An analytic theory of convolutional neural network inverse problems solvers

Minh-Hai Nguyen[1,*], Quoc-Bao Do[1], Edouard Pauwels[2], Pierre Weiss[1]

[1]IRIT, CBI & CNRS, France,[2]TSE, France *Corresponding author: minh-hai.nguyen@irit.fr

ICML International Conference On Machine Learning

Introduction

Inverse problems Let $\boldsymbol{x}$ a random vector in $\mathbb{R}^N$, we observe

$$\boldsymbol{y} = A\boldsymbol{x} + \boldsymbol{e}, \qquad \text{where } \boldsymbol{e} \sim \mathcal{N}\left(0, \sigma^2 \mathrm{Id}\right) \tag{1}$$

In practice, we observe a *realization* y of the random variable $\boldsymbol{y}$:

$$y = A\bar{x} + e$$

- $A : \mathbb{R}^N \to \mathbb{R}^M$: forward linear operator (e.g., convolution, subsampling, tomography, partial Fourier (MRI), etc.)
- $\bar{x}$: true signal to recover (drawn from $\boldsymbol{x}$) and e noise (drawn from $\boldsymbol{e}$)

Learning-based methods: train a neural network ϕ_θ to map y to an estimate $\hat{x}(y) \overset{\text{def}}{=} \phi_\theta(By) \approx \bar{x}$, where B is a back-projection operator (e.g., $A^\top$, A^+ or Id) and ϕ_θ is a neural network with parameters $\theta \in \Theta$.

→ **achieve SOTA performance but usually considered as black-boxes.**

Our contribution: a theoretical framework that relates CNNs to the MMSE estimators, deriving closed-form expression of the theoretical optimal solution, which is shown numerically to match the neural network outputs.

Supervised learning and MMSE estimator

Let $\mathcal{M} \overset{\text{def}}{=} \{\phi : \mathbb{R}^N \to \mathbb{R}^N \text{ measurable}\}$ is the set of all measurable functions.

$\mathcal{M}_\Theta \overset{\text{def}}{=} \{\phi_\theta : \theta \in \Theta\} \subseteq \mathcal{M}$ denote the range of a neural network architecture

Learning: $x^\star = \phi_{\theta^\star} \circ B$,

$$\theta^\star = \underset{\phi_\theta \in \mathcal{M}_\Theta}{\operatorname{argmin}} \mathbb{E}\left[\|\phi_\theta(B\boldsymbol{y}) - \boldsymbol{x}\|^2\right]$$

MMSE: $\hat{x}_{\text{MMSE}} = \phi^\star \circ B$,

$$\phi^\star = \underset{\phi_\theta \in \mathcal{M}}{\operatorname{argmin}} \mathbb{E}\left[\|\phi(B\boldsymbol{y}) - \boldsymbol{x}\|^2\right]$$

(Informal) Neural networks approximate the MMSE estimator. Suppose that

- The network is **sufficiently expressive**: $\mathcal{M}_\Theta \approx \mathcal{M}$
- The optimization problem is solved to optimality: $\phi^\star$ is a global minimizer of the ERM problem.

Then, $\phi^\star$ approximates the MMSE estimator, that is $\phi^\star \approx \hat{x}_{\text{MMSE}}$.

Empirical distribution

$$p_{\boldsymbol{x}} = \frac{1}{|\mathcal{D}|} \sum_{x \in \mathcal{D}} \delta_x \qquad \text{where} \mathcal{D} \text{ is a finite dataset}$$

Theorem 1 (Closed-form of MMSE). The MMSE estimator writes, for any $y \in \mathbb{R}^M$,

$$\hat{x}_{\text{MMSE}}(y) = \sum_{x \in \mathcal{D}} x \cdot w(x|y),$$

where $w(x|y) \propto \mathcal{N}\left(By; BAx, \sigma^2 BB^\top\right)$.

Implication: a neural network with *sufficient capacity* can learn to implement the above formula.

- The formula is exact for the empirical MMSE estimator, which is the optimal solution of the ERM problem.
- In practice, we **could not reproduce** this behavior on real datasets, even with large MLPs or CNNs.
- What is **missing** from this formula? Our answer: **Inductive biases** of the neural network architecture!

CNNs and Inductive Biases

Two well-known inductive biases of CNNs

- **Translation equivariance**: a fully convolutional network is *equivariant* to translations.
- **Locality**: finite receptive fields.

Reducing the search space of the MMSE estimator. Let $\mathcal{M}_{\text{LE}}$ denote the set of all measurable function that are *local and translation equivariant*.

$$\mathcal{M}_\Theta \subseteq \mathcal{M}_{\text{LE}} \subseteq \mathcal{M}$$

Learning: $x^\star = \phi_{\theta^\star} \circ B$,

$$\theta^\star = \underset{\phi_\theta \in \mathcal{M}_\Theta}{\operatorname{argmin}} \mathbb{E}\left[\|\phi_\theta(B\boldsymbol{y}) - \boldsymbol{x}\|^2\right]$$

Constrained MMSE: $\hat{x}_{\text{LE-MMSE}} = \phi^\star \circ B$,

$$\phi^\star = \underset{\phi_\theta \in \mathcal{M}_{\text{LE}}}{\operatorname{argmin}} \mathbb{E}\left[\|\phi(B\boldsymbol{y}) - \boldsymbol{x}\|^2\right]$$

Theorem 2 (Closed-form of LE-MMSE). Suppose that the N matrices $Q_n \overset{\text{def}}{=} \Pi_n B \in \mathbb{R}^{P \times M}$ have the *same rank* $r > 0$ for any n[a]. The LE-MMSE estimator $\hat{x}_{\text{LE-MMSE}}$ admits, for any $y \in \mathbb{R}^M$ the following expression, defined pixel-wise for each pixel n':

$$\hat{x}_{\text{LE-MMSE}}(y)[n'] = \sum_{x \in \mathcal{D}} \sum_{n=1}^{N} x[n] \cdot w_{n',n}(x|y),$$

where $w_{n',n}(x|y) \propto \mathcal{N}\left(Q_{n'}y; Q_n Ax, \sigma^2 Q_n Q_n^\top\right)$.

[a]This assumption can be relaxed: a general formula by rank stratification is also available.

Key insights:

- The value at pixel n' of the LE-MMSE estimator is a weighted average of **all** the pixels in the training images. The reconstruction is a **patchwork** of training patches.
- Can produce images outside $\mathrm{conv}(\mathcal{D})$.
- It is the **orthogonal projection** (in L^2 sense) of the empirical MMSE estimator onto $\mathcal{M}_{\text{LE}}$.

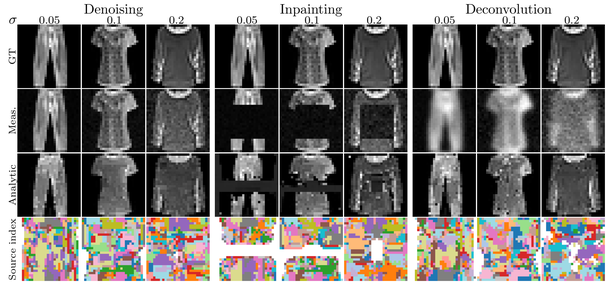


Figure: Pixel index of the reconstruction: contiguous regions coming from the same source image

The role of the back-projection operator B: estimator's variance

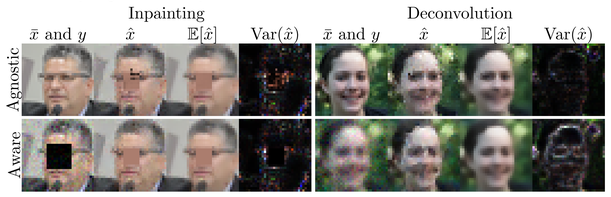


Figure: Physics-aware vs physics-agnostic solvers: pixel-wise variance.

Physics-aware (B = A+, bottom) has lower variance for inpainting (left), while physics-agnostic (B = I, top) has lower variance for deconvolution (right).

Numerical Results

Quantitative comparison: our closed-form matches the neural network outputs

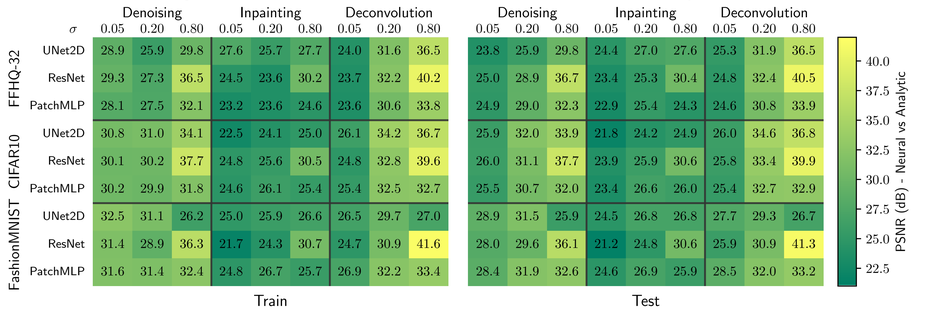


Figure: PSNR between our closed-form formula and neural network outputs

Qualitative comparison: on various inverse problems and datasets

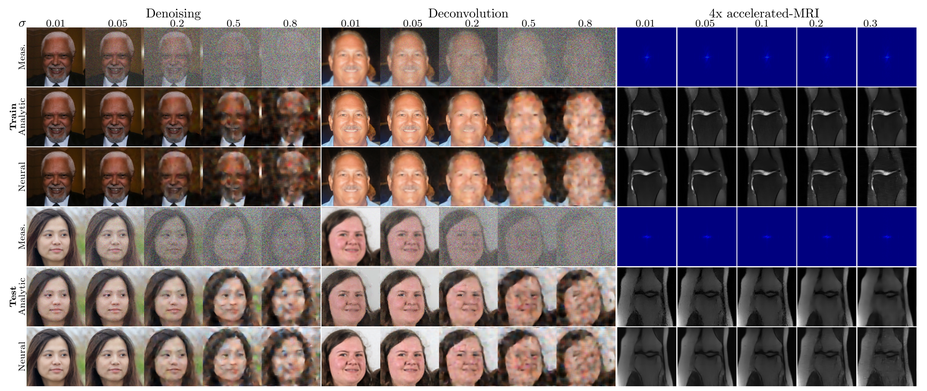
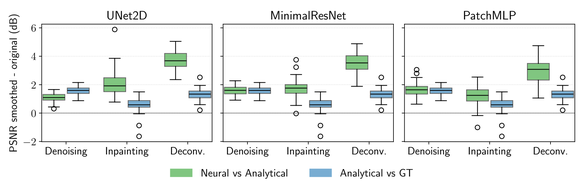


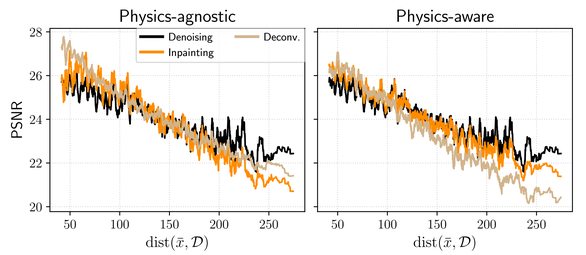
Figure: Qualitative comparison between our closed-form formula and neural network outputs.

Discussion

Spectral bias: CNNs exhibit a bias towards low frequencies, smoother images. Setting $\hat{x}^{\text{smooth}}_{\text{LE-MMSE}}(y) \overset{\text{def}}{=} \mathbb{E}\left[\hat{x}_{\text{LE-MMSE}}(y + \varepsilon \cdot \boldsymbol{z})\right]$ improves the both reconstruction and the match with neural networks.



Limitation: The alignment degrades in low-density regions (eg. out-of-distribution, low-noise levels or large patch sizes).



Next extensions

- Toward recovery guarantee: stability and reconstruction guarantees from the closed-form formula.
- Beyond Gaussian noise and MSE loss and toward more inductive biases.