Approximate Bayesian denoising for deep image reconstruction in the presence of signal-dependent noise

Luis Amador, Laurent Mahieu-Williame, Elie Bretin, Nicolas Ducros

APPROXIMATE BAYESIAN DENOISING FOR DEEP IMAGE RECONSTRUCTION IN THE PRESENCE OF SIGNAL-DEPENDENT NOISE

Luis Amador[1,2], Laurent Mahieu-Williame[1], Elie Bretin[2], Nicolas Ducros[1,3]

[1]Univ Lyon, INSA-Lyon, UCB Lyon 1,CNRS, Inserm, CREATIS UMR 5220, U1206, Lyon, France
[2]Univ Lyon, INSA de Lyon, CNRS UMR 5208, Institut Camille Jordan, F-69621 Villeurbanne, France
[3]Institut universitaire de France (IUF)

ABSTRACT

Recently, a variety of unrolled networks have been proposed for image reconstruction. These can be interpreted as parameter-optimized algorithms that incorporate steps that are traditionally encountered during the optimization of handcrafted objectives . Here, we address the problem of training such networks in the presence of signal-dependent noise, which is more realistic that the common additive Gaussian noise . We focus on algorithms that requires the inversion of large signal-dependent matrices during training, which increases considerably the training time compared signal-independent inversions that can be precomputed before training. In particular, we describe how to approximate the denoising step of the deep expectation-maximization network to reduce the computational cost and memory requirements while limiting the reconstruction error. We present reconstruction results from simulated and experimental data at different noise levels. Our network yields higher reconstruction peak signal-to-noise ratios than other similar approaches and greater robustness in the practical case where the noise level is unknown or is badly estimated.

Index Terms— Image reconstruction, deep learning, unfolded network, signal-dependent noise, Tikhonov regularization.

1. INTRODUCTION

Many biomedical modalities are ill-posed image reconstruction problems that require prior information to stabilize the solution, such as computerized tomography, magnetic resonance, and optical microscopy. In particular, single-pixel imaging is aimed at recovering an image from a few point measurements that correspond to the dot products between an image and a set of functions that are implemented through a spatial light modulator [1]. This is a typical under-determined inverse problem subject to Poisson noise. Single-pixel imaging has been successfully applied to fluorescence microscopy, hyperspectral imaging, image-guided surgery, fluorescence lifetime imaging, and other such techniques.

Image reconstruction problems have long benefited from the theory of compressed sensing, but recent advances in deep learning have revolutionized the field (see [2], and references therein). In particular, unrolled networks that rely on the computation of traditional solutions (e.g., pseudo inverse, ℓ_1-penalized, maximum a posteriori) can be interpreted as iterative schemes that are optimized with respect to a particular task and database [3]. Unrolled networks based on convolutional neural networks can solve efficiently a variety of problems, which includes image reconstruction problems in magnetic resonance imaging [4], computed tomography [5], and optical microscopy [6]. While most unrolled networks relies on gradient steps with no matrix inversion, we have investigated the exploitation of generalized Tikhonov solutions based on covariance matrices. For instance, inspired by the expectation-maximization (EM) algorithm [7], the deep EM network was proposed to solve inverse problems corrupted by signal-dependent noise [8], such as Poisson corrupted measurements encountered in computational optics. However, signal-dependent noise enforces the resolution of a different linear system for each image in the training database, which prevents pre-computation and poses a severe computational issue during the training phase. In practice, approximate (e.g., diagonal) resolutions are implemented, which leads to sub-optimal image quality. Our contribution is to introduce an approximate solution , whose computational complexity is compatible with the training phase. In Section 2, we describe the forward model of single-pixel imaging and the solution of the inverse problem using the deep EM algorithm. In Section 3, we detail the proposed Taylor approximation of the Bayesian solution. In Section 4, we describe how the network is implemented and trained. In Section 5, we report on and analyze our reconstruction results.

2. COMPUTATIONAL OPTICS

2.1. Forward problem

The philosophy of computational optics is to recover an image $\boldsymbol{f} \in [0,1]^N$ from hardware measurements using software $\boldsymbol{m} = \boldsymbol{H}_1 \boldsymbol{f} \in \mathbb{R}^M$ where $\boldsymbol{H}_1 \in \mathbb{R}^{M \times N}$ is a linear opera-

tor. The system matrix $\boldsymbol{H}_1$ collects the patterns that are sequentially uploaded onto a spatial light modulator. Here, we consider the case $M < N$ where we acquire fewer measurements than pixels in the image to speed up acquisitions. We choose the patterns in a Hadamard basis $\boldsymbol{H} \in \mathbb{R}^{N\times N}$ [9]; i.e., $\boldsymbol{H}_1 = \boldsymbol{S}\boldsymbol{H}$ with $\boldsymbol{S} = [\boldsymbol{I}_M, \boldsymbol{0}]$. Experimental implementation of the negative values of $\boldsymbol{H}_1$ is achieved through the use of the positive patterns $\boldsymbol{H}_1^+ \in \mathbb{R}_+^{M\times N}$ and $\boldsymbol{H}_1^- \in \mathbb{R}_+^{M\times N}$ such that $\boldsymbol{H}_1 = \boldsymbol{H}_1^+ - \boldsymbol{H}_1^-$ (see [10] for details). The acquisition is corrupted by Poisson noise $\hat{\mathbf{m}}^{\alpha,\pm} \sim \mathcal{P}(\alpha \boldsymbol{H}_1^{\pm}\boldsymbol{f})$, where α is the intensity (in photons) that is inversely proportional to the noise level. For images of size $N = 64$, even in the case of small image intensities (e.g., $\alpha = 10$ photons), the Hadamard coefficients are sufficiently large ($\hat{m}^{\alpha,\pm} \approx 1{,}000$ photons) to justify the use of the Gaussian approximation intoduced in the next section. We finally preprocess the raw measurements as $\mathbf{m}^\alpha = \frac{1}{\alpha}(\hat{\mathbf{m}}^{\alpha,+} - \hat{\mathbf{m}}^{\alpha,-})$, to get measurements such that $\mathbb{E}\left(\mathbf{m}^\alpha\right) = \boldsymbol{H}_1\boldsymbol{f}$.

2.2. Inverse problem

We aim to compute the maximum a-posteriori solution of the problem $\operatorname{argmax}_{\boldsymbol{f}} \quad \log p(\boldsymbol{m}^\alpha|\boldsymbol{f}) + \log p(\boldsymbol{f})$. We choose $p(\boldsymbol{f}) \propto \exp -\frac{1}{2}\|\boldsymbol{H}\boldsymbol{f}\|^2_{\boldsymbol{\Sigma}^{-1}}$, where $\boldsymbol{\Sigma}$ is a covariance prior, and approximate the likelihood as $p(\boldsymbol{m}^\alpha|\boldsymbol{f}) \propto \exp -\frac{1}{2}\|\boldsymbol{H}_1\boldsymbol{f} - \boldsymbol{m}^\alpha\|^2_{\boldsymbol{\Sigma}_{\boldsymbol{\alpha},\boldsymbol{f}}^{-1}}$ where $\boldsymbol{\Sigma}_{\alpha,\boldsymbol{f}}$ is a signal-dependent noise covariance matrix. While exact log likelihood optimization requires iterative algorithms [11], the problem with an approximate likelihood admits a closed form solution in the case where $\boldsymbol{\Sigma}_{\boldsymbol{\alpha},\boldsymbol{f}}$ is known

$$\mathcal{G}^\alpha_{\text{Bay}}(\boldsymbol{m}^\alpha) = \frac{1}{N}\boldsymbol{H}^\top \begin{bmatrix} \boldsymbol{I}_M \\ \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_1^{-1} \end{bmatrix} \boldsymbol{\Sigma}_1(\boldsymbol{\Sigma}_{\alpha,\boldsymbol{f}} + \boldsymbol{\Sigma}_1)^{-1}\boldsymbol{m}^\alpha, \tag{1}$$

where $\boldsymbol{\Sigma}_1$, $\boldsymbol{\Sigma}_2$ and $\boldsymbol{\Sigma}_{12}$ are the blocks of the covariance matrix defined by $\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_{21}^\top \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_2 \end{bmatrix}$, which can be precomputed as defined in [12]. Note that the left multiplication with $\frac{1}{N}\boldsymbol{H}^\top$ can be implemented using fast inverse Hadamard transform.

2.2.1. *Deep expectation maximization*

Deep-learning-based methods look for a reconstruction mapping $\mathcal{G}_{\boldsymbol{\theta}}$, where $\boldsymbol{\theta}$ represents the parameters of the deep neural network $\mathcal{D}_{\boldsymbol{\theta}}$ that corrects artifacts in the image domain. To solve the maximum a-posteriori problem, we adopt the deep expectation (EM) architecture $\boldsymbol{f}^{(k)} = \mathcal{G}^\alpha_{\boldsymbol{\theta},k}(\boldsymbol{m}^\alpha)$, where the k-th iteration $\boldsymbol{f}^{(k)}$ is defined recursively by [8]

$$\bar{\boldsymbol{f}}^{(k)} = \boldsymbol{f}^{(k)} + \mathcal{G}^\alpha_{\text{Bay}}(\boldsymbol{m}^\alpha - \boldsymbol{H}_1\boldsymbol{f}^{(k)}) \tag{2a}$$

$$\boldsymbol{f}^{(k+1)} = \mathcal{D}_{\boldsymbol{\theta}}(\bar{\boldsymbol{f}}^{(k)}) \tag{2b}$$

with the initialization $\boldsymbol{f}^{(0)} = \boldsymbol{0}$. The parameters of the network are optimized during the training phase by minimizing the mean squared error $\frac{1}{S}\sum_{s=1}^S \|\mathcal{G}_{\boldsymbol{\theta}}(\boldsymbol{m}^\alpha_{(s)}) - \boldsymbol{f}_{(s)}\|_2^2$, over an image database $\{\boldsymbol{f}_{(s)}, \boldsymbol{m}^\alpha_{(s)}\}_{1\le s\le S}$.

2.2.2. *Practical issues*

The cost of the training phase is dominated by the evaluation of Equation (1) and, in particular, by the computation of

$$\boldsymbol{\Sigma}_1(\boldsymbol{\Sigma}_{\alpha,\boldsymbol{f}} + \boldsymbol{\Sigma}_1)^{-1} = \boldsymbol{P}_{\alpha,\boldsymbol{f}}. \tag{3}$$

where the noise covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\alpha},\boldsymbol{f}}$ is estimated from the raw measurement by $\boldsymbol{\Sigma}_{\alpha,\boldsymbol{f}} = \frac{1}{\alpha^2}\text{Diag}\left(\hat{\boldsymbol{m}}^{\alpha,+} + \hat{\boldsymbol{m}}^{\alpha,-}\right)$ [12]. Due to the measurement dependency of $\boldsymbol{\Sigma}_{\alpha,\boldsymbol{f}}$,(3) cannot be precomputed and a different linear system should be solved for each of the S images in the training database. To alleviate this problem, [8] proposed to replace $\boldsymbol{P}_{\alpha,\boldsymbol{f}}$ in Equation (1) by the diagonal approximation

$$\boldsymbol{P}_{\alpha,\boldsymbol{f}} \approx \text{diag}\left(\boldsymbol{\Sigma}_1\right)\left[\boldsymbol{\Sigma}_{\alpha,\boldsymbol{f}} + \text{diag}\left(\boldsymbol{\Sigma}_1\right)\right]^{-1}. \tag{4}$$

As $\boldsymbol{\Sigma}_{\boldsymbol{\alpha},\boldsymbol{f}}$ is diagonal for independent measurements, the computation of Equation (4) is trivial. Despite the simplicity and effectiveness of this approximation, it leads to degraded image quality, as will be shown in Section 5.

3. PROPOSED METHOD

Instead of solving S different linear systems, our idea is to introduce an approximation $\boldsymbol{P}_{\alpha,\boldsymbol{f}} \approx \boldsymbol{Q}_\alpha + \boldsymbol{R}_{\alpha,\boldsymbol{f}}$, where $\boldsymbol{Q}_\alpha$ is independent of $\boldsymbol{f}$ and can be precomputed, and $\boldsymbol{R}_{\alpha,\boldsymbol{f}}$ is a correction term that depends on $\boldsymbol{f}$, but requires no system resolution during the training phase. Then, for a matrix $\widetilde{\boldsymbol{\Sigma}}_\alpha$ independent of $\boldsymbol{f}$ we note that

$$\begin{aligned}\boldsymbol{P}_{\alpha,\boldsymbol{f}} &= \boldsymbol{\Sigma}_1(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_{\alpha,\boldsymbol{f}} + \widetilde{\boldsymbol{\Sigma}}_\alpha - \widetilde{\boldsymbol{\Sigma}}_\alpha)^{-1} \\ &= \boldsymbol{\Sigma}_1(\boldsymbol{I}_M + (\boldsymbol{\Sigma}_1 + \widetilde{\boldsymbol{\Sigma}}_\alpha)^{-1}(\boldsymbol{\Sigma}_{\alpha,\boldsymbol{f}} - \widetilde{\boldsymbol{\Sigma}}_\alpha))^{-1}(\boldsymbol{\Sigma}_1 + \widetilde{\boldsymbol{\Sigma}}_\alpha)^{-1}.\end{aligned}$$

If $(\boldsymbol{\Sigma}_1 + \widetilde{\boldsymbol{\Sigma}}_\alpha)^{-1}(\boldsymbol{\Sigma}_{\alpha,\boldsymbol{f}} - \widetilde{\boldsymbol{\Sigma}}_\alpha)$ is small, then the first order Taylor approximation gives $\boldsymbol{P}_{\alpha,\boldsymbol{f}} \approx \boldsymbol{\Sigma}_1[\boldsymbol{I}_M - (\boldsymbol{\Sigma}_1 + \widetilde{\boldsymbol{\Sigma}}_\alpha)^{-1}(\boldsymbol{\Sigma}_{\alpha,\boldsymbol{f}} - \widetilde{\boldsymbol{\Sigma}}_\alpha)](\boldsymbol{\Sigma}_1 + \widetilde{\boldsymbol{\Sigma}}_\alpha)^{-1}$, which is of the form $\boldsymbol{P}_{\alpha,\boldsymbol{f}} \approx \boldsymbol{Q}_\alpha + \boldsymbol{R}_{\alpha,\boldsymbol{f}}$, where we identify $\boldsymbol{Q}_\alpha = \boldsymbol{\Sigma}_1(\boldsymbol{\Sigma}_1 + \widetilde{\boldsymbol{\Sigma}}_\alpha)^{-1}$ and $\boldsymbol{R}_{\alpha,\boldsymbol{f}} = -\boldsymbol{Q}_\alpha(\boldsymbol{\Sigma}_{\alpha,\boldsymbol{f}} - \widetilde{\boldsymbol{\Sigma}}_\alpha)\boldsymbol{\Sigma}_1^{-1}\boldsymbol{Q}_\alpha$. Note that $\boldsymbol{Q}_\alpha$ can be computed once and for all before the learning phase, while the computation of $\boldsymbol{R}_{\alpha,\boldsymbol{f}}$ only requires three matrix products. We choose the upper bound $\widetilde{\boldsymbol{\Sigma}}_\alpha = \boldsymbol{I}_M\frac{N}{\alpha}$. Indeed, by deriving the variance of the raw measurements, we obtain $\boldsymbol{\Sigma}_{\alpha,\boldsymbol{f}} = \boldsymbol{I}_M\frac{1}{\alpha}\left(\boldsymbol{H}_1^+\boldsymbol{f} + \boldsymbol{H}_1^-\boldsymbol{f}\right)$. Then, using the properties of the Hadamard matrix $\boldsymbol{H}_1^+ + \boldsymbol{H}_1^- = \boldsymbol{1}$, we have $\boldsymbol{\Sigma}_{\alpha,\boldsymbol{f}} = \boldsymbol{I}_M\frac{1}{\alpha}\sum_{n=1}^N \boldsymbol{f}_n \le \widetilde{\boldsymbol{\Sigma}}_\alpha$.

4. EXPERIMENTS

In our simulations and experiments, we consider Hadamard patterns of size $N = 64 \times 64$ pixels and $M = 1024$ measurements. As described in [13], we choose $\mathcal{D}_{\boldsymbol{\theta}}$ as a three-layer convolutional neural network with $K = 1$ iteration for

	Bayesian denoised completion			EM-Net $\alpha_{\text{train}} = \alpha$		EM-Net $\alpha_{\text{train}} = 50$ ph.	
α	Diag	Exact	Proposed	Diag	Proposed	Diag	Proposed
5	20.73 ± 1.37	21.73 ± 1.43	21.61 ± 1.57	22.18 ± 1.71	$\mathbf{22.37 \pm 1.70}$	21.62 ± 1.68	$\mathbf{22.00 \pm 1.77}$
10	21.97 ± 1.31	22.94 ± 1.42	22.82 ± 1.58	23.58 ± 1.80	$\mathbf{23.71 \pm 1.80}$	23.16 ± 1.69	$\mathbf{23.40 \pm 1.78}$

Table 1: Peak signal-to-noise ratio (PSNR) of the reconstructions over the stl-10 test set. The Bayesian denoised completion reconstruction corresponds to Equation (1) using the diagonal ('Diag') approximation, the exact ('Exact') inversion, and the proposed ('Proposed') approximation. The EM-Net reconstructions correspond to Equation (2), either trained with the true image intensity or fixed intensity of 50 photons.

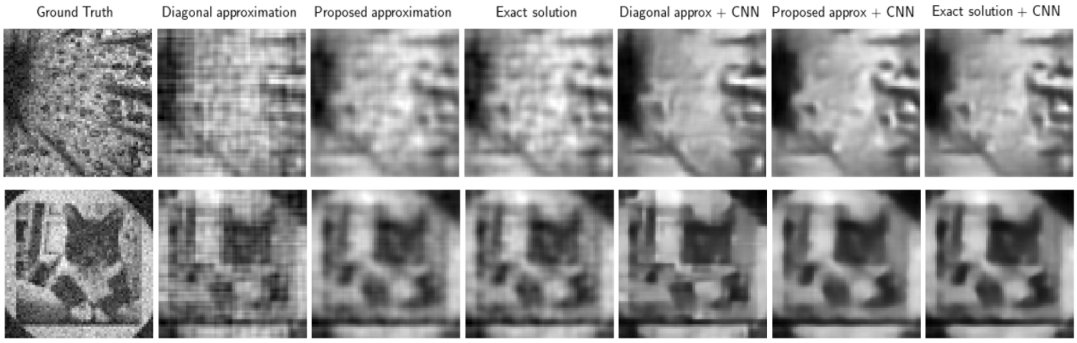

Fig. 1: Reconstructions from Bayesian denoised completion using the diagonal approximation (second column), the proposed approximation (third column) and the exact solution (fourth column), from an EM-Net using the diagonal approximation (fifth column), the proposed approximation (sixth column), and the exact solution (last column). First row: experimental data with $\alpha \simeq 5$ ph and $\alpha_{\text{train}} = 5$ ph; second row: experimental data with $\alpha \simeq 5$ ph and $\alpha_{\text{train}} = 50$ ph.

simplicity. All of the networks are trained using 105,000 images (i.e., the 'unlabeled' and 'train' subsets of the STL10 database; 8,000 images are used for the test (i.e., the 'test' subset of STL10). The original 96×96 images were cropped to 64×64, and normalized between -1 and 1. We implement our methods using the Python SPyRiT package [14], which is based on Pytorch [15]. For training, we consider the ADAM optimizer for 30 epochs. The step size is initialized to 10^{-3} and is reduced by half every 10 epochs. The weight decay regularization parameter is set to 10^{-8}. On a NVIDIA GeForce GTX 1660 Ti, the training phase takes 2 h 25 min considering the exact solution and 38 min considering both the diagonal and proposed approximations. Note that the training of a network that computes the exact solution is infeasible for $M > 1024$ or $K > 1$ due to memory allocation limitations. Therefore, we select the scenario $\{M = 1024, K = 1\}$ for which all approaches can be compared. We also image two samples using our hyperspectral single-pixel camera: a thin slice of a real cherry tomato, and a cat. The latter is an image from the STL-10 test set that we have printed on a transparent sheet. The raw datasets are freely distributed [16]. We reconstruct only one spectral channel per dataset among the 2,048 channels available. In practice, low image intensities correspond to a high noise level, therefore we propose to reconstruct images having intensities of 5 and 10 photons using networks trained with image intensities of 5, 10 and 50 photons (see Table 1 and Fig. 1).

5. RESULTS AND DISCUSSION

First, we compare the simulations of the analytical solutions computed using the different reconstruction methods (see Table 1). We observe that the diagonal approximation leads to reduced image quality compared to the exact solution. For instance, at $\alpha = 5$ photons, the mean PSNR degradation is 1 dB, while the proposed approximation performs very similarly to the exact Bayesian inversion. Then, we assess the robustness of the different reconstruction methods when the estimated light intensity α differs from the noise level used during the training of a network (second row Fig. 1). As expected, the reconstruction quality is lower when the training noise level deviates from the acquisition noise level, for both approximations. Interestingly, the superiority of the proposed approximation is more pronounced in the presence of noise deviation where only an estimate of the image intensity is available. Indeed, while most of the severe artifacts present in the images reconstructed using the diagonal approximation are smoothed out by the convolutionnal layers (compare the second and fifth column of Fig. 1), some of them remain visible. On the contrary, the reconstructions obtained with the

proposed approximation tend to keep details while removing such artifacts (compare the fifth and sixth columns with the last column of Fig. 1).

6. CONCLUSION

We consider the problem of reconstruction of an image from few single-pixel camera measurements. In particular, we focus on a network that is based on a Bayesian formulation, which necessitates the computation of generalized Tikhonov solutions. It requires the resolution of many different linear systems during the training phase, which raises memory and computational cost issues. While a diagonal approximation that does not suffer from these limitations has been proposed, it leads to degraded image quality. In this work, we propose an approximation, which has the same advantages as the diagonal approximation, while leading to improved image quality. Coupled with neural networks, this improves both the quality and the robustness of the reconstruction method.

7. 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 MILYON (ANR-10-LABX-0070) and the LABEX PRIMES (ANR-11-LABX-0063) of Université de Lyon. The images were acquired on the PILoT facility.

8. REFERENCES

[1] P. Edgar et al., "Principles and prospects for single-pixel imaging," Nature Photonics, vol. 13, no. 1, pp. 13–20, Jan. 2019.

[2] Arridge et al., "Solving inverse problems using data-driven models," Acta Numerica, vol. 28, pp. 1–174, 2019.

[3] V. Monga, Y. Li, and Y. Eldar, "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing," IEEE Signal Processing Magazine, vol. 38, no. 2, pp. 18–44, 2021.

[4] H. K. Aggarwal et al., "Modl: Model-based deep learning architecture for inverse problems," IEEE Transactions on Medical Imaging, vol. 38, no. 2, pp. 394–405, 2019.

[5] H. Gupta et al., "Cnn-based projected gradient descent for consistent ct image reconstruction," IEEE Transactions on Medical Imaging, vol. 37, no. 6, pp. 1440–1453, June 2018.

[6] 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.

[7] J. A. Fessler et al., "Space-alternating generalized expectation-maximization algorithm," IEEE Transactions on Signal Processing, vol. 42, no. 10, pp. 2664–2677, 1994.

[8] 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.

[9] M. Ochoa et al., "Assessing patterns for compressive fluorescence lifetime imaging," Opt. Lett., vol. 43, no. 18, pp. 4370–4373, Sep 2018.

[10] A. Lorente Mur et al., "Handling negative patterns for fast single-pixel lifetime imaging," in SPIE Photonics : Molecular-Guided Surgery: Molecules, Devices, and Applications V, 2019, vol. 10862.

[11] F. X Dupé et al., "Inverse problems with poisson noise: Primal and primal-dual splitting.," In 2011 18th IEEE International Conference on Image Processing, pp. 1901–1904.

[12] A. Lorente Mur, P. Leclerc, F. Peyrin, and N. Ducros, "Single-pixel image reconstruction from experimental data using neural networks," in Optics Express, 2021, vol. 29, pp. 17097–17110.

[13] N. Ducros et al., "A completion network for reconstruction from compressed acquisition," in 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), 2020, pp. 619–623.

[14] A. Lorente Mur, F. Peyrin, and N. Ducros, "Single-Pixel Python Image Reconstruction Toolbox (spyrit) Version 1.2," https://github.com/openspyrit/spyrit, 2022.

[15] Adam Paszke et al., "Pytorch: An imperative style, high-performance deep learning library," in Advances in Neural Information Processing Systems 32, pp. 8024–8035. Curran Associates, Inc., 2019.

[16] N. Ducros and L. Mahieu-Williame, "SPIHIM: A collection of single-pixel hyperspectral imaging datasets, Version 1," https://github.com/openspyrit/spihim, 2022.