AFRL-AFOSR-VA-TR-2024-0059

Infrared color-coded aperture optimization for object tracking and spectral classification

HENRY ARGUELLO FUENTES
UNIVERSIDAD INDUSTRIAL DE SANTANDER UIS
CARRERA 27 CALLE 9
BUCARAMANGA, SANTANDER, 680002
COL

12/06/2023
Final Technical Report

Air Force Research Laboratory
Air Force Office of Scientific Research
Arlington, Virginia 22203
Air Force Materiel Command

REPORT DOCUMENTATION PAGE

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED	
20231206	Final	START DATE 20210915	END DATE 20230914

4. TITLE AND SUBTITLE
Infrared color-coded aperture optimization for object tracking and spectral classification

5a. CONTRACT NUMBER	5b. GRANT NUMBER	5c. PROGRAM ELEMENT NUMBER
	FA9550-21-1-0326	
5d. PROJECT NUMBER	**5e. TASK NUMBER**	**5f. WORK UNIT NUMBER**

6. AUTHOR(S)
Henry Arguello Fuentes, Julian Rodriguez

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
UNIVERSIDAD INDUSTRIAL DE SANTANDER UIS
CARRERA 27 CALLE 9
BUCARAMANGA, SANTANDER 680002
COL

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)	11. SPONSOR/MONITOR'S REPORT NUMBER(S)
Air Force Office of Scientific Research 875 N. Randolph St. Room 3112 Arlington, VA 22203	AFRL/AFOSR IOS	AFRL-AFOSR-VA-TR-2024-0059

12. DISTRIBUTION/AVAILABILITY STATEMENT
A Distribution Unlimited: PB Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This report presents the outcomes of the project “Infrared color-coded aperture optimization for object tracking and spectral classification” conducted from September 15th, 2021, to September 14th, 2023. The report outlines the proposed approach for designing near-infrared coded apertures employing an end-to-end method that links the sensing with inference tasks. This methodology consists of a fully differentiable sensing model coupled with deep learning models to perform either spectral reconstruction or spectral classification, directly from the encoded measurements. Specifically, the proposed approach was studied for two different compressive spectral imaging systems: the single-pixel camera and the color-coded filter array sensor. In addition, an infrared spectral image dataset was acquired during this project. This dataset was employed to train the deep learning model. Simulation results show that the designed color-coded apertures can significantly enhance classification and reconstruction performance compared to random or analytical designs, what indicates a promising technology to be applied in the sensing and processing of near-infrared images. Further, test-bed implementations of the optical systems were built to evaluate the effectiveness of the proposed design in controlled laboratory scenarios. The results show that the proposed design improves the spatial-spectral resolution and reduces the number of measurements needed for a suitable reconstruction.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U	SAR	38

19a. NAME OF RESPONSIBLE PERSON	19b. PHONE NUMBER *(Include area code)*
MARIBEL HARMON	000-0000

Standard Form 298 (Rev.5/2020)
Prescribed by ANSI Std. Z39.18

Report: Infrared color-coded aperture optimization for object tracking and spectral classification.

Abstract

This report presents the outcomes of the project "Infrared color-coded aperture optimization for object tracking and spectral classification" conducted from September 15th, 2021, to September 14th, 2023. The report outlines the proposed approach for designing near-infrared coded apertures employing an end-to-end method that links the sensing with inference tasks. This methodology consists of a fully differentiable sensing model coupled with deep learning models to perform either spectral reconstruction or spectral classification, directly from the encoded measurements. Specifically, the proposed approach was studied for two different compressive spectral imaging systems: the single-pixel camera and the color-coded filter array sensor. In addition, an infrared spectral image dataset was acquired during this project. This dataset was employed to train the deep learning model. Simulation results show that the designed color-coded apertures can significantly enhance classification and reconstruction performance compared to random or analytical designs, what indicates a promising technology to be applied in the sensing and processing of near-infrared images. Further, test-bed implementations of the optical systems were built to evaluate the effectiveness of the proposed design in controlled laboratory scenarios. The results show that the proposed design improves the spatial-spectral resolution and reduces the number of measurements needed for a suitable reconstruction.

I. RESEARCH OBJECTIVES:

The main objective of this project is to design, simulate, and implement a coded aperture in an optical-computational system for compressed spectral imaging into the infrared spectrum to improve object tracking and classification. This project considers the following milestones:

1. **To develop statistical prior information of targets for spectral classification and object tracking in the infrared spectrum: (100% of execution).** There are few spectral images dataset in the infrared region. Therefore, we built our spectral image dataset using whiskbroom and pushbroom sensing strategies. Furthermore, statistical analyses were studied for the downloaded and acquired images.

2. **To mathematically design the spectral response and spatial structure of a coded aperture in compressive infrared imaging using as prior the spectral information of the targets for object tracking and spectral classification tasks: (100% of execution).** The spectral response and spatial structures of the coded apertures were modeled considering two different options for the optical devices that could implement them, i.e., the first approach involves using dichroic optical filters, which permit the passage of multiple wavelengths. The second approach is a set of specific band pass filters allowing passage only for a designated wavelength. In either scenario, the near-infrared band chosen for specific tasks such as classification and reconstruction were determined via the end-to-end approach. Addressing the spatial distribution of these selected filters, the challenge was modeled as a binary-coded aperture problem, which was further analyzed using end-to-end optimization techniques.

3. **To simulate a compressive computational optical system with the designed coded apertures for the infrared spectrum: (100% of execution).** The designed CCA was simulated into two computational optical systems: the single-pixel camera and the color-coded filter array sensor. This integration was carried out based on two key concepts: adaptive sensing and optical band selection, both seamlessly incorporated into our proposed end-to-end framework.

4. **To experimentally validate the proposed color-coded aperture designs in object tracking and spectral classification tasks against non-designed colored-coded aperture. (100 % of execution).** According to the project execution two tasks were evaluated, i.e., spectral reconstruction and classification. Simulations and results

from real data show that the designed color-coded aperture can substantially enhance classification and reconstruction performance compared to random or analytical designs, which shows a promising technology to be applied in the sensing and processing of near-infrared images. Also, test-best implementations of the optical systems were implemented in the optical Laboratory to evaluate the effectiveness of the proposed design.

II. ACCOMPLISHED TASKS

Acquisition of Spectral Imaging datasets: This accomplished task is related to objectives 1 and 4. An infrared spectral image dataset was acquired, spanning 512 spectral bands in the range 1000-2500 nm, with 64×64 spatial pixels. This dataset was created using a whiskbroom system implemented at the Optics laboratory of the HDSP (High-Dimensional Signal Processing) Research Group in Universidad Industrial de Santander, Colombia. The selected targets are built using two kinds of structures: homogeneous and heterogeneous material composition. These targets provided a dataset of signatures for different materials. To date, the resulting dataset is composed of 20 spectral data cubes. However, we expect to capture at least 100 additional spectral data cubes to be used in deep-learning tasks.

Coded Aperture Design modeling: This accomplished task is related to objective 2. We proposed a data-driven color-coded aperture (CCA) optimization based on deep learning. Specifically, by modeling the sensor measurements through a fully differentiable image formation model that considers the physics-based propagation of light and its interaction with the CCA, the parameters that define the CCA, and the computational decoder which can be optimized in an end-to-end (E2E) manner. This framework optimizes the CCA by back-propagating a task-related error up to the parameters of the CCA.

Reconstruction of Compressive spectral imaging in the infrared spectrum: This accomplished task is related to objectives 2 and 3. Two compressive spectral imaging systems in the infrared spectrum were implemented, including coded apertures in their setups. The first one is the color-coded filter array which incorporates the principles of compressive sensing by acquiring the entire spectral data cube with just focal plane array (FPA) measurements employing spatial-spectral modulation. The second one is the single-pixel camera with a near-infrared spectrometer as a detector is a hardware compression system that encodes a spectral image using binary-coded apertures.

Simulation and numerical experiments: This accomplished task is related to objectives 3 and 4. We tested our designed method for the tasks of spectral classification and reconstruction in the scenarios of noise, compressive ratio, and adjustment of different spectral ranges. These results have shown an encouraging performance compared with non-designed CCA.

Test-bed Implementation: This accomplished task is related to objective 4. We validated the E2E design to classify single-pixel NIR measurements directly in the compressed domain. Results show that the proposed method enables pixel-level classification of challenging objects, which are not distinguishable in the visible range for different compression levels.

III. DETAILED DESCRIPTION OF THE PROPOSED METHOD

The structure of this report is as follows:

In Section 1, we outline the datasets employed in this project and detail the corresponding analysis conducted to determine the suitable prior information utilized.
In Section 2, we provide a comprehensive explanation of the mathematical formulation underpinning the proposed end-to-end coded aperture design.
In Section 3, we delve into the compressive spectral imaging systems integrated into this project, which incorporate coded apertures into their setups.

In **Section 4,** we delve deeper into the intricacies of the modeling and the approach utilized for designing the coded aperture in the context of the reconstruction task.
In Section 5, we offer a detailed account of the modeling and coded aperture design strategy employed for the classification task.

1. Datasets.

This section describes two spectral datasets used in the project. The first one was obtained from Copernicus-S2 dataset while the second one was acquired in the HDSP lab at the Universidad Industrial de Santander. The former is related to satellite images in different spectrum ranges, i.e., visible, near infrared, and short-wave infrared. The latter is restricted to images in the NIR spectral range (1000-2500 nm) captured by the project team as described in Section 1.2. In addition, image analysis investigating noise, compression, and statistics was performed for the former dataset.

1.1 Available Visible, Near Infrared and Shortwave Infrared Datasets

The public dataset, initially, consisted of 26563 Copernicus-S2 satellite images in each one of the visible (VIS), near-infrared (NIR), and shortwave infrared (SWIR) spectrum ranges. Those images can be acquired from the website https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi/data-formats. Nevertheless, the first 2000 images from the Copernicus dataset were selected to compose the dataset used in this project. Afterwards, a data filtering was performed to remove images corrupted by atmospheric effects, noise, or acquisition issues. At the end, the dataset size for this project was reduced to 1000 images. From those, 100 were selected randomly for performing testing in the proposed deep learning algorithms, while the remaining ones were used for training.
The spatial dimensions of the scenes vary according to the wavelength. For the dataset extracted, the images in the SWIR range have a spatial resolution of 60m, which is three times larger than that from the NIR images, and six times larger than the resolution of images in the visible range. In summary, the spatial resolutions of the captured regions for the satellite images were 60m, 20m, and 10m for the ranges SWIR, NIR, and VIS, respectively. The sizes of the downloaded images in the ranges SWIR, NIR, and VIS are 44x44, 132x132, and 264x264, respectively. Those images cover the same scene for each instance. For the implementation in the proposed deep learning algorithms, the images were cropped to meet the sizes 26x26, 78x78, and 156x156, respectively. This data reduction was necessary to overcome memory storage constraints in the data processing stage. The images in the visible spectrum were defined by combining the bands B2, B3, and B4 generating a grayscale image. The description of those bands is given in the Table 1.1.

This project selected the first 2000 images to compose a subset of the Copernicus dataset that can be accessed through the Github repository https://github.com/zhu-xlab/SSL4EO-S12. The Python script ssl4eo_downloader.py conveys the information for downloading this dataset. The satellite images have datatype uint16 and were selected on the dates 2012-12-21, 2021-09-22, 2021-06-21, and 2023-06-14. Those images were filtered with cloud percentage around 1%. Moreover, those images were constrained to 5 channels, which are formed by three bands in the visible range (B2, B3, and B4) and two bands in the infrared range (B7 and B9). The spatial dimension of the images is 1320 m x 1320 m. For reference, Figure 1.1 illustrates three images of the same region for each spectral range.

Table 1.1 Description of spectral bands from satellite Copernicus-Sentinel 2.

Band	Resolution [m]	Wavelength [nm]	Description
B2	10	490	Blue
B3	10	569	Green
B4	10	665	Red

B7	20	783	NIR
B9	60	950	SWIR

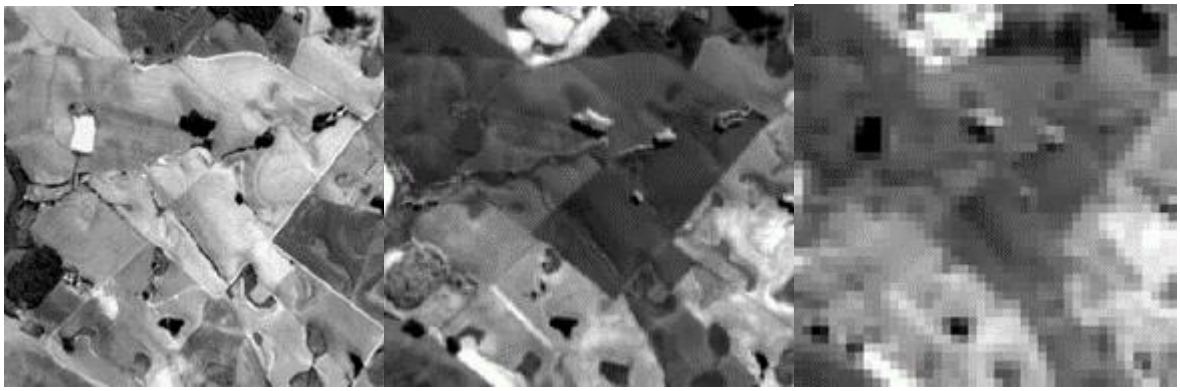
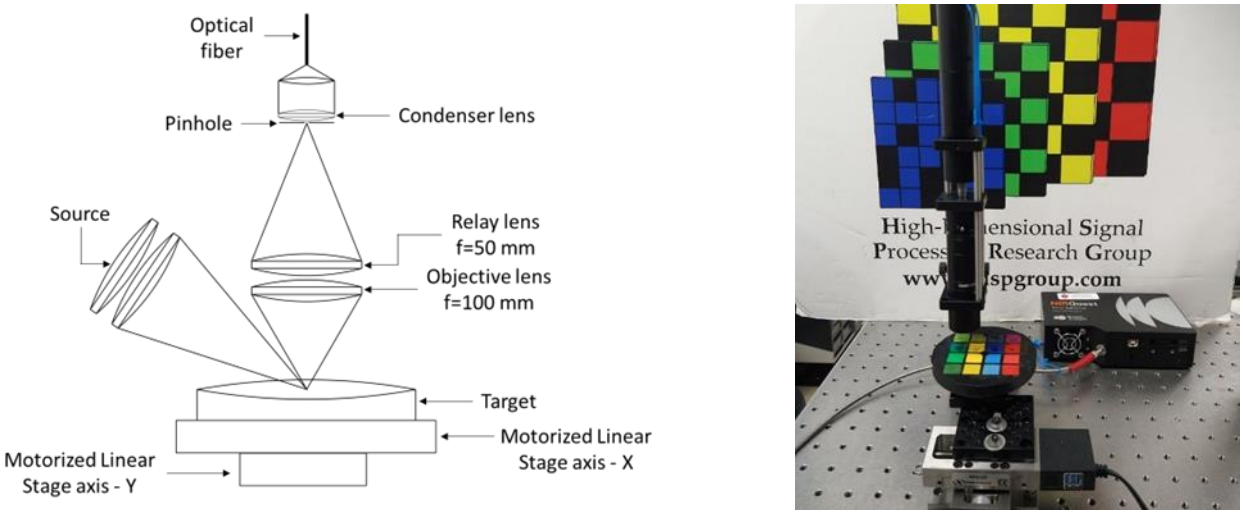


Figure 1.1. Sentinel 2 satellite images from the same region acquired from the Copernicus dataset. (Left) Visible range; (Center) NIR range; (Right) SWIR range.

1.2 Acquired Spectral Near-Infrared Datasets

An optical NIR system was built in our optics laboratory to acquire image datasets in the NIR spectral range (1000-2500 nm). This section reports the study, assembly, and calibration of such a system, which specifically consists of a whiskbroom architecture as illustrated in Fig. 1.2(a). The specifications of employed optical elements are listed in Table 1.2, and the implemented system is depicted in Fig. 1.2(b).



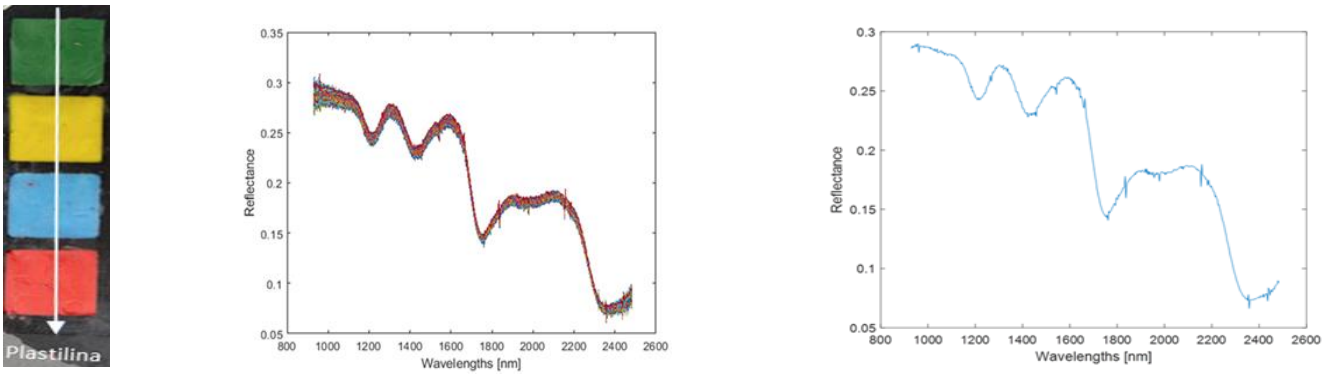
(a) (b)

Figure 1.2. (a) Sketch of the whiskbroom system. (b) Testbed implementation of the whiskbroom system in our optics laboratory.

Table 1.2. List of the optical elements used for the Whiskbroom system in Fig. 1.2(a).

Element	Reference	Quantity	Description	Company
Lens NIR f=50	LB5284	1	Lens in medium IR range 0.18 - 8.0 [µm] with focal length f=50	Thorlabs
Lens NIR f=100	LB5552	1	Lens in medium IR range 0.18 - 8.0 [µm] with focal length f=100	Thorlabs
Pinhole	SM2D25	1	Pinhole with a diameter range from 1.0 mm to 11.9 mm.	Thorlabs
Collimator lens	74-UV	1	Lens optimized for Visible-NIR range (350-2500nm)	Ocean insight
Optical fiber	QP100-2-VIS-NIR	1	Visible-NIR transmission (400 and 2100 nm)	Ocean insight
Spectrometer	NIRQuest + 2.5	1	The NIRQuest+2.5 is a versatile NIR spectrometer for applications ranging from moisture detection to laser characterization. It responds from 900-2500nm.	Ocean insight
Linear stage	UTS100PP	2	Motor-driven linear translation stage with a travel range of 250 mm.	Newport

The main component of the optical system is the NIRQuest+2.5 type spectrometer, which acquires spectral signatures with 512 points in the NIR range (1000-2500 nm). The NIRQuest is composed of a set of mirrors, a NIR grating, and an InGaAs linear array which allows an optical resolution of up to 2.8nm FWHM. To illustrate an example of the spectral signatures acquired by this system, we used the clay-type color target from Fig. 1.3(a). Figure 1.3(b) presents spectral signatures corresponding to different clay sections, while Fig. 1.3(c) presents the average spectral signature for each material.



(a) (b) (c)

Figure 1.3. (a) Illustration of the clay-type target and the (b) spectral signatures acquired using the Whiskbroom. The signatures were acquired by a random spatial scanning of the four clay-type targets. (c) Illustration of the average signature.

Finally, the acquisition of the spectral images was conducted by synchronizing the NIRQuest spectrometer, and the linear stage to scan all the spatial positions of the scene, obtaining the spectral response of the material at each spatial position. This synchronization was used for all the components with a python-based developed software. In this way, we acquired an infrared spectral dataset of 50 objects, each with 128x128 spatial pixels and 500 spectral bands. Figure 1.4 illustrate the acquired data for 5 different objects. Left subfigure depicts the grayscale intensity distribution maps, in the middle, a false composite RGB version of the acquired images is illustrated, and the right column shows the spectral signatures of 3 sample points from each image. This dataset is stored in the digital archives of the HDSP group at the Universidad Industrial de Santander.

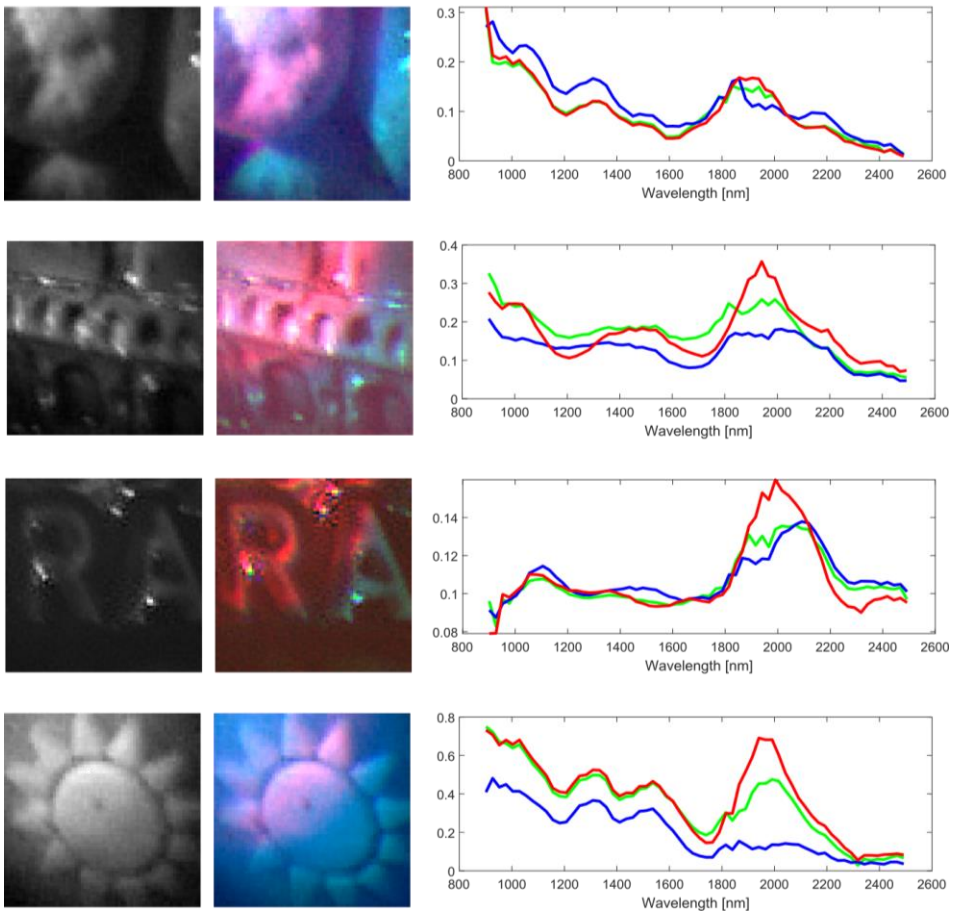


Figure 1.4. (Left) intensity distribution map (grayscale version), (Center) a false composite version, (Right) the spectral signatures in the intensity representation.

1.3. Image Analysis

This section is focused on the image analysis of the spectral images from the Copernicus dataset. It is aimed to understand the difference in the behavior of spectral images, under certain circumstances, depending on the spectrum range. The topics of analyses are the following:

- Image compression analysis.
- Image noise analysis.
- Image statistical analysis.

1.3.1. Compression analysis

The images from the testing dataset were submitted to compression using wavelets. More precisely, Daubechies 1 wavelets with four levels of compression (98%, 80%, 60%, and 50%) were applied to the testing images for each one of the VIS, NIR, and SWIR spectral ranges. However, the images have different sizes depending on the spectrum range. Therefore, to make a fair comparison between the images, 1000 patches of size 26x26 were selected randomly for each one of the 100 images for the testing dataset. It was concluded that the peak signal-to-noise ratio (PSNR) of images in the SWIR range are more sensitive to compression than images in the NIR range, and then those ones in comparison with images in the visible range. The following figures present a comparison of the average values of PSNR among images in the ranges SWIR, NIR, and VIS. It can be noted that the values of PSNR increase as the compression rate increases. However, this increasing is weaker in the range SWIR, followed by the ranges NIR, and VIS. Therefore, it can be inferred that images in the SWIR range are more sensitive to compression. Moreover, the horizontal lines in the figure are references of how much of

compression an image must have so that the PSNR approaches to their values (30 dB, 40 dB, or 50 dB).

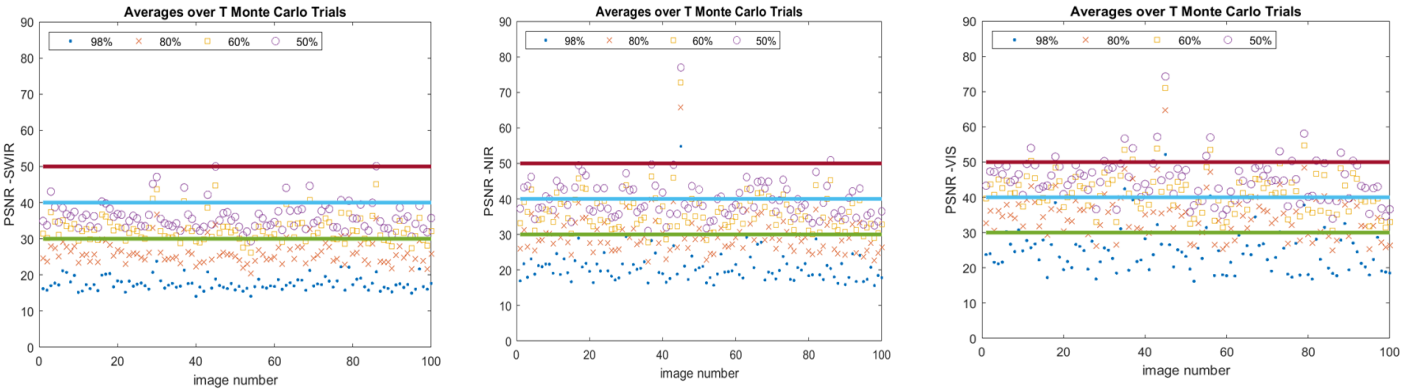


Figure 1.5 Testing images versus average PSNR between T=1000 randomly selected 26x26 image patches and corresponding images in the (left) SWIR, (middle) NIR, and (right) VIS ranges when submitted to four levels of Daubechies 1 wavelet compression, namely 98%, 80%, 60%, and 50%.

Table 1.3. Mean, standard deviation and median values of the PSNR for 100 compressed images in the ranges VIS, NIR, and SWIR averaged over 1000 Monte Carlo trials.

Compression rate	Statistics	VIS	NIR	SWIR
98%	Mean	**24.7**	20.77	17.53
	Standard deviation	6.45	4.99	**2.06**
	Median	**23.84**	19.58	17.13
80%	Mean	**34.65**	29.46	25.92
	Standard deviation	6.25	5.53	**3.10**
	Median	**34.43**	28.2	25.24
60%	Mean	**41.51**	36.00	36.06
	Standard deviation	6.24	5.77	**3.92**
	Median	**41.25**	34.73	35.22
50%	Mean	**45.2**	39.59	36.06
	Standard deviation	6.27	5.92	**3.92**
	Median	**45.2**	38.35	35.22

Table 1.3. presents some statistics of PSNR referring to the points in Figure 1.5. like mean, standard deviation, and median. Notice that the mean and median values of PSNR are higher for the visible range for all compression rates. On the other hand, images in the SWIR range present lower standard deviation.

1.3.2. Noise analysis

Spectral images convey naturally some noises that are produced due the acquisition process. The noises in each spectrum range can differ due to physical constraints and kinds of sensors that are used in the imaging process. The noise distribution of the spectral images is analyzed by means of an estimation of a white noise in the images after applying the denoising algorithm BM3D.

Algorithm 1. Pseudo-code of the noise estimation using the BM3D algorithm.

```
%% Method with BM3D denoiser
c = 0;
Sgmas = linspace(5e-11,5e-10,20);
for sigma =Sgmas
    noo = randn(size(y));
    c = c+1;
    [out] = BM3D(y,sigma);
    error_temp = y - out;
    error_temp = error_temp(:);
    ytemp = y(:);
    estimado_snr(c) = snr(ytemp,error_temp);
    sigma_estimate(c) = (rssq(out(:))./(db2mag(estimado_snr(c)))./rssq(noo(:));
end
error=abs(sigma_estimate-Sgmas);
[p,v]=min(error);
noise=Sgmas(v)
vector_noise(i)=noise;
end
```

The images were denoised by the BM3D method. The range of values of sigma were tuned to restric the values of the estimated noises. Figure 1.4. shows the distribution of those noises for each one of the ranges SWIR, NIR, and VIS.

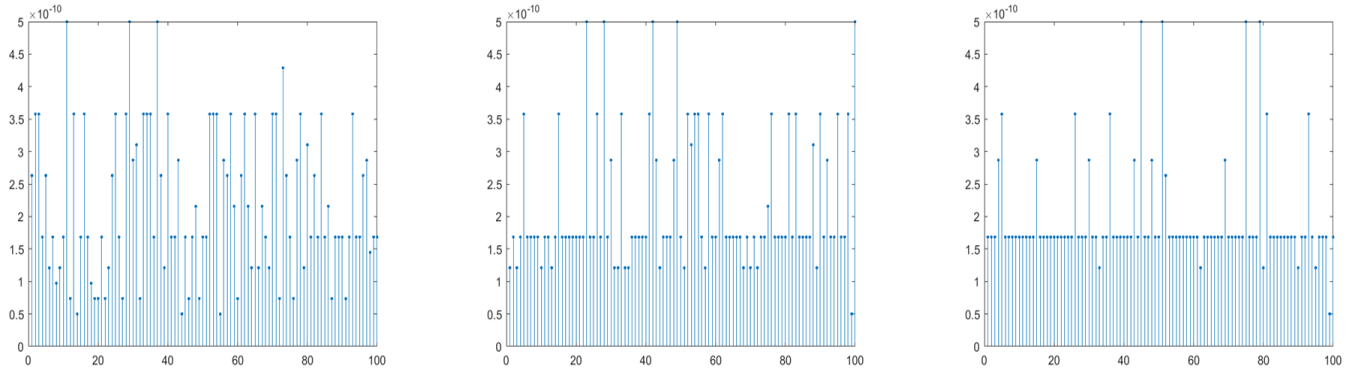


Figure 1.4. Estimated gaussian noise for each one of the 100 satellite images in the (left) SWIR, (middle) NIR, and (right) VIS ranges.

Note that the noise distribution is almost uniform for the images in the visible range but for some outliers. For the range NIR, the same occurs; however, there are more outliers. Nevertheless, for the SWIR range, the distribution follows another pattern rather than uniform. It can be noticed a higher variation in the distribution of the points as corroborated in Table 1.4.

Table 1.4. Statistics of the estimated noise of the satellite images in 1e-10 scale.

	SWIR	**NIR**	**VIS**
Mean	2.1911	2.1626	**1.9566**
Standard deviation	1.1391	1.0326	**0.8186**
Median	1.6842	1.6842	1.6842

The statistics are less than or equal in increasing order for the ranges VIS, NIR, and SWIR. This means that images in the visible range hold less noise. In addition, they present less variation in the distribution of the noise over the images of the dataset. Those pieces of information are inferred via the mean and standard deviation shown in Table 1.4. On the other hand, images in the SWIR range are noisier and present a higher variation from image to image. In fact, in the SWIR range, the standard deviation and mean are higher.

1.3.3. Image statistical analysis

For each one of the ranges SWIR, NIR, and VIS, the mean and standard deviation were calculated by using the MATLAB inbuilt functions mean2 and std2. The statistics of these metrics are presents in the table below.

Table 1.3. Overall averages of the statistics mean and standard deviation of the 100 images per range.

	SWIR	**NIR**	**VIS**
Overall mean	0.4744	0.4529	**0.3362**
Overall standard deviation	0.2142	0.2139	**0.2042**

The images in the visible spectrum show, in average, lower values of mean and standard deviation when compared to the ranges NIR and SWIR.

2. Data-Driven Coded Aperture Design based on E2E framework

This project develops a new method to optimize the coded aperture used in compressive spectral imaging (CSI) in the infrared region. Specifically, the proposed scheme is based on the denominated end-to-end (E2E) and takes advantage of the available data and algorithm capabilities of deep neural networks (DNNs) as shown in [1]. This new framework jointly designs the coding patterns used in CSI and the network parameters to perform a given task directly from the embedded near-infrared compressive measurements, as illustrated in Figure 2.1. More precisely, it simulates the CSI system as a fully differentiable image formation model that considers the physics-based propagation of light and its interaction with the coded aperture (CA). The CA can be represented by learnable parameters and interpreted as an optical layer in this model. In the same way, the overall CSI system can be constructed as an optical encoder composed of different optical layers. Given that deep neural networks (DNNs) represent the forefront in various computer vision tasks, including classification and reconstruction, we suggest an integration strategy. This strategy involves coupling the optical encoder with a DNN, specifically tailored for a defined task. This combination allows for the optimization of ensemble parameters, encompassing both the CA and DNN parameters in an end-to-end (E2E) way. This optimization process leverages a training dataset and employs the back-propagation algorithm. The following section mathematically describes the E2E optimization.

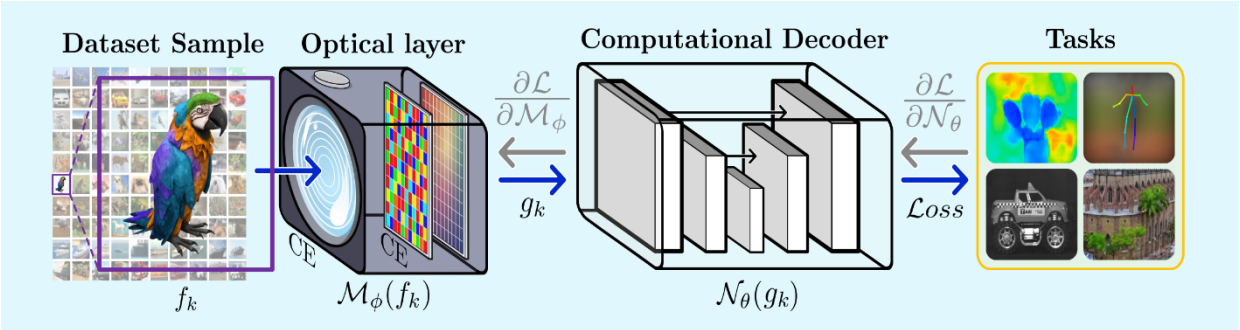


Figure 2.1. E2E scheme where the CSI system is modeled as an optical layer. In training, a set of images passes through the optical system, obtaining the projected measurements that enter the computational decoder producing outputs for an arbitrary task. The estimated task error is propagated from the output of the decoder to the optical layer, updating the weights of the decoder and the optical system. Recovered from [1].

End-to-End Optimization

The key idea for an E2E design is to accurately simulate the forward propagation of a CSI system. In particular, two methodologies are studied based on the kind of filter used in the CCA and considering fabrication constrains. The CSI system is here denoted as M_ϕ and the computational decoder or deep model to perform a given task is denoted as N_θ. Precisely, E2E optimization consists of training the encoder-decoder parameters to perform a given task. This optimization is mathematically expressed as

$$\{\boldsymbol{\phi}^*, \boldsymbol{\theta}^*\} = \sum_k L_{task}\left(N_{\boldsymbol{\theta}}\left(M_{\boldsymbol{\phi}}(\boldsymbol{x}_k)\right), \boldsymbol{d}_k\right) + \rho R_\rho(\boldsymbol{\phi}) + \sigma R_\sigma(\boldsymbol{\theta}), \tag{1}$$

where $\{\boldsymbol{\phi}^*, \boldsymbol{\theta}^*\}$ represent the set of optimal CCA parameters and the optimal weights of the network, respectively, $\{\boldsymbol{x}_k, \boldsymbol{d}_k\}_{k=1}^K$ account for the training database with K elements, where $\boldsymbol{x}_k$ are the input images and $\boldsymbol{d}_k$ are the outputs of the neural decoder that can be a target image, a classification vector, a segmentation map, among others. The loss function L_{task} is linked to a specific inference task. For instance, the mean-squared error (MSE) and cross-entropy metrics are conventionally used for reconstruction and classification tasks [2]. $R_\rho(\boldsymbol{\phi})$ and $R_\sigma(\boldsymbol{\theta})$ denote regularization functions that act on the optical parameters and weights of the decoder, respectively, with ρ and σ as regularization parameters. The regularization functions have been widely used for training deep models to reduce the overfitting problem, a common issue when training deep neural networks. For instance, the l_2 or l_1 norms have been successfully applied[2].

The regularization over the optical parameters $R_\rho(\phi)$ plays a different role than the network's weights since they directly change the values of physical optics. Thus, it is helpful to promote desired properties of the CCA. The main idea of including the regularization in training is that the gradient of the loss function concerning ϕ is calculated using the chain rule as

$$\frac{\partial L}{\partial \boldsymbol{\phi}} = \frac{\partial L_{task}}{\partial N_{\boldsymbol{\theta}}} \frac{\partial N_{\boldsymbol{\theta}}}{\partial g} \frac{\partial \boldsymbol{g}}{\partial \boldsymbol{\phi}} + \rho \frac{\partial R_\rho}{\partial \boldsymbol{\phi}}. \tag{2}$$

Therefore, the design of the optical elements is directly influenced by the loss of the task and the regularization function. For example, physical restrictions of the CCA implementation process that are not addressed in the parametrization of the optical elements impose constraints on the optimization, for instance the CCA entries must be binary. This is addressed in the following subsections. Additionally, the parameter ρ induces a trade-off between the optimal performance and the desired properties imposed in the regularization. In this sense, work in [2] suggests using an exponential increase strategy, where the idea is that at first epochs, the derivative of the loss gives the direction to converge to the desired task values, and then ρ is increased to guarantee regularization performance.

3. Compressive Near-Infrared Spectral Imaging Systems

Utilizing the single-pixel near-infrared imaging and a color-coded filter array as the chosen sensing schemes for this research project presents several noteworthy advantages. Firstly, Single Pixel Imaging (SPI) yields high-resolution spectral data akin to a spectrometer, offering the flexibility to select the most suitable spectral range for specific applications. Secondly, the color-coded filter array provides exceptional spatial resolution, enabling precise identification of the spatial structures within the observed scenes. The use of both systems equipped the research project with a powerful toolset for comprehensive data capture and analysis.

3.1 Single Pixel Near-Infrared Imaging.

SPI has emerged as an extreme case of compressive imaging, since only one sensor is employed as detector. Its unique capability of capturing high-resolution images using a single photodetector or Near-Infrared spectrometer (in the case of Near-Infrared spectral images) has sparked interest among researchers and engineers. Figure 3.1 provides a visual

representation of this concept. In the single pixel imaging system, the spatial light modulators (SLMs) play a vital role in shaping the incident light patterns. Commonly used SLMs include digital micromirror devices (DMDs) and liquid crystal displays (LCDs). The resolution of this imaging system is determined by the characteristics of the SLMs rather than the sensor itself. This attribute makes single pixel imaging particularly attractive for near-infrared imaging applications since the sensor is considerably larger than the pitch of the SLMs. Mathematically, the sensing process of the SPI using a NIR spectrometer as detector is modeled as

$$\boldsymbol{y} = \boldsymbol{h}^{\boldsymbol{T}}\boldsymbol{X}, \tag{3}$$

where $\boldsymbol{X} \in R^{hw \times c}$ denotes the spectral image with h and w representing the horizontal and vertical dimensions and c the number of spectral bands, $\boldsymbol{h} \in R^{hw}$ represents the pattern in the SLM known as binary coded aperture (CA) and $\boldsymbol{y} \in R^{c}$ are the acquired measurements. In SPI, different CA are sequentially displayed on the SLM to acquire different snapshots of the scene. Consequently, the whole measurements in SPI represent the linear projections of $\boldsymbol{X}$ onto the sensing matrix $\mathbf{H} \in R^{m \times n}$ with $n = hw$, where each row of $\mathbf{H}$ represents a CA. Considering the noisy case the whole sensing process is mathematically modeled as

$$\boldsymbol{Y} = \boldsymbol{HX} + \boldsymbol{\eta}, \tag{4}$$

where $\boldsymbol{\eta}$ represents the noise. Common CA patterns include random, Hadamard, and Fourier patterns. However, Hadamard patterns are frequently employed due to their binary nature and invertibility. However, it is important to consider the trade-offs between pattern complexity, acquisition time, and image reconstruction quality when selecting the appropriate CA, i.e., Hadamard matrices are square which means that $m = n$ measurements are needed. However, the idea of compressive systems is reducing the number of measurements $m << n$ by carefully designing the CA.

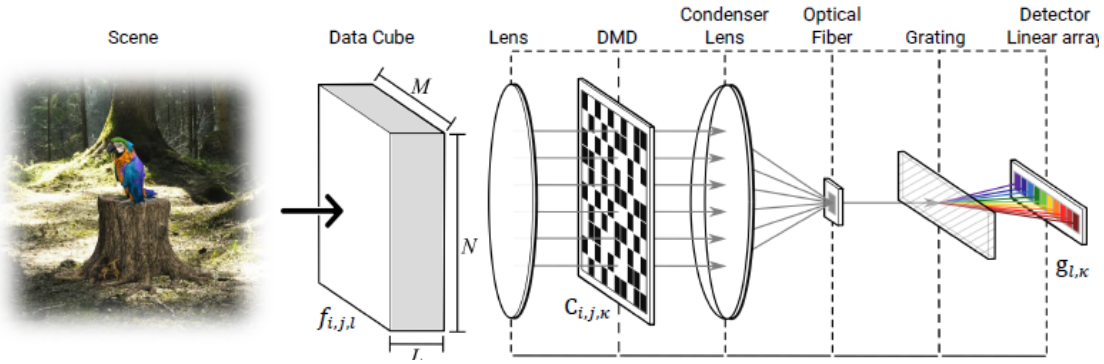


Figure. 3.1 A visual representation of the single pixel imaging system is shown.

3.2 Color-Coded Filter Array

The color-coded filter array (CCFA) is a compressive spectral imaging (CSI) system that allows a spatial-spectral modulation on the spectral data cube using a color-coded aperture (CCA) that employs a different coding pattern for each spectral band, as illustrated in Figure 3.2. Let $\boldsymbol{X} \in R^{M \times N \times L}$ be the spatial-spectral input data cube with $M \times N$ spatial dimensions and L spectral bands. Each voxel of an $M \times N \times L$ array is denoted as $\boldsymbol{x}_{m,n,l}$ where $m = \{0, \dots, M-1\}$ and $n = \{0, \dots, N-1\}$ index the spatial coordinates and $l = \{1, \dots, L-1\}$ represents the spectral bands' index. Specifically, each spectral pixel passes through its corresponding dichroic optical filter that is modeled as a binary coding pattern $\phi^{s}_{m,n,l}$, where $s = \{0, \dots, S-1\}$ indexes the snapshots. Then, the coded spectral scene is relayed into the focal plane array (FPA) detector, where the compressive measurement ($S \ll L$) are acquired by the integration along the spatial dimension of the detector. Mathematically, the output of the sensing process at the (m, n)-th detector pixel and specific snapshots can be expressed as

$$\hat{\boldsymbol{y}}_{m,n}^{s} = \sum_{l=0}^{L-1} \boldsymbol{\phi}_{m,n,l}^{s} \boldsymbol{x}_{m,n,l}. \tag{4}$$

The set of compressive measurements from (4) can be rearranged in a $S \times MN$ matrix $\hat{Y}$, where each column contains the compressive measurements associated to a particular spectral pixel. In practice, the number of coding patterns is less than the number of pixels. Therefore, this work assumes that the number of different coding patterns in a single shot is equal to S and to reduce the redundant information, each pixel is encoded only once by a different coding pattern, i.e., at the end of the sensing procedure, the whole set of CCA encodes all pixels' coding patterns [1].

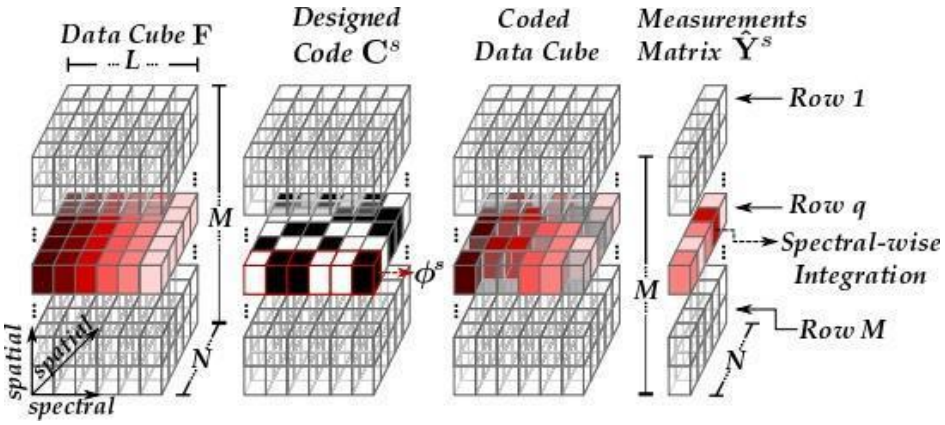


Figure 3.2. CCFA sensing approach at the snapshot s. Recovered from [2].

4. Reconstruction task (Single Pixel Imaging)

In a single pixel imaging system, acquiring a complete observation set $\boldsymbol{Y}$ requires capturing n snapshots. In compression scenarios, compressive strategies are often employed to reduce the number of acquired snapshots (m), where $m << n$. This reduction is achieved by employing a set of Compressive Sensing (CS) techniques which, leverage certain assumptions about the spectral image being acquired, such as sparsity, low-rank, and local correlation, among others. Commonly used CS methods are based on randomness, following the principles of compressive sensing theory. However, the primary aim of this project is to demonstrate that these methods can be customized to achieve optimal performance for the NIR-infrared range. The judicious selection of a reduced number of single-pixel captures introduces a trade-off between resolution and time. Consequently, in this section, we delve into two methodologies for designing the CS methods used in single-pixel imaging systems:

i) **End-to-End Methodology:** In this approach, both, the parameters of the computational optical acquisition system and the parameters of a neural network are treated as trainable variables. These parameters are jointly optimized to enhance image reconstruction quality.

ii) **Adaptive Single-Pixel Imaging**: In this methodology, prior information extracted from previous captures is leveraged to adaptively estimate the CS technique for each spectral image. This is achieved by applying linear transformations to the sensing matrix from two perspectives: left and right linear transformations.

4.1. End-to-End Methodology

To design the coded apertures to reconstruct spectral images in the Visible (VIS), Near infrared (NIR) and Short-wave Infrared (SWIR) ranges, an End-to-End (E2E) approach was implemented. The E2E methodology has two stages: first, the

encoder simulates the optical system. Second, the decoder simultaneously learns certain optical parameters along with the reconstruction of the corresponding spectral images [2].

In the encoding phase, the process involves learning Coded Apertures (CAs) as weights within a neural layer emulating the optical system. This learning is conditioned on a specified compression level. At the end of training, these weights are expected to evolve into binary values. To facilitate this gradual transition, a regularizer is employed. It guides the network to progressively move the weights towards binary values, thus ensuring a seamless learning process that doesn't impose an immediate binary constraint. Conversely, the decoder phase revolves around training a neural network to reconstruct the spectral image. This reconstruction is achieved using the compressed measurements acquired through simulation of the optical system.

The E2E architecture is based on the unrolling algorithm, and it is composed of 9 stages. Internally, it has layers that simulate the Forward, Transpose and Gradient of the Single Pixel Camera [3]. For the network training, random patches are selected from the 900 train images at each epoch, which avoids overfitting. Random selection is also performed on the 100 test images. Finally, for the validation of the method, the test images are systematically traversed through all their patches. Then, inference is performed on each of them.

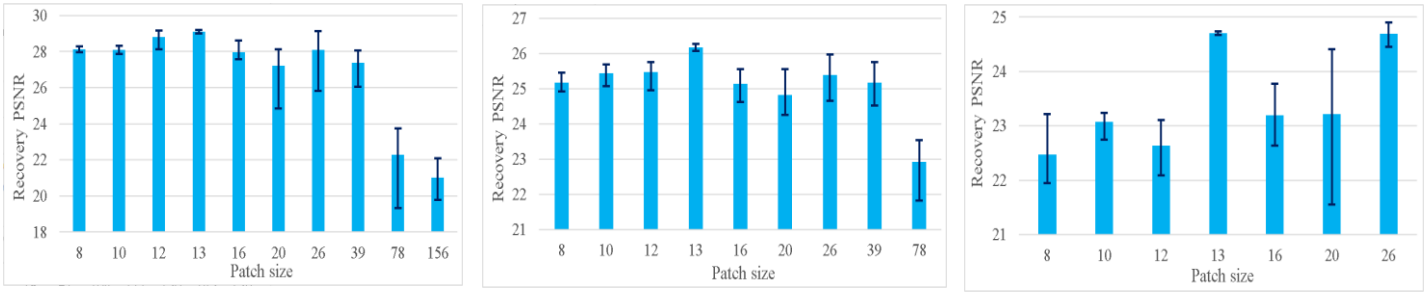
The modifiable parameters of the methodology are the compression ratio, the noise of the compressed measurements, the range of the spectrum, the type of network, the CA transmittance to which the algorithm is expected to converge, the initializations of the regularization parameters of the binarization and transmittance, the magnification factor of the regularizer, magnification step, the number of epochs, the batch size, learning rate, the initialization distribution of the weights representing the encoded aperture, the depth of the neural network, a Boolean value to determine whether to retrain the decoder by freezing the learned values of the encoded aperture.

The number of network parameters is proportional to the patch size and the compression ratio, the latter also determines how many coded apertures will be trained. The compression ratio refers to the percentage of information captured with respect to the information to be reconstructed. For example, for a patch size of 13, and a compression level of 25%, 42 patches are trained, and the number of network parameters is 345,813.

4.1.1 Performance analysis

In this research, the three studied spectral ranges have different characteristics, as described in the Datasets section. Possible biases in the training were identified, such as the number of network parameters, the number of CAs to train, and the different size of the images in each range. To analyze the complexity of learning the CAs, a patch training approach was taken, i.e., the images are subdivided with a specific patch size. This approach makes the number of trained parameters and CAs the same for all ranges, in addition to reducing the required memory and training time. In these experiments a fixed compression ratio value of 25% was used.

A study was performed by varying the patch size for the different image ranges. This investigation involved running five repetitions for each patch size, and subsequently, calculating the mean and standard deviation of these repetitions for each case using testing data, as illustrated in Figure 4.1.



(a) (b) (c)

Figure 4.1. Average PSNR obtained for different repetitions by varying the patch size in the (a) VIS, (b) NIR, and (c) SWIR range.

From Figure 4.1. (a), it can be concluded that the best patch size in the VIS range is 13, since it obtains the best PSNR on average with the lowest standard deviation, although closely followed by 26. In addition, a patch size of 13 divides the image avoiding residual patches, in contrast to 8 and 10, which are not divisors of 156, i.e., the image size.

In the NIR range (b), a considerable improvement is observed in the 5 repetitions of the initial training with the patch size of 13, it is with this value also, that the lowest standard deviation is obtained. It is observed that patch sizes of 10, 12, 13, and 26 have the best quality metrics in terms of PSNR for this spectral range.

The SWIR range in Figure 4.1. (c), having 26x26 images, a smaller number of patches were tested. It can be observed that the best reconstructions on average are obtained with patches of 13 and 26, the latter being the one that in some cases manages to obtain outliers above the average, which makes it a candidate for freezing and retraining.

The best three patch values were selected and shown in Figure 4.2. Note that the best so far is a patch of 13, for the SWIR, NIR and VIS ranges, followed closely by size patch of 26 and finally the 12 patches.

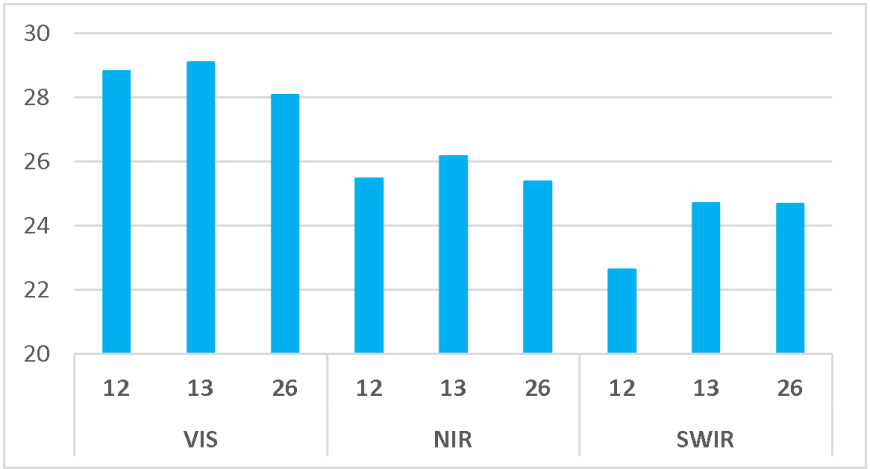
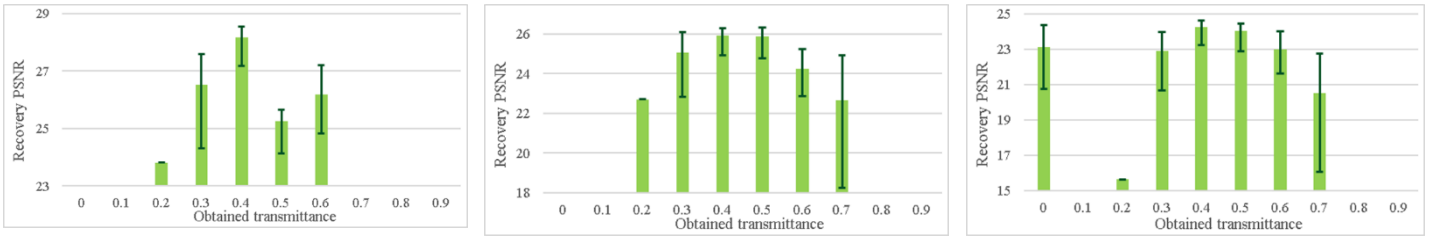


Figure 4.2. Average PSNR obtained in the three ranges with patch size of 12, 13 and 26.

The process of freezing the binary coded apertures learned from the best repetitions or experiments was performed, so that only the decoder can be re-trained. In these results, for a patch size of 13, the reconstruction quality improved to 24.85, 26.48, and 29.41 dB in PSNR, for SWIR, NIR, and VIS, respectively. In the case of a 26 patch, better results were achieved in all cases with 25.14, 26.54, and 29.44 dB, in the three respective ranges. Being 26 the patch size that was finally selected to obtain the real data in the laboratory.

4.1.1. Transmittance analysis

An analysis of the transmittance obtained by the E2E method is carried out for each of the spectral ranges. Transmittance is a measure of the amount of light that passes through the coded aperture. It is determined by the ratio of +1 values in the coded aperture as opposed to -1 values. To obtain a given transmittance value, a regularizer is used whose influence increases with each iteration. With these tests it is possible to observe how the transmittance of a coded aperture affects the quality of the reconstructions. The transmittance values that were tested and are expected to be obtained by the transmittance regularizer are between 0.3 and 0.7, although this is not met in all cases. In these experiments a fixed compression ratio value of 25% was used.



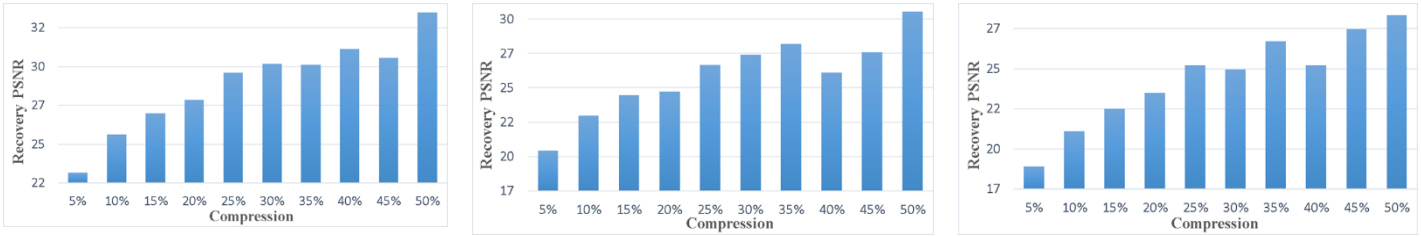
(a) (b) (c)

Figure 4.3. Average PSNR obtained for different transmittance values in the (a) VIS, (b) NIR, and (c) SWIR ranges.

Reconstruction results in Fig. 4.3 (a) show that for the VIS range, the obtained transmittance values range between 0.2 and 0.6, with the best reconstructions for 0.4, and the lowest standard deviation was reached by 0.5 transmittance with almost 26 dB of PSNR. A more spread behavior is presented in the case of the NIR range in Fig. 4.3 (b), where the best results are concentrated around 0.5 as expected. A drop in the quality of the reconstruction is again observed for higher transmittance values. Meanwhile, in the SWIR range, Figure 4.3 (c) shows that transmittances between 0-0.1 and 0.2-0.3 were obtained, which may be due to incorrect convergence due to initialization. In general, it is observed that the best reconstructions were obtained around 0.5 and, for higher values the quality decreases.

4.1.2. Compression ratio analysis

To observe the influence of the compression ratio, as defined in section 4.1, on the quality of the reconstruction, its value was varied from 5% to 50%, with steps of 5%. For each case, 5 repetitions were run, and their average result is here reported. It is important to note that a higher compression level simplifies the reconstruction task, since more information is obtained from the scene, at the cost of requiring a longer capture time and more storage. A lower compression level is a more challenging task, and more required in Compressive Sensing, since fast acquisition processes are desired, intended for applications as remote sensing or spectral video sensing.



(a) (b) (c)

Figure 4.4. Average PSNR obtained with different compression ratio in the (a) VIS, (b) NIR, and (c) SWIR ranges.

The reconstruction results in Figure 4.4 show that increasing the compression ratio yields to reconstruction quality improvement, indicating a reduction in the difficulty of reconstructing the image. In the three ranges, SWIR, NIR and VIS, it is observed that the best reconstructions are obtained with a compression level of 50% (PSNR of 27.82dB, 30.05dB and 32.98dB, respectively). In contrast, for the lowest compression level, i.e., 5%, a PSNR drop of up to 10 dB is observed. This

is mainly due to the difficulty of the inverse problem. Despite this drastic drop, a good quality of the images is maintained, considering the reduced measurements which in turn represents a faster acquisition.

4.2 Adaptive Single Pixel Imaging

Acquiring a complete set of measurements $\boldsymbol{Y_n}$ through SPI, it is necessary to acquire n snapshots, which introduces a trade-off between image resolution and acquisition time. Therefore, compressive sensing strategies are often employed to reduce the amount of acquired snapshots m, i.e., $m \ll n$ by modifying the sensing matrix $\boldsymbol{H}$ to exploit some assumptions of the spectral image, such as sparsity [4], low-rank [5], and non-local representations [6] among others. Current approaches can be split into two categories: Right and Left linear transformations.

Right linear transformation consists in redefining the sensing matrix $\boldsymbol{A}$ as the matrix product between the Hadamard matrix $\boldsymbol{H}$ and a decimation matrix $\boldsymbol{D}$, where the decimation matrix is designed to exploit some assumption on the structural information of the spectral image [7]. Therefore, the image model can be rewritten as

$$\boldsymbol{Y_m = HDX + \eta_m = AX + \eta_m}, \tag{5}$$

where η_m stands for the noise corruption inherent to the image acquisition system.

Left linear transformation modifies the sensing matrix $\boldsymbol{A}$ as the matrix product between a selection matrix $\boldsymbol{S}$, a permutation matrix $\boldsymbol{P}$ and the Hadamard matrix $\boldsymbol{H}$ to select specific rows of the Hadamard matrix, which aims to minimize or maximize some desired properties, such as total variation or cake-cutting on the selected rows [8], [9]. Therefore, the image model can be rewritten as

$$\boldsymbol{Y_m = SPHX + \eta_m = AX + \eta_m}. \tag{6}$$

State-of-the-art approaches focus on developing algorithms for the construction of the permutation matrix $\boldsymbol{P}$ to order the Hadamard matrix, also called Hadamard ordering approaches [10]. While the structure of the selection matrix $\boldsymbol{S}$ consists of a diagonal matrix with m-top elements of its diagonal with ones, and zero otherwise, representing the selection of m-top rows of the ordered Hadamard matrix $\boldsymbol{PH}$. It is important to note that the sensing matrix $\boldsymbol{A}$ is reduced to a m-rows matrix, resulting in a hardware compression ratio of m/n. In the following subsections, we will explain two methodologies which exploit this transformation to enable adaptive acquisition of spectral images from the NIR spectrum in the single pixel imaging system.

4.1.2 Left-Adaptive Hadamard Single Pixel

In the Left-Adaptive Hadamard Single Pixel approach, the design of the sensing matrix A is adaptive, utilizing prior information extracted from complementary acquisition systems, commonly referred to as side-information. In this context, the Hadamard sensing matrix undergoes modification through the incorporation of a decimation matrix, aimed at capturing the structural information within the scene. This modification serves to enhance image quality while reducing the requisite number of measurements [7]. For instance, authors in [11], [12] acquire a grayscale image from a side-information system to design the decimation matrix based on a super-pixel scheme. However, the design of the sensing matrix is detached from the recovery task which relies on a sub-optimal design to acquire spectral images in the NIR spectrum.

4.1.2.1 Proposed Left-Adaptive methodology.

The primary aim of the proposed methodology is to adaptively design the decimation matrix $\boldsymbol{D_\theta}$ by utilizing the information gathered from the VIS region and considering the NIR region through the integration of sensing and recovery processes in the spectral image. Through the integrated network, the sensing parameters are optimized using an end-to-end (E2E) approach. The side-information system involves a grayscale representation $\boldsymbol{X_g}$ of the scene obtained from the VIS spectrum. To achieve this, we employ the grayscale image $\boldsymbol{X_g}$ as the input to the network for adaptive decimation matrix design $\boldsymbol{D_\theta} \in R^{m \times n}$. The employed neural network architecture is the Spixelnet, a state-of-the-art fully convolutional neural network proposed in [13] for super-pixel segmentation. The network is trained using a supervised approach. Once the parameters $\boldsymbol{\theta}$

of the superpixel model $\mathcal{S}(\cdot)$ are fine-tuned, the mathematical description of the decimation matrix $\boldsymbol{D_\theta}$ design can be expressed as follows,

$$\boldsymbol{D_\theta} = \boldsymbol{\mathcal{S}}(\boldsymbol{X_g}, \boldsymbol{\theta}). \tag{7}$$

Then, the spectral image is used in the sensing scheme employing the decimation matrix as

$$\boldsymbol{Y_m} = \boldsymbol{H D_\theta X} = \boldsymbol{AX}. \tag{8}$$

It is important to observe that the structure of the decimation matrix $\boldsymbol{D_\theta}$ is influenced by $\boldsymbol{X_g}$. Consequently, the configuration of the sensing matrix $\boldsymbol{A} = \boldsymbol{H D_\theta}$ is uniquely designed for each spectral scene, adapting according to the specific side information. Furthermore, leveraging the orthogonal characteristics of $\boldsymbol{H}$, the acquisition and subsequent recovery of near-infrared spectral image can be represented using a single decimation process by the following fully differentiable operator.

$$\widehat{\boldsymbol{X}} = D(\boldsymbol{X}, \boldsymbol{D_\theta}) = diag\left(\frac{\mathbf{1}}{\boldsymbol{D_\theta 1}}\right) \boldsymbol{D_\theta}^\top \boldsymbol{D_\theta X}. \tag{9}$$

4.1.2.2 Training procedure

The entire methodology is trained within an E2E framework, where the parameters $\boldsymbol{\theta}$ of the superpixel model $S(\cdot)$ are adjusted based on a loss function denoted as L_{slic} , which is rooted in the SLIC-based [13] approach and incorporates both, a semantic loss term L_{sem} and a spatial compactness of superpixels term. The objective of the L_{sem} term is to confine the deep-superpixels $\boldsymbol{D_\theta}$ within a predefined super-pixel map $\boldsymbol{D_{slic}} \in R^{s \times n}$. The second term enforces spatial compactness of the superpixels, a strategy introduced in [13]. To achieve this, we apply the SLIC algorithm [14], to the grayscale image $\boldsymbol{X_g}$, resulting in the super-pixel map $\boldsymbol{D_{slic}}$= SLIC($\boldsymbol{X_g}$)**.** The selection of SLIC is motivated by its computational efficiency, operating at $\Theta(n)$ complexity, and its ability to regulate the number of superpixels, which distinguishes it from other superpixel algorithms. Consequently, the SLIC-based loss function can be formally defined as follows,

$$L_{slic} = L_{sem} + \tau_1 \|\boldsymbol{C} - \boldsymbol{D}(\boldsymbol{C}, \boldsymbol{D_\theta})\|_2^2, \tag{10}$$

$$L_{sem} = L_{cross}\left(D\left(\boldsymbol{D_{slic}^\top}, \boldsymbol{D_\theta}\right), \boldsymbol{D_{slic}^\top}\right), \tag{11}$$

where $\boldsymbol{C} = [x, y]$, i.e., $\boldsymbol{X}_{x(1+y)}$ represents a position of a pixel by its spatial coordinates. Based on [13], we used the cross-entropy loss L_{cross} as the distance measure. Additionally, to address the design of the decimation matrix $\boldsymbol{D_\theta}$ for the task of recovering the infrared spectral image, we incorporate a reconstruction loss from the non-iterative estimation, which is defined as follows,

$$L_{rec} = \|D(\boldsymbol{X}, \boldsymbol{D_\theta}) - \boldsymbol{X}\|_1. \tag{12}$$

Therefore, the complete objective function for finding the optimal set of network parameters $\boldsymbol{\theta}^*$ is given by.

$$\boldsymbol{\theta}^* = E_X[L_{rec} + \lambda L_{slic}]. \tag{13}$$

Note that the infrared information is only considered in the training step through the backpropagation of the recovery error. In contrast, the visible information is used to adaptively guide the training procedure, i.e., it is the input for the SLIC-based super-pixel method.

4.1.2.3 Results

Initially, we assessed key hyperparameters in the proposed methodology. Specifically, we focused on determining the number of superpixels denoted as m for the decimation matrix and the parameters represented by s for the SLIC-based loss. Our training configuration encompassed 5000 epochs, utilized a batch size of nine, and employed a learning rate of 10^{-3} with the Adam optimizer. This evaluation was carried out using the Washington D.C. Mall (WDCM) dataset, which comprises a high-resolution spectral image resized to dimensions of 1200x360. The dataset preprocessing stage consists of a patch-based format, by extracting 30 random patches. Subsequently, to construct training and testing sets, a subset random split was employed, with 27 images allocated for training and 3 images designated for testing. It is important to highlight that multiple random split trials were performed until achieving a test set that captured the global characteristics of the

training set, in alignment with the approach detailed in Zhu et al. [15]. During the training process, data augmentation strategies, including random horizontal and vertical flips, were implemented to mitigate overfitting. The assessment of the quality of recovered spectral images was performed using three metrics: Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), and Spectral Angle Mapper (SAM).
We compare our proposed methodology with state-of-the-art HSI-based sensing approaches, all operating under the same compression level. This comparison encompasses the following techniques:

- **Uniform**: The uniform spatial decimation design, which lacks scene side-information.
- **Designed**: The designed approach [1], involving optimization of a single decimation matrix for the entire dataset. sampling
- Ordering approaches, including **Cake-Cutting** [9] and **Zigzag** [8] methods for selecting the first m rows of an n-order reordered Hadamard matrix.
- **Sparse**: A sparse method that selects relevant coefficients from the Hadamard spectrum of the grayscale acquisition X_g as modulation patterns for spectral image acquisition [4].
- **SLIC**: The SLIC approach [11], directly employing the SLIC-based decimation matrix of m superpixels.

Quantitative results are presented in Table 4.1, demonstrating the superior performance of the proposed method. In the case of the 225-order Hadamard matrix, the proposed approach outperforms state-of-the-art designs in up to 0.79 dB of PSNR. Additionally, visual results depicted in Figure 4.5 showcase the efficacy of our adaptive design in accurately preserving the structural information of the infrared spectral image for the WDCM dataset.

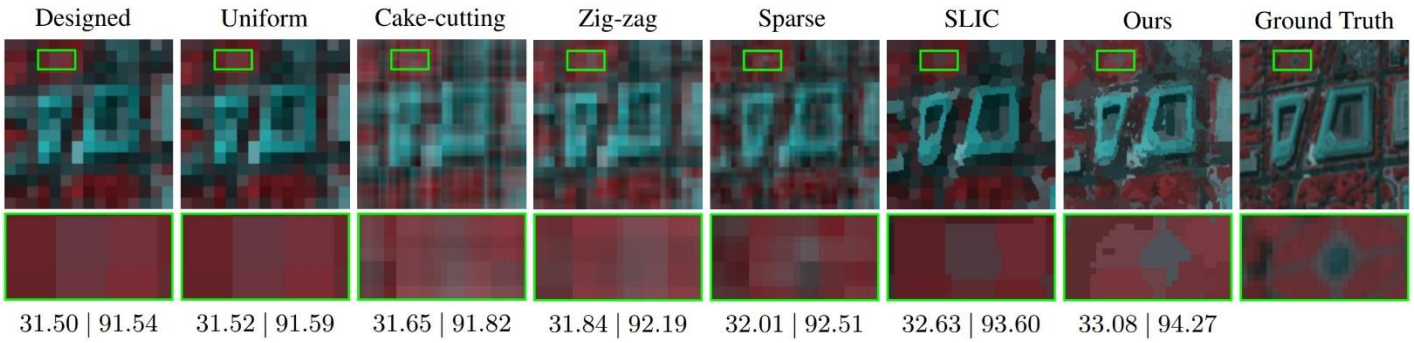


Figure 4.5. Visual comparison with state-of-art sensing design, $m = 225$. Each column corresponds to a false color visualization of the spectral image for each sensing design; the last column corresponds to the reference spectral image. The PSNR and SSIM scores are displayed under each sensing design.

Table 4.1. Comparison results, best and second-best results highlighted in **bold** and underlined 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>
	Ours	**33.56**	**95.00**	**0.125**

400	Designed	32.13	92.96	0.125
	Uniform	32.22	93.13	0.124
	Cake-Cutting	32.42	93.37	0.137
	Zig-Zag	32.79	93.95	0.126
	Sparse	33.02	94.22	0.146
	SLIC	33.67	95.14	0.120
	Ours	**33.83**	**95.31**	**0.116**

*The presented results and further findings were published in the **ICASSP 2023** Conference under the paper titled "**Deep Adaptive Superpixels for Hadamard Single Pixel Imaging in Near-Infrared Spectrum**" https://doi.org/10.1109/ICASSP49357.2023.10095165.

4.1.3 Right-Adaptive Hadamard Single Pixel

By selectively choosing rows from the Hadamard matrix, the SPC system can efficiently capture the crucial information from the input signal while minimizing the number of required measurements or snapshots. Fors instance, we provided the implementation of three state-of-the-art methods for ordering Hadamard matrices in both one-dimensional and two-dimensional forms was conducted, as presented in Figure 4.6. The aim of this implementation was to establish a unified working framework to validate distinct ordering strategies as an initial approach for selecting the first subset of coded apertures to be acquired. Among the noteworthy ordering methods are frequency ordering, block count ordering (Cake-Cutting) [9], and total variation ordering (ZigZag) [8]. These methods hold significant recognition within the scientific community and are considered benchmarks in the field of Hadamard matrix ordering. In all instances, priority is given to selecting modulation patterns linked to lower frequencies, with patterns associated with higher frequencies being incrementally chosen. However, the geometric characteristics differ based on the chosen algorithm. For instance, one such strategy is the Zig-zag method, which follows the rows selection of the Hadamard matrix in a zig-zag pattern. Alternatively, the Cake Cutting approach rearranges the modulation patterns based on an ascendancy ordering of internal block counts. These methods are all grounded on the principle that the upper left portion of the matrix, where the low frequencies are preserved, holds the most relevant information. This observation is substantiated by the mean spectrum values displayed in Figure 4.7, obtained from two well-known datasets, EuroSat and Kaist. By selecting rows primarily from the upper left part of the Hadamard matrix, where the low-frequency components prevail, the SPC system can efficiently capture the significant information of the input signal. This results in compressed sensing advantages such as reduced data storage requirements, faster acquisition times, and decreased power consumption.

However, this assumption is based on the premise that the low-frequency components are consistently concentrated within the same region of the Hadamard spectrum for all sensed images. Unfortunately, this does not hold true for every image. As demonstrated in Figure 4.7, the average position of low-frequency components varies across different datasets. Even within a single dataset like KAIST, as shown in Figure 4.8, the spectrum undergoes significant changes from one image to another. Consequently, it becomes imperative to individually select different sensing matrices $\boldsymbol{S}$ for each image to achieve optimal performance. An effective strategy is to adaptively determine the sensing matrix as the Single Pixel Camera (SPC) operates sequentially. By capturing a few snapshots, a decision can be made for the subsequent measurements. The forthcoming section introduces the proposed methodology for an adaptive design approach.

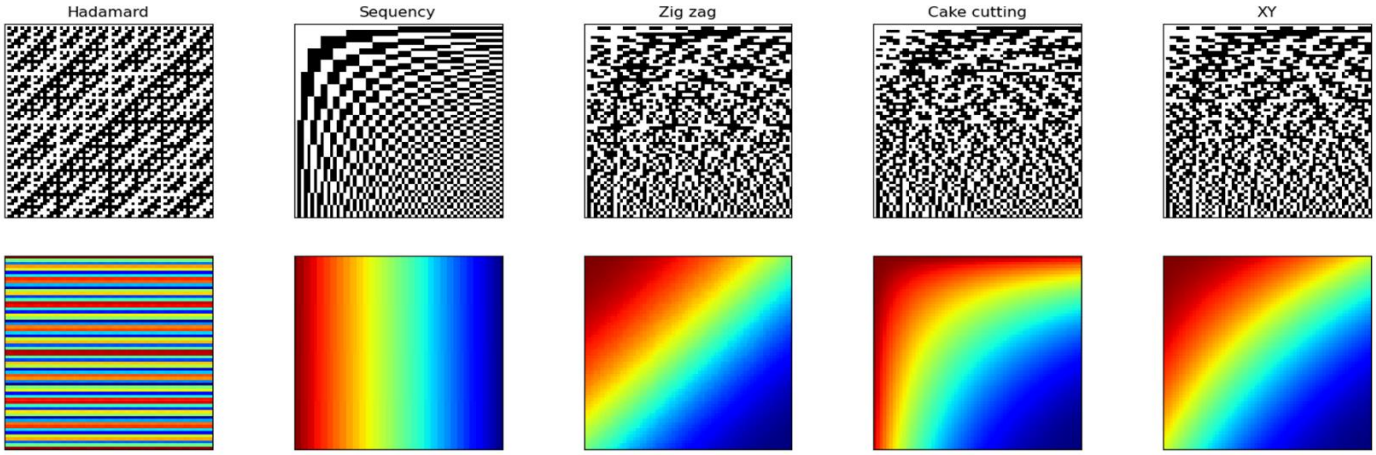


Figure 4.6. Illustrates various sensing strategies in SPC. **(Top)** visually represents different strategies such as random Hadamard, sequentially Hadamard, zigzag, cake cutting, and XY strategy. **(Bottom)** represents the order of rows in the original Hadamard matrix. The color scheme used in this representation assigns red to the first sensing value and blue to the last sensing value, providing a visual indication of the row ordering.

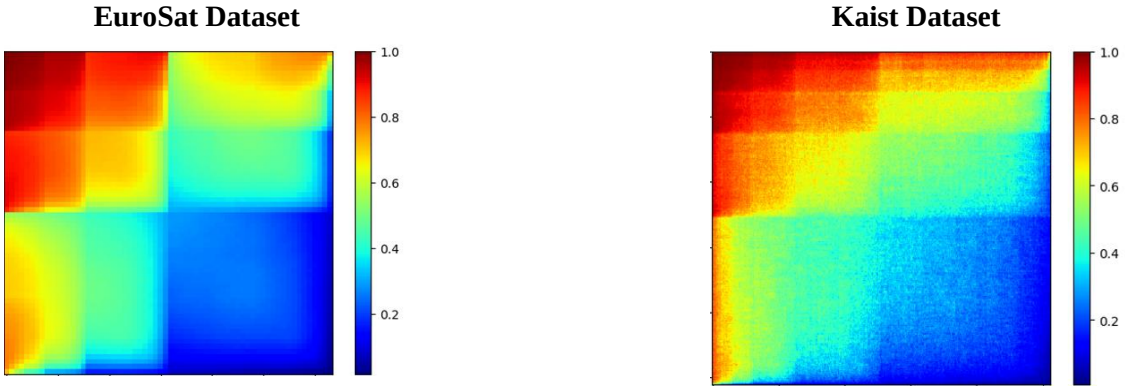


Figure 4.7. Spectral analysis of the Hadamard coefficients for the Eurosat and Kaist datasets. Red values represent the most relevant coefficients while blue represents the least relevant. A different distribution is observed in both datasets.

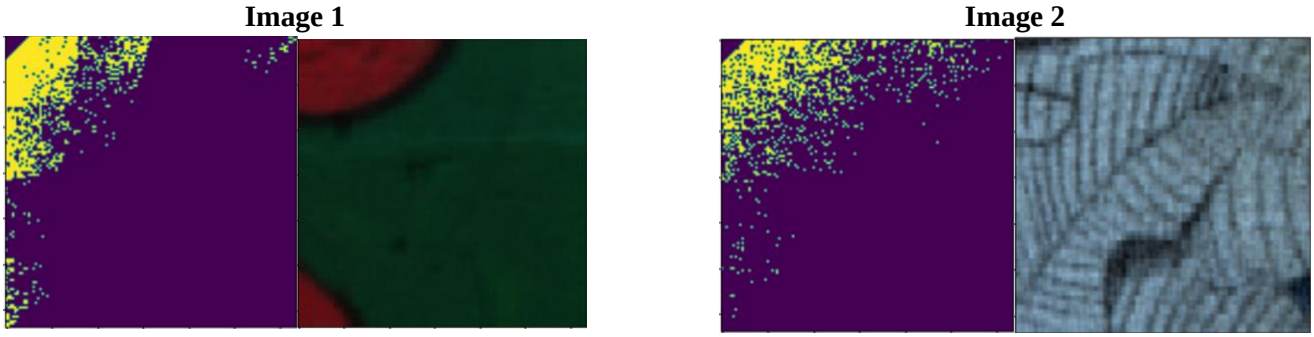


Figure 4.8. Spectral analysis of the Hadamard coefficients for two different images from the Kaist dataset.

4.1.3.1 Proposed Adaptive Methodology

The proposed adaptive methodology is illustrated in Figure 4.9, it involves a structured approach comprising five stages. Firstly, it initiates with a fixed structure based on the zig-zag methodology [8]. Following the initial acquisition, a quick estimate of the scene is generated. This estimate is then utilized to predict the new patterns to be acquired using a neural

network in the third stage. Subsequently, the newly predicted patterns are acquired. Lastly, both, the initially acquired measurements and the newly acquired measurements are combined to obtain the final estimate of the scene. This adaptive methodology ensures an iterative and refined approach for scene estimation, leveraging the benefits of initial estimation and subsequent pattern predictions for improved accuracy and quality of the final scene reconstruction. Specifically, it consists of the following stages:

1. *First fixed acquisition:* in this step we acquire a first set of observation $\boldsymbol{Y_1}$ based on a pre-defined sensing path $\boldsymbol{S_1 P}$ given by the zig-zag methodology.

$$\boldsymbol{Y_1} = \boldsymbol{S_1 P H_n X} + \boldsymbol{\eta_m}, \tag{14}$$

2. *Fast reconstruction:* we perform a fast reconstruction of the spectral image,

$$\boldsymbol{X_1} = \boldsymbol{H_n^\top Y_1}, \tag{15}$$

3. *Estimation of the next rows of the Hadamard:* exploiting the high-level features acquired from the first acquisition we employ a deep neural network to estimation the second set of modulation patterns to be used,

$$\boldsymbol{S_2} = S_\theta(\boldsymbol{X_1}) = (1 - \boldsymbol{S_1}) \cdot diag(\, sigmoid\big(\boldsymbol{S_{\theta:-1}}(\boldsymbol{X_1})\big)\,), \tag{16}$$

where S_θ represents a deep neural network model with θ as the trained parameters**.**

4. *Second Adaptive acquisition:* acquire the second set of adaptive coefficients based on the estimated selection matrix S_2.

$$\boldsymbol{Y_2} = \boldsymbol{S_2 P H_n X} + \boldsymbol{\eta_m}, \tag{17}$$

5. *Final reconstruction Step:* Since both fixed and adaptive acquisitions are related to the same Hadamard basis, we perform the fast reconstruction of the whole coefficients, as follows.

$$\widehat{\boldsymbol{X}} = \boldsymbol{H_n^\top}(\boldsymbol{Y_1} + \boldsymbol{Y_2}\,)\,. \tag{18}$$

This two-stage sensing, and recovery process is unrolled in a single network, and with a spectral dataset, the parameters of the S_θ can be adjusted by solving the following optimization problem,

$$L\big(\boldsymbol{X}, \widehat{\boldsymbol{X}};\, \boldsymbol{S_\theta}\big), \tag{19}$$

where $\mathcal{L}$ is a distance metric which aims to measure the error between the reference images and the image estimation from adaptive methodology sensing. In this report we aim to analysis various alternatives to adjust the network parameters by exploiting different perspective for the training of adaptive estimation coefficients, from a conventional regression perspective and from a classification perspective. For this we define the following training cost functions

- The **MSE** cost function involves computing the Euclidean norm between the image estimations and a reference, with the introduction of a transmittance regularizer. This regularizer is introduced to ensure the predefined sampling ratio.

$$MSE = \big\|\boldsymbol{X} - \widehat{\boldsymbol{X}}\big\|_{\boldsymbol{2}}^{\boldsymbol{2}} + (\mathrm{E}[\boldsymbol{S_2}] - \delta)^2$$

- The **BCE** cost function, which consists in view the adaptive selection coefficients as a classification problem, where the k-top coefficients are pre-defined by the k-top ordering of these coefficients in the Hadamard spectrum. Therefore, the training cost function consists in learn this adaptive ordering by the neural network as a classification problem, employing the binary cost entropy cost function between the selection matrix estimation and the k-top binary selection matrix defined by the sparse representation of the reference images.

$$BCE = BCE(\boldsymbol{S_2}, \mathrm{ktop}(|\mathbf{HX}|))$$

- The **Order** cost function, like the BCE cost function, focuses on learning the relevance of selection coefficients based on their distribution within the sparse representation. In this context, the neural network's objective is to directly acquire knowledge of the index of sorted coefficients. It utilizes the MSE and MAE cost functions to minimize the disparity between the estimated ordering index and the one predefined by sorting the Hadamard coefficients by magnitude.

$$MSE\ Order = \|S_2 - argsort(\mathbf{HX})\|_2^2$$
$$MAE\ Order = \|S_2 - argsort(\mathbf{HX})\|_1$$

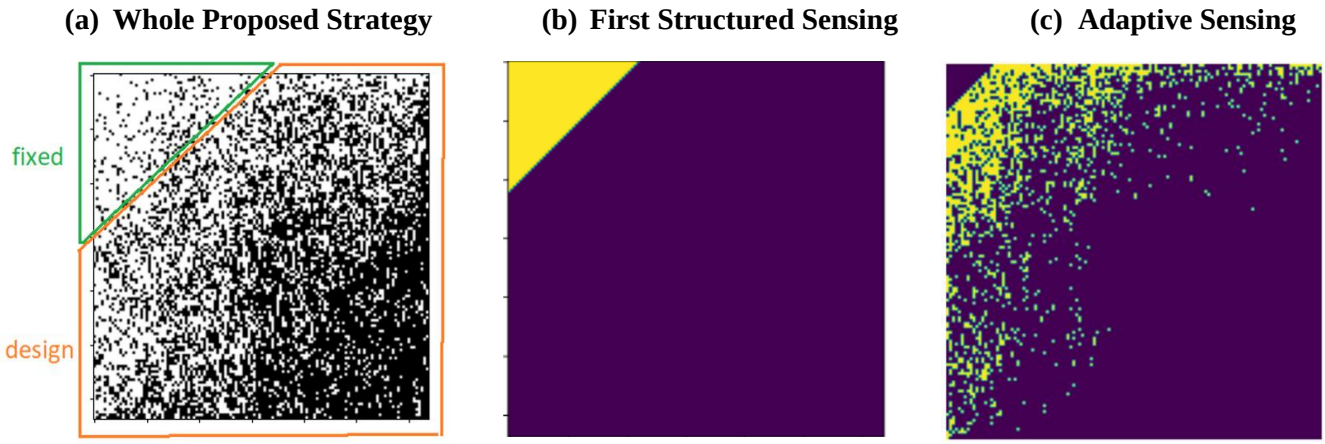


Figure 4.9. (a) Proposed strategy example for the Hadamard coefficient analysis of images from the Kaist dataset. (b) The proposed method starts with a fixed structure using the zig-zag methodology. (c) An adaptive approach is utilized to estimate the missing details. This adaptive step aims to enhance the analysis by incorporating additional information and refining the coefficient representation.

4.1.3.2 Reconstruction results

To assess the impact of the different training strategies in the quality of reconstructed spectral images, the zigzag ordering technique was employed as the acquisition method for the initial set of coefficients. Furthermore, for measure reconstruction quality in a 10% sampling scenario, which involves a subset of 409 coefficients from a total of 4096 possible coefficients, evaluation metrics such as PSNR, SSIM, and SAM were chosen. Figure 4.10 illustrates the outcomes under various configurations, with the x-axis representing the number of coefficients captured in the first instance using zigzag ordering. It becomes apparent that the use of the adaptive acquisition system significantly enhances the quality of acquired images within the same sampling scenario for all training strategies.

Figure 4.10. Influence of training strategy for different subsampling ratios.

Furthermore, it can be noted that training with the binary cross-entropy (BCE) and mean square error (MSE) regression strategies yielded the most favorable outcomes in terms of the PSNR metric. Conversely, for SSIM and SAM metrics, the binary cross-entropy (BCE) strategy and regression strategies utilizing l_1 and l_2 normalization on the ordered index matrix (mean absolute error ordering, MAE order, and mean square error ordering, MSE order) produced the most promising results. This suggests that binary classification strategies hold promise for adaptively selecting the most significant coded coefficients during the reconstruction of compressed spectral images. Notably, these strategies displayed signs of overfitting compared to the transmittance regression strategy, which is commonly employed in E2E training protocols.

As mentioned previously in Section 4.2.2, the geometric characteristics of the sensing path described in the Hadamard spectrum differ based on the chosen algorithm, which directly impacts the final image quality for each specific compression scenario. To quantitatively analyze these discrepancies, an examination of the ordering strategies across various sampling

rate scenarios is presented in Figure 4.11, accompanied by PSNR, SSIM, and SAM metrics for each reconstruction. The zigzag and XY ordering strategies exhibit almost identical behavior across most sampling rate scenarios for the three chosen metrics. Yet, variations in performance emerge for sampling values exceeding 30%, where the XY strategy demonstrates superior PSNR performance compared to the zigzag strategy. Regarding the cake-cutting strategy, its underperformance is evident at sampling values below 25%, yet it outperforms at values beyond 25%. This 25% threshold marks a turning point in the behavior of the three strategies.

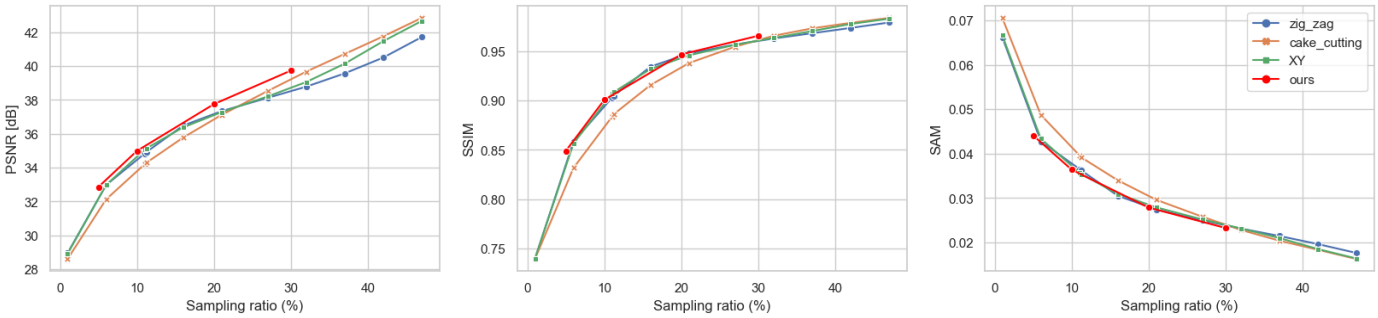


Figure 4.11. Performance of the different ordering strategies for different sampling rate values. Each graph corresponds to the performance under the PSNR, SSIM and SAM metrics used for the evaluation of the reconstructed images, from left to right, respectively.

These findings allow us to discern how prevalent fixed sampling strategies exhibit limitations concerning their efficacy in relation to the type of data to be acquired and the specific sampling scenario of application. It is important to emphasize that to date, a fixed strategy cannot universally guarantee optimal performance across most scenarios. Lastly, Figure 4.12 offers visual insights into employing the three ordering strategies on a chosen image. Reconstruction quality outcomes are reported based on PSNR and SSIM metrics for sampling scenarios of 1%, 5%, 7%, 10%, 15%, and 20%. Like Figure 4.11, images estimated using zigzag and XY ordering strategies exhibit closely aligned values and comparable structural characteristics in the chosen metrics. However, notable differences arise with the cake-cutting ordering strategy in low sampling rate scenarios. As the sampling rate escalates, a reduction in the quality disparity between the three sorting strategies becomes evident, both quantitatively and visually, indicating a more consistent behavior at higher sampling rates.

Finally, Figure 4.13 showcases visual outcomes that facilitate a comparison between the acquired learned selection matrix S_2 and the most relevant Hadamard coefficients within the sparse representation. Across diverse input spectral images, the deep neural model effectively approximates a coefficient distribution that closely aligns with the expectations set by the sparse representation. Notably, these distributions exhibit significant distinctions for each scene, underscoring the adaptive nature inherent in our proposed methodology. Furthermore, we offer a visual contrast between the ground truth images and the reconstructed images, providing tangible evidence of the exceptional quality achieved in visual reconstructions.

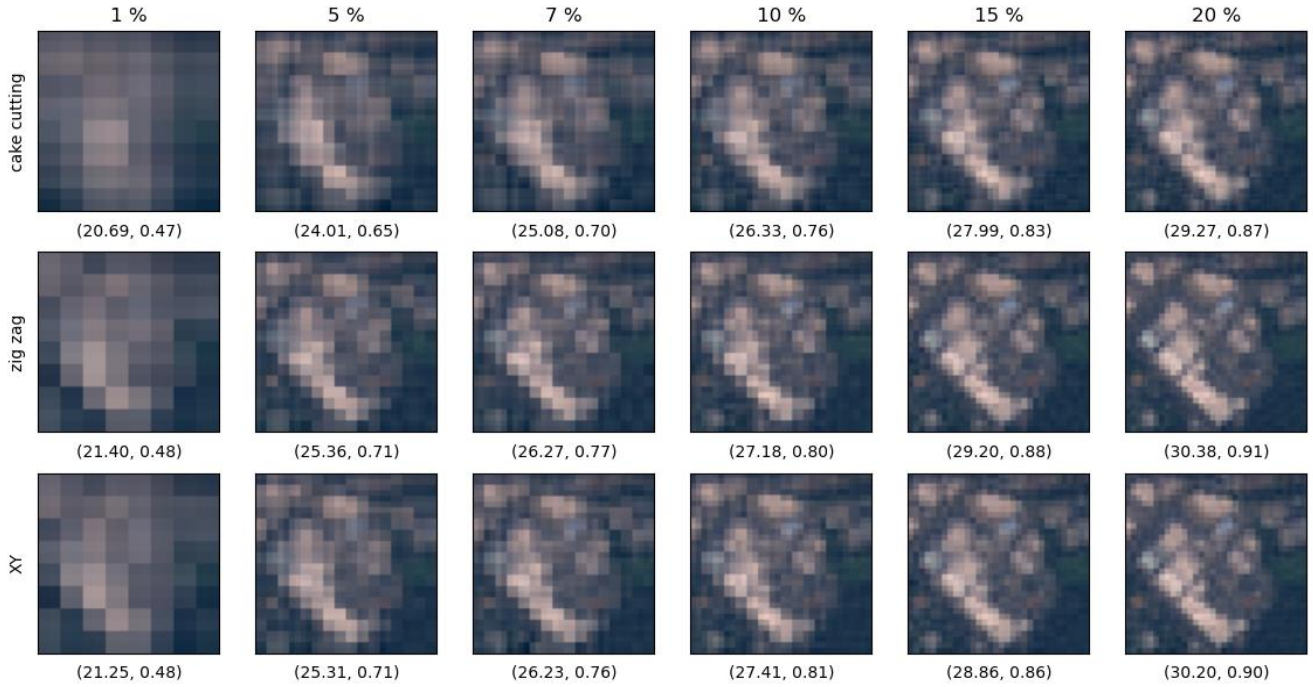


Figure 4.12. Visual comparison of ordering Hadamard basis algorithms under different compression scenarios.

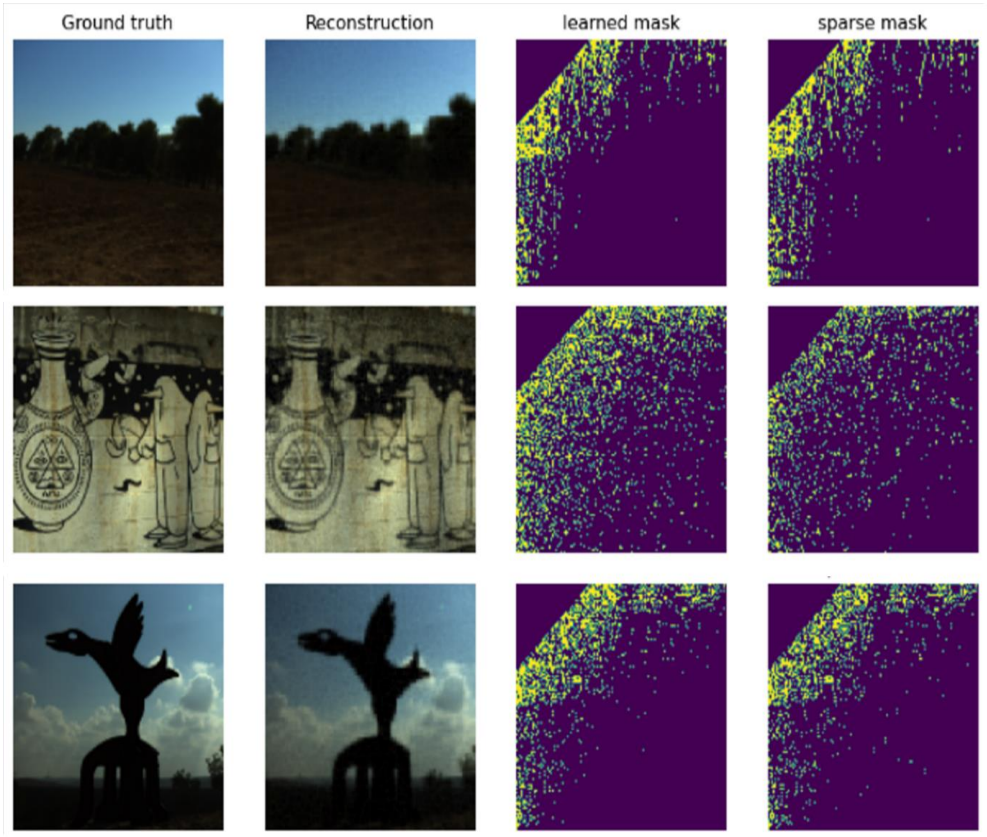


Figure 4.13. Visual results of the proposed adaptive sensing methodology for three different images of the Arad Dataset. Learned Mask correspond to the estimated selection matrix S_2 and sparse mask refers to the k-top binary selection matrix of the most relevant Hadamard Coefficients $ktop(|HX|)$

4.3. Results from real data of SPC implementation

This section evaluates the proposed method with experimental data obtained in the HDSP optics lab with the testbed implementation shown in Figure 4.14. The optical setup uses multiples lens LB1471-ML, a dichroic beam splitter, an optimized NIR DMD LC4500-NIR-EKT, two broadband mirrors PF10-03-P01, a collimator lens from ocean insight, a raptor owl 640T, an optical fiber and a NIRQUEST spectrometer. Remarking that the spectrometer is the main spectral sensor, which defines the system as a single-pixel camera. Specifically, the implementation allows the acquisition of spectral images in a maximum spatial resolution of 1280x800 with 512 spectral bands up to 2500[nm] in the NIR range. To illustrate the measurements captured using the implemented system, with binary {-1,1} coded apertures, Figure 4.15 presents the first 256 measurements for an arbitrary scene. On the other hand, the algorithm presented in section 4.2 was tested employing measurements acquired with our implemented SPC optimized for the NIR spectral range, which is shown in the following subsections. It is important to remark that to calculate the metrics, for each scene and approach, a ground truth was also acquired by employing a full-rank sensing matrix for SPC.

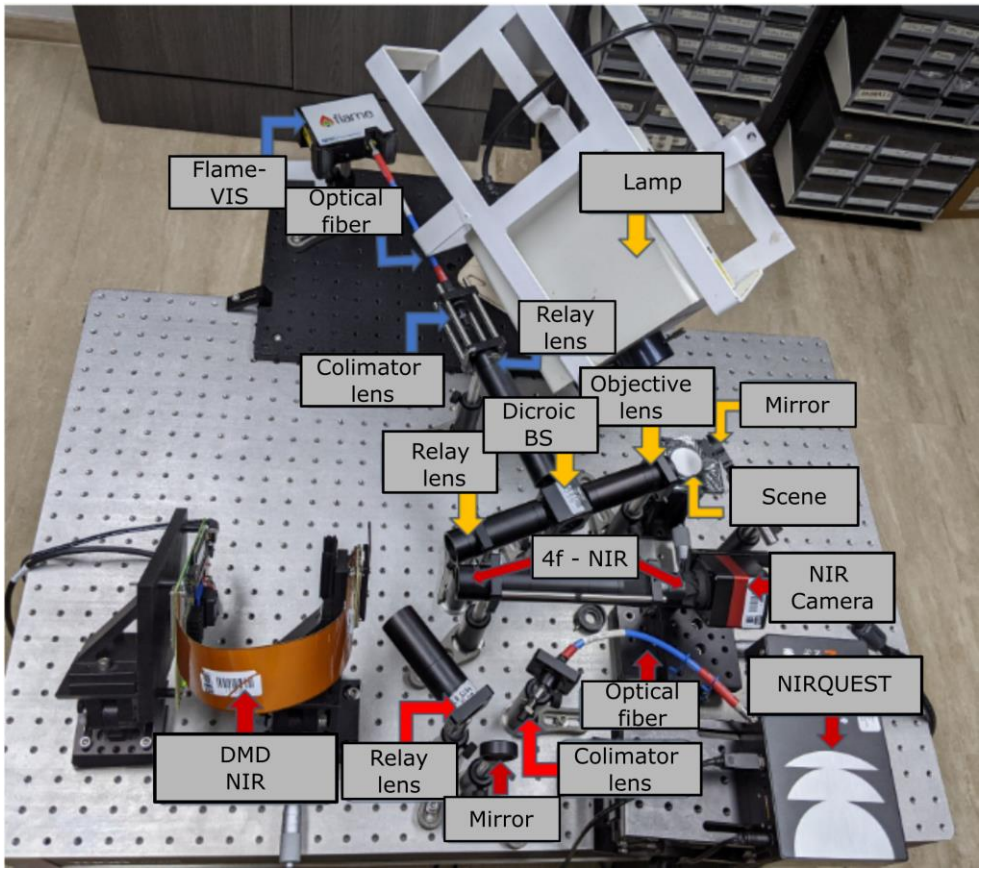


Figure 4.14. Testbed implementation of the NIR SPC imaging system.

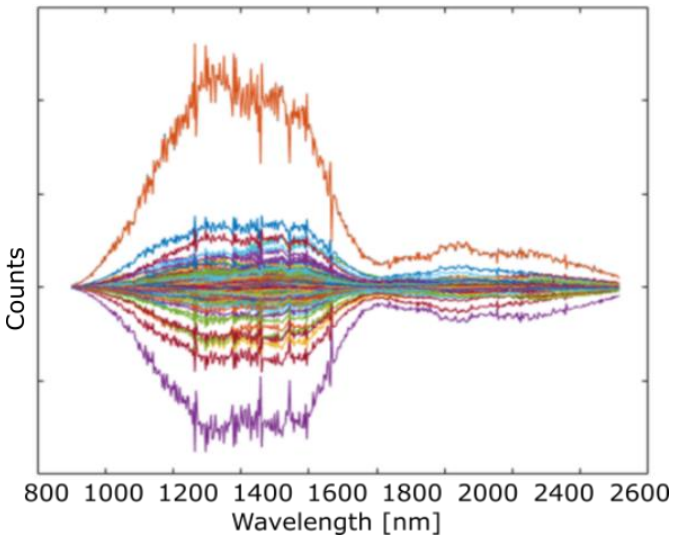


Figure 4.15. Captured measurements examples employing the SPC optical system for 256 binary {1,-1} coded apertures, each different color represents a different measurement.

4.3.1. Adaptive decimation approach testing with real SPC NIR data.

In this section, we experimentally validate the adaptive decimation sensing methodology described in Section 4.2 using real data. To achieve this, we employ the test bed implementation of the single-pixel imaging acquisition system in the NIR spectrum from Figure 4.14. Initially, we acquire a grayscale (panchromatic) image from the NIR spectrum, which is then employed for the adaptive estimation of super-pixel maps. These maps are subsequently used in the coded aperture designs to acquire the decimated Hadamard measurements. Visual and quantitative comparisons for the three NIR scenes are presented in Figures 4.16, 4.17, and 4.18. For each scene, we assess spatial and spectral quality using metrics such as Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Spectral Angle Mapper (SAM). Additionally, we select two points for each scene to display their spectral reflectance for both the recovered spectral images and the reference image acquired from full sampling of the Hadamard basis.

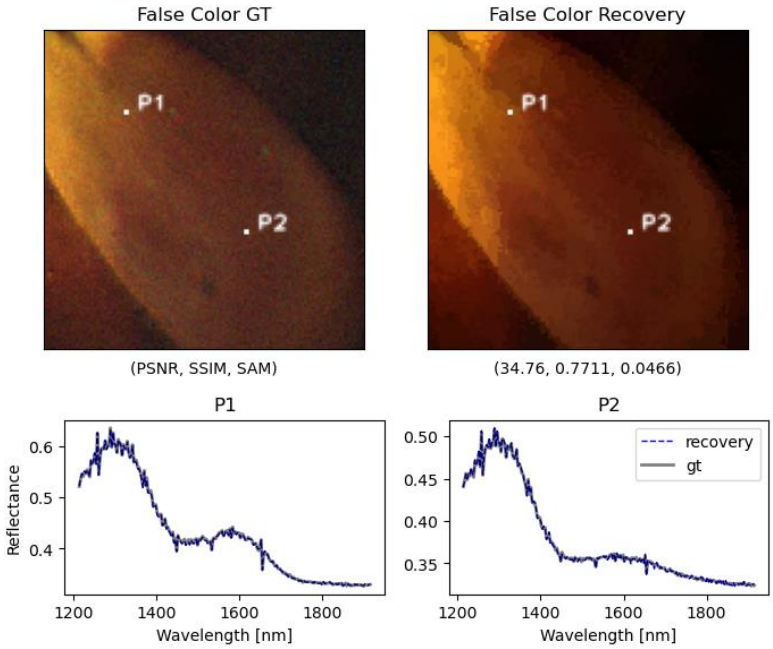


Figure 4.16. Visual and quantitative comparison of full Hadamard reconstruction, assumed as ground truth (GT), and adaptive superpixels for Scene 01.

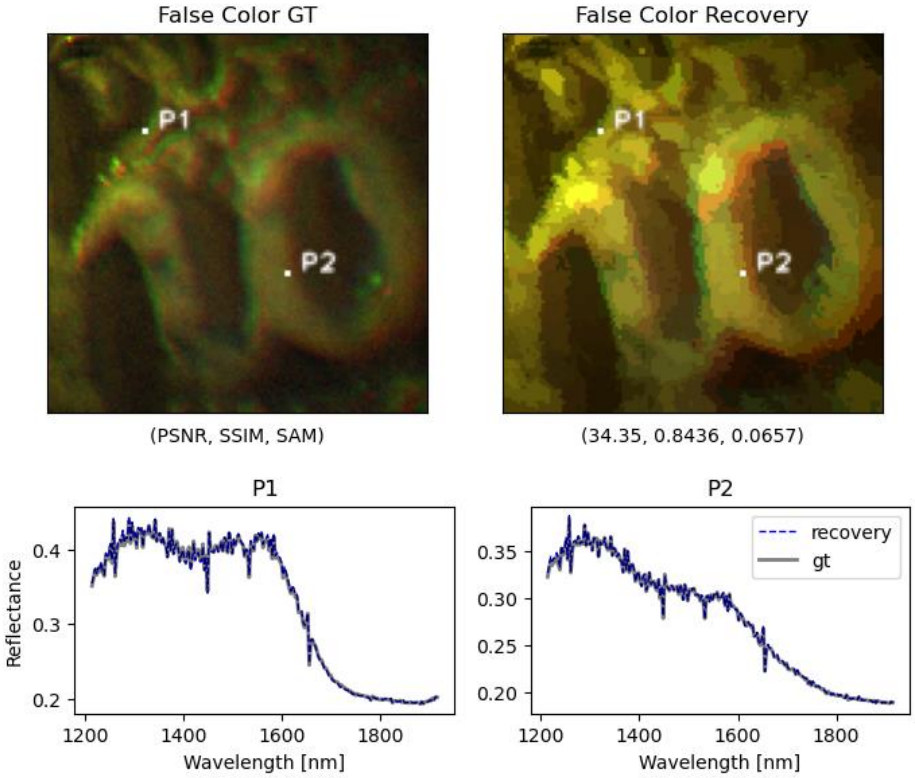


Figure 4.17. Visual and quantitative comparison of full Hadamard reconstruction, assumed as ground truth (GT), and adaptive superpixels for Scene 02.

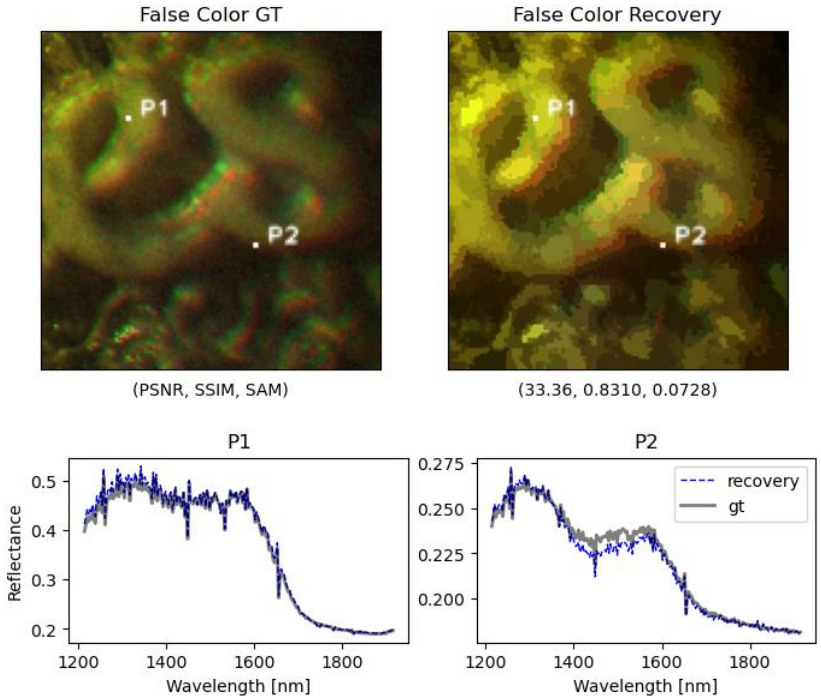


Figure 4.18. Visual and quantitative comparison of full Hadamard reconstruction, assumed as ground truth (GT), and adaptive superpixels for Scene 03.

4.3.2 E2E coded aperture design approach testing with real SPC NIR data.

In this section, we experimentally validate the End-to-End methodology described in Section 4.2 using real scenes. To achieve this, we utilize the test bed implementation of the single-pixel imaging acquisition system in the NIR spectrum shown in Figure 4.3.1. We selected one scene for testing, and varied the binary coded apertures (CAs) learned in simulations for SWIR, NIR and VIS. The metrics used to measure the reconstruction quality are PSNR, SSIM and MSE. Note that in this case, the grayscale representation is shown, since only one band of the spectrum is reconstructed.

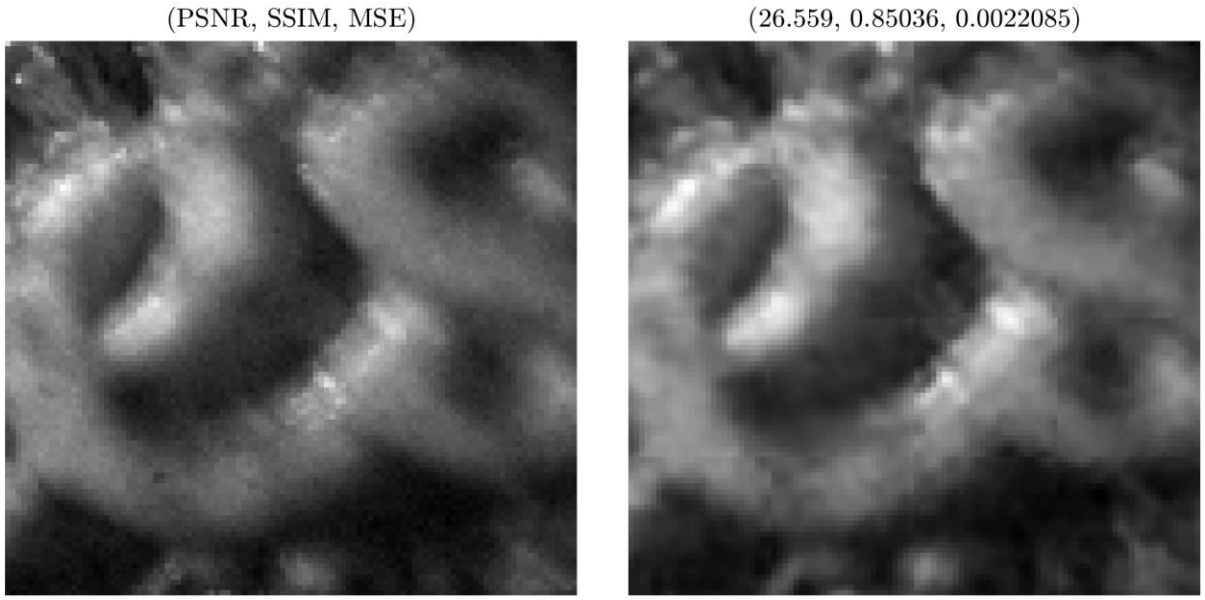


Figure 4.19. Visual and quantitative of (Left) full Hadamard reconstruction, assumed as GT, and (Right) End-to-End methodology with SWIR CAs.

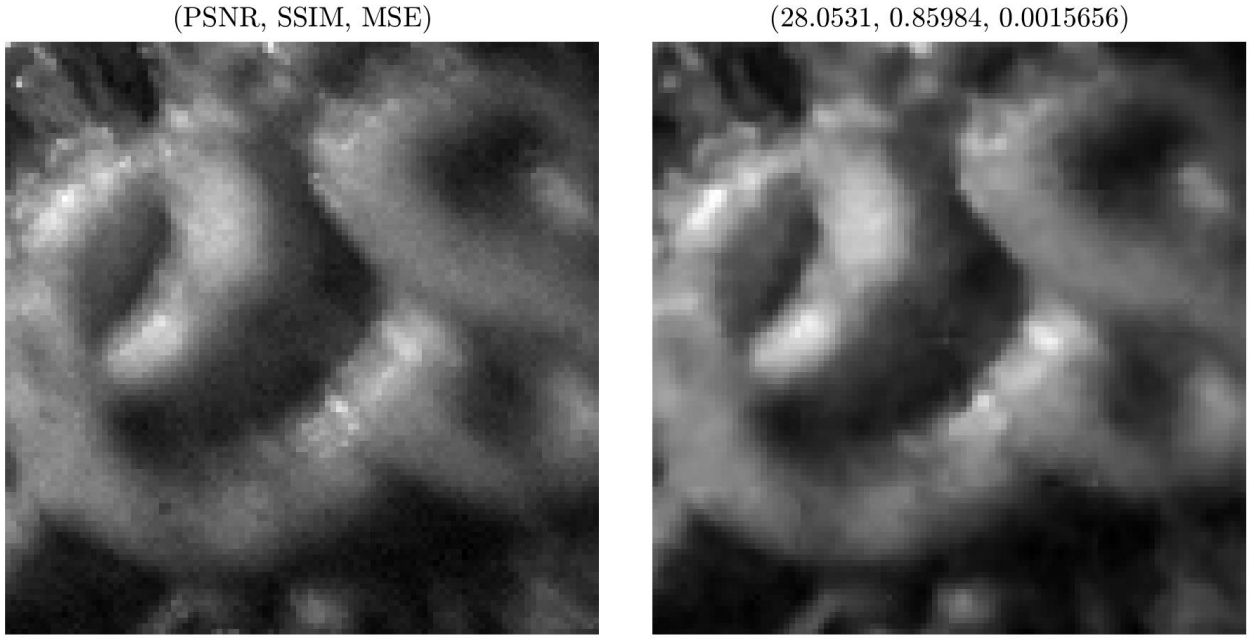


Figure 4.20. Visual and quantitative of (Left) full Hadamard reconstruction, assumed as GT, and (Right) End-to-End methodology with NIR CAs.

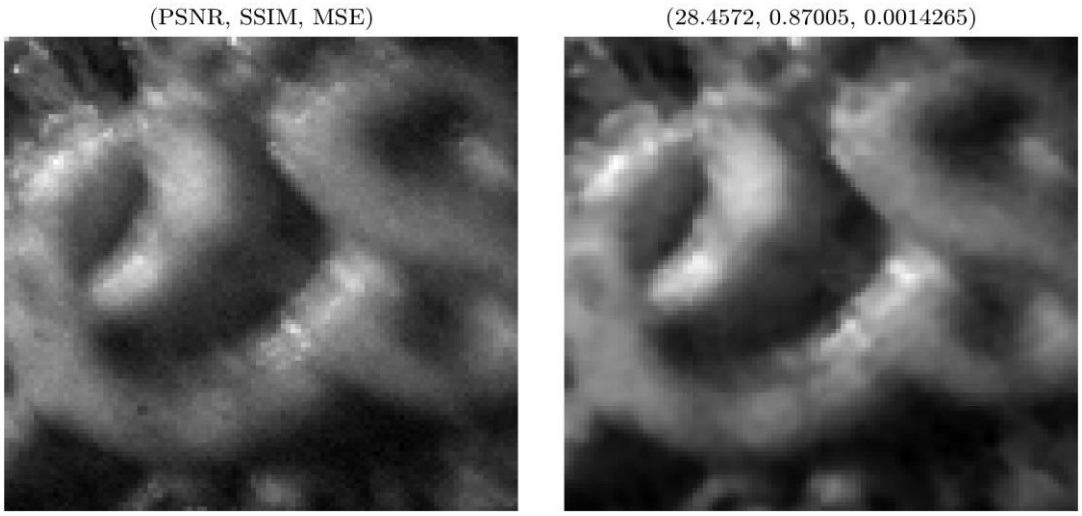


Figure 4.21. Visual and quantitative of (Left) full Hadamard reconstruction, assumed as GT, and (Right) End-to-End methodology with VIS CAs.

4.3.3 Testing a classification material approach with real SPC NIR data.

The proposed approaches to reconstruct NIR spectral images from SPC measurements can be extended to a classification task, specifically, the proposed classification strategy in the NIR region was tested in the implementation of the SPC presented in Figure 4.3.1. This prototype was described in Section 4.3 to obtain 512 spectral values in the wavelength range 900nm- 2500nm. To develop the classification task in our approach, first, a reconstruction step is performed with the algorithms presented in Chapter 4, after the reconstruction step a classification network is employed for training, which consists of three convolutional layers with batch normalization and rectified linear unit (ReLU) as the activation function, followed by a dropout to avoid overfitting, and finally, two fully connected layers, where the last one has the SoftMax activation. This network was trained using 10% of the total extracted spectral features. The experiment evaluated the "White" scene composed of four homogeneous color materials: salt, milk powder, sugar, and bicarbonate. These materials were selected to show the effect of the NIR region. An RGB picture of the White target acquired with a cellphone camera and the ground truth map is illustrated in Figure 4.22. We evaluate different compression ratios {99.6%, 98.43%, 93.75%, 75 %}, respectively. Figure 4.22(a). shows one false mapping of the NIR extracted features and Figure 4.22(b). reports the classification map using the trained network. The increase in the sub-sampling ratio allows us to obtain a more detailed classification finding 32 as the optimal value. Furthermore, notice that 64 provided more detailed features; however, the classification is not the best due to the variability increase.

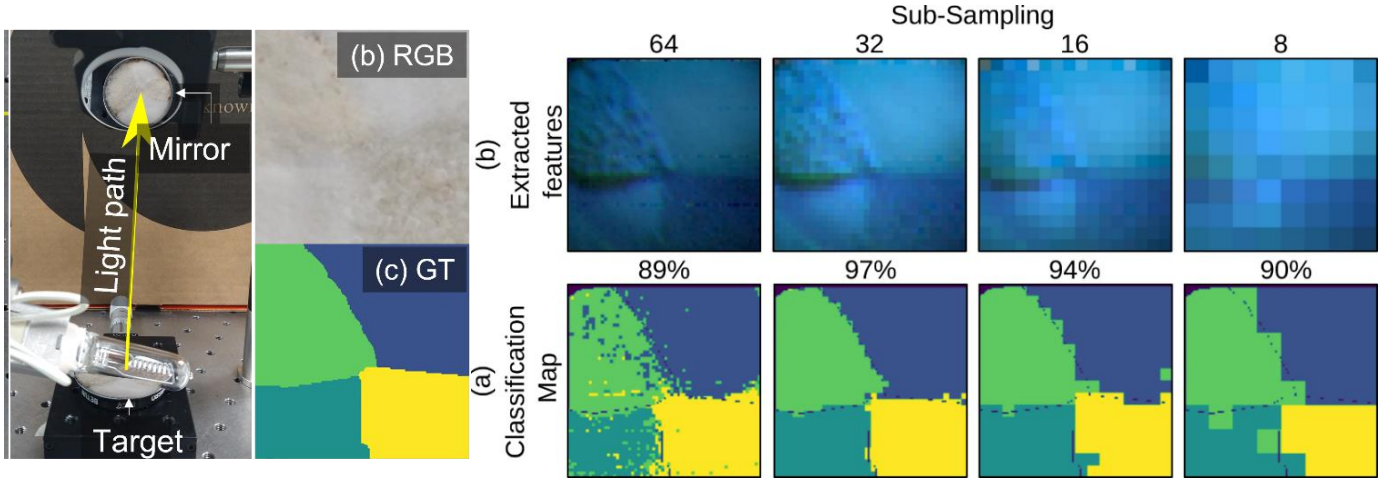


Figure 4.22. (a) Features extracted directly from the SP measurements in the NIR range, (b) Classification map obtained for different subsampling scales. Recovered from [16].

5. Color Coded Aperture Modeling for Classification Task

The color-coded aperture (CCA) is an optical filter matrix as illustrated in Fig. 5.1 Depending on the manufacturing procedure, there are two main optical architectures. The first approach involves using dichroic optical filters, which permit the passage of multiple wavelengths. However, constructing such a filter matrix is challenging, as it requires combining different materials in the same pixel. The second approach employs specific optical filters (band selection), which allows only a particular wavelength to pass through. This type of CCA involves fewer manufacturing complexities and is commercially available in numerous spectral imaging cameras. Therefore, according to these two technologies, some mathematical and optical models are proposed to design the coded aperture. It should be noted that both types of CCA were trained following the E2E methodology explained in the previous report.

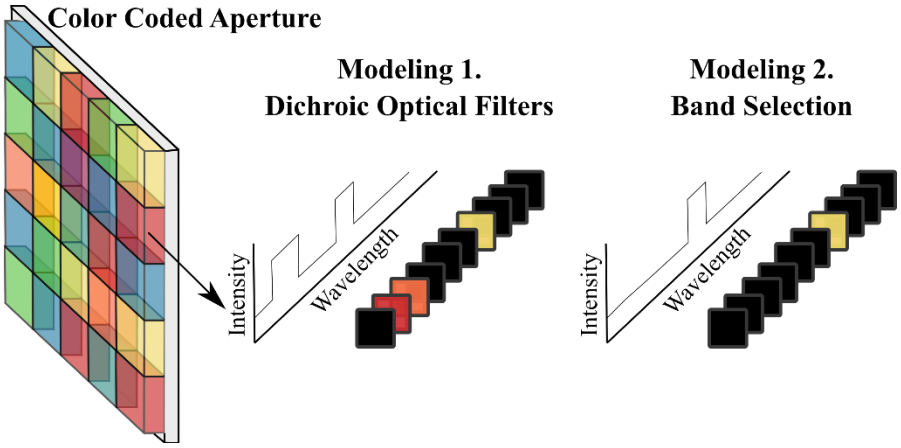


Figure 5.1. Visual representation of a Color-Coded Aperture CCA. There are two main modeling approaches depending on the manufacturing procedure, 1) composed of dichroic optical filters and 2) of band pass filters.

5.1 Color Coded Aperture Modeling Composed of Dichroic Optical Filters

The proposed method models the color-coded filter array system as a fully differentiable optical encoder. Considering the rearrange step, the forward sensing model can be treated as a fully connected layer, where the number of patterns is equivalent to the number of neurons as illustrated in the sensing stage of Figure 5.2. Therefore, the optical system can be modeled as an optical encoder,

$$\boldsymbol{Y} = \boldsymbol{HX} \coloneqq M_{\phi}(\boldsymbol{X}), \tag{20}$$

where ϕ denotes the arrangements of dichroic optical filters, which are the learnable parameters in the fully connected model M_{ϕ}. Notice that the main difference with the traditional fully connected layer is that the entries of the coding patterns must be binary. This binary constraint can be addressed via the inclusion of a binary regularization $R_{\rho}(\phi)$ in the E2E optimization [17]. In particular, the regularization function is included in the optimization problem as:

$$R_{\rho}(\boldsymbol{\phi}) = \frac{1}{n}\sum_{l=1}^{n}(\boldsymbol{\phi}_l)^2\,(\boldsymbol{\phi_l} - 1)^2, \tag{21}$$

which is minimized when the elements of the coding patterns are either 0 or 1. The proposed E2E model, which simultaneously learns the CCA, and the parameters of the classification network, is summarized in Figure 26.

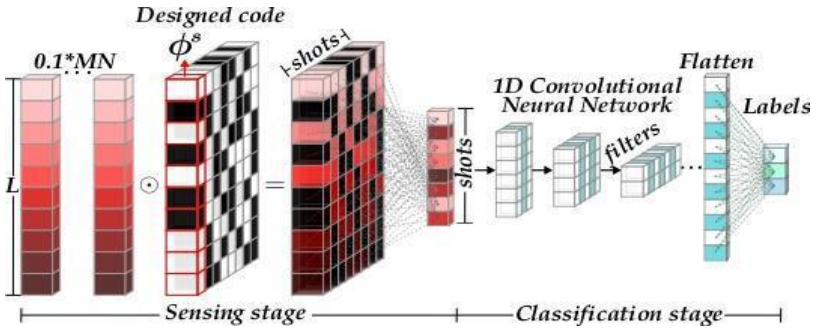


Figure 5.2. Proposed deep learning scheme. In the sensing stage, the designed dichroic optical filters are learned and implemented to acquire the compressive measurements, which are the input to train the classification neural network. Recovered from [2].

The classification network structure used in the E2E model was inspired by the work presented in [18] for NIR images. Specifically, the network consists of three convolutional layers with batch normalization and rectified linear unit (ReLU) as the activation function, followed by a dropout to avoid overfitting, and finally, two fully connected layers, where the last one has the softmax activation.

5.1.1 Simulations and Experimental Results

The proposed coding pattern design for classification was tested on two hyperspectral datasets in the NIR region. The first dataset, called Cereals, contains the reflectance of cereal samples in the spectral range of 943-1643 nm with an interval of 7 nm leading to 101 spectral bands [19]. The spatial dimension of this image is 241× 181 pixels, which includes three types of puffed cereals labeled as corn, wheat, and rice, according to their main components. The second NIR dataset, Yatsuhashi, available at https://www.kaggle.com/hacarus/near-infrared-hyperspectral-image, called the Near Infrared Hyperspectral Image Dataset, contains 192 × 256 pixels with 96 spectral bands covering 1293-2215 nm and is composed of three types of yatsuhashi sweets commercialized by different companies. This network was trained using 10% of the total spectral signatures. The proposed deep coding patterns design was compared with a random pattern (Random-design), and the coded design proposed in [18] was denoted as (Traditional-design). Additionally, the classification results when using the whole spectral data cube (Full data) are included, i.e., no compression is performed, and the sensing matrix is $H = I$.

Numerical tests were conducted to demonstrate the proposed coding patterns design under different sensing ratio $\%_{3D} = [0.1, 0.2, 0.3, 0.4, 0.5]$ where 0.1 is the extreme case of compression evaluated. Tables 5.1 and 5.2 summarize the classification accuracy obtained for both datasets by selecting the best experiment from a total of 25 training trials. For all the scenarios, the proposed design outperforms the other designs by up to 10% accuracy. Additionally, the main gain is obtained with the highest compression value$\%_{3D} = 0.1$ which is the desired performance of senseless data. The result employing the full data is also presented, i.e., without compression$\%_{3D} = 1(\%_{3D} = 0.5)$, where it can be observed that the compressive classification in the NIR spectrum is possible, and the accuracy difference is less than 2% in comparison with the Deep design for $\%_{3D} = 0.5$.

Table 5.1. Quantitative evaluation for the coding patterns design in terms of overall accuracy for the Cereals Dataset.

%3D	Random design	Traditional design	Deep design	Full data
0.1	0.8003	0.8856	**0.9117**	-
0.2	0.8663	0.9136	**0.9275**	-
0.3	0.8399	0.9232	**0.9386**	-
0.4	0.8753	0.9359	**0.9440**	-
0.5	0.8772	0.9398	**0.9596**	-
1	-	-	-	**0.9785**

Table 5.2. Quantitative evaluation for the coding patterns design in terms of overall accuracy for the Atsushi Dataset.

%3D	Random design	Traditional design	Deep design	Full data
0.1	0.9192	0.9322	**0.9671**	-
0.2	0.9328	0.9488	**0.9681**	-
0.3	0.9458	0.9483	**0.9684**	-
0.4	0.9542	0.9528	**0.9708**	-
0.5	0.9610	0.9652	**0.9714**	-
1	-	-	-	**0.9894**

Finally, to see a visual representation of the compressive NIR spectral classification, the bottom part of figures 5.3 and 5.4 shows the results obtained for $\%_{3D} = 0.1$. The proposed method is more accurate compared to traditional methods in the classification task for both datasets. Also, the coding patterns are shown in the upper part of the figures. The proposed deep design converges to special bandpass filters, like the traditional approach. However, the deep design contains more elements, resulting in the optimal transmittance for the NIR dataset used.

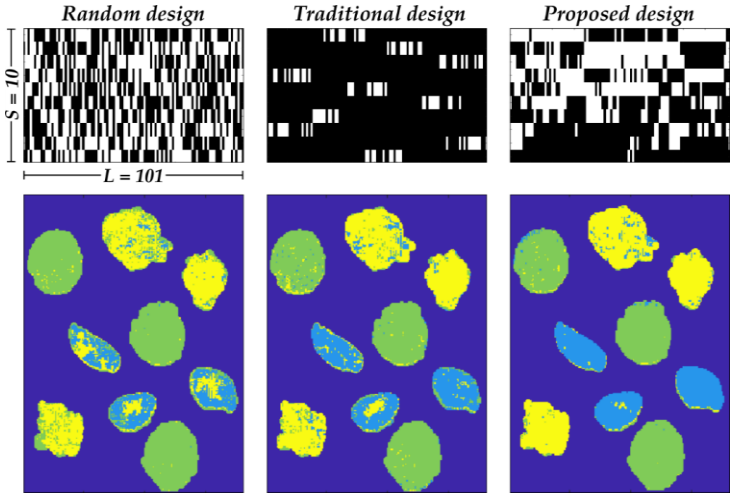


Figure 5.3. Visual representation of the classification results and coding patterns for each design using the $\%3D = 0.1$ scenario in the Cereals Datasets. Recovered from [2].

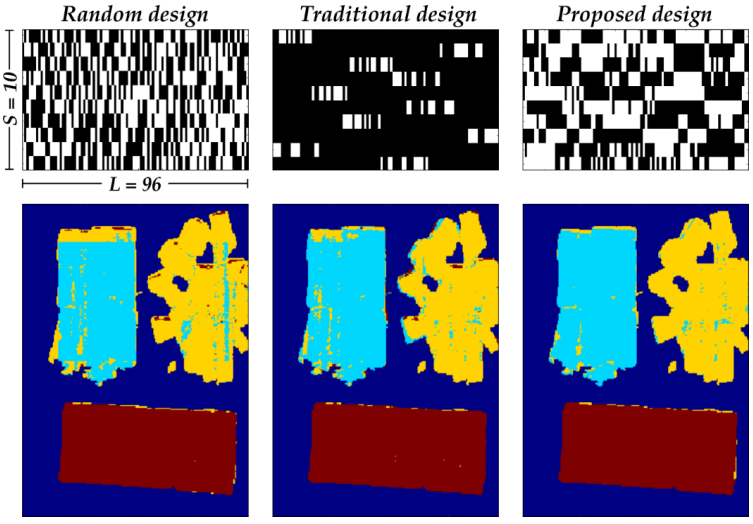


Figure 5.4. Visual representation of the classification results and coding patterns for each design using the $\%3D = 0.1$ scenario in the Yatsuhashi Dataset. Recovered from [2].

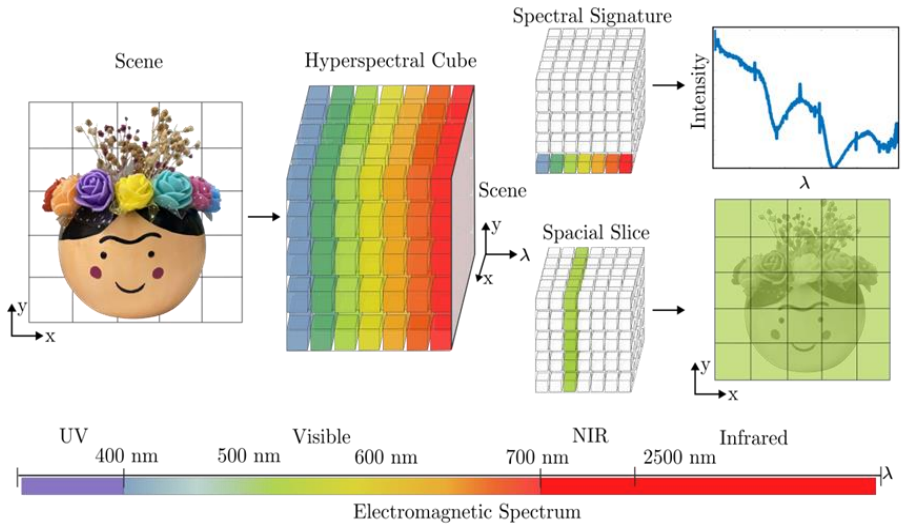


Figure 5.5. Spectral data acquisition model. Demonstration of the differences between spectral signatures, and spatial slice (spatial representation in one spectral band).

5.2 Color Coded Aperture Modeling based on Band-Selection

The acquisition of spectral data can be captured as spectral signature or spatial slices as shown in Fig. 5.5. The proposed method presented in Fig. 5.6 consists of modeling the color-coded aperture design as a band selection problem. For this proposed in [20] we considered train all the selected wavelength at the time, and then put in the CCA. Specifically, we consider the following operator M_ϕ as

$$y = M_{\boldsymbol{\phi}}(x) = \boldsymbol{\phi} \odot x \quad where \quad ||y||_0 = N \ll L, \tag{4}$$

where $x \in R^L$ is the full bands spectral signature and y is the selected bands by the binary weight $\phi \in \{0,1\}^L$ with N one values. The proposed method consists of jointly designing ϕ and the weights θ of a classification network $N_\theta(\cdot)$, considering a training label dataset $\{x_k, c_k\}_{k=1}^K$ by solving the following optimization problem:

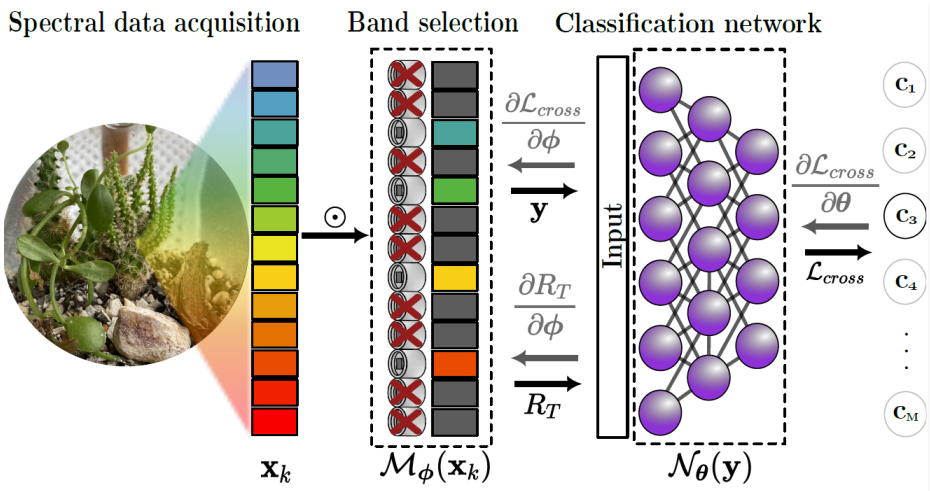


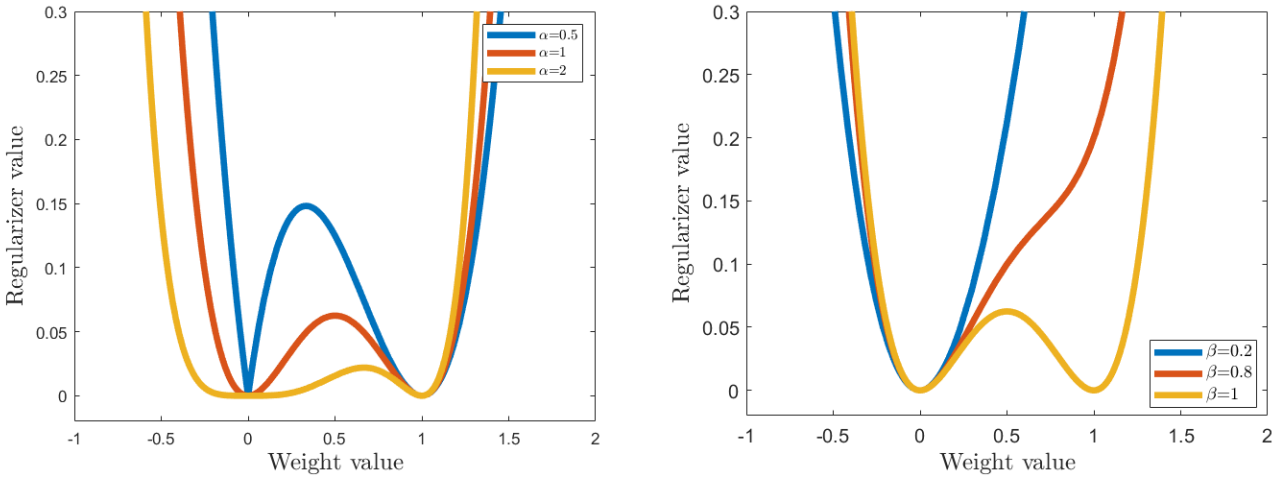
Figure 5.6. The computational pipeline of the proposed method. The element-wise product $\odot$ is performed between x_k and ϕ obtaining only the information of the selected bands. The training is performed over the L_{cross} including also the regularizer R_T over the binary weight ϕ.

$$\{\phi^*, \theta^*\} = arg \min_{\phi,\theta} \sum_{k=1}^{K} L_{cross}\left(N_\theta\left(M_\phi(x_k)\right), c_k\right) + \mu R_T(\phi), \tag{22}$$

Where μ is a trade-off hyper-parameter between the cross-entropy loss function, L_{cross} and the proposed regularizer, $R_T(\phi)$. Observe that $R_T(.)$ operates only over ϕ; therefore, the proposed regularizer aims to obtain binary values and the number of selected bands as follows.

$$R_T(\phi) = \underbrace{\left(\sum_{n=1}^{L} \beta\,(\phi_n^2)^{\alpha}(1-\phi_n)^2 + (1-\alpha)(\phi_n^2)^{\alpha}\right)}_{R_{binary}} + \rho\underbrace{\left(N - \sum_{n=1}^{L}\phi_n\right)^2}_{R_{bands}}. \tag{23}$$

Notice that the regularizers R_{binary} and R_{bands} are minimized when the entries of ϕ are binary and $||\phi||_1 = N$, respectively.



a) *Alpha* (α) parameter

b) *Beta* (β) parameter

Figure 5.7. The behavior of regularization functions with different parameters that establish their convergence. a) Behavior of the alpha parameter. b) Behavior of the beta parameter.

The parameters β and α help to control the number of resulting bands N in the training step. Note that the above method selects the most relevant spectral bands for the classification task with the selected dataset. However, to design a CCA, it is necessary to design optical filters that acquire the most relevant spectral bands. For the regularization parameters it is sought to have two specific characteristics of the customized layer, first that it is binary and second that the binarized values in one are the desired number of bands, for this different values of the same were visualized illustrated in Figure 5.7, where we see how alpha is decreasing its weight to the zeros, this first value is taken because the initialization starts with all in 1, and gradually reaches the desired balance where the best bands take strength and are selected.

Also, we analyzed the behavior of the hyperparameter with each number of selected bands for 3 bands, 5 bands and 10 bands using the spectral dataset Indian Pines. The analysis was made using the accuracy and loss metrics for each case in Figure 5.8 for N = 3, Figure 5.9 for N=5 and Figure 5.10 for N=10. The loss in high values (in yellow), means that the algorithm does not converge in the regularization terms, that is why tends to a very large number of the loss.

3 bands

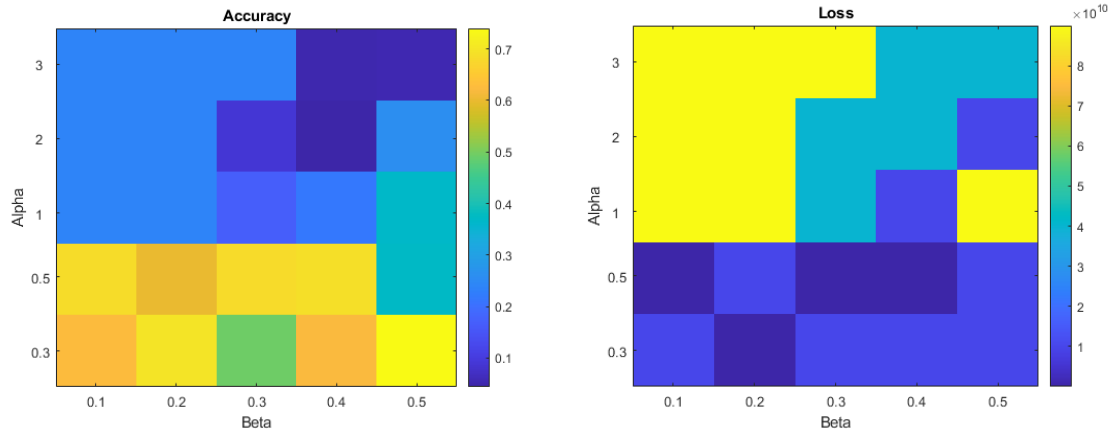


Figure 5.8. Accuracy and loss training for N= 3 bands. Small values of alpha and high values of beta benefit this number of bands, an alpha greater than 1 is a very strong parameter that does not allow to select as few bands as three.

5 bands

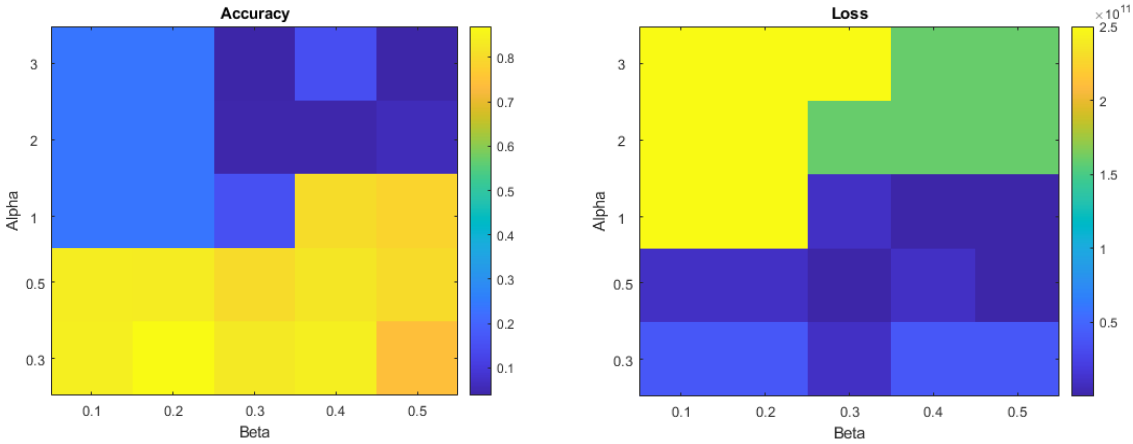


Figure 5.9. Accuracy and loss training for N= 5 bands. The convergence of the network covers more values of alpha, with more freedom to select the bands, the classification starts to be more stable.

10 bands

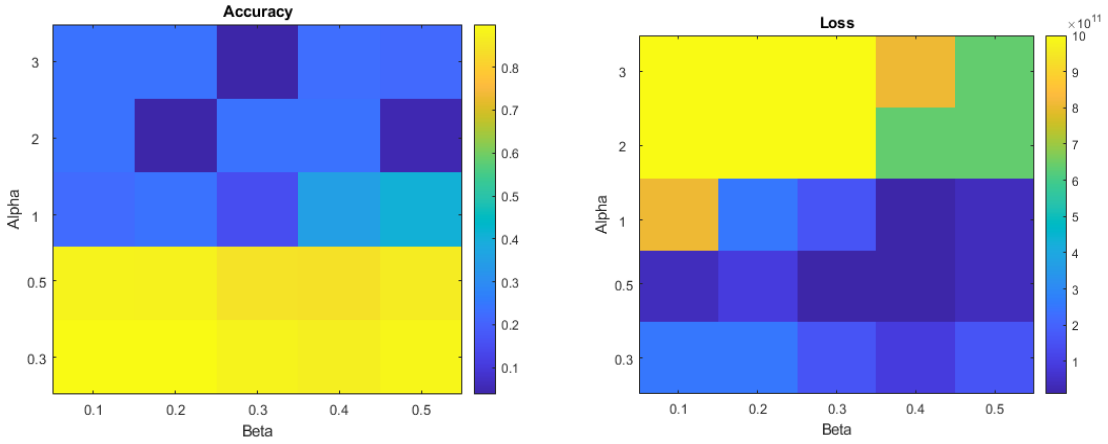


Figure 5.10. Accuracy and loss training for N= 10 bands. Values greater than 1 for the alpha parameter do not generate a good prediction of the learned weights, however, in this number of N=10, the beta values that gave the best results were the smallest, contrary to the previous tests.

The bandwidth is given by 20 nm, representing a standard bandwidth in commercial filters as shown in Fig. 5.11. However, this is a parameter we can vary in the current implementation. Thus, the sensing method, with contain the CCA can be designed by multiplying the original image with each designed filter.

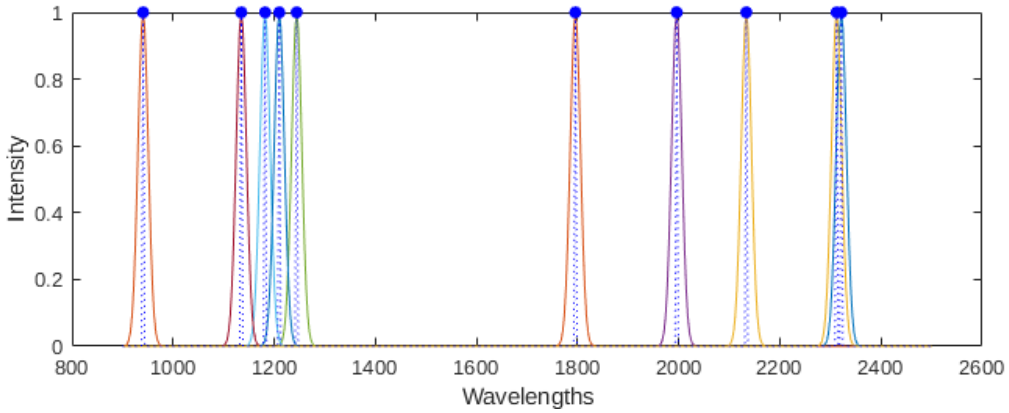


Figure 5.11 The design scheme of optical filters is to be put into the color-coded aperture using the end-to-end scheme.

5.2.1 Simulation Results

We use the Indian Pines dataset (L=200) with a proportion of 70%, 10%, and 20% for training, validation, and testing. The proposed optical band selection system is trained using each spatial point as an independent measurement. The parameters used are initialized in $\beta = 0.2, \alpha = 1, \mu = 10^{-6}, \rho = 10$. The parameters α *and* β increase by a factor of 2 every 4 epochs. Table 4 summarizes the results for 20 trials for different seeds varying the number of bands N= {3,5,7,10,15,20} for the SOTA and the proposed method. Note that the proposed approach outperforms the other SOTA methods for $N \geq 5$ y and presents comparable results for the extreme evaluated case $N = 3$. All methods were tested under the same conditions.
Some methods for hyperspectral band selection have been proposed, including Reinforcement Learning [21], Top-Rank Cut (TRC) method utilizing Optimal Clustering (OC), and Density-Peak-Based Clustering (FDPC). Other approaches involve combining Normalized Cut (NC) with Information Entropy (IE) and Maximum-Variance Principal Component Analysis (MVPCA) [22]. Although these methods aim to preserve spectral information, they are not designed to select characteristic information to realize a specific task such as classification.

Table 5.3. State of the art methods comparison with classification overall accuracy and standard deviation for 20 trials, remarking that for $N = L = 200$ the classification accuracy is $0{,}930 \pm 0{,}006$. The best results are indicated in bold, while the second best are underlined.

	Number of bands					
Methods	**3**	**5**	**7**	**10**	**15**	**20**
RL	0.670±0.006	0.686±0.005	0.740±0.008	0.793±0.007	0.838±0.007	0.869±0.004
TRC OC FDPC	0.661±0.005	0.767±0.062	<u>0.847±0.005</u>	0.855±0.005	<u>0.874±0.007</u>	<u>0.893±0.005</u>
NC OC MVPCA	**0.734±0.006**	<u>0.816±0.004</u>	**0.849±0.005**	<u>0.858±0.006</u>	0.871±0.005	0.886±0.006
NC OC IE	**0.734±0.006**	<u>0.816±0.004</u>	0.845±0.006	0.855±0.005	0.861±0.009	0.882±0.006
Proposed	<u>0.723±0.005</u>	**0.8173±0.004**	**0.849±0.005**	**0.878±0.004**	**0.894±0.005**	**0.904±0.005**

IV. IMPACTS:

The direct impact of this research is to provide infrared-color-coded aperture optimization designs used to acquire near-infrared spectral images with compressive spectral imaging systems that allow high-level tasks such as classification to be performed. By designing and validating the color-coded aperture design in a real system, not only shows how what was developed improves the quality obtained but also opens a wide range of possibilities for other disciplines to use this data for future research. Furthermore, the end-to-end scheme developed in this work can be extended to various image computing systems, allowing its impact to be expanded to different areas of science.

Social impact:
1. The development of the project contributed to the formation of Colombian Human Capital of different academic levels between undergraduate students in systems engineering to doctors in areas of computer science.
2. Increased interest in the national scientific community in developments related to computational images.

Economic Impact:
1. Reduction in the costs of implementing equipment for image acquisition through the design of optical elements used in computer imaging systems.
2. Reduction in the cost related to the capture of images necessary in computational imaging systems.

V. FUTURE WORK:

The work conducted during this project has leveraged new research directions related to coded aperture NIR spectral imaging. Some of the most relevant include:

1. **NIR spectral Imaging Dataset:** Following the steps taken in this project to build a NIR spectral dataset, we plan to acquire a plurality of scenes, so that the dataset with rich spatial and spectral information can be shared to the scientific community.
2. **Test-bed single pixel implementation:** Evaluate alternatives to improve the optical implementation of the single pixel camera to increase the field of view (FOV) while maintaining the high SNR of the system.
3. **Implementation of Dichroic Filters in CASSI system:** The CCA design based on the dichroic filters has shown promising results for reconstruction and classification. However, its main limitation lies in that its implementation cost increases with the number of bands. Therefore, we are interested in studying coding strategies to emulate the effect of this type of coding without increasing manufacturing price. For example, recent work has shown that increasing the integration time while changing a binary coded aperture can transform the time dependence into wavefront information if a prism is included in the setups.
4. **Evaluation of the CCA design with real data:** The experiments conducted in this project employing colored CA involved simulated and real data. Nonetheless, based on richer spatial and spectral NIR data acquired in the laboratory, we plan to assess the performance of the designed CCA with a larger real dataset. Further, we plan to implement real and commercial CSI systems that consider the CCA design and evaluate the performance of the error introduced in the calibration process.
5. **Testing of the CCA design in other high-level task:** Some of the results presented so far have been evaluated in image reconstruction and pixel classification tasks. Further studies should evaluate the effect of the designed coding strategies in high-level tasks such as unmixing and object detection.

VI. REFERENCES

[1] H. Arguello *et al.*, “Deep Optical Coding Design in Computational Imaging: A data-driven framework,” *IEEE Signal Process. Mag.*, vol. 40, no. 2, pp. 75–88, Mar. 2023.
[2] J. Bacca, T. Gelvez-Barrera, and H. Arguello, “Deep Coded Aperture Design: An End-to-End Approach for Computational Imaging Tasks,” *IEEE Transactions on Computational Imaging*, vol. 7, pp. 1148–1160, 2021.
[3] L. Wang, C. Sun, Y. Fu, M. H. Kim, and H. Huang, “Hyperspectral image reconstruction using a deep spatial-spectral prior,” in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, 2019, pp. 8032–8041.
[4] E. J. Candes and M. B. Wakin, “An Introduction To Compressive Sampling,” *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.
[5] W. Dong, G. Shi, X. Li, Y. Ma, and F. Huang, “Compressive sensing via nonlocal low-rank regularization,” *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3618–3632, Aug. 2014.
[6] W. Cui, S. Liu, F. Jiang, and D. Zhao, “Image compressed sensing using non-local neural network,” *IEEE Trans. Multimedia*, vol. 25, pp. 816–830, 2023.
[7] H. Garcia, C. V. Correa, and H. Arguello, “Optimized sensing matrix for single pixel multi-resolution compressive spectral imaging,” *IEEE Trans. Image Process.*, 2020.
[8] X. Yu, R. I. Stantchev, F. Yang, and E. Pickwell-MacPherson, “Super Sub-Nyquist Single-Pixel Imaging by Total Variation Ascending Ordering of the Hadamard Basis,” *Sci. Rep.*, vol. 10, no. 1, p. 9338, Jun. 2020.
[9] W.-K. Yu, “Super Sub-Nyquist Single-Pixel Imaging by Means of Cake-Cutting Hadamard Basis Sort,” *Sensors* , vol. 19, no. 19, Sep. 2019.
[10] P. G. Vaz, D. Amaral, L. F. Requicha Ferreira, M. Morgado, and J. Cardoso, “Image quality of compressive single-pixel imaging using different Hadamard orderings,” *Opt. Express*, vol. 28, no. 8, pp. 11666–11681, Apr. 2020.
[11] H. Garcia, C. V. Correa, and K. Sánchez, “Multi-resolution coded apertures based on side information for single pixel spectral reconstruction,” *2018 26th European*, 2018.
[12] B. Monroy, J. Bacca, and H. Arguello, “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)*, 2023, pp. 1–5.
[13] F. Yang, Q. Sun, H. Jin, and Z. Zhou, “Superpixel segmentation with fully convolutional networks,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 13964–13973.
[14] J.-L. Xu, C. Riccioli, A. Herrero-Langreo, and A. Gowen, “Deep learning classifiers for near infrared spectral imaging: a tutorial,” *J. Spectr. Imaging*, Dec. 2020.
[15] F. Zhu, Y. Wang, S. Xiang, B. Fan, and C. Pan, “Structured Sparse Method for Hyperspectral Unmixing,” *ISPRS J. Photogramm. Remote Sens.*, vol. 88, pp. 101–118, Feb. 2014.
[16] J. Bacca, M. Marquez, and H. Arguello, “Single pixel near-infrared imaging for spectral classification,” in *Imaging and Applied Optics Congress 2022 (3D, AOA, COSI, ISA, pcAOP)*, Vancouver, British Columbia, 2022, p. CW1B.2.
[17] J. Bacca, A. Hernandez-Rojas, and H. Arguello, “Deep Coding Patterns Design for Compressive Near-Infrared Spectral Classification,” in *2022 30th European Signal Processing Conference (EUSIPCO)*, 2022, pp. 548–552.
[18] H. Arguello, S. Pinilla, Y. Peng, H. Ikoma, J. Bacca, and G. Wetzstein, “Shift-variant color-coded diffractive spectral imaging system,” *Optica*, vol. 8, no. 11, p. 1424, Nov. 2021.
[19] C. Hinojosa, J. Bacca, and H. Arguello, “Coded Aperture Design for Compressive Spectral Subspace Clustering,” *IEEE J. Sel. Top. Signal Process.*, vol. 12, no. 6, pp. 1589–1600, Dec. 2018.
[20] K. Fonseca, H. Garcia, F. Da Silva, H. Arguello, and J. Bacca, “Joint Deep Learning Optical Band Selection and Classification Method for Spectral Data,” in *2023 Optica Imaging Congress*, Boston, Massachusetts.
[21] L. Mou, S. Saha, Y. Hua, F. Bovolo, L. Bruzzone, and X. X. Zhu, “Deep Reinforcement Learning for Band Selection in Hyperspectral Image Classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, 2022.
[22] Q. Wang, F. Zhang, and X. Li, “Optimal Clustering Framework for Hyperspectral Band Selection,” *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 5910–5922, Oct. 2018.