FROM OBSERVATIONS TO THEORETICAL CONSISTENCY: DECODER RECOVERY IN CODED APERTURE IMAGING USING CONVOLUTIONAL NEURAL NETWORKS

Jocelyn Ornelas Munoz, Erica M. Rutter, Roummel F. Marcia

Department of Applied Mathematics, University of California, Merced

2024 IEEE Conference on Computational Imaging Using Synthetic Apertures (CISA) | 979-8-3503-0864-8/24/$31.00 ©2024 IEEE | DOI: 10.1109/CISA60639.2024.10576438

ABSTRACT

Coded aperture imaging, crucial for low-light imaging in challenging conditions, requires specific decoders for image reconstruction. Traditional image reconstruction methods can be complex and focus on reconstruction rather than learning the underlying decoder. Our work introduces a one-layer CNN network for interpretable decoder recovery, without prior knowledge of encoding or decoding arrays. Using observed detector images, the network produces reconstructed images using a learned decoder. We train our network using the MNIST dataset and report high accuracy in image reconstruction even for images with high signal to noise ratio. To validate the generalizability of the method, we show that the MNIST-trained CNN-learned decoder is able to accurately reconstruct images from the grayscale FashionMNIST dataset.

***Index Terms*—** Coded aperture imaging, interpretable deep learning, image reconstruction, convolutional neural network, deep learning

1. INTRODUCTION

Deep learning has become ubiquitous in the modern world. As deep learning models become more complex and sophisticated, the need to provide explanations regarding its predictions becomes more important. Interpretability and generalizability continues to be a focus of much research in deep learning [1]. Interpretability in deep learning extends beyond the models themselves to encompass various facets, including the input data and model parameters. Moreover, a recognized limitation of deep learning models lies in their ability to generalize effectively to new, unseen data. The research presented specifically emphasizes the *mathematical* interpretability and generalizability of convolutional neural networks within the domain of coded aperture imaging.

Coded aperture (CA) imaging emerged from the need to increase the photon count reaching a detector in optical systems without compromising resolution, such as by enlarging the diameter of a pinhole. The basic idea is to create a mask pattern that introduces a more complex point spread function compared to that of a pinhole, leveraging this pattern to produce high-quality image reconstructions.

Significant advancements have been made in CA imaging with the creation of Modified Uniformly Redundant Arrays (MURAs) [2] – designed to enhance decoding capabilities and improve image reconstruction. MURAs are mathematically tailored to increase the redundancy of the coded aperture pattern, enabling better noise suppression, higher imaging resolution, and more accurate scene reconstruction based on detector array measurements. Our focus in this study will involve studying images encoded using MURA.

Typically, when radiation emitted from a source (denoted as $\mathbf{S}$) interacts with a binary aperture mask ($\mathbf{A}$), it casts a shadow of an object. This mask is designed with a pattern of openings that allow a significant portion of photons to reach a position-sensitive detector, capturing spatial information from the source. The resulting recorded image ($\mathbf{D}$) is unrecognizable as it represents a transformed version of $\mathbf{S}$ and lacks resemblance to the original source structure. To be meaningful, the observed image ($\mathbf{D}$) must undergo a reconstruction process to identify the location and intensity of each source within the field of view, thus producing an approximation ($\hat{\mathbf{S}}$) of the original source image.

Coded aperture technology originated to address challenges in x-ray and gamma-ray imaging, although it has now been widely adopted in astronomy, remote sensing, surveillance systems, and biomedical imaging [3, 4, 5,] To successfully reconstruct images without access to the underlying decoder, complex mathematical algorithms and computational methods, including the Maximum Entropy Method (MEM) [6, 7], wavelet-based techniques [8, 9], and deep convolutional neural networks (CNNs) [10], have been used. Our approach differs from conventional methods as our objective is to recover the decoding function itself, rather than solely focusing on image reconstruction.

2. PROBLEM FORMULATION

In the context of coded aperture imaging, the arrival of photons at the detector are modeled by the following process:

$$\mathbf{D} = \mathbf{S} * \mathbf{A} + \mathbf{B}$$

This work is supported by National Science Foundation grant DMS 1840265.

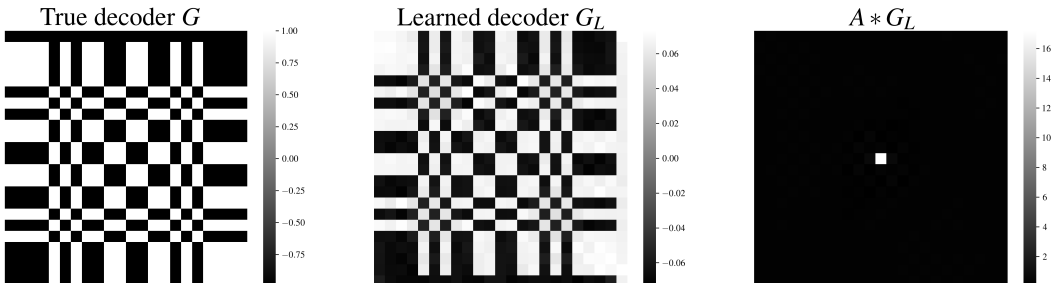


Fig. 1. (a) Depiction of true decoder $\mathbf{G}$. (b) Learned decoder $\mathbf{G}_L$. (c) Convolution $\mathbf{A} * \mathbf{G}_L$ (right).

where $\mathbf{A} \in \{0,1\}^{n\times n}$ is the binary coded aperture, $\mathbf{S} \in \mathbb{R}^{n\times n}$ is the source signal, $\mathbf{B} \in \mathbb{R}^{n\times n}$ is the background noise, $\mathbf{D} \in \mathbb{R}^{n\times n}$ is the observed image at the detector stage, and $*$ denotes the convolution operator (see [2]).

The reconstruction is given by

$$\hat{\mathbf{S}} = \mathbf{D} * \mathbf{G}$$

where $\mathbf{G} \in \mathbb{R}^{n\times n}$ represents the mathematical function or decoding procedure used to approximate the original source. Our interest is recovering the decoding function given coded observations, without knowledge of the encoding array. That is, we seek to solve the unconstrained minimization problem:

$$\mathbf{G}_L = \arg\min_{\hat{\mathbf{G}}} \Phi(\hat{\mathbf{S}}, \mathbf{S}) \tag{1}$$

where we seek to adjust the weights $\hat{\mathbf{G}}$ in order to improve the quality of the approximate output $\hat{\mathbf{S}}$. The loss function Φ is minimized using backpropagation [11].

3. METHODOLOGY

We propose to use a single convolutional layer with no activation function and no bias with the purpose of evaluating the potential of deep learning in obtaining an interpretable and mathematically accurate approximation for the decoding function by solving $\hat{\mathbf{S}} = \mathbf{D} * \mathbf{G}_L$. In doing so, we develop a fully data-driven method without explicit knowledge of the imaging system. For this reason, we do not impose constraints on the decoding function. The output $\hat{\mathbf{S}}$ is compared to the target or source image $\mathbf{S}$ by letting Φ be the mean squared error (MSE) loss function given by

$$\Phi(\hat{\mathbf{S}}, \mathbf{S}) = \frac{1}{|\mathcal{S}|} \sum_{i=1}^{|\mathcal{S}|} \|\hat{\mathbf{S}}_i - \mathbf{S}_i\|_F^2 \tag{2}$$

where $\mathcal{S}$ is the dataset and $|\mathcal{S}|$ is its cardinality.

4. NUMERICAL EXPERIMENTS

Training Datasets. We train our method on $60,000$ images from the MNIST dataset. MNIST is composed of $70,000$ grey-scale 28×28 pixel images. We used 2-D modified uniformly redundant arrays to encode the data [2]. The size of each MURA is defined by a prime integer p.

Testing Datasets. We evaluate our method on two datasets: (i) MNIST and (ii) FashionMNIST. We test on the $10,000$ held-out images from MNIST. To test the generalizability of the learned decoder, we then directly apply our trained CNN to encoded FashionMNIST images, which consists of $10,000$ greyscale images of size 28×28.

Performance. We trained our model exclusively on the MNIST training data for 10 epochs, utilizing the mean squared error (MSE) defined in Equation (2) as our loss function. The model was tested on both MNIST and FashionMNIST datasets. The model has p^2 parameters, which represent the learned decoder. Following the reconstruction of the images using both methods, we utilized two metrics to assess the quality of the reconstructions: mean squared error (MSE) and structural similarity index measure (SSIM) [12]. Metrics are calculated between the reconstructed image $\hat{\mathbf{S}}$ and the original image $\mathbf{S}$. SSIM measures perceived changes between images and is given by

$$\text{SSIM}(\hat{\mathbf{S}}, \mathbf{S}) = \frac{(2\mu_{\hat{\mathbf{S}}}\mu_{\mathbf{S}} + c_1)(2\sigma_{\hat{\mathbf{S}}\mathbf{S}} + c_2)}{(\mu_{\hat{\mathbf{S}}}^2 + \mu_{\mathbf{S}}^2 + c_1)(\sigma_{\hat{\mathbf{S}}}^2 + \sigma_{\mathbf{S}}^2 + c_2)} \tag{3}$$

where $\mu_{\hat{\mathbf{S}}}$ and $\mu_{\mathbf{S}}$ represent the mean values and $\sigma_{\hat{\mathbf{S}}}$ and $\sigma_{\mathbf{S}}$ denote the corresponding variances of $\hat{\mathbf{S}}$ and $\mathbf{S}$, respectively. c_1 and c_2 are used to stabilize the denominator.

To evaluate the learned decoder $\mathbf{G}_L$, we consider the two constraints posed by Gottesman and Fenimore [2]: (1) the decoding function be the correlational inverse of $\mathbf{A}$, i.e. $\mathbf{A} * \mathbf{G}_L \approx \delta$ where δ is the Kronecker-delta function, and (2) the constraint that $\mathbf{G}_L$ be unimodular, i.e., the elements of $\mathbf{G}_L$ have equal magnitude. All codes are implemented in Pytorch and available at `github.com/jornelasmunoz/coded-aperture`.

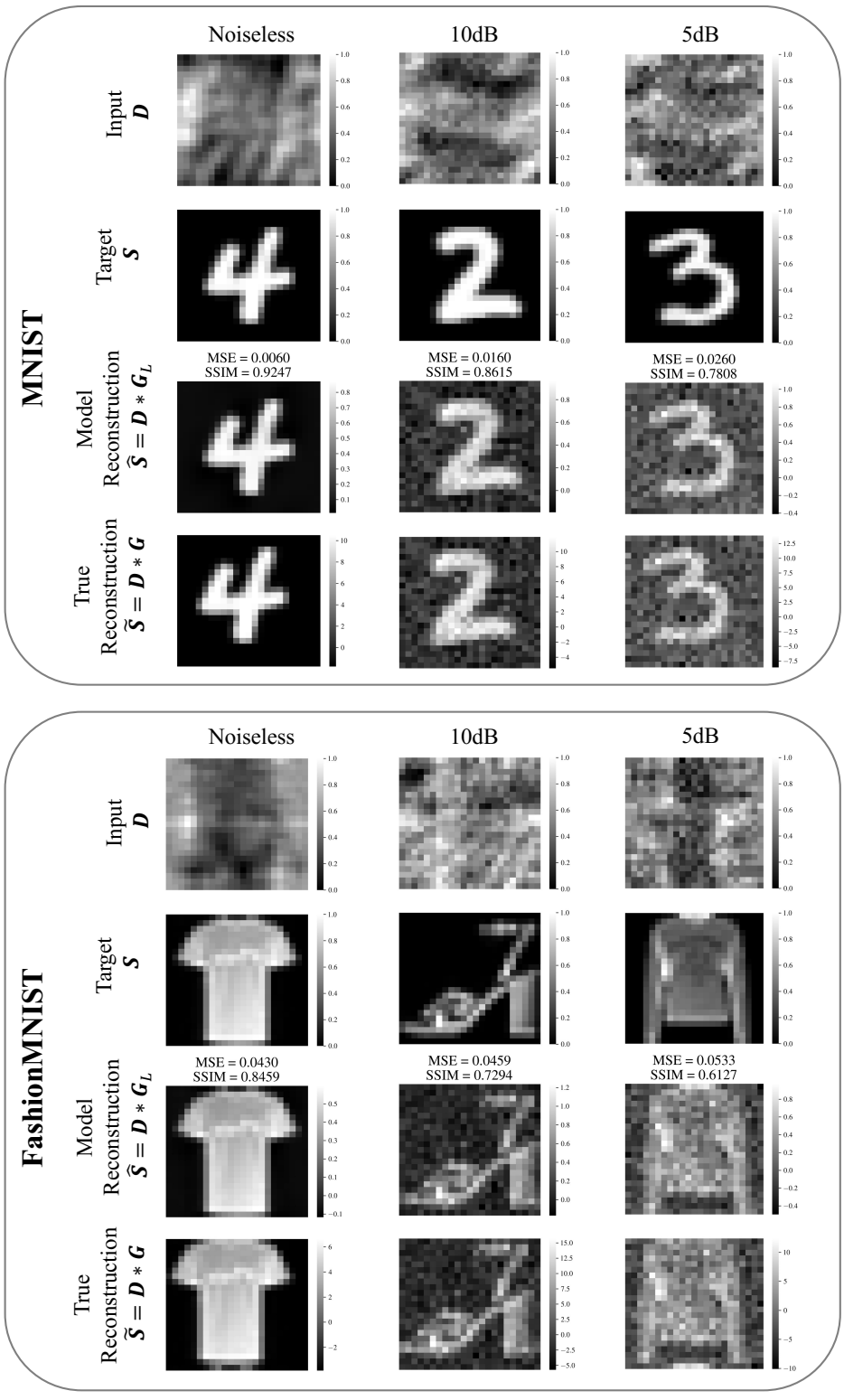
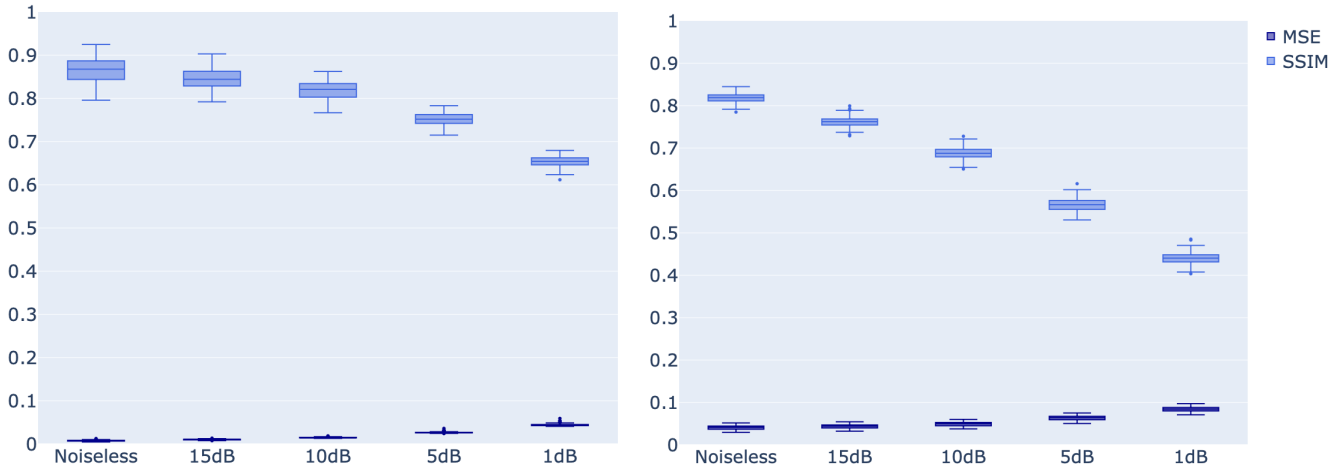


Fig. 2. Numerical experiments on 3 images from the MNIST dataset and FashionMNIST dataset at varying SNR levels. Row 1 and 5: input images $\mathbf{D}$. Rows 2 and 6: Ground truth images $\mathbf{S}$. Rows 3 and 7: Final reconstructions $\hat{\mathbf{S}}$ using the learned decoder $\mathbf{G}_L$. Rows 4 and 8: Final reconstructions $\tilde{\mathbf{S}}$ using the true decoder $\mathbf{G}$. MSE and SSIM values between $\hat{\mathbf{S}}$ and $\mathbf{S}$ are presented for each model reconstruction.



(a) MNIST (a) FashionMNIST

Fig. 3. Box plots showing image reconstruction quality at varying Signal-to-Noise Ratio levels measured in decibels (dB) using learned decoder with $p = 23$. (a) MNIST dataset and (b) FashionMNIST dataset. Dark blue shows mean squared error (MSE) and light blue shows (SSIM).

5. RESULTS

In this section we present results from our approach. We will evaluate the performance based on two criteria: (1) accuracy of the learned decoder and (2) mean-squared error between the learned decoded images and the source images.

Figure 1(b) displays the learned weights for $p = 23$. We can see that the learned decoder $\mathbf{G}_L$ is a rotated version of the true decoder (Figure 1(a)). This makes sense mathematically since Pytorch implements a cross-correlation function but calls it convolution for convolutional neural networks. A cross-correlation and a convolution function are equivalent, but the kernel is flipped [13]. Further, the learned decoder satisfies the constraint that $\mathbf{A} * \mathbf{G}_L \approx \delta$ (Figure 1(c)). We observe that the elements of matrix $\mathbf{G}_L$ exhibit unequal magnitudes, attributed to the stochastic nature of gradient descent methods. The variability introduced by stochastic gradient descent methods, such as Adam [14], adds complexity to the convergence process, making it challenging to reach the exact minimum [15].

Next we turn to evaluating the reconstructions. The example reconstructions, illustrated in Figure 2, showcase the performance of the learned decoder across different signal-to-noise ratios: noiseless, 10dB, and 5dB. The first and fifth rows present the coded images, while the third and seventh rows display the corresponding reconstructions achieved with the learned decoder. As reference, the second and sixth row show the ground truth images and the fourth and eighth rows show reconstructions depicting the true decoded images.

Figure 3 presents overall MSE and SSIM on $10,000$ MNIST testing images and $10,000$ FashionMNIST testing images for the model with $p = 23$. A single model was trained on noiseless MNIST data and subsequently tested on varying signal-to-noise (SNR) levels measured in decibels (dB) for both MNIST and FashionMNIST data. We observe that the pixel values of the reconstructions do not fall within the range $[0, 1]$, as we did not apply an activation function. The average MSE value and average SSIM value for MNIST noiseless data was 7.82×10^{-3} and 0.86, respectively. For FashionMNIST noiseless data, the average MSE value was 4.10×10^{-2} and the average SSIM value was 0.82.

6. CONCLUSIONS

This paper presents a one-layer convolutional neural network grounded in mathematical principles for reconstructing recorded observations obtained from a coded aperture detector. In addition to reconstructing the images, the learned weights of our one-layer CNN represent the decoder. We demonstrate that a single model, solely trained on binary MNIST images, exhibits robust performance when directly applied to grayscale images from FashionMNIST with varying noise levels. Our results demonstrate the network's capacity to generalize effectively across diverse datasets, underscoring its proficiency in capturing underlying patterns in new, unseen data obtained from coded aperture systems. Furthermore, our investigation reveals that the learned model weights effectively represent a decoding function that satisfies requisite mathematical constraints. This validation underscores the model's fidelity in adhering to fundamental principles.

7. REFERENCES

[1] Zachary C Lipton, "The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery.," *Queue*, vol. 16, no. 3, pp. 31–57, 2018.

[2] Stephen R Gottesman and E Edward Fenimore, "New family of binary arrays for coded aperture imaging," *Applied optics*, vol. 28, no. 20, pp. 4344–4352, 1989.

[3] Jing Chen, Yongtian Wang, and Hanxiao Wu, "A coded aperture compressive imaging array and its visual detection and tracking algorithms for surveillance systems," *Sensors (Basel, Switzerland)*, vol. 12, pp. 14397 – 14415, 2012.

[4] Ezio Caroli, JB Stephen, G Di Cocco, L Natalucci, and A Spizzichino, "Coded aperture imaging in x-and γ-ray astronomy," *Space Science Reviews*, vol. 45, pp. 349–403, 1987.

[5] Ana Ramirez, Henry Arguello, Gonzalo R Arce, and Brian M Sadler, "Spectral image classification from optimal coded-aperture compressive measurements," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 6, pp. 3299–3309, 2013.

[6] Trevor J. Ponman, "Maximum entropy methods," *Nuclear Instruments and Methods in Physics Research*, vol. 221, no. 1, pp. 72–76, 1984, Proceedings of the International Workshop on X- and γ-Ray Imaging Techniques.

[7] Richard Willingale, Mark R. Sims, and Martin J.L. Turner, "Advanced deconvolution techniques for coded aperture imaging," *Nuclear Instruments and Methods in Physics Research*, vol. 221, no. 1, pp. 60–66, 1984.

[8] Roummel F Marcia and Rebecca M Willett, "Compressive coded aperture superresolution image reconstruction," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2008, pp. 833–836.

[9] Mário AT Figueiredo, Robert D Nowak, and Stephen J Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE Journal of selected topics in signal processing*, vol. 1, no. 4, pp. 586–597, 2007.

[10] Rui Zhang, Pin Gong, Xiaobin Tang, Peng Wang, Cheng Zhou, Xiaoxiang Zhu, Le Gao, Dajian Liang, and Zeyu Wang, "Reconstruction method for gamma-ray coded-aperture imaging based on convolutional neural network," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 934, pp. 41–51, 2019.

[11] Jürgen Schmidhuber, "Deep learning in neural networks: An overview," *Neural networks*, vol. 61, pp. 85–117, 2015.

[12] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[13] Ian Goodfellow, Yoshua Bengio, and Aaron Courville, *Deep Learning*, MIT Press, 2016, `http://www.deeplearningbook.org`.

[14] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[15] Sebastian Ruder, "An overview of gradient descent optimization algorithms," *arXiv preprint arXiv:1609.04747*, 2016.