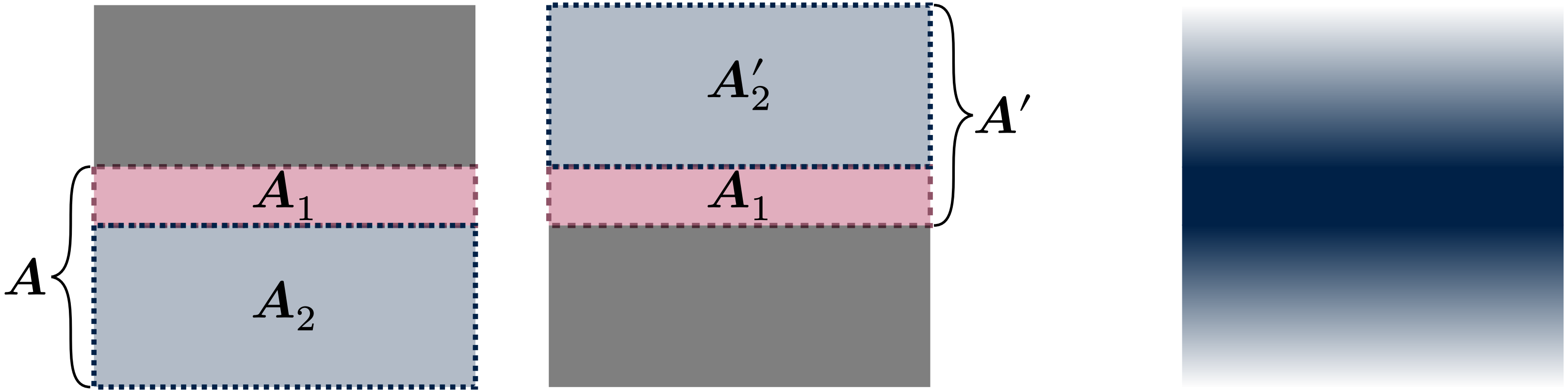


A visual interpretation of Q_{A_1}

Solution with $Q_{A_1} \triangleq \mathbb{E}_{A \mid A_1}\{A^\top A\}$ and v any function in:

$$f^*(y_1, A_1) = Q_{A_1}^\dagger Q_{A_1} \underbrace{\mathbb{E}_{x \mid y_1, A_1}\{x\}}_{\text{MMSE estimator}} + \underbrace{\left(I - Q_{A_1}^\dagger Q_{A_1}\right) v(y_1)}_{\text{Error}}$$

If Q_{A_1} **invertible**: $\quad f^*(y_1, A_1) = \mathbb{E}_{x \mid y_1, A_1}\{x\}$



$\mathbb{E}_{A \mid A_1}[A^\top A]$

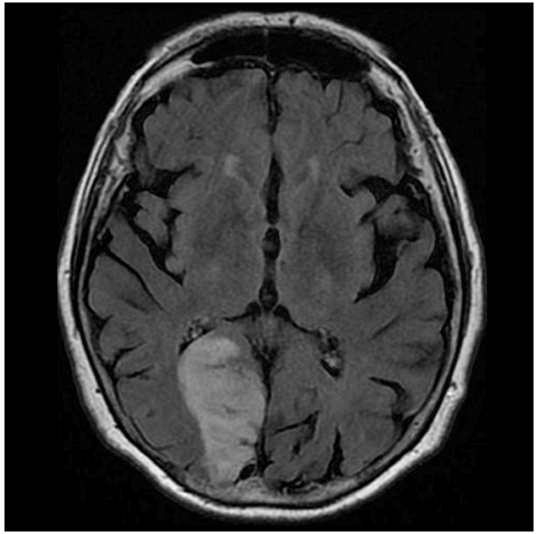
Invariance

- **A priori:** Invariance to certain transformations: $\forall g \in \mathcal{G} : p\big(\boldsymbol{T}_g \boldsymbol{x}\big) = p(\boldsymbol{x})$.

Invariance

- **A priori:** Invariance to certain transformations: $\forall g \in \mathcal{G} : p\big(\boldsymbol{T}_g \boldsymbol{x}\big) = p(\boldsymbol{x})$.

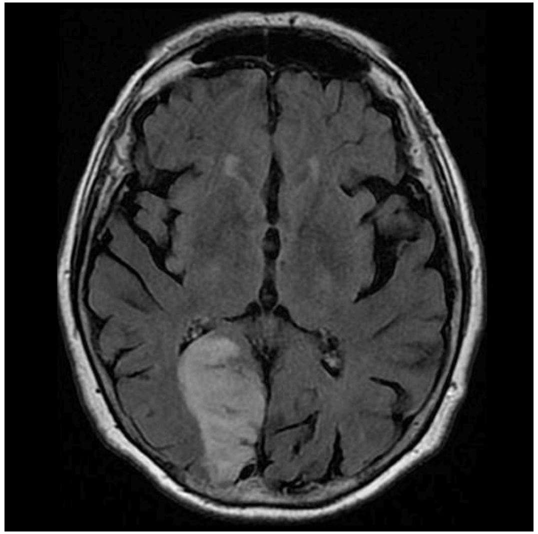
$$\boldsymbol{x} \in \mathcal{X}$$



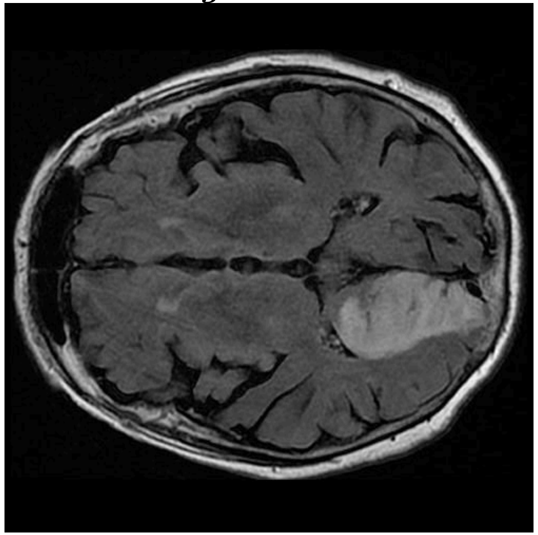
Invariance

- **A priori:** Invariance to certain transformations: $\forall g \in \mathcal{G} : p\big(\boldsymbol{T}_g \boldsymbol{x}\big) = p(\boldsymbol{x})$.

$\boldsymbol{x} \in \mathcal{X}$



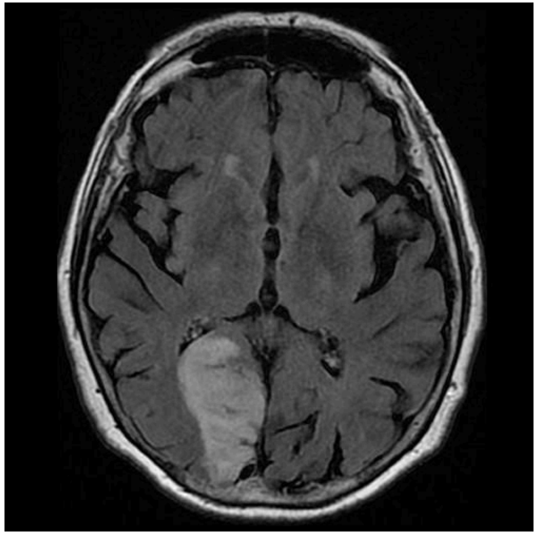
$\boldsymbol{T}_g \boldsymbol{x} \in \mathcal{X}$



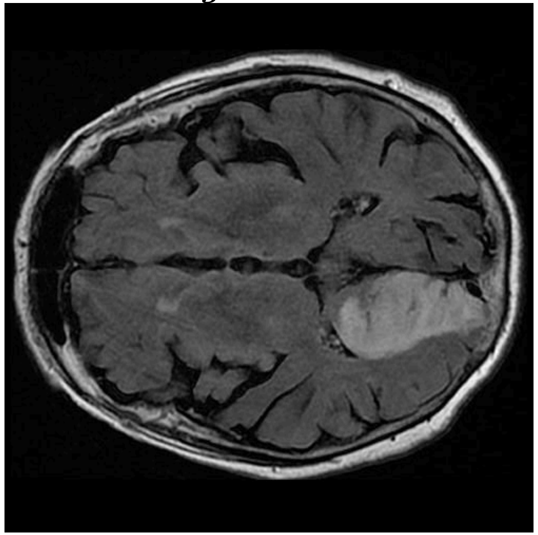
Invariance

- **A priori:** Invariance to certain transformations: $\forall g \in \mathcal{G} : p\big(\boldsymbol{T}_g \boldsymbol{x}\big) = p(\boldsymbol{x})$.

$\boldsymbol{x} \in \mathcal{X}$



$\boldsymbol{T}_g \boldsymbol{x} \in \mathcal{X}$



$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}\boldsymbol{T}_g \boldsymbol{T}_g^{-1}\boldsymbol{x} = \boldsymbol{A}_g \boldsymbol{x}'$$

Equivariant splitting

- **Assumptions:**
  - **Single** operator $\boldsymbol{A}$.
  - **Noiseless** measurements: $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$.
-

Equivariant splitting

- **Assumptions:**
  - **Single** operator $\boldsymbol{A}$.
  - **Noiseless** measurements: $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$.
- **Method:**
  - **Split** measurements $\boldsymbol{y} = [\boldsymbol{y}_1, \boldsymbol{y}_2]$ for **different** operators $\boldsymbol{A}\boldsymbol{T}_g = [\boldsymbol{A}_1, \boldsymbol{A}_2]$.

$$f^* \in \underset{f}{\operatorname{argmin}}\, \mathbb{E}_{\boldsymbol{y},} \{\mathcal{L}_{\text{ES}}(\boldsymbol{y}, \boldsymbol{A}, f)\}$$

$$\begin{aligned}\mathcal{L}_{\text{ES}}(\boldsymbol{y}, \boldsymbol{A}, f) &\triangleq \mathbb{E}_g\left\{\mathcal{L}_{\text{SPLIT}}\left(\boldsymbol{y}, \boldsymbol{A}\boldsymbol{T}_g, f\right)\right\} \\ &= \mathbb{E}_g \mathbb{E}_{\boldsymbol{y}_1, \boldsymbol{A}_1 \mid \boldsymbol{y}, \boldsymbol{A}\boldsymbol{T}_g}\left\{\left\|\boldsymbol{A}\boldsymbol{T}_g f(\boldsymbol{y}_1, \boldsymbol{A}_1) - \boldsymbol{y}\right\|^2\right\}\end{aligned}$$

A visual interpretation of $\boldsymbol{Q}_{\boldsymbol{A}_1}$

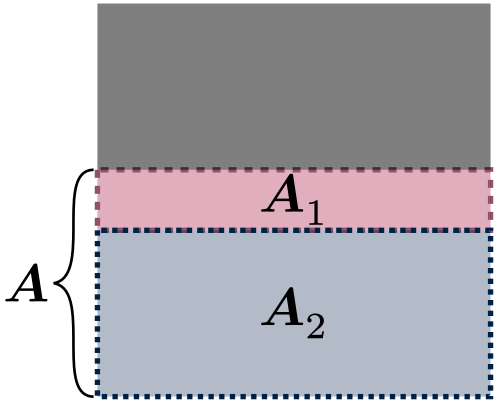
Solution with $\boldsymbol{Q}_{\boldsymbol{A}_1} \triangleq \mathbb{E}_{g \mid \boldsymbol{A}_1}\left\{\left(\boldsymbol{A}\boldsymbol{T}_g\right)^\top \boldsymbol{A}\boldsymbol{T}_g\right\}$:

$$f^*(\boldsymbol{y}_1, \boldsymbol{A}_1) = \boldsymbol{Q}^\dagger_{\boldsymbol{A}_1}\boldsymbol{Q}_{\boldsymbol{A}_1}\underbrace{\mathbb{E}_{\boldsymbol{x} \mid \boldsymbol{y}_1, \boldsymbol{A}_1}\{\boldsymbol{x}\}}_{\text{MMSE estimator}} + \underbrace{\left(\boldsymbol{I} - \boldsymbol{Q}^\dagger_{\boldsymbol{A}_1}\boldsymbol{Q}_{\boldsymbol{A}_1}\right)v(\boldsymbol{y}_1)}_{\text{Error}}$$

A visual interpretation of $\boldsymbol{Q}_{\boldsymbol{A}_1}$

Solution with $\boldsymbol{Q}_{\boldsymbol{A}_1} \triangleq \mathbb{E}_{g \mid \boldsymbol{A}_1}\left\{\left(\boldsymbol{A}\boldsymbol{T}_g\right)^\top \boldsymbol{A}\boldsymbol{T}_g\right\}$:

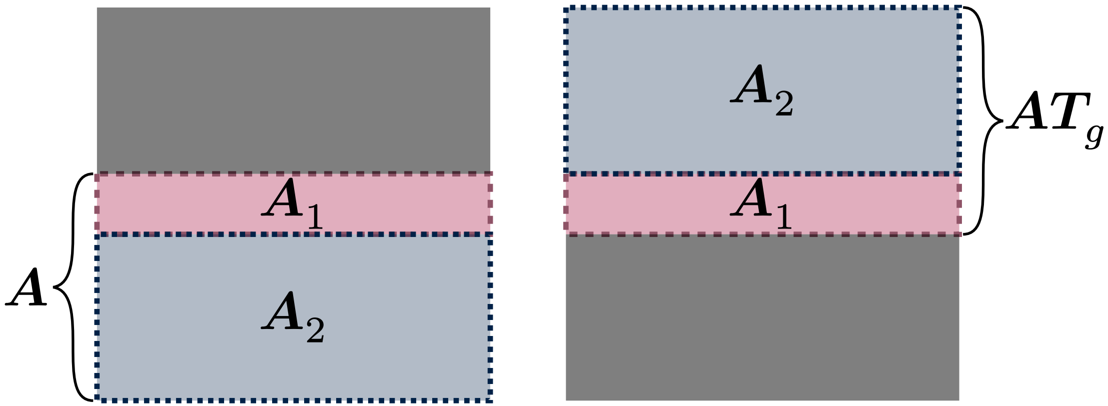
$$f^*(\boldsymbol{y}_1, \boldsymbol{A}_1) = \boldsymbol{Q}_{\boldsymbol{A}_1}^\dagger \boldsymbol{Q}_{\boldsymbol{A}_1} \underbrace{\mathbb{E}_{\boldsymbol{x} \mid \boldsymbol{y}_1, \boldsymbol{A}_1}\{\boldsymbol{x}\}}_{\text{MMSE estimator}} + \underbrace{\left(\boldsymbol{I} - \boldsymbol{Q}_{\boldsymbol{A}_1}^\dagger \boldsymbol{Q}_{\boldsymbol{A}_1}\right) v(\boldsymbol{y}_1)}_{\text{Error}}$$



A visual interpretation of Q_{A_1}

Solution with $\boldsymbol{Q}_{\boldsymbol{A}_1} \triangleq \mathbb{E}_{g \mid \boldsymbol{A}_1}\left\{\left(\boldsymbol{A}\boldsymbol{T}_g\right)^\top \boldsymbol{A}\boldsymbol{T}_g\right\}$:

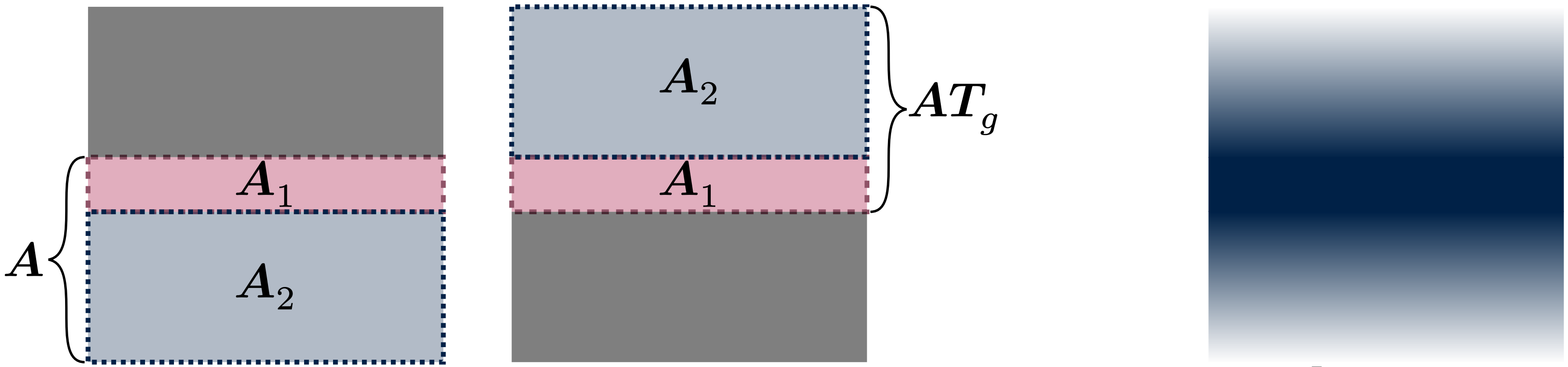
$$f^*(\boldsymbol{y}_1, \boldsymbol{A}_1) = \boldsymbol{Q}_{\boldsymbol{A}_1}^\dagger \boldsymbol{Q}_{\boldsymbol{A}_1} \underbrace{\mathbb{E}_{\boldsymbol{x} \mid \boldsymbol{y}_1, \boldsymbol{A}_1}\{\boldsymbol{x}\}}_{\text{MMSE estimator}} + \underbrace{\left(\boldsymbol{I} - \boldsymbol{Q}_{\boldsymbol{A}_1}^\dagger \boldsymbol{Q}_{\boldsymbol{A}_1}\right) v(\boldsymbol{y}_1)}_{\text{Error}}$$



A visual interpretation of Q_{A_1}

Solution with $Q_{\boldsymbol{A}_1} \triangleq \mathbb{E}_{g \mid \boldsymbol{A}_1}\left\{\left(\boldsymbol{A}\boldsymbol{T}_g\right)^\top \boldsymbol{A}\boldsymbol{T}_g\right\}$:

$$f^*(\boldsymbol{y}_1, \boldsymbol{A}_1) = Q^\dagger_{\boldsymbol{A}_1} Q_{\boldsymbol{A}_1} \underbrace{\mathbb{E}_{\boldsymbol{x} \mid \boldsymbol{y}_1, \boldsymbol{A}_1}\{\boldsymbol{x}\}}_{\text{MMSE estimator}} + \underbrace{\left(\boldsymbol{I} - Q^\dagger_{\boldsymbol{A}_1} Q_{\boldsymbol{A}_1}\right) v(\boldsymbol{y}_1)}_{\text{Error}}$$



$$\mathbb{E}_{g \mid \boldsymbol{A}_1}\left[\left(\boldsymbol{A}\boldsymbol{T}_g\right)^\top \boldsymbol{A}\boldsymbol{T}_g\right]$$

Inference

$$f^* \in \underset{f}{\operatorname{argmin}}\, \mathbb{E}_{\boldsymbol{y}}\{\mathcal{L}_{\text{ES}}(\boldsymbol{y}, \boldsymbol{A}, f)\}$$

- **At test time:** several splits $\boldsymbol{A}_1 \longrightarrow$ **Average** the reconstructions.
-
-

$$f^* \in \operatorname*{argmin}_f \mathbb{E}_{\boldsymbol{y}}\{\mathcal{L}_{\mathrm{ES}}(\boldsymbol{y}, \boldsymbol{A}, f)\}$$

- **At test time:** several splits $\boldsymbol{A}_1 \longrightarrow$ **Average** the reconstructions.
- If the average $\overline{\boldsymbol{Q}}_{\boldsymbol{A}} \triangleq \mathbb{E}_{\boldsymbol{A}_1 \mid \boldsymbol{A}}\left\{\boldsymbol{Q}_{\boldsymbol{A}_1}\right\}$ **is invertible**:

$$\begin{aligned}\overline{f}(\boldsymbol{y}, \boldsymbol{A}) &\triangleq \mathbb{E}_{\boldsymbol{y}_1, \boldsymbol{A}_1 \mid \boldsymbol{y}, \boldsymbol{A}}\left\{\overline{\boldsymbol{Q}}_{\boldsymbol{A}}^{-1} \boldsymbol{Q}_{\boldsymbol{A}_1} f^*(\boldsymbol{y}_1, \boldsymbol{A}_1)\right\} \\ &= \underbrace{\mathbb{E}_{\boldsymbol{y}_1, \boldsymbol{A}_1 \mid \boldsymbol{y}, \boldsymbol{A}}\left\{\overline{\boldsymbol{Q}}_{\boldsymbol{A}}^{-1} \boldsymbol{Q}_{\boldsymbol{A}_1} \mathbb{E}_{\boldsymbol{x} \mid \boldsymbol{y}_1, \boldsymbol{A}_1}\{\boldsymbol{x}\}\right\}}_{\text{Convex combination of MMSE estimators}}.\end{aligned}$$

-

$$f^* \in \operatorname*{argmin}_f \mathbb{E}_{\boldsymbol{y}}\{\mathcal{L}_{\mathrm{ES}}(\boldsymbol{y}, \boldsymbol{A}, f)\}$$

- **At test time:** several splits $\boldsymbol{A}_1 \longrightarrow$ **Average** the reconstructions.
- If the average $\overline{\boldsymbol{Q}}_{\boldsymbol{A}} \triangleq \mathbb{E}_{\boldsymbol{A}_1 \mid \boldsymbol{A}}\left\{\boldsymbol{Q}_{\boldsymbol{A}_1}\right\}$ **is invertible**:

$$\begin{aligned}\overline{f}(\boldsymbol{y}, \boldsymbol{A}) &\triangleq \mathbb{E}_{\boldsymbol{y}_1, \boldsymbol{A}_1 \mid \boldsymbol{y}, \boldsymbol{A}}\left\{\overline{\boldsymbol{Q}}_{\boldsymbol{A}}^{-1} \boldsymbol{Q}_{\boldsymbol{A}_1} f^*(\boldsymbol{y}_1, \boldsymbol{A}_1)\right\} \\ &= \underbrace{\mathbb{E}_{\boldsymbol{y}_1, \boldsymbol{A}_1 \mid \boldsymbol{y}, \boldsymbol{A}}\left\{\overline{\boldsymbol{Q}}_{\boldsymbol{A}}^{-1} \boldsymbol{Q}_{\boldsymbol{A}_1} \mathbb{E}_{\boldsymbol{x} \mid \boldsymbol{y}_1, \boldsymbol{A}_1}\{\boldsymbol{x}\}\right\}}_{\text{Convex combination of MMSE estimators}}.\end{aligned}$$

- If $\overline{\boldsymbol{Q}}_{\boldsymbol{A}}$ cannot be computed efficiently $\longrightarrow$ **non-weighted average**:

$$\overline{f}(\boldsymbol{y}, \boldsymbol{A}) = \frac{1}{J} \sum_{j=1}^{J} f^*\left(\boldsymbol{y}_1^{(j)}, \boldsymbol{A}_i^{(j)}\right) \quad \text{with} \quad \left(\boldsymbol{y}_1^{(j)}, \boldsymbol{A}_1^{(j)}\right) \sim p\big(\boldsymbol{y}_1, \boldsymbol{A}_1 \mid \boldsymbol{y}, \boldsymbol{A}\boldsymbol{T}_g\big).$$

Noisy measurements

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\varepsilon}$$

- **Divide** the $\mathcal{L}_{\text{ES}}$ into 2 terms:

$$\mathcal{L}_{\text{ES}}(\boldsymbol{y}, \boldsymbol{A}, f) = \mathbb{E}_g \mathbb{E}_{\boldsymbol{y}_1, \boldsymbol{A}_1 \mid \boldsymbol{y}, \boldsymbol{A}\boldsymbol{T}_g} \Big\{ \underbrace{\left\| \boldsymbol{A}_1 f(\boldsymbol{y}_1, \boldsymbol{A}_1) - \boldsymbol{y}_1 \right\|^2}_{\text{Measurement consistency (MC)}} + \underbrace{\left\| \boldsymbol{A}_2 f(\boldsymbol{y}_1, \boldsymbol{A}_1) - \boldsymbol{y}_2 \right\|^2}_{\text{prediction accuracy}} \Big\}.$$

-

-

Noisy measurements

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\varepsilon}$$

- **Divide** the $\mathcal{L}_{\text{ES}}$ into 2 terms:

$$\mathcal{L}_{\text{ES}}(\boldsymbol{y}, \boldsymbol{A}, f) = \mathbb{E}_g \mathbb{E}_{\boldsymbol{y}_1, \boldsymbol{A}_1 \mid \boldsymbol{y}, \boldsymbol{A}\boldsymbol{T}_g} \Big\{ \underbrace{\|\boldsymbol{A}_1 f(\boldsymbol{y}_1, \boldsymbol{A}_1) - \boldsymbol{y}_1\|^2}_{\text{Measurement consistency (MC)}} + \underbrace{\|\boldsymbol{A}_2 f(\boldsymbol{y}_1, \boldsymbol{A}_1) - \boldsymbol{y}_2\|^2}_{\text{prediction accuracy}} \Big\}.$$

- **Replace** MC by a self-supervised **denoising** loss:

$$\mathbb{E}_g \mathbb{E}_{\boldsymbol{y}_1, \boldsymbol{A}_1, \boldsymbol{\omega} \mid \boldsymbol{y}, \boldsymbol{A}\boldsymbol{T}_g} \Bigg\{ \underbrace{\Bigg\| \boldsymbol{A}_1 f(\boldsymbol{y}_1 + \alpha\boldsymbol{\omega}, \boldsymbol{A}_1) - \left(\boldsymbol{y}_1 - \frac{\boldsymbol{\omega}}{\alpha} \right) \Bigg\|^2}_{\text{R2R (denoising) loss}} + \|\boldsymbol{A}_2 f(\boldsymbol{y}_1 + \alpha\boldsymbol{\omega}, \boldsymbol{A}_1) - \boldsymbol{y}_2\|^2 \Bigg\}.$$

-

Noisy measurements

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\varepsilon}$$

- **Divide** the $\mathcal{L}_{\text{ES}}$ into 2 terms:

$$\mathcal{L}_{\text{ES}}(\boldsymbol{y}, \boldsymbol{A}, f) = \mathbb{E}_g \mathbb{E}_{\boldsymbol{y}_1, \boldsymbol{A}_1 \mid \boldsymbol{y}, \boldsymbol{A}\boldsymbol{T}_g} \left\{ \underbrace{\|\boldsymbol{A}_1 f(\boldsymbol{y}_1, \boldsymbol{A}_1) - \boldsymbol{y}_1\|^2}_{\text{Measurement consistency (MC)}} + \underbrace{\|\boldsymbol{A}_2 f(\boldsymbol{y}_1, \boldsymbol{A}_1) - \boldsymbol{y}_2\|^2}_{\text{prediction accuracy}} \right\}.$$

- **Replace** MC by a self-supervised **denoising** loss:

$$\mathbb{E}_g \mathbb{E}_{\boldsymbol{y}_1, \boldsymbol{A}_1, \boldsymbol{\omega} \mid \boldsymbol{y}, \boldsymbol{A}\boldsymbol{T}_g} \left\{ \underbrace{\left\| \boldsymbol{A}_1 f(\boldsymbol{y}_1 + \alpha\boldsymbol{\omega}, \boldsymbol{A}_1) - \left(\boldsymbol{y}_1 - \frac{\boldsymbol{\omega}}{\alpha} \right) \right\|^2}_{\text{R2R (denoising) loss}} + \|\boldsymbol{A}_2 f(\boldsymbol{y}_1 + \alpha\boldsymbol{\omega}, \boldsymbol{A}_1) - \boldsymbol{y}_2\|^2 \right\}.$$

- $\boldsymbol{Q}_{\boldsymbol{A}_1}$ invertible $\Rightarrow f^*(\boldsymbol{y}_1, \boldsymbol{A}_1) = \mathbb{E}_{\boldsymbol{x} \mid \boldsymbol{y}_1, \boldsymbol{A}_1}\{\boldsymbol{x}\}$.

Computing the training loss efficiently

- **In many methods:** $f(\boldsymbol{y}, \boldsymbol{A}) = \varphi_{\boldsymbol{\theta}}\big(\boldsymbol{A}^\top \boldsymbol{y}\big)$ or $f(\boldsymbol{y}, \boldsymbol{A}) = \varphi_{\boldsymbol{\theta}}\big(\boldsymbol{A}^\dagger \boldsymbol{y}\big)$
- **Image-to-image equivairant** network: $\varphi\big(\boldsymbol{T}_g \boldsymbol{x}\big) = \boldsymbol{T}_g \varphi(\boldsymbol{x})$

Computing the training loss efficiently

- **In many methods:** $f(\boldsymbol{y}, \boldsymbol{A}) = \varphi_{\boldsymbol{\theta}}\big(\boldsymbol{A}^\top \boldsymbol{y}\big)$ or $f(\boldsymbol{y}, \boldsymbol{A}) = \varphi_{\boldsymbol{\theta}}\big(\boldsymbol{A}^\dagger \boldsymbol{y}\big)$
- **Image-to-image equivairant** network: $\varphi\big(\boldsymbol{T}_g \boldsymbol{x}\big) = \boldsymbol{T}_g \varphi(\boldsymbol{x})$

$$\mathcal{L}_{\mathrm{ES}}(\boldsymbol{y}, \boldsymbol{A}, f) = \mathbb{E}_g \mathbb{E}_{\boldsymbol{y}_1, \boldsymbol{A}_1 \mid \boldsymbol{y}, \boldsymbol{A}\boldsymbol{T}_g} \Big\{ \big\| \boldsymbol{A}\boldsymbol{T}_g f(\boldsymbol{y}_1, \boldsymbol{A}_1) - \boldsymbol{y} \big\|^2 \Big\}$$

Computing the training loss efficiently

- **In many methods:** $f(\boldsymbol{y},\boldsymbol{A}) = \varphi_{\boldsymbol{\theta}}\big(\boldsymbol{A}^\top \boldsymbol{y}\big)$ or $f(\boldsymbol{y},\boldsymbol{A}) = \varphi_{\boldsymbol{\theta}}\big(\boldsymbol{A}^\dagger \boldsymbol{y}\big)$
- **Image-to-image equivairant** network: $\varphi\big(\boldsymbol{T}_g \boldsymbol{x}\big) = \boldsymbol{T}_g \varphi(\boldsymbol{x})$

$$
\begin{aligned}
\mathcal{L}_{\mathrm{ES}}(\boldsymbol{y},\boldsymbol{A},f) &= \mathbb{E}_g \mathbb{E}_{\boldsymbol{y}_1,\boldsymbol{A}_1 \mid \boldsymbol{y},\boldsymbol{A}\boldsymbol{T}_g} \left\{ \left\| \boldsymbol{A}\boldsymbol{T}_g f(\boldsymbol{y}_1,\boldsymbol{A}_1) - \boldsymbol{y} \right\|^2 \right\} \\
&= \mathbb{E}_g \mathbb{E}_{\boldsymbol{M} \mid \boldsymbol{y},g} \left\{ \left\| \boldsymbol{A}\boldsymbol{T}_g f\big(\boldsymbol{M}\boldsymbol{y}, \boldsymbol{M}\boldsymbol{A}\boldsymbol{T}_g\big) - \boldsymbol{y} \right\|^2 \right\}
\end{aligned}
$$

Computing the training loss efficiently

- **In many methods:** $f(\boldsymbol{y}, \boldsymbol{A}) = \varphi_{\boldsymbol{\theta}}(\boldsymbol{A}^\top \boldsymbol{y})$ or $f(\boldsymbol{y}, \boldsymbol{A}) = \varphi_{\boldsymbol{\theta}}(\boldsymbol{A}^\dagger \boldsymbol{y})$
- **Image-to-image equivairant** network: $\varphi(\boldsymbol{T}_g \boldsymbol{x}) = \boldsymbol{T}_g \varphi(\boldsymbol{x})$

$$\begin{aligned}\mathcal{L}_{\mathrm{ES}}(\boldsymbol{y}, \boldsymbol{A}, f) &= \mathbb{E}_g \mathbb{E}_{\boldsymbol{y}_1, \boldsymbol{A}_1 \mid \boldsymbol{y}, \boldsymbol{A}\boldsymbol{T}_g} \left\{ \left\| \boldsymbol{A}\boldsymbol{T}_g f(\boldsymbol{y}_1, \boldsymbol{A}_1) - \boldsymbol{y} \right\|^2 \right\} \\ &= \mathbb{E}_g \mathbb{E}_{\boldsymbol{M} \mid \boldsymbol{y}, g} \left\{ \left\| \boldsymbol{A}\boldsymbol{T}_g f(\boldsymbol{M}\boldsymbol{y}, \boldsymbol{M}\boldsymbol{A}\boldsymbol{T}_g) - \boldsymbol{y} \right\|^2 \right\} \\ &= \mathbb{E}_g \mathbb{E}_{\boldsymbol{M} \mid \boldsymbol{y}, g} \left\{ \left\| \boldsymbol{A}\boldsymbol{T}_g \varphi_{\boldsymbol{\theta}}(\boldsymbol{T}_g^{-1} \boldsymbol{A}^\top \boldsymbol{M}^\top \boldsymbol{M}\boldsymbol{y}) - \boldsymbol{y} \right\|^2 \right\}\end{aligned}$$

Computing the training loss efficiently

- **In many methods:** $f(\boldsymbol{y}, \boldsymbol{A}) = \varphi_{\boldsymbol{\theta}}(\boldsymbol{A}^\top \boldsymbol{y})$ or $f(\boldsymbol{y}, \boldsymbol{A}) = \varphi_{\boldsymbol{\theta}}(\boldsymbol{A}^\dagger \boldsymbol{y})$
- **Image-to-image equivairant** network: $\varphi(\boldsymbol{T}_g \boldsymbol{x}) = \boldsymbol{T}_g \varphi(\boldsymbol{x})$

$$
\begin{aligned}
\mathcal{L}_{\mathrm{ES}}(\boldsymbol{y}, \boldsymbol{A}, f) &= \mathbb{E}_g \mathbb{E}_{\boldsymbol{y}_1, \boldsymbol{A}_1 \mid \boldsymbol{y}, \boldsymbol{A}\boldsymbol{T}_g} \left\{ \left\| \boldsymbol{A}\boldsymbol{T}_g f(\boldsymbol{y}_1, \boldsymbol{A}_1) - \boldsymbol{y} \right\|^2 \right\} \\
&= \mathbb{E}_g \mathbb{E}_{\boldsymbol{M} \mid \boldsymbol{y}, g} \left\{ \left\| \boldsymbol{A}\boldsymbol{T}_g f(\boldsymbol{M}\boldsymbol{y}, \boldsymbol{M}\boldsymbol{A}\boldsymbol{T}_g) - \boldsymbol{y} \right\|^2 \right\} \\
&= \mathbb{E}_g \mathbb{E}_{\boldsymbol{M} \mid \boldsymbol{y}, g} \left\{ \left\| \boldsymbol{A}\boldsymbol{T}_g \varphi_{\boldsymbol{\theta}}(\boldsymbol{T}_g^{-1} \boldsymbol{A}^\top \boldsymbol{M}^\top \boldsymbol{M}\boldsymbol{y}) - \boldsymbol{y} \right\|^2 \right\} \\
&= \mathbb{E}_g \mathbb{E}_{\boldsymbol{M} \mid \boldsymbol{y}, g} \left\{ \left\| \boldsymbol{A}\cancel{\boldsymbol{T}_g \boldsymbol{T}_g^{-1}} \varphi_{\boldsymbol{\theta}}(\boldsymbol{A}^\top \boldsymbol{M}^\top \boldsymbol{M}\boldsymbol{y}) - \boldsymbol{y} \right\|^2 \right\} \quad \text{if } \varphi \text{ is equivariant!}
\end{aligned}
$$

Computing the training loss efficiently

- **In many methods:** $f(\boldsymbol{y}, \boldsymbol{A}) = \varphi_{\boldsymbol{\theta}}(\boldsymbol{A}^\top \boldsymbol{y})$ or $f(\boldsymbol{y}, \boldsymbol{A}) = \varphi_{\boldsymbol{\theta}}(\boldsymbol{A}^\dagger \boldsymbol{y})$
- **Image-to-image equivairant** network: $\varphi(\boldsymbol{T}_g \boldsymbol{x}) = \boldsymbol{T}_g \varphi(\boldsymbol{x})$

$$
\begin{aligned}
\mathcal{L}_{\mathrm{ES}}(\boldsymbol{y}, \boldsymbol{A}, f) &= \mathbb{E}_g \mathbb{E}_{\boldsymbol{y}_1, \boldsymbol{A}_1 \mid \boldsymbol{y}, \boldsymbol{A}\boldsymbol{T}_g} \left\{ \left\| \boldsymbol{A}\boldsymbol{T}_g f(\boldsymbol{y}_1, \boldsymbol{A}_1) - \boldsymbol{y} \right\|^2 \right\} \\
&= \mathbb{E}_g \mathbb{E}_{\boldsymbol{M} \mid \boldsymbol{y}, g} \left\{ \left\| \boldsymbol{A}\boldsymbol{T}_g f(\boldsymbol{M}\boldsymbol{y}, \boldsymbol{M}\boldsymbol{A}\boldsymbol{T}_g) - \boldsymbol{y} \right\|^2 \right\} \\
&= \mathbb{E}_g \mathbb{E}_{\boldsymbol{M} \mid \boldsymbol{y}, g} \left\{ \left\| \boldsymbol{A}\boldsymbol{T}_g \varphi_{\boldsymbol{\theta}}(\boldsymbol{T}_g^{-1} \boldsymbol{A}^\top \boldsymbol{M}^\top \boldsymbol{M}\boldsymbol{y}) - \boldsymbol{y} \right\|^2 \right\} \\
&= \mathbb{E}_g \mathbb{E}_{\boldsymbol{M} \mid \boldsymbol{y}, g} \left\{ \left\| \boldsymbol{A}\cancel{\boldsymbol{T}_g \boldsymbol{T}_g^{-1}} \varphi_{\boldsymbol{\theta}}(\boldsymbol{A}^\top \boldsymbol{M}^\top \boldsymbol{M}\boldsymbol{y}) - \boldsymbol{y} \right\|^2 \right\} \quad \text{if } \varphi \text{ is equivariant!} \\
&= \mathbb{E}_{\boldsymbol{M} \mid \boldsymbol{y}} \{ \| \boldsymbol{A} f(\boldsymbol{M}\boldsymbol{y}, \boldsymbol{M}\boldsymbol{A}) - \boldsymbol{y} \|^2 \}
\end{aligned}
$$

Computing the training loss efficiently

- **In many methods:** $f(\boldsymbol{y}, \boldsymbol{A}) = \varphi_{\boldsymbol{\theta}}(\boldsymbol{A}^\top \boldsymbol{y})$ or $f(\boldsymbol{y}, \boldsymbol{A}) = \varphi_{\boldsymbol{\theta}}(\boldsymbol{A}^\dagger \boldsymbol{y})$
- **Image-to-image equivairant** network: $\varphi(\boldsymbol{T}_g \boldsymbol{x}) = \boldsymbol{T}_g \varphi(\boldsymbol{x})$

$$
\begin{aligned}
\mathcal{L}_{\text{ES}}(\boldsymbol{y}, \boldsymbol{A}, f) &= \mathbb{E}_g \mathbb{E}_{\boldsymbol{y}_1, \boldsymbol{A}_1 \mid \boldsymbol{y}, \boldsymbol{A}\boldsymbol{T}_g} \left\{ \left\| \boldsymbol{A}\boldsymbol{T}_g f(\boldsymbol{y}_1, \boldsymbol{A}_1) - \boldsymbol{y} \right\|^2 \right\} \\
&= \mathbb{E}_g \mathbb{E}_{\boldsymbol{M} \mid \boldsymbol{y}, g} \left\{ \left\| \boldsymbol{A}\boldsymbol{T}_g f\left(\boldsymbol{M}\boldsymbol{y}, \boldsymbol{M}\boldsymbol{A}\boldsymbol{T}_g\right) - \boldsymbol{y} \right\|^2 \right\} \\
&= \mathbb{E}_g \mathbb{E}_{\boldsymbol{M} \mid \boldsymbol{y}, g} \left\{ \left\| \boldsymbol{A}\boldsymbol{T}_g \varphi_{\boldsymbol{\theta}}\left(\boldsymbol{T}_g^{-1} \boldsymbol{A}^\top \boldsymbol{M}^\top \boldsymbol{M}\boldsymbol{y}\right) - \boldsymbol{y} \right\|^2 \right\} \\
&= \mathbb{E}_g \mathbb{E}_{\boldsymbol{M} \mid \boldsymbol{y}, g} \left\{ \left\| \boldsymbol{A}\cancel{\boldsymbol{T}_g \boldsymbol{T}_g^{-1}} \varphi_{\boldsymbol{\theta}}(\boldsymbol{A}^\top \boldsymbol{M}^\top \boldsymbol{M}\boldsymbol{y}) - \boldsymbol{y} \right\|^2 \right\} \quad \text{if } \varphi \text{ is equivariant!} \\
&= \mathbb{E}_{\boldsymbol{M} \mid \boldsymbol{y}} \{ \| \boldsymbol{A} f(\boldsymbol{M}\boldsymbol{y}, \boldsymbol{M}\boldsymbol{A}) - \boldsymbol{y} \|^2 \} \\
&= \mathcal{L}_{\text{SPLIT}}(\boldsymbol{y}, \boldsymbol{A}, f)
\end{aligned}
$$

Equivariant reconstructor

Definition

We say that the reconstruction function $f(\boldsymbol{y}, \boldsymbol{A})$ is an equivariant reconstructor if

$$f\big(\boldsymbol{y}, \boldsymbol{A}\boldsymbol{T}_g\big) = \boldsymbol{T}_g^{-1} f(\boldsymbol{y}, \boldsymbol{A}), \quad \forall \boldsymbol{y} \in \mathbb{R}^m, \quad \forall g \in \mathcal{G}, \quad \forall \boldsymbol{A} \in \mathbb{R}^{m \times n}.$$

Definition

We say that the reconstruction function $f(\boldsymbol{y}, \boldsymbol{A})$ is an equivariant reconstructor if

$$f\big(\boldsymbol{y}, \boldsymbol{A}\boldsymbol{T}_g\big) = \boldsymbol{T}_g^{-1} f(\boldsymbol{y}, \boldsymbol{A}), \quad \forall \boldsymbol{y} \in \mathbb{R}^m, \quad \forall g \in \mathcal{G}, \quad \forall \boldsymbol{A} \in \mathbb{R}^{m \times n}.$$

- **Artifact removal and unrolled network.** $f(\boldsymbol{y}, \boldsymbol{A}) = \boldsymbol{x}_L$ with

$$\boldsymbol{x}_{k+1} = \varphi\Big(\boldsymbol{x}_k - \gamma \nabla_{\boldsymbol{x}_k} d(\boldsymbol{A}\boldsymbol{x}_k, \boldsymbol{y})\Big).$$

-
-
-

Definition

We say that the reconstruction function $f(\boldsymbol{y}, \boldsymbol{A})$ is an equivariant reconstructor if

$$f\left(\boldsymbol{y}, \boldsymbol{A}\boldsymbol{T}_g\right) = \boldsymbol{T}_g^{-1} f(\boldsymbol{y}, \boldsymbol{A}), \quad \forall \boldsymbol{y} \in \mathbb{R}^m, \quad \forall g \in \mathcal{G}, \quad \forall \boldsymbol{A} \in \mathbb{R}^{m \times n}.$$

- **Artifact removal and unrolled network.** $f(\boldsymbol{y}, \boldsymbol{A}) = \boldsymbol{x}_L$ with

$$\boldsymbol{x}_{k+1} = \varphi\left(\boldsymbol{x}_k - \gamma \nabla_{\boldsymbol{x}_k} d(\boldsymbol{A}\boldsymbol{x}_k, \boldsymbol{y})\right).$$

- **Reynolds averaging.**

$$f(\boldsymbol{y}, \boldsymbol{A}) = \frac{1}{|G|} \sum_{g \in \mathcal{G}} \boldsymbol{T}_g r\left(y, \boldsymbol{A}\boldsymbol{T}_g\right).$$

-
-

Definition

We say that the reconstruction function $f(\boldsymbol{y}, \boldsymbol{A})$ is an equivariant reconstructor if

$$f\big(\boldsymbol{y}, \boldsymbol{A}\boldsymbol{T}_g\big) = \boldsymbol{T}_g^{-1} f(\boldsymbol{y}, \boldsymbol{A}), \quad \forall \boldsymbol{y} \in \mathbb{R}^m, \quad \forall g \in \mathcal{G}, \quad \forall \boldsymbol{A} \in \mathbb{R}^{m \times n}.$$

- **Artifact removal and unrolled network.** $f(\boldsymbol{y}, \boldsymbol{A}) = \boldsymbol{x}_L$ with

$$\boldsymbol{x}_{k+1} = \varphi\Big(\boldsymbol{x}_k - \gamma \nabla_{\boldsymbol{x}_k} d(\boldsymbol{A}\boldsymbol{x}_k, \boldsymbol{y})\Big).$$

- **Reynolds averaging.**

$$f(\boldsymbol{y}, \boldsymbol{A}) = \frac{1}{|G|} \sum_{g \in \mathcal{G}} \boldsymbol{T}_g r\big(y, \boldsymbol{A}\boldsymbol{T}_g\big).$$

- **Maximum a posteriori (MAP).**

$$f(\boldsymbol{y}, \boldsymbol{A}) = \underset{\boldsymbol{x} \in \mathbb{R}^n}{\operatorname{argmax}} \{p(\boldsymbol{x} \mid \boldsymbol{y}, \boldsymbol{A})\}.$$

-

Equivariant reconstructor

Definition

We say that the reconstruction function $f(\boldsymbol{y}, \boldsymbol{A})$ is an equivariant reconstructor if

$$f\big(\boldsymbol{y}, \boldsymbol{A}\boldsymbol{T}_g\big) = \boldsymbol{T}_g^{-1} f(\boldsymbol{y}, \boldsymbol{A}), \quad \forall \boldsymbol{y} \in \mathbb{R}^m, \quad \forall g \in \mathcal{G}, \quad \forall \boldsymbol{A} \in \mathbb{R}^{m \times n}.$$

- **Artifact removal and unrolled network.** $f(\boldsymbol{y}, \boldsymbol{A}) = \boldsymbol{x}_L$ with

$$\boldsymbol{x}_{k+1} = \varphi\Big(\boldsymbol{x}_k - \gamma \nabla_{\boldsymbol{x}_k} d(\boldsymbol{A}\boldsymbol{x}_k, \boldsymbol{y})\Big).$$

- **Reynolds averaging.**

$$f(\boldsymbol{y}, \boldsymbol{A}) = \frac{1}{|G|} \sum_{g \in \mathcal{G}} \boldsymbol{T}_g r\big(y, \boldsymbol{A}\boldsymbol{T}_g\big).$$

- **Maximum a posteriori (MAP).**

$$f(\boldsymbol{y}, \boldsymbol{A}) = \underset{\boldsymbol{x} \in \mathbb{R}^n}{\operatorname{argmax}} \{p(\boldsymbol{x} \mid \boldsymbol{y}, \boldsymbol{A})\}.$$

- **Minimum mean squared error (MMSE).**

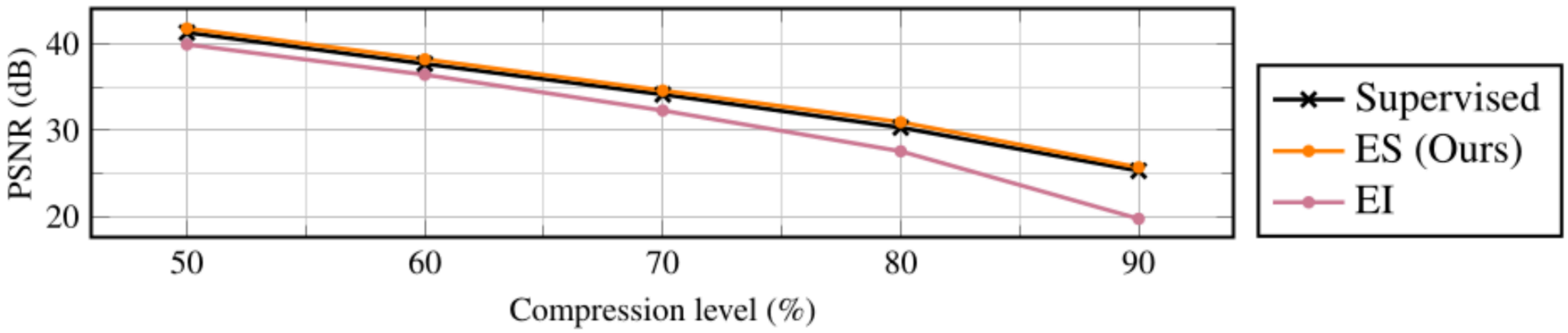
$$f(\boldsymbol{y}, \boldsymbol{A}) = \mathbb{E}[\boldsymbol{x} \mid \boldsymbol{y}, \boldsymbol{A}].$$

Experiments

Problem	**Equivariant architecture**	**Dataset**	**Transformation**
Compressive sensing	Unrolled equivariant UNet	MNIST 20000 images	Translations
Image inpainting	Equivariant UNet	DIV2K 800 images	Translations
MRI	Unrolled UNet + reynold averaging	FastMRI 900 images	Rotation + Flip

Experiments

Problem	**Equivariant architecture**	**Dataset**	**Transformation**
Compressive sensing	Unrolled equivariant UNet	MNIST 20000 images	Translations
Image inpainting	Equivariant UNet	DIV2K 800 images	Translations
MRI	Unrolled UNet + reynold averaging	FastMRI 900 images	Rotation + Flip

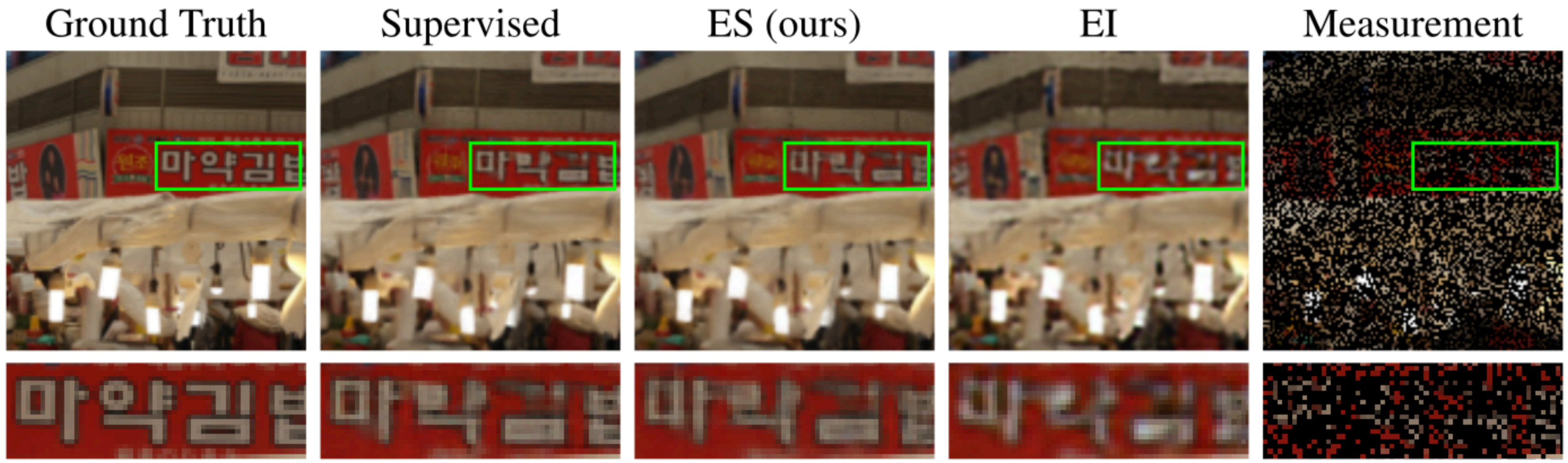


Inpainting and MRI

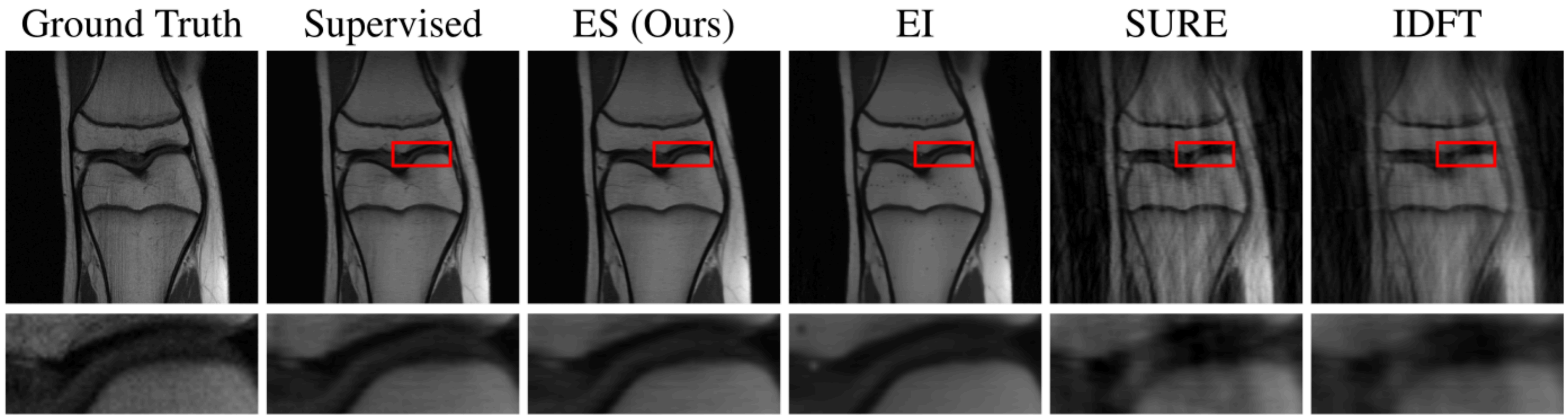
- **Inpainting:** image size: 128×128, $\boldsymbol{A}$: keep 30% of the pixels, no noise.
- **MRI:** image size: 265×256, acceleration: $\times$ 8, noise: $\sigma = 0.05$.

		PSNR ↑	
Training loss	Eq. arch.	Inpainting	MRI
Supervised	✓	28.46	28.74
	×	28.62	28.48
Splitting (Ours)	✓	**27.45**	**28.54**
	×	27.20	28.18
EI	✓	25.89	27.88
	×	26.33	-
Incomplete image/ IDFT	-	8.22	23.62

Images

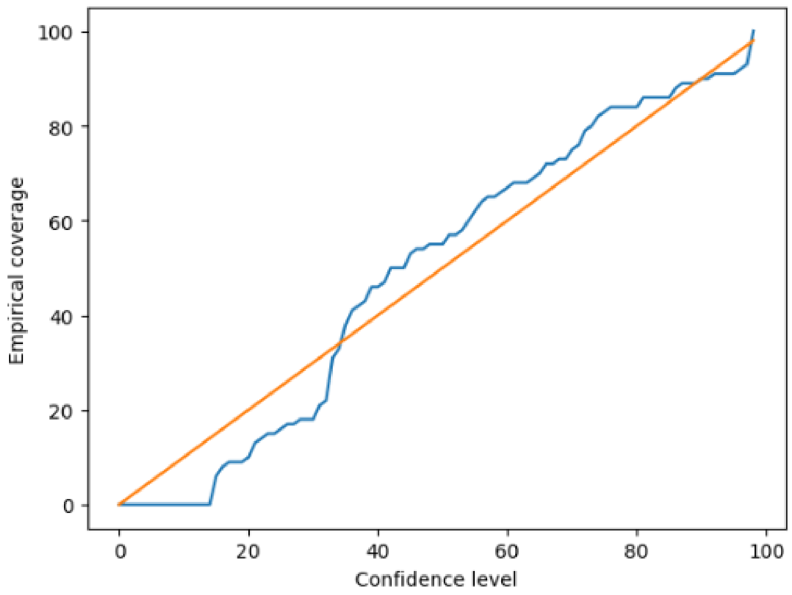


Ground Truth
Supervised
ES (ours)
EI
Measurement

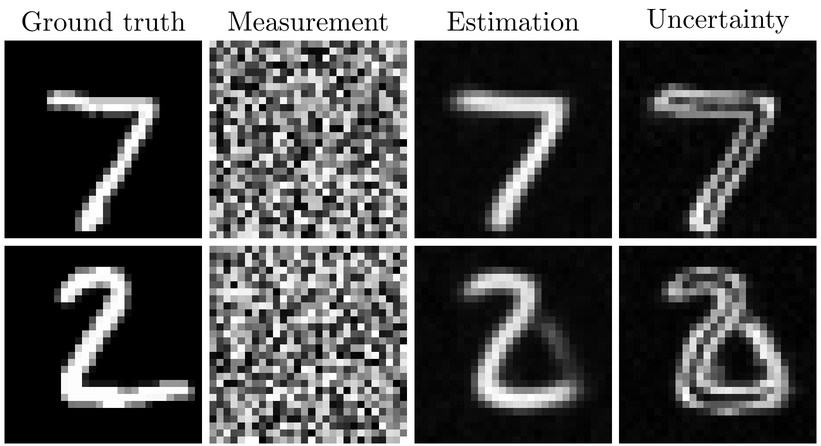


Ground Truth
Supervised
ES (Ours)
EI
SURE
IDFT

- Collaborative project with Marcelo Pereyra: **Uncertainty Quantification**.



Coverage plot

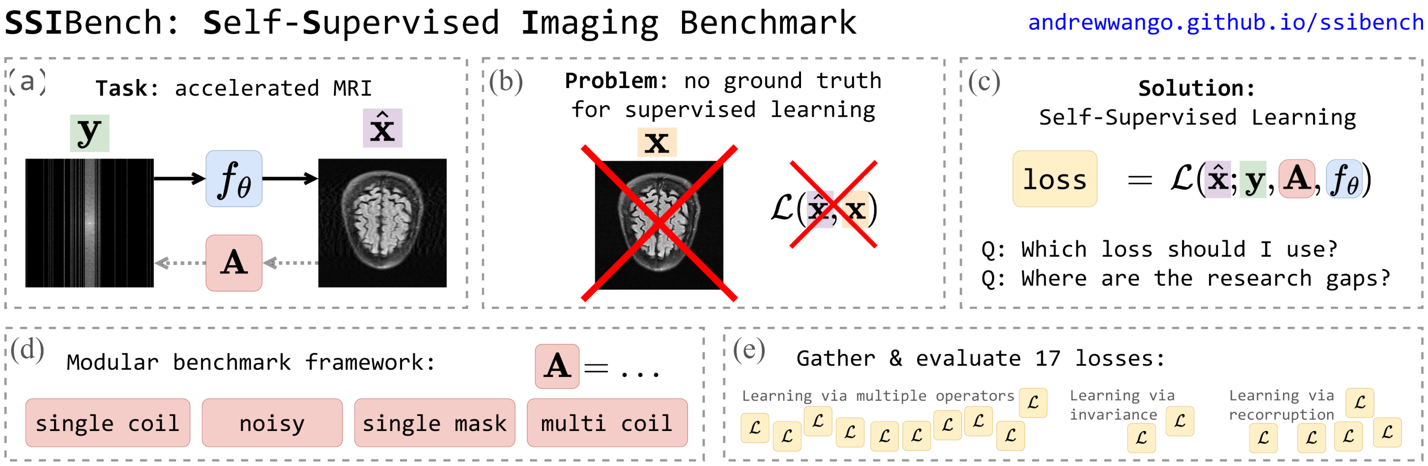


Predicted error

Code & acknowledgment

-**Code:** https://github.com/vsechaud/Equivariant-Splitting

-**MRI**:



-**Photographs of flamingos:** Émile Sechaud

Appendix

Complementaries results

Method	PSNR $\uparrow$	SSIM $\uparrow$
Supervised	28.46 ± 2.97	0.8982 ± 0.0411
ES (ours)	27.45 ± 2.86	0.8737 ± 0.0461
EI	25.89 ± 2.65	0.8332 ± 0.0521
Incomplete image	8.22 ± 2.47	0.0973 ± 0.0542

Method	PSNR $\uparrow$	SSIM $\uparrow$
Supervised	28.74 ± 2.81	0.6445 ± 0.1094
ES (ours)	28.54 ± 2.75	0.6195 ± 0.1188
EI	27.88 ± 2.64	0.5731 ± 0.1299
SURE	24.45 ± 1.86	0.5479 ± 0.0740
IDFT	23.62 ± 1.90	0.5052 ± 0.0900

Complementaries results

		Image inpainting	
Training loss	Eq. arch.	PSNR ↑	SSIM ↑
Supervised	✓	28.46 ± 2.97	0.8982 ± 0.0411
	×	28.62 ± 3.03	0.9002 ± 0.0414
Splitting (Ours)	✓	27.45 ± 2.86	0.8737 ± 0.0461
	×	27.20 ± 2.83	0.8652 ± 0.0461
		MRI ($\times$ 8 Accel.)	
Training loss	Eq. arch.	PSNR ↑	SSIM ↑
Supervised	✓	28.74 ± 2.81	0.6445 ± 0.1094
	×	28.48 ± 2.68	0.6381 ± 0.1082
Splitting (Ours)	✓	28.54 ± 2.75	0.6195 ± 0.1188
	×	28.18 ± 2.58	0.6104 ± 0.1176

What is $Q_{\boldsymbol{A}_1}$ for equivariant splitting?

- $\boldsymbol{Q}_{\boldsymbol{A}_1}$ is a **weight** depending on all possible $\boldsymbol{T}_g$, knowing $\boldsymbol{A}_1$.
- The more **probable** $\boldsymbol{T}_g$ is, the more it **impacts** $\boldsymbol{Q}_{\boldsymbol{A}_1}$.
- $\boldsymbol{Q}_{\boldsymbol{A}_1}$ invertible $\Rightarrow$ the entire space $\mathbb{R}^n$ is covered by all $\left(\boldsymbol{A}\boldsymbol{T}_g\right)^\top \boldsymbol{A}\boldsymbol{T}_g$.