



HAL
open science

Deep expectation-maximization for single-pixel image reconstruction with signal-dependent noise

Antonio Lorente Mur, Françoise Peyrin, Nicolas Ducros

► **To cite this version:**

Antonio Lorente Mur, Françoise Peyrin, Nicolas Ducros. Deep expectation-maximization for single-pixel image reconstruction with signal-dependent noise. *IEEE Transactions on Computational Imaging*, 2022, 10.1109/TCI.2022.3200841 . hal-03522038v3

HAL Id: hal-03522038

<https://hal.science/hal-03522038v3>

Submitted on 13 Sep 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Deep expectation-maximization for single-pixel image reconstruction with signal-dependent noise

A. Lorente Mur, F. Peyrin, *Member, IEEE*, N. Ducros, *Member, IEEE*,

Abstract—Image reconstruction from a sequence of a few linear measurements that are corrupted by signal-dependent normally distributed noise is an inverse problem with many biomedical imaging applications, such as computerized tomography and optical microscopy. In this study, we focus on single-pixel imaging, where the set-up acquires a down-sampled Hadamard transform of an image of the scene. Deep learning is a computationally efficient framework to solve inverse problems in imaging. Several neural-network architectures provide a link between deep and optimization-based image reconstruction methods. These deep-learning methods rely on a forward operator and lead to more interpretable networks. Here, we propose a novel network architecture obtained by unrolling an heuristic expectation-maximization algorithm. In particular, we compute the maximum *a posteriori* estimate of the unknown image given measurements corrupted by normally distributed signal-dependent noise. We show that the so-called expectation-maximization reconstruction network (EM-Net) applies to mixed Skellam-Gaussian noise models that are common in single-pixel imaging. We present reconstruction results from simulated and experimental single-pixel acquisitions. We show that EM-Net generalizes very well to noise levels not seen during training, despite having fewer learned parameters than alternative methods. The proposed EM-Net generally reconstructs images with fewer artifacts and higher signal-to-noise ratios, in particular in high-noise situations compared to other state of the art reconstruction algorithms that do not estimate the noise covariance.

Index Terms—Image reconstruction, deep learning, expectation-maximization, iterative algorithm, single-pixel imaging, Skellam-Gaussian noise.

I. INTRODUCTION

SINGLE-pixel imaging is a computational imaging configuration where a single point detector is used to recover an image [1]. It has been successfully applied to fluorescence microscopy [2], hyperspectral imaging [3], [4], diffuse optical tomography [5], image-guided surgery [6], and short-wave infrared imaging [7]. A single-pixel camera acquires the dot product between the image of a scene and some two-dimensional light patterns that are displayed sequentially using a spatial light modulator [8]. The image of the scene is then reconstructed from the raw measurements. To limit acquisition times, it is highly desirable to reduce the number of light patterns, which leads to an under-determined inverse problem.

This study was supported by the French National Research Agency (ANR), under Grant ANR-17-CE19-0003 (ARMONI Project), and performed within the framework of the LABEX PRIMES (ANR-11-LABX-0063) of Université de Lyon. This material is based upon work done on the PILoT facility (PILoT, INSA LYON, Bât. Blaise Pascal, 7 Av. Jean Capelle 69621 Villeurbanne).

The authors are with the University of Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, UJM Saint-Etienne, CREATIS CNRS UMR 5220, Inserm U1294, F-69621, Lyon, France.

(e-mail: antonio.lorente-mur@creatis.insa-lyon.fr; peyrin@esrf.fr; corresponding author : nicolas.ducros@creatis.insa-lyon.fr).

Recent advances in deep learning have revolutionized image reconstruction [9], [10]. In particular, convolutional neural networks (CNNs) are very efficient for solving the computed tomography problem, either by learning direct inverse mapping [11], or through the use of adversarial neural networks [12]. Much effort is being devoted to bridging the gap between more traditional model-based approaches for image reconstruction and deep-learning-based approaches for inverse problems. CNNs can provide sparsifying transforms [13] and model the manifold of natural images [14], or of a projector onto this space [15]. The CNN priors are then plugged into an unrolled optimization algorithm [16], [17], [18] or used in series expansions, like the Neumann series [19].

This trend also greatly benefits computational optics [20], [21]. In single-pixel imaging, in particular, fully learned mapping from the measurement space to the image space can outperform compressed sensing reconstructions [22]. [23] used a deep auto-regressive model that is plugged into an alternating direction method of multiplier algorithm for single-pixel images. In [24], we proposed a simple reconstruction CNN where the first layer computes the conditional expectation of the unknown image, given noiseless measurements. We generalized this idea to the case of noisy data with varying noise levels in [25], and we introduced an iterative architecture in [26].

Single-pixel measurements result from the difference between two Poisson variables. Therefore single-pixel data are corrupted by mixed Skellam-Gaussian noise, which has no closed-form expression. For similar problems, such as mixed Poissonian-Gaussian denoising, this difficulty is circumvented by using variance stabilizing transforms, such as the generalized Anscombe transform [27], before processing the Gaussian noise [28]. However, this introduces a bias [29]. Alternative approaches rely on approximations. In [30], [31] and [32], the likelihood of mixed Poisson-Gaussian is approximated as a finite sum. The Poisson-Gaussian unbiased risk-estimator linear expansion of thresholds [33] seeks to approximate an unbiased estimate of the mean squared error with linear combinations of Wiener filters followed by a thresholding of Haar wavelet coefficients. Other approaches simplify the problem by either approximating the Poisson contribution by normally distributed signal-dependent noise [34] or by neglecting the normal distribution. The maximum *a posteriori* (MAP) estimate can then be computed using an expectation-maximization (EM) algorithm [35]. However, the resulting optimization problem is usually solved using hand-crafted priors (*e.g.*, Hessian-based or total variation [TV]-based penalization).

A. Contribution

In this paper, we derive and analyze a deep neural-network architecture based on a heuristic EM algorithm. The so-called EM reconstruction network (EM-Net) can tackle inverse problems with signal-dependent noise, such as those involved in computational optics. We significantly extend from our preliminary results [26] in three directions. First, we model single-pixel reconstruction as a mixed Skellam-Gaussian reconstruction problem, for which we derive the corresponding likelihood. As this is intractable numerically, we propose a normal approximation that leads to signal-dependent covariance that cannot be exactly estimated. Secondly, we propose a five-step iterative algorithm, where the image prior is learnt from training examples, and we show how to reduce the number of learnable parameters. We also propose a training strategy where most of the parameters of our EM-Net can be precomputed. Thirdly, we conduct extensive numerical comparisons with data-driven reconstruction methods designed for single-pixel imaging, and also other modalities. As the noise level is usually unknown, in practical experiments we focus on the robustness of the methods in the presence of noise, where the level differs from that used during training. We consider both simulated measurements corrupted by mixed Skellam-Gaussian noise and experimental data.

The choice of the EM algorithm is motivated by the unknown noise covariance matrix (although we have access to an estimation of it). This makes the use of standard methods challenging, as the MAP cost function depends on the target image and the unknown signal-dependent noise covariance matrix. The EM algorithm allows us to separate the unknown noise covariance into an approximated covariance and the covariance of the complete data. As the block of the covariance of the complete data corresponding to the acquired measurement is learned, the proposed method learns how to correct the estimated noise covariance. It is also important to note that the EM algorithm leads to a generalized Tikhonov regularization problem for the data-consistency step, while most unrolled methods use “normal” Tikhonov regularization; *i.e.*, the measurement covariance, and sometimes also the image covariance, are assumed to be identity matrices. This provides a stronger prior for the data consistency step of the unrolled algorithm.

Upon acceptance, our reconstruction method will be made available in the Python toolbox SPyRiT[36].

B. Organization of the paper

In Section II, we model single-pixel acquisition and describe the associated reconstruction problem. In Section III, we introduce a deep network based on the EM algorithm. In Section IV, we describe the numerical experiments and experimental data that we considered to validate our approach. Finally, we analyze our reconstruction results in Section V.

C. Notations

Throughout this paper, deterministic variables are indicated by italics. In particular, we use normal font letters to denote

scalars (*e.g.*, $x \in \mathbb{R}$), lowercase bold letters for vectors (*e.g.*, $\mathbf{x} \in \mathbb{R}^N$), and capital bold letters for matrices (*e.g.*, $\mathbf{X} \in \mathbb{R}^{N \times M}$). The i -th element of the vector $\mathbf{x} \in \mathbb{R}^N$ is denoted by $(\mathbf{x})_i$. $\text{Diag}(\mathbf{x}) \in \mathbb{R}^{N \times N}$ is a diagonal matrix, where the diagonal elements are given by $\mathbf{x} \in \mathbb{R}^N$, while $\text{diag}(\mathbf{A}) \in \mathbb{R}^N$ is a vector, where the elements are taken from the main diagonal of the matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$. We define the weighted squared norm as $\|\mathbf{x}\|_{\mathbf{A}}^2 = \mathbf{x}^\top \mathbf{A} \mathbf{x}$, where \mathbf{A} is a positive definite matrix.

We use nonitalic symbols to denote random vectors. An N -dimensional random vector \mathbf{x} that follows a multivariate normal distribution with mean $\boldsymbol{\mu} \in \mathbb{R}^N$ and covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{N \times N}$ is written as $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. An N -dimensional random vector \mathbf{x} is said to follow a Poisson distribution with mean $\boldsymbol{\mu} \in \mathbb{R}^N$, in notations $\mathbf{x} \sim \mathcal{P}(\boldsymbol{\mu})$, if it contains N independent Poisson-distributed random variables. We denote the expected value of a random variable \mathbf{x} by $\mathbb{E}(\mathbf{x})$. $\mathbb{E}(\mathbf{x} | \mathbf{y})$ is shorthand for $\mathbb{E}(\mathbf{x} | \mathbf{y} = \mathbf{y})$.

II. COMPRESSIVE SINGLE-PIXEL ACQUISITION

A. Compressed single-pixel acquisition

Let $\mathbf{f} \in [0, 1]^N$ be the image to be acquired. The main idea of compressive single-pixel imaging is to measure $\mathbf{y} = \mathbf{H}\mathbf{f}$ using hardware, and to recover \mathbf{f} using software. The matrix $\mathbf{H} = [\mathbf{h}_1, \dots, \mathbf{h}_N] \in \mathbb{R}^{N \times N}$ collects the patterns that are sequentially uploaded in a spatial light modulator. These patterns can be chosen as a basis matrix, such as Fourier, discrete cosines, wavelets, and Hadamard bases [37]. In practice, to accelerate acquisition times, only a few patterns in a basis are acquired. As the spatial light modulator cannot implement negative values, the patterns \mathbf{H} are split into positive and negative patterns \mathbf{H}_+ and \mathbf{H}_- , such that $(\mathbf{H}_+)_{i,n} = \max((\mathbf{H})_{i,n}, 0)$ and $(\mathbf{H}_-)_{i,n} = \max(-(\mathbf{H})_{i,n}, 0)$ (see [38] for details).

We model the raw measurement as mixed Poisson-Gaussian noise [39]

$$\hat{\mathbf{m}}_{+,-}^\alpha \sim g\mathcal{P}(\alpha\mathbf{S}\mathbf{H}_{+,-}\mathbf{f}) + \mathcal{N}(\mu_{\text{dark}}\mathbf{1}, \sigma_{\text{dark}}^2\mathbf{I}) \quad (1)$$

where \mathcal{P} and \mathcal{N} are the Poisson and Gaussian distributions, $\mathbf{S} = [\mathbf{I}_M, \mathbf{0}] \in \mathbb{R}^{M \times N}$ is a down-sampling matrix (where $M \leq N$), g is a constant that represents the overall system gain (in counts/electron), α is the intensity (in photons) of the image, μ_{dark} is the dark current (in counts), and σ_{dark} is the dark noise (in counts). We further hypothesize that g , μ_{dark} and σ_{dark} are scalars independent of the image intensity α , which can be estimated by calibration. We choose a Hadamard matrix ordered by decreasing variance (*i.e.*, we chose \mathbf{H} such that $\mathbb{E}(\|\mathbf{S}\mathbf{H}\mathbf{f}\|_2^2)$ is maximal), as suggested in [40].

Rather than working on the raw measurements directly, they are first combined and normalized

$$\mathbf{m}^\alpha = (\hat{\mathbf{m}}_+^\alpha - \hat{\mathbf{m}}_-^\alpha) / (\alpha g). \quad (2)$$

As the problem now reads $\mathbb{E}(\mathbf{m}^\alpha) = \mathbf{S}\mathbf{H}\mathbf{f}$, where \mathbf{H} is a Hadamard matrix, we will benefit from fast forward and inverse transforms. This will reduce the computational burden of the reconstruction algorithm at the cost of a loss of information.

Plugging the noise model of the raw measurements given by Equation (1) into the previous equation leads to the mixed Skellam-Gaussian noise model

$$\mathbf{m}^\alpha \sim \frac{1}{\alpha} \mathcal{S}_k(\alpha \mathbf{S} \mathbf{H}_+ \mathbf{f}, \alpha \mathbf{S} \mathbf{H}_- \mathbf{f}) + \frac{2}{\alpha K} \mathcal{N}(\mathbf{0}, \sigma_{\text{dark}}^2 \mathbf{I}), \quad (3)$$

where the Skellam distribution [41] $\mathcal{S}_k(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)$ is the distribution of the difference between two Poisson-distributed variables with means of $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$.

B. Normal approximation of the mixed Skellam-Gaussian noise model

The likelihood of the mixed Skellam-Gaussian noise model involves an infinite sum of exponentials multiplied by modified Bessel functions (see Appendix A for the derivation). Exploitation of this analytical expression is numerically intractable, as it involves an infinite series and the evaluation of Bessel functions.

$$p(\mathbf{m}|\mathbf{f}) \propto \prod_{i=1}^M \sum_{n \in \mathbb{N}} e^{-\alpha(\mathbf{h}_{i,+}^\top \mathbf{f} + \mathbf{h}_{i,-}^\top \mathbf{f})} \left(\frac{\mathbf{h}_{i,+}^\top \mathbf{f}}{\mathbf{h}_{i,-}^\top \mathbf{f}} \right)^{\frac{n - \alpha \mathbf{h}_{i,+}^\top \mathbf{f}}{2\alpha}} \mathcal{I}_{|\frac{n}{\alpha} - \mathbf{h}^\top \mathbf{f}|} \left(2\alpha \sqrt{(\mathbf{h}_{i,+}^\top \mathbf{f})(\mathbf{h}_{i,-}^\top \mathbf{f})} \right) e^{-\frac{(m_i - \frac{n}{\alpha})^2}{\beta^2}}. \quad (4)$$

Similar to work that has addressed mixed Poisson-Gaussian noise, we propose to introduce an approximation. As the Gaussian distributions are good approximations of Poisson distributions with means greater than 5 (see [42]) – single-pixel measurements are typically several orders of magnitude above this – we use the normal approximation to the Poisson distributions in Equation (3). We obtain

$$\mathbf{m}^\alpha = \mathbf{S} \mathbf{H} \mathbf{f} + \boldsymbol{\epsilon}^\alpha, \text{ with } \boldsymbol{\epsilon}^\alpha \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\alpha(\mathbf{f})) \quad (5)$$

where the noise covariance $\boldsymbol{\Sigma}_\alpha(\mathbf{f})$ is given by

$$\boldsymbol{\Sigma}_\alpha(\mathbf{f}) = \text{Diag}(\boldsymbol{\sigma}_\alpha(\mathbf{f})), \text{ with} \quad (6)$$

$$\boldsymbol{\sigma}_\alpha(\mathbf{f}) = \frac{1}{\alpha} \mathbf{S} \mathbf{H}_+ \mathbf{f} + \frac{1}{\alpha} \mathbf{S} \mathbf{H}_- \mathbf{f} + \frac{2\sigma_{\text{dark}}^2}{g^2 \alpha^2}. \quad (7)$$

Note that the noise variance $\boldsymbol{\sigma}_\alpha(\mathbf{f})$ is unknown, as it depends on the unknown image \mathbf{f} .

III. PROPOSED DEEP EM NETWORK

A. Objective

We adopt a Bayesian framework and reconstruct \mathbf{f} from \mathbf{m}^α by computing a point-wise estimator of $p(\mathbf{f}|\mathbf{m}^\alpha)$. In particular, we aim to compute the MAP solution that solves

$$\underset{\mathbf{f}}{\text{argmax}} \log p(\mathbf{m}^\alpha|\mathbf{f}) + \log p(\mathbf{f}), \quad (8)$$

where we assume the probability density function $p(\mathbf{m}^\alpha|\mathbf{f}) \propto \exp\left(-\frac{1}{2}\|\mathbf{S} \mathbf{H} \mathbf{f} - \mathbf{m}^\alpha\|_{\boldsymbol{\Sigma}_\alpha^{-1}(\mathbf{f})}^2\right)$ according to Equation (5), while $p(\mathbf{f})$ is an unknown probability density function.

Since $\boldsymbol{\Sigma}_\alpha(\mathbf{f})$ is a signal-dependent noise, we cannot use signal-independent strategies for solving this problem.

B. The EM algorithm

Classical derivations of EM algorithms [35] have commonly been used to estimate the MAP for image reconstruction tasks [43]. EM algorithms are iterative algorithms that produce a sequence of estimations $\{\mathbf{f}^{(k)}\}$ such that the sequence $\{\log p(\mathbf{f}^{(k)}|\mathbf{m}^\alpha)\}_k$ is monotonically nondecreasing. Here, we make two heuristic modifications that lead to an algorithm called EM-Net. First, we apply signal-independent formulas to signal-dependent noise covariances. Second, we introduce a deep neural network. This leads to a practically useful algorithm; however, it lacks the convergence guarantees offered by the classical EM algorithm.

Every iteration of the EM algorithm is based on two steps: the expectation step, and the maximization step. The expectation step computes the conditional expectation of the log-likelihood of \mathbf{f} with respect to an auxiliary random variable \mathbf{x} , given the current estimate $\mathbf{f}^{(k)} \in \mathbb{R}^N$ and the measurements $\mathbf{m}^\alpha \in \mathbb{R}^M$

$$\mathcal{Q}(\mathbf{f}|\mathbf{f}^{(k)}) = \mathbb{E}_{\mathbf{x}}[\log p(\mathbf{x}|\mathbf{f})|\mathbf{m}^\alpha, \mathbf{f}^{(k)}] + p(\mathbf{f}). \quad (9)$$

In the EM literature, the variable $\mathbf{x} \in \mathbb{R}^N$ is referred to as the complete data, as it may contain some unobserved data.

This conditional expectation is then maximized with respect to \mathbf{f} during the maximization step, to produce the next iteration

$$\mathbf{f}^{(k+1)} = \underset{\mathbf{f}}{\text{argmax}} \mathcal{Q}(\mathbf{f}|\mathbf{f}^{(k)}). \quad (10)$$

The EM algorithm requires the complete data \mathbf{x} to satisfy some admissibility properties. As $p(\mathbf{m}^\alpha|\mathbf{f})$ is Gaussian, \mathbf{x} and $\mathbf{m}^\alpha|\mathbf{x}$ can be chosen as Gaussian vectors [44]. In particular, we will choose to interpret \mathbf{x} as the full measurement vector and $\mathbf{m}^\alpha|\mathbf{x}$ as the subsampled measurements by setting

$$\mathbf{x} \sim \mathcal{N}(\mathbf{H} \mathbf{f}, \mathbf{C}) \quad \text{and} \quad (11)$$

$$\mathbf{m}^\alpha|\mathbf{x} \sim \mathcal{N}(\mathbf{S} \mathbf{x}, \tilde{\boldsymbol{\Sigma}}), \quad (12)$$

where \mathbf{C} and $\tilde{\boldsymbol{\Sigma}}$ are two design covariance matrices that must follow some consistency conditions. We further detail the choice of these design matrices and the consistency condition in Section III-D.

C. Unrolling the EM algorithm

Under the Gaussian assumption with signal-independent covariance of Equation (11)–Equation (12), the expectation step of Equation (9) simplifies to [44]

$$\bar{\mathbf{x}}^{(k)} = \mathbb{E}(\mathbf{x}|\mathbf{m}^\alpha, \mathbf{f}^{(k)}), \quad (13a)$$

$$\mathcal{Q}(\mathbf{f}|\mathbf{f}^{(k)}) = \log p(\bar{\mathbf{x}}^{(k)}|\mathbf{f}) + \log p(\mathbf{f}). \quad (13b)$$

Using classical properties of Gaussian vectors (see Chapter 5 of [45]), we can rewrite these steps as

$$\bar{\mathbf{x}}^{(k)} = \underset{\mathbf{x}}{\text{argmin}} \|\mathbf{S} \mathbf{x} - \mathbf{m}^\alpha\|_{(\tilde{\boldsymbol{\Sigma}}^{(k)})^{-1}}^2 + \|\mathbf{x} - \mathbf{H} \mathbf{f}^{(k)}\|_{(\mathbf{C}^{(k)})^{-1}}^2 \quad (14a)$$

$$\mathbf{f}^{(k+1)} = \underset{\mathbf{f}}{\text{argmin}} \|\bar{\mathbf{x}}^{(k)} - \mathbf{H} \mathbf{f}\|_{(\mathbf{C}^{(k)})^{-1}}^2 - \log p(\mathbf{f}). \quad (14b)$$

Note that the design covariance matrices can be chosen for each iteration independently. As $p(\mathbf{f})$ is unknown, we propose to replace Equation (14b) by a nonlinear model $\mathcal{D}_{\omega^{(k)}}$

$$\tilde{\mathbf{x}}^{(k)} = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{S}\mathbf{x} - \mathbf{m}^\alpha\|_{(\tilde{\Sigma}^{(k)})^{-1}}^2 + \|\mathbf{x} - \mathbf{H}\mathbf{f}^{(k)}\|_{\mathbf{C}^{(k)}^{-1}}^2 \quad (15a)$$

$$\mathbf{f}^{(k+1)} = \mathcal{D}_{\omega^{(k)}}\left(\frac{1}{N}\mathbf{H}^\top \tilde{\mathbf{x}}^{(k)}\right), \quad (15b)$$

where $\omega^{(k)}$ represents the parameters of the model \mathcal{D} , which will be optimized during a training phase, such that Equation (15b) solves Equation (14b).

In the literature on deep unrolled methods, Equation (15a) is commonly called the data-consistency layer. Introducing the variable $\mathbf{y}^{(k)} = \tilde{\mathbf{x}}^{(k)} - \mathbf{H}\mathbf{f}^{(k)}$, the analytical solution of Equation (15a) is given by (see Appendix B)

$$\mathbf{y}^{(k)} = \begin{bmatrix} \mathbf{C}_1^{(k)} \\ \mathbf{C}_2^{(k)} \end{bmatrix} (\tilde{\Sigma}^{(k)} + \mathbf{C}_1^{(k)})^{-1} (\mathbf{m}^\alpha - \mathbf{S}\mathbf{H}\mathbf{f}^{(k)}), \quad (16)$$

where $\mathbf{C}_1^{(k)} \in \mathbb{R}^{M \times M}$ and $\mathbf{C}_2^{(k)} \in \mathbb{R}^{(N-M) \times M}$ are the blocks of the covariance $\mathbf{C}^{(k)}$ defined by

$$\mathbf{C}^{(k)} = \begin{bmatrix} \mathbf{C}_1^{(k)} & (\mathbf{C}_2^{(k)})^\top \\ \mathbf{C}_2^{(k)} & \mathbf{C}_3^{(k)} \end{bmatrix}. \quad (17)$$

D. Choice of the design covariance matrices

The covariance matrices $\mathbf{C}^{(k)}$ and $\tilde{\Sigma}^{(k)}$ can be chosen in any way provided that the variance of $\mathbf{m}^\alpha | \mathbf{f}$ derived from Equations (11)–(12) equals that from Equation (7). Therefore, the following consistency condition applies to each iteration:

$$\Sigma_\alpha(\mathbf{f}) = \tilde{\Sigma}^{(k)} + \mathbf{S}\mathbf{C}^{(k)}\mathbf{S}^\top \quad (18)$$

$$= \tilde{\Sigma}^{(k)} + \mathbf{C}_1^{(k)}. \quad (19)$$

Here, we choose $\tilde{\Sigma}^{(k)}$ as the noise covariance of the iterate $\mathbf{f}^{(k)}$ obtained from Equation (7), i.e.,

$$\tilde{\Sigma}^{(k)} = \Sigma_\alpha(\mathbf{f}^{(k)}). \quad (20)$$

Our idea is to approximate the true variance from the variance of the current iterate, while $\mathbf{C}_1^{(k)}$ is a correction term. Note that $\mathbf{C}_1^{(k)}$ depends on \mathbf{f} ; however, we chose to neglect this dependence to simplify the derivation of Equation (13).

Moreover, this choice both reduces the memory footprint and computational burden of Equation (16). Indeed, as $\Sigma_\alpha(\mathbf{f})$ is diagonal, choosing a diagonal $\tilde{\Sigma}^{(k)}$ implies that $\mathbf{C}_1^{(k)}$ is diagonal too. Hence, we choose

$$\mathbf{C}_1^{(k)} = \operatorname{Diag}(\mathbf{c}_1^{(k)}) \quad (21)$$

and Equation (16) simplifies to

$$\mathbf{y}_1^{(k)} = \mathbf{c}_1^{(k)} / (\mathbf{c}_1^{(k)} + \tilde{\sigma}^{(k)}) (\mathbf{m}^\alpha - \mathbf{S}\mathbf{H}\mathbf{f}^{(k)}) \quad (22a)$$

$$\mathbf{y}_2^{(k)} = \mathbf{C}_2^{(k)} \operatorname{Diag}(1/\mathbf{c}_1^{(k)}) \mathbf{y}_1^{(k)} \quad (22b)$$

where the division and multiplication between vectors apply element-wise. As there is no consistency condition on $\mathbf{C}_2^{(k)}$, we choose it such that

$$\mathbf{y}_2^{(k)} = \mathbf{R}^{(k)} \mathbf{y}_1^{(k)}, \quad (23)$$

where $\mathbf{R}^{(k)} \in \mathbb{R}^{(N-M) \times M}$ is a matrix that is precomputed (see Section III-F for more details).

E. Algorithm summary and interpretation

We can summarize our algorithm by the following five steps

$$\tilde{\sigma}^{(k)} = \frac{1}{\alpha} (\mathbf{S}\mathbf{H}_+ \mathbf{f}^{(k)} + \mathbf{S}\mathbf{H}_- \mathbf{f}^{(k)}) + \frac{2\sigma_{\text{dark}}^2}{g^2 \alpha^2} \mathbf{I}_M, \quad (24a)$$

$$\mathbf{y}_1^{(k)} = \mathbf{c}_1^{(k)} / (\mathbf{c}_1^{(k)} + \tilde{\sigma}^{(k)}) (\mathbf{m}^\alpha - \mathbf{S}\mathbf{H}\mathbf{f}^{(k)}), \quad (24b)$$

$$\mathbf{y}_2^{(k)} = \mathbf{R}^{(k)} \mathbf{y}_1^{(k)}, \quad (24c)$$

$$\tilde{\mathbf{f}}^{(k)} = \mathbf{f}^{(k)} + \frac{1}{N} \mathbf{H}^\top \mathbf{y}^{(k)}, \quad (24d)$$

$$\mathbf{f}^{(k+1)} = \mathcal{D}_{\omega^{(k)}}(\tilde{\mathbf{f}}^{(k)}). \quad (24e)$$

which can be seen as the unrolled network depicted in Fig. 1. At each iteration, we estimate the measurement variance, the first layer denoises the measurements, the second completes the missing measurements, the third maps the completed measurements to the image space, and the last denoises the solution in the image space.

We refer to this network as EM-Net, and denote it by

$$\mathbf{f}^{(K)} = \mathcal{G}_\theta^K(\mathbf{m}^\alpha) \quad (25)$$

where K is the number of EM iterations. For $\mathbf{f}^{(0)} = \mathbf{0}$, the case $K = 1$ corresponds to the denoised completion network (i.e., the DC-Net) proposed in [25].

F. EM-Net training strategy and initialization

Given an image database $\{\mathbf{f}_s\}$, $1 \leq s \leq S$ and the corresponding measurements $\{\mathbf{m}_s^\alpha\}$ and $1 \leq s \leq S$, simulated via Equation (2)–Equation (3), we train the neural network in an end-to-end fashion

$$\min_{\theta} \frac{1}{S} \sum_{s=1}^S \|\mathcal{G}_\theta^{(K)}(\mathbf{m}_s^\alpha) - \mathbf{f}_s\|_2^2 + \lambda \|\theta\|_2^2, \quad (26)$$

where θ represents the learnable parameters of the network, K is the number of iterations, and λ is the weight decay parameter. We set $\mathbf{f}^{(0)} = \mathbf{0}$ and choose to learn the variance correction $\mathbf{c}_1^{(k)}$, for all iterations except the initial one, and the parameters of the image-domain denoiser for all iterations, i.e.,

$$\theta = \{\mathbf{c}^{(k)}, \omega_1^{(k')}\}, \quad 1 \leq k \leq K \text{ and } 0 \leq k' \leq K. \quad (27)$$

To precompute $\mathbf{c}^{(0)}$ and $\mathbf{R}^{(k)}$, we exploit the interpretation of Equation (14), i.e., $\mathbf{H}\mathbf{f}^{(k)}$ and $\mathbf{C}^{(k)}$ are the mean and covariance of the prior $\mathbf{x}^{(k)}$. We extract $\mathbf{c}^{(0)}$ from the diagonal of $\operatorname{Cov}(\mathbf{H}\mathbf{f})$, which is estimated prior to training as

$$\mathbf{C}^{(0)} = \frac{1}{S-1} \sum_{s=1}^S \mathbf{H}(\mathbf{f}_s - \boldsymbol{\mu}^{(0)})(\mathbf{f}_s - \boldsymbol{\mu}^{(0)})^\top \mathbf{H}^\top, \quad (28)$$

where $\boldsymbol{\mu}^{(0)} = \frac{1}{S} \sum_{s=1}^S \mathbf{f}_s$. Accordingly, we compute $\mathbf{R}^{(0)}$ as $\mathbf{C}_2^{(0)}(\mathbf{C}_1^{(0)})^{-1}$. The initial measurement variance is estimated from the raw measurements as in [25]

$$\tilde{\sigma}^{(0)} = \frac{1}{\alpha^2 K} (\hat{\mathbf{m}}_+^\alpha + \hat{\mathbf{m}}_-^\alpha - 2\mu_{\text{dark}} \mathbf{1}) + \frac{2\sigma_{\text{dark}}^2}{K^2 \alpha^2} \mathbf{1}. \quad (29)$$

Then, we precompute $\mathbf{R}^{(k)}$, $1 \leq k \leq K$, after having trained a one-iteration EM-Net and estimated $\mathbf{C}^{(1)}$ as

$$\mathbf{C}^{(1)} = \frac{1}{S-1} \sum_{s=1}^S \mathbf{H}(\mathbf{f}_s - \mathbf{f}_s^{(1)})(\mathbf{f}_s - \mathbf{f}_s^{(1)})^\top \mathbf{H}^\top, \quad (30)$$

where $\mathbf{f}_s^{(1)} = \mathcal{G}_\theta^{(1)}(\mathbf{m}_s^\alpha)$. Finally, we compute $\mathbf{R}^{(k)}$, $1 \leq k \leq K$, as $\mathbf{C}_2^{(1)}(\mathbf{C}_1^{(1)})^{-1}$. As there is no consistency condition on $\mathbf{C}_2^{(k)}$, $\mathbf{R}^{(k)}$ can be chosen as a constant. This helps to reduce the computational burden of the algorithm. Note that we do not enforce the consistency condition on $\mathbf{c}_1^{(k)}$ explicitly, but we rely on end-to-end training to find appropriate values of $\mathbf{c}_1^{(k)}$ for a given $\mathbf{R}^{(k)}$. Nevertheless, the consistency Equation (19) justifies our choice for a diagonal matrix \mathbf{C}_1 .

IV. EXPERIMENTS

A. Comparison methods

We implemented two variants of the proposed EM-Net. The first one EM-Net shared (EM-Net-s) considers that the weights of the image-domain network in Equation (24e) are shared across iterations (i.e., $\omega^{(k)} = \omega$, for all k), while the weights of the second variant, EM-Net un-shared (EM-Net-u) are not shared (i.e., $\omega^{(k)}$ can differ for different values of k). This greatly reduces the number of parameters to be optimized (see Table I).

The two variants of the proposed EM-Net are compared with four methods. We first consider a hand-crafted MAP solution where the prior $\log p(\mathbf{f})$ is the total variation, which we compute using the EM algorithm and the split-Bregman algorithm [46] (EM-TV) implemented with Pylops [47]. The other three methods are data-driven approaches based on deep neural networks. First, we postprocess the Moore-Penrose pseudo-inverse using a neural network chosen as either the CNN presented in Fig. 1.b or a U-Net [48]. We refer to the two variants as CNN and U-Net, respectively. Secondly, we consider the model-based reconstruction with deep learned priors (MoDL) [14]. MoDL is an unrolled algorithm with a convolutional neural-network-based regularization prior. Finally, we consider a Neumann network (NN) [19], as a neural-network architecture based on the Neumann series decomposition.

Model-based reconstruction with deep learned priors is a particular case of the proposed deep-EM algorithm, which assumes additive signal-independent Gaussian noise; i.e., $\Sigma_2 = \mathbf{0}$ and $\Sigma_1 = \lambda \mathbf{I}$ in Equation (24), where λ is a regularization term. CNN and U-Net are one-iteration EM-Nets that further assume no noise; i.e., $K = 1$ and $\Sigma^\alpha = \mathbf{0}$.

B. Datasets

We train the networks using two different datasets, and also consider an experimental dataset to test them.

1) *STL-10*: The STL-10 [49] dataset is an image dataset that contains 120,000 images, as 100,000 unlabeled images, 5,000 labeled images called the ‘training’ set, and 8,000 images called the ‘test’ set. Our methods are trained using the 105,000 images from the training and unlabeled sets, and are tested on the test set. We randomly crop 64×64 patches from the original 96×96 images, and transform these into gray-scale images.

2) *ImageNet*: The ImageNet dataset [50] is a large-scale dataset that includes a training set of 1,281,167 images, and 50,000 validation images. Here, we consider the ‘down-sampled ImageNet’ [51] dataset, where all of the images have been resized to 64×64 images.

3) *SPIHIM*: The SPIHIM dataset [52] contains 20 hyperspectral cubes acquired with a single-pixel camera. Each hyperspectral image has 2,048 channels in the visible range; however, we limit ourselves to the channel $\lambda = 610$ nm. As the dataset has a small number of hyperspectral images, we did not train any neural network based on this dataset; instead, we used this dataset to validate the generalization of our algorithms.

C. Metrics

We evaluate our reconstructed images with the peak signal-to-noise ratio (PSNR) and the structural similarity (SSIM). Given the ground-truth image \mathbf{f} , we compute the PSNR of a reconstructed image \mathbf{f}^* as

$$\text{PSNR}(\mathbf{f}^*, \mathbf{f}) = 10 \log_{10} \frac{2^2}{\|\mathbf{f}^* - \mathbf{f}\|_2^2}. \quad (31)$$

Similarly, we compute the SSIM as

$$\text{SSIM}(\mathbf{f}, \mathbf{f}^*) = \frac{(2\mu\mu_* + d_1)(2\tilde{\sigma} + d_2)}{(\mu^2 + \mu_*^2 + d_1)(\sigma^2 + \sigma_*^2 + d_2)} \quad (32)$$

where μ is the average of \mathbf{f} , μ_* is the average of \mathbf{f}^* , σ is the variance of \mathbf{f} , σ_* is the variance of \mathbf{f}^* , $\tilde{\sigma}$ is the covariance of \mathbf{f} and \mathbf{f}^* , $d_1 = (k_1 L)^2$ and $d_2 = (k_2 L)^2$ are two variables that are meant to stabilize the division, where L is the dynamic range of pixel values, and $k_1 = 0.01$ and $k_2 = 0.03$ by default.

D. Implementation and training details

In our experiments, we consider $M = 512$ and $M = 1,024$ Hadamard patterns of size $N = 64 \times 64$ pixels (which corresponds to a compression rate (CR= M/N) of 1/8 for $M = 512$ and 1/4 for $M = 1024$ respectively) from both the STL-10 and ImageNet databases. We set the number of iterations to $K = 5$ for all of the iterative networks (i.e., MoDL, NN, EM-Net). MoDL, NN, and the proposed EM-Net-u and EM-Net-s rely on a learned image-domain denoiser. For MoDL and NN, we chose this as a U-Net, which is the state-of-the-art architecture. For EM-Net, we considered a tiny CNN with fewer parameters. The numbers of learned parameters were 28,993 for CNN, 499,985 for U-Net, 499,986 for MoDL, 499,986 for NN, $28,993 + M \times (K - 1)$ for EM-Net-s and $28,993 \times K + M \times (K - 1)$ for EM-Net-u and are reported in Table I. The weights of the image-domain denoiser in MoDL, NN, and EM-Net were initialised as the weights of a Gaussian denoiser. All of the networks were implemented using Pytorch [53], and trained in an end-to-end fashion for the same noise level corresponding to $\alpha = 10$ photons. We considered the ADAM optimiser for 100 epochs, with an initial step size of 10^{-3} that we divided by a factor of 5 every 10 epochs. We set the weight decay regularization parameter to 10^{-6} . We stopped the training early after 60 epochs if the networks had reached convergence (i.e., the relative variations of the cost function were below 1%).

V. RESULTS AND DISCUSSION

A. Reconstruction metric for the STL-10 and ImageNet datasets

Table II gives the comparisons of the reconstruction qualities of the four methods on 200 images from the STL-10

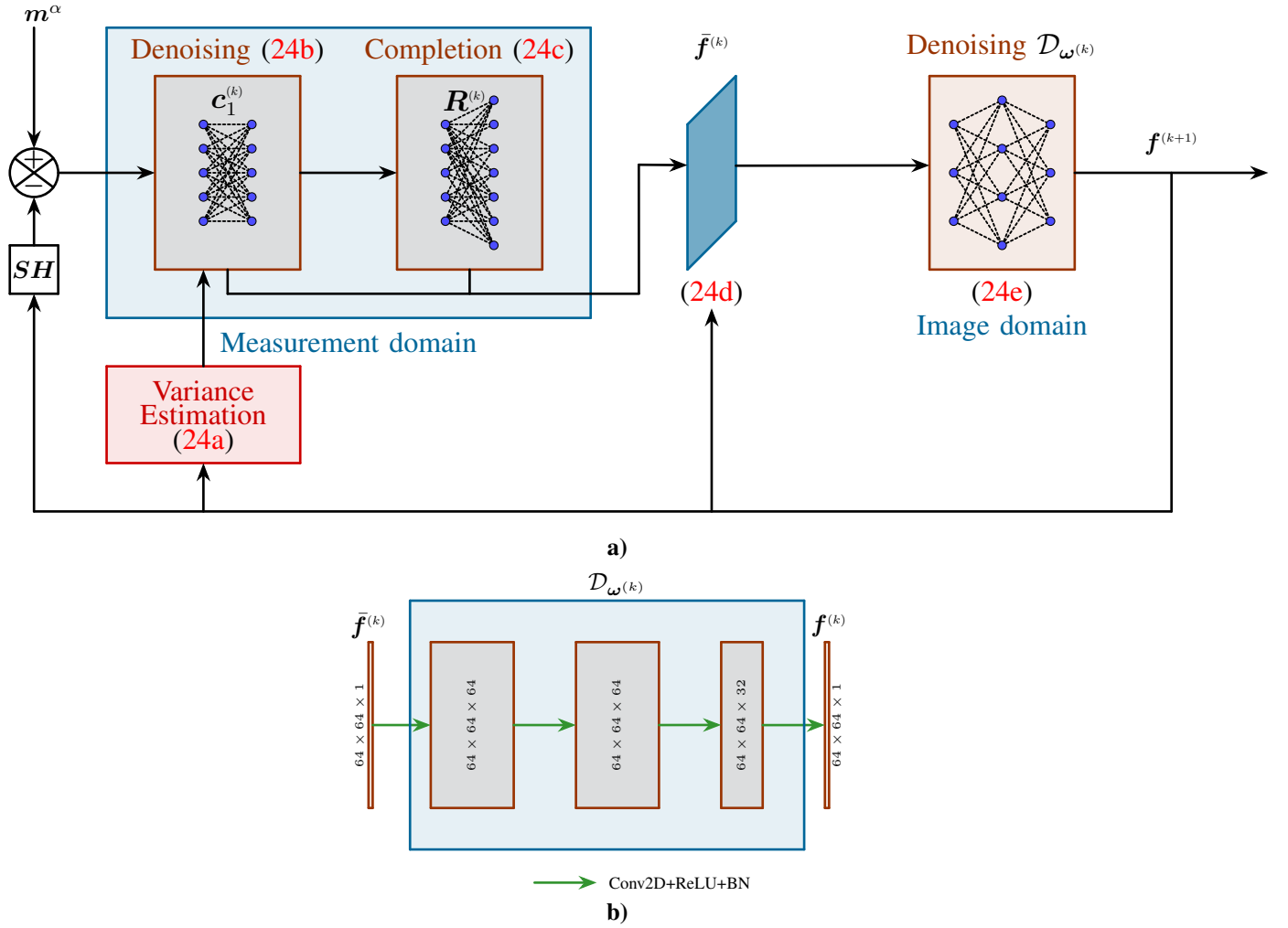


Fig. 1. Proposed EM-Net framework for single-pixel image reconstruction. (a) The recursive algorithm that alternates between completion and denoising in the measurement domain for Equations (24b),(24c), and denoising in the image domain for Equation (24e). (b) Architecture of the neural network in charge of the image domain denoising step.

Method	$M = 512$	$M = 1024$
EM-TV	0	0
CNN	28,993	28,993
U-Net	499,985	499,985
MoDL	499,986	499,986
NN	499,986	499,986
EM-Net-s	31,041	33,089
EM-Net-u	147,013	149,061

TABLE I
NUMBER OF LEARNED PARAMETERS FOR THE DIFFERENT RECONSTRUCTION METHODS AT TWO DIFFERENT COMPRESSION RATES ($M = 512$ (CR = 1/8), $M = 1024$ (CR = 1/4)).

test set. All of the networks were trained using the STL-10 dataset for $\alpha = 10$ photons. Similarly, the four methods are compared for 500 images of the ImageNet test in Table III. All of the networks were trained with the ImageNet dataset for $\alpha = 10$ photons. To explore the robustness of the methods in the presence of noise, the levels of which are unknown in practice, all of the networks were tested at a higher noise level ($\alpha = 3$ photons) and a lower noise level ($\alpha = 50$ photons).

According to our simulations at $M = 1,024$, the best performing methods are EM-Net-u at higher noise ($\alpha = 3$ photons) and MoDL at lower noise ($\alpha = 50$ photons). For $M = 1,024$, EM-TV, CNN, and MoDL perform relatively well at low levels of noise, with PSNR values typically less than 1 dB below the best PSNR (-1.67 dB for CNN, -1.39 dB for EM-TV, -0.56 dB for U-Net, with MoDL as the best at 20.02 dB). At the high level of noise, their reconstruction PSNRs are significantly lower than the best one (-6.12 dB for CNN, -3.46 dB for EM-TV, -1.03 dB for U-Net, -3.71 dB for MoDL). The opposite was seen for NN, which yields good PSNR for higher noise (only -1.12 dB compared to EM-Net-u), but performs poorly for lower noise (-2.85 dB compared to MoDL). The same trend is observed for $M = 512$, except that NN also strongly degrades at higher noise (-2.2 dB compared to EM-Net-u). For high levels of noise, the PSNR might be worse for $M = 1,024$ than for $M = 512$. This is explained on the basis that mostly noise is acquired for the high frequency coefficients. This addition of noisy information can highly degrade the image quality despite the higher number of acquisitions.

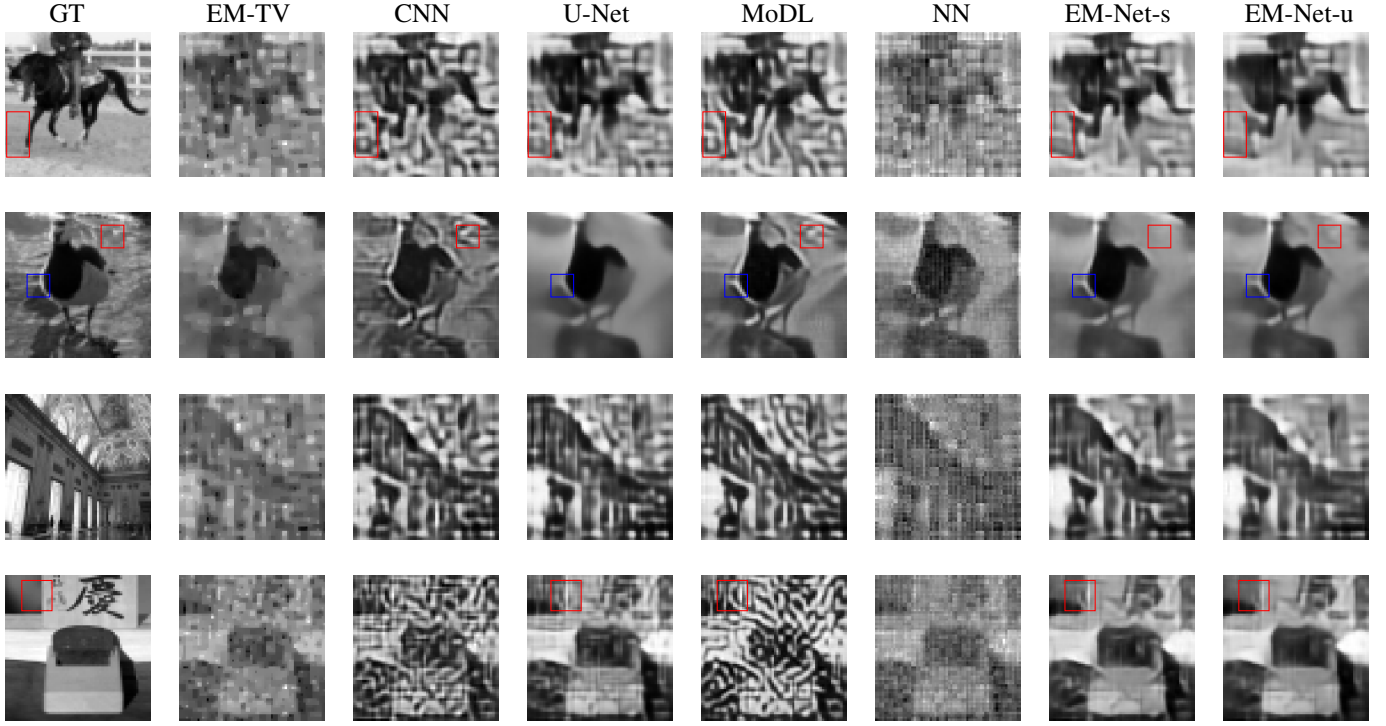


Fig. 2. Reconstructions of the simulated images using the different methods. Top row: STL-10 horse image with $M = 512$ ($CR = 1/8$) and $\alpha = 3$ photons. Second row: STL-10 duck with $M = 1,024$ ($CR = 1/4$) and $\alpha = 30$ photons. Third row: ImageNet hall with $M = 512$ ($CR = 1/8$) and $\alpha = 2$ photons. Fourth row: ImageNet couch with $M = 1,024$ ($CR = 1/4$) and $\alpha = 5$ photons. The images were reconstructed from simulated measurements from the ground-truth (GT) image. The following columns show reconstructions using an expectation-maximization total variation (EM-TV) algorithm, the ConvNet reconstructor presented in Fig. 1b (CNN), a U-Net reconstructor [48] (U-Net), a model-based reconstruction with deep learned priors [14] (MoDL), and the proposed method (EM-Net-s) and (EM-Net-u).

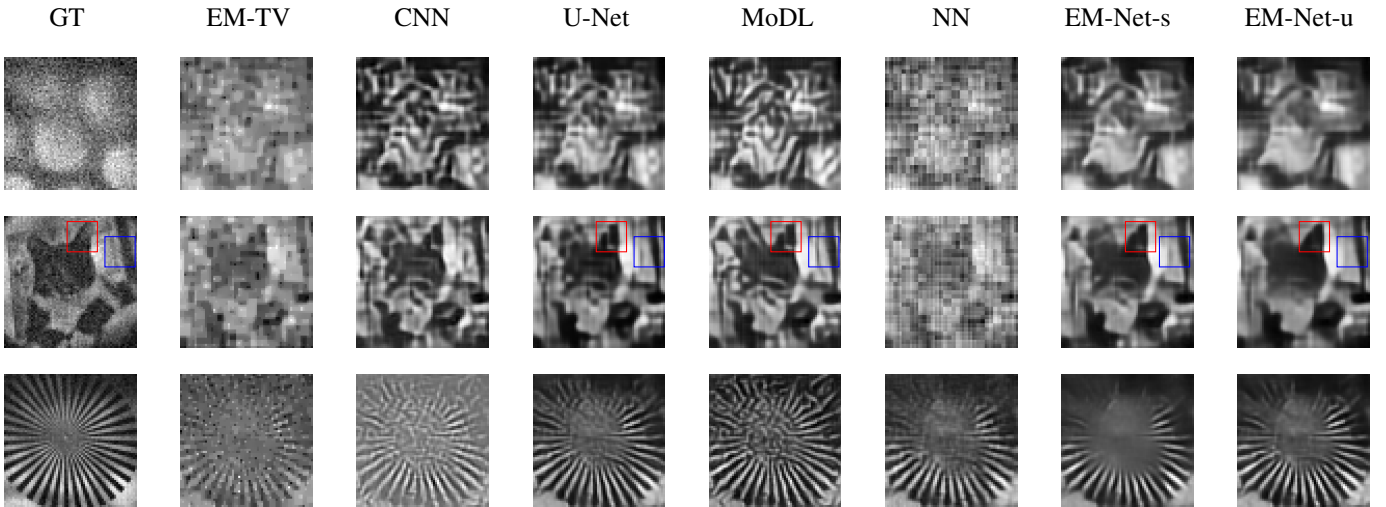


Fig. 3. Reconstructions of the three experimental datasets using the six different methods. Top row: LED lamp with $M = 512$ ($CR = 1/8$) measurements. Second row: STL-10 cat with $M = 512$ ($CR = 1/8$) measurements. Third row: Siemens star with $M = 2,048$ ($CR = 1/2$) measurements. The ground-truth (GT) images were reconstructed from a fully sampled dataset acquired with high image intensity (first column: $\alpha = 148$ photons, $\alpha = 195$ photons, $\alpha = 295$ photons). The following columns show reconstructions from the down-sampled measurements acquired from a fully sampled dataset with lower image intensity ($\alpha = 9$ ph, $\alpha = 10$ photons and $\alpha = 14$ photons) using expectation-maximization total variation (EM-TV) algorithm, the ConvNet reconstructor presented in Fig. 1b (CNN), a U-Net reconstructor [48] (U-Net), a model-based reconstruction with deep learned priors [14] (MoDL), and the proposed method (EM-Net-s) and (EM-Net-u).

The proposed EM-Net-u method is very effective in a wide variety of situations. At high noise (*i.e.*, $\alpha = 3$ ph), EM-Net-u is the best performing method in terms of PSNR and SSIM for both the STL-10 and ImageNet images, whatever the compression ratio. At lower noise (*i.e.*, $\alpha = 50$ ph), EM-Net-u is the best for $M = 512$, and second-best performing method for $M = 1,024$, only slightly behind the best performing method (-0.16 dB compared to MoDL).

EM-Net-s yields reconstructions with slightly lower metrics than EM-Net-u for $M = 512$ (-0.79 dB for $\alpha = 3$ photons and 0.13 dB for $\alpha = 50$ photons) and for $M = 1024$ (-0.57 dB for $\alpha = 3$ photons and -0.6 dB for $\alpha = 50$ photons). Despite being outperformed by EM-Net-u, EM-Net-s always outperforms U-Net, CNN, NN and EM-TV. Finally, MoDL only outperforms EM-Net-s at high photon counts (*i.e.*, $\alpha = 50$ photons), with a modest PSNR gain of 0.03 dB and 0.22 dB for $M = 512$ and $M = 1024$, respectively.

In summary, the metrics indicate that EM-Net-u is the best method when the data is heavily under-sampled, or under high noise levels. Furthermore, EM-Net-s is among the best methods that we assessed here. Both variants of the proposed EM-Net algorithm offer excellent reconstruction under noise levels that differ from that used during training, which is often the case in real-life applications, and also under different compression ratios.

B. Visual assessment of the reconstructed images

1) *Simulated data*: Fig. 2 shows the assessment of our deep EM-Net with data simulated from the STL-10 and ImageNet databases. In particular, we consider two images from the STL-10 dataset, a horse and a duck. We simulate the acquisition of the horse image with $M = 512$ measurements and $\alpha = 3$ ph, and that of the duck with $M = 1,024$ measurements and $\alpha = 30$ photons. We also consider two images from the ImageNet test set, a couch and a hall. We simulate the acquisition of the couch with $M = 1,024$ measurements and $\alpha = 5$ ph, and that of the hall with $M = 512$ measurements and $\alpha = 2$ photons.

In all of the simulations, the EM-TV reconstruction produces images that are piece-wise linear and lack some fine details. NN suffers from grid-like or staircase artefacts. While it preserves more image details at the lower compression ratio (*i.e.*, $M = 1,024$), the NN reconstruction presents strong artifacts at the high compression ratio (*i.e.*, $M = 512$). CNN, U-Net, MoDL, EM-Net-s and EM-Net-u perform visually better than EM-TV and NN. However, these reconstructions have different behaviors, as indicated in Fig. 2 using red and blue rectangles. The red rectangles highlight areas where the ground-truth images are mostly flat, while the blue rectangle highlights an area where a detail is present. In the red areas, the CNN, U-Net, and MoDL reconstructions tend to include worm-like artifacts that are not seen for either EM-Net (*e.g.*, compare the image homogeneity within the red rectangles of the horse image in Fig. 2). On the other hand, despite low noise conditions, U-Net blurs fine details, such as the white detail in the blue rectangle, while MoDL and both EM-Net variants mostly keep these fine details intact.

While both variants of the proposed method produce similar results, EM-Net-u appears better at removing artifacts (such

as those in the red rectangle in the horse and couch images) while retaining more details (such as the white spot in the red rectangle in the duck image).

In summary, U-Net, MoDL, and the proposed EM-Net variants are visually best performing. Among these methods, the EM-Net variants do not show most of the grid-like and worm-like artifacts at low photon counts, while they preserve the fine details at higher photon counts, which can be blurred with U-Net and MoDL. This visual assessment confirms the excellent robustness of EM-Net to a wide variety of simulation parameters.

2) *Experimental Data*: Fig. 3 shows the assessment of the different methods using experimental data from the SPIHIM experimental dataset. In particular, this includes the images of a LED lamp, a cat from the STL-10 dataset printed on a transparent sheet, and the 32 branch Siemens star [54]. We display the images that we reconstruct with fully sampled measurements (*i.e.*, $M = 4,096$) in the first and second columns of Fig. 3. The first column corresponds to very low noise acquisitions, and the second to the higher noise acquisitions that we assessed for these methods. We estimated the image intensities for each of the acquisitions, and obtained $\tilde{\alpha} = 9$ photons for the LED, $\tilde{\alpha} = 10$ photons for the cat, and $\tilde{\alpha} = 14$ photons for the Siemens star. Then, we down-sampled the acquisitions *a posteriori*, keeping $M = 512$ measurements for the LED and for the cat, and $M = 1,024$ for the Siemens star.

For the LED lamp, as for the simulated measurements, there are grid-like artefacts in the EM-TV and NN reconstructions, which are most visible at the borders of the LED disks and in the background. On the other hand, the CNN, MoDL, U-Net, and EM-Net reconstructors are more smooth. Here again, CNN, U-Net, and MoDL present worm-like artefacts, while EM-Net-u and EM-Net-s leads to a more homogeneous background and LED disks.

For the STL-10 cat, as for the LED lamp, high-frequency grid artefacts are seen for the EM-TV and NN reconstructions, while the other methods are visually more similar. However, EM-Net removes some of the reconstruction artifacts seen in CNN, U-Net, and EM-TV (*e.g.*, compare the red rectangles in the second row of Fig. 3; the ear of the cat is mostly artifact free using EM-Net). We further note that EM-Net preserves image details well (*e.g.*, compare the blue rectangles in the second row of Fig. 3; EM-Net preserves the edge of the door while keeping the image intensity closer to the ground-truth grayscale value).

For the Siemens star, U-Net and EM-Net-u best reconstruct the branches of the target. This is particularly visible for the upper left branches that are completely removed by the other methods. Both methods also create some artifacts in the center of the target, but these are less present than with other reconstruction methods such as MoDL or CNN. U-Net and EM-Net-u are also less prone to oversmoothing than EM-Net-s, NN and EM-TV. This object has a high-frequency content, and structures that are not like any of the images in the training set, which make it a challenging sample.

		$M = 512$ (CR = 1/8)		$M = 1024$ (CR = 1/4)	
		$\alpha = 3$	$\alpha = 50$	$\alpha = 3$	$\alpha = 50$
EM-TV	PSNR	13.91±1.39	18.29±3.46	12.33±0.95	18.63±3.61
	SSIM	0.68±0.14	0.90±0.05	0.59±0.16	0.92±0.04
CNN	PSNR	13.2±1.05	19.19±3.75	9.67±0.82	18.35±3.23
	SSIM	0.66±0.16	0.94±0.05	0.43±0.19	0.92±0.05
U-Net	PSNR	16.17±2.04	19.29±3.75	14.76±1.43	19.46±3.83
	SSIM	0.79±0.13	0.92±0.07	0.72±0.16	0.93±0.07
MoDL	PSNR	14.8±1.55	19.47±3.87	12.08±0.78	20.02±4.19
	SSIM	0.73±0.15	0.93±0.07	0.58±0.18	0.95±0.05
NN	PSNR	14.87±1.58	18.3±3.26	15.27±2.1	17.07±3.21
	SSIM	0.6±0.27	0.71±0.33	0.39±0.43	0.48±0.47
EM-Net-s	PSNR	16.28±2.06	19.45±3.89	15.82±1.8	19.8±4.08
	SSIM	0.8±0.13	0.93±0.06	0.78±0.15	0.94±0.05
EM-Net-u	PSNR	17.07±2.49	19.58±3.94	16.39±2.1	19.86±4.07
	SSIM	0.83±0.11	0.94±0.06	0.8±0.15	0.95±0.05

TABLE II

PEAK SIGNAL-TO-NOISE RATIOS (PSNR) AND STRUCTURAL SIMILARITIES (SSIM) FOR THE DIFFERENT RECONSTRUCTION METHODS. THE METRICS ARE COMPUTED FROM 200 SIMULATED IMAGES FROM THE STL-10 DATASET. ALL OF THE NEURAL NETWORKS WERE TRAINED USING THE STL-10 DATASET WITH $\alpha = 10$ PHOTONS, AND WERE TESTED FOR $\alpha = 3$ AND $\alpha = 50$ PHOTONS FOR DIFFERENT COMPRESSION RATES ($M = 512$ (CR = 1/8), $M = 1,024$ (CR = 1/4)). RESULTS ARE SHOWN FOR AN EXPECTATION-MAXIMIZATION TOTAL VARIATION (EM-TV) ALGORITHM, DIRECT U-NET RECONSTRUCTOR (U-NET), MODEL-BASED RECONSTRUCTION WITH DEEP LEARNED PRIORS (MoDL), NEUMANN NETWORKS (NN), AND THE TWO VARIANTS OF THE PROPOSED METHOD (EM-NET-U AND EM-NET-S). **BLUE** INDICATES THE HIGHEST PSNRs; **GREEN** INDICATES THE HIGHEST SSIMs.

		$M = 512$ (CR = 1/8)		$M = 1024$ (CR = 1/4)	
		$\alpha = 3$	$\alpha = 50$	$\alpha = 3$	$\alpha = 50$
EM-TV	PSNR	13.4±1.75	17.17±3.78	11.99±1.29	17.51±3.93
	SSIM	0.66±0.14	0.88±0.06	0.58±0.14	0.90±0.06
CNN	PSNR	12.94±1.35	17.89±4.12	10.49±0.82	17.3±3.66
	SSIM	0.65±0.17	0.92±0.06	0.49±0.17	0.91±0.06
U-Net	PSNR	15.27±2.35	17.97±4.09	14.16±1.79	18.2±4.21
	SSIM	0.77±0.14	0.91±0.08	0.71±0.15	0.91±0.08
MoDL	PSNR	14.03±1.83	18.22±4.28	10.88±0.82	18.65±4.53
	SSIM	0.7±0.17	0.91±0.08	0.51±0.17	0.94±0.06
NN	PSNR	14.82±2.35	16.2±3.18	14.34±2.01	17.42±3.79
	SSIM	0.51±0.35	0.6±0.34	0.6±0.26	0.72±0.32
EM-Net-s	PSNR	15.53±2.47	18.08±4.22	15.27±2.28	18.42±4.41
	SSIM	0.78±0.14	0.91±0.08	0.77±0.15	0.93±0.07
EM-Net-u	PSNR	16.12±2.86	18.22±4.29	15.58±2.49	18.45±4.35
	SSIM	0.81±0.13	0.92±0.07	0.79±0.14	0.93±0.07

TABLE III

PEAK SIGNAL-TO-NOISE RATIOS (PSNR) AND STRUCTURAL SIMILARITIES (SSIM) FOR THE DIFFERENT RECONSTRUCTION METHODS. THE METRICS WERE COMPUTED FROM 200 SIMULATED IMAGES FROM THE IMAGENET DATASET. ALL OF THE NEURAL NETWORKS WERE TRAINED USING THE IMAGENET DATASET WITH $\alpha = 10$ PHOTONS, AND WERE TESTED FOR $\alpha = 3$ AND $\alpha = 50$ PHOTONS FOR DIFFERENT COMPRESSION RATES ($M = 512$ (CR = 1/8), $M = 1,024$ (CR = 1/4)). RESULTS ARE SHOWN FOR AN EXPECTATION-MAXIMIZATION TOTAL VARIATION (EM-TV) ALGORITHM, DIRECT U-NET RECONSTRUCTOR (U-NET), MODEL-BASED RECONSTRUCTION WITH DEEP LEARNED PRIORS (MoDL), NEUMANN NETWORKS (NN), AND THE TWO VARIANTS OF THE PROPOSED METHOD (EM-NET-U AND EM-NET-S). **BLUE** INDICATES THE HIGHEST PSNRs; **GREEN** INDICATES THE HIGHEST SSIMs.

C. Comparison of the number of learned parameters

Table I shows that EM-Net-s has the second lowest number of learned parameters. EM-Net-s is very close to CNN in terms of the number of parameters, and has about 16 times fewer parameters compared to U-Net, MoDL, and NN. Table I also shows that EM-Net-u has about 3 times fewer learned parameters than U-Net, NN and MoDL. As shown in the previous section, the metrics and the visuals indicate that the two EM-Net variants are among the best of the methods assessed. EM-Net-u is usually the best performing method, while it has 3 times fewer parameters than its main competitors,

e.g., U-Net and MoDL. EM-Net-s performs slightly poorer than EM-Net-u; however, it has a comparable performance to reconstructors that have 16 times more learned parameters. EM-Net-s clearly outperforms CNN that has roughly the same number of parameters.

VI. CONCLUSIONS AND PERSPECTIVES

We propose an iterative network based on the EM algorithm. This deep EM network can solve linear under-determined inverse problems where the measurements are corrupted by normally distributed signal-dependent noise (e.g.,

Poisson noise, mixed Poisson-Gaussian noise, mixed Skellam-Gaussian noise). The EM-Net alternates over five steps, which have straightforward interpretation, as estimation of the noise covariance, measurement denoising, measurement completion, measurement to image domain mapping, and image denoising.

Most unrolled networks assume additive Gaussian noise with constant variance. Therefore, they tend to generalize poorly to different noise levels, and might be ill-suited to applications where the noise levels cannot be predicted beforehand. On the contrary, the proposed EM-Net method explicitly estimates the noise covariance to denoise the measurement domain. As a consequence, EM-Net generalizes very well to noise levels that differ to that used during training. In particular, it outperforms all of the other methods assessed here when in the presence of noise with variance higher than that used during training, and it is comparable to the best method in the presence of noise with lower variance.

VII. ACKNOWLEDGMENTS

This work was supported by the French National Research Agency (ANR), under Grant ANR-17-CE19-0003 (ARMONI Project), and performed within the framework of the LABEX PRIMES (ANR-11-LABX-0063) of Université de Lyon. This material is based upon work carried out at the PILoT facility (PILoT, INSA LYON, Bât. L. de Vinci, 21 Av. Jean Capelle 69621 Villeurbanne). This work was granted access to the HPC resources of IDRIS under the allocation 2019-101203 made by GENCI (Jean Zay computing center).

APPENDIX A

LIKELIHOOD OF SKELLAM-GAUSSIAN MEASUREMENTS

We first rewrite the i -th component of \mathbf{m}^α in Equation (3), denoted m for simplicity, as

$$m = \mathbf{h}^\top \mathbf{f} + \epsilon_f + \epsilon, \quad (33)$$

where

$$\epsilon_f \sim \frac{1}{\alpha} \mathcal{S}_k(\alpha \mathbf{h}_+^\top \mathbf{f}, \alpha \mathbf{h}_-^\top \mathbf{f}) - \mathbf{h}^\top \mathbf{f}, \quad (34)$$

$$\epsilon \sim \mathcal{N}(0, \beta^2), \quad \beta^2 = \frac{2\sigma_{\text{dark}}^2}{K^2\alpha^2}. \quad (35)$$

where $\mathbf{h} = \mathbf{h}_+ - \mathbf{h}_-$ represents a row of the Hadamard matrix.

To derive the probability mass function of ϵ_f , we recall that of a Skellam-distributed random variable $Z \sim \mathcal{S}_k(\mu_1, \mu_2)$ [41]

$$\mathbb{P}(Z = k; \mu_1, \mu_2) = e^{-(\mu_1 + \mu_2)} \left(\frac{\mu_1}{\mu_2} \right)^{\frac{k}{2}} \mathcal{I}_{|k|}(2\sqrt{\mu_1\mu_2}), \quad (36)$$

where \mathcal{I}_\cdot are the modified Bessel functions of the first kind. By linear transformation of Equation (36) according to Equation (34), we obtain $p_n = \mathbb{P}(\epsilon_f = \frac{n}{\alpha} - \mathbf{h}^\top \mathbf{f}), \forall n \in \mathbb{N}$,

$$p_n = e^{-\alpha(\mathbf{h}_+^\top \mathbf{f} + \mathbf{h}_-^\top \mathbf{f})} \left(\frac{\mathbf{h}_+^\top \mathbf{f}}{\mathbf{h}_-^\top \mathbf{f}} \right)^{\frac{n - \alpha \mathbf{h}^\top \mathbf{f}}{2\alpha}} \mathcal{I}_{|\frac{n}{\alpha} - \mathbf{h}^\top \mathbf{f}|} \left(2\alpha \sqrt{(\mathbf{h}_+^\top \mathbf{f})(\mathbf{h}_-^\top \mathbf{f})} \right). \quad (37)$$

Therefore, the probability density function of ϵ_f is

$$p(\epsilon_f) = \sum_{n \in \mathbb{N}} p_n \delta(\epsilon_f - \frac{n}{\alpha} + \mathbf{h}^\top \mathbf{f}), \quad (38)$$

where δ denotes the Dirac delta function. As the probability density function of $E = \epsilon + \epsilon_f$ is the convolution of Equation (38) with the probability density function of ϵ [55], we have

$$p_E(E) = \sum_{n \in \mathbb{N}} p_n \frac{1}{\sqrt{2\pi\beta^2}} \exp\left(-\frac{(E - \frac{n}{\alpha} - \mathbf{h}^\top \mathbf{f})^2}{\beta^2}\right). \quad (39)$$

Next, we write the likelihood of the measurement as

$$p(m|\mathbf{f}) = p_E(m - \mathbf{h}^\top \mathbf{f}) \quad (40)$$

$$= \sum_{n \in \mathbb{N}} p_n \frac{1}{\sqrt{2\pi\beta^2}} \exp\left(-\frac{(m - \frac{n}{\alpha})^2}{\beta^2}\right). \quad (41)$$

Assuming independent measurements, the chain rule leads to

$$p(\mathbf{m}|\mathbf{f}) = \prod_{i=1}^M p(\mathbf{m}_i|\mathbf{f}). \quad (42)$$

Substituting Equation (41) and the expression of p_n in Equation (37) into Equation (42), we finally obtain (4).

APPENDIX B

QUADRATIC SOLUTION OF UNDER-SAMPLED ORTHOGONAL ACQUISITION

The analytical solution to the quadratic minimization problem of Equation (15a) is given by (see p. 66 of [45])

$$\bar{\mathbf{x}}^{(k)} = \mathbf{H} \mathbf{f}^{(k)} + \mathbf{C} \mathbf{S}^\top (\mathbf{S} \mathbf{C} \mathbf{S}^\top + \tilde{\Sigma})^{-1} (\mathbf{m}^\alpha - \mathbf{S} \mathbf{H} \mathbf{f}^{(k)}). \quad (43)$$

We first introduce $\mathbf{y}^{(k)} = \bar{\mathbf{x}}^{(k)} - \mathbf{H} \mathbf{f}^{(k)}$, which leads to

$$\mathbf{y}^{(k)} = \mathbf{C} \mathbf{S}^\top (\mathbf{S} \mathbf{C} \mathbf{S}^\top + \tilde{\Sigma})^{-1} (\mathbf{m}^\alpha - \mathbf{S} \mathbf{H} \mathbf{f}^{(k)}). \quad (44)$$

Then, recalling that $\mathbf{S} = [\mathbf{I}_M, \mathbf{0}]$, where \mathbf{I}_M is the identity matrix, we have the following two identities

$$\mathbf{C} \mathbf{S}^\top = \begin{bmatrix} \mathbf{C}_1 \\ \mathbf{C}_2 \end{bmatrix}, \quad \text{and} \quad \mathbf{S} \mathbf{C} \mathbf{S}^\top = \mathbf{C}_1, \quad (45)$$

where \mathbf{C}_1 and \mathbf{C}_2 are the blocks of \mathbf{C} defined according to Equation (17). Finally, substituting Equation (45) into Equation (44), we obtain

$$\mathbf{y}^{(k)} = \begin{bmatrix} \mathbf{C}_1 \\ \mathbf{C}_2 \end{bmatrix} (\tilde{\Sigma} + \mathbf{C}_1)^{-1} (\mathbf{m}^\alpha - \mathbf{S} \mathbf{H} \mathbf{f}^{(k)}). \quad (46)$$

REFERENCES

- [1] M. F. Duarte et al., "Single-pixel imaging via compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 83–91, 2008.
- [2] V. Studer, J. Bobin, M. Chahid, H. S. Mousavi, E. Candes, and M. Dahan, "Compressive fluorescence microscopy for biological and hyperspectral imaging," *Proceedings of the National Academy of Sciences*, vol. 109, no. 26, pp. E1679–E1687, 2012.
- [3] F. Rousset, N. Ducros, F. Peyrin, G. Valentini, C. D'Andrea, and A. Farina, "Time-resolved multispectral imaging based on an adaptive single-pixel camera," *Opt. Express*, vol. 26, no. 8, pp. 10 550–10 558, Apr 2018.

- [4] G. R. Arce, D. J. Brady, L. Carin, H. Arguello, and D. S. Kittle, "Compressive coded aperture spectral imaging: An introduction," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 105–115, Jan 2014.
- [5] Q. Pian, R. Yao, N. Sinsuebphon, and X. Intes, "Compressive hyperspectral time-resolved wide-field fluorescence lifetime imaging," *Nature Photonics*, vol. 11, pp. 411–414, 2017.
- [6] E. Aguénonon, F. Dadouche, W. Uhring, N. Ducros, and S. Gioux, "Single snapshot imaging of optical properties using a single-pixel camera: a simulation study," *Journal of Biomedical Optics*, vol. 24, no. 7, pp. 1–6, 2019.
- [7] Y. Zhang, G. M. Gibson, M. P. Edgar, G. Hammond, and M. J. Padgett, "Dual-band single-pixel telescope," *Opt. Express*, vol. 28, no. 12, pp. 18 180–18 188, Jun 2020.
- [8] M. P. Edgar, G. M. Gibson, and M. J. Padgett, "Principles and prospects for single-pixel imaging," *Nature Photonics*, vol. 13, no. 1, pp. 13–20, Jan. 2019.
- [9] G. Wang, J. C. Ye, K. Mueller, and J. A. Fessler, "Image reconstruction is a new frontier of machine learning," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1289–1296, 2018.
- [10] S. Arridge, P. Maass, O. Öktem, and C.-B. Schönlieb, "Solving inverse problems using data-driven models," *Acta Numerica*, vol. 28, p. 1–174, 2019.
- [11] M. T. McCann, K. H. Jin, and M. Unser, "Convolutional neural networks for inverse problems in imaging: A review," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 85–95, 2017.
- [12] E. Kang, H. J. Koo, D. H. Yang, J. B. Seo, and J. C. Ye, "Cycle-consistent adversarial denoising network for multiphase coronary ct angiography," *Medical Physics*, vol. 46, p. 550–562, 2019.
- [13] I. Y. Chun and J. A. Fessler, "Convolutional analysis operator learning: Acceleration and convergence," *IEEE Transactions on Image Processing*, vol. 29, pp. 2108–2122, 2020.
- [14] H. K. Aggarwal, M. P. Mani, and M. Jacob, "Modl: Model-based deep learning architecture for inverse problems," *IEEE Transactions on Medical Imaging*, vol. 38, no. 2, pp. 394–405, 2019.
- [15] H. Gupta, K. H. Jin, H. Q. Nguyen, M. T. McCann, and M. Unser, "Cnn-based projected gradient descent for consistent ct image reconstruction," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1440–1453, June 2018.
- [16] J. Adler and O. Öktem, "Learned primal-dual reconstruction," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1322–1332, 2018.
- [17] A. J. Reader, G. Corda, A. Mehranian, C. d. Costa-Luis, S. Ellis, and J. A. Schnabel, "Deep learning for pet image reconstruction," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 5, no. 1, pp. 1–25, 2021.
- [18] H. Lim, I. Y. Chun, Y. K. Dewaraja, and J. A. Fessler, "Improved low-count quantitative pet reconstruction with an iterative neural network," *IEEE Transactions on Medical Imaging*, vol. 39, no. 11, pp. 3512–3522, 2020.
- [19] D. Gilton, G. Ongie, and R. Willett, "Neumann networks for linear inverse problems in imaging," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 328–343, 2020.
- [20] G. Barbastathis, A. Ozcan, and G. Situ, "On the use of deep learning for computational imaging," *Optica*, vol. 6, no. 8, pp. 921–943, Aug 2019.
- [21] M. R. Kellman, E. Bostan, N. A. Repina, and L. Waller, "Physics-based learned design: Optimized coded-illumination for quantitative phase imaging," *IEEE Transactions on Computational Imaging*, vol. 5, no. 3, pp. 344–353, 2019.
- [22] C. Higham, R. Murray-Smith, M. Padgett, and M. Edgar, "Deep learning for real-time single-pixel video," *Scientific Reports*, no. 8, Feb 2018.
- [23] A. Dave, A. K. Vadathya, R. Subramanyam, R. Baburajan, and K. Mitra, "Solving inverse computational imaging problems using deep pixel-level prior," *IEEE Transactions on Computational Imaging*, vol. 5, pp. 37–51, 2019.
- [24] N. Ducros, A. Lorente Mur, and F. Peyrin, "A completion network for reconstruction from compressed acquisition," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, 2020, pp. 619–623.
- [25] A. Lorente Mur, P. Leclerc, F. Peyrin, and N. Ducros, "Single-pixel image reconstruction from experimental data using neural networks," *Opt. Express*, vol. 29, no. 11, pp. 17 097–17 110, May 2021.
- [26] A. Lorente Mur, P. Bataille, F. Peyrin, and N. Ducros, "Deep expectation-maximization for image reconstruction from under-sampled poisson data," in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, 2021, pp. 1535–1539.
- [27] F. J. Anscombe, "The transformation of Poisson, binomial and negative-binomial data," *Biometrika*, vol. 35, no. 3–4, pp. 246–254, 12 1948.
- [28] F. Murtagh, J. luc Starck, and A. Bijaoui, "Image restoration with noise suppression using a multiresolution support," *Astronomy and Astrophysics, Suppl. Ser.*, vol. 112, pp. 179–189, 1995.
- [29] M. Makitalo and A. Foi, "Optimal inversion of the anscombe transformation in low-count poisson image denoising," *IEEE Transactions on Image Processing*, vol. 20, no. 1, pp. 99–109, 2011.
- [30] F. Benvenuto, A. L. Camera, C. Theys, A. Ferrari, H. Lanteri, and M. Bertero, "The study of an iterative method for the reconstruction of images corrupted by poisson and gaussian noise," *Inverse Problems*, vol. 24, no. 3, p. 035016, apr 2008.
- [31] E. Chouzenoux, A. Jeziarska, J.-C. Pesquet, and H. Talbot, "A convex approach for image restoration with exact poisson-gaussian likelihood," *SIAM Journal on Imaging Sciences*, vol. 8, no. 4, pp. 2662–2682, 2015.
- [32] M. Ghulyani and M. Arigovindan, "Fast roughness minimizing image restoration under mixed poisson-gaussian noise," *IEEE Transactions on Image Processing*, vol. 30, pp. 134–149, 2021.
- [33] J. Li, F. Luisier, and T. Blu, "Pure-let image deconvolution," *IEEE Transactions on Image Processing*, vol. 27, no. 1, pp. 92–105, 2018.
- [34] J. Li, Z. Shen, R. Yin, and X. Zhang, "A reweighted ℓ_2 method for image restoration with poisson and mixed poisson-gaussian noise," *Inverse Problems and Imaging*, vol. 9, no. 3, pp. 875–894, Jul. 2015.
- [35] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 1, pp. 1–22, Sep. 1977.
- [36] A. Lorente Mur, F. Peyrin, and N. Ducros, "Single-Pixel pYthon Image Reconstruction Toolbox (spyrit) Version 1.0," https://github.com/openspyrit/spyrit/tree/ieee_tci_branch, 2020.
- [37] M. Ochoa, Q. Pian, R. Yao, N. Ducros, and X. Intes, "Assessing patterns for compressive fluorescence lifetime imaging," *Opt. Lett.*, vol. 43, no. 18, pp. 4370–4373, Sep 2018.
- [38] A. Lorente Mur, M. Ochoa, J. E. Cohen, X. Intes, and N. Ducros, "Handling negative patterns for fast single-pixel lifetime imaging," in *SPIE Photonics : Molecular-Guided Surgery: Molecules, Devices, and Applications V*, vol. 10862, 2019.
- [39] A. Foi, M. Trimeche, V. Katkovnik, and K. Egiazarian, "Practical poissonian-gaussian noise modeling and fitting for single-image raw-data," *IEEE Transactions on Image Processing*, vol. 17, no. 10, pp. 1737–1754, Oct. 2008.
- [40] L. Baldassarre, Y. Li, J. Scarlett, B. Gözcü, I. Bogunovic, and V. Cevher, "Learning-based compressive subsampling," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 4, pp. 809–822, June 2016.
- [41] J. G. Skellam, "The frequency distribution of the difference between two poisson variates belonging to different populations," *Journal of the Royal Statistical Society*, vol. 109, no. 3, p. 296, 1946.
- [42] K. Huang, *Introduction to statistical physics*. London New York: Taylor & Francis, 2001.
- [43] J. Zhou, J.-L. Coatrieux, A. Bousse, H. Shu, and L. Luo, "A bayesian map-em algorithm for pet image reconstruction using wavelet transform," *IEEE Transactions on Nuclear Science*, vol. 54, no. 5, pp. 1660–1669, 2007.
- [44] J. Fessler and A. Hero, "Space-alternating generalized expectation-maximization algorithm," *IEEE Transactions on Signal Processing*, vol. 42, no. 10, pp. 2664–2677, 1994.
- [45] A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation*. Society for Industrial and Applied Mathematics, 2005.
- [46] T. Goldstein and S. Osher, "The split bregman method for l_1 -regularized problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 323–343, Jan. 2009.
- [47] M. Ravasi and I. Vasconcelos, "Pylops—a linear-operator python library for scalable algebra and optimization," *SoftwareX*, vol. 11, p. 100361, 2020.
- [48] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 234–241.
- [49] A. Coates, A. Ng, and H. Lee, "An analysis of single-layer networks in unsupervised feature learning," in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, G. Gordon, D. Dunson, and M. Dudík, Eds., vol. 15. Fort Lauderdale, FL, USA: PMLR, 11–13 Apr 2011, pp. 215–223.
- [50] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.

- [51] P. Chrabaszcz, I. Loshchilov, and F. Hutter, “A downsampled variant of imagenet as an alternative to the CIFAR datasets,” CoRR, vol. abs/1707.08819, 2017.
- [52] N. Ducros and A. Lorente Mur, “Single-Pixel Hyperspectral Imaging dataset Version 1.0,” <https://github.com/openspyrit/spihim>, 2020.
- [53] A. Paszke et al., “Pytorch: An imperative style, high-performance deep learning library,” in Advances in Neural Information Processing Systems 32. Curran Associates, Inc., 2019, pp. 8024–8035.
- [54] Photography — Electronic still picture imaging — Resolution and spatial frequency responses, International Organization for Standardization, Geneva, CH, 2019, <https://www.iso.org/standard/71696.html>.
- [55] V. V. Petrov, Sums of Independent Random Variables. Springer Berlin Heidelberg, 1975.