Unsupervised Learning to Solve Inverse Problems: Application to Single-Pixel Imaging

Julián TACHELLA[1], Dongdong CHEN[2], Mike DAVIES[2]

[1]ENSL, CNRS, Laboratoire de Physique, Lyon, France
46, allée d'Italie, F-69364 Lyon Cedex 07, France

[2]University of Edinburgh, School of Engineering, UK
julian.tachella@ens-lyon.fr, d.chen@ed.ac.uk, mike.davies@ed.ac.uk

Résumé – Ces dernières années, les approches basées sur l'apprentissage profond ont obtenu des performances de pointe pour solutionner de nombreux problèmes d'imagerie inverse allant de l'imagerie médicale à la photographie computationnelle. Ces méthodes nécessitent généralement des paires de signaux et mesures pour l'apprentissage. Cependant, pour divers problèmes d'imagerie, nous avons généralement accès qu'à des mesures comprimées des signaux sous-jacents, rendant l'approche basée sur l'apprentissage compliquée, voire impossible car les observations compressées ne contiennent pas d'information dans l'espace nul de l'opérateur de détection directe. Le récent cadre d'imagerie équivariante surmonte cette limitation en exploitant l'invariance aux transformations (translations, rotations, etc.) présentes dans les signaux naturels. Dans cet article, nous exploitons ce nouveau cadre d'apprentissage non-supervisé pour reconstruire des données d'une camera mono-pixel à partir de mesures compressées uniquement. Une série d'expériences montre que la méthode proposée a des performances comparables à celles de l'approche supervisée standard.

Abstract – In recent years, learning-based approaches have obtained state-of-the-art performance in multiple imaging inverse problems ranging from medical imaging to computational photography. These methods generally require pairs of signals and associated measurements for training. However, in various imaging problems, we usually only have access to compressed measurements of the underlying signals, hindering this learning-based approach. Learning from measurement data only is impossible in general, as the compressed observations do not contain information in the nullspace of the forward sensing operator. The recent equivariant imaging framework overcomes this limitation by exploiting the invariance to transformations (translations, rotations, etc.) present in natural signals. In this paper, we leverage this novel unsupervised learning framework for reconstructing single-pixel imaging data from compressed measurements alone. A series of experiments show that the proposed method performs comparably to the standard supervised approach.

1 Introduction

Linear inverse problems consist of reconstructing a signal $x \in \mathcal{X} \subset \mathbb{R}^n$ from incomplete and noisy measurements $y \in \mathbb{R}^m$, that is

$$y = Ax + \epsilon \tag{1}$$

where $A \in \mathbb{R}^{n \times m}$ is the forward sensing operator and ϵ is the noise affecting the measurements. This is generally an ill-posed task due to the incomplete forward operator A with $m < n$ (or $m = n$ with a large condition number) and the noise affecting the measurements. In order to reconstruct the signals, we need knowledge about the signal distribution and its support $\mathcal{X}$.

Classical approaches assume a signal distribution using some prior knowledge about the underlying signals x. For example, the well-known total variation model [1] is built on the prior belief that natural images are approximately piecewise smooth. This strategy often yields a loose description of the true model, providing biased and/or suboptimal reconstructions. In recent years, this approach has been replaced by learning the reconstruction mapping $y \mapsto x$ directly from data.

Despite the appeal and better performance of the learning-based approach [2], in many sensing applications we can only access incomplete measurements y, resulting in a chicken-and-egg problem : in order to reconstruct x we require knowledge about the signal model, but to learn this model we require ground truth training data x. Moreover, if the measurement process A is incomplete, it is fundamentally impossible to learn the signal distribution through only measurements y, as there is no information about the set of signals $\mathcal{X}$ in the nullspace of A. The recent *equivariant imaging* (EI) framework [3, 4], showed that this fundamental limitation can be overcome by exploiting the invariance of typical signal sets to transformations, such as translation, rotation or scaling.

In this paper, we extend the EI framework to compressive single-pixel imaging [5, 6]. This modality relies on a single detector and is used in applications where standard cameras with arrays of detectors are very expensive or impossible to build, such as sensing in the non-visible spectrum or ultrafast imaging [6]. In these applications, ground-truth signals x may be expensive or impossible to obtain, thus learning from single-pixel measurements alone is very important. The proposed method only requires compressed measurements for training and obtains a similar performance to fully supervised approaches.

2 Preliminaries

We begin with some basic definitions. As the number of measurements is lower than the dimension of the signal space, i.e., $m < n$, the operator A has a non trivial nullspace, denoted by $\mathcal{N}_A \subset \mathbb{R}^n$. The pseudo-inverse of A is denoted by $A^\dagger \in \mathbb{R}^{n\times m}$. The learning task consists of estimating the parameters $\theta \in \mathbb{R}^p$ of a reconstruction function $f_\theta : \mathbb{R}^m \to \mathbb{R}^n$ such that $f_\theta(y) \approx x$ for all $x \in \mathcal{X}$, by minimising a training loss $\mathcal{L} : \mathbb{R}^p \mapsto \mathbb{R}$:

$$\arg\min_\theta \mathcal{L}(\theta). \tag{2}$$

The reconstruction mapping f_θ is modeled as a deep neural network where θ are the weights and biases. We now discuss the choice of $\mathcal{L}(\theta)$ in the context of supervised and unsupervised learning.

Supervised Learning Standard learning methods learn the parameters θ using a dataset composed of N pairs of signals and associated measurements (x_i, y_i), by minimising the following supervised loss

$$\mathcal{L}_{\text{Sup}}(\theta) = \sum_{i=1}^{N} \|x_i - f_\theta(y_i)\|^2. \tag{3}$$

However, in many real-world settings, such as medical and astronomical imaging, obtaining ground-truth signals x_i can be very expensive or even impossible.

Unsupervised Learning In settings where we have access to only measurement data y_i, it is still possible to train the reconstruction function by enforcing measurement consistency, i.e.,

$$\mathcal{L}_{\text{MC}}(\theta) = \sum_{i=1}^{N} \|y_i - Af_\theta(y_i)\|^2. \tag{4}$$

Unfortunately, this approach is ill-fated if A has a non-trivial nullspace, as there are multiple possible reconstruction functions that satisfy measurement consistency :

Proposition 1 (Proposition 1 in [3])**.** *Any reconstruction function $f_\theta(y) : \mathbb{R}^m \mapsto \mathbb{R}^n$ of the form*

$$f_\theta(y) = A^\dagger y + v(y) \tag{5}$$

where $v(y) : \mathbb{R}^m \mapsto \mathcal{N}_A$ is any function whose image belongs to the nullspace of A verifies the measurement consistency requirement.

In other words, the measurement consistency loss doesn't contain any information about $\mathcal{X}$ in $\mathcal{N}_A$. We require some additional prior information about $\mathcal{X}$ in order to learn from measurement data alone, as we will show next.

3 Equivariant Imaging

Recently, the EI framework [3] showed that invariance to transformations, such as translations, rotations or reflections, can be enough to learn from measurement data alone. A signal set $\mathcal{X}$ is invariant to a group of invertible transformations $T_1, \dots, T_{|\mathcal{G}|} \in \mathbb{R}^{n\times n}$, if for all $x \in \mathcal{X}$, then $T_g x$ also belongs to $\mathcal{X}$ for all $g = 1, \dots, |\mathcal{G}|$. Most natural signals present invariance to a certain group of transformations. For example, sets of natural 2D images are generally assumed to be invariant to shifts, rotations and/or reflections.

Under the invariance assumption, we have

$$y = Ax = AT_gT_g^{-1}x = A_g x' \tag{6}$$

for $g = 1, \dots, |\mathcal{G}|$, where $A_g = AT_g$ and $x' = T_g^{-1}x$ belongs to the signal set. Thus, the invariance of $\mathcal{X}$ provides *implicit* access to the operators A_g with potentially different nullspaces, allowing us to learn beyond the nullspace of A.

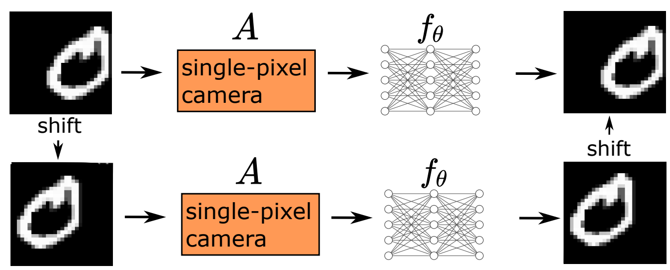

FIGURE 1 – **Equivariant imaging systems.** If the set of signals is invariant to a certain set of transformations, the composition of imaging operator A with the reconstruction function f_θ should be equivariant to these transformations.

As illustrated in Figure 1, due to the invariance of $\mathcal{X}$, the composition of the forward operator and reconstruction function $f_\theta \circ A$ must be *equivariant* to the group of transformations, that is

$$f_\theta(AT_gx) = T_g f_\theta(Ax) \tag{7}$$

for all $x \in \mathcal{X}$ and $g = 1, \dots, |\mathcal{G}|$. The equivariant constraint on $f_\theta \circ A$ can be enforced using the following (unsupervised) loss [3] :

$$\mathcal{L}_{\text{EQ}}(\theta) = \sum_{i=1}^{N} \frac{1}{|\mathcal{G}|} \sum_{g=1}^{|\mathcal{G}|} \|f_\theta(AT_g\tilde{x}_i) - T_g\tilde{x}_i\|^2 \tag{8}$$

where $\tilde{x}_i = f_\theta(y_i)$ for all $i = 1, \dots, N$. Taking into account both measurement consistency and equivariance of $f_\theta \circ A$, the *equivariant imaging* (EI) loss [3] is given by $\mathcal{L}_{\text{EI}}(\theta) = \mathcal{L}_{\text{MC}}(\theta) + \alpha\mathcal{L}_{\text{EQ}}(\theta)$, where α is a hyperparameter which controls the trade-off between equivariance and measurement consistency.

Analysis Under the assumption of invariance and noiseless measurements, i.e., $y = Ax$, we have

$$\mathbb{E}_y\{\mathcal{L}_{\text{MC}}(\theta)\} = \mathbb{E}_x\{\frac{1}{|\mathcal{G}|} \sum_{g\in\mathcal{G}} \|AT_gx - Af_\theta(AT_gx)\|^2\}. \tag{9}$$

Using the equivariance constraint in (7) we get

$$\mathbb{E}_y\{\mathcal{L}_{\text{MC}}(\theta)\} = \mathbb{E}_x\{\frac{1}{|\mathcal{G}|}\sum_{g\in\mathcal{G}} \|AT_g x - AT_g f_\theta(Ax)\|^2\} \quad (10)$$

$$= \mathbb{E}_{(x,y)}\{\|M\,(x - f_\theta(y))\,\|^2\} \quad (11)$$

where

$$M = \frac{1}{|\mathcal{G}|}\begin{bmatrix} AT_1 \\ \vdots \\ AT_{|\mathcal{G}|}\end{bmatrix}. \quad (12)$$

It is easy to see that the standard supervised loss in (3) is obtained by setting M as the identity matrix. If M has rank smaller than n, the unsupervised loss does not penalize reconstruction error in the nullspace of M. Thus, a necessary condition for unsupervised learning is that $M \in \mathbb{R}^{m|\mathcal{G}|\times n}$ is of rank n :

Proposition 2 (Theorem 1 in [3])**.** *A necessary condition for identifying the signal model $\mathcal{X}$ from compressed observations is that the matrix in* (12) *is of rank n.*

A detailed analysis of necessary and sufficient conditions for model identification can be found in [7].

3.1 Handling Noise with SURE

In most sensing applications, the observed measurements are corrupted by noise. The performance of the standard EI method degrades with the amount of noise affecting the observations, as the the measurement consistency loss $\mathcal{L}_{\text{MC}}(\theta)$ does not prevent f_θ from overfitting the noise. However, if the noise distribution is known, we can correct the measurement consistency loss using Stein's unbiased risk estimator (SURE) [4], such that the corrected loss is an unbiased estimator of the oracle measurement consistency with pairs of noiseless and noisy measurements (u_i, y_i), i.e.,

$$\mathbb{E}_y\{\mathcal{L}_{\text{MC-SURE}}(\theta)\} = \mathbb{E}_{(u,y)}\{\frac{1}{m}\|u - Af_\theta(y)\|^2\}$$

where $u_i := Ax_i$ are the noiseless measurements. If the noise is iid Gaussian, the SURE loss is given by

$$\mathcal{L}_{\text{MC-SURE}}(\theta) = \sum_{i=1}^{N}\frac{1}{m}\|y_i - Af_\theta(y_i)\|^2 - \sigma^2 + \frac{2\sigma^2}{m}\nabla\cdot h_\theta(y_i) \quad (13)$$

where $h_\theta = A\circ f_\theta$ and $\nabla\cdot$ is the divergence operator. For certain f_θ, computing the divergence can be challenging, so we instead approximate it using the Monte Carlo approach introduced by Ramani et al. [8], which only requires an additional evaluation of f_θ.

The SURE method can be applied to a large family of noise distributions : for example, expressions for Poisson and Poisson-Gaussian noise distributions can be found in [4]. The *robust equivariant imaging* (REI) loss is obtained by replacing $\mathcal{L}_{\text{MC}}(\theta)$ by the SURE-corrected one in $\mathcal{L}_{\text{EI}}(\theta)$, that is,

$$\mathcal{L}_{\text{REI}}(\theta) = \mathcal{L}_{\text{MC-SURE}}(\theta) + \alpha\mathcal{L}_{\text{EQ}}(\theta). \quad (14)$$

4 Single-Pixel Imaging

The single-pixel camera [5] is an imaging modality inspired by compressed sensing which uses a single sensor to measure incoming light modulated by spatial masks. These masks are programmable binary patterns which are set using spatial light modulators, digital micromirror devices or LED arrays. Each measurement obtained by the detector corresponds to the scene filtered by a binary pattern. As measuring more patterns requires longer acquisition times and more storage, generally only a few patterns $m \ll n$ are used to reconstruct the scene. Single-pixel cameras are particularly useful in settings where standard detector arrays are prohibitively expensive such as imaging non-visible spectra and ultrafast imaging. See [6] for a recent survey about this modality.

There are multiple ways to select the patterns : popular choices are iid random entries or Hadamard basis vectors [6]. Here we focus on patterns with random iid entries, such that the entries of the sensing operator are given by

$$A_{i,j} = \begin{cases} \frac{1}{\sqrt{n}} & \text{with probability 0.5} \\ -\frac{1}{\sqrt{n}} & \text{with probability 0.5} \end{cases} \quad (15)$$

for $i = 1,\ldots,m$ and $j = 1,\ldots,n$. Due to the iid patterns, the resulting forward operator is highly incoherent with most groups of transformations, thus $m \geq n/|\mathcal{G}|$ measurements are generally enough to verify the necessary condition in Proposition 2. For example, consider images of 28×28 pixels and the group of 2D shifts ($|\mathcal{G}| = 784$). The necessary condition is fulfilled with probability 0.94 with just $m = 1$ pattern, and probability 1 with $m \geq 2$ patterns (estimated using 100 random realizations of A).

5 Experiments

We evaluate the proposed method using the '0' and '1' digits of the MNIST dataset. The dataset consists of 1000 images for training, and 200 images for testing. The images have 28×28 pixels, i.e., $n = 784$, and are normalised to have intensity values in $[0, 1]$. The single-pixel measurements are obtained using patterns sampled according to (15) and corrupted by iid Gaussian noise with standard deviation $\sigma = 0.05$. We compare the following learning methods :

- **Pseudo-inverse** ($A^\dagger y$) : Linear reconstruction by applying the pseudo-inverse to the observed measurements $y_i = A^\dagger x_i$ (baseline of no learning).
- **Measurement Consistency** (MC) : Training from only compressed measurements y_i using $\mathcal{L}_{\text{MC}}(\theta)$ in (4).
- **Robust Equivariant Imaging** (REI) : Training from only compressed measurements y_i using $\mathcal{L}_{\text{REI}}(\theta)$ in (14) with the group of 2D shifts [1].
- **Supervised** (Sup) : Standard training using ground truth pairs (x_i, y_i) using $\mathcal{L}_{\text{Sup}}(\theta)$ in (3).

1. Note that the set of images of digits is approximately shift invariant.

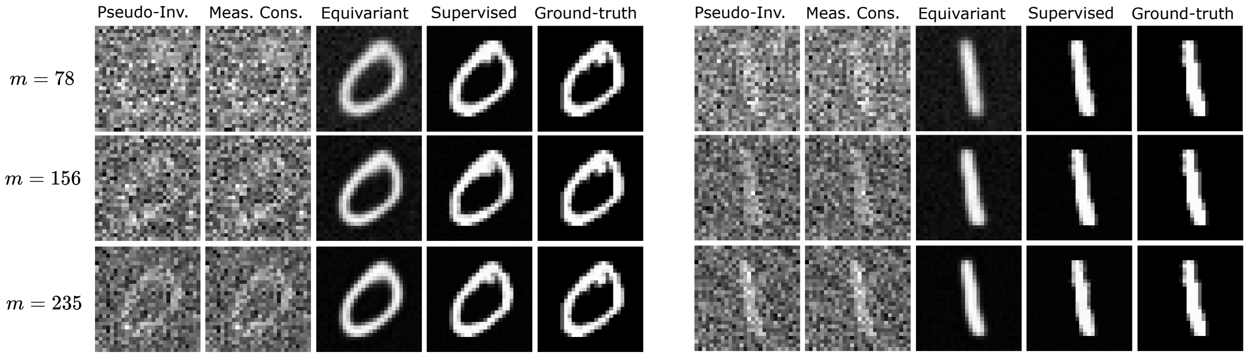

FIGURE 2 – Reconstructed test images with $m = 78, 156, 235$ patterns for the evaluated learning approaches. Networks trained with measurement consistency only fail to provide good reconstructions, performing similarly to the linear pseudo-inverse. By enforcing equivariance, the proposed unsupervised method obtains reconstructions close to the fully supervised setting.

We model f_θ as the U-Net network in [3] and use the Adam optimizer for all learning approaches. In all cases, each gradient step is performed using mini-batches of 4 images, and 3 transformations per mini-batch in the case of REI (chosen randomly from the set of $|\mathcal{G}|$ transformations). For the REI approach we use $\alpha = 5$. The experiment was performed for 3 different sensing matrices with 78, 156 and 235 patterns respectively.

TABLE 1 – Average test PSNR in dB obtained by the evaluated training methods for different number of patterns m.

m	$A^\dagger y$	MC	REI	Sup
78	10.5 ± 2.2	10.5 ± 2.2	20.3 ± 2.9	23.3 ± 3.8
156	10.9 ± 2.3	10.9 ± 2.3	23.0 ± 3.0	25.3 ± 3.6
235	11.5 ± 2.2	11.5 ± 2.2	24.4 ± 2.9	26.7 ± 3.4

Table 1 shows the average and standard deviation of the test peak signal-to-noise ratio (PSNR) obtained by the different methods. Figure 2 shows reconstructed test images. The unsupervised MC training approach obtains the same performance as the linear pseudo-inverse $A^\dagger y$, effectively failing to learn the signal set. As discussed in Section 2, enforcing only measurement consistency is not enough for learning the reconstruction function. On the other hand, the proposed unsupervised REI approach obtains a 10 dB improvement over the linear pseudo-inverse, with an average PSNR which is approximately 2.5 dB behind the fully supervised approach. Moreover, the REI approach obtains good reconstructions with as few as $m = 78$ measurements per image.

6 Conclusions and Future Work

We present a novel unsupervised learning approach for single-pixel imaging, where the reconstruction mapping is learned from measurement data alone. The proposed method can reconstructing single-pixel measurement data by relying only on weak invariance properties, and without imposing strong assumptions on the signal distribution. The unsupervised approach is especially important for scientific imaging applications where the goal is to recover information with the smallest amount of prior assumptions about the underlying signals.

We leave the implementation of the proposed method in a real single-pixel imaging system for future work.

Références

[1] Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D : nonlinear phenomena*, 60(1-4) :259–268, 1992.

[2] Gregory Ongie, Ajil Jalal, Christopher A Metzler, Richard G Baraniuk, Alexandros G Dimakis, and Rebecca Willett. Deep learning techniques for inverse problems in imaging. *IEEE Journal on Selected Areas in Information Theory*, 1(1) :39–56, 2020.

[3] Dongdong Chen, Julián Tachella, and Mike E. Davies. Equivariant imaging : Learning beyond the range space. In *Proc. Int. Conf. Comp. Vis. (ICCV)*, pages 4379–4388, Oct 2021.

[4] Dongdong Chen, Julián Tachella, and Mike E Davies. Robust equivariant imaging : a fully unsupervised framework for learning to image from noisy and partial measurements. In *Proc. Conf. Comp. Vis. and Patt. Recog. (CVPR)*, 2022.

[5] Marco F Duarte, Mark A Davenport, Dharmpal Takhar, Jason N Laska, Ting Sun, Kevin F Kelly, and Richard G Baraniuk. Single-pixel imaging via compressive sampling. *IEEE Signal Processing Magazine*, 25(2) :83–91, 2008.

[6] Graham M. Gibson, Steven D. Johnson, and Miles J. Padgett. Single-pixel imaging 12 years on : a review. *Opt. Express*, 28(19) :28190–28208, Sep 2020.

[7] Julián Tachella, Dongdong Chen, and Mike Davies. Sampling theorems for unsupervised learning in linear inverse problems. *arXiv preprint arXiv :2203.12513*, 2022.

[8] Sathish Ramani, Thierry Blu, and Michael Unser. Monte-Carlo SURE : A black-box optimization of regularization parameters for general denoising algorithms. *IEEE Transactions on image processing*, 17(9) :1540–1554, 2008.