Vol. 23, n.°4, pp. 17-30, 2024
Revista UIS Ingenierías
https://revistas.uis.edu.co/index.php/revistauisingenierias

# Spectral Classification using a Dual Optical Setup and Deep Neural Networks

# Clasificación espectral mediante una configuración óptica dual y redes neuronales profundas

**Andrés Jerez [1a], Geison Blanco [2a], Sergio Urrea [1b], Hans García [1c], Henry Arguello [2b]**

[1] Escuela de Ingeniería Eléctrica, Electrónica y Telecomunicaciones, Universidad Industrial de Santander, Colombia.
Orcid: 0000-0002-0078-9166 [a], 0000-0001-6699-1538 [b], 0000-0001-6542-1680 [c]. Emails:
andres.jerez1@correo.uis.edu.co [a], sergio2228328@correo.uis.edu.co [b], hans.garcia@saber.uis.edu.co [c]
[2] Escuela de Ingeniería de Sistemas e Informática, Universidad Industrial de Santander, Colombia.
Orcid: 0000-0003-4911-8344 [a], 0000-0002-2202-253X [b]. Emails: geison2180045@correo.uis.edu.co [a],
henarfu@uis.edu.co [b]

Received: 3 April 2024. Accepted: 11 august 2024. Final version: 20 November 2024.

## Abstract

Spectral classification allows material labeling based on spectral information. Single-pixel cameras (SPCs) have been used as a low-cost solution for acquiring spectral images, providing high-resolution spectral and low-resolution spatial information. Also, diffractive optical cameras (DOCs) based on multilevel phase masks (MPMs) can acquire spectral features to perform classification tasks. Traditional spectral classification approaches have not incorporated SPCs and DOCs into a single optical architecture. This work proposes a dual optical system based on SPC and DOC for spectral classification. Specifically, the height map in the MPM and the deep neural network parameters are jointly learned from end-to-end (E2E) optimization. The proposed method contains an optical layer that describes the dual system, a fusion layer that estimates the spectral image, and a classification network that labels the materials over spectral datasets. The simulation results show an improvement of up to 3% in classification metrics compared to other optical architectures.

**Keywords:** spectral classification; single pixel camera; diffractive optical camera; multilevel phase maskt; end-to-end optimization; deep neural networks.

## Resumen

La clasificación espectral permite etiquetar materiales basándose en información espectral. Las cámaras de un solo píxel (SPC) se utilizan como una solución de bajo costo para adquirir imágenes espectrales, proporcionando información espectral de alta resolución y espacial de baja resolución. Además, las cámaras ópticas difractivas (DOC) basadas en máscaras de fase multinivel (MPM) pueden adquirir características espectrales para realizar tareas de clasificación. Los enfoques tradicionales de clasificación espectral no han incorporado SPC y DOC en una única arquitectura óptica. Este trabajo propone un sistema óptico dual basado en SPC y DOC para la clasificación espectral. Específicamente, el mapa de altura en MPM y los parámetros de la red neuronal profunda se aprenden conjuntamente a partir de la optimización de un extremo a otro (E2E). El método propuesto contiene una capa óptica que describe el

ISSN Online: 2145 - 8456

How to cite: A. Jerez, G. Blanco, S. Urrea, H. García, H. Arguello, “Spectral Classification using a Dual Optical Setup and Deep Neural Networks,” *Rev. UIS Ing.*, vol. 23, no. 4, pp. 17-30, 2024, doi: https://doi.org/10.18273/revuin.v23n4-2024002

sistema dual, una capa de fusión que estima la imagen espectral y una red de clasificación que etiqueta los materiales en conjuntos de datos espectrales. Los resultados de la simulación muestran una mejora de hasta un 3% en las métricas de clasificación en comparación con otras arquitecturas ópticas.

**Palabras clave:** clasificación espectral; cámara de un solo píxel; cámara óptica difractiva; máscara de fase multinivel; optimización de extremo a extremo; redes neuronales profundas.

## 1. Introduction

Spectral imaging systems collect multiple wavelengths from a target across the electromagnetic spectrum. These images have been used for numerous applications, such as environmental monitoring [1], precision agriculture [2], aerospace [3], defense [4], and biomedicine [5]. Recently, spectral imaging systems have included diffractive imaging methods [6], named phase-coded spectral imaging systems. These phasecoded spectral systems have been designed relying on a diffractive optical camera (DOC) [6]. This architecture incorporates a diffractive optical element (DOE), also known as a multilevel phase mask (MPM), which can be easily integrated into a camera to modulate the input light phase for each wavelength. Specifically, the phase-coded spectral imaging formulates the image formation as a convolution process between the wavelength-specified point spread function (PSF) and the monochrome object image at each wavelength [6]. Here, the DOE manipulates the phase term of the PSF, distinguishing spectral signatures as light propagates. A suitable DOE design influences the PSF behavior, thereby leading to more accurate spectral reconstruction results [6]. Compared with amplitude coded spectral imaging, the phase-coded approach can greatly increase the light throughput.

On the other hand, the single-pixel camera (SPC) is widely used in amplitude-coded spectral imaging, providing reliable spectral images at lower-cost hardware compared to bidimensional spectral sensors [7]. The SPC integrates the spatial information into a single measurement, resulting in spectral images with high spectral resolution and low spatial resolution [8]. To address the spatial resolution limitation in SPC, the state-of-the-art has introduced dual optical architectures incorporating complementary monochromatic sensors. These architectures have significantly improved spectral image recovery performance, offering high spatial and spectral resolution [9], [10].

In addition to the reconstruction task inherent in spectral systems, spectral classification represents an important computational task [11], [12], [13]. Spectral classification involves categorizing objects or materials based on spectral signatures, which depict their interactions with electromagnetic radiation across different wavelengths [14]. These signatures are acquired using spectral imaging techniques, enabling the differentiation and identification of materials based on their distinct spectral properties [14]. In classification tasks, DOCs allow useful feature extraction [15]. Particularly, the height map in the MPM can be modeled through Zernike polynomials and optimized via an end-to-end (E2E) optimization framework. The E2E optimization approach jointly learns the sensing and the deep neural network parameters to address different computational imaging tasks, such as recovery, detection, classification, segmentation, and image fusion, among others [16], [17], [18]. In traditional spectral imaging techniques, dual optical architectures integrating both DOCs and SPCs for spectral classification tasks into an E2E approach have not been reported; combining phase encoding from DOC with other encoding architectures as SPC can boost classification performance.

Therefore, this work proposes a dual optical architecture based on SPC and DOC for spectral classification based on E2E optimization. The E2E deep spectral classification framework is used to jointly optimize the MPM and the network parameters to classify materials in land cover spectral datasets. The proposed E2E scheme is composed of three stages: an optical stage based on incoherent light that describes the propagation model tailored to the dual spectral setup leveraging the SPC and DOC; a spectral image fusion stage that computes the spectral image from the acquired measurements via the plug-and-play alternating direction method of multipliers (ADMM) algorithm; and a three-dimensional (3D) convolutional neural network (CNN) that labels the materials defined within the spectral datasets. The proposed E2E spectral classification based on the dual optical system achieves better classification performance than conventional optical setups. The experimental results validate the E2E deep spectral classification from the spatial-spectral measurements acquired by the implemented dual optical testbed.

## 2. Diffractive image forward model

This section introduces the diffraction image formation model by assuming spatially incoherent light. The diffractive imaging system often consists of an $MPM \phi(x', y')$ and a sensor separated by a distance $\mathcal{Z}$, as shown in Figure 1. Specifically, the MPM produces a phase delay in the incident wave $\varphi(x',y')$ before reaching the sensor.

Mathematically, the phase delay induced by an MPM can be expressed as

$$\phi(x',y') = \frac{2\pi\Delta\eta}{\lambda} h(x',y'), \tag{1}$$

where $\Delta\eta$ is the refractive index difference between air and the MPM material [19], $\lambda$ is the wavelength, and $h(x',y')$ is the height map. A wave field $\gamma_\lambda(x',y';z=0)$ with amplitude $A$ and phase $\varphi$ incident on the MPM will be affected as

$$\gamma_\lambda(x',y';z=0) = (x',y')e^{[j(\varphi(x',y')+\phi(x',y'))]}, \tag{2}$$

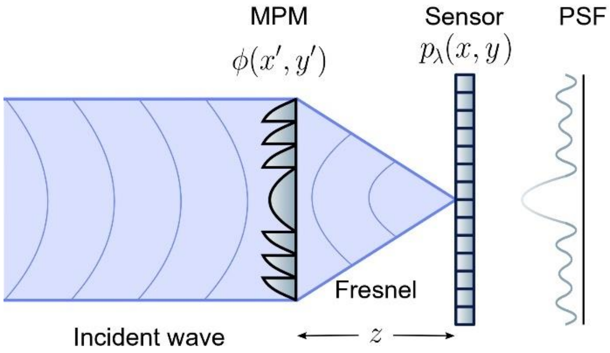


Figure 1. Diffractive imaging system based on an MPM. A light wave $\varphi(x',y')$ with a wavelength $\lambda$ incidents on the aperture plane containing an MPM $\phi(x',y')$. This MPM shifts the phase information present on the incident wavefront. The resulting wavefront is propagated to a distance $\mathcal{Z}$ until the sensor. The collected intensities into the sensor describe the PSF $p\lambda$ $(x,y)$.

where $j=\sqrt{-1}$, such that $\gamma_\lambda(x',y';0)$ is the wave field right after it passes through the MPM. As illustrated in Figure 1, the optical field is propagated a distance $\mathcal{Z}$ in free space, which can be modeled by the Fresnel approximation such that $\lambda \ll \mathcal{Z}$.

$$\gamma_\lambda(x,y;z) = \frac{e^{jk_0z}}{j\lambda z}\iint \gamma_\lambda(x',y';0)\, e^{\left[\frac{jk_0z}{2z}((x-x')^2+(y',y')^2)\right]}dx'\,dy', \tag{3}$$

Where $k_0 = \frac{2\pi}{\lambda}$ is the wavenumber.

From (3), the point spread function (PSF) $p\lambda$ is formulated as

$$p_\lambda(x,y) \propto \left|\mathcal{F}\left\{\gamma_\lambda(x',y',0)e^{\left[\frac{j\pi}{\lambda z}(x'^2+y'^2)\right]}\right\}\right|^2, \tag{4}$$

where $\mathcal{F}\{\cdot\}$ denotes the Fourier transform, and $|\cdot|$ denotes the magnitude operator.

## 3. Proposed Spectral Classification Methodology

The proposed methodology is mainly composed of three stages, as shown in Figure 2, where an optical acquisition stage models the forward propagation model according to the DOC and SPC; a spectral image fusion stage computes the spectral image; and a classification network labels the elements within the spectral image [14].

### 3.1. Dual optical forward model

Here, the proposed dual optical architecture is presented. More precisely, the proposed dual system consists of a double optical path architecture, where spatial-spectral features from the scene of interest are collected through an optical path described by a diffractive system; and another optical path characterized by a single pixel system.

### 3.1.1. Diffractive optical camera

After formulating the PSF in (4), the DOC image formation model can be modeled as [20].

$$Y = \sum_{\kappa=1}^{K} \mathrm{b}P_\kappa * X_\kappa + \Omega, \tag{5}$$

where $Y \in R^{n\times n}$ is the spatial measurement, $X_\kappa \in R^{n\times n}$ is the $k$-th spectral band, $P_\kappa \in R^{n\times n}$ denotes the discretized version of the PSF in (4), and $\Omega \in R^{n\times n}$ is the spatial sensor noise, and $*$ is the convolution operator. From (5), the DOC acquisition process can be rewritten as

$$\ddot{y} = \ddot{\Psi}x + \ddot{\omega}, \tag{6}$$

where $\ddot{\boldsymbol{y}} \in R^{n^2}$ is the vectorization of the spatial measurements $Y$, $\ddot{\boldsymbol{\Psi}} \in R^{n^2\times n^2K}$ is the DOC sensing matrix, and $\ddot{\boldsymbol{\omega}} \in R^{n^2}$ is the vectorization of the matrix $\Omega$.

### 3.1.2. Single pixel camera

From a spectral scene $\{X_\kappa\}_{\kappa=1}^{K}$ with $K$ as the spectral bands, the SPC image formation at the $l$-th snapshot can be modeled as [21].

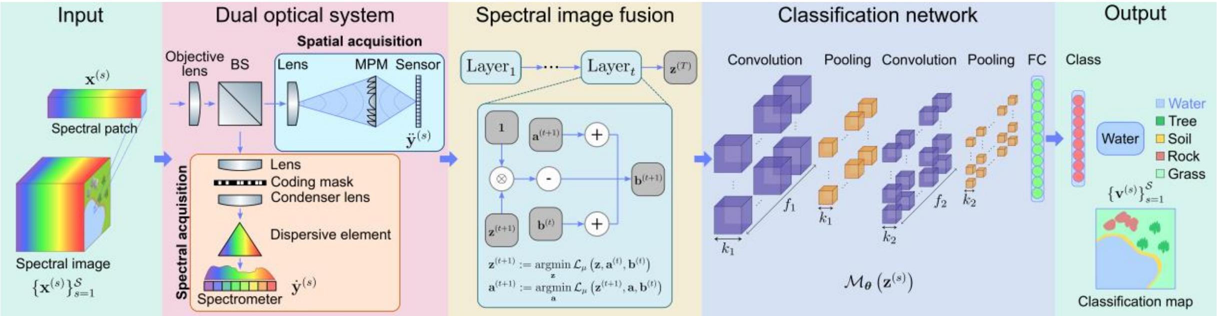


Figure 2. Proposed E2E spectral classification approach based on the SPC and DOC. The spectral patch $\mathbf{x}^{(s)}$ is equally divided by a beam splitter (BS), where two optical branches separately acquire the spatial-spectral measurements by using a diffractive camera and single-pixel architecture, respectively. Then, a spectral patch $\mathbf{z}^{(s)}$ is calculated from the acquired measurements $\dot{\mathbf{y}}^{(s)}$ and $\ddot{\mathbf{y}}^{(s)}$ by using the Plug-and-Play ADMM algorithm $\mathfrak{T}\theta^{(\cdot)}$. This spectral patch is the input of the classification network $\mathcal{M}\theta^{(\cdot)}$ based on 3D CNN, which is composed of 3D convolutions, 3D MaxPooling, and fully connected (FC) layers. Finally, the classification network returns the corresponding class $\mathbf{v}^{(s)}$.

$$(y_l)_\kappa = \sum_{p=1}^{n}\sum_{q=1}^{n}(D_l)_{p,q}(X_\kappa)_{p,q} + (\omega_l)_\kappa, \tag{7}$$

for $\kappa \in \{1, \cdots, K\}$ and $l \in \{1, \cdots, L\}$ with $L$ as the total number of snapshots ($L = n^2$), where $y_l \in R^K$ is the spectral measurement, $D_l \in \{-1,1\}^{n\times n}$ corresponds to a Hadamard pattern [8], and $\omega_l \in R^K$ is the spectral sensor noise. From (7), the SPC acquisition process is given by

$$(y_l)_\kappa = \psi_l^T \boldsymbol{x_\kappa} + (\omega_l)_\kappa, \tag{8}$$

where $\psi \in R^{n^2}$ is the vectorization of the l-th coding mask $D_l$, $x_\kappa \in R^{n^2}$ is the vectorization of the κ-th spectral band $X_k$, and $(\cdot)^{\mathcal{T}}$ is the transpose operator. The sensing model, including the total number of snapshots $L$, is defined as

$$y_k = \Psi x_k + \omega_k \tag{9}$$

with $y_\kappa = [(y_1)_\kappa, \cdots, (y_L)_\kappa]^{\mathcal{T}}$ and $\omega_\kappa = [(\omega_1)_\kappa, \cdots, (\omega_L)_\kappa]^{\mathcal{T}}$, where $\Psi \in R^{L\times n^2}$, specifically, $\Psi = [\boldsymbol{\Psi}_1, \cdots, \boldsymbol{\Psi}_L]^{\mathcal{T}}$. Finally, staking the measurements and the spectral noise from all spectral bands in a single vector $\dot{y} = [y_1^T, \cdots, y_K^T]^{\mathcal{T}}$ and $\dot{\omega} = [\omega_1^T, \cdots, \omega_K^T]^{\mathcal{T}}$, respectively, the sensing model for all spectral bands $K$ and snapshots $L$ can be expressed as

$$\dot{y} = \dot{\Psi} x + \dot{\omega}, \tag{10}$$

where $\dot{\Psi} \in R^{LK\times n^2 K}$ is the SPC sensing matrix, and $x \in R^{n^2 K}$ is the spectral image.

### 3.2. Multilevel phase mask parameterization

To train the MPM values (1) using the proposed E2E optimization, the height map in the MPM is parameterized by using Zernike polynomials [22].

$$H = \sum_{i=1}^{O} \alpha_i \Xi_i \tag{11}$$

where $H \in R^{n'\times n'}$ corresponds to the discretized version of the height map in (1), Ξi denotes the *i*-th Zernike polynomial in Noll notation, and αi is the corresponding coefficient [23]. Each Zernike polynomial describes a wavefront aberration [20]. Thus, the MPM is formed by the linear combination of $O$ aberrations.

### 3.3. Spectral image fusion

Once the spectral information (10) and spatial information (6) are acquired, the spectral image $z \in R^{n^2 K}$ can be calculated from the following optimization problem

$$\text{argmin} \frac{1}{2}\left\|\dot{\boldsymbol{y}} - \dot{\boldsymbol{\Psi}}_{\boldsymbol{z}}\right\|_2^2 + \frac{1}{2}\left\|\ddot{\boldsymbol{y}} - \ddot{\boldsymbol{\Psi}}_{\boldsymbol{z}}\right\|_2^2 + \alpha\|\boldsymbol{z}\|_{TV}, \boldsymbol{z} \in \mathbb{R}^{n^2 k} \tag{12}$$

where $\dot{y} \in R^{LK}$ and $\ddot{y} \in R^{n^2}$ represent the spectral and spatial measurements obtained from SPC and DOC, respectively, $\alpha > 0$ is a regularization parameter and $\|\cdot\|_{TV}$ represents a TV regularization. The minimization problem in (12) can be solved by using the alternating direction method of multipliers (ADMM) algorithm [24], where the objective function and the constraint are split from auxiliary variables as below

$$\begin{gathered}\operatorname{argmin}\frac{1}{2}\left\|\dot{\boldsymbol{y}}-\dot{\boldsymbol{\Psi}}a_1\right\|_2^2+\frac{1}{2}\left\|\ddot{\boldsymbol{y}}-\ddot{\boldsymbol{\Psi}}a_2\right\|_2^2\\ +\alpha\|a_3\|_{TV},\\ \text{Subject } a_i=\boldsymbol{z},\end{gathered} \tag{13}$$

where $a_i \in R^{n^2K}$ is the split variable with $i \in \{1,2,3\}$. For notational convenience, we introduce $\mathrm{a} = [\mathrm{a}_1^T, \mathrm{a}_2^T, \mathrm{a}_3^T]^{\mathcal{T}}$. Then, the objective function is defined as

$$\begin{gathered}f(a)=\frac{1}{2}\left\|\dot{y}-\dot{\Psi}a_1\right\|_2^2+\frac{1}{2}\left\|\ddot{y}-\ddot{\Psi}a_2\right\|_2^2\\ +\alpha\|a_3\|_{TV}.\end{gathered} \tag{14}$$

As a consequence, (13) can be reduced to

$$\begin{gathered}f(a)=\frac{1}{2}\left\|\dot{y}-\dot{\Psi}a_1\right\|_2^2+\frac{1}{2}\left\|\ddot{y}-\ddot{\Psi}a_2\right\|_2^2\\ +\alpha\|a_3\|_{TV}.\end{gathered} \tag{15}$$

where $1 \in R^3$ is an all-ones vector and $\otimes$ represents the Kronecker product. The augmented Lagrangian associated with (15) is given by

$$\mathcal{L}_\mu(z,a,b)=f(a)+\frac{\mu}{2}\|a-1\otimes z+b\|_2^2, \tag{16}$$

where $b \in R^{3n^2K}$ is the dual variable, and $\mu \geq 0$ corresponds to the dual regularizer parameter. The Plug-and-Play ADMM [25] procedure is summarized in Algorithm 1. In line 2, the split $a^{(0)}$ and dual $b^{(0)}$ variables are initialized as all-zeros vectors. The Plug-and-Play ADMM iterations are computed in lines 4, 5, and 6. It is important to highlight that the sub-problem associated with the split variable $a_3$ in (14) can be solved using a proximal operator of the TV regularization, then, this variable is computed by using a deep proximal operator at each iteration [26]. Finally, the spectral image $z$ is returned in line 8.

Algorithm 1 Plug-and-Play ADMM spectral fusion

```
1:  Input: Acquired spectral data for SPC $\{\dot{y},\dot{\Psi}\}$ and DOC $\ddot{y},\ddot{\Psi}$, α, μ, and maximum number of iterations $T$.
2:  Initialize: $a^{(0)}=0$ and $b^{(0)}=0$.
3:  for $t = 1: T-1$ do
4:      $z^{(t+1)} := \underset{z}{\operatorname{argmin}}\, \mathcal{L}_\mu\left(z, a^{(t)}, b^{(t)}\right)$.
5:      $\boldsymbol{a}^{(t+1)} := \underset{\boldsymbol{a}}{\operatorname{argmin}}\, \mathcal{L}_\mu\left(\boldsymbol{z}^{(t+1)}, \boldsymbol{a}, \boldsymbol{b}^{(t)}\right)$.
6:      $\boldsymbol{b}^{(t+1)} := \boldsymbol{a}^{(t+1)} + \boldsymbol{b}^{(t)} - \mathbf{1}\otimes \boldsymbol{z}^{(t+1)}$
7:  end for
8:  Return: Fused spectral image z.
```

### 3.4. Classification network

The spectral classification network, denoted as $\mathcal{M}_\theta(\cdot)$, receives as input a spectral patch **z**. Specifically, this neural network corresponds to a 3D CNN that integrates spatial and spectral features into a joint spatial-spectral classification framework [14]. This CNN contains two 3D convolutions, two 3D MaxPooling operators, and a fully connected layer to conduct the labeling step. This neural network learns the parameters θ mapping the input (spectral patches) and the output (corresponding classes). The classification loss function can be expressed by the generalized cross-entropy (CE) loss over multiple classes as

$$\mathcal{L}_{\mathcal{CE}} := -\sum_{i=1}^{C}(u)_i\log((v)_i), \tag{17}$$

where $u \in N^C$ is the ground truth and $v \in N^C$ is the prediction.

### 3.5. End-to-end spectral classification approach

The proposed spectral classification scheme, illustrated in Figure 2, jointly learns the height map values $\boldsymbol{H}$ in the MPM and the parameters $\theta_1$ in the classification network, including ρ, α, and μ. From $\mathcal{S}$ patches extracted from a spectral image, the E2E optimization problem can be defined as

$$\begin{aligned}&\{\boldsymbol{H}^*,\boldsymbol{\theta}_1^*\}\\ &\in\backslash\, argmin_{H,\theta_1}LL(\boldsymbol{H},\boldsymbol{\theta}_1|\mathbf{u},\mathbf{v}),(\boldsymbol{H},\boldsymbol{\theta}_1|\mathbf{u},\mathbf{v})\\ &=\frac{\mathbf{1}}{LS}\sum_{s=\mathbf{1}}^{L,S}LL_{CE}(\mathbf{u},\mathbf{v})+\rho LR(\mathbf{H})\end{aligned} \tag{18}$$

where $\rho > 0$ is a regularization parameter, and $\mathcal{R}(\cdot)$ is a regularization function based on the Zernike polynomials (11). Algorithm 2 summarizes the proposed E2E optimization for spectral classification. This algorithm receives as input the spectral image $x^{(s)}$ and the corresponding class $u^{(s)}$ in line 1. Then, the height map $H$ is initialized as an all-zeros matrix in line 2. In lines 5 and 6, the spatial-spectral measurements $\ddot{y}^{(s)}$ and $\dot{\boldsymbol{y}}^{(s)}$ are computed by using (6) and (10), respectively. The spectral image fusion $z^{(s)}$ is calculated in line 7 by using the Plug-and-Play ADMM algorithm $\mathcal{I}_{\theta_2}(\cdot)$ described in the Algorithm 1. The labeling $v^{(s)}$ is obtained in line 8. Line 9 evaluates the loss function. In addition, the gradients of $H$ and θ are estimated in lines 10 and 11, respectively, which are used in the Adam update $\mathcal{A}_{dam}(\cdot)$ weighted by $\beta_1$ and $\beta_2$, respectively. Finally, the optimal height map and network parameters of the spectral classification network are returned in line 14.

```
Algorithm 2 Proposed E2E spectral classification
1:  Input: Training set $\{x^{(s)}\}_{s=1}^{\mathcal{S}} \to \{u^{(s)}\}_{s=1}^{\mathcal{S}}$ with $\mathcal{S}$ spectral patches, $\rho$, $\beta_1$ and $\beta_2$.
2:  Initialize: Set the height map $H$ as an all-zeros matrix.
3:  for epoch= 1:$\mathcal{E}$ do                                   ▷ $\mathcal{E}$ epochs
4:      for $s = 1:\mathcal{S}$ do                                  ▷ $S$ patches
5:          $\dot{\boldsymbol{y}}^{(s)} = \dot{\boldsymbol{\Psi}}\mathbf{x}^s + \dot{\boldsymbol{\omega}}$.            ▷ $SPC$
6:          $\ddot{\boldsymbol{y}}^{(s)} = \ddot{\boldsymbol{\Psi}}\mathbf{x}^s + \ddot{\boldsymbol{\omega}}$.            ▷ $DOC$
7:          $z^{(s)} \leftarrow \mathcal{I}_{\theta_2}(\{\dot{\boldsymbol{y}}^{(s)}, \dot{\Psi}\}, \{\ddot{\boldsymbol{\Psi}}\mathbf{x}^s + \ddot{\boldsymbol{\omega}}\})$.   ▷ ADMM 1
8:          $v^{(s)} \leftarrow \mathcal{M}_{\theta}(z^{(s)})$.                ▷Classification 3D CNN
9:          $\mathcal{L}_{\mathcal{H},\theta_1} = \frac{1}{\mathcal{S}}\sum_{s=1}^{\mathcal{S}} \mathcal{L}_{CE}(u^{(s)}, v^{(s)}) + \rho\mathcal{R}(H)$.
10:         $H \leftarrow \mathcal{A}_{dam}(H, \beta_1 \nabla_H \mathcal{L}_{\mathcal{H},\theta_1})$.
11:         $\theta_1 \leftarrow \mathcal{A}_{dam}(\theta_1, \beta_2 \nabla_\theta \mathcal{L}_{\mathcal{H},\theta_1})$.
12:     end for
13: end for
14: Return: Optimal height map $H$ and network parameters $\theta_1$.
```

## 4. Numerical results

This section presents the classification results from the proposed dual architecture under noiseless and noisy scenarios. Specifically, the noisy cases in spatial-spectral measurements are fixed by different signal-to-noise (SNR) values, where

$$\mathrm{SNR} = \begin{cases} 10\log_{10}\left(\|\ddot{y}\|_2^2/(n^2\sigma)\right) & \text{for DOC} \\ 10\log_{10}\left(\|\dot{y}\|_2^2/(LK\sigma)\right) & \text{for SPC} \end{cases}, \tag{19}$$

with $\sigma^2$ as the variance of the noise. The proposed E2E approach was trained using Tensorflow on a GPU Tesla T4 with 15 GB VRAM and 12.7 GB RAM. Several experiments were conducted to analyze the proposed methodology's performance through evaluating metrics, such as accuracy, precision, recall, and F1 score.

### 4.1. Spectral image dataset

Three public spectral datasets (The hyperspectral datasets are available on https://ehu.eus/ccwintco/index.php/Hyperspectral) were selected to train, validate and test the proposed E2E spectral classification methodology. From each spectral dataset, random patches were uniformly extracted. In particular, the training, validating, and testing samples correspond to 80%, 10%, and 10%, respectively, from the total number of spectral patches extracted at each dataset, as shown in Table 1.

Table 1. Summary of the number of patches selected for training, validation, and testing according to each spectral dataset

| **Dataset** | **Training ($S$)** | **Validation** | **Testing** | **Total** |
|---|---|---|---|---|
| Indian Pines | 8,199 | 1,025 | 1,025 | 10,249 |
| Pavia U | 34,220 | 4,278 | 4,278 | 42,776 |
| Pavia Center | 118,521 | 14,815 | 14,816 | 148,152 |

- **Indian Pines:** The Indian Pines dataset contains a spatial resolution of 145×145 pixels and a spectral resolution of 200 bands distributed along the spectral range 400 - 2500 [*nm*] for 16 different classes.
- **Pavia University:** The Pavia U dataset contains a spatial resolution of 610×340 pixels and a spectral resolution of 103 bands distributed along the spectral range 430-850 [*nm*] for 9 different classes.
- **Pavia Center:** The Pavia Center dataset contains a spatial resolution of 1096×715 pixels and a spectral resolution of 102 bands distributed along the spectral range 450-850 [*nm*] for 9 different classes.

Figure 3 illustrates the spectral datasets: Indian Pines, Pavia University, and Pavia Center. The false RGB representation, the classification map, the spectral signatures for each land cover class, and the classes are shown for each spectral dataset. In the Indian Pines dataset, the classes are stone-steel-towers, buildings-grass-trees-drives, woods, what, soybean-clean, soybean-min-till, soybean-no-till, oats, hay-windrowed, grass-pasture-mowed, grass-trees, grass-pasture, corn, corn-min-till, corn-no-till, and alfalfa. In the Pavia University dataset, the classes are shadows, self-blocking bricks, bitumen, bare soil, painted metal sheets, trees, gravel, meadows, and asphalt. Finally, in the Pavia Center dataset, the classes are bare soil, meadows, shadows, tiles, bitumen, self-blocking bricks, asphalt, trees, and water.

### 4.2. Classification network analysis

The proposed E2E classification approach was trained for each spectral dataset during *LE*= 100 epochs from a batch size of 64 spectral patches by using the Adam optimizer with a learning rate of $1 \times 10^{-3}$ $\beta_1 = 0.9$ and $\beta_2 = 0.999$. To analyze the classification performance, the noise level, the number of iterations, and the patch size were varied at each training of the proposed network.

Figure 4 presents the classification performance behavior by varying the noise level SNR $\in \{10,20,30\}$ [dB], the number of iterations $T \in \{1,2,3,4,5,6\}$, and the patch size $W \in \{8 \times 8, 12 \times 12, 16 \times 16\}$ pixels in terms of accuracy. For this experiment, the number of spectral bands was fixed at $K$=12 across all spectral datasets, with uniform selection.

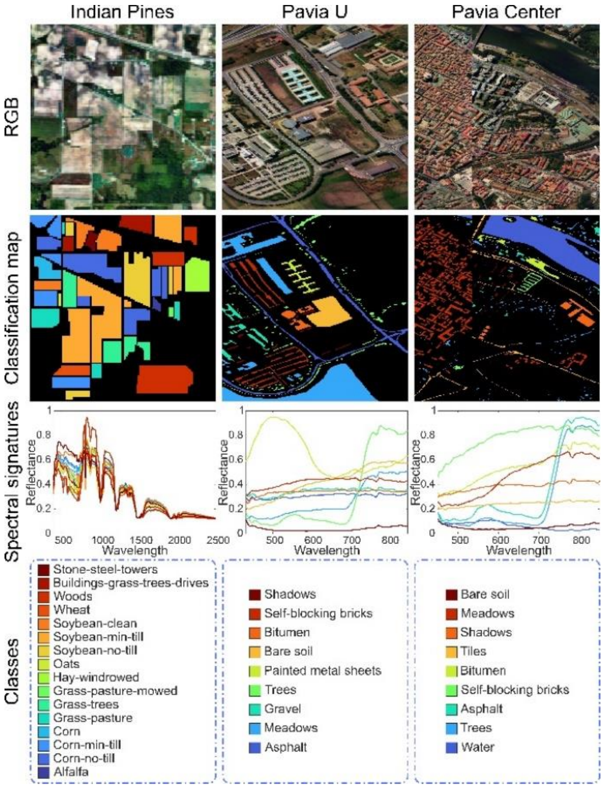


Figure 3. Indian Pines, Pavia U, and Pavia Center spectral datasets were used for the classification task. Columns: Spectral datasets. Row 1: False RGB representation. Row 2: Classification map. Row 3: Spectral signatures for each land cover class. Row 4: Land cover classes.

Notice that the E2E spectral classification approach for the Pavia Center dataset performs better than the Indian Pines and Pavia U datasets. Specifically, the Pavia Center dataset presents stable accuracy behavior across different noise levels when the patch size is $W = 8 \times 8$ pixels. Additionally, the Indian Pines and Pavia U datasets exhibit an improvement in classification accuracy with a patch size of $W = 8 \times 8$. It can be observed that conducting $T$=2 iterations is sufficient to obtain a suitable classification performance when the patch size is $W = 8 \times 8$.

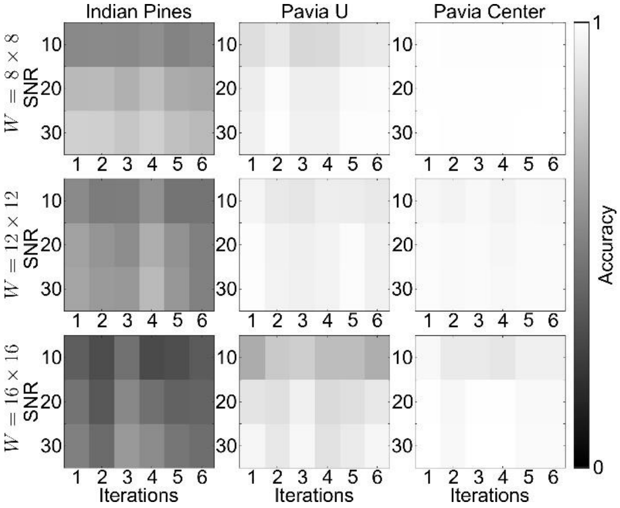


Figure 4. Classification network analysis in terms of accuracy by varying the noise level, the number of iterations, and patch size at each spectral dataset. Rows: Patch size. Columns: Spectral dataset. A lighter color indicates a better spectral classification.

Based on these results, we fixed the patch size at $W = 8 \times 8$ and the number of iterations at $T$=2 during the training stage of the proposed classification network for the experiments discussed below: the phase mask analysis over the dual optical architecture and the classification performance concerning the optical setup impact.

### 4.3. Multilevel phase mask analysis

Figure 5 illustrates the learned and non-larned MPMs with the resulting PSFs for spectral classification across different spectral datasets using the proposed dual optical setup. These MPMs were obtained by fixing $O = 15$ Zernike optical aberrations in the height map during the proposed E2E method training. It is important to note that the height map in the learned MPM encompasses different optical aberrations and adjusts depending on the spectral dataset. In contrast, the height map in the non-learned MPM was configured to consistently simulate the Fresnel lens across each spectral dataset, with the defocus coefficient set to one, and the remaining coefficients set to zero.

### 4.4. Optical setup impact on classification performance

The proposed E2E classification method was assessed across various optical setups, where the classification network parameters θ were optimized for each optical setup. Specifically, the proposed dual optical architecture was simulated using both learned and non-learned MPM. Furthermore, the spectral classification was evaluated separately using the SPC and the DOC, where the DOC involved both learned and non-learned MPM.

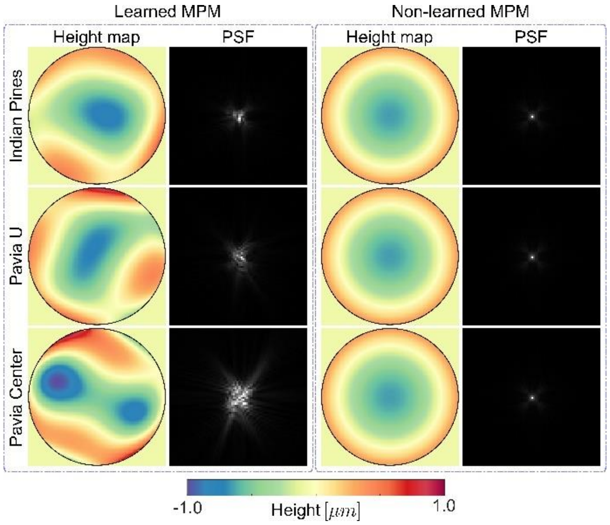


Figure 5. The resulting learned and non-learned MPM, along with their corresponding PSF, by training the proposed E2E classification method across different spectral datasets.

Table 2 summarizes the classification quantitative results. The best results are highlighted in the table with the green color and the second-best with yellow. The accuracy, precision, recall, and F1 score metrics were computed to evaluate the classification performance. Note that the proposed method is the best setup for all the spectral datasets, achieving up to 3% in the evaluated metrics above the second-best setup. Moreover, Figure 6 shows the classification maps for the proposed spectral classification methodology using the dual optical architecture by learning the MPM (Learned Dual), and non-learning the MPM (Non-learned Dual), the SPC, and the DOC by learning the MPM (Learned DOC) and non-learning the MPM (Non-learned DOC). The overall accuracy (OA) and F1 score were assessed across all spectral patches within each dataset. In particular, the proposed dual architecture with the learned MPM achieves better classification performance than the other configurations.

## 5. Experimental setup

This section presents the implemented dual optical architecture, as illustrated in Figure 7. This experimental setup mainly involves a halogen light source 3900e Illumination Technology with a power of 150 [$W$]; a CCD camera Stingray F-080B 1032 × 778 with a pixel size of $\Delta_{\text{CCD}} = 4.65$ [$\mu m$]; a piezoelectric DM Thorlabs DMP40-P01 with an aperture of 10 [$mm$]; a DMD Texas Instruments D4120 1024 × 768 with a pixel size of $\Delta_{\text{DMD}} = 13.6$ [$\mu m$]; and an Ocean Optics Flame S-VIS-NIR-ES spectrometer.

In the implemented optical testbed, the scene is illuminated by a light source, then, an objective lens guarantees the image formation in front of a lens with focal distance $f = 50$ [$mm$] (L1).

Table 2. Quantitative results for the classification task in terms of accuracy, precision, recall, and F1 score through the proposed E2E method based on different optical setups with SNR = 30 [dB]. The best result is highlighted in green, and the second best is highlighted in yellow. This experiment exhibits the classification performance across the testing dataset

| **Dataset** | **Metric** | **Optical Setup** | | | | |
|---|---|---|---|---|---|---|
| | | **Dual** | | **SPC** | **DOC** | |
| | | **Learned** | **Non-learned** | | **Learned** | **Non-learned** |
| Indian Pines | Accuracy ↑ | 0.8839 | 0.8566 | 0.7834 | 0.4985 | 0.4234 |
| | Precision ↑ | 0.8919 | 0.8550 | 0.7883 | 0.4509 | 0.3417 |
| | Recall ↑ | 0.8839 | 0.8566 | 0.7834 | 0.4985 | 0.4234 |
| | F1 score ↑ | 0.8833 | 0.8497 | 0.7640 | 0.4355 | 0.3481 |
| Pavia U | Accuracy ↑ | 0.9974 | 0.9822 | 0.9778 | 0.8817 | 0.8476 |
| | Precision ↑ | 0.9974 | 0.9824 | 0.9783 | 0.8793 | 0.8166 |
| | Recall ↑ | 0.9974 | 0.9818 | 0.9778 | 0.8717 | 0.8476 |
| | F1 score ↑ | 0.9974 | 0.9819 | 0.9776 | 0.8779 | 0.8303 |
| Pavia Center | Accuracy ↑ | 0.9931 | 0.9931 | 0.9469 | 0.8896 | 0.8431 |
| | Precision ↑ | 0.9934 | 0.9934 | 0.9193 | 0.8703 | 0.8250 |
| | Recall ↑ | 0.9931 | 0.9931 | 0.9469 | 0.8896 | 0.8431 |
| | F1 score ↑ | 0.9932 | 0.9932 | 0.9293 | 0.8684 | 0.8144 |

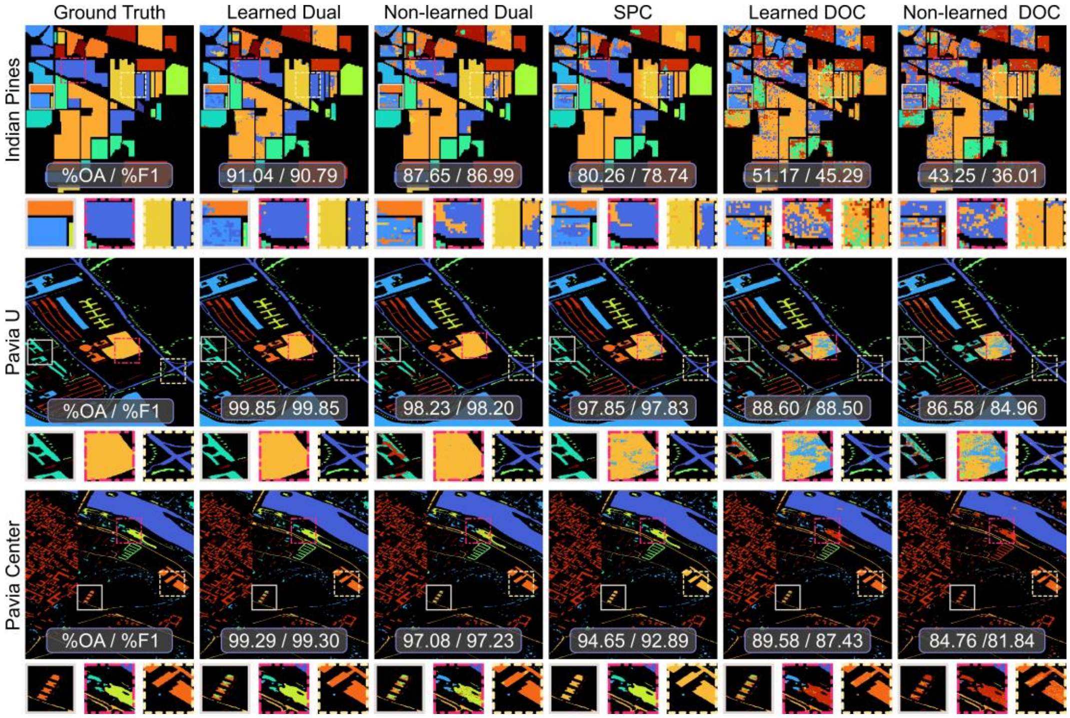


Figure 6. Classification maps for Pavia U, Pavia Center, and Indian Pines spectral datasets were obtained from the proposed E2E methodology using the dual optical setup for learned MPM and non-learned MPM, the single pixel architecture, and the diffractive system using learned MPM and non-learned MPM with SNR = 30 [dB]. This experiment exhibits the classification performance across the whole dataset.

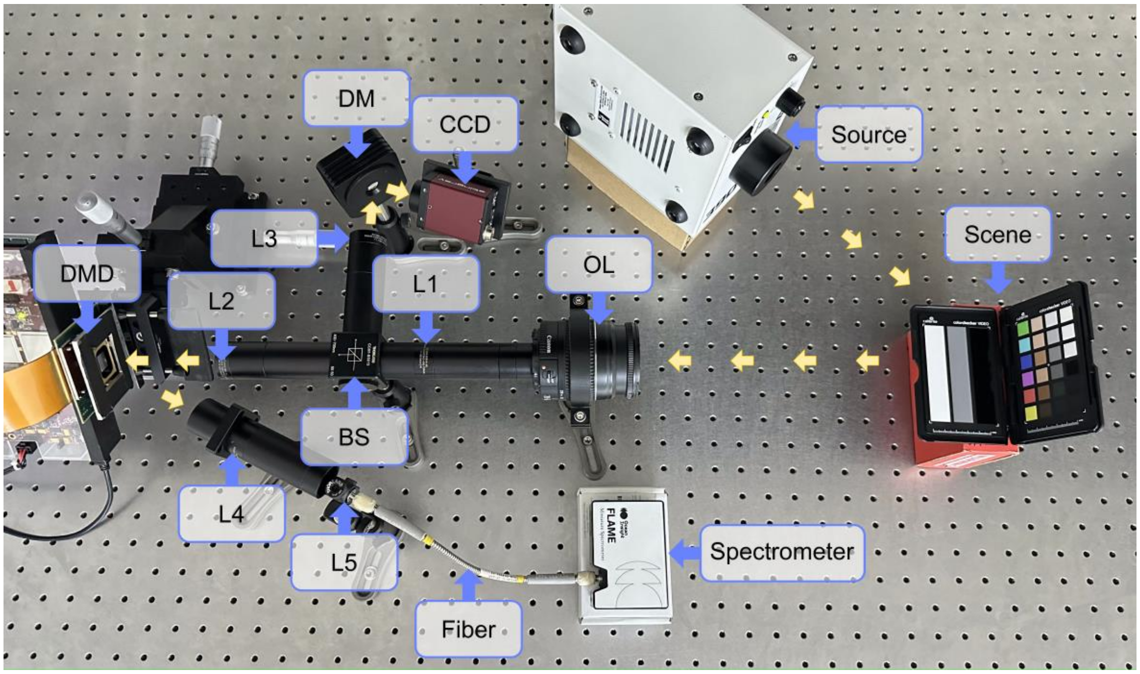


Figure 7. Dual optical setup for spectral classification. Optical elements: Halogen light source, illumination; OL, objective lens; L1, L2, L3, L4, and L5, lenses; DM, phase modulator; DMD, magnitude modulator; BS, beam splitter; CCD, registration camera; spectrometer, spectral sensor.

The incoming light is 50:50 divided by a non-polarizing BS Thorlabs CCM1-BS013, resulting in two optical paths, each of which is composed of a lens with a focal distance of $f = 100$ $[mm]$ (L2 and L3). Notice that the two optical paths are based on a $4f$ system. The first optical path incorporates a DM that emulates the phase mask, then, the modulated image is recorded by a CCD sensor. The second optical path employs a DMD that emulates the magnitude mask, then, a lens with a focal distance of $50$ $[mm]$ (L4) ensures the image formation over a condenser lens F220SMA-A (L5) connected through an optical fiber VIS-NIR with a core diameter of $1$ $[mm]$ until a spectrometer. The spectral classification algorithm was tuned for labeling the spectral scene from the coded spatial-spectral measurements acquired through the implemented optical testbed.

Figure 8 presents the classification results using the implemented dual optical testbed. The acquired dataset contains 9 different classes: pink synthetic rubber, yellow cold porcelain, blue felt, light green clay, light pink porcelain, light yellow synthetic rubber, pink cold porcelain, and light blue plastic. From this dataset, 10,997 spectral patches with $K$=12 bands and $W = 8 \times 8$ pixels were extracted, where 8,797, 1,100, and 1,100 patches were selected for training, testing, and validating, respectively. Note that the proposed dual architecture with the learned MPM achieves a more accurate classification map than the non-learned MPM.

Finally, Figure 9 presents the confusion matrix across the testing dataset for the proposed E2E spectral classification method based on the implemented optical architecture using the learned and non-learned MPM. The proposed dual system using the learned MPM obtains better performance in terms of the evaluated metrics compared to the non-learned MPM.

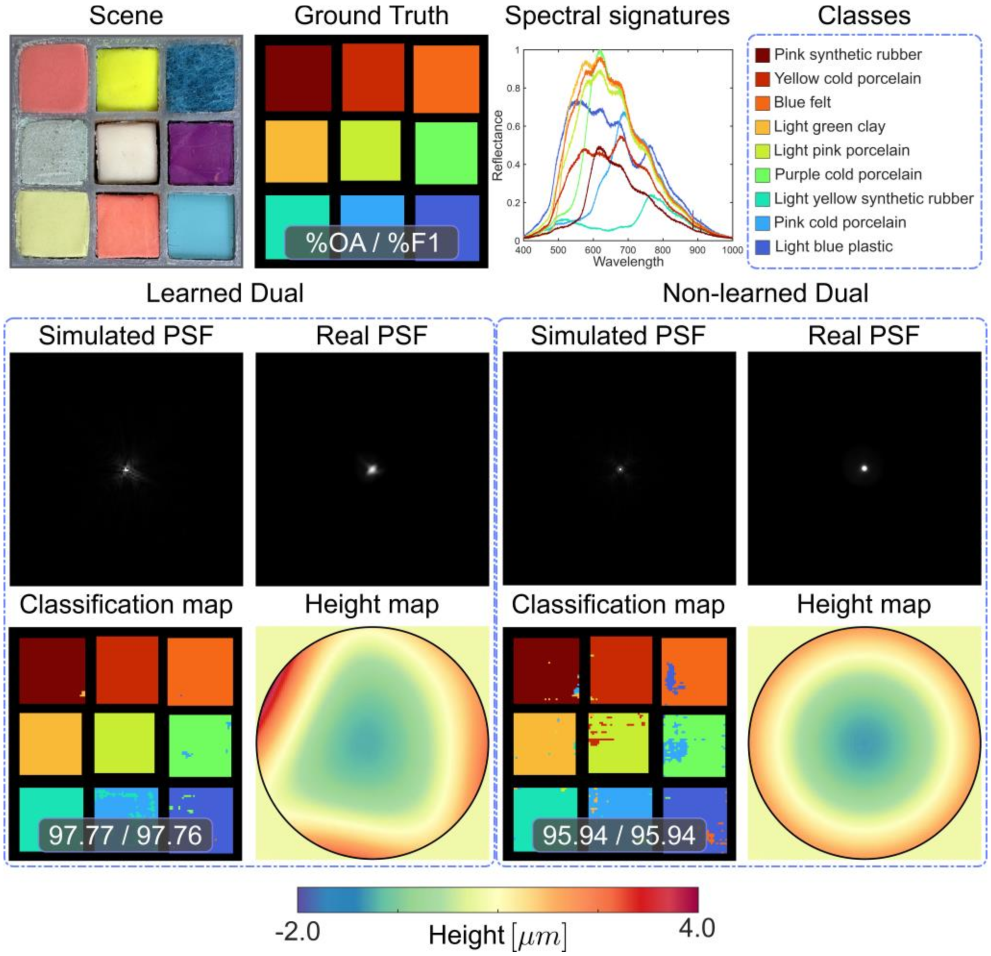
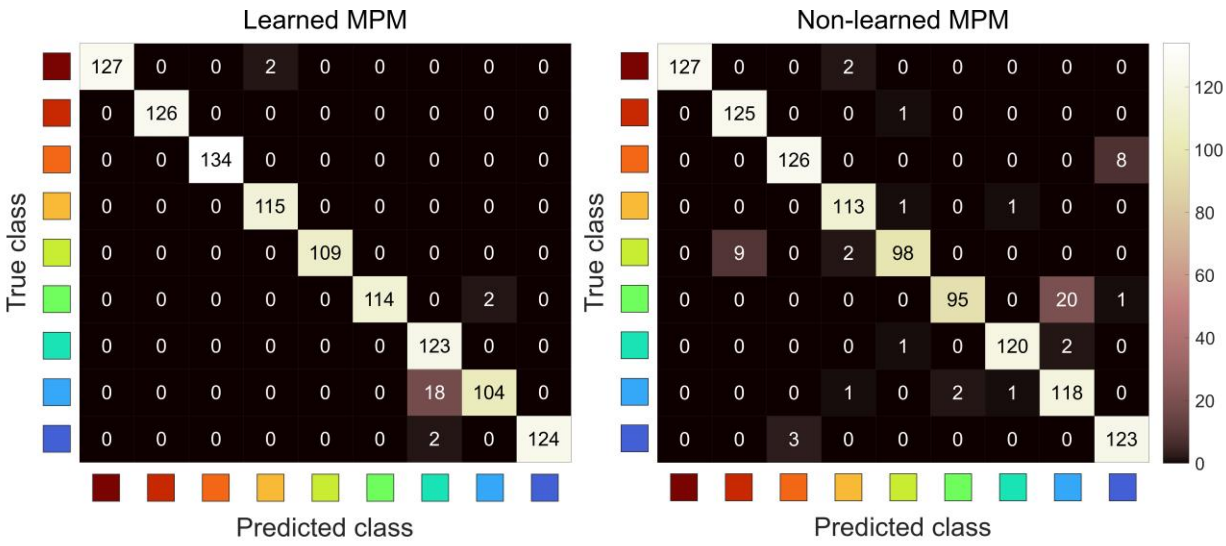


Figure 8. Classification results using the implemented dual optical setup. Row 1: RGB scene, classification map, spectral signatures for each material, and classes. Row 2: Simulated and real PSFs for the learned and non-learned MPM. Row 3: Classification results and height map for the learned and nonlearned MPMusing the whole dataset.



Learned MPM

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 127 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 |
| 0 | 126 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 134 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 115 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 109 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 114 | 0 | 2 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 123 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 18 | 104 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 124 |

Non-learned MPM

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 127 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 |
| 0 | 125 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 0 | 0 | 126 | 0 | 0 | 0 | 0 | 0 | 8 |
| 0 | 0 | 0 | 113 | 1 | 0 | 1 | 0 | 0 |
| 0 | 9 | 0 | 2 | 98 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 95 | 0 | 20 | 1 |
| 0 | 0 | 0 | 0 | 1 | 0 | 120 | 2 | 0 |
| 0 | 0 | 0 | 1 | 0 | 2 | 1 | 118 | 0 |
| 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 123 |

**Accuracy:** 97.82; **Precision:** 98.00
**Recall:** 97.82; **F1 - score:** 97.81

**Accuracy:** 95.00; **Precision:** 95.29
**Recall:** 95.00; **F1 - score:** 94.98

Figure 9. Confusion matrix for the dual optical setup based on the learned and non-learned MPMusing the testing dataset.

## 6. Conclusions

A dual optical architecture for spectral classification based on an E2E deep learning approach was proposed. The E2E network jointly learns the height map values in an MPM and the network parameters in a classification CNN. The proposed methodology contains three stages: first, an optical stage based on incoherent light describes the dual optical system, which incorporates an SPC and a DOC to acquire spatial-spectral information; second, a spectral image fusion stage calculates the spectral image acquired through the dual optical setup by using the plug-and-play ADMM algorithm; third, a classification network labels each material presented in the estimated spectral image from a 3D CNN. The proposed dual scheme improves the classification performance compared to conventional optical setups. Finally, an experimental setup for the proposed dual architecture was implemented to acquire real-world spatial-spectral measurements. The spectral classification was evaluated over real-world measurements.

## Funding

This work was supported by ICETEX and MINCIENCIAS through the CTO 2022-0716: “Sistema óptico computacional tipo pushbroom en el rango visible e infrarrojo cercano (VNIR), para la clasificación de frutos cítricos sobre bandas transportadoras mediante aprendizaje profundo, desarrollado en alianza con citricultores de Santander”, under Grant 8284.

## Autor Contributions

A. Jerez: Conceptualization, Formal Analysis, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. G. Blanco: Conceptualization, Formal Analysis, Investigation, Software, Validation, Writing – original draft. S. Urrea: Conceptualization, Formal Analysis, Investigation. H. García: Conceptualization, Formal Analysis, Investigation. H. Arguello: Conceptualization, Formal Analysis, Supervision.

## Conflicts of Interest

The authors declare no conflict of interest.

## Institutional Review Board Statement

Not applicable.

## Informed Consent Statement

Not applicable.

## References

[1] Q. Weng, “Thermal infrared remote sensing for urban climate and environmental studies: Methods, applications, and trends,” *ISPRS Journal of photogrammetry and remote sensing*, vol. 64, no. 4, pp. 335–344, 2009, doi: https://doi.org/10.1016/j.isprsjprs.2009.03.007

[2] M. J. Khan, H. S. Khan, A. Yousaf, K. Khurshid, and A. Abbas, “Modern trends in hyperspectral image analysis: A review,” *Ieee Access*, vol. 6, p. 14118–14129, 2018, doi: https://doi.org/10.1109/ACCESS.2018.2812999

[3] D. Guzzi, A. Barducci, P. Marcoionni, I. Pippi, “An atmospheric correction iterative method for high spectral resolution aerospace imaging spectrometers,” *in 2009 IEEE International Geoscience and Remote Sensing Symposium*, 2009, doi: https://doi.org/10.1109/IGARSS.2009.5418004

[4] M. Shimoni, R. Haelterman, C. Perneel, “Hypersectral imaging for military and security applications: Combining myriad processing and sensing techniques,” *IEEE Geoscience and Remote Sensing Magazine*, vol. 7, no. 2, pp. 101– 117, 2019, doi: https://doi.org/10.1109/MGRS.2019.2902525

[5] M. H. Tran, B. Fei, “Compact and ultracompact spectral imagers: technology and applications in biomedical imaging,” *Journal of biomedical optics*, vol. 28, no. 4, pp. 040 901–040 901, 2023, doi: https://doi.org/10.1117/1.JBO.28.4.040901

[6] L. Huang, R. Luo, X. Liu, and X. Hao, “Spectral imaging with deep learning,” *Light: Science & Applications*, vol. 11, no. 1, p. 61, 2022, doi: https://doi.org/10.1038/s41377-022-00743-6

[7] O. Denk, A. Musiienko, K. Žídek, “Differential single-pixel camera enabling low-cost microscopy in near-infrared spectral region,” *Optics express*, vol. 27, no. 4, pp. 4562–4571, 2019.

[8] H. Garcia, C. V. Correa, and H. Arguello, “Multi-resolution compressive spectral imaging reconstruction from single pixel measurements,” *IEEE Transactions on Image Processing*, vol. 27, no. 12, pp. 6174–6184, 2018, doi: https://doi.org/10.1109/TIP.2018.2867273

[9] A. Jerez, H. Garcia, and H. Arguello, “Single pixel spectral image fusion with side information from a grayscale sensor,” *in 2018 IEEE 1st Colombian Conference on Applications in Computational Intelligence (ColCACI)*, 2018, doi: https://doi.org/10.1109/ColCACI.2018.8484848

[10] L. Galvis, D. Lau, X. Ma, H. Arguello, G. R. Arce, “Coded aperture design in compressive spectral imaging based on side information,” *Applied optics*, vol. 56, no. 22, pp. 6332–6340, 2017, doi: https://doi.org/10.1364/AO.56.006332

[11] M. Imani and H. Ghassemian, “An overview on spectral and spatial information fusion for hyperspectral image classification: Current trends and challenges,” *Information fusion*, vol. 59, pp. 59–83, 2020, doi: http://dx.doi.org/10.1016/j.inffus.2020.01.007

[12] C. Hinojosa, K. Sanchez, H. Garcia, and H. Arguello, “C-3spcd: coded aperture similarity constrained design for spatio-spectral classification of single-pixel measurements,” *Applied Optics*, vol. 61, no. 8, pp. E21–E32, 2022, doi: https://doi.org/10.1364/AO.445326

[13] J. Bacca, E. Martinez, and H. Arguello, “Computational spectral imaging: a contemporary overview,” *JOSA A*, vol. 40, no. 4, pp. C115–C125, 2023, doi: https://doi.org/10.1364/JOSAA.482406

[14] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, “Deep feature extraction and classification of hyperspectral images based on convolutional neural networks,” *IEEE transactions on geoscience and remote sensing*, vol. 54, no. 10, pp. 6232–6251, 2016, doi: https://doi.org/10.1109/TGRS.2016.2584107

[15] X. Lin, Y. Rivenson, N. T. Yardimci, M. Veli, Y. Luo, M. Jarrahi, and A. Ozcan, “All-optical machine learning using diffractive deep neural networks,” *Science*, vol. 361, no. 6406, pp. 1004– 1008, 2018, doi: https://doi.org/10.1126/science.aat8084

[16] H. Arguello, S. Pinilla, Y. Peng, H. Ikoma, J. Bacca, and G. Wetzstein, “Shift-variant colorcoded diffractive spectral imaging system,” *Optica*, vol. 8, no. 11, pp. 1424–1434, 2021, doi: https://doi.org/10.1364/OPTICA.439142

[17] R. Jacome, J. Bacca, and H. Arguello, “Deepfusion: An end-to-end approach for compressive spectral image fusion,” in *2021 IEEE International Conference on Image Processing (ICIP)*, IEEE, 2021, doi: https://doi.org/10.1109/ICIP42928.2021.9506692

[18] J. Bacca, T. Gelvez-Barrera, 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, doi: https://doi.org/10.48550/arXiv.2105.03390

[19] L. Li, L. Wang, W. Song, L. Zhang, Z. Xiong, and H. Huang, "Quantization-aware deep optics for diffractive snapshot hyperspectral imaging," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, doi: https://doi.org/10.1109/CVPR52688.2022.01916

[20] V. Sitzmann, S. Diamond, Y. Peng, X. Dun, S. Boyd, W. Heidrich, F. Heide, and G. Wet- zstein, "End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging," *ACM Transactions on Graphics (TOG)*, vol. 37, no. 4, pp. 1–13, 2018, doi: https://doi.org/10.1145/3197517.3201333

[21] H. Garcia, C. V. Correa, and H. Arguello, "Optimized sensing matrix for single pixel multiresolution compressive spectral imaging," *IEEE Transactions on Image Processing*, vol. 29, pp. 4243–4253, 2020, doi: https://doi.org/10.1109/TIP.2020.2971150

[22] C. Hinojosa, J. C. Niebles, and H. Arguello, "Learning privacy-preserving optics for human pose estimation," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, doi: https://doi.org/10.1109/ICCV48922.2021.00257

[23] M. Born and E. Wolf, *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light*. Elsevier, 2013.

[24] E. Vargas, H. Arguello, J. Y. Tourneret, "Spectral image fusion from compressive measurements using spectral unmixing and a sparse representation of abundance maps," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 7, pp. 5043–5053, 2019, doi: https://doi.org/10.1109/TGRS.2019.2895822

[25] S. H. Chan, X. Wang, O. A. Elgendy, "Plugand-play admm for image restoration: Fixedpoint convergence and applications," *IEEE Transactions on Computational Imaging*, vol. 3, no. 1, pp. 84–98, 2016, doi: https://doi.org/10.48550/arXiv.1605.01710

[26] R. Jacome, J. Bacca, H. Arguello, "D 2uf: Deep coded aperture design and unrolling algorithmfor compressive spectral image fusion," *IEEE Journal of Selected Topics in Signal Processing*, 2022, doi: https://doi.org/10.1109/JSTSP.2022.3207663