Deep Adaptive Hadamard Single Pixel Imaging in Near-Infrared Spectrum

Brayan Esneider Monroy Chaparro

Master thesis to qualify for the title of

Master of Systems Engineering and Computer Science

Advisor

Jorge Luis Bacca Quintero

PhD. in Computer Science

Co-Advisor

Henry Arguello Fuentes

PhD. in Electrical and Computer Engineering

Universidad Industrial de Santander

Faculty of Physicomechanical

Department of Systems Engineering

Bucaramanga

2024

Dedicatory

To my family.

Acknowledgments

Developing this work would have been impossible without the support and assistance of numerous amazing people. Here is a list of those I would like to thank, in no specific order.

- To my advisor, Professor Jorge Bacca, and my co-advidsor, Professor Henry Arguello, for their constant support and guidance in my academic and personal growth.
- To Tatiana Gelvez-Barrera, Claudia Correa, Laura Galvis and Andres Jerez, for your invaluable guidance in improving my presentation skills and communication of findings.
- To all members of the HDSP group who have helped me with research projects, administrative issues and made long hours more pleasant with their friendly company.
- To the HDSP optical laboratory team, especially to Hans Garcia and Ana Karina, who offered invaluable help.
- To my family who has supported me in all my decisions during this important stage of my life.
- To Leonel, my first mentor and compass in my professional career.
- To my friends, Camila, Valentina, Natalia, Karen and Maria, who have provided unconditional friendship and support.
- To my cat Ciri, who ensured that I never slept more than necessary.

Table of Contents

List of Figures

Abstract

Title: Deep Adaptive Hadamard Single Pixel Imaging in Near-Infrared Spectrum [*]

Autor: Brayan Esneider Monroy Chaparro [**]

Keywords: Adaptive Sensing, Hadamard Single Pixel, Deep Learning.

Description: Compressed Spectral Imaging (CSI) is an efficient technology used to obtain spatial and spectral information. Unlike conventional spectral scanning techniques, CSI captures only a few snapshots, avoiding long acquisition times and expensive specialized sensors. This approach has applications in precision agriculture and remote sensing. In the near-infrared (NIR) range, beyond the visible spectrum, CSI is particularly valuable since sensors with high spatial and spectral resolutions are not available. Specifically, the Hadamard Single-Pixel Imaging (HSPI) presents cost-effective acquisition systems to capture high-quality NIR images. However, capturing compressed images leads to information loss, requiring image restoration algorithms for spectral image reconstruction. In the HSPI, information loss occurs when using a subset of snapshots (some rows of the Hadamard matrix) instead of the full Hadamard matrix, directly affecting the quality of acquired spectral images. Therefore, it is crucial to develop a sensing system that optimizes the subset choosed by maximization information while minimizes the required snapshots. While existing literature focuses on computational algorithms and deep learning models for a fixed sensing system, the challenge lies in designing a system that effectively addresses the unique features of each spectral image to extract the most relevant information.

Consequently, this work presents two adaptive deep learning models for the selection of adaptive modulation

---

[*] Master Thesis

[**] Faculty of Sciences. Department of Physics. Advisor: Henry Arguello Fuentes, Ph.D. in Electrical and Computer Engineering.

patterns in the Hadamard-based single-pixel imaging system for Near-Infrared Spectral Imaging. Specifically, the proposed models perform adaptive selection of modulation patterns in two modalities. In cases where side information is available, the estimation of image superpixels is proposed to enhance spatial-structural information in images. For scenarios without side information, an adaptive ordering strategy is suggested, where an initial fixed subset is acquired, and a deep learning model guides the estimation of subsequent subset. This involves optimizing information extraction, reducing the number of captures, and improving modulation pattern selection. This approach incorporates modeling the sensing system's propagation process, designing an adaptive deep learning architecture, defining a suitable cost function and training strategy for the jointly optimization of network parameters and the imaging system, and validating the model performance using spatial and spectral quality metrics in NIR spectral images. The EuroSAT dataset from the Sentinel-2 mission serves as the basis for validation. The performance of the adaptive model is analyzed based on acquisitions carried out in the HDSP Optical Laboratory. The outcome of this research is a deep learning model capable of adaptively acquiring spectral images and enhancing their spatial and spectral quality by extracting relevant information.

Resumen

Título: Adquisición adaptativa de imágenes espectrales comprimidas en el infrarrojo cercano basada en el sistema de captura de único píxel mediante aprendizaje profundo. *

Autor: Brayan Esneider Monroy Chaparro **

Palabras Clave: Muestreo Adaptativo, Camara de unico pixel Hadamard, Aprendizaje profundo.

Descripción: La adquisición de imagenes espectrales comprimidas (CSI, por sus siglas en inglés) son una tecnología eficiente utilizada para obtener información espacial y espectral. A diferencia de las técnicas convencionales de escaneo espectral, CSI captura solo unas pocas imágenes instantáneas, evitando largos tiempos de adquisición y sensores especializados costosos. Este enfoque tiene aplicaciones en la agricultura de precisión y en la teledetección. En el rango del infrarrojo cercano (NIR), más allá del espectro visible, CSI es particularmente valiosa ya que no se dispone de sensores con alta resolución espacial y espectral. Específicamente, la técnica de imagen de un solo píxel basada en Hadamard (HSPI) presenta sistemas de adquisición rentables para capturar imágenes NIR de alta calidad. Sin embargo, la captura de imágenes comprimidas conlleva una pérdida de información, lo que requiere algoritmos de restauración de imágenes para la reconstrucción espectral. En HSPI, la pérdida de información ocurre al utilizar solo un subconjunto de capturas en lugar de todo el conjunto de la matriz Hadamard, lo que afecta directamente la calidad de las imágenes espectrales adquiridas. Por lo tanto, es crucial desarrollar un sistema de adquisición que optimice la extracción de información y minimice el número de capturas necesarias. Mientras que la literatura existente se centra en algoritmos computacionales y modelos de aprendizaje profundo para un sistema de captura fijo,

---

* Trabajo de grado

** Facultad de Ingenierias Fisicomecánicas. Escuela de Ingenieria de Sistemas e Informatica. Director: Jorge Luis Bacca Quintero, Doctorado en Ciencias de la Computación.

el desafío radica en diseñar un sistema que aborde eficazmente las características únicas de cada imagen espectral para extraer la información más relevante.

En consecuencia, este trabajo presenta dos modelos adaptativos de aprendizaje profundo para la selección de patrones de modulación adaptativa en el sistema de imágenes de un solo píxel basado en Hadamard para imágenes espectrales en el infrarrojo cercano. En concreto, los modelos propuestos realizan la selección adaptativa de patrones de modulación en dos modalidades. En los casos en los que se dispone de información previa, se propone la estimación de superpíxeles de imagen para mejorar la información espacial-estructural de las imágenes. Para escenarios sin información previa, se propone una estrategia de ordenación adaptativa, donde se adquiere un subconjunto fijo inicial, y un modelo de aprendizaje profundo guía la estimación del subconjunto subsiguiente. Este enfoque incluye el modelado del proceso de propagación del sistema de detección, el diseño de una arquitectura de aprendizaje profundo adaptativo, definir una función de costes y una estrategia de entrenamiento adecuadas para optimizar conjuntamente los parámetros de la red y el sistema de adquisición de imágenes, y la validación del rendimiento del modelo mediante métricas de calidad espacial y espectral en imágenes espectrales NIR. El conjunto de datos EuroSAT de la misión Sentinel-2 se emplea para la validación. El rendimiento del modelo adaptativo se analizo a partir adquisiciones realizadas en el Laboratorio Óptico HDSP. El resultado de esta investigación es un modelo de aprendizaje profundo capaz de adquirir imágenes espectrales de forma adaptativa y mejorar su calidad espacial y espectral a través de la extracción de información relevante.

Research Products

Contributions of the thesis

- Adaptive Learning framework for Hadamard Single Pixel imaging system in two scenarios.
    - Access to side-information: Adaptive sensing based on deep learned superpixels.
    - Non-access to side-information: Adaptive sensing based on adaptive ordering of the Hadamard matrix.
- We proposed a re-formulation of the coventional end-to-end training as a binary classification problem to improve the performance of the model in the HSPI system.
- We validate the proposed adaptive framework with both simulation and real acquisition, where our method outperforms the non-adaptive methods.
- We present an efficient Hadamard row-wise generation algorithm with $\mathcal{O}(2^n)$ asymptotic computational complexity.

Publications

The developments of this thesis have been published in various international journals and conferences.

Journal papers:

1. **Monroy, B.**, Sanchez, K., Arguello, P., Estupiñán, J., Bacca, J., Correa, C. V., ... & Rojas-Morales, F. (2023). Automated chronic wounds medical assessment and tracking framework based on deep learning. Computers in Biology and Medicine, 165, 107335: 10.1016/j.compbiomed.2023.107335

2. **Monroy, B.**, Estupiñan, J., Gelvez-Barrera, T., Bacca, J., & Arguello, H. (2024). Designed Dithering Sign Activation for Binary Neural Networks. IEEE Journal of Selected Topics in Signal Processing. 10.1109/JSTSP.2024.3467926

3. **Monroy, B.**, Garcia, H., Arguello, H., Bacca, J. (To be Submitted). Deep Adaptive Ordering of Hadamard Basis for Near-Infrared Single-Pixel Imaging.

Conference papers:

1. **Monroy, B.**, Bacca, J.,& Arguello, H. (2024, April). Predicting The Spectrum: Deep Adaptive Sensing for Hadamard Single Pixel Spectral Imaging. In 2024 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW). IEEE. doi:10.1109/ICASSPW62465.2024.10627208

2. Bacca, J., **Monroy, B.** & Arguello, H. (2024, April). Deep Plug-and-Play Algorithm for Unsaturated Imaging. In 2024 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP). IEEE. doi: 10.1109/ICASSP48485.2024.10446495

3. J Bacca, M Carlsson, **B Monroy**, & Arguello, H. (2024, April). Plug-And-Play Al-

gorithm Coupled with Low-Rank Quadratic Envelope Regularization for Compressive Spectral Imaging. In 2024 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) IEEE. doi:10.1109/ICASSP48485.2024.10447145

4. Jacome, R., **Monroy, B.**, Bacca, J., & Arguello, H. (2023, June). Stochastic Deep Image Prior for Multishot Compressive Spectral Image Fusion. In 2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW) (pp. 1-5). IEEE. doi:ICASSPW59220.2023.10193340

5. **Monroy, B**., Bacca, J., & Arguello, H. (2023, June). Deep Adaptive Superpixels For Hadamard Single Pixel Imaging In Near-Infrared Spectrum. In ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 1-5). IEEE. doi:ICASSP49357.2023.10095165

6. **Monroy, B**., Contreras, K., and Bacca, J. (2024). Autoregressive High-Order Finite Difference Modulo Imaging: High-Dynamic Range for Computer Vision Applications. In ECCVW European Conference on Computer Vision Workshops (ECCVW).

7. Contreras, K., **Monroy, B**., and Bacca, J. (2024). High Dynamic Range Modulo Imaging for Robust Object Detection in Autonomous Driving. In ECCVW European Conference on Computer Vision Workshops (ECCVW).

8. **Monroy, B**., and Bacca, J. (Submitted). OptimUS: Optimization-based Unlimited Sampling Algorithm.

1. Introduction

Spectral images contain spatial information of scenes along multiple wavelengths of the electromagnetic spectrum. This data allows us to identify specific characteristics of materials and objects by analyzing the response to absorption or reflectance at different wavelengths of light Garini et al. (2006); Bacca et al. (2023). The unique spectral behavior of materials and/or objects, known as the spectral signature, provides useful information for high-level tasks such as detection, classification, and segmentation. These high-level tasks allow applications in diverse areas, including remote sensing, medical imaging, and precision agriculture Shaw and Burke (2003). In particular interest is the near-infrared (NIR) range, which spans approximately 780 to 2500 nanometers Davies (2005). Spectral imaging in NIR extends beyond the visible spectrum, enabling the identification of additional properties that are different from those captured in the visible spectrum. For example, a substantial portion of the thermal radiation emitted by objects near room temperature falls within the infrared spectrum Teena and Manickavasagan (2014). In addition, spectral images that contain NIR information offer invaluable information for vegetation-related applications, facilitating the estimation of the normalized difference vegetation index, which is crucial for environmental monitoring and agricultural management Zhu et al. (2018). However, challenges remain in acquiring high-resolution spectral images in the NIR spectrum, as the resolution of commercial systems is directly tied to the cost of the sensors; specifically, current commercial scanning infrared sensors have a sensor pitch size of around 10 μm and are about 100 times

more expensive than digital visible cameras with comparable resolution Rogalski (2022); Rogalski et al. (2016). This relationship directly affects the quality of commercial solutions.

The Single-Pixel Imaging (SPI) system has emerged as an outstanding and cost-effective solution to acquire spectral images, leveraging the principles of Compressive Sensing (CS) Duarte et al. (2008). In contrast to conventional scanning methods, which sequentially capture subsets of spectral images, SPI systems capture inner products between the entire scene and a set of modulation patterns resulting in a substantial reduction in hardware complexity and cost. Specifically, in the SPI system, the spatial resolution is defined from modulation patterns, such as coded apertures implemented through a Digital Mirror Device (DMD) Sampsell (1994), and the spectral resolution is defined by the resolution of the spectrometer. This split between spatial and spectral resolutions offers a cost-effective approach, particularly beneficial in the NIR range, where conventional 2D sensors face technological and physical limitations Centrone (2015); Gibson et al. (2020).

However, a well-known limitation of SPI systems is the need to use a substantial number of modulation patterns, resulting in long acquisition times Garcia et al. (2020); Monroy et al. (2023); Bacca et al. (2020). Consequently, significant efforts have been made to reduce the required number of snapshots to speed up image acquisition and recovery processes. In particular, Hadamard SPI (HSPI) introduces an innovative approach in which modulation patterns are derived from rows of the orthogonal Hadamard matrix, which are later reshaped into two-dimensional matrices Vaz et al. (2020). The binary and orthogonal properties of the Hadamard matrix enable feasible optical implementations and fast image

recovery through matrix multiplication Zhang et al. (2017b). In compression scenarios, a subset of the Hadamard matrix rows is selected, resulting in a reduction in the number of modulation patterns employed and a subsequent decrease in acquisition time. Various strategies have been explored in the ordering of the Hadamard matrix to select modulation patterns that preserve the maximum image quality of the observed scene Vaz et al. (2020). These include order based on the count of blocks of Hadamard patterns Yu (2019); Sun et al. (2017), maximization of total variation Yu et al. (2020), and utilization of the structural geometry of the sensing path López-García et al. (2022); Cai et al. (2022).

Various alternatives have been proposed to adapt a different modulation pattern for each sensed image. For example, the authors in Diaz et al. (2018) propose a gradient thresholding algorithm to compute the consecutive color-coded aperture from a low-resolution estimation. Specifically, in the case of SPI, some work has proposed adaptive methodologies for selecting codification functions based on wavelet-based scans. For instance, Dekel (2008) proposes a tree-structured gradual selection that considers a *father-son* relationship between wavelet coefficients to predict the relevant coefficients at finer and finer scales. The extended work in Averbuch et al. (2012) changed the prediction strategy of the dictionaries, the estimation of discrete probability distributions, and mutual information to decide whether the coefficient was significant. However, these methodologies do not employ binary modulation patterns, which are not trivial to implement in SPI with block-unblock light modulation optical elements, unlike binary modulation patterns used in the Hadamard matrix.

On the other hand, the integration of deep learning techniques with compressive

sensing has brought significant advancements in spectral imaging Huang et al. (2022). In particular, deep neural networks have improved the reconstruction of spectral images in compressive spectral imaging by learning non-linear data representations Choi et al. (2017) or its incorporation into optimization recovery algorithms Zheng et al. (2021); Wang et al. (2019). In addition, deep learning architectures have been used to improve compressive spectral imaging systems directly to acquisition by designing optical coding elements using the well-known deep optical coding design (DOCD) data-driven framework. Specifically, optical elements are treated as adjustable parameters in the design of acquisition systems, allowing joint optimization of hardware and software components Arguello et al. (2023).The optimization methods used by DOCD, focused on enhancing image quality, lead to a general boost in performance across various computational imaging applications, including compressed spectral imaging Bacca et al. (2021), privacy-preserving pose estimation Hinojosa et al. (2022) and privacy-scene captioning Arguello et al. (2022).

Both Hadamard ordering techniques and the DOCD approach have limitations in their practical applicability across various scenes. Hadamard ordering often assumes fixed spectrum geometries for all images, overlooking the unique attributes inherent in each spectral image. This approach tends to converge to a standardized design, neglecting the diverse requirements of individual scenes and potentially compromising the optimization of selection patterns for optimal image acquisition and recovery Monroy et al. (2023). Similarly, DOCD, once the optical parameters are trained using a specific dataset, converges towards a generalized design for the training scenes. This approach overlooks the unique characte-

ristics of individual scenes, failing to achieve the optimal spectral image quality tailored to each scene's specific demands. These limitations highlight the need for adaptive strategies that consider the distinct characteristics and attributes of each spectral image to ensure the selection of modulation patterns optimized for individual scenes.

2. Problem statement

In the field of spectral image acquisition, techniques based on compressed sensing have been proposed to acquire a multiplexed measurements containing both the spatial and spectral information of the scene, to reduce acquisition times and the cost of the acquisition system employed Arce et al. (2013). However, acquiring a compressed version of the scene results in an unavoidable loss of information, which is mitigated by computational algorithms and deep learning models that take advantage of certain properties present in the desired images, such as sparse representations and low-rank structures Arce et al. (2013); Arguello et al. (2023). These assumptions have also been used for the optimal design of the acquisition scheme to improve the quality of the acquired images.

In the particular case of the single-pixel camera, multiple binary or grayscale modulation patterns are used to sequentially measure the total intensity of the modulated scene for each pattern Edgar et al. (2019). In addition, the optimal design of this set of modulation patterns has been investigated using orthogonal representation basis, such as Hadamard or Fourier Zhang et al. (2017b). In this way, a simple but effective method for obtaining the spectral image reconstruction is by applying the inverse linear transform of the basis used during acquisition, thus reducing the computational cost. The literature has addressed strategies for the ordering of the Hadamard matrix Vaz et al. (2020), where the rows of the Hadamard matrix are permuted following some selection priority order. For example, some approaches such as block count ordering Yu (2019), frequency order, recursive Russian doll-

based orderings Sun et al. (2017), and the use of total variation scores Yu et al. (2020). These ordering methods seek to optimize data acquisition by efficiently selecting the modulation patterns which captures the most relevant spatial information, resulting in an improvement in the quality of the reconstructed images.

However, the current ordering strategies assume that a single acquisition design will obtain satisfactory performance throughout the data domain, which can lead to suboptimal solutions. Each spectral image has its particularities that must be considered to achieve an optimal acquisition system design. There are outstanding challenges in design optimization, reduction of the number of captures, and improvement of the quality of spatio spectral information retrieval Diaz et al. (2018). The optimal sensing matrix, which allows preserving the most relevant information of each scene, varies significantly for the case of subsets of the Hadamard matrix, as presented in Figure 1. Therefore, the research question arises: How can we design a deep learning model to achieve an adaptive design and acquisition of compressed spectral images in the single-pixel capture system?

Figure 1. Effect of Hadamard ordering on Kodak images. The ideal Hadamard ordering technique (Sequency, Cake Cutting, Zig Zag, XY) that maintains the most significant spatial information and image quality differs depending on the specific scene.

3. Objectives

General Objective

To develop a deep-learning model for the adaptive design of the sensing matrix for the acquisition and recovery of near-infrared spectral images in the Hadamard single-pixel imaging system.

Specific Objectives

1. To model the adaptive acquisition protocol based on the Hadamard single-pixel imaging system for compressed near-infrared spectral imaging.

2. To design an adaptive deep-learning model that jointly considers the network trainable parameters and the Hadamard single-pixel imaging system.

3. To define the cost function and training strategy for the optimization of trainable parameters of the adaptive deep learning model.

4. To validate the performance of the proposed model in a set of defined metrics in terms of spatial and spectral quality with state-of-the-art methods.

4. Computational Spectral Imaging

Computational spectral imaging, as discussed in Garini et al. (2006), involves the capture and analysis of spectral images. These images provide detailed information on different wavelengths of light, enabling advanced analysis beyond conventional imaging methods. Various scanning spectral imaging techniques are employed, such as whisk broom scanners, push broom scanners, and spectral band scanners Fowler (2014). Each of these techniques offers unique advantages in terms of data acquisition and image quality. Using these methods, computational spectral imaging enhances the ability to capture comprehensive spectral data for diverse applications, ranging from medicine Levenson and Mansfield (2006); Aloupogianni et al. (2021) to environmental monitoring Stuart et al. (2019); De Biasio et al. (2010).

Compression sensing (CS)-based computational spectral imaging systems, such as single-pixel cameras Duarte et al. (2008) and compressed coded apertures Arce et al. (2013), offer a promising alternative to traditional scanning techniques. By acquiring compressed measurements, CS-based imaging systems significantly reduce the number of snapshots or the overall acquisition time required for high-quality spectral imaging Arce et al. (2013). The mathematical model of spectral imaging systems often leads to ill-conditioned problems, which pose substantial challenges for accurate image reconstruction Ribes and Schmitt (2008). Then, in a linear-approximation setting the image formation in the CS-based computational spectral imaging system can be defined as follows

$$\mathbf{y} = \mathbf{A}\mathbf{x}, \tag{1}$$

where $\mathbf{y} \in \mathbb{R}^m$ is the vector of compressed measurements, $\mathbf{A} \in \mathbb{R}^{m \times n}$ is the linear approximation of the imaging system, and $\mathbf{x} \in \mathbb{R}^n$ represents the vectorization of the spectral image. Here, in most imaging models, $m < n$, which means that a significantly smaller number of measurements have been captured compared to the number of desired variables, imposing an ill-conditioned problem.

Spectral image reconstruction is fundamentally formulated as an optimization problem, leveraging the principles of CS to effectively sense signals Candès and Wakin (2008). This process typically includes an image formation model, represented as the fidelity term $f(\mathbf{x})$, that aims to minimize the difference between the observed data $\mathbf{y}$ and the reconstructed signal $\mathbf{A}\mathbf{x}$, typically minimizing the mean square error, that is, $f(\mathbf{x}) = \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$, as presented by the authors in Boyd and Vandenberghe (2004). To improve the quality of reconstruction, exploiting prior knowledge about the signal, a regularization term $g(\mathbf{x})$ is introduced. Then, the image restoration optimization problem can be expressed as

$$\hat{\mathbf{x}} \in \arg\min_{\mathbf{x}} f(\mathbf{x}) + g(\mathbf{x}). \tag{2}$$

Regularization strategies focus on incorporating prior knowledge about the natural signal $\mathbf{x}$, such as Total Variation, Sparsity, and Low-Rank assumptions Elad et al. (2023). Notable regularization methods include Tikhonov regularization $g(\mathbf{x}) = \|\mathbf{x}\|_2^2$, spatial smoothness

$g(\mathbf{x}) = \|\mathbf{D}\mathbf{x}\|_2^2$ with $\mathbf{D}$ a spatial finite difference matrix, wavelet sparsity $\mathbf{g}(\mathbf{x}) = \|\mathbf{W}\mathbf{x}\|_1$ with $\mathbf{W}$ an orthonormal basis, and various sparsity methods $g(\mathbf{x}) = \|\boldsymbol{\alpha}\|_0$ st. $\mathbf{x} = \mathbf{W}^\top \boldsymbol{\alpha}$.

4.1. Single Pixel Spectral Imaging

In a single-pixel spectral imaging system, a spectrometer captures the inner product between a set of modulation patterns $\{\mathbf{a}_i\}_{i=1}^n$ and the spatial vectorization of a spectral image Duarte et al. (2008). In practice, modulation of light using these patterns can be achieved in two ways. The first method involves placing a light modulator in the image plane of a camera lens. This setup masks images of the scene, and the coded intensities are measured by the single-pixel detector Dong et al. (2019). The second method involves modulating the illumination source to structure the incident rays on the scene, a technique known as structured illumination Zhang et al. (2017a).

The spectral image, denoted as $\mathbf{X} \in \mathbb{R}^{n \times c}$, is a matrix where $n = hw$ represents the total number of spatial pixels, with h and w as the spatial dimensions, and c denotes the number of spectral bands. Here, $\mathbf{X}$ is the horizontal stack of vectorized spatial information by bands. In this context, the image formation model can be described as

$$\mathbf{Y} = \mathbf{A}\mathbf{X}. \tag{3}$$

In this context, reducing the number of required patterns and measurements is necessary for single-pixel imaging systems to be widely adopted. There are several strategies for constructing the sensing matrix $\mathbf{A}$. Typically, binary patterns or low-bit-precision grayscale patterns

are preferred. Examples include random binary patterns, Hadamard patterns, Fourier patterns, or patterns learned through neural networks.

4.1.1. Hadamard Single Pixel. In the Hadamard case, the modulation patterns consist of rows of a Hadamard matrix of n order, denoted as $\mathbf{H} \in \{-1, 1\}^{n \times n}$, with each row of $\mathbf{H}$ representing a modulation pattern $\mathbf{h}_i$ of the set $\{\mathbf{h}_i\}_{i=1}^{n}$, as presented in Figure 2. Consequently, the acquisition process of the HSPI system involves the linear transformation of the NIR spectral image using the orthogonal Hadamard basis. This transformation is equivalent to acquiring the Hadamard coefficients $\mathbf{Y} \in \mathbb{R}^{n \times c}$ for each spectral band of the NIR image spectral, which can be defined mathematically as follows.

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{Z}, \tag{4}$$

where, $\mathbf{Z}$ denotes noise inherited from the physical imaging system. In this sense, using the complete set of modulation patterns corresponds to a full acquisition of the NIR spectral image in the Hadamard spectrum, which means that no compression occurs. However, images in the Hadamard spectrum often exhibit a sparse representation, characterized by a high number of near-to-zero coefficients, which implies an unnecessary amount of data capture. Therefore, different strategies have been explored to select the minimum number of modulation patterns that capture the most relevant information on the spectral image. Specifically, when access to previous snapshots of the desired scene is available, such as remote sensing images or complementary information from low-cost imagers like visible grayscale digital

cameras (also known as side information), the Hadamard patterns can be adapted to the structural information using superpixel maps. In cases where side information is not available, the selection of modulation patterns can be addressed by ordering the Hadamard matrix based on some priority order.

Figure 2. Hadamard Single Pixel Imaging. The starting spectral image is spatially modulated using a specific pattern and then focused through a collimating lens, after which the spectral information is detected and recorded by a sensor array.

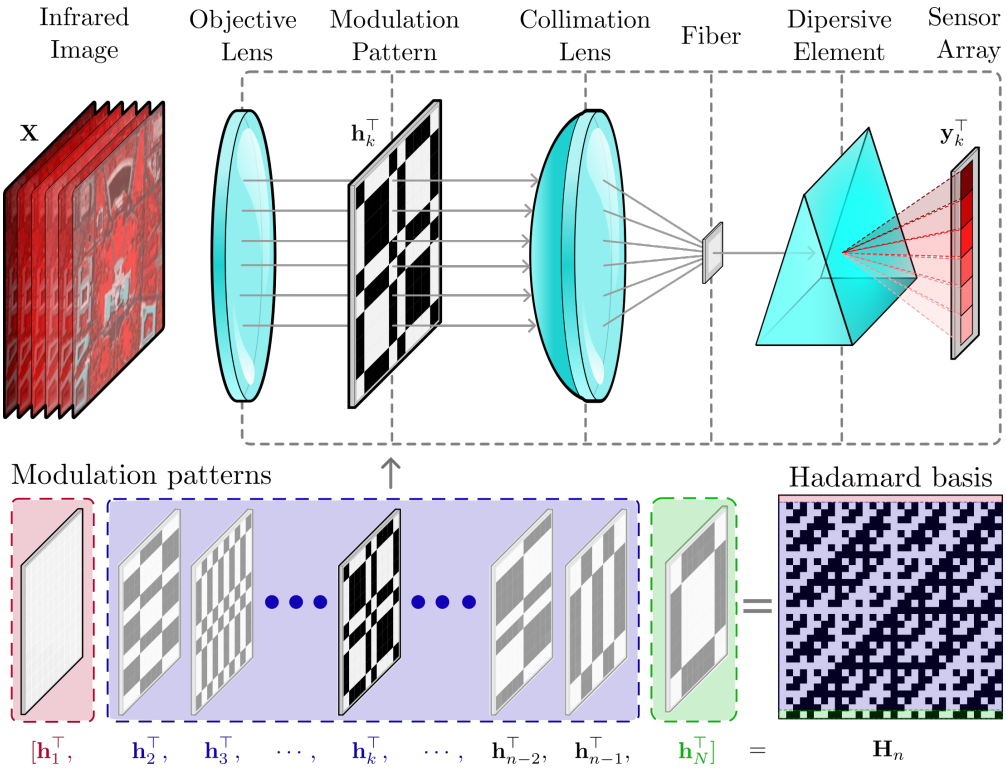
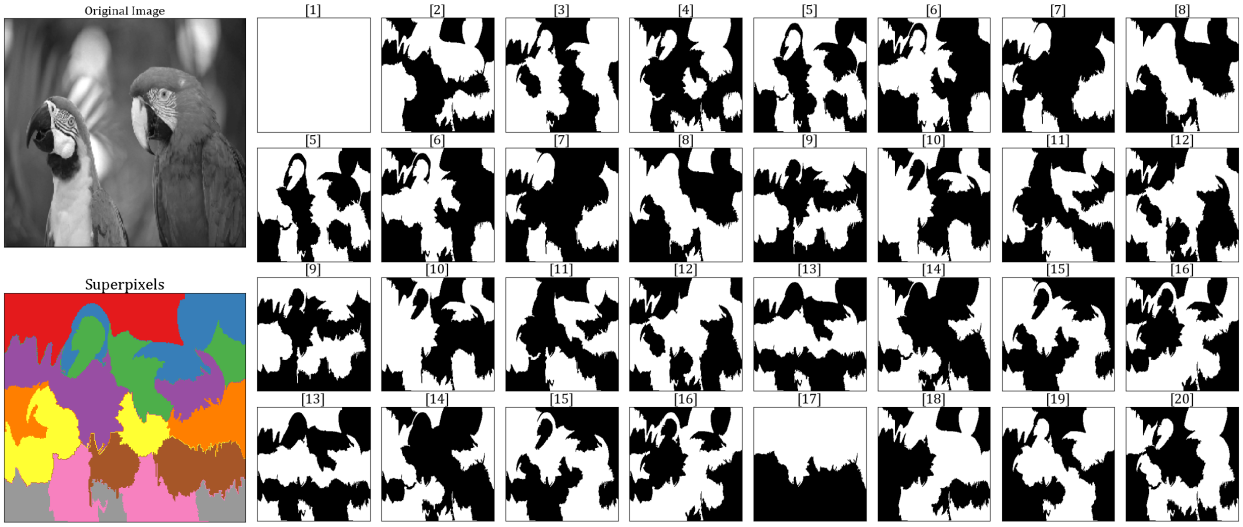


Figure 3. Hadamard superpixels for the Kodak image of a parrot. Twenty superpixels are computed for the reference image. These superpixel maps are then demultiplexed across all modulation patterns to preserve 50 % transmittance.



Hadamard Single Pixel by Superpixels

The superpixels method relies on reformulating the design of the sensing matrix as the composition of a Hadamard matrix and a decimation matrix, **HD**, where the decimation matrix $\mathbf{D} \in \{0,1\}^{m \times n}$ is designed based on prior information of the target scene Garcia et al. (2020). In this sense, the decimation matrix **D** commonly follows a superpixel structure as an irregular decimation strategy to group local regions of the image with similar spectral information Garcia et al. (2020). This approach reduces the number of captures from n to m, where m stands for the number of superpixels, preserving the most relevant spectral-structural information. Initially, it is possible to employ the decimation matrix **D** as a set of modulation patterns without projecting onto the Hadamard matrix. However, the natural transmittance of these modulation patterns will be significantly low, approximately with

a transmittance value of $1/m$. In this context, the inclusion of the Hadamard matrix is a useful strategy to preserve a $50\,\%$ transmittance for all modulation patterns regardless of the number of superpixels, serving as a 'demultiplexing' strategy, as illustrated in Figure 3. In this sense, mathematically the image formation model from Equation (3) can be rewritten as follows

$$\mathbf{Y} = \mathbf{HDX} + \mathbf{Z}. \tag{5}$$

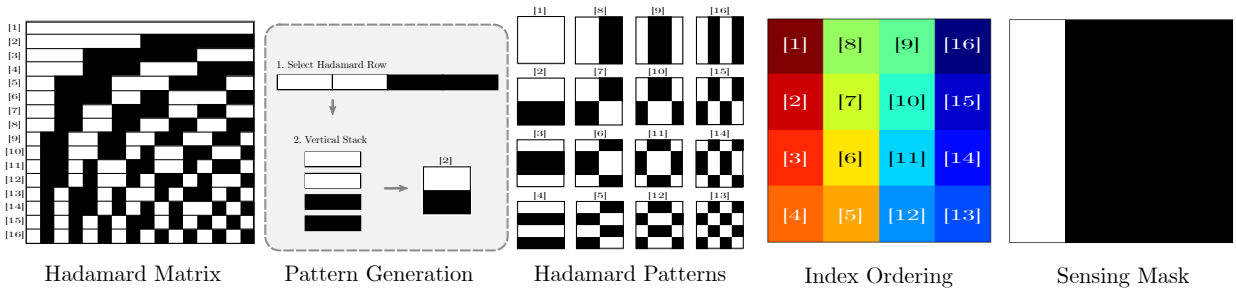
Consequently, the optimization problem of the spectral image reconstruction by minimization of the ℓ_2 cost function can be performed from the pseudo-inverse computation as follows

$$\hat{\mathbf{X}} \approx (\mathbf{HD})^{\dagger}\mathbf{Y} = \mathbf{D}^{\dagger}\mathbf{H}^{\dagger}\mathbf{Y}. \tag{6}$$

Based on the fact that the Hadamard matrix, its an orthogonal basis, its pseudo-inverse is given as $\mathbf{H}^{\dagger} = \frac{1}{m}\mathbf{H}^{\top}$. Additionally, the decimation of the decimation matrix $\mathbf{D}^{\dagger}$ can be efficiently approximated as $\mathbf{D}^{\dagger} \approx \text{diag}\Big(\frac{1}{\mathbf{D1}_n}\Big)$ if $\mathbf{1}_m^{\top}\mathbf{D} = 1$, since, the superpixels map are mutually exclusive. Therefore, Equation (6) can be rewritten as

$$\hat{\mathbf{X}} = \frac{1}{m}\mathbf{D}^{\top}\text{diag}\Big(\frac{1}{\mathbf{D1}_n}\Big)\mathbf{H}^{\top}\mathbf{Y}. \tag{7}$$

Figure 4. Generation of Hadamard patterns following a sequence order. Each row of the Hadamard matrix is assigned with an index ordering. The Hadamard pattern is generated as a vertical stack of segments for each row of the Hadamard matrix. All Hadamard patterns are re-arranged in an index ordering grid. Finally, the sensing masks are constructed by selecting the top-k pattern with the first index values.



Hadamard Single Pixel by Ordering

The ordering method relies on the construction of the sensing matrix as the composition of $\mathbf{A} = \mathbf{P}^\top\mathbf{SPH}$, where $\mathbf{P} \in \{0,1\}^{n\times n}$ is a row permutation matrix, and $\mathbf{S} = \text{diag}(\mathbf{s})$ is a diagonal matrix that selects a set of m-top rows of the Hadamard matrix based on the entries of the binary mask $\mathbf{s} \in \{0,1\}^n$, as represented in Figure 4. This construction consists of the location of the most relevant modulation patterns in the upper rows to be selected by the binary mask $\mathbf{s}$. In this sense, mathematically the image formation model from Equation (3) can be rewritten as follows

$$\mathbf{Y} = \mathbf{P}^\top\mathbf{SPHX}. \tag{8}$$

The sampling ratio is given by $\delta = \|\mathbf{s}\|_0/n$. Once captured, a coarse spectral image reconstruction based on the solution of $||\mathbf{Y} - \mathbf{P}^\top\mathbf{SPHX}||_F$ can be obtained using the orthogonal characteristics of the Hadamard matrix, which acts as the transpose of the sensing matrix

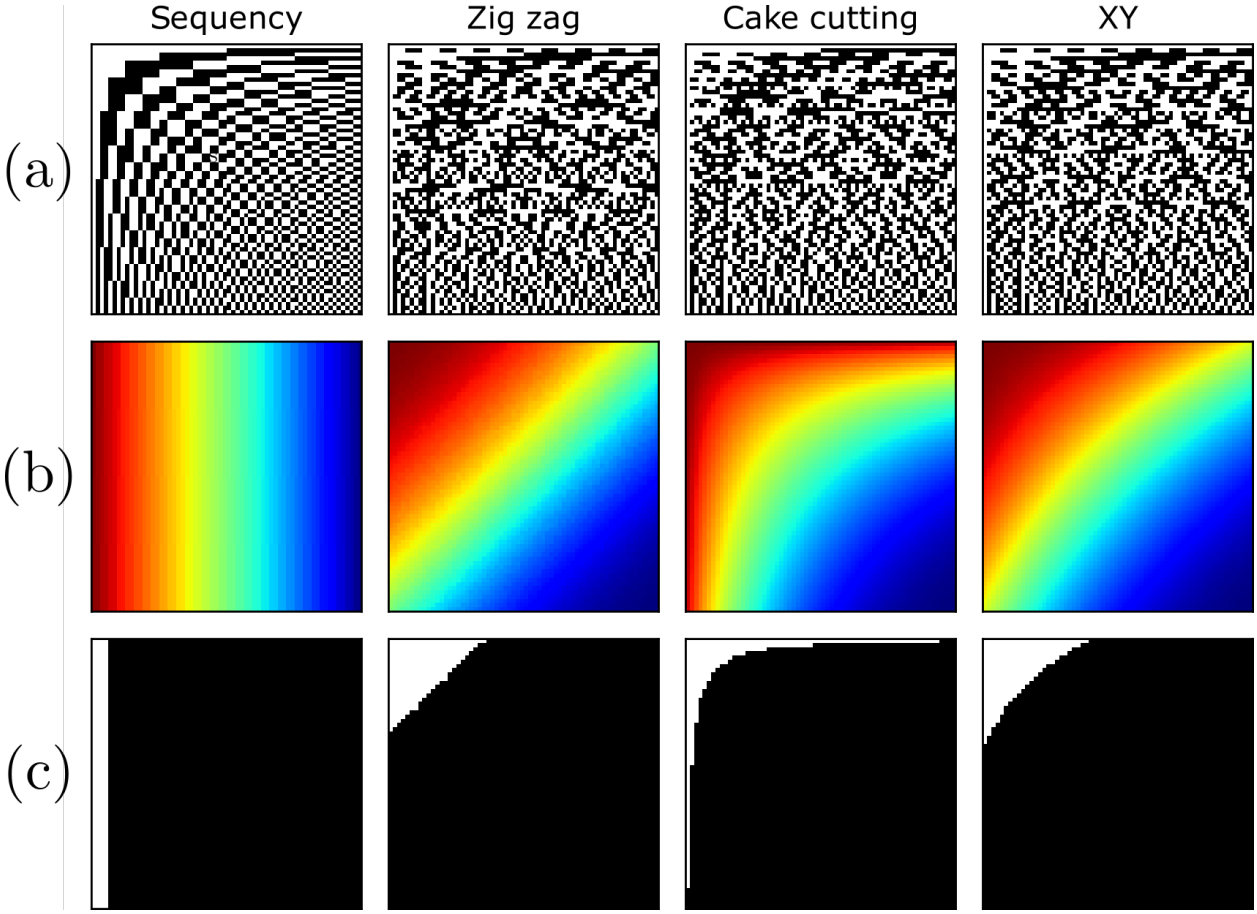
as

$$\hat{\mathbf{X}} = \frac{1}{n}(\mathbf{H})^{\top}\mathbf{Y}. \tag{9}$$

This premise assumes that the non-acquired elements are substituted with zeros and that the Y matrix preserves the structure of a complete acquisition.

Various ordering methods have been proposed, each exploiting some structure of natural images or modulation patterns. Among the most prominent methods are frequency ordering, block count ordering Yu (2019), and total variation score Yu et al. (2020). Frequency sorting involves rearranging the rows of the Hadamard matrix so that the number of sign changes in each row increases gradually. On the other hand, block-count ordering considers the resizing of each modulation pattern as a matrix and rearranges the rows of the Hadamard matrix so that the number of blocks in each row increases incrementally Yu (2019). In terms of total variation ordering, similar to the block counting approach, the ordering is performed on the rescaled representation, but in this case the rows of the Hadamard matrix are ordered ascendingly according to the value of the total variation regularizer Yu et al. (2020). When analyzing the geometry of the Hadamard spectrum in its two-dimensional representation, considering the order of the indices for each type of order, a similar behavior can be observed. In all cases, the selection of modulation patterns associated with low frequencies is prioritized and gradually the modulation patterns associated with higher frequencies are chosen. However, there are differences in the geometry depending on the selected algorithm, as shown in Figure 5.

Figure 5. Hadamard Ordering Strategies. (a) The first row illustrates diverse orderings of the Hadamard matrix. (b) The second row demonstrates geometry of coefficients relevance on the sensing path of the 2D Hadamard spectrum. (c) The final row displays the binary masks used for selecting the 2D Hadamard spectrum to be sensed.



4.2. End to End Optical Design

The End-to-End Optical Design of acquisition systems involves incorporating the imaging system $\mathbf{A}$ as an additional layer in the network architecture by modeling it as an *optical layer*. Specifically, $\mathcal{A}_\phi \approx \mathbf{A}$, where ϕ represents trainable optical parameters associated with physical optical elements of the imaging systems (such as the modulation patterns in the SPI system). Consequently, this framework aims for the joint optimization of trainable optical parameters, encapsulated in the *optical layer* $\mathcal{A}_\phi$, and the neural network parameters of a *computational decoder* $\mathcal{N}_\theta$, through an end-to-end (E2E) optimization.

Specifically, E2E optimization consists of training the optical-ϕ and decoder-θ parameters to minimize some cost function $\mathcal{L}_{task}$ and simultaneously address the physical constraints in the optical parameters $\mathcal{R}_\rho(\phi)$. Then, the final objective consists of obtaining the set of optimal parameters of the acquisition system and the optimal weights of the computational decoder $\{\phi^*, \theta^*\}$ as follows

$$\{\phi^*, \theta^*\} \in \arg\min_{\phi,\theta} \mathbb{E}_{\mathbf{x}}[\mathcal{L}_{task}(\mathcal{N}_\theta(\mathcal{A}_\phi(\mathbf{x})), \mathbf{y}) + \rho\mathcal{R}_\rho(\phi) + \sigma\mathcal{R}_\sigma(\theta)], \tag{10}$$

with, the set $\{\mathbf{x}_k, \mathbf{y}_k\}_{k=1}^K$ as our training database, comprising K elements. In this set, $\mathbf{x}_k$ represents the input image, while $\mathbf{y}_k$ denotes the output generated by the computational decoder. The output from the computational decoder can encompass a target image, a classification vector, or even a segmentation map, among other possibilities that are contemplated in the cost function $\mathcal{L}_{task}$. The terms $\mathcal{R}_\rho(\phi)$ and $\mathcal{R}_\sigma(\theta)$ weighted by ρ and σ controls the

regularization on optical parameters and network parameters, respectively. Regularization $\mathcal{R}_\sigma(\theta)$ over the network parameters is commonly used to train deep neural networks and is effective for overfitting and training stability. Notably, ℓ_1 and ℓ_2 norm regularization have found success in regularizing the computational decoder, as discussed in Bacca et al. (2021). On the other hand, the regularization of optical parameters via $\mathcal{R}_\rho$ serves a distinct purpose compared to the regularization of network parameters. This distinction arises from the direct association of optical parameters with the values and/or parameters in the optical acquisition system. Consequently, promoting desired properties in optical parameters is a valuable practice. For instance, optimization of binary weights is desired for the design of binary coded apertures in compressive image classification and reconstruction architectures Bacca et al. (2020). The central concept behind incorporating regularization into training can be evidenced by the analysis of the gradient of the cost function, which involves utilizing the chain rule as follows

$$\frac{\partial \mathcal{L}}{\partial \phi} = \frac{\partial \mathcal{L}_{task}}{\partial \mathcal{N}_\theta} \frac{\partial \mathcal{N}_\theta}{\partial \mathcal{A}_\phi} \frac{\partial \mathcal{A}_\phi}{\partial \phi} + \rho \frac{\partial \mathcal{R}_\rho}{\partial \phi}. \tag{11}$$

Consequently, the design of optical elements is directly impacted by the cost function associated with $\mathcal{L}_{task}$ and the regularization function $\mathcal{R}_\rho$. This influence extends to the inclusion of physical constraints on the coded apertures utilized in the acquisition process. The specific properties of these constraints can be tailored to align with the structure of the regularization function $\mathcal{R}_\rho$. Additionally, the weighted parameter (referred to as ρ) plays

a pivotal role in striking an optimal balance between task performance and the desired properties imposed by the regularization, as detailed in Bacca et al. (2021). However, once the optical parameters have been established through the training step, they remain fixed during subsequent captures. This inflexibility can limit the performance of DOCD-based systems when utilized with scenes significantly different from those found during training.

5. Hadamard row-wise generation algorithm

5.1. Algorithm.

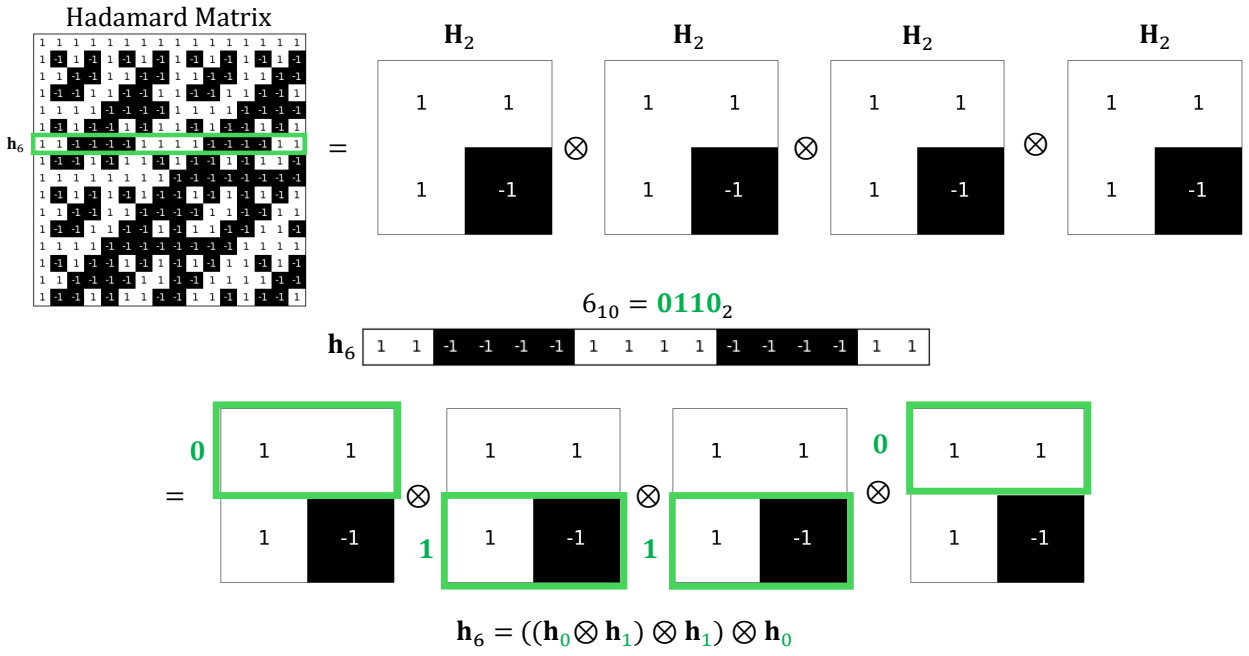
Generating the i-th row $\mathbf{h}_i$ of a Hadamard matrix $\mathbf{H}$ of order 2^n typically requires pre-computing the entire matrix. This can be memory-intensive, especially when 2^n is large. However, when capturing single pixel measurements sequentially, only one Hadamard row is required at a time as per the predefined acquisition protocol. To address this, a row-wise generation algorithm is developed to calculate the specific coefficients of the i-th row without the need to generate the whole matrix Monroy and Bacca (2024). Specifically, following Sylvester's construction, a Hadamard matrix of order 2^n can be constructed from the base matrix of 2-order and Kronecker products as follows

$$\mathbf{H}_{2^n} = \begin{bmatrix} \mathbf{H}_{2^{n-1}} & \mathbf{H}_{2^{n-1}} \\ \mathbf{H}_{2^{n-1}} & -\mathbf{H}_{2^{n-1}} \end{bmatrix} = \mathbf{H}_2 \otimes \mathbf{H}_{2^{n-1}}, \tag{12}$$

with $2 \leq n \in \mathbb{N}$ where $\otimes$ denotes the Kronecker product.

Analyzing the above equation reveals that each row of the Hadamard matrix of order 2^n can be expressed as the Kronecker product between the first or the second row of the Hadamard matrix of order 2. In particular, the sequence of Kronecker products is derived from the binary representation $i_{10} = (b_n b_{n-1} \dots b_1 b_0)_2$ of the i-th row as follows

Figure 6. Hadamard Row-Wise Generation Algorithm. In the case of $\mathbf{h}_6$, the 6th index has a binary representation of $\mathbf{0110}_2$. The digits in this binary representation can be used to index the Kronecker product of n, 2-order Hadamard matrices, where 0/1 corresponds to using the first or second row, respectively.



$$\mathbf{h}_i = \bigotimes_{k=n}^{0} \mathbf{h}_{b_k} = \mathbf{h}_{b_n} \otimes \mathbf{h}_{b_{n-1}} \otimes \cdots \otimes \mathbf{h}_{b_1} \otimes \mathbf{h}_{b_0}. \tag{13}$$

In this sense, it is possible to develop an algorithm for the sequential computation of each of the rows of the Hadamard matrix, without the need to calculate the complete matrix, decreasing computational resources in terms of memory and allowing the acquisition of measurements for the single pixel acquisition system for higher-order matrices as presented in Figure 6. The Hadamard Row Wise Generation is summarized in Algorithm 1. Line 1 sets the base 2-order Hadamard matrix, $\mathbf{H}_2$. Line 2 maps the specified i-th index to its binary representation (i.e., $6 \rightarrow$ 0b0110) and initializes the Hadamard row $\mathbf{h}_i$ as a scalar 1.

Subsequently, lines 4-7 consist of an iterative loop over the digits of the binary representation of the i-th index from left to right. This is done to select either the first or second row of the 2-order Hadamard matrix. Finally, the Hadamard row, $\mathbf{h}_i$, is constructed by the cumulative Kronecker product between the currently selected Hadamard row and the previous output.

Algorithm 1 Hadamard Row Wise Generation

Input: i, n
Output: $\mathbf{h}_i$
1: Compute $\mathbf{H}_2$
2: Set $indxs = \texttt{dec2bin}(i, n)$
3: Set $\mathbf{h}_i = 1$
4: **for** $j = 0$ to $\texttt{length}(indxs)$ **do**
5: Set $\mathbf{h}_{temp} = \mathbf{H}_2[indxs[j], :]$
6: Set $\mathbf{h}_i = \texttt{kron}(\mathbf{h}_i, \mathbf{h}_{temp})$
7: **end for**
8: **return** $\mathbf{h}_i$

It is important to highlight that although the specific idea of generating a row of a Hadamard matrix directly, without constructing the entire matrix, is not as commonly discussed in the literature as the general construction of Hadamard matrices themselves, it is a natural extension of the recursive nature of Hadamard matrices proposed in Sylvester (1867). In this context, the proposed algorithm offers a more detailed documentation of this strategy. Furthermore, the algorithm can be easily adapted to other Hadamard ordering strategies, as it primarily depends on permuting the ordering indices. The code implementation for generating the Hadamard matrix row by row is available on GitHub. `https://github.com/bemc22/hadamard-spc`.

5.2. Computational complexity.

The computational complexity of Algorithm 1 relies on the complexity in compute n times Kronecker product of vector of size 2. In these sense, for each Kronecker product between the vector of size 2^{k-1} (resulting from the previous $k-1$ Kronecker products) and a vector of size 2, the number of multiplication required is $2 \times 2^{k-1} = 2^k$. Thus, the total computational complexity $\mathcal{C}(n)$ for performing n Kronecker products is the sum of the number of multiplications for each step

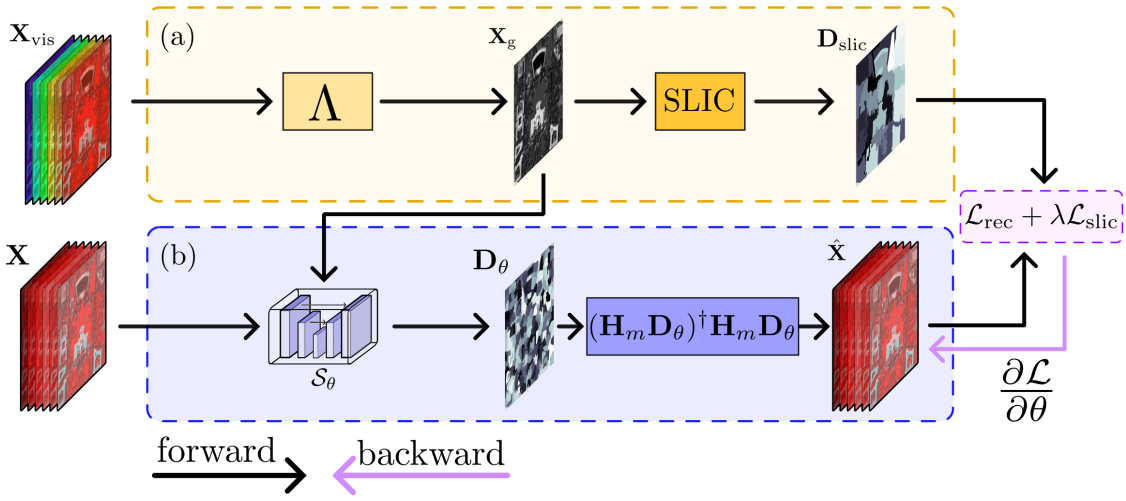
$$\mathcal{C}(n) = 2^1 + 2^2 + 2^3 + \cdots + 2^n = \sum_{k=1}^{n} 2^k \tag{14}$$

which consists of a geometric series that can be simplified as $\mathcal{C}(n) = 2^{n+1} - 2$, with the dominant term in $\mathcal{C}(n)$ being 2^{n+1}, so the asymptotic computational complexity of Algorithm 1 is $\mathcal{O}(2^{n+1}) \sim \mathcal{O}(2^n)$. On the other hand, an alternative to generating a specific row consists of applying the fast Hadamard transform (FHT) to the one-hot encoding of a specific index i-th. For a vector of size 2^n, the FHT algorithm has a computational complexity of $\mathcal{O}(n \times 2^n)$, which is higher by a factor of n compared to the presented row-wise generation algorithm. Moreover, the construction of a Hadamard row can be optimized beyond scalar multiplication by employing sign flips or evaluations of XOR gates, given that Hadamard entries are binary values.

6. Hadamard Single Pixel by Adaptive Superpixels

This chapter introduces an adaptive methodology that uses superpixels for the sensing and recovery of spectral images. The proposed approach employs a deep neural network to design an adaptive decimation matrix, $\mathbf{D}$, guided by side information from the visible (VIS) spectrum, while targeting the near-infrared (NIR) region. Specifically, the framework uses a state-of-the-art fully convolutional neural network, SpixelNet, trained in an unsupervised manner to generate superpixels, which are then used to design $\mathbf{D}$. The adaptive decimation matrix is optimized end-to-end (E2E) within the deep neural network framework, ensuring that the structure of the sensing matrix $\mathbf{HD}$ adapts to each spectral image based on its side information VIS, as presented in Figure 7.

Figure 7. Schematic diagram of the proposed methodology; (a) a grayscale image $\mathbf{X}_g$ in the visible spectrum is aquired given sensor response $\mathbf{\Lambda}$, (b) $\mathbf{X}_g$ is used to adaptively design the decimation matrix $\mathbf{D}_\theta$ by the deep learning model $\mathcal{S}_\theta$.



6.1. Method

Acquisition protocol and Adaptive Model. The proposed adaptive acquisition protocol based on superpixels consists of the adaptive design of the decimation matrix $\mathbf{D}$ using the side information acquired from the VIS region considering the NIR region by its guidance during the jointly trained sensing and recovery of the spectral image. Both stages are integrated into a deep neural model, where the trainable parameters are adjusted through end-to-end (E2E) optimization. For this, a grayscale image $\mathbf{X}_g$ serves as input to the network $\mathcal{S}_\theta$ to adaptively design the decimation matrix $\mathbf{D}$. The neural network consists of a state-of-the-art fully convolutonal neural network, Spixelnet Yang et al. (2020), which is trained in an unsupervised scheme. Once the network parameters θ are adjusted, the adaptive estimation of the decimation matrix $\mathbf{D}$ can be mathematically described as

$$\begin{aligned} \mathbf{F}_\theta &= \mathcal{S}_\theta(\mathbf{X}_g), \\ \mathbf{D}_\theta &= \texttt{binary-argmax}(\mathbf{F}_\theta) \end{aligned} \tag{15}$$

where $\texttt{binary-argmax}(\cdot)$ is defined as follows. For a given matrix $\mathbf{F}_\theta \in \mathbb{R}^{m\times n}$, the function $\texttt{binary-argmax}(\mathbf{F}_\theta)$ returns a binary matrix $\mathbf{D}_\theta \in \{0,1\}^{m\times n}$ such that

$$\mathbf{D}_{\theta,ij} = \begin{cases} 1 & \text{if } \mathbf{F}_{\theta,ij} = \max_k(\mathbf{F}_{\theta,ik}) \text{ for } k = 1,\ldots,n \\ 0 & \text{otherwise.} \end{cases} \tag{16}$$

Here, $\texttt{binary-argmax}(\cdot)$ is applied along the second dimension of $\mathbf{F}_\theta$. Following this,

the compressed projection of the spectral image is obtained using the decimation matrix $\mathbf{D}\theta$ as

$$\mathbf{Y}_{NIR} = \mathbf{H}\mathbf{D}_\theta\mathbf{X}_{NIR}. \tag{17}$$

Notice that, since the decimation matrix depends on $\mathbf{X}_g$, the structure of the sensing matrix $\mathbf{H}\mathbf{D}_\theta$ is adapted for each spectral image based on its side information. Furthermore, based on the orthogonal properties of the Hadamard matrix, spetral image acquisition and recovery can be expressed as in a fully differentiable oprator $\mathcal{D}(\cdot)$ as

$$\hat{\mathbf{X}}_{NIR} = \mathcal{D}(\mathbf{X}, \mathbf{D}_\theta) = \text{diag}\Big(\frac{1}{\mathbf{D}_\theta\mathbf{1}}\Big)\mathbf{D}_\theta^\top\mathbf{D}_\theta\mathbf{X}_{NIR}. \tag{18}$$

Network Architecture. The architecture of the adaptive superpixel estimation model $\mathcal{S}_\theta(\cdot)$ is based on the superpixel network presented by Yang et al. (2020). This network is structured as a fully convolutional network (FCN) with a U-net encoder-decoder style. The input to the network consists in the grayscale image $\mathbf{X}_g$, which is processed through a sequence of convolutional layers designed to extract relevant features that aid in the accurate assignment of pixels to superpixels. The network employs an encoder-decoder architecture with integrated skip connections to ensure that critical spatial information is preserved throughout the feature extraction and reconstruction process.

The encoder reduces the spatial resolution of the input image while increasing the depth of its feature representations using multiple convolutional layers followed by Leaky ReLU activations. The encoder compresses spatial information to extract high-level features

that describe the structure of the image. Skip connections transfer high-resolution features from the encoder to the decoder to preserve fine spatial details crucial for defining superpixel boundaries.

In the final stage of the network, a softmax activation function with nine classes is applied to generate a pixel-superpixel association map. This map represents the probabilistic assignment of each pixel to a neighboring superpixel center, with the highest probability determining the final assignment. Each of the nine classes corresponds to a specific pixel neighbor in a 3×3 grid, ranging from top left to bottom right. This approach ensures that the resulting superpixels are both compact and well-aligned with image boundaries. The code implementation of the adaptive deep learning model designed based on superpixels is available on GitHub `https://github.com/bemc22/AdaHSI`.

Training procedure. To adjust the adaptive estimation model $\mathcal{S}_\theta$ the whole methodology is implemented following the E2E scheme where the outputs of the neural network are regularized following a SLIC-based function $\mathcal{L}_{slic}$ and the simulated non-iterative of the spectral image. The $\mathcal{L}_{\text{slic}}$ loss consists of two terms, $\mathcal{L}_{\text{sem}}$ term which aims to constraint deep-superpixels $\mathbf{D}_\theta$ in a given super-pixel map $\mathbf{D}_{\text{slic}} \in \mathbb{R}^{s \times n}$ and a second term which enforces the superpixels to be spatially compact, as suggested by authors in Yang et al. (2020). For this,the SLIC algorithm is employed Achanta et al. (2012) over $\mathbf{X}_{\text{g}}$, i.e., $\mathbf{D}_{\text{slic}} = \text{SL IC}(\mathbf{X}_{\text{g}})$. SLIC is chosen due to its low computational complexity $\mathcal{O}(n)$, also, this algorithm is able to control over the number of superpixels, unlike other superpixels algorithms. Therefore, the

SLIC-based loss function can be defined as

$$\begin{aligned}\mathcal{L}_{\text{slic}} &= \mathcal{L}_{\text{sem}} + \tau_1 \|\mathbf{p} - \mathcal{D}(\mathbf{p}; \mathbf{D}_\theta)\|_2^2, \\ \mathcal{L}_{\text{sem}} &= \mathcal{L}_{\text{cross}}(\mathbf{F}_\theta^\dagger \mathbf{F}_\theta \mathbf{D}_{\text{slic}}^\top,\ \mathbf{D}_{\text{slic}}^\top),\end{aligned} \tag{19}$$

where $\mathbf{p} = [\mathbf{c}_x, \mathbf{c}_y]$, i.e, $\mathbf{X}_{\mathbf{c}_x^{(i)}(1+\mathbf{c}_y^{(j)})}$ represents a pixel position by its image coordinates and $\mathbf{F}_\theta^\dagger$ is the psudo-inverse of $\mathbf{F}_\theta$. Based on Yang et al. (2020), the cross-entropy loss $\mathcal{L}_{\text{cross}}$ is used as the distance measure. Additionally, to address the design of the decimation matrix $\mathbf{D}_\theta$ for the recovery the infrared spectral image, the reconstruction loss from the non-iterative estimation is incorporated, defined as follows

$$\mathcal{L}_{\text{rec}} = \|\mathcal{D}(\mathbf{X}; \mathbf{D}_\theta) - \mathbf{X}\|_1. \tag{20}$$

Therefore, the complete objective function for finding the optimal set of network parameters θ^* is given by

$$\theta^* = \arg\min_\theta \quad \mathbb{E}_{\mathbf{X}}[\mathcal{L}_{\text{rec}} + \lambda \mathcal{L}_{\text{slic}}]. \tag{21}$$

Notice that the infrared information is only considered in the training step through the back-propagation of the recovery error. In contrast, the visible information is used to adaptively guide the training procedure, i.e., is the input for the SLIC-based superpixels.

6.2. Experiments

First, the main hyperparameters of the proposed methodology are evaluated, specifically the number of superpixels in the decimation matrix and the SLIC-based loss, i.e., m and s, respectively. The training configuration consists of 5000 epochs with nine batch-size and 10^{-3} learning rate with the Adam optimizer in the Washington D.C. Mall (WDCM) dataset. This dataset is composed of a high-resolution spectral image which is resized to 1200×360 to extract 30 random patches for the patch-based dataset. The dataset is random split into 27 and 3 images for training and testing, respectively (Several random split trials are carry out until achieved a test set which contains the global information of training set) following the methodology in Zhu et al. (2014). During training, a data augmentation (random horizontal and vertical flip) was used to avoid overfitting. The quality of recovery SIs was evaluated using the peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM), and spectral angle mapper (SAM) metrics. Table 1 shows the estimation of the infrared image using 225 and 400 m-ord Hadamard matrix, namely $m = \{225, 400\}$ and a varying of $s = \{16, ..., 48\}$ associated to a s-ord Hadamard matrix, which is the number of superpixels in the SLIC. Then it can see that, for both values of m, the best performance is achieved using around $s = 32$ SLIC superpixels in the evaluated dataset.

Table 1
Experimental setup for different values of m-ord Hadamard matrix and s SLIC superpixels. Best and second best results highlighted in ***bold*** *and* <u>underline</u> *respectively.*

Metric	s	m 16	24	28	32	36	40	48
PSNR (↑)	225	33.11	<u>33.51</u>	33.42	**33.56**	33.17	33.07	33.00
	400	33.40	33.62	**34.00**	<u>33.83</u>	33.56	33.73	33.40
SSIM (↑)	225	94.41	<u>94.91</u>	94.83	**95.00**	94.50	94.40	94.28
	400	94.82	95.08	**95.52**	<u>95.31</u>	95.01	95.21	94.81
SAM (↓)	225	0.133	<u>0.124</u>	0.126	**0.125**	0.129	0.129	0.127
	400	0.116	0.119	**0.112**	<u>0.116</u>	0.119	0.117	0.119

Table 2
Ablation settings. Each row corresponds to removing a methodology component listed in the w/o column; the This work case means using the full proposed methodology.

m	w/o	PSNR (↑)	SSIM (↑)	SAM (↓)
225	$\mathcal{L}_{\text{slic}}$	33.27	94.63	0.129
	$\mathcal{L}_{\text{rec}}$	31.25	91.25	0.144
	DA	32.22	93.09	0.144
	This work	**33.56**	**95.00**	**0.125**
400	$\mathcal{L}_{\text{slic}}$	33.47	94.91	0.116
	$\mathcal{L}_{\text{rec}}$	32.05	92.82	0.127
	DA	32.44	93.47	0.131
	This work	**33.83**	**95.31**	**0.116**

Ablation study

In this section, the influence of the two main loss functions is evaluated, $\mathcal{L}_{\text{slic}}$ which leads to the learning of superpixels and $\mathcal{L}_{\text{rec}}$ which leads the training to achieve optimal performance in the recovery of the spectral image, and the influence of incorporate data augmentation (DA) by removing each of them. Note in Table 2, the most relevant component of the proposed methodology is $\mathcal{L}_{\text{rec}}$, which is consistent since this term directly measures

the error of the estimated decimation matrix for spectral image recovery.

Table 3
Comparison results, best and second best results highlighted in ***bold*** *and* <u>underline</u> *respectively.*

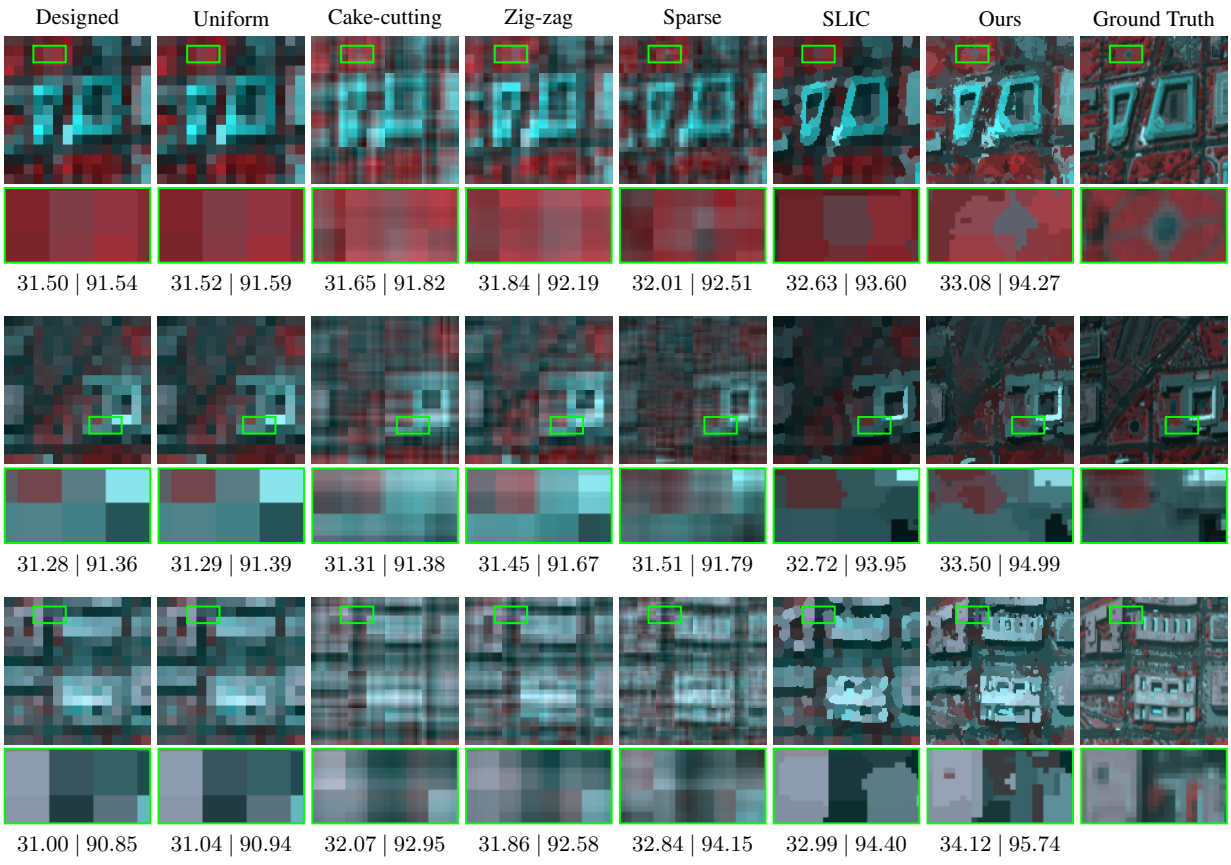
m	Method	PSNR (↑)	SSIM (↑)	SAM (↓)
225	Designed	31.25	91.25	0.144
	Uniform	31.28	91.30	0.143
	Cake-cutting	31.67	92.05	0.148
	Zig-zag	31.71	92.14	0.143
	Sparse	32.11	92.81	0.153
	SLIC	<u>32.77</u>	<u>93.98</u>	<u>0.135</u>
	This work	**33.56**	**95.00**	**0.125**
400	Designed	32.13	92.96	0.125
	Uniform	32.22	93.12	0.124
	Cake-cutting	32.42	93.37	0.137
	Zig-zag	32.79	93.95	0.126
	Sparse	33.02	94.22	0.142
	SLIC	<u>33.67</u>	<u>95.14</u>	<u>0.120</u>
	This work	**33.83**	**95.31**	**0.116**

Comparison with state-of-art sensing design

Finally, the proposed methodology is compared with state-of-the-art HSI-based sensing methodologies under the same level of compression. The comparison considers the uniform spatial decimation design, where no side information about the scene is available Bacca et al. (2022). Designed approach Arguello et al. (2023), which consists in optimizing a single decimation matrix for the whole dataset. Ordering approaches include cake-cutting Yu (2019) and zig-zag Zhang et al. (2017b) for choosing the first m-rows of a n-ord reordered Hadamard matrix. A sparse method that selects the most relevant coefficients from the Hadamard spectrum of the grayscale acquisition $\mathbf{X}_{\mathrm{g}}$ as modulation patterns to acquire the

spectral image Candès and Wakin (2008), and SLIC approach Garcia et al. (2020), which consists of directly employing the SLIC-based decimation matrix of m superpixels. Quantitative results are shown in Table 3, where it can see how the proposed method outperforms state-of-the-art designs by up 0.79 dB in PSNR for the case of the 225-ord Hadamard matrix. Additionally, Fig. 8 shows visual results where the proposed adaptive design correctly preserves the structural information of the infrared spectral image for the WDCM dataset.

Figure 8. Visual comparison with state-of-art sensing design, $m = 225$. Each column corresponds to a false color visualization of the spectral image for each sensing design; the last column corresponds to the reference spectral image. The PSNR and SSIM scores are displayed under each sensing design.



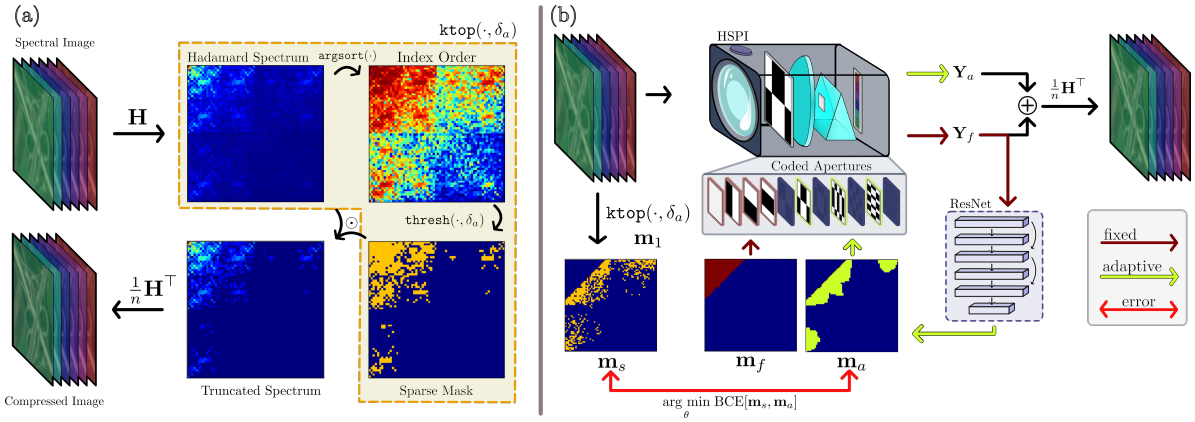
7. Hadamard Single Pixel by Adaptive Ordering

This chapter presents an adaptive ordering approach for the Hadamard basis in Hyperspectral Single-Pixel Imaging (HSPI), integrating both fixed and adaptive sensing stages to enhance image acquisition. Initially, a subset of Hadamard coefficients is acquired using state-of-the-art ordering algorithms in the fixed sensing stage. This initial data is then utilized by a deep neural network to adaptively estimate the ordering of the remaining coefficients, thereby selecting a second subset of modulation patterns. The proposed acquisition protocol is designed to optimize the overall sampling rate by combining both fixed and adaptively selected rows. The adaptive sensing stage employs the coarse features obtained from the fixed sensing to guide the selection of the most relevant coefficients in the unacquired spectrum, enhancing the image reconstruction process. The optimization of the adaptive estimation model is formulated as a binary classification problem, where the model predicts the adaptive sensing mask based on the initial acquisition. Finally, the model is trained using an end-to-end framework to minimize the binary cross-entropy loss, ensuring the effective reconstruction of spectral images.

7.1. Method

The proposed adaptive ordering of the Hadamard basis divides the HSPI imaging system into sequential sensing stages: i) fixed sensing and ii) adaptive sensing. In the fixed sensing stage, an initial subset of the Hadamard spectrum is acquired following a predefined sensing path established by state-of-the-art Hadamard ordering algorithms. Subsequently, in

Figure 9. Adaptive Hadamard Single Pixel Imaging. (a) Construction of reference sensing mask $\mathbf{m}_s$: the reference sensing mask $\mathbf{m}_s$ are constructed by selection of the δ_a-top coefficients in the Hadamard spectrum. (b) Proposed adaptive acquisition protocol, a fixed sensing acquisition $\mathbf{Y}_f$ based on sensing mask $\mathbf{m}_f$ guides the selection of adaptive sensing mask $\mathbf{m}_a$.



the adaptive sensing stage, the results from the initial sensing serve as inputs for a deep neural network, which adaptively estimates the ordering of the non-navigated coefficients to select a second subset of modulation patterns in an end-to-end manner. The following subsection explores the details of the acquisition protocol, the architecture of the deep neural network, and the training procedure.

Adquisition protocol. The acquisition protocol entails obtaining the Hadamard coefficients from the adaptive row selection matrix $\mathbf{M} = \text{diag}(\mathbf{m}_f + \mathbf{m}_a)$, where $\mathbf{m}_f$ represents the fixed selection rows, and $\mathbf{m}_a$ corresponds to the adaptively selected rows, as illusted in Figure 9. The overall sampling rate is defined as $\delta = (\delta_f + \delta_a)/n$, where $\delta_f = \|\mathbf{m}_f\|_0$ and $\delta_a = \|\mathbf{m}_a\|_0$ denote the quantities of fixed and adaptive selected rows, respectively. For convenience in the formulation, the binary selection matrix is denoted as $\mathbf{M} = \mathbf{P}^\top \mathbf{S} \mathbf{P}$ based on the model in Equation (8). Since the permutation of the Hadamard matrix where the

binary mask is fixed is equal to the permutation of the binary mask where the Hadamard matrix is fixed, both notations are equivalent.

Fixed Sensing. The fixed sensing step consists of acquiring a subset of the Hadamard matrix using a consistent approach across all images, following the current state-of-the-art sensing method (typically Zig-Zag López-García et al. (2022)). It is worth noting that statistical differences in the ordering of coefficients are assumed to be negligible at this stage. Mathematically, this process can be represented as

$$\mathbf{Y}_f = \mathbf{M}_f \mathbf{H} \mathbf{X}, \tag{22}$$

where $\mathbf{M}_f = \text{diag}(\mathbf{m}_f)$ is the fixed selection matrix, which selects the δ_f-th rows of the Hadamard matrix based on a predefined Hadamard ordering algorithm.

Adaptive Hadamard Ordering. Following the initial acquisition $\mathbf{Y}_f$, a deep neural model $\mathcal{P}_\theta$ estimates the ordering of the most relevant coefficients in the no-acquired spectrum. Mathematically, this estimation is represented as:

$$\mathbf{m}_a = \mathcal{P}_\theta(\mathbf{Y}_f), \tag{23}$$

Using this approach, the selection of a second Hadamard coefficient ordering is induced from the first acquisition ($\mathbf{Y}_f$). The presence of coarse features within the spectral images enables us to adaptively estimate the subsequent Hadamard coefficients. This is possible because

initial acquisition provides a rough approximation of spectral content, allowing the model to predict which coefficients will most likely contain significant information in the unacquired spectrum.

Adaptive Sensing. The adaptive sensing consists of acquiring the remaining coefficients based on the previously estimated ordering. This acquisition process can be formulated as

$$\mathbf{Y}_a = \mathbf{M}_a \mathbf{H} \mathbf{X}. \tag{24}$$

Subsequently, as both fixed and adaptive acquisitions are associated with the same Hadamard basis, and the captured coefficients are mutually exclusive, the overall recovery can be carried out from both acquisitions, as defined in Equation (9), in the following manner,

$$\hat{\mathbf{X}} = \frac{1}{n} \mathbf{H}^\top (\mathbf{Y}_f + \mathbf{Y}_a). \tag{25}$$

Network Architecture. In the approach to neural network design, a standard model is used to dynamically estimate the Hadamard coefficients. This choice highlights the importance of the adaptive methodology while avoiding the computational overhead associated with more complex neural networks. The deep neural network architecture is centered around a ResNet-based backbone, incorporating 5 residual blocks alongside batch normalization for enhanced stability. Starting with an initial feature size of 64, the architecture progressively expands through the hidden layers. To effectively reduce spatial dimensions, max-pooling operators are strategically inserted between the residual blocks.

Post extraction of activation features from the ResNet backbone, they are used as input into a fully connected head comprising 3 layers. Within this head, the final transformation is applied to the selection matrix vector $\mathbf{m}_a$, where the last activations consist of a sigmoid activation. Throughout the training phase, dropout regularizers are introduced between fully connected layers, with a dropout probability parameter set at 90 % to mitigate overfitting risks. The Rectified Linear Unit (ReLU) serves as the chosen nonlinear activation function throughout the architecture of the neural network. In particular, the adaptive estimation model boasts approximately 5.5 million trainable parameters. The code implementation of adaptive deep learning designed based on Hadamard matrix ordering is available on GitHub `https://github.com/bemc22/adaptive-ordering-spc`.

Training procedure.

To address the optimization of the adaptive estimation model, denoted as $\mathcal{P}_\theta$, the E2E framework is employed, encompassing the joint modeling of the two sensing acquisitions and the reconstruction of spectral images within a network architecture, as illustrated in Figure 9. In essence, the final goal is to optimize the trainable parameters, represented by θ, within the adaptive estimation model $\mathcal{P}_\theta$. This optimization process aims to minimize a defined cost function, denoted $\mathcal{L}(\cdot)$, in the training data set. This training optimization problem can be formally expressed as follows.

$$\theta^* \in \arg\min_{\theta} \quad \mathbb{E}_{\mathbf{X}}[\mathcal{L}(\mathbf{X}, \hat{\mathbf{X}}; \mathcal{P}_\theta)]. \tag{26}$$

Adaptive Sensing as Binary Classification.

To address the selection of modulation patterns with the adaptive estimation model, denoted $\mathcal{P}_\theta$, the problem of selecting the most relevant coefficients is modeled as a binary classification problem. Each spectral image can be represented in the Hadamard spectrum as a linear combination of vector basis elements. $\mathbf{h}_k$ and intensities $\mathbf{y}_k$ as follows

$$\mathbf{X} = \sum_{k=1}^{n} \mathbf{h}_k \otimes \mathbf{y}_k, \tag{27}$$

where $\otimes$ denotes the outer product. We propose splitting the Hadamard coefficients into three classes: i) the fixed coefficients, acquired in the first step of the scheme, ii) the top δ_a Hadamard coefficients by magnitude, which serve as the ground-truth labels learned by the adaptive selection model, and iii) the least relevant coefficients. Correspondingly, the modulation patterns are divided into $\{\hat{\mathbf{h}}_i\}_{i=1}^{\delta_f}$ for the fixed patterns, $\{\hat{\mathbf{h}}_j\}_{j=\delta_f+1}^{\delta_a+\delta_f}$ for the adaptive patterns, and $\{\hat{\mathbf{h}}_k\}_{k=\delta_a+\delta_f}^{n}$ for the unused patterns, allowing the spectral image to be represented as follows

$$\mathbf{X} = \underbrace{\sum_{i=1}^{\delta_f} \hat{\mathbf{h}}_i \otimes \mathbf{y}_i}_{fixed} + \underbrace{\sum_{j=\delta_f+1}^{\delta_a+\delta_f} \hat{\mathbf{h}}_j \otimes \mathbf{y}_j}_{most\ relevant} + \underbrace{\sum_{k=\delta_a+\delta_f+1}^{n} \hat{\mathbf{h}}_k \otimes \mathbf{y}_k}_{less\ relevant}. \tag{28}$$

Hence, the binary classification mask can be computed as position-threshold in the ordered Hadamard spectrum $\mathbf{m}_s = \texttt{ktop}(\mathbf{M}_f^r|\mathbf{Y}|\Sigma, \delta_a)$, discarding the already sensed coefficients in the fixed sensing by the reverse mask $\mathbf{M}_f^r$, as presented in Figure 9(b). Finally, the recons-

truction optimization problem in Equation (26) can reformulated as a binary classification problem where the deep model $\mathcal{P}_\theta(\cdot)$ estimates the adaptive sensing mask $\mathbf{m}_a$ from a fixed acquisition $\mathbf{Y}_f$ which are minimized with a binary cross-entropy loss function as follows

$$\theta^* \in \arg\min_{\theta} \mathcal{L} := \mathrm{BCE}[\mathbf{m}_s, \underbrace{\mathcal{P}_\theta(\mathbf{Y}_f)}_{\mathbf{m}_a}]. \tag{29}$$

7.2. Experiments

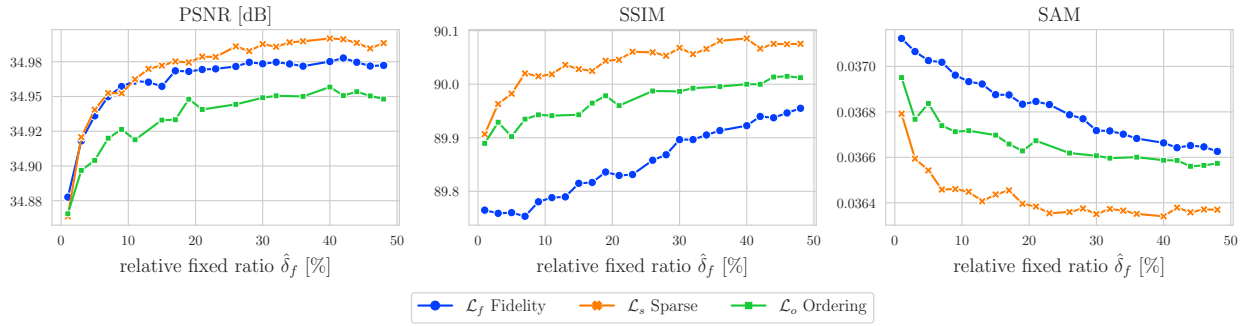
In terms of training configuration, 500 epochs were conducted with an initial learning rate of 0.001, and the model parameters were optimized using the Adam optimizer. The EuroSAT dataset was used, partitioned into a train-test split of 0.9/0.1 for model evaluation. The hyperspectral EuroSAT dataset comprises 27k Sentinel-2 satellite images with a spatial resolution of 64×64, covering 13 spectral bands spanning the visible, near-infrared, and short-wave infrared segments of the spectrum.

Evaluation of loss functions.

In this section, a comprehensive evaluation of the proposed method is presented, considering a range of different training strategies. Specifically, the proposed **Sparse** loss is compared against two alternatives: **Data Fidelity** loss and **Ordering** loss, each of which will be examined in detail.

Data Fidelity Loss. In the data-fidelity loss, the primary objective is to minimize the reconstruction error between the estimated image and the ground truth image while ensures the desired amount of modulation patterns, namely, $\mathbb{E}[\mathbf{m}_2] = \delta_a/n$. To address this,

Figure 10. Performance of different training cost functions for different split sensing configurations, relative fixed ratio is defined as $\hat{\delta}_f = \delta_f/m$.



the DOCD framework proposed double costs function composed of a MSE/MAE error and a transmittance regularization on the sensing mask $\mathbf{m}_2$. This regularization ensures a low transmittance level and, consequently, guarantees the desired compression ratio. In this sense, the loss function for this perspective, denoted as $\mathcal{L}_f$, consists of two terms

$$\mathcal{L}_f(\mathbf{X}, \hat{\mathbf{X}}) = \|\mathbf{X} - \hat{\mathbf{X}}\|_F^2 + \|\mathbb{E}[\mathbf{m}_2] - \delta_a/n\|_2^2. \tag{30}$$

Ordering Loss. The ordering loss consists of an index regression training directly between the estimated measurement matrix $\mathbf{M}_a$ and the indexed sparse ordering matrix $\mathbf{m}_o = \texttt{argsort}(|\mathbf{Y}|\boldsymbol{\Sigma})$. This approach does not constrain a specific transmittance level, as it learns the entry ordering of the coefficients for the spectral images. The loss function for this perspective, denoted as $\mathcal{L}_o$, is based on the mean squared error

$$\mathcal{L}_o(\mathbf{m}_a, \mathbf{m}_o) = \|\mathbf{m}_a - \mathbf{m}_o\|_2^2. \tag{31}$$

The analysis and discussion are based on Figure 10. The performance of the proposed method is evaluated across several metrics. The *fidelity* loss secures the second-best PSNR but performs lower in SSIM and SAM, especially for low δ_f. In contrast, *sparse* loss shows better performance in all scenarios and metrics. The ordering loss places third in the PSNR scores. In particular, **fidelity** requires two optimization terms, which are tricky to tune, while both **sparse** and **ordering** use a single optimization term, satisfying the compression ratio and consistency of fidelity in the Hadamard spectrum. These findings highlight the strengths and limitations of each strategy, with *sparse* being the most effective and chosen for the next simulations.

We explore the influence of the parameter of the relative fixed ratio $\hat{\delta}_f = \delta_f/m$, the ratio of fixed coefficients δ_f to the total sensed coefficients m, in various compression configurations. The δ values range from 10 % to 50 %. Within each compression scenario, δ is evaluated from 1 % to 95 %. Spatial and spectral metrics are shown in Figure 11. The experiments show a trade-off between fixed and adaptive sensing. The optimal value of δ is typically between 50 % and 80 % as the compression ratio increases. A critical threshold occurs when δ exceeds 80 %, where over 80 % of coefficients are from a fixed algorithm, using less than 20 % for adaptive measurements. This balance is crucial for optimal performance. Conversely, when $\hat{\delta}_f$ ranges from 1 % to 20 % for fixed coefficients, overall performance significantly improves. This suggests that more fixed coefficients help distinguish coarse features in spectral images.

Figure 11. Relative performance on PSNR, SSIM and SAM scores under different transmittance setup and relative fixed ratios (color bars) and different relative fixed rations, the relative fixed ratio is defined as $\hat{\delta}_f = \delta_f/m$.

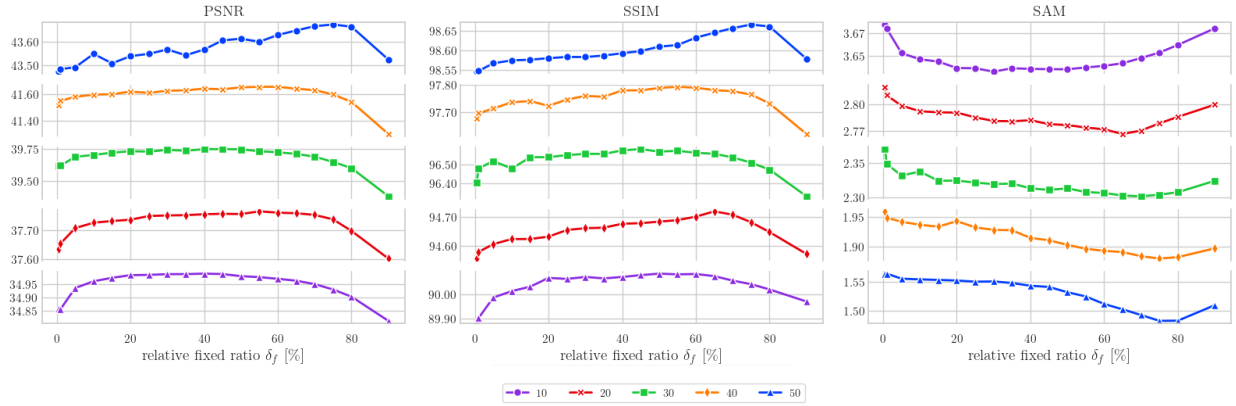
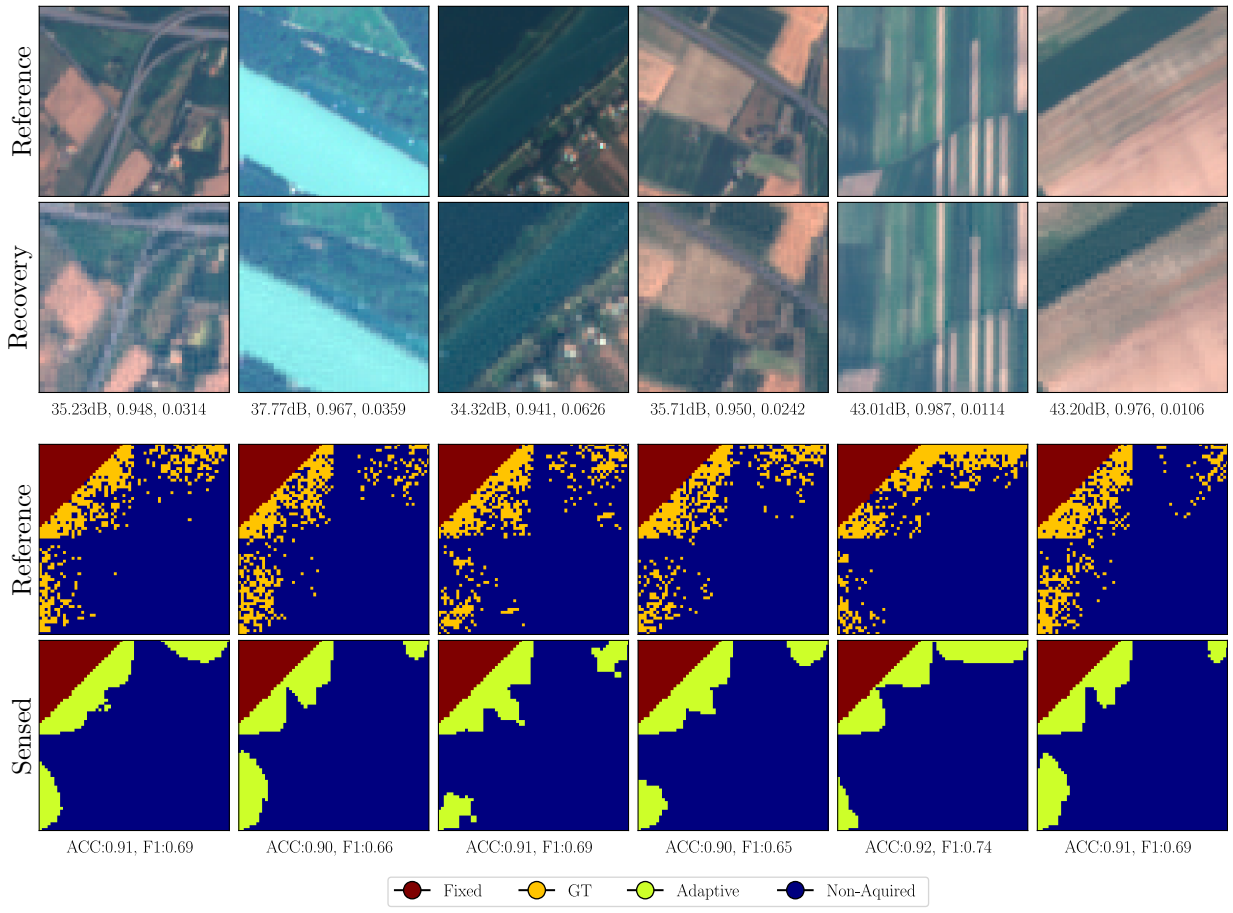


Figure 12, a visual validation of the adaptive behavior of the proposed methodology is presented. Here, reference images and reconstructed images derived from the sparse strategy in the image domain are showcased in Figure 12(a), and the Hadamard spectrum domain is displayed in Figure 12(b). In the context of the reference sparse strategy, the fixed mask $\mathbf{m}_f$, the reference coefficient mask $\mathbf{m}_s$, and the non-acquired coefficients are differentiated by red, orange, and blue, respectively. Consequently, the mask of the adaptive coefficients $\mathbf{m}_a$ is depicted in yellow. The estimated mask of coefficients for each image varies according to the geometry of the reference coefficients, which confirms the adaptive nature of the proposed sensing methodology.

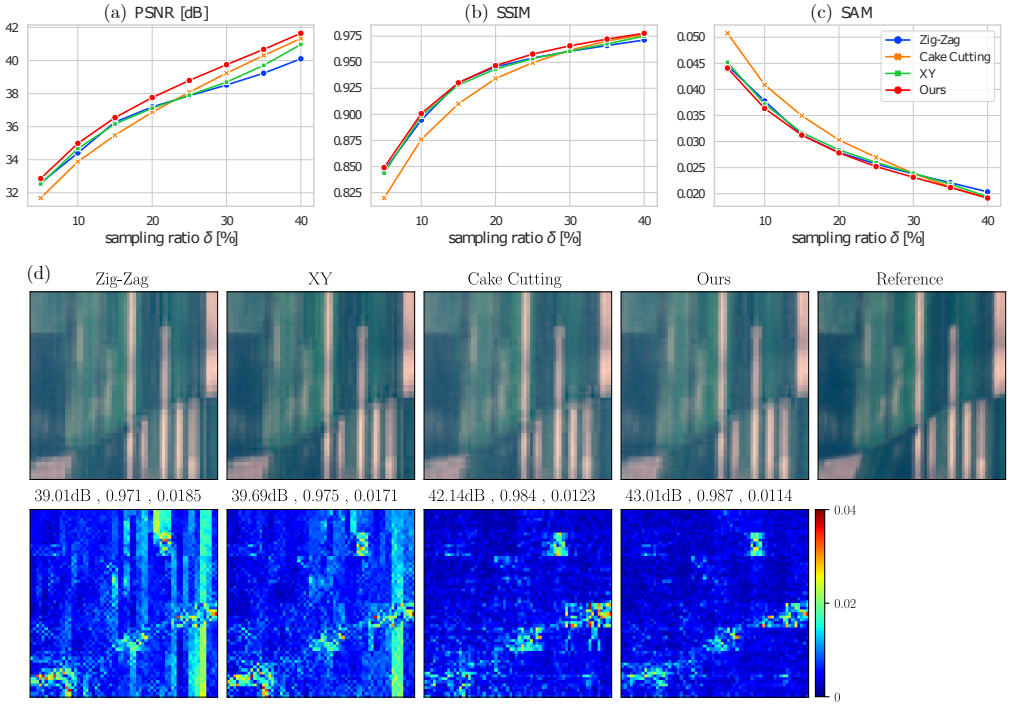
We report accuracy and F1-score between the reference sparse coefficients (orange) and the adaptive coefficients (yellow) as metrics. The simulations reveal approximately 91 % accuracy and 70 % F1 score. The high accuracy may be due to the class imbalance from sampling ratios below 50 %. An F1 score of 70 % indicates satisfactory performance and

Figure 12. Image reconstruction from the proposed adaptive sensing methodology. a) presents reference images and reconstructed images in the natural domain while b) presents the reference sparse mask $\mathbf{m}_s$ and estimated adaptive coefficients $\mathbf{m}_a$.



quality improvement, but also allows for refinement of adaptive Hadamard coefficients. A comparative analysis of coefficients is proposed as a robust benchmark for future adaptive Hadamard sensing research. By examining the F1-score within the spectrum, depending on each dataset's spatial distribution, it is possible to establish thresholds for using adaptive sensing compared to other methods.

Figure 13. Image reconstruction comparison in EUROSAT hyperspectral dataset. (a-c) Presents the comparison in spatial and spectral quality metrics as a function of the sampling ratio.(d) Example of the reconstructed image and error maps using different Hadamard ordering methodologies at the sampling ratio of 25 %.



Comparison with Hadamard Ordering methods

In this section, the effectiveness of the proposed method is evaluated by a comparative analysis with established Hadamard ordering methodologies, namely ZigZag López-García et al. (2022), CakeCutting Yu (2019), and XY Cai et al. (2023). A fixed sampling ratio of 0.4 is maintained for the proposed method. Performance curves are reported in Figure 13(a). It is important to note that different ordering methodologies exhibit varying levels of performance depending on the sampling ratio. For example, the XY and ZigZag ordering methods show superior performance at low sampling ratios, whereas the CakeCutting ordering method yields

better results at higher sampling ratios. In particular, the adaptive methodology consistently outperforms traditional ordering methodologies in all sampling scenarios and metrics. It is important to note that the PSNR metric exhibits greater sensitivity to changes in the ordering strategy, while both the SSIM and the SAM metrics yield comparable results for both the ordering and adaptive methodologies. Furthermore, a significant improvement in PSNR quality is evident when using the proposed adaptive methodology within the sampling ratio range of 20 % to 25 %.

7.3. Real implementation performance analysis

To validate the effectiveness of the proposed adaptive sensing approach, it is imperative to construct an optical testbed capable of facilitating performance analysis in realistic noise scenarios. Note that for near-infrared (NIR) implementations, specialized optical elements optimized for this wavelength range are required. In this context, Figure 14 shows the testbed setup in the HDSP laboratory. It is built with an NIR lamp (3900e-Ilumination technology) that illuminates the sample scene, in which light is redirected by a broadband mirror through the objective lens and the relay lens, employing Thorlabs LB5552 biconvex lenses; it is important to note that a dichroic mirror is employed to filter the NIR spectrum, the setup then constructs an imaging plane on the DMD for the NIR range (Vialux GmbH vd65), enabling the reflection surface to encode the scene. Subsequently, another relay lens, in conjunction with a collimator lens, directs the ray lights to an optical fiber (QP1000-025-VIS-NIR). Finally, the set-up includes a NIRQUEST spectrometer from Ocean Insight, which allows the acquisition of up to 512 spectral bands. Using the NIR single-photon count

(SPC) technique, a series of eight scenes were captured using a complete Hadamard set of patterns, serving as the ground truth for subsequent performance analysis.

Implemented optical system acquisition protocol

In the acquisition protocol, a set of SPC measurements $\{\mathbf{y}_i\}_i^n$ is sequentially obtained, corresponding to a series of modulation patterns $\{\mathbf{h}_i\}_i^n$. In practical implementations, considering that each modulation pattern on the Hadamard basis encompasses values of both negative and positive magnitudes, that is, $\{-1, 1\}$, and the optical light modulation element (DMD) exclusively accepts binary values $\{0, 1\}$, the SPC measurements are obtained by subtracting the *positive* and *negative* sides of the modulation pattern $\mathbf{h}$. Consequently, the model for each Near-Infrared (NIR) spectral response associated with a modulation pattern is expressed as

$$\mathbf{y}_i^\top = (\mathbf{h}_i^+)^\top \mathbf{X} - (\mathbf{h}_i^-)^\top \mathbf{X}, \tag{32}$$

where $\mathbf{h}_i^+$ and $\mathbf{h}_i^-$ represent binary modulation patterns constructed from the positive and negative sides of the Hadamard coefficients $\mathbf{h}_i$, respectively. As each SPC measurement is linked to a specific Hadamard basis vector $\mathbf{h}_i$, real-time reconstruction becomes straightforward through the cumulative addition of the current set of acquisitions. Then, the cumulative estimation of the NIR spectral image in the k-th snapshot is expressed mathematically as

$$\hat{\mathbf{X}}_k = \frac{1}{n} \sum_i^k \mathbf{h}_k \otimes \mathbf{y}_k, \tag{33}$$

where $\otimes$ denotes the outer product. It is noteworthy that, given that optical light modulation occurs along the spatial dimension, this modulation is inherently broadcasted along the spectral dimensions. Therefore, the extension to different numbers of bands is dictated by the specifications of the employed spectrometer.

Figure 14. Photo of the optical system implementation.

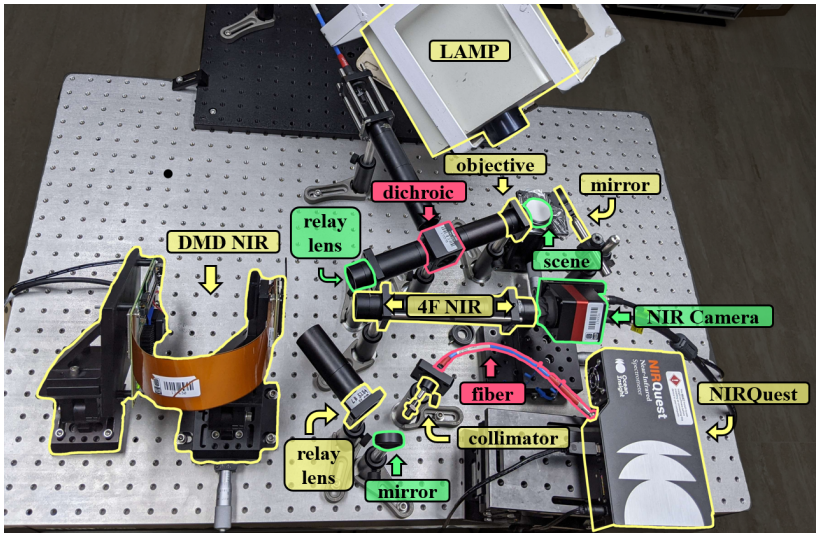
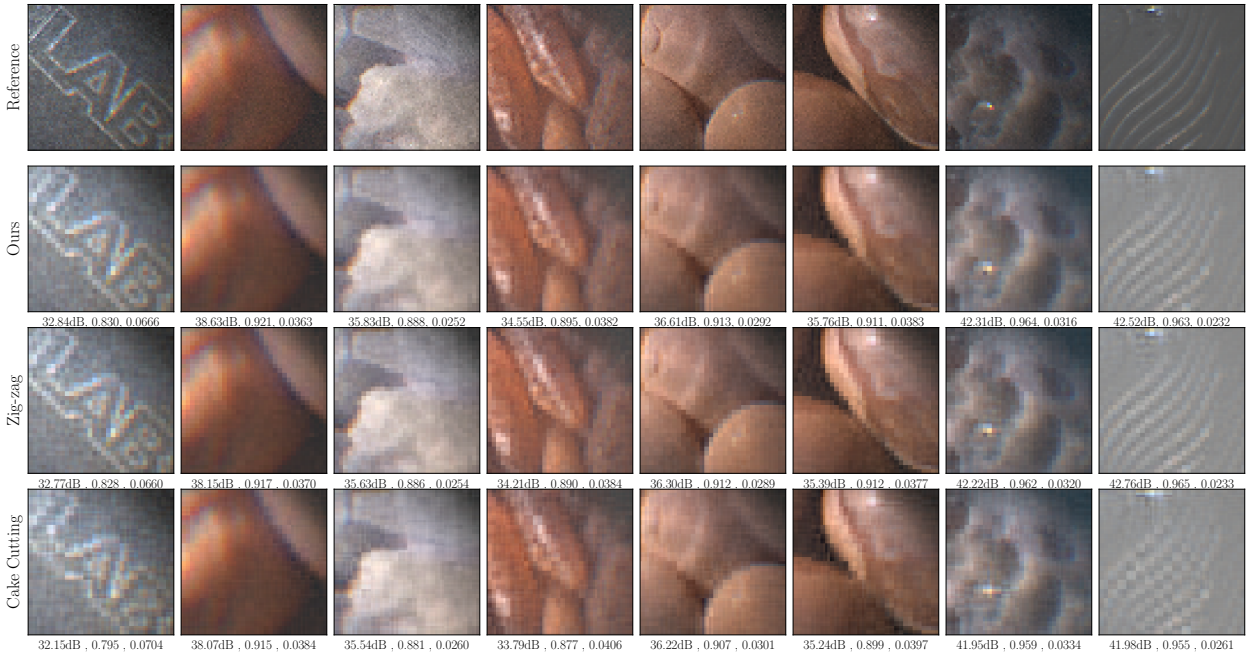


Figure 15. Real implementation of adaptive sensing methodology and ordering methodologies on NIR spectral images. First rows correspond to reference NIR spectral images reconstructed from a full-sensing of the Hadamard Single Pixel, the remaining rows corresponds to the proposed adaptive sensing and Zig-zag and Cake Cutting ordering methods, respectively.



Spectral infrarred image reconstruction from real acquisitions

The NIR spectral images consist of eight scenes displaying various materials, such as metal, different cereals and grains, rocks, and a coin, captured for analysis (Figure 15). Hadamard coefficients are collected for each scene to facilitate the emulation of different Hadamard orderings and the proposed methodology. This amounts to a total of 16384 Hadamard patterns for a spatial resolution of 128×128.

In Figure 15, a comparison is presented between Zig Zag and Cake cutting, which are the most predominant ordering methodologies in the state of the art, and the proposed approach, using a sampling rate of 25 %. The adaptive estimation model of Hadamard

coefficients is applied using model parameters obtained directly from training on the EuroSAT dataset under the same sampling rate configuration, eliminating the need for additional training.

The results indicate that the proposed adaptive methodology outperforms other ordering strategies in around 0.5 dB of the PSNR and 0.5 % of the SSIM for all scenes, with the second-best results alternating between the Zig-Zag and Cake Cutting orderings, mirroring the trends observed in the simulation results. It should be emphasized that in the visual results, the proposed method tends to provide more smooth structures within the objects and more detailed edges such as in the last image in Figure 15. However, there is still an image quality gap that needs to be improved for high-frequency details, as presented in the LAB image. Since these borders are diagonal, the associated Hadamard coefficients are the last to be captured in all ordering methods.

In particular, the proposed adaptive model, trained in spectral images from remote sensing datasets, demonstrates adequate generalization capabilities for spatial information, effectively ensuring acceptable image quality for images obtained from the Optical Laboratory.

8. Conclusions and Dicussion

This work proposes two adaptive deep-learning models for the design of the sensing matrix used in the acquisition and recovery of near-infrared spectral images within the Hadamard single-pixel imaging system. The two models address scenarios with and without access to side information. In cases where side information is available, the estimation of image superpixels is proposed to enhance spatial-structural information in images. For scenarios without side information, an adaptive ordering strategy is suggested, where an initial fixed subset is acquired, and a deep learning model guides the estimation of the subsequent subset.

The adaptive deep learning models demonstrate strong performance in both simulations and real spectral infrared images. A key advantage of the proposed adaptive ordering method is its ability to provide stable performance across a broad range of compression scenarios, unlike existing ordering strategies that tend to excel only in specific cases. Additionally, the conventional end-to-end optical design loss function has been reformulated to incorporate deep learning models that adaptively adjust optical elements based on predefined side information or task-specific requirements.

From the perspective of single-pixel imaging, the image restoration task is redefined as a classification problem in the Hadamard domain. This reformulation simplifies the minimization cost function to a single term that addresses both data fidelity and the desired compression factor, thereby avoiding the computational and numerical challenges associated

with directly learning binary layers. Furthermore, this classification approach facilitates the evaluation of adaptive deep learning models using image classification metrics in the Hadamard domain, providing a valuable benchmark tool for adaptive Hadamard single-pixel strategies and highlighting potential areas for improving current solutions.

Some interesting research directions that can be highlighted from the findings in this work include the strategies presented, where structural information of the scene guides the selection of specific coefficients on the Hadamard spectrum. Furthermore, the inclusion of other image restoration algorithms beyond the image recovery from the ℓ_2 norm could potentially improve image recovery quality.

Another intriguing area of research involves increasing the frequency in adaptive acquisition. While the present work evaluates the case of one fixed and one adaptive estimation, an interesting strategy would be to initialize with a single fixed coefficient and employ an autoregressive estimation of the next coefficient based on previous acquisitions in a one-by-one adaptive estimation, these could be implemented following the state-of-the-art techniques in word estimation from natural lenguage processing field or adaptive agents in the Reinforcement Learning community. However, while these strategies could potentially provide the best image quality restoration and adaptive capabilities, they also pose additional challenges in terms of computational feasibility. Training these models and the computational cost for model inference in real applications would be significantly increased.

References

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., and Süsstrunk, S. (2012). Slic superpixels compared to state-of-the-art superpixel methods. 34(11):2274–2282.

Aloupogianni, E., Ishikawa, M., Ichimura, T., Sasaki, A., Kobayashi, N., and Obi, T. (2021). Design of a hyper-spectral imaging system for gross pathology of pigmented skin lesions. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 3605–3608. IEEE.

Arce, G. R., Brady, D. J., Carin, L., Arguello, H., and Kittle, D. S. (2013). Compressive coded aperture spectral imaging: An introduction. *IEEE Signal Processing Magazine*, 31(1):105–115.

Arguello, H., Bacca, J., Kariyawasam, H., Vargas, E., Marquez, M., Hettiarachchi, R., Garcia, H., Herath, K., Haputhanthri, U., Ahluwalia, B. S., et al. (2023). Deep optical coding design in computational imaging: a data-driven framework. *IEEE Signal Processing Magazine*, 40(2):75–88.

Arguello, P., Lopez, J., Hinojosa, C., and Arguello, H. (2022). Optics lens design for privacy-preserving scene captioning. In *2022 IEEE International Conference on Image Processing (ICIP)*, pages 3551–3555. IEEE.

Averbuch, A., Dekel, S., and Deutsch, S. (2012). Adaptive compressed image sensing using dictionaries. *SIAM Journal on Imaging Sciences*, 5(1):57–89.

Bacca, J., Galvis, L., and Arguello, H. (2020). Coupled deep learning coded aperture design for compressive image classification. *Optics express*, 28(6):8528–8540.

Bacca, J., Gelvez-Barrera, T., and Arguello, H. (2021). Deep coded aperture design: An end-to-end approach for computational imaging tasks. *IEEE Transactions on Computational Imaging*, 7:1148–1160.

Bacca, J., Marquez, M., and Arguello, H. (2022). Single pixel near-infrared imaging for spectral classification. In *Computational Optical Sensing and Imaging*, pages CW1B–2. Optica Publishing Group.

Bacca, J., Martinez, E., and Arguello, H. (2023). Computational spectral imaging: a contemporary overview. *JOSA A*, 40(4):C115–C125.

Boyd, S. and Vandenberghe, L. (2004). *Convex optimization*. Cambridge university press.

Cai, Y., Li, S., Zhang, W., Wu, H., Yao, X., and Zhao, Q. (2023). A detail-enhanced sampling strategy in hadamard single-pixel imaging. *Chinese Optics Letters*, 21(7):071101.

Cai, Y., Li, S., Zhang, W., Wu, H., Yao, X.-r., and Zhao, Q. (2022). A detail-enhanced sampling strategy in hadamard single-pixel imaging. *arXiv preprint arXiv:2209.04449*.

Candès, E. J. and Wakin, M. B. (2008). An introduction to compressive sampling. *IEEE signal processing magazine*, 25(2):21–30.

Centrone, A. (2015). Infrared imaging and spectroscopy beyond the diffraction limit. *Annual review of analytical chemistry*, 8:101–126.

Choi, I., Jeon, D. S., Nam, G., Gutierrez, D., and Kim, M. H. (2017). High-quality hyperspectral reconstruction using a spectral prior. 36(6).

Davies, A. (2005). An introduction to near infrared spectroscopy. *NIR news*, 16(7):9–11.

De Biasio, M., Arnold, T., Leitner, R., McGunnigle, G., and Meester, R. (2010). Uav-based environmental monitoring using multi-spectral imaging. In *Airborne Intelligence, Surveillance, Reconnaissance (ISR) Systems and Applications VII*, volume 7668, pages 331–337. SPIE.

Dekel, S. (2008). Adaptive compressed image sensing based on wavelet-trees. *preprint*.

Diaz, N., Rueda, H., and Arguello, H. (2018). Adaptive filter design via a gradient thresholding algorithm for compressive spectral imaging. *Applied Optics*, 57(17):4890–4900.

Dong, X., Xiao, X., Pan, Y., Wang, G., and Yu, Y. (2019). Dmd-based hyperspectral imaging system with tunable spatial and spectral resolution. *Optics express*, 27(12):16995–17006.

Duarte, M. F., Davenport, M. A., Takhar, D., Laska, J. N., Sun, T., Kelly, K. F., and

Baraniuk, R. G. (2008). Single-pixel imaging via compressive sampling. *IEEE signal processing magazine*, 25(2):83–91.

Edgar, M. P., Gibson, G. M., and Padgett, M. J. (2019). Principles and prospects for single-pixel imaging. *Nature photonics*, 13(1):13–20.

Elad, M., Kawar, B., and Vaksman, G. (2023). Image denoising: The deep learning revolution and beyond—a survey paper. *SIAM Journal on Imaging Sciences*, 16(3):1594–1654.

Fowler, J. E. (2014). Compressive pushbroom and whiskbroom sensing for hyperspectral remote-sensing imaging. In *2014 IEEE international conference on image processing (ICIP)*, pages 684–688. IEEE.

Garcia, H., Correa, C. V., and Arguello, H. (2020). Optimized sensing matrix for single pixel multi-resolution compressive spectral imaging. *IEEE Transactions on Image Processing*, 29:4243–4253.

Garini, Y., Young, I. T., and McNamara, G. (2006). Spectral imaging: principles and applications. *Cytometry part a: the journal of the international society for analytical cytology*, 69(8):735–747.

Gibson, G. M., Johnson, S. D., and Padgett, M. J. (2020). Single-pixel imaging 12 years on: a review. *Optics express*, 28(19):28190–28208.

Hinojosa, C., Marquez, M., Arguello, H., Adeli, E., Fei-Fei, L., and Niebles, J. C. (2022).

Privhar: Recognizing human actions from privacy-preserving lens. In *European Conference on Computer Vision*, pages 314–332. Springer.

Huang, L., Luo, R., Liu, X., and Hao, X. (2022). Spectral imaging with deep learning. *Light: Science & Applications*, 11(1):61.

Levenson, R. M. and Mansfield, J. R. (2006). Multispectral imaging in biology and medicine: slices of life. *Cytometry Part A: the journal of the International Society for Analytical Cytology*, 69(8):748–758.

López-García, L., Cruz-Santos, W., García-Arellano, A., Filio-Aguilar, P., Cisneros-Martínez, J. A., and Ramos-García, R. (2022). Efficient ordering of the hadamard basis for single pixel imaging. *Optics Express*, 30(8):13714–13732.

Monroy, B. and Bacca, J. (2024). Hadamard row-wise generation algorithm.

Monroy, B., Bacca, J., and Arguello, H. (2023). Deep adaptive superpixels for hadamard single pixel imaging in near-infrared spectrum. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE.

Ribes, A. and Schmitt, F. (2008). Linear inverse problems in imaging. *IEEE Signal Processing Magazine*, 25(4):84–99.

Rogalski, A. (2022). Scaling infrared detectors—status and outlook. *Reports on Progress in Physics*, 85(12):126501.

Rogalski, A., Martyniuk, P., and Kopytko, M. (2016). Challenges of small-pixel infrared detectors: a review. *Reports on Progress in Physics*, 79(4):046501.

Sampsell, J. B. (1994). Digital micromirror device and its application to projection displays. *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena*, 12(6):3242–3246.

Shaw, G. A. and Burke, H. K. (2003). Spectral imaging for remote sensing. *Lincoln laboratory journal*, 14(1):3–28.

Stuart, M. B., McGonigle, A. J., and Willmott, J. R. (2019). Hyperspectral imaging in environmental monitoring: A review of recent developments and technological advances in compact field deployable systems. *Sensors*, 19(14):3071.

Sun, M.-J., Meng, L.-T., Edgar, M. P., Padgett, M. J., and Radwell, N. (2017). A russian dolls ordering of the hadamard basis for compressive single-pixel imaging. *Scientific reports*, 7(1):3464.

Sylvester, J. J. (1867). Lx. thoughts on inverse orthogonal matrices, simultaneous signsuccessions, and tessellated pavements in two or more colours, with applications to newton's rule, ornamental tile-work, and the theory of numbers. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 34(232):461–475.

Teena, M. and Manickavasagan, A. (2014). Thermal infrared imaging. In *Imaging with Electromagnetic Spectrum: Applications in Food and Agriculture*, pages 147–173. Springer.

Vaz, P. G., Amaral, D., Ferreira, L. R., Morgado, M., and Cardoso, J. (2020). Image quality of compressive single-pixel imaging using different hadamard orderings. *Optics express*, 28(8):11666–11681.

Wang, L., Sun, C., Fu, Y., Kim, M. H., and Huang, H. (2019). Hyperspectral image reconstruction using a deep spatial-spectral prior. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8032–8041.

Yang, F., Sun, Q., Jin, H., and Zhou, Z. (2020). Superpixel segmentation with fully convolutional networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 13964–13973.

Yu, W.-K. (2019). Super sub-nyquist single-pixel imaging by means of cake-cutting hadamard basis sort. *Sensors*, 19(19):4122.

Yu, X., Stantchev, R. I., Yang, F., and Pickwell-MacPherson, E. (2020). Super sub-nyquist single-pixel imaging by total variation ascending ordering of the hadamard basis. *Scientific Reports*, 10(1):9338.

Zhang, Z., Wang, X., Zheng, G., and Zhong, J. (2017a). Fast fourier single-pixel imaging via binary illumination. *Scientific reports*, 7(1):12029.

Zhang, Z., Wang, X., Zheng, G., and Zhong, J. (2017b). Hadamard single-pixel imaging versus fourier single-pixel imaging. *Optics Express*, 25(16):19619–19639.

Zheng, S., Liu, Y., Meng, Z., Qiao, M., Tong, Z., Yang, X., Han, S., and Yuan, X. (2021). Deep plug-and-play priors for spectral snapshot compressive imaging. *Photonics Research*, 9(2):B18–B29.

Zhu, F., Wang, Y., Xiang, S., Fan, B., and Pan, C. (2014). Structured sparse method for hyperspectral unmixing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 88:101–118.

Zhu, L., Suomalainen, J., Liu, J., Hyyppä, J., Kaartinen, H., Haggren, H., et al. (2018). A review: Remote sensing sensors. *Multi-purposeful application of geospatial data*, pages 19–42.