LEARNING OPTIMAL PARAMETERS FOR BINARY SENSING IMAGE RECONSTRUCTION ALGORITHMS

Renán A. Rojas[⋆], Wangyu Luo[†], Victor Murray[⋆‡], and Yue M. Lu[†]

[⋆]Department of Electrical Engineering, Universidad de Ingeniería y Tecnología, Lima, Perú
[†]School of Engineering and Applied Sciences, Harvard University, MA, USA
[‡]Department of Electrical and Computer Engineering, University of New Mexico, NM, USA

I. Abstract

A novel data-driven reconstruction algorithm for quantum image sensors (QIS) is proposed. Observations are efficiently decoded by modeling the reconstruction structure as a two-layer neural network, where optimal coefficients are obtained via error backpropagation. Our model encapsulates the structure of state-of-the-art algorithms, yet it presents a faster alternative which adapts to input examples without a priori statistical information. Simulations on natural and synthetic datasets show accurate reconstructions consistent with the state of the art, while requiring 5 times less computational cost.

II. Theoretical Background

Let $\mathbf{c} = \{c_0, c_1, \ldots, c_{N-1}\}^T$ be a set of ground-truth coefficients representing the image to be encoded by a QIS linear array [1]. Let the array contain M pixels covering $x \in [0,1]$ in a uniform fashion. Then, the total light exposure value for each pixel becomes:

$$s_m = \alpha \cdot \sum_{n=0}^{N-1} c_n \cdot g_{m-Kn},$$

where α is a gain factor, g_m is a linear filter which depends on a nonnegative interpolation kernel $\varphi_{\mathrm{QIS}}(x)$ and the *box function* $\beta(x)$, and $K \triangleq \frac{M}{N}$, $K \in \mathbb{Z}^+ \backslash \{1\}$ is the spatial oversampling factor.

Photons hitting each pixel surface are denoted by realizations of a Poisson random variable Y_m. Then, the QIS observations are defined as $b_m \triangleq Q(y_m)$, where $Q(y)$ is a binary quantifier with an integer threshold q. Consequently, for random variable $B_m \triangleq Q(Y_m)$, the probability distribution $p_{b_m}(s) \triangleq \mathbb{P}(B_m = b_m, s_m)$, is defined by:

$$p_0(s) \triangleq \sum_{k=0}^{q-1} \frac{s^k}{k!} e^{-s}, \qquad p_1(s) \triangleq 1 - \sum_{k=0}^{q-1} \frac{s^k}{k!} e^{-s}.$$

Figures 1 and 2 show the imaging model and a sensing example, respectively, for the scenario of interest: $\varphi_{\mathrm{QIS}}(x) = \beta(x)$ and $q = 1$.

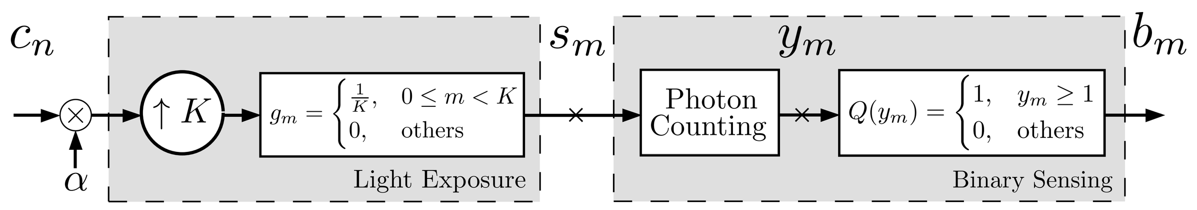
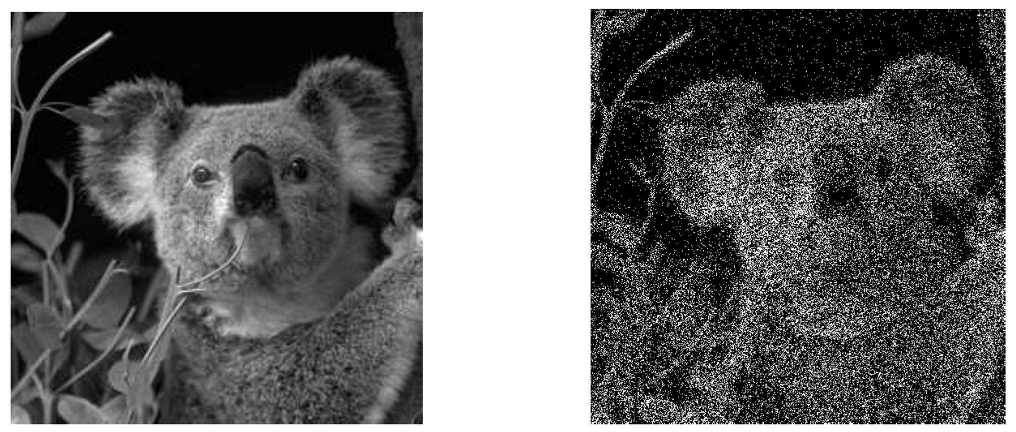


Figure 1: QIS imaging scheme for $\varphi_{\mathrm{QIS}}(x) = \beta(x)$ and $q = 1$.



(a) Ground-truth image $\mathbf{c}$. (b) QIS binary observation $\mathbf{b}$.

Figure 2: QIS imaging example for $K = 4$, $\varphi_{\mathrm{QIS}}(x) = \beta(x)$ and $q = 1$.

III. Problem Formulation

A. Parametric Representation and Optimality Criterion

A two-layer structure comprised of a concatenation of linear-shift-invariant systems and pointwise nonlinearities is proposed. Figure 3 shows the proposed structure and the following elements:

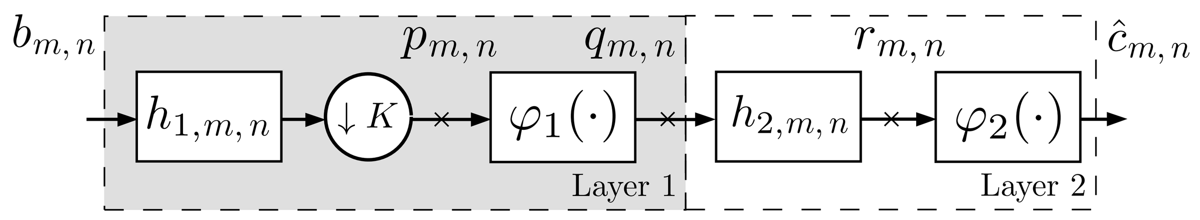


Figure 3: Proposed two-layer reconstruction algorithm.

(i) a downsampling process of factor K ($\hat{K} \triangleq \sqrt{K}$ in each dimension [2]), (ii) two linear, shift-invariant systems, $h_{1,m,n}$, $h_{2,m,n}$, and (iii) two pointwise nonlinearities $\varphi_1(\cdot)$, $\varphi_2(\cdot)$, parametrized by: $\varphi_i(z) = \sum_k w_{i,k} \cdot \beta_3(\frac{z}{\Delta_i} - k)$, where β_3 corresponds to *cubic B-splines*, and $\Delta_i \in \mathbb{R}$ is a scaling factor. Therefore, the system is characterized by:

$$\hat{c}_{m,n} = \sum_k w_{2,k} \cdot \beta_3 \left\{ \frac{r_{m,n}}{\Delta_2} - k \right\}, \qquad r_{m,n} = \sum_{s,t} h_{2,s,t} \cdot q_{m-s,n-t},$$

$$q_{m,n} = \sum_k w_{1,k} \cdot \beta_3 \left\{ \frac{p_{m,n}}{\Delta_1} - k \right\}, \qquad p_{m,n} = \sum_{s,t} h_{1,s,t} \cdot b_{\hat{K}m-s, \hat{K}n-t}.$$

The fine-tuning of each parameter is similar to the trainable structures presented in [3, 4]: motivated by accurate QIS reconstruction schemes [1, 2], a parametrized feed-forward architecture is adapted to a representative training set $\{\mathbf{b}_\ell, \mathbf{c}^*_\ell\}, \ell \in \{0, L-1\}$. Let $\mathbf{a} = \mathrm{vec}(\{\mathbf{h_1}, \mathbf{w_1}, \mathbf{h_2}, \mathbf{w_2}\})$ be the column vector formed by the system components. Then, the system parameters are optimized by:

$$\hat{\mathbf{a}} = \underset{\mathbf{a} \in \mathcal{A}}{\operatorname{argmin}} \ \frac{1}{L} \sum_\ell \varepsilon(\mathbf{a}, \mathbf{b}_\ell, \mathbf{c}^*_\ell),$$

where $\mathcal{A}$ is the feasible set and $\varepsilon(\mathbf{a}, \mathbf{b}, \mathbf{c}^*) \triangleq \frac{1}{2} \|\mathbf{c}^* - \hat{\mathbf{c}}(\mathbf{a}, \mathbf{b})\|^2_{\ell_2}$ is the cost function. Figure 4 describes the proposed optimization scenario.

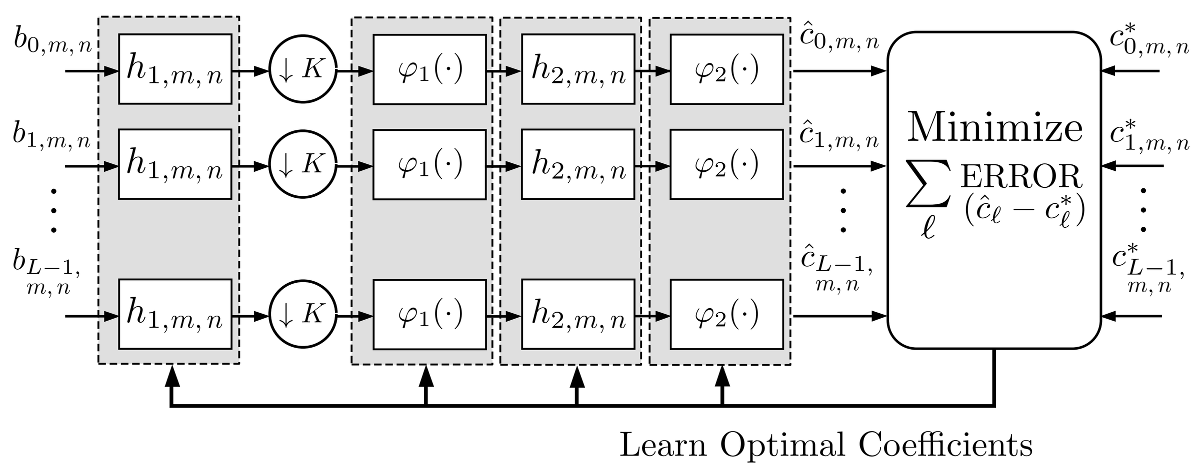


Figure 4: Proposed learning approach based on the MMSE.

B. Optimization Problem and Algorithm Initialization

Let $\hat{\mathbf{c}}(\mathbf{a})$ be the vectorized version of output $\hat{c}_{m,n}$ for parameters $\mathbf{a}$, sorted in column-wise order. Then, $\hat{\mathbf{a}}$ can be iteratively computed as:

$$\mathbf{a}^{(i)} = \operatorname{proj}_{\mathcal{A}} \{ \mathbf{a}^{(i-1)} - \mu \nabla \varepsilon (\mathbf{a}^{(i-1)}) \},$$

where $\operatorname{proj}_{\mathcal{A}}$ is an orthogonal projection operator onto set $\mathcal{A}$ and μ the step size. For each cost function in (1), the gradient $\nabla\varepsilon(\mathbf{a})$ and the Jacobian matrix $\frac{d}{d\mathbf{a}}\hat{\mathbf{c}}(\mathbf{a})$ are expressed as:

$$\nabla \varepsilon(\mathbf{a}) = \left[\frac{d}{d\mathbf{a}} \hat{\mathbf{c}}(\mathbf{a}) \right]^T (\hat{\mathbf{c}}(\mathbf{a}) - \mathbf{c}^*),$$

$$\frac{d}{d\mathbf{a}} \hat{\mathbf{c}}(\mathbf{a}) \triangleq \left[\frac{d\hat{\mathbf{c}}(\mathbf{a})}{d\mathbf{h_1}}, \frac{d\hat{\mathbf{c}}(\mathbf{a})}{d\mathbf{w_1}}, \frac{d\hat{\mathbf{c}}(\mathbf{a})}{d\mathbf{h_2}}, \frac{d\hat{\mathbf{c}}(\mathbf{a})}{d\mathbf{w_2}} \right].$$

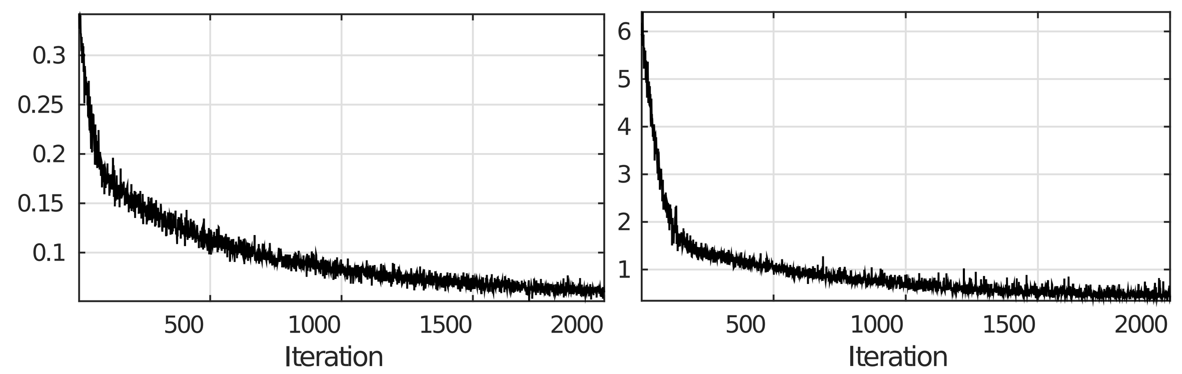
Initial weights $\mathbf{a}^{(0)}$ are set according to the non-iterative reconstruction algorithm [2] with a lowpass filter as denoising method. For the 2D scenario, let L^1 be re-defined as: $L^1_{m,n} \triangleq \sum_{s=0}^{\hat{K}-1} \sum_{t=0}^{\hat{K}-1} b_{\hat{K}m+s, \hat{K}n+t}$. Then, $h_{1,m,n}$ is initialized as follows:

$$h^{(0)}_{1,m,n} = \begin{cases} 1, & -\hat{K} < m < 1 \wedge -\hat{K} < n < 1 \\ 0, & \text{other cases} \end{cases}.$$

Similarly, $h_{2,m,n}$ is initialized as a Gaussian lowpass filter with standard deviation σ heuristically selected between $[0.25, 1]$. $\varphi_1(z)$ is initialized as the *Anscombe transform* $\mathcal{T}$. Finally, $\varphi_2(z)$ is initialized as the *inverse Anscombe transform* $\mathcal{T}^{-1}$, followed by the logarithmic function $-\log(1 - \frac{x}{K})$ and the factor $\frac{K}{\alpha}$: $\varphi^{(0)}_2(z) = -\frac{K}{\alpha} \log\left(1 - \frac{\mathcal{T}^{-1}(z)}{K}\right)$.

IV. Numerical Results

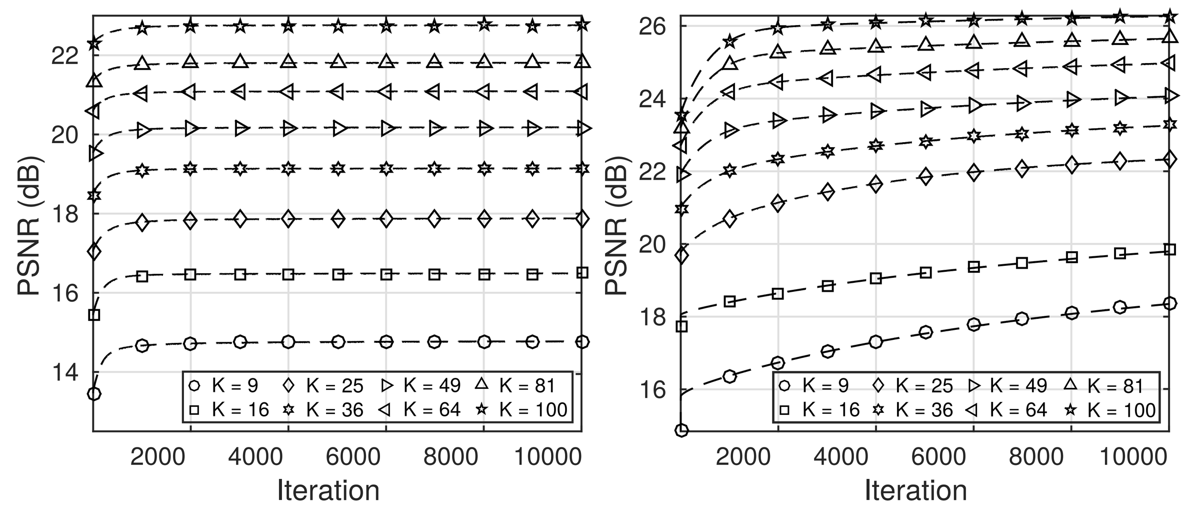
For the synthetic scenario, c^* is generated as a 32×32 random matrix with standard uniform distribution. Online learning is performed on 128 training samples per iteration and a test set of 512 samples. Figure 5 shows the cost function values and its gradient norm at each iteration for $K = 36$. Their behavior reflects the reconstruction improvement along the learning procedure. Figure 6a shows the average PSNR of the reconstructed data set for different K values.



(a) Cost function. (b) Gradient norm.

Figure 5: Gradient descent for the synthetic scenario ($K = 36$).

For the natural scenario, the Cifar-10 dataset ($5 \cdot 10^4$ training samples and 10^4 test samples of 32×32 pixels) is considered. Online learning is performed on 128 training samples per iteration. Figure 6b shows the average PSNR of the reconstructed data set for different K values. Also, performance on larger images is evaluated on the Berkeley Segmentation Dataset (200 training samples and 100 test samples of 481×321 pixels). Figure 7 shows results for the proposed algorithm (MMSE), the maximum likelihood estimation (MLE) [1] and the non-iterative method (NI) with BM3D as denoising algorithm [2, 5]. MMSE yields results comparable with NI, while better preserving the scene structure.



(a) Synthetic scenario. (b) Natural scenario.

Figure 6: Learning process for synthetic and natural scenarios.

Processing time is compared using Matlab-only code (1.2 GHz Intel core i7, L2: 256K, RAM: 4G) for $K = 16$ and the described setup on the Berkeley dataset. On average, MLE requires 52 ms to process a single sample, whereas NI and MMSE require 2377 ms and 453 ms, respectively. This shows that the proposed algorithm achieves a comparable reconstruction quality but is 5.25 times faster than NI.

(a) MLE output.	(b) NI output.	(c) MMSE output.
SNR= 8.29 dB	SNR= 15.58 dB	SNR= 15.51 dB
PSNR= 16.34 dB	PSNR= 23.63 dB	PSNR= 23.56 dB
SSIM= 0.37	**SSIM= 0.55**	**SSIM= 0.60**

(d) MLE output.	(e) NI output.	(f) MMSE output.
SNR= 8.87 dB	SNR= 15.78 dB	SNR= 15.6 dB
PSNR= 13.28 dB	PSNR= 20.19 dB	PSNR= 20.01 dB
SSIM= 0.37	**SSIM= 0.41**	**SSIM= 0.51**

Figure 7: Reconstruction accuracy for the proposed algorithm and alternative methods on natural images ($K = 16$).

V. Conclusions

The proposed QIS image reconstruction algorithm obtains accurate estimations consistent with state-of-the-art methods, while showing a more efficient design. Modeled as a simple neural network, optimal components are learned directly from examples without any statistical assumptions, achieving reconstructions with coherent structural similarity 5 times faster than alternative methods, as experimentally demonstrated. Further work will focus on adding layers in the structure while preserving its computational efficiency to explore more complex initial settings and alternative binary sensing scenarios.

References

[1] F. Yang, Y. M. Lu, L. Sbaiz, and M. Vetterli, "Bits from photons: Oversampled image acquisition using binary poisson statistics," *IEEE Transactions on image processing*, vol. 21, no. 4, pp. 1421–1436, 2012.

[2] S. H. Chan, O. A. Elgendy, and X. Wang, "Images from Bits: Non-Iterative Image Reconstruction for Quanta Image Sensors," *Sensors*, vol. 16, no. 11, p. 1961, 2016.

[3] A. M. Bronstein, P. Sprechmann, and G. Sapiro, "Learning Efficient Structured Sparse Models," in *Proceedings of the 29th International Conference on Machine Learning*, 2012, p. 33.

[4] U. S. Kamilov and H. Mansour, "Learning optimal nonlinearities for iterative thresholding algorithms," *IEEE Signal Processing Letters*, vol. 23, no. 5, pp. 747–751, 2016.

[5] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Transactions on image processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

Aknowledgements

This work was supported by the UTEC-Harvard Academic Collaboration Funds under the Grant Join Research Seed Fund 2015-03, and by the US NSF under grant CCF-1319140.