# CODED APERTURE DESIGN FOR COMPRESSIVE SPECTRAL IMAGE CLASSIFICATION USING DEEP LEARNING

NELSON MAURICIO SILVA MALDONADO

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FISICOMECÁNICAS
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
BUCARAMANGA
2022

# CODED APERTURE DESIGN FOR COMPRESSIVE SPECTRAL IMAGE CLASSIFICATION USING DEEP LEARNING

NELSON MAURICIO SILVA MALDONADO

Trabajo de Grado para optar al título de
Magister en Ingeniería de Sistemas

Advisor
Ph.D Henry Arguello Fuentes
Co-advisor
Ph.D Laura Viviana Galvis Carreño

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FISICOMECÁNICAS
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
BUCARAMANGA
2022

## ACKNOWLEDGMENTS

The author expresses his acknowledge to HDSP Research group, specially to professors Henry Arguello and Laura Galvis for their support throughout this research work. Special thanks to the fellow students of the HDSP Optical Laboratory for their inconditional support and those who contributed with their ideas and comments to this document.

# Contents

# List of Figures

**page**



---

[1] Qingli Li et al. "Review of spectral imaging technology in biomedical engineering: achievements and challenges". In: *Journal of Biomedical Optics* 18.10 (2013), p. 100901. DOI: 10.1117/1.jbo.18.10.100901.

[2] Gonzalo R. Arce et al. *Compressive coded aperture spectral imaging: An introduction*. 2014. DOI: 10.1109/MSP.2013.2278763.

[3] Vivienne Sze et al. "Efficient Processing of Deep Neural Networks: A Tutorial and Survey". In: *Proceedings of the IEEE* 105.12 (2017), pp. 2295–2329. DOI: 10.1109/JPROC.2017.2761740. arXiv: 1703.09039.

---

[4] Han Xiao, Kashif Rasul, and Roland Vollgraf. “Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms”. In: *CoRR* abs/1708.07747 (2017). arXiv: 1708.07747.

# List of Tables

**page**

# RESUMEN

**TÍTULO:** CODED APERTURE DESIGN FOR COMPRESSIVE SPECTRAL IMAGE CLASSIFICATION USING DEEP LEARNING [*]

**AUTOR:** NELSON MAURICIO SILVA MALDONADO [**]

**PALABRAS CLAVE:** sensado compresivo, optimización, imágenes espectrales, classificación, aprendizaje profundo.

**DESCRIPCIÓN:**

La teoría del sensado compresivo habilita la reconstrucción de imágenes espectrales usando un número menor de observaciones que las dictadas por el enfoque tradicional basado en el teorema de Shannon-Nyquist a través de sistemas de imágenes compresivos (CSI). Estos sistemas CSI se apoyan en un montaje óptico basado en un elemento dispersivo acoplado a una o más aperturas codificadas para capturar y comprimir una escena espectral de manera simultánea. Después, la reconstrucción de la escena subyacente se obtiene a través de algoritmos computacionales. Luego las tareas de procesamiento como clasificación, detección de objetos y segmentación son ejecutadas sobre las imágenes reconstruidas. Sin embargo, este proceso de reconstrucción es costoso desde el punto de vista computacional. La descompresión hace que se requiera más tiempo y recursos para realizar este tipo de tareas . En este trabajo de investigación la clasificación espectral se realiza directamente sobre las medidas comprimidas que se adquirieron a través de una arquitectura óptica que sigue los lineamientos de la teoría de sensado compresivo (CS). Se propone un método de extremo a extremo para la optimización conjunta de las aperturas codificadas y los parámetros del modelo de aprendizaje profundo que se usará para la clasificación. Este enfoque se aplicó a la clasificación de afecciones particulares del limón Tahití (Citrus latifolia), pero puede ser usado para distintos productos agrícolas. Además, con el objetivo de comparar los resultados obtenidos, se encontró que nuestros experimentos mejoraron hasta en un 6% la precisión en clasificación cuando las aperturas codificadas fueron optimizadas respecto al uso de aperturas aleatorias.

---

[*] Trabajo de grado

[**] Facultad de Ingenierías Físico-Mecánicas. Escuela de Ingeniería de Sistemas e Informática. Director: Ph.D Henry Arguello Fuentes, Co-directora: Ph.D Laura Viviana Galvis Carreño.

# ABSTRACT

**TITLE:** CODED APERTURE DESIGN FOR COMPRESSIVE SPECTRAL IMAGE CLASSIFICATION USING DEEP LEARNING [*]

**AUTHOR:** NELSON MAURICIO SILVA MALDONADO [**]

**KEYWORDS:** compressed sensing, deep learning, optimization, spectral images, spectral classification, Tahiti lime.

**DESCRIPTION:**

Compressed sensing (CS) theory enables the reconstruction of spectral images (SI) using a lower number of measurements than the traditional Shannon-Nyquist sampling approach, through compressive spectral imaging (CSI) systems. These CSI systems rely on a dispersive-based optical setup coupled to one or more coded apertures to capture and compress a spectral scene simultaneously. Afterward, the reconstruction of the underlying scene is obtained through computational algorithms. Then, processing tasks like classification, object detection, or segmentation are performed over the reconstructed images. However, this reconstruction process is computationally expensive, which introduces a time overhead for these tasks. In this research work, spectral classification is directly performed over compressed measurements acquired through an optical architecture following the CS framework. An end-to-end method to optimize both coded apertures and deep learning model parameters is proposed. This approach has been applied to the grading of particular features of Tahiti lime (Citrus latifolia), but can be used for different agricultural materials. For the purpose of comparison, our experiments improved up to 6% in classification accuracy over a testing database when the optimized coded apertures were used instead of the random ones.

---

[*] Master Thesis

[**] Facultad de Ingenierías Físico-Mecánicas. Escuela de Ingeniería de Sistemas e Informática. Director: Ph.D Henry Arguello Fuentes, Co-directora: Ph.D Laura Viviana Galvis Carreño

# INTRODUCTION

Spectral images have been used in several fields and industries such as agriculture [1], food quality [2], biomedical images [3], remote sensing [4], archaeology [5], among others. However, SI are expensive to capture due to the cost of the sensors and equipment involved [6], limiting the widespread use of this promising technology.

Many approaches have been proposed to acquire SI at a lower cost, such as those based on compressive sensing (CS) theory [7]. CS states that a signal can be recovered with a number of measurements lower than those stated by the Shannon-Nyquist sampling theorem, as long as the signal can be represented in a known basis such as wavelet, Discrete Cosine Transform (DCT), among others. There are several

---

[1] Aasha Nandhini, S. Hemalatha, R. Radha, et al. "Web Enabled Plant Disease Detection System for Agricultural Applications Using WMSN". in: *Wireless Personal Communications* (2018). DOI: https://doi.org/10.1007/s11277-017-5092-4.

[2] Andjela Draganić et al. "Compressive sensing approach in the table grape cold chain logistics". In: *2017 6th Mediterranean Conference on Embedded Computing (MECO)*. 2017, pp. 1–4. DOI: 10.1109/MECO.2017.7977143.

[3] Yunyun Yang et al. "A New Compressed Sensing Model Based on Median Filter with Application to Reconstruct Brain MR Images". In: *2018 IEEE 3rd International Conference on Signal and Image Processing (ICSIP)*. 2018, pp. 116–120. DOI: 10.1109/SIPROCESS.2018.8600479.

[4] Alireza Mirrashid and Ali Asghar Beheshti. "Compressed Remote Sensing by Using Deep Learning". In: *2018 9th International Symposium on Telecommunications (IST)*. 2018, pp. 549–552. DOI: 10.1109/ISTEL.2018.8661112.

[5] Yan Zhang et al. "Compressive Sensing based Software Defined GPR for Subsurface Imaging". In: *2021 IEEE Radar Conference (RadarConf21)*. 2021, pp. 1–6. DOI: 10.1109/RadarConf2147009.2021.9455291.

[6] *TE-Cooled Fluorescence Spectrometer | Edmund Optics*.

[7] D. L. Donoho. "Compressed sensing". In: *IEEE Transactions on Information Theory* 52.4 (2006), pp. 1289–1306.

optical setups that demonstrate the CS theory, namely compressive spectral imaging (CSI) systems such as the single-pixel camera [8], the coded aperture snapshot spectral imager (CASSI) [9], and its variants like dual-dispersion architecture (DD-CASSI) [9], single-disperser architecture (SD-CASSI) [10], and colored coded apertures CASSI (3D-CASSI) [11]. More recently, a dual-camera compressive hyperspectral imaging system (DCCHI) was proposed by Xing Lin *et al*. [12]. These CSI systems rely on a dispersive-based optical setup coupled to one or more coded apertures to capture and compress a spectral scene simultaneously.

Compressed spectral images acquired with these optical systems must be reconstructed to apply further processing tasks such as classification, segmentation, and object detection. This reconstruction process is computationally expensive because an ill-posed, under-determined linear system of equations must be solved [13]. There have been different computational approaches to solve this problem, for example, the gradient projection sparse reconstruction (GPSR) [14], the orthogonal matching

---

[8] M. F. Duarte et al. "Single-pixel imaging via compressive sampling". In: *IEEE Signal Processing Magazine* 25.2 (2008), pp. 83–91.

[9] M. E. Gehm et al. "Single-shot compressive spectral imaging with a dual-disperser architecture". In: *Optics Express* 15.21 (Oct. 2007), p. 14013. DOI: 10.1364/oe.15.014013.

[10] Ashwin Wagadarikar et al. "Single disperser design for coded aperture snapshot spectral imaging". In: *Applied Optics* 47.10 (2008). DOI: 10.1364/AO.47.000B44.

[11] Xing Lin et al. "Spatial-spectral encoded compressive hyperspectral imaging". In: *ACM Transactions on Graphics* 33.6 (2014). DOI: 10.1145/2661229.2661262.

[12] Lizhi Wang et al. "Adaptive Nonlocal Sparse Representation for Dual-Camera Compressive Hyperspectral Imaging". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39.10 (2017), pp. 2104–2111. DOI: 10.1109/TPAMI.2016.2621050.

[13] ChenXingyi, ZhangYujie, and QiRui. "Block Sparse Signals Recovery Algorithm for Distributed Compressed Sensing Reconstruction". In: *Journal of Information Processing Systems* 15.2 (Apr. 2019), pp. 410–421.

[14] Mário A.T. Figueiredo, Robert D. Nowak, and Stephen J. Wright. "Gradient projection for sparse

pursuit (OMP) [15], and the iterative hard thresholding (IHT) algorithms [16]. More recently, reconstruction algorithms make use of CNNs like the proposed by Chen Zhao *et al*. [17].

However, the reconstruction process introduces a time overhead that makes CS methods unsuitable for critical response-time applications like object detection, motion detection, etc. The reconstruction step, required as an additional process before the inference task is the main problem of CS aimed to address with this work. There have been different approaches avoiding the reconstruction step. For example, [18] successfully made motion detection, image classification, and noise filtering from compressive measurements. More recently, Kuldeep Kulkarni et al.[19] proposed a convolutional neural network (CNN) to classify compressive measurements of the

---

reconstruction: Application to compressed sensing and other inverse problems". In: *IEEE Journal on Selected Topics in Signal Processing* 1.4 (2007), pp. 586–597. DOI: 10.1109/JSTSP.2007.910281.

[15] J. A. Tropp and A. C. Gilbert. "Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit". In: *IEEE Transactions on Information Theory* 53.12 (2007), pp. 4655–4666.

[16] Thomas Blumensath and Mike E. Davies. "Iterative hard thresholding for compressed sensing". In: *Applied and Computational Harmonic Analysis* 27.3 (2009), pp. 265–274. DOI: 10.1016/j.acha.2009.04.002. arXiv: 0805.0510.

[17] Chen Zhao et al. "CREAM: CNN-REgularized ADMM Framework for Compressive-Sensed Image Reconstruction". In: *IEEE Access* 6 (2018), pp. 76838–76853. DOI: 10.1109/ACCESS.2018.2882990.

[18] Mark A. Davenport et al. "Signal processing with compressive measurements". In: *IEEE Journal on Selected Topics in Signal Processing* 4.2 (2010), pp. 445–460. DOI: 10.1109/JSTSP.2009.2039178.

[19] Kuldeep Kulkarni et al. "ReconNet: Non-iterative reconstruction of images from compressively sensed measurements". In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 2016-December (2016), pp. 449–458. DOI: 10.1109/CVPR.2016.55.

MNIST [20] database. Finally, [21] proposed a coupled deep learning architecture for coded aperture and feature extraction optimization using both MNIST and CIFAR-10 [22] databases. The latter reporting the improvement achieved when the coded apertures are optimized.

Nonetheless, that inference on compressive measurements has been explored by several authors. Still, to our knowledge, no approach has been proposed for spectral imaging that uses a deep learning model to optimize the coded apertures and optical elements of the acquisition system in an end-to-end approach and for a real application, which is the hypothesis and main contribution of this work.

Traditional fruit grading has been labor-intensive, requiring lots of people to perform this task. However, advances in computer vision algorithms, like the one exposed in this work, have made the development of automatic grading and sorting machines possible.

The grading process goal is to select those fruits compliant with regulatory and market standards. These standards generally reject those fruits with diseases produced by mites, fungus, etc. Also, fruits that do not comply with market-specific characteristics like size, color, flavor, etc., are rejected.

In this work, it is proposed an end-to-end scheme for the classification of spectral images, which takes the SI compressive measurements as input and the categorical variable as the output of the system with no intermediate steps in the training process of the deep learning model.

---

[20] Yann LeCun and Corinna Cortes. “MNIST handwritten digit database”. In: (2010).

[21] Jorge Bacca, Laura Galvis, and Henry Arguello. “Coupled deep learning coded aperture design for compressive image classification”. In: *Opt. Express* 28.6 (), pp. 8528–8540. DOI: 10.1364/OE.381479.

[22] Alex Krizhevsky, Vinod Nair, and Geoffrey Hinton. “CIFAR-10 (Canadian Institute for Advanced Research)”. In: ().

A spectral database of Tahiti lime (Citrus latifoia) has been chosen to demonstrate the proposed approach. The spectral images of a set of real limes have been acquired in an optical setup implemented in the HDSP research group's optical laboratory. Tahiti lime is a fruit that grows in tropical zones like Colombia, and it has a great potential for exporting to countries where quality products are highly appreciated. To this end, crops need to be organic, and fruits must comply with specific market standards and agricultural regulatory policies.

# 1. OBJECTIVES

## General Objective

- To optimize the coded apertures of a compressive spectral imaging classification system based on an end-to-end deep learning scheme.

## Specific Objectives

- To capture a spectral imaging database to be labelled and used in a deep learning classification system. Also to select a compressive spectral imaging optical architecture to obtain the respective compressive measurements.
- To parametrize the optical elements of the compressive spectral imaging classification system such that they can be optimized by an end-to-end deep learning scheme.
- To optimize the coded apertures of the compressive spectral imaging classification by using the end-to-end deep learning approach and the captured data base.
- To verify the performance of the end-to-end optimized coded apertures and to compare the results with non-optimized sets of coded apertures in the deep learning compressive spectral imaging classification task.

## 2. PUBLICATIONS

Some results of this work have been presented in the following conference:

- Silva-Maldonado, M., Galvis, L., & Arguello, H. (2022). End-to-End Compressive Spectral Classification: A Deep Learning Approach Applied to the Grading of Tahiti Lime. In International Conference on Smart Technologies, Systems and Applications (pp. 44-57). Springer, Cham.

# 3. ORGANIZATION OF THE MASTER THESIS

This thesis has been organized as follows:

A theoretical introduction is presented in chapter 4 where topics like spectral imaging, compressive sensing for spectral imaging, spectral image classification with CNN neural networks and coded aperture design are introduced. Later in this chapter, the end-to-end approach for optimization is introduced.

In chapter 5, an end-to-end classification methodology is proposed jointly with the general classification workflow. The steps required to build and annotate the spectral dataset are also introduced, particularly, the patch-wise approach for specimen sampling is depicted.

In chapter 6 some simulations of the optimization approach are presented for three types of CNN architectures and same number of datasets.

Later in chapter 7, a new dataset is build using a CASSI optical setup. Several real measurements are acquired and some experimental results are shown for a particular CNN architecture and this dataset.

Finally, the conclusions of this work are presented.

# 4. THEORETICAL BACKGROUND

## 4.1. Spectral Imaging

Spectral image (SI) sensors detect light reflectancy beyond the visible spectrum (400-700nm) such as the near-infrared (NIR) spectrum (800 - 1700nm) and short-wave infrared (SWIR) spectrum (1800 - 2500nm). Several techniques have been developed to acquire spectral scenes depending on how spatial and spectral information is captured. Whiskbroom sensors capture the spectral signature one pixel at a time [23]. Pushbroom or line sensors scan the spectral signature of one scene line by line. Stare imagers capture a whole 2D scene at a time for a given wavelength. Snapshot imagers acquire the whole 3D data cube at a time. These approaches as well as their main optical components are depicted in fig 1. As a result of this process, a three-dimension $(x, y, \lambda)$ data cube is obtained, as shown in fig. 2. Note the resulting spectral signature when the spectral dimension is evaluated in a certain point $(x, y)$.

## 4.2. Compressive Sensing for Sectral Imaging

Different architectures have been proposed to capture spectral images using the compressive sensing principles. The coded aperture snapshot spectral imager (CASSI) proposed by Wagadarikar et al. [10] exploits CS principles to capture a snapshot spectral image using a single disperser. The whole process is shown in Fig 3 and the

---

[23] Qingli Li et al. "Review of spectral imaging technology in biomedical engineering: achievements and challenges". In: *Journal of Biomedical Optics* 18.10 (2013), p. 100901. DOI: 10.1117/1.jbo.18.10.100901.

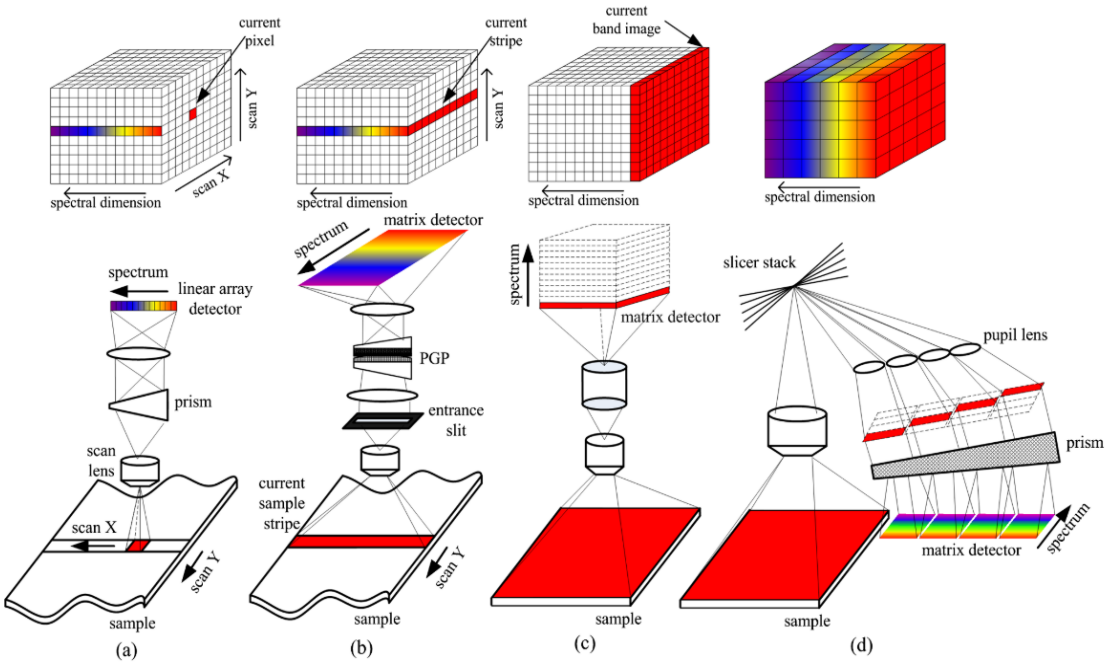



Figure 1. Typical spectral imaging approaches. (a) Whiskbroom. (b) Pushbroom. (c) Staring. (d) Snapshot. Source [24].

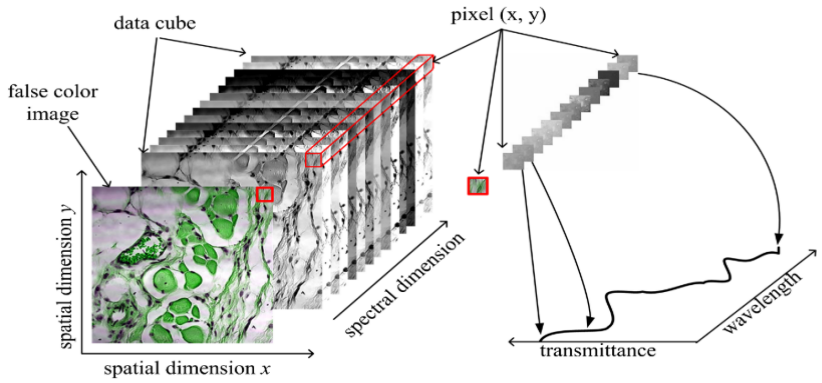



Figure 2. Spectral data cube. Source [24].

acquisition process is modeled as,

$$\mathbf{Y}^{l}_{mn} = \sum_{k=0}^{L-1} \mathbf{F}_{m(n-k)k} \mathbf{T}^{l}_{m(n-k)} + \boldsymbol{\omega}_{mn}, \tag{1}$$

where $\mathbf{Y}$ represents the measurements, $\mathbf{F}$ is referring to the data cube, $\mathbf{T}$ represents

a coded aperture, which modulates the scene and $\boldsymbol{\omega}$ is the noise in the system. The dimensions of the data cube are $M \times N \times L$, $\mathbf{T}$ is $M \times N$, and $\mathbf{Y}$ is $M \times (N + L - 1)$. $L$ represents the number of spectral bands. It is important to mention that the coded aperture is the most important element in a compressive optical system and it is the only that can be optimized.

Once the measurements have been acquired, the scene can be reconstructed by solving the following optimization problem,

$$\hat{\mathbf{f}} = \boldsymbol{\Psi} \left( \arg\min_{\boldsymbol{\theta}} \|\mathbf{y} - \mathbf{H}\boldsymbol{\Psi}\boldsymbol{\theta}\|_2^2 + \lambda \|\boldsymbol{\theta}\|_1 \right), \tag{2}$$

where $\hat{\mathbf{f}}$ refers to the reconstructed scene, $\mathbf{y}$ represents the vectorized measurements, $\mathbf{H}$ is the acquisition matrix, $\boldsymbol{\Psi}$ is the basis representation, $\boldsymbol{\theta}$ represents the sparse coefficients in the given basis representation, and $\lambda$ is a regularization constant.

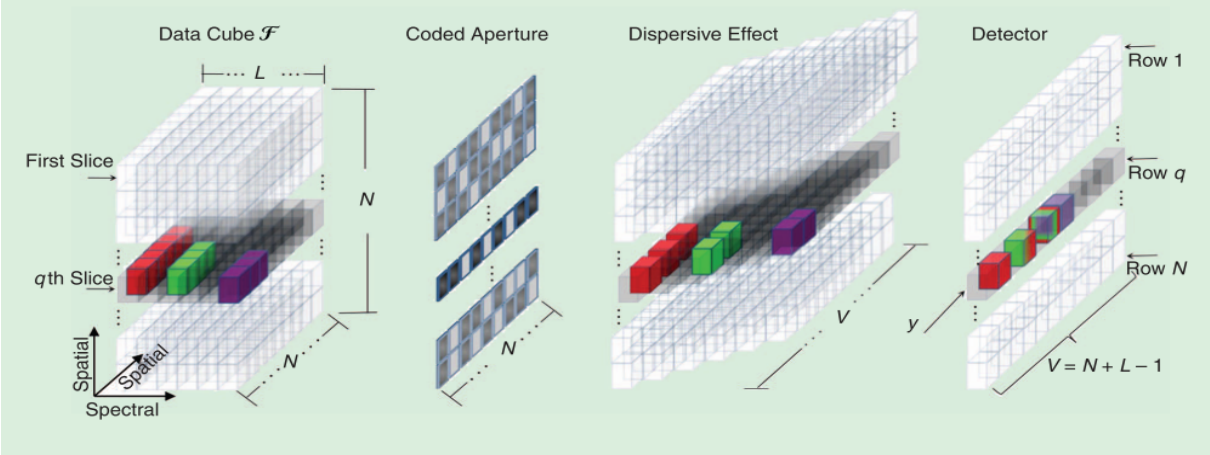



Figure 3. Physical sensing phenomena in CASSI architecture. Source [25] .

There are several methods to solve this optimization problem, like the gradient projection for sparse reconstruction (GSPR) algorithm proposed by [14], the LARS/LASSO

algorithm proposed by [26], among others. Besides the iterative approaches, non-iterative methods have also been proposed, such as [19], which uses a CNN and an image denoiser to reconstruct a scene from compressive measurements.

### 4.3. Spectral Image Classification with Deep Learning

Advances in machine learning and deep learning techniques have led to an increase in the efficacy of image processing algorithms. Specifically CNN's have been proven to be a valuable tool for classification tasks [27].

Basically a CNN takes and image data cube as an input, then a series of convolutions are performed over the image to extract the low and high level features. Finally, fully connected layers extract the global features required to generate a predicted class. The whole process is depicted in figure 4.

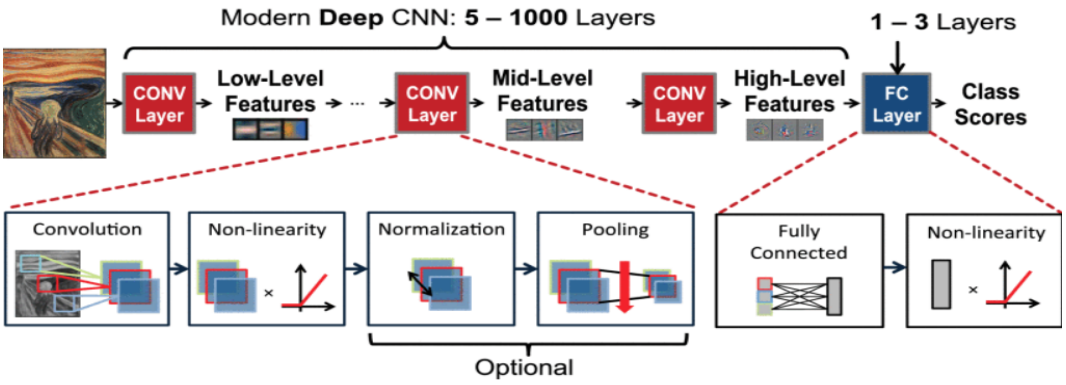



Figure 4. CNN Architecture. Source [28].

Several methods have been developed for spectral image classification using deep

---

[26] David L. Donoho and Yaakov Tsaig. "Fast solution of l1-Norm minimization problems when the solution may be sparse". In: *IEEE Transactions on Information Theory* 54.11 (2008), pp. 4789–4812. DOI: 10.1109/TIT.2008.929958.

[27] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. *ImageNet Classification with Deep Convolutional Neural Networks*. Tech. rep.

learning such as 1D-CNN, 2D-CNN[29] and 3D-CNN's [30]. Also, ensemble techniques have been used, which combines two o more of the aforementioned methods.

The classification process over spectral images can be done using a spectral, spatial or spatio-spectral approach. Spectral classification takes a set of 1D spectra as input. This is the preferred approach for low-spatial, high-spectral resolution images [31]. Spatio-spectral classification on the other hand takes both 1D spectra and 2D images as inputs. This approach is possible with high-spectral, high-spatial resolution images.

In general the HSI classification process using a CNN starts by finding a non-linear transformation function through a process called network training. This trainning process solves the following optimization problem [32],

$$\min_{\phi} \frac{1}{K} \sum_{i=0}^{K} \mathcal{L}(y^i, f(\mathbf{x}^i, \phi)), \tag{3}$$

where $K$ is the number of training examples, $\phi$ is the parameter of the mapping function, $\mathbf{x}^i$ is the feature vector of the i-th sample, $y^i$ is the corresponding label, and $\mathcal{L}$ is the loss function. To solve this optimization problem several algorithms based on

---

[29] Shutao Li et al. "Deep Learning for Hyperspectral Image Classification: An Overview". In: *IEEE Transactions on Geoscience and Remote Sensing* 57.9 (2019), pp. 6690–6709. DOI: 10.1109/TGRS.2019.2907932.

[30] Yushi Chen et al. "Deep Feature Extraction and Classification of Hyperspectral Images Based on Convolutional Neural Networks". In: *IEEE Transactions on Geoscience and Remote Sensing* 54.10 (2016), pp. 6232–6251. DOI: 10.1109/TGRS.2016.2584107.

[31] Nicolas Audebert et al. "Deep Learning for Classification of Hyperspectral Data: A Comparative Review. IEEE geoscience and remote sensing magazine Deep Learning for Classification of Hyperspectral Data: " in: *IEEE* 7.2 (2019), pp. 159–173. DOI: 10.1109/MGRS.2019.2912563Ãŕ.

[32] Shiliang Sun et al. "A Survey of Optimization Methods From a Machine Learning Perspective". In: *IEEE Transactions on Cybernetics* (2019), pp. 1–14. DOI: 10.1109/tcyb.2019.2950779. arXiv: 1906.06821.

the gradient descent [33] can be used, like stochastic gradient descent [34], mini-batch gradient descent, stochastic gradient descent with momentum [35], adaptive moment estimation (Adam)[36], among others. Currently, ensamble methods[37] wich combine machine learning, deep learning, residual learning and novel feature extraction are the state of the art.

### 4.4. Coded aperture Design

A key component of a compressive sensing setup is the coded aperture (CA) whose design will determine the quality of the reconstruction of the underliying or compressed scene. There are several approaches to find an optimal CA design. By one hand, analitycal methods make use of Compressed sensing principles like RIP, sparse basis, among others. By other hand, there are interative methods like those based on deep learning that make use of CNN for this task. Some examples of optimized CA's are shown in figure 5. It can be shown that coded aperture designs take a particular pattern for each task. For example, in figure 5a the pattern is random in the middle but some holes appear in the upper sides. In figure 5b the pattern is

---

[33] Yann LeCun et al. "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2323. DOI: 10.1109/5.726791.

[34] Léon Bottou. "Large-Scale Machine Learning with Stochastic Gradient Descent". In: *Proceedings of COMPSTAT'2010*. Ed. by Yves Lechevallier and Gilbert Saporta. Heidelberg: Physica-Verlag HD, 2010, pp. 177–186.

[35] Ilya Sutskever et al. *On the importance of initialization and momentum in deep learning*. Tech. rep. 2013.

[36] Diederik P. Kingma and Jimmy Lei Ba. "Adam: A method for stochastic optimization". In: *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. International Conference on Learning Representations, ICLR, Dec. 2015. arXiv: 1412.6980.

[37] Shutao Li et al. "Deep Learning for Hyperspectral Image Classification: An Overview". In: *IEEE Transactions on Geoscience and Remote Sensing* 57.9 (2019), pp. 6690–6709. DOI: 10.1109/TGRS.2019.2907932.

composed of clusters of pixels. In figures 5c and 5d the patterns have some sparcity.

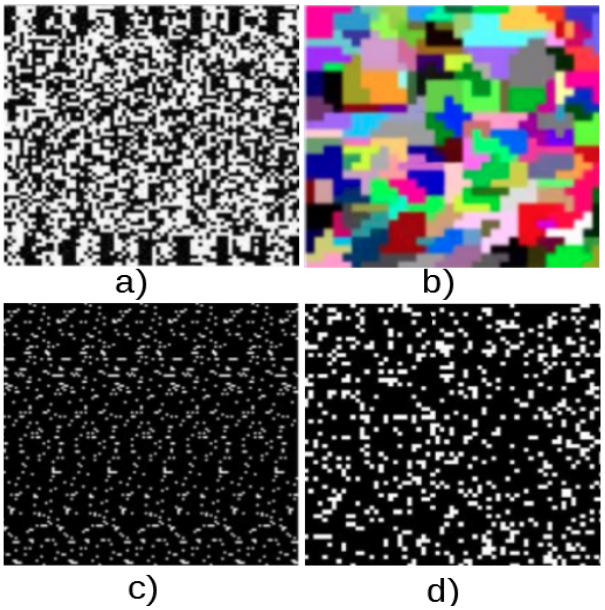



Figure 5. Task-optimized coded apertures. a) Tomography. b) Deep learning. c) Superresolution. d) Spectral selectivity.

### 4.5. End-to-End Optimization approach.

In traditional approaches, deep learning optimization is performed over a set of variables at a time. Recently, the end-to-end methods can optimize two or more sets of variables simultanously by integrating these variables in the deep learning optimization scheme. Figure 6 depicts a model for the optimization of CNN parameters and coded aperture design. It starts by taking an spectral image as input, then the coded aperture and CNN's parameters are optimized jointly. Finally, the optimized model process the input to determine a categorical variable or class as output.

This technique has been proven to be effective in fields like optics[38] where tradition-

[38] Vincent Sitzmann et al. "End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging". In: *ACM Transactions on Graphics* 37.4 (2018). DOI: 10.1145/3197517.3201333.

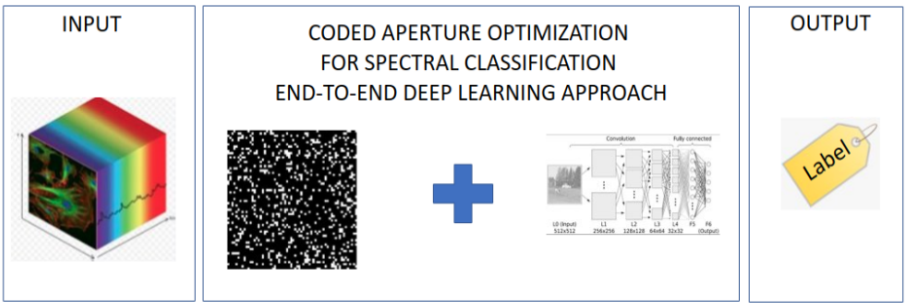



Figure 6. End-to-End optimization approach. Source: Author.

ally designs the optical system, then the image processing algoritms are optmized. Nowadays, both optical system and processing algoritms are optimized simultaneously. Likewise, Long *et al*. [39] managed to improve the field of semantic segmentation by proposing an end-to-end method for training fully convolutional networks.

[39] J. Long, E. Shelhamer, and T. Darrell. "Fully convolutional networks for semantic segmentation". In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2015, pp. 3431–3440.

# 5. PROPOSED METHODOLOGY

## 5.1. End-to-end approach for the Tahiti Lime Classification

This section describes the proposed methodology for the classification of Tahiti lime in a compressive spectral imaging framework. The approach is divided into a learning or training process and a testing part. The learning process is proposed as an end-to-end approach that jointly optimizes the acquisition and classification tasks. A coded aperture pattern used to modulate the scene in the compressive imaging framework is optimized for the acquisition. In the classification part, the CNN parameters are also optimized following an optimization problem described below. After the training, in the testing part, a real measurement is acquired using the coded aperture designed, and the trained CNN classifies the lime patches in their corresponding class. A flowchart of the end-to-end approach is depicted in Fig. 7.

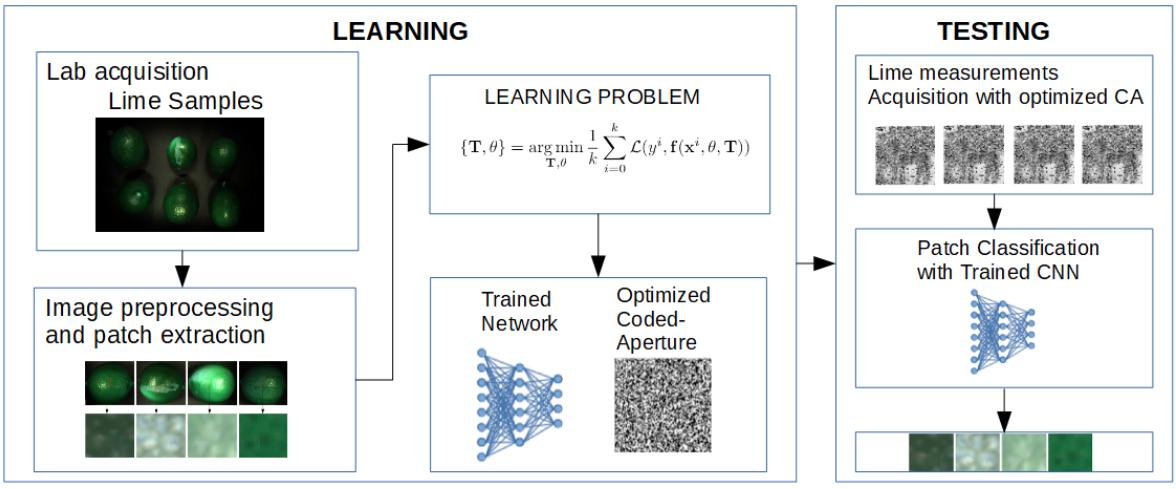



Figure 7. Flowchart of the general steps of the proposed end-to-end model for the simultaneously coded aperture design and lime patch classification. The two main steps are the learning and testing parts. Source: Author.

## 5.2. Coded aperture Design

The well-known coded aperture snapshot spectral imager (CASSI) [9] architecture is the compressive spectral system used in this work because it is suitable for the spectral classification task, as reported in the literature. The CASSI is one example of a CS architecture that captures 2D projections by multiplexing the spatio-spectral information of a scene through a coded aperture and a dispersive element such as a prism or a diffractive device. The coded apertures are implemented by use of a photomask, an SLM, or a digital micromirror device (DMD). The latter two provide versatility as the coding is easily modified with subsequent shots or snapshots. As a result, a multidimensional cube is generated representing the whole compressed scene.

The Coded aperture (CA) design is a key component in the compressive acquisition system [40]. It modulates the scene spatially and/or spectrally such that an optimal image reconstruction or an optimal classification accuracy can be achieved. Scene modulation is made by applying one or multiple CA patterns or designs to the scene. A color CA is an arrangement of more than one CAs that modulates specific bands of the scene. On the other hand, a binary CA is a single CA design whose pattern is applied to every spectral band of the scene. In this work, only binary CAs will be used, resulting in only spatial modulation. The CA is defined as,

$$\mathbf{T} \in \mathbb{R}^{MN}, \tag{4}$$

where $M$ and $N$ are the spatial dimensions. Since only spatial modulation will be performed, the same code design $\mathbf{T}$ is applied to all the spectral bands of the scene.

---

[40] Laura Galvis et al. “Shifting colored coded aperture design for spectral imaging”. In: *Appl. Opt.* 58.7 (2019), B28–B38. DOI: 10.1364/AO.58.000B28.

The acquisition of scene $\mathbf{F}$, in the CASSI system and applying the spatial modulation $\mathbf{T}$ can be expressed as,

$$\mathbf{Y}_{mn} = \sum_{k=0}^{L-1} \mathbf{F}_{m(n-k)k} \mathbf{T}_{mn} + \boldsymbol{\omega}_{mn}, \tag{5}$$

where $\mathbf{Y}$ represents the measurements, $\mathbf{F}$ is referring to the data cube, $\mathbf{T}$ defines the coded aperture, and $\boldsymbol{\omega}$ is the noise in the system. The dimensions of the data cube $\mathbf{F}$ are $M \times N \times L$, $\mathbf{T}$ is $M \times N$ , and $\mathbf{Y}$ is $M \times (N + L - 1)$. $L$ represents the number of spectral bands. The resulting shifting in the measurements is the dispersive element effect, characteristic in the CASSI system.

### 5.3. Tahiti Lime Classification

The goal of the grading process of Tahiti lime is to accept or reject a whole fruit based on its healthiness and compliance with target-market standards and agriculture regulators. However, the fact that fruit has been affected by a fungus or mite or any other kind of defect does not mean that it be automatically rejected. Some levels of affectation are accepted. To reject a whole fruit, it must have affectation levels beyond the target market or regulator standards. In some cases, 10% of mite or fungus affectation is permitted. Due to this, a one-shot fruit classification will require huge quantities of training samples, which is a drawback of traditional classification systems. Instead, a patch-wise approach was implemented. This approach has two stages: the first one is used to build the training and testing database. The second to perform automated classification. In the first stage, segments of known anomalies are extracted by cropping the whole lime images with the help of a lime farmer who identifies the place, size, and type of the anomaly. Those segments are cropped again to create the patches of $16 \times 16 \times 40$ pixels using an image processing tool. Then, each patch is labeled, and finally stored as the training and testing database

of the CNN classifier. In the second stage, once the CNN has been simultaneously trained and the coded aperture optimized, it is possible to perform a lime classification in the patch basis. For this, a background removal algorithm [41] is applied to the lime image sample. Then, the Compact Whatershed unsupervised segmentation algorithm [42] is applied. Once the lime image is segmented, the trained network can be applied to every segment, thus resulting in a more precise and efficient classification. This image-processing workflow is depicted in Fig.8. In this approach both anomaly and amount of surface affected can be inferred. It is up to the farmer to set the affectation allowance in order to meet its target-market requirements. Since the image processing techniques reported before have been widely studied, only the trained network used for the patch classification will be developed in this work.

### 5.4. Proposed training model

In order to optimize both the coded aperture pattern and the parameters of the classification CNN, the training architecture depicted in Fig. 9 is proposed. The input data, which can be gray-scale, RGB, or spectral images, is pre-processed to extract the patches with some anomalies. Then, the CASSI system is simulated using an initial random coded aperture, which spatially modulates the patches. Then, a CNN is used to find both optimal coded aperture design and CNN parameters by solving the following optimization problem,

---

[41] Andrews Sobral and Antoine Vacavant. “A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos”. In: *Computer Vision and Image Understanding* 122 (2014), pp. 4–21. DOI: https://doi.org/10.1016/j.cviu.2013.12.005.

[42] Peer Neubert and Peter Protzel. “Compact Watershed and Preemptive SLIC: On Improving Trade-offs of Superpixel Segmentation Algorithms”. In: *2014 22nd International Conference on Pattern Recognition*. 2014, pp. 996–1001. DOI: 10.1109/ICPR.2014.181.

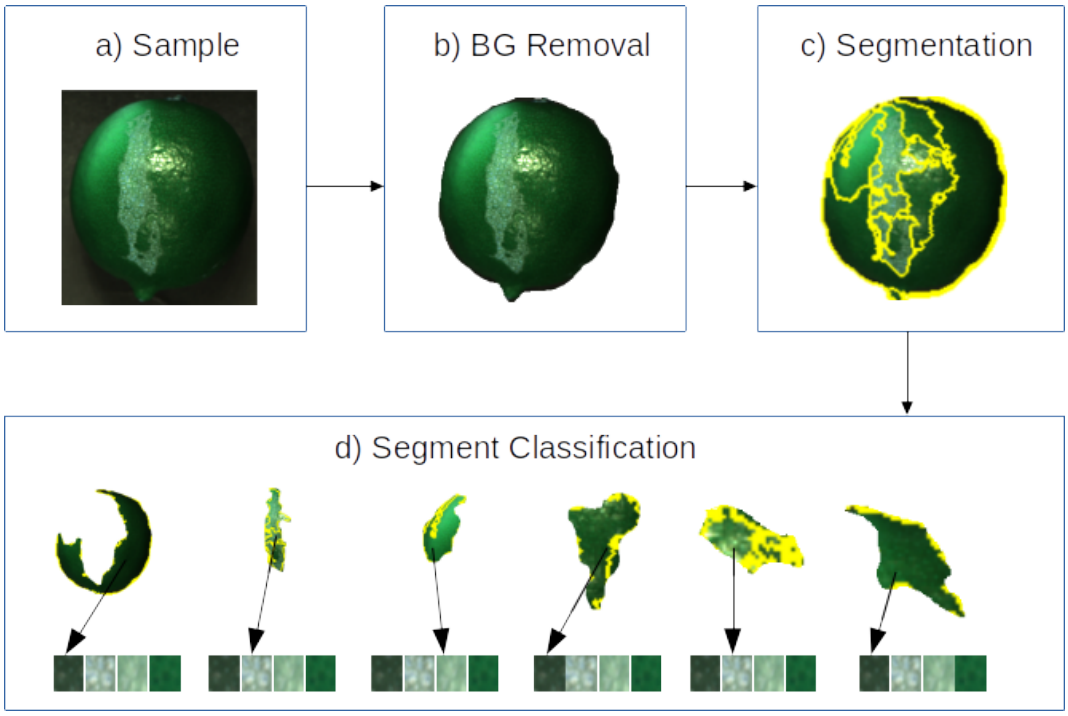



Figure 8. Lime classification workflow. a) A lime image is acquired. b) Image background is removed. c) Image segmentation algorithm is applied. e) Segment classification using a trained network is performed over each segment. Source: Author.

$$\{\mathbf{T}, \theta\} = \arg\min_{\mathbf{T},\theta} \frac{1}{k} \sum_{i=0}^{k} \mathcal{L}(y^i, \mathbf{f}(\mathbf{x}^i, \theta, \mathbf{T})), \tag{6}$$

where $\mathcal{L}$ is a loss or cost function, $\mathbf{f}$ is the output of the trained model, $\mathbf{T}$ is the coded aperture, $\theta$ are the CNN parameters, k is the number of training samples, $y^i$ are the training labels, and $\mathbf{x}^i$ are the training samples.

A regularizer [43] can be added to Eq. 6 to make coded aperture entries to converge to 1's or 0's as shown in Eq. 7,

---

[43] Catherine F. Higham et al. "Deep learning for real-time single-pixel video". In: *Scientific Reports* 8.1 (2018), p. 2369. DOI: 10.1038/s41598-018-20521-y.

$$\{\mathbf{T}, \theta\} = \arg\min_{\mathbf{T},\theta} \frac{1}{k} \sum_{i=0}^{k} \mathcal{L}(y^i, f(\mathbf{x}^i, \theta, \mathbf{T})) + \mu(\mathbf{1} + \mathbf{T})^2(\mathbf{T})^2, \tag{7}$$

where $\mu$ is a regularizer constant. This regularization is included with the aim to produce implementable coded aperture designs.

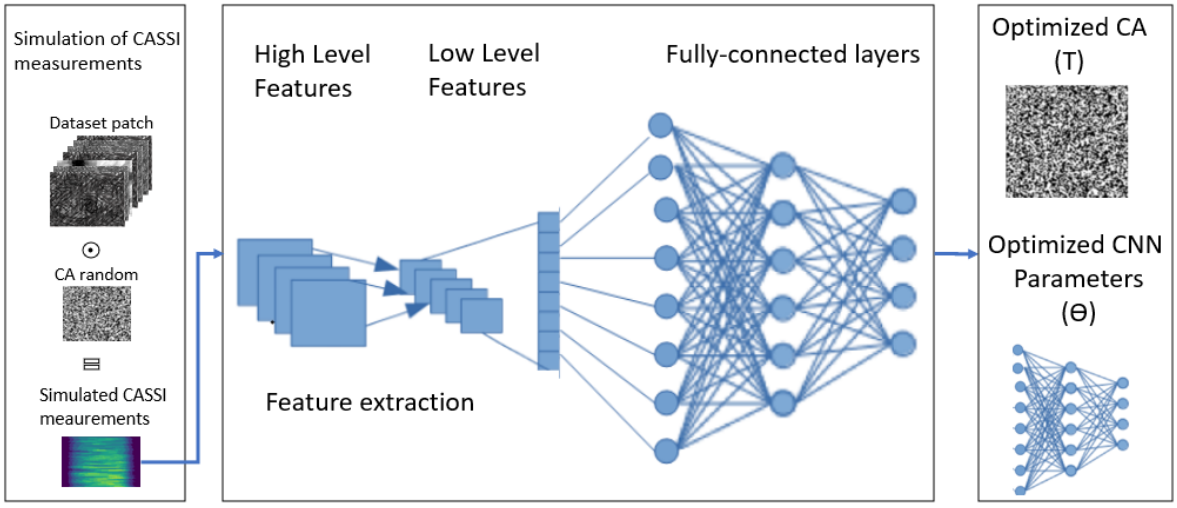



Figure 9. Proposed CNN training model. The input for the training is the set of lime patches. A complete CNN architecture is used to optimize the coded aperture pattern, as well as the classification CNN parameters. The output is an optimized coded aperture and the optimized CNN parameters to be used in the testing step. Source: Author.

### 5.5. Proposed inference/testing model

Once an optimal coded aperture design and the trained network are achieved from the training stage, real lime measurements can be acquired in the lab to perform their respective classification. The CASSI system can be implemented or simulated to acquire the lime measurements in a single shot, using the designed coded aperture obtained in the training. Then, the resulting compressive measurements are used as the input of the trained network, which classifies the observed patch. The label of

the lime class is the output of the inference pipeline. Figure 10 depicts the described testing process.

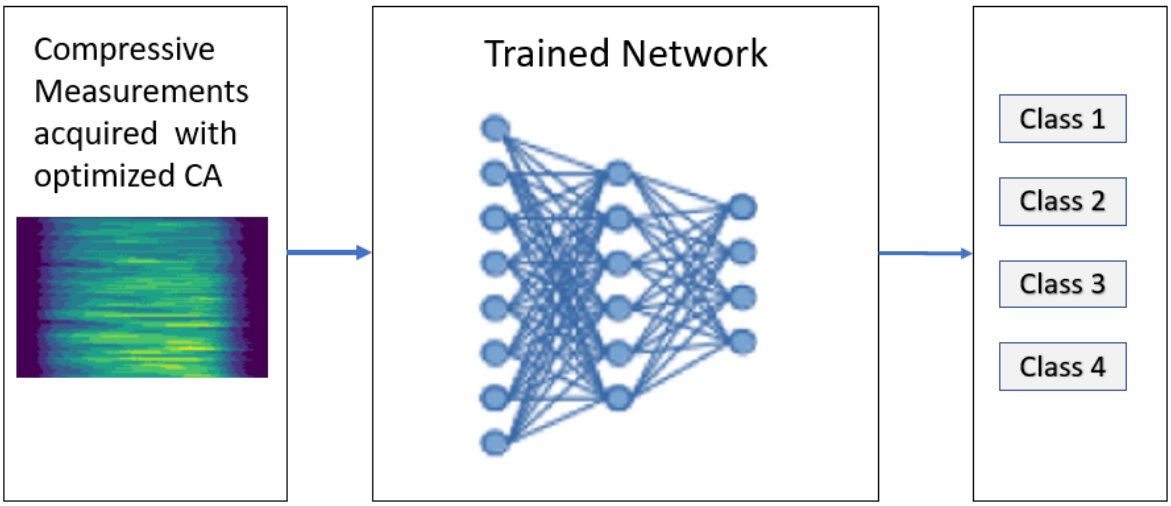



Figure 10. Proposed CNN inference model. The input of this step is the output of the training model. Real measurements can be acquired using the CASSI system and the designed coded aperture. Then these measurements are used to classify the lime patch in one class using the trained CNN. Source: Author.

# 6. SIMULATIONS AND RESULTS

In this chapter, data is acquired for training and testing of the proposed model for Tahiti lime classification. To this end, an optical setup, samples of different lime categories, a labelling scheme, and different CNN networks are implemented.

## 6.1. Optical setup

An optical setup was assembled in the HDSP research group's laboratory to acquire the lime spectral database. The sample scene was focused by a CCD Monochrome Camera (Stingray camera F-080B) coupled with a macro lens (Tamron, Co., Ltd; 8mm, 1.1", C mount Lens). A tunable light source (Oriel Instruments, TLS-300XR) was used to decompose the source light in 40 spectral bands ranging from 400 nm to 700 nm. The whole setup is depicted in Figure 17.

## 6.2. Hardware and software for the database acquisition

The CCD Monochrome Camera and the tunable light source were controlled by a Desktop-PC equipped with an Intel Core i7 processor and 16 GB of RAM. Matlab (The MathWorks, Inc., Natick, Massachusetts, United States.) scripts were used to control the acquisition process.

## 6.3. Data acquisition and pre-processing

Instead of using several images, a patch-wise approach has been applied to the $72$ image samples acquired in the lab. The complete size of these images is $1388 \times 1038 \times 40$ pixels. This approach let us construct a spectral database of $2442$ patches of $16 \times 16 \times 40$ pixels size, with which the proposed deep learning model was trained.

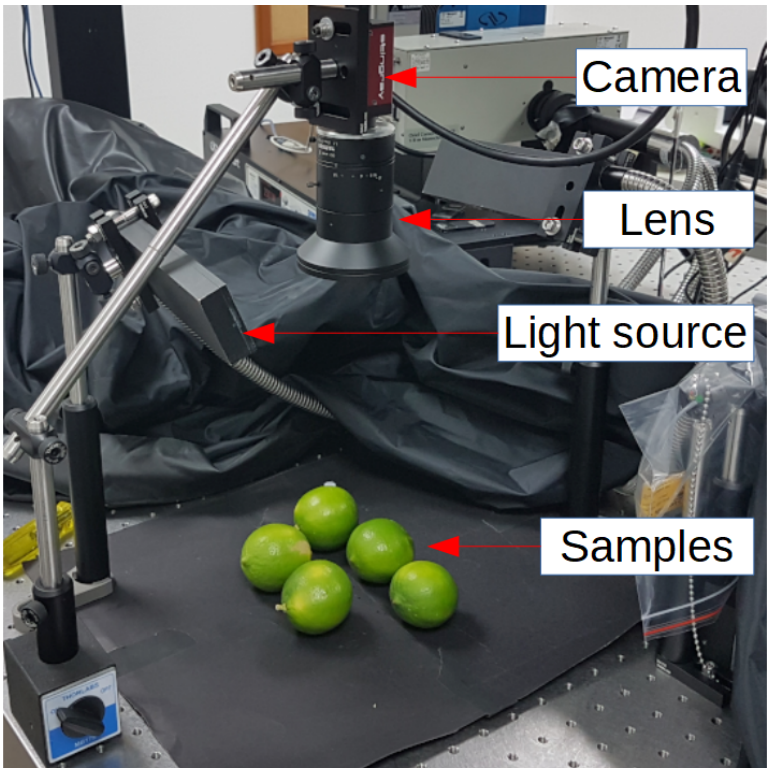



Figure 11. Optical setup for dataset acquisition. The samples are illuminated with a tunable light source, which allows the 40 spectral bands acquisition. Source: Author.

The database spectral range is the visible spectrum from $400nm$ to $700nm$. The following patch classes were defined: Healthy or European-market-ready fruits, fungus or mite-affected fruits, shadow-grown fruits, and wood-pocket affected fruits. These classes are shown in Figure 12, and they were suggested by a Tahiti lime exporter[44], who performs manual grading of their fruits. Patch samples of both training and testing databases are shown in Table 1.

In order to tackle the category imbalance problem we have applied oversampling techniques[45] to the whole lime spectral database. This technique let us to copy

---

[44] Orange Export S.A.S. www.orange-export.com

[45] Mateusz Buda, Atsuto Maki, and Maciej A. Mazurowski. "A systematic study of the class imbalance problem in convolutional neural networks". In: *Neural Networks* 106 (2018), pp. 249–259.

some training and testing samples in least populated categories in order to match the cardinality of most populated category. This way a same number of samples per category can be achieved avoiding the most populated classes to bias the results.

### 6.4. CNN implementation

CNN networks were implemented in Python programming language using the Pytorch library [46]. Lime patches required for the training and testing database were manually extracted with the Gimp (https://www.gimp.org/) image processing software and Python's Numpy library. Parameters of each CNN tested were trained with an Adam based optimizer, an initial learning rate of $0.0001$, and a Cosine Annealing learning rate scheduler.

Table 1. Database of lime samples used for training and testing. The number of samples used for training and testing for each of the classes is reported.

| **Class Name** | **Training** | **Testing** |
|---|---|---|
| Healthy | 576 | 100 |
| Mite affected | 488 | 102 |
| Shadow-grown | 484 | 100 |
| Wood pocket affected | 496 | 96 |
| | —- | —- |
| **Totals** | 2044 | 398 |

DOI: https://doi.org/10.1016/j.neunet.2018.07.011.

[46] Adam Paszke et al. "PyTorch: An Imperative Style, High-Performance Deep Learning Library". In: *Advances in Neural Information Processing Systems 32*. Ed. by H. Wallach et al. Curran Associates, Inc., 2019, pp. 8024–8035.

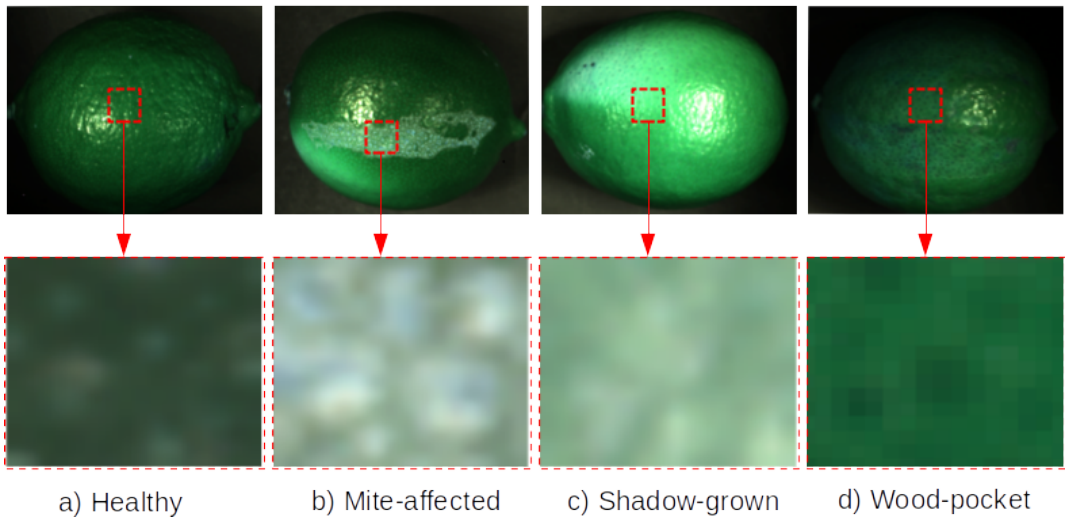



Figure 12. Patch extraction process. For the database acquisition, several patches out of a whole fruit sample are extracted. This process is performed over lima samples belonging to the four different classes. The zoomed sections are patches examples. Source: Author.

### 6.5. Simulations and Results

Three different CNN architectures were selected to test the accuracy of the proposed end-to-end approach: a simple CNN vanilla architecture, Alexnet [47], and Resnet34 [48]. Vanilla is a custom 6-layer architecture intended to be a baseline for benchmarking deeper architectures. Alexnet is an 11-layer, well-known, and proven deep learning architecture. Also, a state-of-the-art Resnet architecture with 34-layers was used for performance testing. A block diagram for each architecture is depicted in figure 13.

Regarding the datasets, besides the custom 40-spectral-band database of Tahiti lime

---

[47] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "ImageNet Classification with Deep Convolutional Neural Networks". In: *Advances in Neural Information Processing Systems*. Ed. by F. Pereira et al. Vol. 25. Curran Associates, Inc., 2012.

[48] Kaiming He et al. "Deep Residual Learning for Image Recognition". In: *CoRR* abs/1512.03385 (2015). arXiv: 1512.03385.

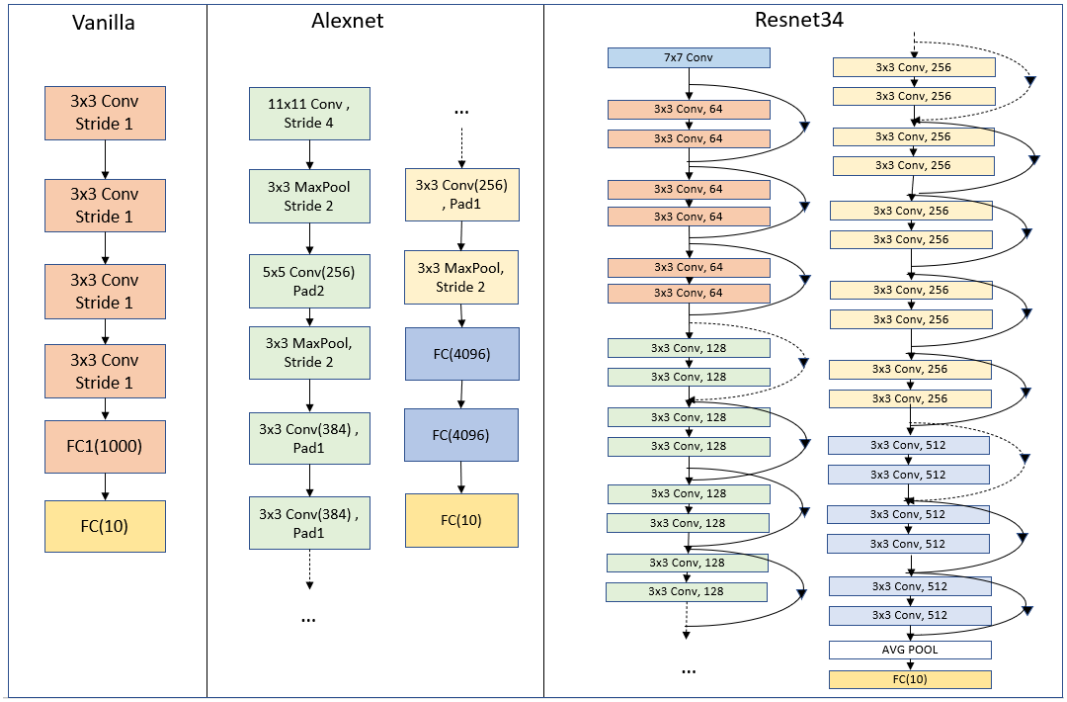



Figure 13. CNN architectures used for lime classification. Source: Author.

aforementioned, a new database was built by selecting three spectral bands out of the 40 of the original spectral database. This experiment is intended to show the contribution of spectral information. In addition, the MNIST-Fashion database [49] also will be used for testing purposes of the end-to-end approach, this dataset is reported in figure 14.

An optimization process was performed for each CNN architecture using the proposed model and the dataset shown in table 1. Once the coded apertures were optimized, several measurements were simulated using the CASSI model. Figure 15 shows the simulated measurements for four different affectations using the optimized coded apertures and the 40-band dataset. Finally, these measurements were classified and compared to the ground truth class labels.

---

[49] Han Xiao, Kashif Rasul, and Roland Vollgraf. "Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms". In: *CoRR* abs/1708.07747 (2017). arXiv: 1708.07747.

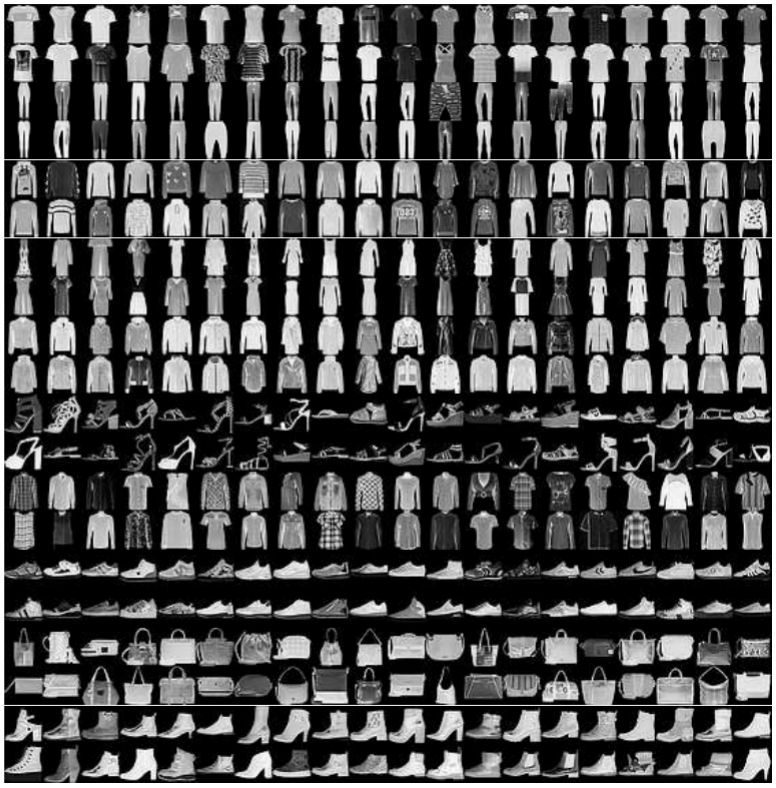
| Label | Description | Examples |
|---|---|---|
| 0 | T-Shirt/Top |  |
| 1 | Trouser | |
| 2 | Pullover | |
| 3 | Dress | |
| 4 | Coat | |
| 5 | Sandals | |
| 6 | Shirt | |
| 7 | Sneaker | |
| 8 | Bag | |
| 9 | Ankle boots | |

Figure 14. Fashion MNIST dataset. Source: [50].

The Resnet34 architecture outperformed the vanilla and Alexnet architectures over all three datasets in up to $20\%$ accuracy, as shown in table 2. However, the vanilla architecture outperformed the Alexnet in the two Tahiti lime datasets, nonetheless that it was five layers shallower. The performance of the different architectures over the MNIST-fashion dataset shows us that the deeper the network, the better the results. The performance achieved by the three CNN architectures is higher when using the 40 spectral bands, which demonstrates the advantage of using spectral instead of RGB data.

An additional set of simulations were performed to demonstrate the improvement achieved by the optimized coded apertures. The same approach but using a random pattern instead of the designed CA is also presented in table 2. These results are

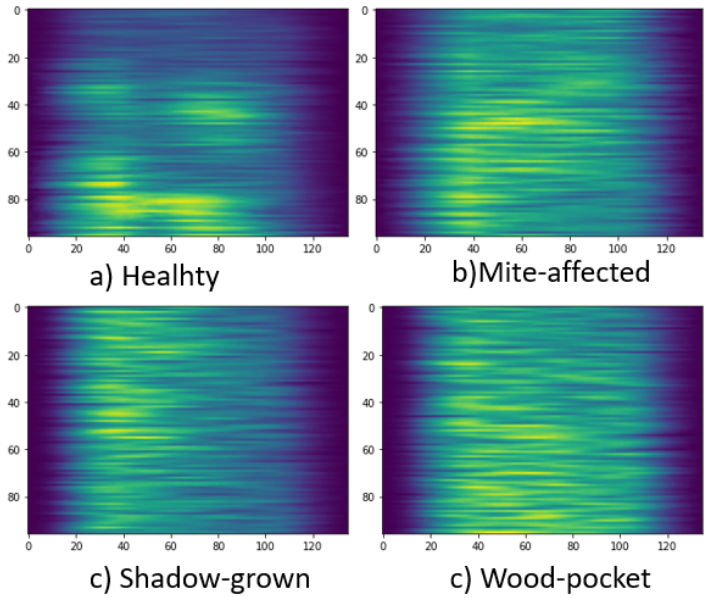



Figure 15. Simulated measurements. a) Healthy. b) Mite-affected. c) Shadow grown. d) Wood-pocket. Source: Author.

Table 2. CNN Accuracy over test datasets. The customs 40-band and 3-band correspond to the acquired lime datasets. A comparison of the classification results with the optimized and non-optimized CA is presented.

| **Dataset** | **Coded aperture** | **CNN Architecture** | | |
|---|---|---|---|---|
| | | **Vanilla** | **Alexnet** | **Resnet34** |
| Custom 40-band | Optimized CA | **96**% | **84**% | **99**% |
| | Non-Optimized CA | 94% | 77% | 97% |
| Custom 3-band | Optimized CA | **94**% | **80**% | **98**% |
| | Non-Optimized CA | 93% | 76% | 94% |
| Fashion MNIST | Optimized CA | **91**% | **92**% | **94**% |
| | Non-Optimized CA | 90% | 91% | 93% |

referred to as Non-optimized CA in the table. As can be seen, in all the experiment combinations (CNN and dataset), the optimization of the CA design reported the highest accuracy.

Figure 16 presents the optimized coded aperture patterns for each CNN architecture and for the 40-band lime dataset. It can be seen that patterns did not converge to a

particular shape, due to the little spatial variability of the dataset. The confusion matrices also reported in Figure 16 shows the mite-affected and shadow-grown patches were the most challenging anomalies to classify. Figure 16 also presents the accuracy behavior for the different network architectures. Notice that the Resnet34 only needs 6 epochs to converge, whereas Alexnet and Vanilla need more than 10 to obtain a similar behavior.

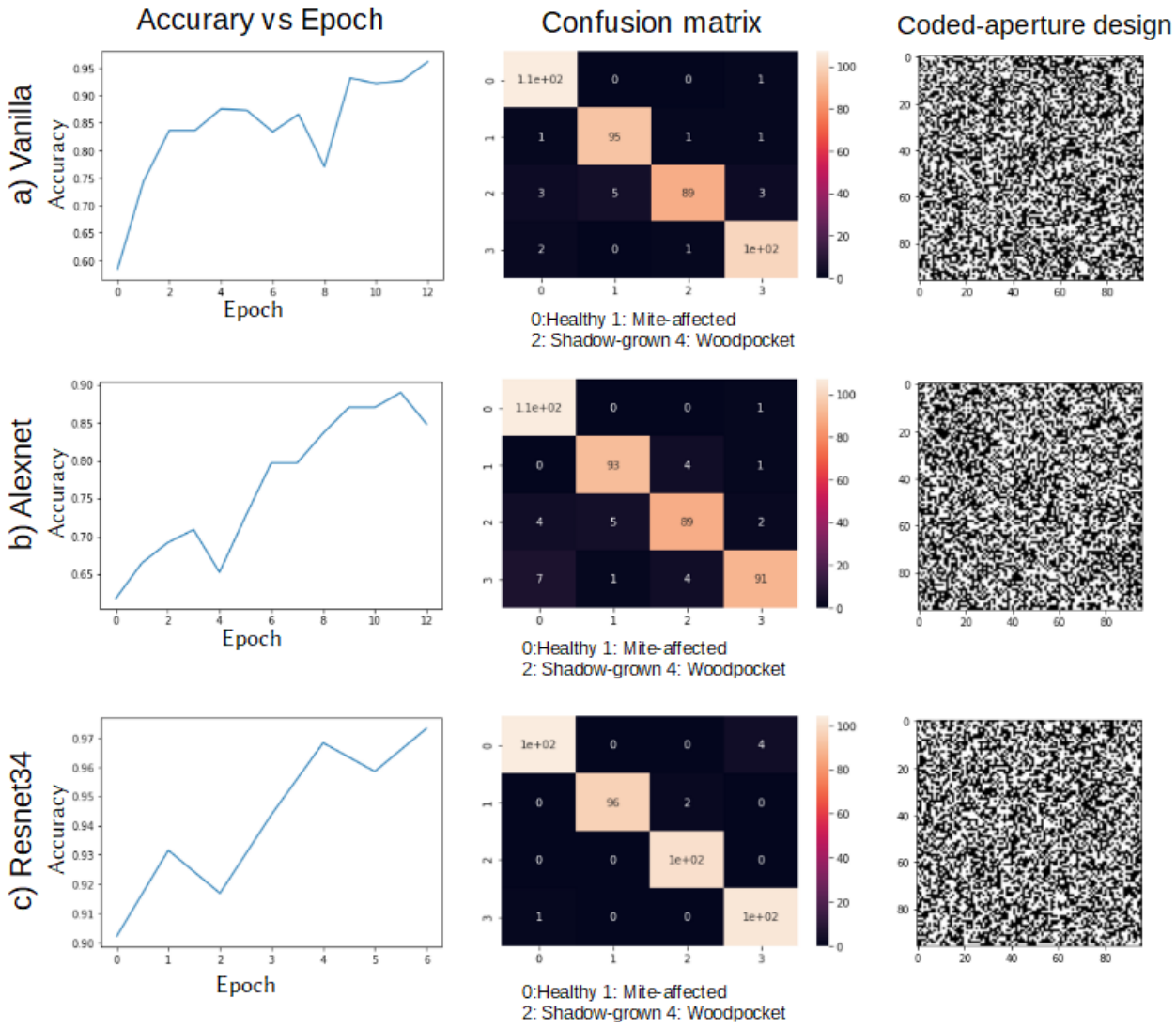



Figure 16. Accuracy, confusion matrix, and CA design for each CNN architecture selected using the 40-band lime dataset. Source: Authors.

# 7. EXPERIMENTAL RESULTS

To verify the performance of the proposed method a real experiment was carried out over a new database composed of samples acquired using the side-information CASSI optical setup.

## 7.1. Optical Setup

To acquire the database with the aforementioned architecture, the setup shown in figure 17 was used: The sample scenes were focused by a CCD Monochrome Camera (Stingray camera F-080B) coupled with a Amici Prism (Shanghai Optics), and a macro lens (Tamron, Co., Ltd; 8mm, 1.1”, C mount Lens). A high-speed digital micro-mirror device (DMD) (Texas Instruments-DLI4130) was used as spatial modulator. A tunable light source (Oriel Instruments, TLS-300XR) was used to decompose the source light in 40 spectral bands ranging from 400 nm to 700 nm.

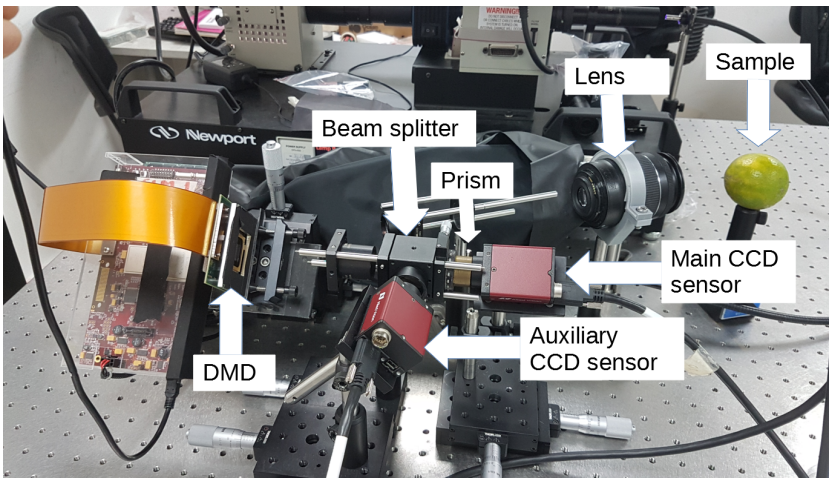



Figure 17. Optical setup for dataset acquisition. The samples are illuminated with a tunable light source, which allows the 40 spectral bands acquisition. Source: Author.

### 7.2. Dataset

A new dataset was built using the same approach used for simulations, i.e., extracting patches out of whole samples. The size of the real patches used is $96 \times 96$ pixels. The number of samples of both training and testing dataset are shown in table 3. A set of $16$ samples of the three classes were taken to verify the performance of the proposed methodology over real measurements. Each sample was acquired with the optimized and the non-optimized coded apertures. Some samples of the patches in specific bands are shown in figure 18. The real measurements for three of the testing samples are reported in figure 19.

Table 3. Database of lime samples used for training and testing. The number of samples used for training and testing for each of the classes is reported.

| **Class Name** | **Training** | **Testing** |
|---|---|---|
| Healthy | 621 | 8 |
| Mite affected | 184 | 4 |
| Wood pocket affected | 450 | 4 |
| | —- | —- |
| **Totals** | 1255 | 16 |

### 7.3. Coded Aperture optimization

The non-optimized coded aperture and the optimized coded aperture for the Vanilla CNN using the end-to-end approach are shown in figure 20. It can be seen that the optimized coded aperture shape does not change drastically with respect to the non-optimized due to the low spatial variability of the samples; however, as will be reported, the performance achieved with each code is different. The transmittance of both coded apertures is $50\%$.

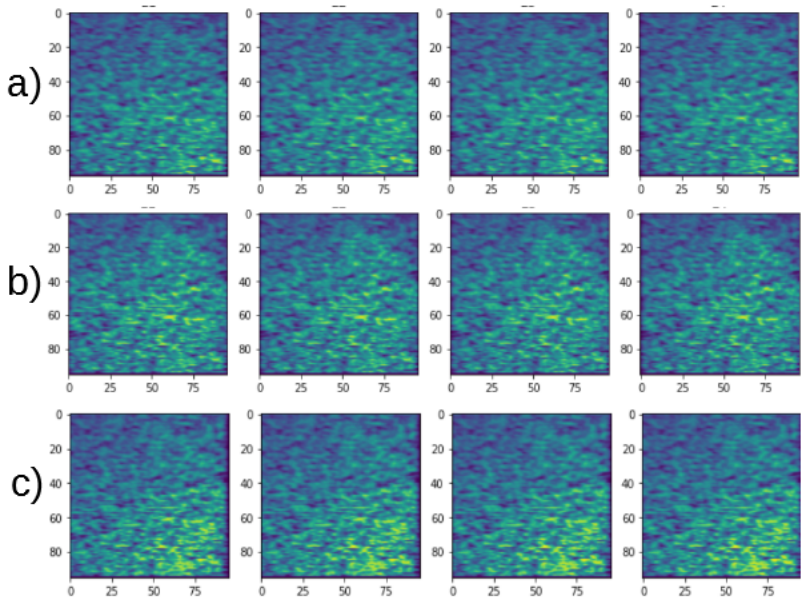



Figure 18. Samples of real measurements for each lime class in the spectral bands 21, 22, 23, and 24. a) Healthy. b) Mite-affected. c) Wood-pocket. Source: Author.

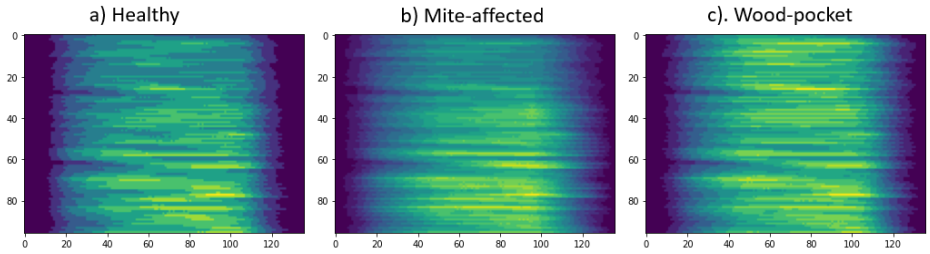



Figure 19. Real measurement samples. a) Healthy. b) Mite-affected. c) Wood-pocket. Source: Author.

### 7.4. CNN implementation

In this experiment, only a Vanilla CNN architecture was used. The CNN implementation was made using the Pytorch library and the traning was perfomed using the Colab platform. Hyperparameters of this CNN were trained with an Adam based optimizer, an initial learning rate of $0.0001$, and a Cosine Annealing learning rate scheduler. The number of epochs was fixed to $20$. The average training time per epoch was $10$ minutes.

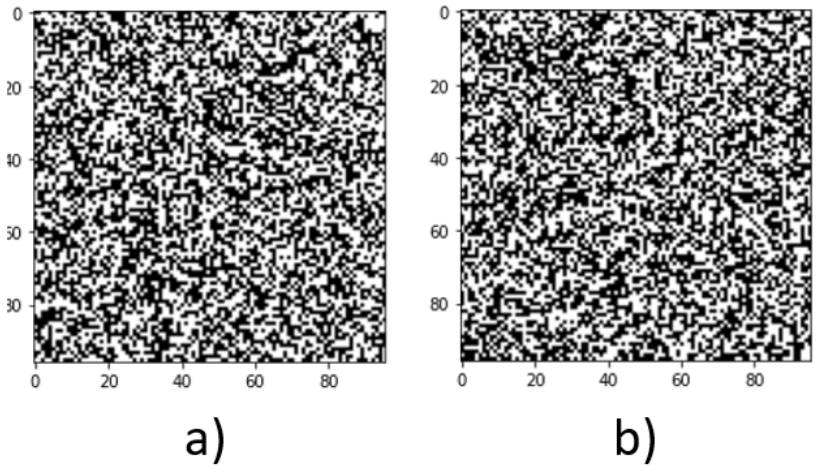



Figure 20. Coded apertures. a) Non-optimized. b) Optimized for CNN Vanilla. Source: Author.

### 7.5. Results

Both Alexnet and Resnet34 networks were affected by overfitting, that is, the testing accuracy was lower that training accuracy. Because of this, the experimental results were perfomed using the CNN Vanilla, which is the simpler architecture and had a good behavior in simulations. The score over the 3-classes dataset is shown in table 4. The accuracy of this CNN increased up to $6\%$ when an optimized CA is used. Since the optical architecture is an spectral setup and this experiment uses the real measurements, only the 40-band results are reported. Figure 21 shows the confusion matrices for the CNN using the optimized and non-optimized coded apertures. These results demonstrate the classification accuracy increases when the coded apertures are optimized, as was shown in simulations, and additionally verifies that classification can be performed over compressed measurements without requiring the previous reconstruction of the spectral image.

Table 4. CNN Accuracy over 40-band test dataset. A comparison of the classification results with the optimized and non-optimized CA is presented.

| **Dataset** | **Coded aperture** | **CNN Vanilla** |
|---|---|---|
| Custom 40-band 3-Classes | Optimized CA | **94**% |
| | Non-Optimized CA | 88% |

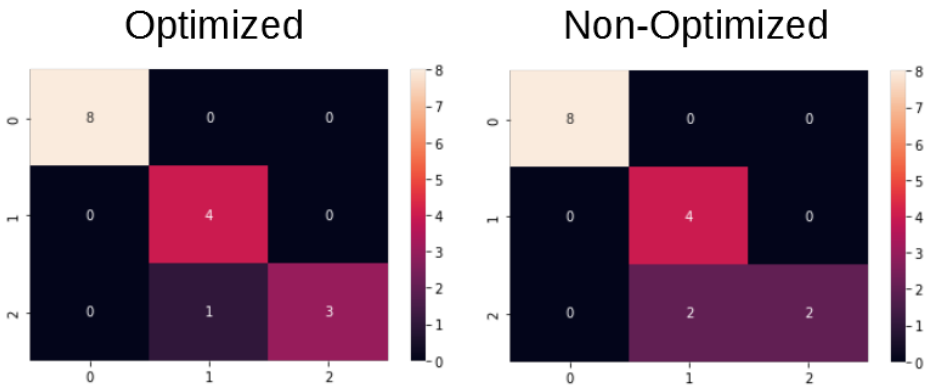



Figure 21. Confusion Matrix for testing experiment using real measurements and three classes: healthy, mite-affected, and wood-pocket. Source: Author.

# 8. CONCLUSIONS

The proposed methodology for spectral image classification using an end-to-end deep learning optimization approach has been developed, simulated, and tested in a real setup for an agricultural application. Instead of building the database using whole specimen samples, which are more complex and time-consuming to acquire, a patch-wise approach was performed. The experiments show that optimizing the coded apertures of the compressive sensing optical setup improved the classification accuracy by $7\%$ compared with the non-optimized ones when a simulation setup was used and up to $6\%$ when real measurements were tested. The CASSI system was implemented and used to prove the methodology, the model, and the optimizations. The accuracy of the model, when tested in simulations over the 40-band dataset, outperformed the 3-band dataset, which confirms the intuition that additional spectral information improves the classification accuracy. As to the CNN architectures used, it can be said that the Resnet34 performed the best on simulations but tended to overfit in real experiments. The Alexnet network performed well on simulations but also tended to overfit in real experiments. The shallow architecture of the CNN Vanilla performed well both in simulations and real experiments. It can be inferred that this CNN's behavior is tied to the particular nature of the dataset. Finally, it was demonstrated through simulations and real experiments that the classification could be performed over compressed measurements instead of reconstructing first the spectral image.

## BIBLIOGRAPHY


Arce, Gonzalo R. et al. *Compressive coded aperture spectral imaging: An introduction*. 2014. DOI: 10.1109/MSP.2013.2278763 (cit. on p. 22).

Audebert, Nicolas et al. "Deep Learning for Classification of Hyperspectral Data: A Comparative Review. IEEE geoscience and remote sensing magazine Deep Learning for Classification of Hyperspectral Data: " in: *IEEE* 7.2 (2019), pp. 159–173. DOI: 10.1109/MGRS.2019.2912563Ãŕ (cit. on p. 24).

Bacca, Jorge, Laura Galvis, and Henry Arguello. "Coupled deep learning coded aperture design for compressive image classification". In: *Opt. Express* 28.6 (), pp. 8528–8540. DOI: 10.1364/OE.381479 (cit. on p. 15).

Blumensath, Thomas and Mike E. Davies. "Iterative hard thresholding for compressed sensing". In: *Applied and Computational Harmonic Analysis* 27.3 (2009), pp. 265–274. DOI: 10.1016/j.acha.2009.04.002. arXiv: 0805.0510 (cit. on p. 14).

Bottou, Léon. "Large-Scale Machine Learning with Stochastic Gradient Descent". In: *Proceedings of COMPSTAT'2010*. Ed. by Yves Lechevallier and Gilbert Saporta. Heidelberg: Physica-Verlag HD, 2010, pp. 177–186 (cit. on p. 25).

Buda, Mateusz, Atsuto Maki, and Maciej A. Mazurowski. "A systematic study of the class imbalance problem in convolutional neural networks". In: *Neural Networks* 106 (2018), pp. 249–259. DOI: https://doi.org/10.1016/j.neunet.2018.07.011 (cit. on p. 36).

Chen, Yushi et al. “Deep Feature Extraction and Classification of Hyperspectral Images Based on Convolutional Neural Networks”. In: *IEEE Transactions on Geoscience and Remote Sensing* 54.10 (2016), pp. 6232–6251. DOI: 10.1109/TGRS.2016.2584107 (cit. on p. 24).

ChenXingyi, ZhangYujie, and QiRui. “Block Sparse Signals Recovery Algorithm for Distributed Compressed Sensing Reconstruction”. In: *Journal of Information Processing Systems* 15.2 (Apr. 2019), pp. 410–421 (cit. on p. 13).

Davenport, Mark A. et al. “Signal processing with compressive measurements”. In: *IEEE Journal on Selected Topics in Signal Processing* 4.2 (2010), pp. 445–460. DOI: 10.1109/JSTSP.2009.2039178 (cit. on p. 14).

Donoho, D. L. “Compressed sensing”. In: *IEEE Transactions on Information Theory* 52.4 (2006), pp. 1289–1306 (cit. on p. 12).

Donoho, David L. and Yaakov Tsaig. “Fast solution of l1-Norm minimization problems when the solution may be sparse”. In: *IEEE Transactions on Information Theory* 54.11 (2008), pp. 4789–4812. DOI: 10.1109/TIT.2008.929958 (cit. on p. 23).

Draganić, Andjela et al. “Compressive sensing approach in the table grape cold chain logistics”. In: *2017 6th Mediterranean Conference on Embedded Computing (MECO)*. 2017, pp. 1–4. DOI: 10.1109/MECO.2017.7977143 (cit. on p. 12).

Duarte, M. F. et al. “Single-pixel imaging via compressive sampling”. In: *IEEE Signal Processing Magazine* 25.2 (2008), pp. 83–91 (cit. on p. 13).

Figueiredo, Mário A.T., Robert D. Nowak, and Stephen J. Wright. “Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse

problems". In: *IEEE Journal on Selected Topics in Signal Processing* 1.4 (2007), pp. 586–597. DOI: 10.1109/JSTSP.2007.910281 (cit. on pp. 13, 22).

Galvis, Laura et al. "Shifting colored coded aperture design for spectral imaging". In: *Appl. Opt.* 58.7 (2019), B28–B38. DOI: 10.1364/AO.58.000B28 (cit. on p. 29).

Gehm, M. E. et al. "Single-shot compressive spectral imaging with a dual-disperser architecture". In: *Optics Express* 15.21 (Oct. 2007), p. 14013. DOI: 10.1364/oe.15.014013 (cit. on pp. 13, 29).

He, Kaiming et al. "Deep Residual Learning for Image Recognition". In: *CoRR* abs/1512.03385 (2015). arXiv: 1512.03385 (cit. on p. 38).

Higham, Catherine F. et al. "Deep learning for real-time single-pixel video". In: *Scientific Reports* 8.1 (2018), p. 2369. DOI: 10.1038/s41598-018-20521-y (cit. on p. 32).

Kingma, Diederik P. and Jimmy Lei Ba. "Adam: A method for stochastic optimization". In: *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. International Conference on Learning Representations, ICLR, Dec. 2015. arXiv: 1412.6980 (cit. on p. 25).

Krizhevsky, Alex, Vinod Nair, and Geoffrey Hinton. "CIFAR-10 (Canadian Institute for Advanced Research)". In: () (cit. on p. 15).

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E Hinton. *ImageNet Classification with Deep Convolutional Neural Networks*. Tech. rep. (cit. on p. 23).

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E Hinton. "ImageNet Classification with Deep Convolutional Neural Networks". In: *Advances in Neural Information*

*Processing Systems*. Ed. by F. Pereira et al. Vol. 25. Curran Associates, Inc., 2012 (cit. on p. 38).

Kulkarni, Kuldeep et al. “ReconNet: Non-iterative reconstruction of images from compressively sensed measurements”. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 2016-December (2016), pp. 449–458. DOI: 10.1109/CVPR.2016.55 (cit. on pp. 14, 23).

LeCun, Yann and Corinna Cortes. “MNIST handwritten digit database”. In: (2010) (cit. on p. 15).

LeCun, Yann et al. “Gradient-based learning applied to document recognition”. In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2323. DOI: 10.1109/5.726791 (cit. on p. 25).

Li, Qingli et al. “Review of spectral imaging technology in biomedical engineering: achievements and challenges”. In: *Journal of Biomedical Optics* 18.10 (2013), p. 100901. DOI: 10.1117/1.jbo.18.10.100901 (cit. on pp. 20, 21).

Li, Shutao et al. “Deep Learning for Hyperspectral Image Classification: An Overview”. In: *IEEE Transactions on Geoscience and Remote Sensing* 57.9 (2019), pp. 6690–6709. DOI: 10.1109/TGRS.2019.2907932 (cit. on p. 24).

— “Deep Learning for Hyperspectral Image Classification: An Overview”. In: *IEEE Transactions on Geoscience and Remote Sensing* 57.9 (2019), pp. 6690–6709. DOI: 10.1109/TGRS.2019.2907932 (cit. on p. 25).

Lin, Xing et al. “Spatial-spectral encoded compressive hyperspectral imaging”. In: *ACM Transactions on Graphics* 33.6 (2014). DOI: 10.1145/2661229.2661262 (cit. on p. 13).

Long, J., E. Shelhamer, and T. Darrell. “Fully convolutional networks for semantic segmentation”. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2015, pp. 3431–3440 (cit. on p. 27).

Mirrashid, Alireza and Ali Asghar Beheshti. “Compressed Remote Sensing by Using Deep Learning”. In: *2018 9th International Symposium on Telecommunications (IST)*. 2018, pp. 549–552. DOI: 10.1109/ISTEL.2018.8661112 (cit. on p. 12).

Nandhini, Aasha, S. Hemalatha, R. Radha, et al. “Web Enabled Plant Disease Detection System for Agricultural Applications Using WMSN”. In: *Wireless Personal Communications* (2018). DOI: https://doi.org/10.1007/s11277-017-5092-4 (cit. on p. 12).

Neubert, Peer and Peter Protzel. “Compact Watershed and Preemptive SLIC: On Improving Trade-offs of Superpixel Segmentation Algorithms”. In: *2014 22nd International Conference on Pattern Recognition*. 2014, pp. 996–1001. DOI: 10.1109/ICPR.2014.181 (cit. on p. 31).

Paszke, Adam et al. “PyTorch: An Imperative Style, High-Performance Deep Learning Library”. In: *Advances in Neural Information Processing Systems 32*. Ed. by H. Wallach et al. Curran Associates, Inc., 2019, pp. 8024–8035 (cit. on p. 37).

Sitzmann, Vincent et al. “End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging”. In: *ACM Transactions on Graphics* 37.4 (2018). DOI: 10.1145/3197517.3201333 (cit. on p. 26).

Sobral, Andrews and Antoine Vacavant. “A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos”. In: *Computer*

*Vision and Image Understanding* 122 (2014), pp. 4–21. DOI: https://doi.org/10.1016/j.cviu.2013.12.005 (cit. on p. 31).

Sun, Shiliang et al. “A Survey of Optimization Methods From a Machine Learning Perspective”. In: *IEEE Transactions on Cybernetics* (2019), pp. 1–14. DOI: 10.1109/tcyb.2019.2950779. arXiv: 1906.06821 (cit. on p. 24).

Sutskever, Ilya et al. *On the importance of initialization and momentum in deep learning*. Tech. rep. 2013 (cit. on p. 25).

Sze, Vivienne et al. “Efficient Processing of Deep Neural Networks: A Tutorial and Survey”. In: *Proceedings of the IEEE* 105.12 (2017), pp. 2295–2329. DOI: 10.1109/JPROC.2017.2761740. arXiv: 1703.09039 (cit. on p. 23).

*TE-Cooled Fluorescence Spectrometer | Edmund Optics* (cit. on p. 12).

Tropp, J. A. and A. C. Gilbert. “Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit”. In: *IEEE Transactions on Information Theory* 53.12 (2007), pp. 4655–4666 (cit. on p. 14).

Wagadarikar, Ashwin et al. “Single disperser design for coded aperture snapshot spectral imaging”. In: *Applied Optics* 47.10 (2008). DOI: 10.1364/AO.47.000B44 (cit. on pp. 13, 20).

Wang, Lizhi et al. “Adaptive Nonlocal Sparse Representation for Dual-Camera Compressive Hyperspectral Imaging”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39.10 (2017), pp. 2104–2111. DOI: 10.1109/TPAMI.2016.2621050 (cit. on p. 13).

Xiao, Han, Kashif Rasul, and Roland Vollgraf. “Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms”. In: *CoRR* abs/1708.07747 (2017). arXiv: 1708.07747 (cit. on pp. 39, 40).

Yang, Yunyun et al. “A New Compressed Sensing Model Based on Median Filter with Application to Reconstruct Brain MR Images”. In: *2018 IEEE 3rd International Conference on Signal and Image Processing (ICSIP)*. 2018, pp. 116–120. DOI: 10.1109/SIPROCESS.2018.8600479 (cit. on p. 12).

Zhang, Yan et al. “Compressive Sensing based Software Defined GPR for Subsurface Imaging”. In: *2021 IEEE Radar Conference (RadarConf21)*. 2021, pp. 1–6. DOI: 10.1109/RadarConf2147009.2021.9455291 (cit. on p. 12).

Zhao, Chen et al. “CREAM: CNN-REgularized ADMM Framework for Compressive-Sensed Image Reconstruction”. In: *IEEE Access* 6 (2018), pp. 76838–76853. DOI: 10.1109/ACCESS.2018.2882990 (cit. on p. 14).