

**METODOLOGÍA DE CLASIFICACIÓN EN IMÁGENES ESPECTRALES A PARTIR  
DEL USO DE REDES NEURONALES EN UN SISTEMA CON MÚLTIPLES  
VISTAS ANGULARES.**

**JUNIOR ALEJANDRO CÓRDOBA CARRERO**

**UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE INGENIERÍAS FÍSICOMECÁNICAS  
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA  
BUCARAMANGA**

**2022**

**METODOLOGÍA DE CLASIFICACIÓN EN IMAGÉNES ESPECTRALES A PARTIR  
DEL USO DE REDES NEURONALES EN UN SISTEMA CON MÚLTIPLES  
VISTAS ANGULARES**

**JUNIOR ALEJANDRO CÓRDOBA CARRERO**

**Una tesis presentada en cumplimiento de los requisitos para el grado de:  
Ingeniero de Sistemas e Informática**

**Director:**

**Miguel Angel Marquez Castellanos  
Ph.D(c) en Física**

**Codirector:**

**Henry Arguello Fuentes  
Doctor en Ingeniería Eléctrica y Computación**

**UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE INGENIERÍAS FÍSICOMECÁNICAS  
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA  
BUCARAMANGA**

**2022**

## **AGRADECIMIENTOS**

Agradezco a mis padres Elsa y Pedro por su apoyo, buenos valores y esfuerzo por enseñarme a ser una mejor persona cada día. A mis hermanos Diego y David por todo el apoyo durante el desarrollo de mi carrera. También agradezco a mis amigos y compañeros, en especial a Natalia Gómez por las experiencias personales y académicas, una persona admirable que con su amistad, personalidad e intelecto hicieron de esta una experiencia inolvidable.

Un reconocimiento y agradecimiento importante lo realizo a mi director de trabajo de grado Miguel Angel Marquez, por dedicar su tiempo, experiencia y conocimiento en la guía de mi proyecto.

## CONTENIDO

	<b>pág.</b>
<b>INTRODUCCIÓN</b> . . . . .	<b>11</b>
<b>1. OBJETIVOS</b> . . . . .	<b>16</b>
<b>2. MARCO TEÓRICO</b> . . . . .	<b>17</b>
2.1. MUESTREO ESPECTRAL . . . . .	17
2.1.1. IMÁGENES ESPECTRALES: . . . . .	17
2.1.2. Métodos convencionales para capturar imágenes espectrales: . . . . .	19
2.1.3. Teoría de muestreo compresivo espectral: . . . . .	21
2.2. Arquitecturas para la adquisición imágenes espectrales compresivas: . . . . .	23
2.2.1. Sistema de adquisición de imágenes espectrales basado en aperturas codificadas (CASSI). . . . .	24
2.2.2. Sistema de adquisición de imágenes espectrales con detectores basa- dos en filtros ópticos (SCCSI). . . . .	28
2.3. CLASIFICACIÓN DE IMÁGENES ESPECTRALES . . . . .	32
2.3.1. Redes neuronales artificiales, aprendizaje profundo. . . . .	34
2.3.2. Red neuronal basada en una estructura U-Net. . . . .	38
2.4. DISEÑO ÓPTICO . . . . .	39
2.4.1. Configuración básica: . . . . .	39
2.4.2. Enfoque de micro lentes al infinito óptico: . . . . .	40
2.4.3. Sistemas tradicionales de campo de luz espectral de compresión: . . . . .	41
2.4.4. Cámara plenoptica: . . . . .	42
<b>3. METODOLOGÍA</b> . . . . .	<b>45</b>

3.1. MODELOS DE ADQUISICIÓN Y CLASIFICACIÓN DE SI . . . . .	45
3.1.1. Dispersor dual de imágenes espectrales de instantáneas de apertura codificada (DD - CASSI) . . . . .	46
3.1.2. Modelo de detección continuo y discreto . . . . .	48
3.2. MODELO DE APRENDIZAJE PROFUNDO PARA LA SEGMENTACIÓN DE IMÁGENES . . . . .	52
3.2.1. Arquitectura Implementada . . . . .	52
<b>4. RESULTADOS . . . . .</b>	<b>55</b>
4.1. Selección de base de datos. . . . .	55
4.2. Marco tecnológico . . . . .	57
4.2.1. Software . . . . .	57
4.2.2. Hardware . . . . .	58
4.3. Desempeño de la arquitectura propuesta . . . . .	60
<b>5. CONCLUSIONES . . . . .</b>	<b>67</b>
<b>BIBLIOGRAFÍA . . . . .</b>	<b>69</b>

## LISTA DE FIGURAS

	<b>pág.</b>
Figura 1. Representación de una imagen espectral . . . . .	17
Figura 2. Métodos convencionales para capturar SI. . . . .	19
Figura 3. Arquitectura CASSI . . . . .	25
Figura 4. Propagación de luz a través de CASSI . . . . .	25
Figura 5. Arquitectura SCCSI . . . . .	29
Figura 6. Funcionamiento de una CNN. . . . .	36
Figura 7. Funcionamiento de U-Net . . . . .	39
Figura 8. Esquema conceptual de una cámara . . . . .	40
Figura 9. Arquitectura Lytro-CASSI. . . . .	41
Figura 10. Múltiples capturas durante el proceso de cristalización. . . . .	42
Figura 11. Esquema del uso de filtros de paso de banda múltiple. . . . .	43
Figura 12. Funcionamiento de una cámara plenoptica convencional . . . . .	44
Figura 13. Imagen de rayos del sistema. . . . .	46
Figura 14. Arquitectura DD-CASSI . . . . .	47
Figura 15. Ejemplo de las medidas en el CCD. . . . .	48
Figura 16. Esquema ilustrativo del método propuesto. . . . .	51
Figura 17. Esquema ilustrativo del código de apertura. . . . .	52
Figura 18. Clases de anotación introducidas . . . . .	56
Figura 19. Ejemplos de conjuntos de datos y anotaciones . . . . .	57
Figura 20. Visualización de bandas espectrales . . . . .	58
Figura 21. Captura sitio web del dataset . . . . .	59
Figura 22. Resultado de exactitud para la clasificación de imágenes . . . . .	61

Figura 23. Firmas espectrales . . . . . 62

Figura 24. Ejemplo 1 de los resultados de la clasificación . . . . . 63

Figura 25. Ejemplo 2 de los resultados de la clasificación . . . . . 64

Figura 26. Ejemplo 3 de los resultados de la clasificación . . . . . 65

## LISTA DE TABLAS

	<b>pág.</b>
Tabla 1. Distribución del conjunto de datos . . . . .	56
Tabla 2. Desempeño del sistema óptico-computacional . . . . .	60
Tabla 3. Resultados para 1, 4 y 8 medidas . . . . .	63

## RESUMEN

**TÍTULO:** METODOLOGÍA DE CLASIFICACIÓN EN IMÁGENES ESPECTRALES A PARTIR DEL USO DE REDES NEURONALES EN UN SISTEMA CON MÚLTIPLES VISTAS ANGULARES. \*

**AUTOR:** JUNIOR ALEJANDRO CÓRDOBA CARRERO \*\*

**PALABRAS CLAVE:** CLASIFICACIÓN, REDES NEURONALES CONVOLUCIONALES, PROCESAMIENTO DE IMAGEN, APRENDIZAJE PROFUNDO.

**DESCRIPCIÓN:** La clasificación se ha convertido en una herramienta indispensable en áreas comerciales como de la investigación al ser usada en el reconocimiento de objetivos y la interpretación de imágenes, más aún al usar información espectral, porque el comportamiento de la reflectancia a diferentes frecuencias electromagnéticas es característico de cada material, siendo posible identificarlos y analizarlos usando procesamiento de imágenes. Las SI son capturadas mediante sistemas ópticos, estos permiten obtener imágenes espectrales con una alta resolución espectral, pero su uso se ve limitado a escenarios donde los objetos poseen poco o nulo movimiento y su adquisición requiere altas capacidades de almacenamiento y de transmisión de datos. Para solucionar estos problemas, surge el muestreo compresivo espectral, que establece que es posible estimar la información espectral de una escena a partir de menos proyecciones que las requeridas por los sistemas ópticos tradicionales. A pesar de que estas técnicas permiten solucionar problemas de almacenamiento, transferencia de información, altos tiempos de adquisición, entre otros, tienen como desventaja la degradación de la información espacial. Para mejorar la calidad de reconstrucción y de clasificación se usan algunos métodos tradicionales para la adquisición de múltiples imágenes como son el barrido por fila espectral, pixel por pixel y por banda espectral, se han modificado los sistemas ópticos a su vez de desarrollar e implementar aplicaciones y algoritmos basados en técnicas de DL con pocas proyecciones. En este trabajo se propone el estudio de un sistema compresivo óptico computacional que permite mejorar las tareas de clasificación a partir de una única medida comprimida la cual se obtiene mediante un arreglo de lentes y de una red neuronal basada en aprendizaje profundo.

---

\* Trabajo de grado

\*\* Facultad de Ingenierías Físico-Mecánicas. Escuela de Ingeniería de Sistemas e Informática. Director: Miguel Angel Marquez, Ph.D(c). Codirector: Henry Arguello, Ph.D.

## ABSTRACT

**TITLE:** CLASSIFICATION METHODOLOGY IN SPECTRAL IMAGES FROM THE USE OF NEURAL NETWORKS IN A SYSTEM WITH MULTIPLE ANGULAR VIEWS. \*

**AUTHOR:** JUNIOR ALEJANDRO CÓRDOBA CARRERO. \*\*

**KEYWORDS:** CLASSIFICATION, CONVOLUTIONAL NEURAL NETWORKS, IMAGE PROCESSING, DEEP LEARNING.

**DESCRIPTION:** Classification has become an indispensable tool in commercial areas such as research when used in target recognition and image interpretation, even more so when using spectral information, because the behavior of reflectance at different electromagnetic frequencies is characteristic of each material, being possible to identify and analyze them using image processing. SI are captured by optical systems, these allow obtaining spectral images with high spectral resolution, but their use is limited to scenarios where objects have little or no movement and their acquisition requires high data storage and transmission capacities. To solve these problems, spectral compression sampling arises, which establishes that it is possible to estimate the spectral information of a scene from fewer projections than those required by traditional optical systems. Although these techniques allow solving problems of storage, information transfer, high acquisition times, among others, they have the disadvantage of spatial information degradation. To improve the quality of reconstruction and classification, some traditional methods are used for the acquisition of multiple images, such as scanning by spectral row, pixel by pixel and by spectral band, optical systems have been modified in turn to develop and implement applications and algorithms based on DL techniques with few projections. This paper proposes the study of a computational optical compressive system that allows improving classification tasks from a single compressed measure, which is obtained by means of an array of lenses and a neural network based on deep learning.

---

\* Degree work

\*\* Faculty of Physics-Mechanics Engineering. School of Systems Engineering and Informatics. Advisor: Miguel Angel Marquez, Ph.D(c). Co-advisor: Henry Arguello, PhD.

## INTRODUCCIÓN

La clasificación es un área de la investigación que se centra en la identificación, extracción y organización de elementos que componen una escena. Tradicionalmente existen diferentes métodos de clasificación tales como: Soporte de máquinas vectoriales (SVM), clasificadores lineales, redes neuronales convolucionales (CNN), entre otros, los cuales explotan la información espacial y espectral de una escena. Cabe resaltar que, entre los métodos de clasificación, los que han tenido un mayor impacto en la comunidad científica son los métodos de clasificación basados en el uso de la información espectral ya que cada elemento en la naturaleza absorbe, emite y refleja diferentes cantidades de energía dependiendo de la frecuencia electromagnética con la que se ilumina <sup>1</sup>. Los cuales, se han convertido en una herramienta indispensable en áreas tanto comerciales como de la investigación, por ejemplo, en la agricultura de precisión <sup>2</sup>, detección de enfermedades <sup>3</sup>, análisis de imágenes espaciales <sup>4</sup>. Específicamente, el comportamiento de la reflectancia a diferentes frecuencias electromagnéticas es característico de cada material, es decir, cada material posee una respuesta espectral única. Debido a esto, la clasificación a partir de la adquisición de imágenes espectrales (SI por sus siglas del inglés) permite obtener

---

<sup>1</sup> CAMACHO-VELASCO, Ariolfo, *et al.* “Aplicaciones y retos del sensado remoto hiperespectral en la geología colombiana”. En: *Facultad de Ingeniería* 24.40 (2015), págs. 17-29.

<sup>2</sup> ZHANG, Naiqian; WANG, Maohua y WANG, Ning. “Precision agriculture—a worldwide overview”. En: *Computers and electronics in agriculture* 36.2-3 (2002), págs. 113-132.

<sup>3</sup> GAVHALE, Kiran R y GAWANDE, Ujwalla. “An overview of the research on plant leaves disease detection using image processing techniques”. En: *IOSR Journal of Computer Engineering (IOSR-JCE)* 16.1 (2014), págs. 10-16.

<sup>4</sup> FAUVEL, Mathieu, *et al.* “Advances in spectral-spatial classification of hyperspectral images”. En: *Proceedings of the IEEE* 101.3 (2012), págs. 652-675.

un mejor desempeño en las tareas de detección, y/o clasificación <sup>5</sup>, en comparación a las técnicas tradicionales de adquisición de SI las cuales se basan en la extracción de características espaciales <sup>6</sup>.

Convencionalmente las SI son capturadas mediante el uso de sistemas ópticos de barrido, los cuales se pueden clasificar en tres categorías principales: barrido por fila espectral (comúnmente conocido en inglés como pushbroom), barrido pixel-por-pixel espectral (comúnmente conocido en inglés como whiskbroom) y barrido por banda espectral (comúnmente conocido en inglés como filter wheel). El principal objetivo de estos sistemas ópticos es adquirir la información espacial y espectral de una escena mediante el muestreo de cada una de sus dimensiones adyacentes, las cuales son fusionadas para obtener la imagen espectral. A pesar de que estos protocolos de muestreo permiten obtener imágenes espectrales con una alta resolución espectral, su uso se ve limitado a escenarios donde los objetos poseen poco o nulo movimiento, es decir, estos procedimientos únicamente son viables para escenarios estáticos <sup>7</sup>. Además, estos procesos de adquisición requieren altas capacidades de almacenamiento y de transmisión de datos <sup>8</sup>.

Para dar solución a estos problemas, surge la teoría de muestreo compresivo espec-

---

<sup>5</sup> GHAMISI, Pedram, *et al.* "Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art". En: *IEEE Geoscience and Remote Sensing Magazine* 5.4 (2017), págs. 37-78.

<sup>6</sup> CHUANG, Keh-Shih, *et al.* "Fuzzy c-means clustering with spatial information for image segmentation". En: *computerized medical imaging and graphics* 30.1 (2006), págs. 9-15.

<sup>7</sup> BALAS, Costas, *et al.* "A novel hyper-spectral imaging apparatus for the non-destructive analysis of objects of artistic and historic value". En: *Journal of Cultural Heritage* 4 (2003), págs. 330-337.

<sup>8</sup> VO-DINH, Tuan; CULLUM, Brian y KASILI, Paul. "Development of a multi-spectral imaging system for medical applications". En: *Journal of Physics D: Applied Physics* 36.14 (2003), pág. 1663.

tral (de sus siglas en inglés CSI), la cual establece que es posible estimar la información espectral de una escena a partir de menos proyecciones que las requeridas por los sistemas ópticos tradicionales. Específicamente, los sistemas CS realizan un proceso de codificación y multiplexación de la información espacio-espectral de la escena en el proceso de detección, esto se logra a partir del uso de elementos ópticos, tales como, los códigos de apertura <sup>9</sup>, espejos deformables <sup>10</sup>, prismas <sup>11</sup>, rejillas de difracción <sup>12</sup>, códigos de color <sup>13</sup>, entre otros. A pesar de que la teoría de CSI permite reducir el número de muestras requeridas, el proceso de codificación y multiplexación genera distorsiones en la dimensión espacial, es decir, estas técnicas limitan la calidad espacial. Por lo tanto, para mejorar la calidad de reconstrucción y por ende de clasificación, trabajos recientes en CSI han modificado los sistemas ópticos para admitir múltiples adquisiciones <sup>14</sup>, lo cual se logra al usar elementos ópticos de codificación dinámica <sup>15</sup>. A pesar que los sistemas CSI permiten reducir

- 
- <sup>9</sup> ARGUELLO, Henry y ARCE, Gonzalo R. "Code aperture optimization for spectrally agile compressive imaging". En: *JOSA A* 28.11 (2011), págs. 2400-2413.
- <sup>10</sup> POLANS, James , *et al.* "Compressed wavefront sensing". En: *Optics letters* 39.5 (2014), págs. 1189-1192.
- <sup>11</sup> GAO, Hao, *et al.* "Multi-energy CT based on a prior rank, intensity and sparsity model (PRISM)". En: *Inverse problems* 27.11 (2011), pág. 115012.
- <sup>12</sup> QU, Lele, *et al.* "Diffraction tomographic ground-penetrating radar multibistatic imaging algorithm with compressive frequency measurements". En: *IEEE Geoscience and Remote Sensing Letters* 12.10 (2015), págs. 2011-2015.
- <sup>13</sup> YUAN, Xin, *et al.* "Low-cost compressive sensing for color video and depth". En: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2014, págs. 3318-3325.
- <sup>14</sup> CORREA, Claudia V; ARGUELLO, Henry y ARCE, Gonzalo R. "Spatiotemporal blue noise coded aperture design for multi-shot compressive spectral imaging". En: *JOSA A* 33.12 (2016), págs. 2312-2322.
- <sup>15</sup> LEÓN-LÓPEZ, Kareth M; CARRENO, Laura V Galvis y FUENTES, Henry Arguello. "Temporal colored coded aperture design in compressive spectral video sensing". En: *IEEE Transactions on Image Processing* 28.1 (2018), págs. 253-264.

los tiempos de adquisición, en comparación con los sistemas de muestreo SI tipo barrido, esta técnica requiere una reducción en la resolución temporal en función del número de disparos. Por otro lado, el uso de estos elementos ópticos de modulación variable incrementa los costos de implementación, el tamaño del sistema óptico, el consumo y/o rendimiento energético y la complejidad de calibración.

Para reducir el número de proyecciones requeridas en CSI y mantener la calidad de la clasificación, se han desarrollado aplicaciones y algoritmos basados en técnicas de aprendizaje profundo (DL por sus siglas del inglés) <sup>16, 17</sup>. Específicamente, DL es un subcampo del aprendizaje automático <sup>18</sup>, con algoritmos que utilizan “redes neuronales artificiales” inspirados en la estructura y funcionamiento del cerebro humano. Consta de varias capas dentro de sus redes, las cuales junto con datos de referencia y por medio de jerarquía transmiten y procesan información a través de entrenamiento que mejora a su vez la precisión del resultado <sup>19,20</sup>. A partir de un conjunto de entrenamiento, tomando gran cantidad de imágenes como referencia y entrenando los pesos de la red, DL permite realizar estimaciones o clasificación de imágenes espectrales, permitiendo de esta manera hacer una tarea en específico

---

<sup>16</sup> LOUIZOS, Christos; ULLRICH, Karen y WELLING, Max. “Bayesian compression for deep learning”. En: *Advances in neural information processing systems*. 2017, págs. 3288-3298.

<sup>17</sup> CHAN, Tsung-Han, *et al.* “PCANet: A simple deep learning baseline for image classification?” En: *IEEE transactions on image processing* 24.12 (2015), págs. 5017-5032.

<sup>18</sup> LI, Shutao, *et al.* “Deep learning for hyperspectral image classification: An overview”. En: *IEEE Transactions on Geoscience and Remote Sensing* 57.9 (2019), págs. 6690-6709.

<sup>19</sup> SCHMIDHUBER, Jürgen. “Deep learning in neural networks: An overview”. En: *Neural networks* 61 (2015), págs. 85-117.

<sup>20</sup> KUSSUL, Nataliia, *et al.* “Deep learning classification of land cover and crop types using remote sensing data”. En: *IEEE Geoscience and Remote Sensing Letters* 14.5 (2017), págs. 778-782.

<sup>21</sup>. Debido a esto, se han desarrollado una gran variedad de algoritmos de clasificación basados en DL que permiten obtener altos rendimiento de clasificación y/o detección en sistemas compresivos con una única captura. Sin embargo, debido al auge del DL los trabajos recientes se han enfocado únicamente en soluciones computacionales, dejando a un lado el estudio y desarrollo de sistemas ópticos que permitan mejorar el rendimiento en las tareas de clasificación. Debido a esto, el desarrollo de un sistema óptico junto con una metodología DL que permita obtener resultados de clasificación a partir del uso en un sistema compresivo es altamente deseado.

Por lo tanto, en este trabajo se propuso un sistema compresivo óptico computacional que permite mejorar las tareas de clasificación a partir del uso de un arreglo de lentes y de una red neuronal basada en aprendizaje profundo. Específicamente, aprovechamos la propiedad de los lentes permitiendo ver diferentes secciones de la imagen con base en la codificación. Por lo tanto, al generarse esta redundancia de información generada por el arreglo de lentes se realizan múltiples codificaciones a partir de una única captura.

---

<sup>21</sup> MAKANTASIS, Konstantinos, *et al.* "Deep supervised learning for hyperspectral data classification through convolutional neural networks". En: *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE. 2015, págs. 4959-4962.

## 1. OBJETIVOS

### Objetivo general

- Diseñar, modelar y simular una metodología que permita mejorar la clasificación en imágenes espectrales a partir de la extracción de características espaciales y espectrales en un sistema con múltiples vistas angulares.

### Objetivos específicos

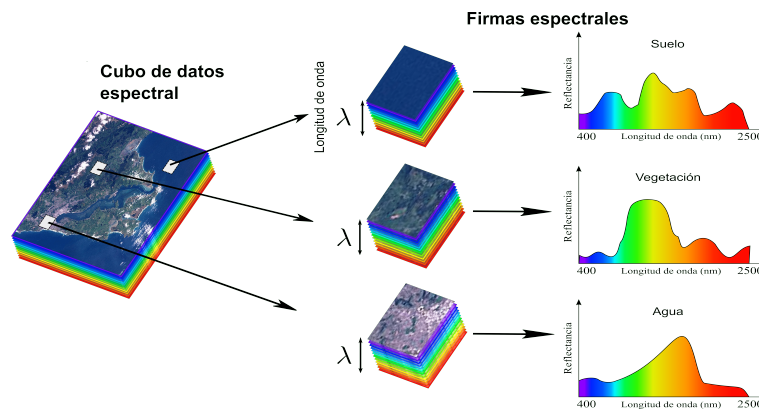
1. Estudiar diferentes modelos tradicionales de adquisición y clasificación de imágenes espectrales, y sus requerimientos técnicos.
2. Seleccionar y modelar una metodología de muestreo compresivo espectral en el cual sea viable la incorporación de un arreglo de lentes para la adquisición de múltiples versiones de la escena con diferente codificación.
3. Diseñar y modelar un sistema espectral compresivo para adquirir múltiples versiones comprimidas de la escena a partir de una única captura.
4. Implementar una red neuronal para la clasificación de imágenes espectrales a partir del uso de las medidas comprimidas adquiridas por el sistema propuesto.
5. Evaluar el desempeño del sistema óptico-computacional y de la red neuronal, realizando la clasificación de varias escenas a partir del uso de base de datos disponibles y/o adquiridas en laboratorio.

## 2. MARCO TEÓRICO

### 2.1. MUESTREO ESPECTRAL

**2.1.1. IMÁGENES ESPECTRALES:** Las imágenes espectrales son conjuntos de imágenes bidimensionales que contienen información de una escena y utilizan múltiples bandas lo largo de todo el espectro electromagnético incluidos sus 3 rangos: región espectral del infrarrojo cercano (NIR), ultravioleta-visible (VIS) y la radiación infrarroja (IR). Específicamente, la versión discreta de una imagen espectral puede representarse como un cubo de datos figura-1 en el que dos dimensiones representan el espacio y una tercera dimensión las bandas espectrales <sup>22</sup>.

**Figura 1.** Representación de una imagen espectral. Tomado de<sup>23</sup>.



Las imágenes espectrales proveen más información que la de la visión humana, debido a que contienen bandas espectrales fuera del espectro visible. Se tiene que, al agregar algunas bandas espectrales adicionales a los canales de color rojo (R), verde (G) y azul (B) resulta útil en múltiples aplicaciones de diferentes campos de

<sup>22</sup> SHAW, Gary A y BURKE, Hsiaohua K. "Spectral imaging for remote sensing". En: *Lincoln laboratory journal* 14.1 (2003), págs. 3-28.

investigación, tales como medicina <sup>24</sup>. Agricultura <sup>25</sup>, medio ambiente o minería <sup>26</sup>. Las imágenes espectrales (SI) son espectralmente suaves y espacialmente lisas por partes: los valores en ubicaciones vecinas y longitudes de onda están altamente correlacionados. Esto se puede observar mediante matrices de covarianza no diagonales y funciones de autocorrelación extremadamente amplias <sup>27</sup>. Una interpretación equivalente de las SI viene dada por la adquisición de una pila de imágenes que representan la intensidad en la banda respectiva (intervalo de longitud de onda). Debido a esta interpretación, los SI también se denominan cubos de datos hiperespectrales.

Los datos de detección remota hiperespectral con ancho de banda de nivel nanométrico (nm) tienen decenas o incluso varios cientos de canales y contienen abundante información espectral. Los diferentes canales tienen sus propias propiedades y muestran las características espectrales de varios objetos en la imagen. Esto se puede ver en la figura-1, muestra que una SI con canales mostrando propiedades de diferentes materiales (en este caso vegetación, agua y suelo) a diferentes longitudes de onda  $\lambda$  y mediante la firma espectral única, se puede sustraer de este tipo de imágenes. La selección racional de características de las variedades de canales

---

<sup>24</sup> EWERLÖF, Maria; LARSSON, Marcus y SALERUD, E Göran. "Spatial and temporal skin blood volume and saturation estimation using a multispectral snapshot imaging camera". En: *Imaging, Manipulation, and Analysis of Biomolecules, Cells, and Tissues XV*. Vol. 10068. International Society for Optics y Photonics. 2017, pág. 1006814.

<sup>25</sup> VIBHUTE, Anup y BODHE, Shrikant K. "Applications of image processing in agriculture: a survey". En: *International Journal of Computer Applications* 52.2 (2012).

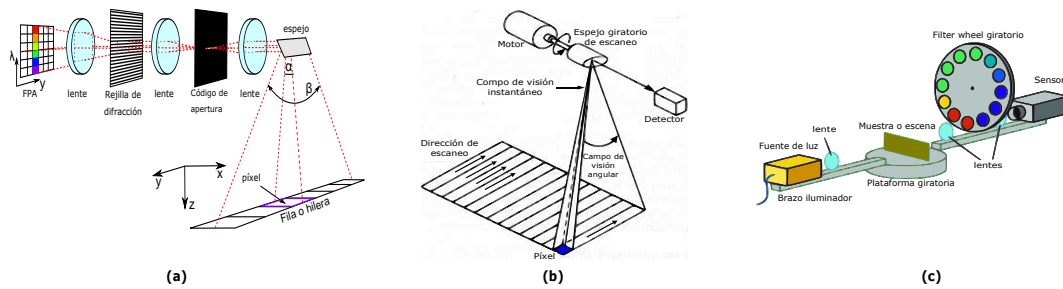
<sup>26</sup> CONSTANTIN, Dragos, *et al.* "Detection of crop properties by means of hyperspectral remote sensing from a micro UAV". En: *Bornimer Agrartechnische Berichte* 88.CONF (2015), págs. 129-137.

<sup>27</sup> BIOUCAS-DIAS, José M, *et al.* "Hyperspectral remote sensing data analysis and future challenges". En: *IEEE Geoscience and remote sensing magazine* 1.2 (2013), págs. 6-36.

es muy importante para el análisis efectivo y la extracción de información de datos hiperspectrales.

**2.1.2. Métodos convencionales para capturar imágenes espectrales:** En esta subsección se estudiarán los tres métodos más representativos del estado del arte para la captura de imágenes espectrales ejecutados mediante vehículos aéreos, los cuales son: barrido por fila espectral, barrido píxel-por-píxel espectral y por banda espectral. Su funcionamiento se describe a continuación.

**Figura 2.** Métodos convencionales para capturar SI: (a) Pushbroom, (b) Whiskbroom, (c) Filter wheel. Adaptada de <sup>28,29,30</sup>.



■ **Barrido por fila espectral (Pushbroom)**

El sensor pushbroom es un escáner sin ningún espejo de escaneo mecánico, pero con una matriz lineal de elementos semiconductores sólidos que le permite grabar una línea de una imagen simultáneamente, tiene una lente óptica a través de la cual se detecta una imagen lineal simultáneamente perpendicular a la posición del detector <sup>31</sup>.

En el sensor pushbroom, el espectro del píxel sombreado gris se asigna a la columna de color en la matriz del plano focal figura-2(a). El movimiento a lo largo de la dirección de vuelo  $x$  permite la formación de una imagen mediante

<sup>31</sup> GUPTA, Rajiv y HARTLEY, Richard I. "Linear pushbroom cameras". En: *IEEE Transactions on pattern analysis and machine intelligence* 19.9 (1997), págs. 963-975.

la combinación de huellas consecutivas. El campo de visión  $\beta$  determina el ancho de la franja, mientras que el campo de visión instantáneo está relacionado con el tamaño de los píxeles individuales <sup>32</sup>.

- **Barrido píxel-por-píxel espectral (Whiskbroom)**

Los escáneres whiskbroom, también conocidos como reflectores o escáneres de riel, usan un espejo para reflejar la luz en un solo detector. El espejo se mueve hacia adelante y hacia atrás para recopilar mediciones de un píxel en la imagen a la vez figura-2(b). Las partes móviles hacen que este tipo de sensor sea costoso y más propenso a desgastarse. Este tipo de arquitectura presenta un sistema mecánico complejo y a su vez un sistema óptico simple, utilizando filtros y sensores, permitiendo menor tiempo de permanencia, aunque con una mayor distorsión de píxeles <sup>33</sup>.

- **Por banda espectral (Filter wheel)**

Esta arquitectura de escaneo multiespectral hace parte de los llamados filtros ajustables, que, al capturar una imagen de una banda espectral a la vez, se produce una imagen multiespectral completa después de una secuencia de exposiciones <sup>34</sup>. Filter wheel figura-2(c), instala una serie de filtros ópticos deseados. Cuando se combina con una cámara monocromática y se inserta típicamente en el camino óptico, dicha rueda de filtro puede funcionar de forma sincronizada con la cámara, de modo que una exposición corresponde a

---

<sup>32</sup> KÖHLER, Claas Henning. "Airborne imaging spectrometer hypex". En: *Journal of large-scale research facilities JLSRF* 2.A93 (2016), págs. 1-6.

<sup>33</sup> BHATTA, Suman. "Band to Band Calibration and Relative Gain Analysis of Satellite Sensors Using Deep Convective Clouds". En: (2015).

<sup>34</sup> GAT, Nahum. "Imaging spectroscopy using tunable filters: a review". En: *Wavelet Applications VII*. Vol. 4056. International Society for Optics y Photonics. 2000, págs. 50-64.

un cierto tipo de filtro, que produce una región o banda espectral dada <sup>35</sup>.

**2.1.3. Teoría de muestreo compresivo espectral:** La teoría de muestreo compresivo es una forma de muestrear y comprimir cierto tipo de señales. En el muestreo compresivo, el modelo tradicional limitado en banda planteado por el teorema de Shannon-Nyquist-Whittaker-Kotelnikov es remplazado por un modelo disperso o escaso, el cual, asume que una señal puede ser eficientemente representada usando solamente un número pequeño de proyecciones lineales no adaptativas presentes en el dominio de alguna transformada <sup>36</sup>. Por lo tanto, la teoría de muestreo compresivo afirma que se puede reconstruir la aproximación de una imagen espectral a partir de menos mediciones o muestras aleatorias o semi aleatorias que los métodos tradicionales, suponiendo que la señal es de baja densidad acorde a la base de representación dada. Para hacer esto posible, muestreo compresivo se fundamenta en dos principios: la escasez (sparsity en inglés), que corresponde a las señales de interés, y la incoherencia, que pertenece a la modalidad de detección <sup>37</sup>.

En términos matemáticos, cuando el proceso de adquisición de información es lineal, el problema se reduce a resolver un sistema lineal de ecuaciones tal que:

$$\mathbf{Ax} = \mathbf{y}, \quad (1)$$

donde  $\mathbf{y} \in \mathbb{R}^M$  son las medidas adquiridas,  $\mathbf{x} \in \mathbb{R}^N$  una imagen de entrada repre-

---

<sup>35</sup> LAPRAY, Pierre-Jean, *et al.* "Multispectral filter arrays: Recent advances and practical implementation". En: *Sensors* 14.11 (2014), págs. 21626-21659.

<sup>36</sup> MORENO, Rodolfo G, *et al.* "Desarrollo de algoritmos para muestreo compresivo aplicado a señales de audio". En: *Información tecnológica* 25.5 (2014), págs. 137-146.

<sup>37</sup> CANDÈS, Emmanuel J, *et al.* "Compressive sampling". En: *Proceedings of the international congress of mathematicians*. Vol. 3. Madrid, Spain. 2006, págs. 1433-1452.

sentada como un vector y  $\mathbf{A} \in \mathbb{R}^{M \times N}$  es la matriz de muestreo con  $M < N$ <sup>38</sup>.

- **Escasez:** El concepto de escasez expresa la idea de que la "tasa de información" de una señal de tiempo continua puede ser mucho más pequeña que la sugerida por su ancho de banda, o que una señal de tiempo discreto depende de un número de grados de libertad que es comparativamente mucho menor que su longitud (finita)<sup>37</sup>. En términos matemáticos se tiene que la señal  $f$ , cuando es expresada en una base apropiada  $\Psi \in \mathbb{R}^{n \times n}$  (también conocida como la base de representación), da como resultado señales más ligeras. Teniendo que la señal  $f$  es expandida sobre la base  $\Psi$  de la siguiente manera,

$$f = \Psi\theta, \quad (2)$$

Donde  $\theta \in \mathbb{R}^n$  denota la representación escasa de  $f$  en la base de representación  $\Psi$ . Por lo general representación escasa en pocas palabras es la característica de la señal  $\theta$  en la cual la mayoría de sus coeficientes son despreciables y solo unos pocos coeficientes relativamente grandes obtienen la mayoría de la información. Por lo tanto, se pueden eliminar coeficientes despreciables sin mucha pérdida perceptual. La escasez se puede tener en cuenta como una medida para determinar con que eficiencia se pueden adquirir señales no adaptativas, porque las señales con alta escasez pueden ser adquiridas por un número menor de coeficientes a los que requieren las señales poco escasas<sup>39</sup>.

- **Incoherencia:** La incoherencia expresa la idea de que los objetos que tienen una representación dispersa en  $\Psi$  deben extenderse en el dominio en el que

---

<sup>38</sup> FOUCART, Simon y RAUHUT, Holger. "An invitation to compressive sensing". En: *A mathematical introduction to compressive sensing*. Springer, 2013, págs. 1-39.

<sup>39</sup> LU, Pei, *et al.* "Digital image information encryption based on compressive sensing and double random-phase encoding technique". En: *Optik* 124.16 (2013), págs. 2514-2518.

se adquieren, al igual que un pico en el dominio del tiempo se extiende en el dominio de la frecuencia. Dicho de otra manera, la incoherencia dice que, a diferencia de la señal de interés, las formas de onda de muestreo tienen una representación extremadamente densa en  $\Psi$  <sup>37</sup>. Se tienen las ortobases  $\Phi \in \mathbb{R}^{N \times N}$  que se usa para muestrear un objeto  $f$  y  $\Psi \in \mathbb{R}^{N \times N}$  para representar ese objeto. La coherencia entre la base de muestreo  $\Phi$  y la base de representación  $\Psi$  es representada por:

$$\mu(\Phi, \Psi) = \sqrt{n} \cdot \text{máx}(| \Phi \Psi |), \quad (3)$$

donde  $\text{máx}(\cdot)$  retorna al valor máximo de su argumento y  $\mu(\Phi, \Psi) \in [1, \sqrt{N}]$ . La coherencia mide la mayor correlación entre pares de  $\Phi$  y  $\Psi$ . Si estas dos bases contienen elementos correlacionados su valor de coherencia será grande, de lo contrario será pequeño y cumpliría las condiciones de CS, debido a que entre mayor sea el valor de incoherencia se necesitaran menos medidas para capturar una señal  $f$  <sup>38</sup>.

## 2.2. Arquitecturas para la adquisición imágenes espectrales compresivas:

Basado en la teoría de muestreo compresivo para la captura de imágenes espectrales se deriva a imágenes espectrales compresivas (CSI de su sigla en inglés Compressive Spectral Imaging). Generalmente, las técnicas de CSI capturan una escena utilizando un dispositivo de única captura sin tener la necesidad de escanearla. Contrario a los métodos tradicionales, CSI no realiza el muestreo directamente sobre los vóxeles, a su vez captura proyecciones codificadas aleatoriamente en la escena. Esa escena se reconstruye a través del uso de un algoritmo de optimización convexa, determinando una aproximación a la imagen espectral original.

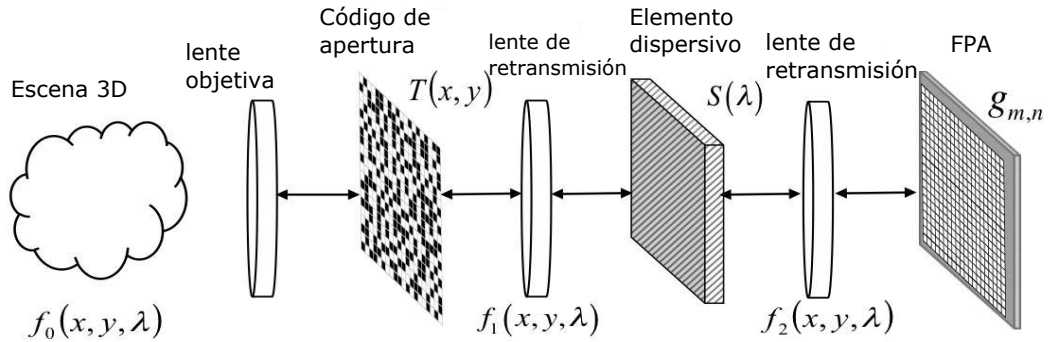
**2.2.1. Sistema de adquisición de imágenes espectrales basado en aperturas codificadas (CASSI).** El sistema de imágenes espectrales de instantáneas de apertura codificada (CASSI de su sigla en inglés Coded Aperture Snapshot Spectral Imager) es una arquitectura animada que detecta eficazmente la información espectral tridimensional (3D) de una escena, utilizando una única medición de proyección aleatoria codificada en 2D, detecta la información espectral de una escena tridimensional mediante el uso de proyecciones de matriz de plano focal codificado (FPA), un detector bidimensional y un elemento dispersivo. Las proyecciones en CASSI están localizadas de tal manera que cada medición contiene información espectral solo de una región espacial específica del cubo de datos. Dado un conjunto de mediciones de compresión de matriz de plano focal (FPA), la teoría de detección de compresión (CS) se explota para recuperar el cubo de datos espectrales 3D subyacente al encontrar la aproximación más escasa con la distancia euclidiana mínima a las mediciones de proyección aleatoria 2D. Normalmente, la salida discretizada en el detector CASSI se modela como la suma de los segmentos espectrales de vóxel subyacentes que han sido modulados previamente por una apertura codificada y posteriormente dispersados espacialmente por un prisma. La resolución espacial en CASSI depende en gran medida de la resolución que exhibe el detector FPA; por lo tanto, las imágenes de alta resolución requieren detectores de alta resolución que exigen altos costos <sup>40</sup>.

**Modelo continuo CASSI** Sea  $f_0(x, y, \lambda)$  la densidad de la fuente de energía espacio-espectral, donde  $x$  y  $y$  indexan las coordenadas espaciales y  $\lambda$  indexan la longitud de onda. La densidad de la fuente de imagen espacio-espectral  $f_0(x, y, \lambda)$

---

<sup>40</sup> RUEDA CHACÓN, Hoover Fabian y ARGUELLO FUENTES, Henry. "Spatial super-resolution in coded aperture-based optical compressive hyperspectral imaging systems". En: *Revista Facultad de Ingeniería Universidad de Antioquia* 67 (2013), págs. 7-18.

**Figura 3.** Arquitectura óptica CASSI. Tomado de<sup>41</sup>.

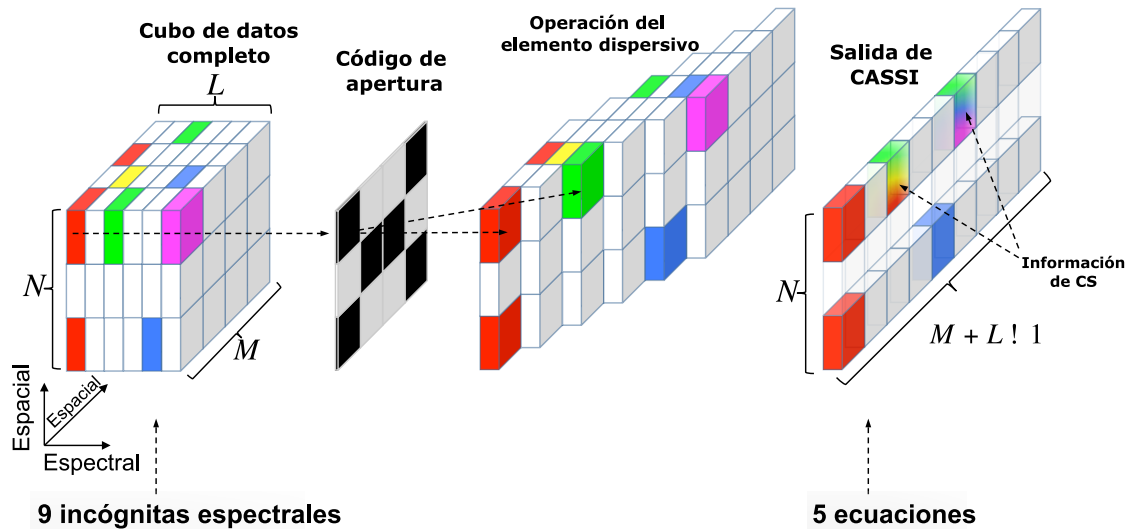


se codifica en primer lugar espacialmente mediante una apertura codificada  $T(x, y)$ , lo que da como resultado el campo codificado,

$$f_1(x, y, \lambda) = T(x, y)f_0(x, y, \lambda) \quad (4)$$

donde  $f_1(x, y, \lambda)$  representa el frente de onda modulado espacialmente.

**Figura 4.** Propagación de luz a través de la arquitectura CASSI. Tomado de<sup>42</sup>.



Específicamente la función de transmisión de la apertura codificada se describe

como,

$$T(x, y) = \sum_{i,j} t_{i,j} \text{rect} \left( \frac{x}{\Delta_c} - i, \frac{y}{\Delta_c} - j \right), \quad (5)$$

para  $i = 1, \dots, M$ , y  $j = 1, \dots, N$ , donde  $M \times N$  es la resolución espacial de la apertura codificada,  $t_{i,j} \in \{0, 1\}$  representa un elemento translúcido (1) o de bloqueo (0),  $\Delta_c$  es el tamaño de píxel del apertura codificada y  $\text{rect}(\cdot)$  representa la función rectángulo; seguidamente el frente de onda espacialmente codificado  $f_1(x, y, \lambda)$  pasa a través de un elemento dispersivo(prisma), el cual introduce una dispersión horizontal. La densidad espectral a la salida del elemento dispersivo, se puede expresar como,

$$f_2(x, y, \lambda) = \iint T(x', y') f_0(x', y', \lambda') \delta(x' - x - S(\lambda), y' - y) dx' dy', \quad (6)$$

donde  $\delta(x' - x - S(\lambda), y' - y)$  representa la respuesta de impulso óptico del sistema, tal que  $S(\lambda) = \alpha(\lambda)(\lambda - \lambda_c)$  es la dispersión inducida por el prisma a lo largo de  $x$  - axis, que está centrada en la longitud de onda  $\lambda_c$  y tiene un coeficiente de dispersión  $\alpha(\lambda)$ . La imagen de intensidad resultante en el FPA es la integración del campo  $f_2(x, y, \lambda)$  sobre la sensibilidad del rango espectral del detector  $\Lambda$  que se puede representar como,

$$\begin{aligned} g(x, y) &= \int_{\Lambda} f_2(x, y, \lambda) d\lambda, \\ &= \int_{\Lambda} \iint T(x', y') f_0(x', y', \lambda') \delta(x' - x - S(\lambda), y' - y) dx' dy' d\lambda, \\ &= \int_{\Lambda} T(x + S(\lambda), y) f_0(x + S(\lambda), y, \lambda) d\lambda, \end{aligned} \quad (7)$$

donde el último paso se deriva de la suposición de que se utilizan elementos ópticos

ideales <sup>43</sup>.

Usando un modelo discreto CASSI de primer orden <sup>44,45,46</sup>, una medida CASSI en el píxel  $(m, n)^{th}$  viene dada por,

$$g_{mn} = \iint p(m, n; x, y)g(x, y)dx dy + w_{mn}, \quad (8)$$

donde  $w_{mn}$  representa ruido aditivo y,

$$p(m, n; x, y) = \text{rect} \left( \frac{x}{\Delta_d} - m, \frac{y}{\Delta_d} - n \right), \quad (9)$$

representa la pixelación del detector, siendo  $\Delta_d$  el tamaño de píxel del detector, para  $m = 1, \dots, M, n = 1, \dots, N + L - 1$ , donde  $M \times N + L - 1$  es la resolución espacial FPA. Usando ecuación-(5) y ecuación- (7), la medición discreta del sistema CASSI

---

<sup>43</sup> RUEDA-CHACON, Hoover F. "Higher order discretization model for coded aperture spectral imaging systems". Tesis doct. University of Delaware, 2015.

<sup>44</sup> WAGADARIKAR, Ashwin, *et al.* "Single disperser design for coded aperture snapshot spectral imaging". En: *Applied optics* 47.10 (2008), B44-B51.

<sup>45</sup> WAGADARIKAR, Ashwin A, *et al.* "Video rate spectral imaging using a coded aperture snapshot spectral imager". En: *Optics express* 17.8 (2009), págs. 6368-6388.

<sup>46</sup> WAGADARIKAR, Ashwin Ashok. *Compressive spectral and coherence imaging*. Duke University, 2010.

se puede escribir como,

$$\begin{aligned}
g_{mn} &= \iint p(m, n; x, y)g(x, y)dxdy + w_{mn}, \\
&= \iiint_{\Lambda} \sum_{i,j} t_{i,j} \text{rect} \left( \frac{x}{\Delta_c} - i, \frac{y}{\Delta_c} - j \right) f_0(x + S(\lambda), y, \lambda) \times, \\
&\quad \text{rect} \left( \frac{x}{\Delta_d} - m, \frac{y}{\Delta_d} - n \right) d\lambda dxdy, \\
&= \sum_k t_{i-k,j} F_{i-k,j,k},
\end{aligned} \tag{10}$$

donde el último paso se deriva del hecho de que  $f_{i,j,k} = \iiint_{\Omega_{ijk}} f_0(x, y, \lambda)dxdyd\lambda$  es la energía del vóxel del cubo de datos  $(i, j, k)^{th}$  delimitado por  $\Omega_{ijk}$  y  $k = \{1, \dots, L\}$  indexa las bandas espectrales del cubo de datos. Por otro lado, la fuente espacio-espectral se puede representar como un cubo de datos discretizados calculando la energía en cada vóxel con la expresión

$$F_{i,j,k} = \int_{\lambda_k}^{\lambda_{k+1}} \int_{i\Delta}^{(i+1)\Delta} \int_{j\Delta}^{(j+1)\Delta} f_0(x, y, \lambda)dxdyd\lambda, \tag{11}$$

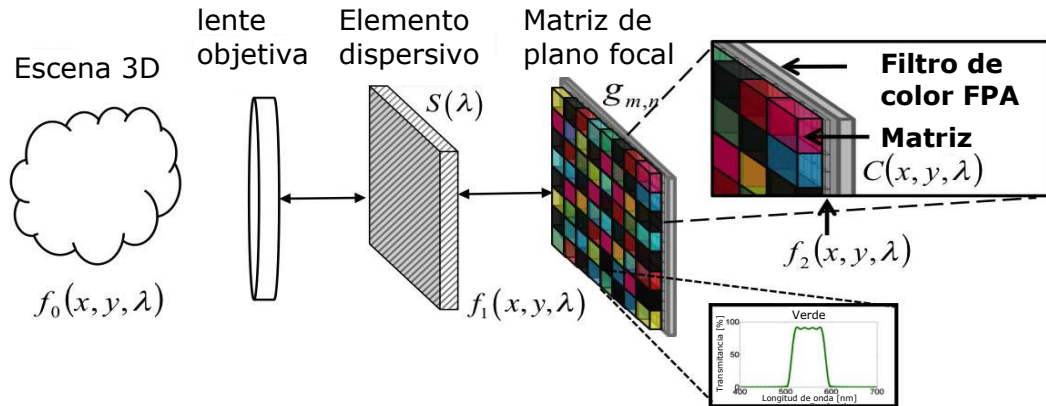
donde  $k = \{0, \dots, L - 1\}$  indexa las bandas espectrales. Esta expresión da como resultado un cubo de datos  $M \times N \times L$ , donde  $N$  corresponde a la resolución espacial y  $L$  es el número de bandas espectrales.

### 2.2.2. Sistema de adquisición de imágenes espectrales con detectores basados en filtros ópticos (SCCSI).

El sistema de adquisición de imágenes espectrales con detectores basados en filtros ópticos (SCCSI de su sigla en inglés Snapshot Colored Compressive Spectral Imager) basado en los conceptos CSI reemplaza el detector FPA tradicional por un detector con un arreglo de filtros ópticos, teniendo a su vez un elemento dispersivo que permite adquirir la información espacio espectral de la escena en un solo disparo. Consta también de un filtro óptico

cuya función es permitir el paso de la luz con determinadas propiedades, eliminando o atenuando la luz restante. Sus filtros más comunes son los filtros de color, los cuales bloquean o admiten el paso de luz a determinadas longitudes de onda <sup>47</sup>.

**Figura 5.** Esquema de la captura de imágenes espectrales de compresión en color (SCCSI). Tomado de<sup>48</sup>.



**Modelo SCCSI continuo** Denote la densidad de la fuente espacio-espectral en la figura-5 como  $f_0(x, y, \lambda)$ , donde  $(x, y)$  indexa las coordenadas espaciales y  $\lambda$  indexa la dimensión espectral. Como se mencionó anteriormente, la fuente  $f_0(x, y, \lambda)$  se dispersa primero por el elemento dispersivo que produce  $f_1(x, y, \lambda)$ . El campo disperso resultante se puede expresar como

$$f_1(x, y, \lambda) = \iint f_0(x', y', \lambda') h(x' - x - S(\lambda), y' - y) dx' dy', \quad (12)$$

donde  $h(\cdot)$  es la respuesta de impulso óptico del sistema y  $S(\lambda)$  es la función de dispersión del elemento dispersivo que opera a lo largo del eje de coordenadas  $x$ . Luego, la fuente dispersa es codificada por el detector con patrón de colores, que puede modelarse como una matriz de filtros ópticos  $T(x, y, \lambda)$ , seguido de un

<sup>47</sup> CORREA, Claudia V; ARGUELLO, Henry y ARCE, Gonzalo R. "Snapshot colored compressive spectral imager". En: *JOSA A* 32.10 (2015), págs. 1754-1763.

detector de irradiancia. Más específicamente, cada ubicación espacial de  $T(x, y, \lambda)$  está asociada con una respuesta espectral específica que modula la luz incidente en esa posición particular. Además, existe una correspondencia uno a uno entre los elementos de la matriz de filtros de color y los del FPA; por lo tanto, la energía en un píxel del detector se ve afectada por un solo elemento de filtro óptico. Por lo tanto, la fuente de irradiancia codificada y dispersa que incide en el detector se puede escribir como

$$f_2(x, y, \lambda) = f_1(x, y, \lambda)T(x, y, \lambda). \quad (13)$$

La medida de compresión indicada como  $g(x, y)$  se obtiene finalmente mediante la integración del campo  $f_2(x, y, \lambda)$  sobre la sensibilidad del rango espectral del detector. Más específicamente, usando las ecuaciones- (12) y (13), las medidas de compresión SCCSI se pueden escribir como

$$\begin{aligned} g(x, y) &= \int_{\Lambda} f_2(x, y, \lambda) d\lambda \\ &= \int_{\Lambda} \iint f_0(x', y', \lambda) h(x' - x - S(\lambda), y' - y) \times C(x, y, \lambda) dx' dy' d\lambda, \end{aligned} \quad (14)$$

donde  $\Lambda$  es la sensibilidad del rango espectral del detector.

**Modelo SCCSI discreto** Sea  $\Delta$  el tamaño de píxel del detector, de modo que el área ocupada por el elemento de píxel  $(m, n) - th$  se pueda representar como  $p(n, m; x, y) = \text{rect}\left(\frac{x}{\Delta} - n, \frac{y}{\Delta} - m\right)$ . Luego, la energía capturada en el píxel  $(m, n) - th$  del detector se puede escribir como

$$\begin{aligned} g_{mn} &= \iint g(x, y) p(m, n; x, y) dx dy \\ &= \int_{\Lambda} \iiint f_0(x', y', \lambda) h(x' - x - S(\lambda), y' - y) \times C(x, y, \lambda) \\ &\quad \text{rect}\left(\frac{x}{\Delta} - n, \frac{y}{\Delta} - m\right) dx' dy' dx dy d\lambda. \end{aligned} \quad (15)$$

El ancho de las bandas espectrales está determinado por la función de dispersión del prisma  $S(\lambda)$  y el paso de píxeles del detector  $\Delta$ . Teniendo en cuenta que se requieren columnas adicionales de píxeles en el detector debido a los efectos de dispersión a lo largo del eje  $x$ . En general, se necesita un detector FPA  $N \times V$ , donde  $V = (N + L - 1)$ . De manera similar, debido a que existe una coincidencia de 1 a 1 entre los filtros ópticos y los píxeles en el detector, la matriz de filtros de color se puede discretizar en términos del paso de píxeles de FPA,  $\Delta$ , como

$$C(x, y, \lambda) = \sum_{m,n,k} C_{m,n,k} \text{rect} \left( \frac{x}{\Delta} - n, \frac{y}{\Delta} - m, \frac{\lambda}{\Delta} - k \right), \quad (16)$$

La discretización en la ecuación-16 da como resultado una matriz 3D  $C_{m,n,k}$  con dos dimensiones espaciales y una espectral. Dada la correspondencia 1-1 entre los elementos de la matriz de filtros de color y el detector FPA, las dimensiones espaciales del primero se corresponden con las del segundo. Además, las respuestas espectrales de los filtros ópticos en  $C_{m,n,k}$  en el detector se discretizan en valores  $L$ . Suponiendo respuestas espectrales normalizadas de los filtros, las entradas de la matriz de filtros de color,  $C_{m,n,k}$ , están en el conjunto  $\{0, \dots, 1\}$ . Por el contrario, cuando se asumen respuestas espectrales ideales, las entradas de  $C_{m,n,k}$  son cero o uno. La transmitancia de la matriz de filtros de color, que es la cantidad de energía que pasa, se puede escribir como

$$Tr = \frac{1}{NVL} \sum_{m,n,k} C_{m,n,k}. \quad (17)$$

Cada píxel espacial del detector se puede asociar a un filtro óptico diferente. En otras palabras, una matriz de filtros de color  $N \times V$  puede verse como un mosaico de filtros ópticos  $NV$ . En la práctica, tal cantidad de filtros diferentes no es alcanzable porque el número máximo de posibles respuestas está determinado por el número de bandas espectrales  $L$  y viene dado por  $\sum_{i=1}^L \binom{L}{i}$ . Por lo tanto,  $C_{m,n,k}$  se limita a un

conjunto de filtros disponibles, normalmente con un máximo de  $L$  filtros diferentes que se utilizan para recuperar la misma cantidad de bandas espectrales.

Usando la ecuación-16 y, asumiendo una PSF ideal, la salida en el detector en la ecuación-15 se puede reescribir como

$$g_{mn} = \sum_{m,n,k} C_{m,n,k} \int_{\Lambda} \iint f_0(x + S(\lambda), y, \lambda) \times \text{rect} \left( \frac{x}{\Delta} - n, \frac{y}{\Delta} - m, \frac{\lambda}{\Delta} - k \right) dx dy d\lambda. \quad (18)$$

Ahora, usando la ecuación-11 y teniendo en cuenta la función de transferencia del elemento dispersivo, ecuación-18 puede, a su vez, expresarse como

$$g_{m,n} = \sum_{m,n,k} F_{m,(n-k),k} C_{m,n,k} \times \iiint \text{rect} \left( \frac{x}{\Delta} - n, \frac{y}{\Delta} - m, \frac{\lambda}{\Delta} - k \right) dx dy d\lambda. \quad (19)$$

Tenga en cuenta que el cambio en los índices de  $F$  representa el efecto del elemento dispersivo. Además, la función de dispersión del prisma  $S(\lambda)$  tiene un efecto crítico en la incidencia de energía desde un vóxel fuente a cada uno de los píxeles detectores. Más específicamente, después de la dispersión, el vóxel de cubo cuadrado entrante se corta de tal manera que su energía incidirá no solo en un solo píxel FPA sino en el conjunto de hasta tres píxeles vecinos. La salida discretizada en el detector FPA, por lo tanto, se puede expresar sucintamente como

$$g_{m,n} = \sum_{k=0}^{L-1} \sum_{u=0}^2 w_{m,n,k,u} F_{m,(n-k-u),k} C_{m,n,k}. \quad (20)$$

### 2.3. CLASIFICACIÓN DE IMÁGENES ESPECTRALES

En las imágenes espectrales (SI) los valores de radiación contenidos en cada uno de los píxeles plotados a múltiples y continuas longitudes de onda origina una firma espectral característica. Las firmas espectrales son ampliamente utilizadas en tra-

bajos para detectar, diferenciar, discriminar o clasificar entre dos o más elementos. La gran cantidad de información espectral que proporcionan las SI permite que sean ampliamente usadas en áreas como la clasificación derivándose esta en supervisada y no supervisada <sup>49</sup>.

A continuación, se presentan dos tipos de clasificación, la clasificación supervisada y no supervisada que permiten explorar diferentes tipos de atributos o clases por medio del análisis estadístico multivariado, este proceso identifica los valores de cada píxel de una o varias bandas de una CSI, crea y evalúa las clases o clúster (firmas) y finalmente reclasifica de acuerdo a las probabilidades de cada clase.

- **Clasificación no supervisada:** A comparación de la clasificación supervisada la clasificación no supervisada identifica los valores de cada píxel inscritos en un cubo de datos a distintas longitudes de onda creando una búsqueda automática de grupos con valores semejantes comúnmente llamados clúster. La ventaja de este método se deriva en que no se requiere ningún conocimiento previo para obtener las clases, es decir lo realiza en base a probabilidades. En la técnica de clasificación de agrupamiento (Clustering – Clúster) la imagen es segmentada en clases desconocidas que posteriormente serán etiquetadas. Ésta se emplea en diversos métodos, por ejemplo: K-medias, C-medias e ISODATA (Interactive Self-Organizing Data Analysis Technique (Algorithm)). Básicamente este método busca minimizar los errores cuadráticos estimando la distancia mínima euclidiana ( $\ell_1 - norm$ ) entre cada pixel que conforman una muestra con el fin de obtener un centroide <sup>49</sup>.

- **Clasificación supervisada:** La clasificación supervisada utiliza un conjunto

---

<sup>49</sup> GUTIÉRREZ, Mauricio Arango. "Clasificación no supervisada de coberturas Vegetales sobre imágenes digitales de sensores Remotos: "landsat-etm+"". En: *Revista Facultad Nacional de Agronomía Medellín* 58.1 (2005), págs. 2611-2634.

de muestras previamente entrenadas que consiste en la etiqueta de clase de cada pixel mediante su firma espectral que sirven para delimitar las clases de materiales presentes como criterio de clasificación. A pesar de la limitación que presenta este método debido a la ausencia de información la precisión en la clasificación es mayor comparable con los métodos no supervisados <sup>50</sup>.

**2.3.1. Redes neuronales artificiales, aprendizaje profundo.** Las redes neuronales y aprendizaje profundo (Deep learning) son redes interconectadas masivamente en paralelo de elementos simples y con organización jerárquica compuestas de capas de entrada y de salida, por las cuales se procesa la información para la identificación de objetos o características <sup>51</sup>. El aprendizaje profundo (DL) permite que los modelos computacionales que se componen de múltiples capas de procesamiento aprendan representaciones de datos con múltiples niveles de abstracción. Estos métodos han mejorado dramáticamente el estado del arte en reconocimiento de voz <sup>52</sup>, reconocimiento de objetos visuales <sup>53</sup>, detección de objetos <sup>54</sup>, entre otros

---

<sup>50</sup> PÉREZ, Aritz; LARRANAGA, Pedro e INZA, Inaki. "Modelos gráficos probabilísticos para la clasificación supervisada empleando la estimación basada en kernels Gaussianos esféricos". En: *III Taller Nacional de Minería de Datos y Aprendizaje* (2005), págs. 125-134.

<sup>51</sup> MATICH, Damián Jorge. "Redes Neuronales: Conceptos básicos y aplicaciones". En: *Universidad Tecnológica Nacional, México* (2001).

<sup>52</sup> DENG, Li; HINTON, Geoffrey y KINGSBURY, Brian. "New types of deep neural network learning for speech recognition and related applications: An overview". En: *2013 IEEE international conference on acoustics, speech and signal processing*. IEEE. 2013, págs. 8599-8603.

<sup>53</sup> CICHY, Radoslaw Martin, *et al.* "Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence". En: *Scientific reports* 6 (2016), pág. 27755.

<sup>54</sup> LIU, Li, *et al.* "Deep learning for generic object detection: A survey". En: *International journal of computer vision* 128.2 (2020), págs. 261-318.

<sup>55, 56, 57</sup>. El aprendizaje profundo descubre una estructura compleja en grandes conjuntos de datos mediante el uso del algoritmo de retropropagación para indicar cómo una máquina debe cambiar sus parámetros internos que se utilizan para calcular la representación en cada capa a partir de la representación en la capa anterior. Las redes convolucionales profundas han logrado avances en el procesamiento de imágenes, video, voz y audio, mientras que las redes recurrentes han iluminado datos secuenciales como el texto y la voz <sup>58</sup>.

Específicamente, la clasificación se ha enfocado en la asignación de un objeto o un fenómeno físico a una de las diversas categorías o clases especificadas, siendo una técnica muy útil para el reconocimiento de patrones, en donde las redes convolucionales han sido usadas para tareas de clasificación sobre todo en el área de imágenes, tomándolas como entradas y asignándoles importancias (pesos) a ciertos elementos en la imagen para así poder diferenciar sus características <sup>59</sup>.

### ■ Redes neuronales convolucionales

Las redes neuronales convolucionales (CNN) son un algoritmo de DL que es-

---

<sup>55</sup> GAWEHN, Erik; HISS, Jan A y SCHNEIDER, Gisbert. "Deep learning in drug discovery". En: *Molecular informatics* 35.1 (2016), págs. 3-14.

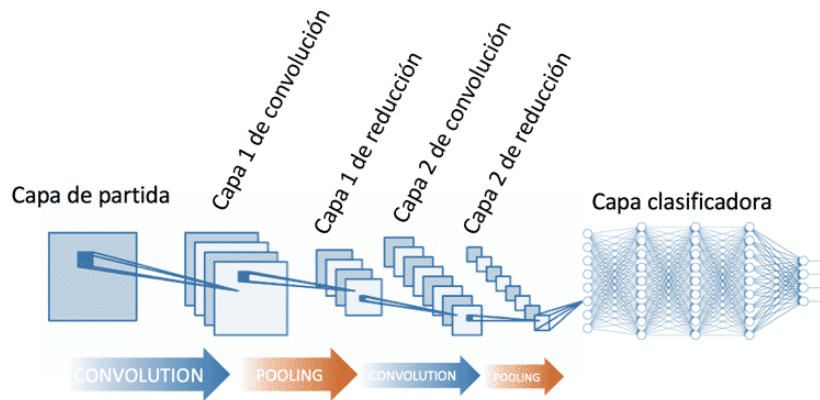
<sup>56</sup> ZOU, James, *et al.* "A primer on deep learning in genomics". En: *Nature genetics* 51.1 (2019), págs. 12-18.

<sup>57</sup> YANG, Zhi, *et al.* "Deep transfer learning for military object recognition under small training set condition". En: *Neural Computing and Applications* 31.10 (2019), págs. 6469-6478.

<sup>58</sup> LECUN, Yann; BENGIO, Yoshua e HINTON, Geoffrey. "Deep learning". En: *nature* 521.7553 (2015), págs. 436-444.

<sup>59</sup> GRAVES, Alex; MOHAMED, Abdel-rahman e HINTON, Geoffrey. "Speech recognition with deep recurrent neural networks". En: *2013 IEEE international conference on acoustics, speech and signal processing*. IEEE. 2013, págs. 6645-6649.

**Figura 6.** Funcionamiento de una CNN. Adaptada de <sup>60</sup>.



tán diseñadas para procesar datos que vienen en forma de múltiples matrices, por ejemplo, una imagen en color compuesta por tres matrices 2D que contienen intensidades de píxeles en los tres canales de color. Muchas modalidades de datos tienen la forma de múltiples matrices: 1D para señales y secuencias, incluido el lenguaje; 2D para imágenes o espectrogramas de audio; y 3D para video o imágenes volumétricas. Hay cuatro ideas clave detrás de CNN que aprovechan las propiedades de las señales naturales: conexiones locales, pesos compartidos, agrupación y el uso de muchas capas <sup>61</sup>.

La arquitectura de una CNN típica figura-6 está estructurada como una serie de etapas. Las primeras etapas se componen de dos tipos de capas: capas convolucionales y capas de agrupación. Las unidades en una capa convolucional están organizadas en mapas de características, dentro de los cuales cada unidad está conectada a parches locales en los mapas de características de la capa anterior a través de un conjunto de pesos llamado banco de filtros. El resultado de esta suma ponderada local se pasa a través de una no lineali-

---

<sup>61</sup> KRIZHEVSKY, Alex; SUTSKEVER, Ilya e HINTON, Geoffrey E. "Imagenet classification with deep convolutional neural networks". En: *Advances in neural information processing systems* 25 (2012), págs. 1097-1105.

dad. Todas las unidades en un mapa de características comparten el mismo banco de filtros <sup>62</sup>. Los diferentes mapas de características en una capa usan diferentes bancos de filtros. La razón de esta arquitectura es doble. Primero, en los datos de matriz, como las imágenes, los grupos locales de valores a menudo están altamente correlacionados, formando motivos locales distintivos que se detectan fácilmente <sup>59</sup>. En segundo lugar, las estadísticas locales de imágenes y otras señales son invariables para la ubicación. En otras palabras, si un motivo puede aparecer en una parte de la imagen, podría aparecer en cualquier lugar, de ahí la idea de unidades en diferentes ubicaciones que compartan los mismos pesos y detecten el mismo patrón en diferentes partes de la matriz. Matemáticamente, la operación de filtrado realizada por un mapa de características es una convolución discreta, de ahí el nombre <sup>63</sup>.

- Enfoque de aprendizaje de diccionario

Debido a que las señales e imágenes de interés pueden ser escasas o comprimibles en algún diccionario se ha tenido mucho interés en el procesamiento de imágenes, señales y visión. El diccionario puede estar basado en un modelo matemático de los datos o puede aprenderse directamente de los datos. Se ha observado que aprender un diccionario directamente a partir de los datos de entrenamiento en lugar de usar un diccionario predeterminado (es decir, wavelet) generalmente conduce a una representación más compacta y, por lo tanto, puede proporcionar mejores resultados en muchas aplicaciones prácticas de

---

<sup>62</sup> TOMPSON, Jonathan J, *et al.* "Joint training of a convolutional network and a graphical model for human pose estimation". En: *Advances in neural information processing systems*. 2014, págs. 1799-1807.

<sup>63</sup> LEUNG, Michael KK, *et al.* "Deep learning of the tissue-regulated splicing code". En: *Bioinformatics* 30.12 (2014), págs. i121-i129.

procesamiento de imágenes, como restauración y clasificación <sup>64</sup>.

**2.3.2. Red neuronal basada en una estructura U-Net.** Esta arquitectura fue propuesta por Ronneberger et al. <sup>65</sup> como un modelo para realizar una segmentación semántica de imágenes de forma rápida y precisa, que permite obtener una tasa de error baja y resultados destacables basandose en las características de los píxeles de la imagen

A esta arquitectura se concatena el modelo óptico propuesto para realizar múltiples capturas. Hasta ahora, la U-Net ha superado el mejor método anterior (una red convolucional de ventana deslizante) en el desafío ISBI para la segmentación de estructuras neuronales en pilas de microscopía electrónica.

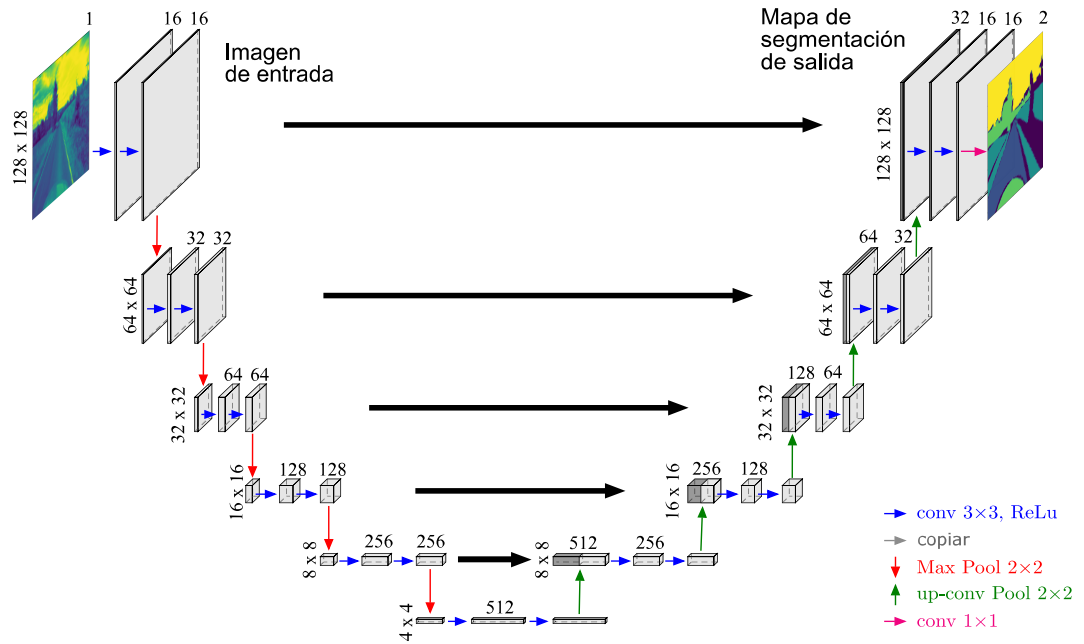
Siendo una arquitectura simétrica para la segmentación semántica, la U-Net consiste en dos partes principales: un camino de contracción y un camino de expansión. La ruta de contracción es un componente de agrupación de convolución que sigue la arquitectura típica de una red convolucional, en esta se usan bloques de dos capas convolucionales de 3x3 que aumentan la profundidad de la imagen, seguidas de una capa de agrupación máxima de 2x2 que reduce a la mitad el tamaño de la imagen. Además, los mapas de características después de cada bloque se duplican para que la arquitectura aprenda efectivamente estructuras más complejas. Entre el camino de contracción y expansión se encuentra un bloque compuesto por dos capas de convolución de 3x3 seguidas de una capa de convolución superior de 2x2. Finalmente, en la expansión cada bloque consta de dos capas de convolución de 3x3 seguidas de una capa de convolución superior de 2x2 . Después de cada

---

<sup>64</sup> VAN NGUYEN, Hien, et al. "Kernel dictionary learning". En: *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2012, págs. 2021-2024.

<sup>65</sup> RONNEBERGER, O.; PFISCHER y BROX, T. "U-Net: Convolutional Networks for Biomedical Image Segmentation". En: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Vol. 9351. LNCS. (available on arXiv:1505.04597 [cs.CV]). Springer, 2015, págs. 234-241.

**Figura 7.** Arquitectura U-Net para la segmentación de características. El número de canales se indica encima de los cuboides y los números verticales indican el tamaño de la fila y la columna en píxeles. Adaptada de<sup>66</sup>.



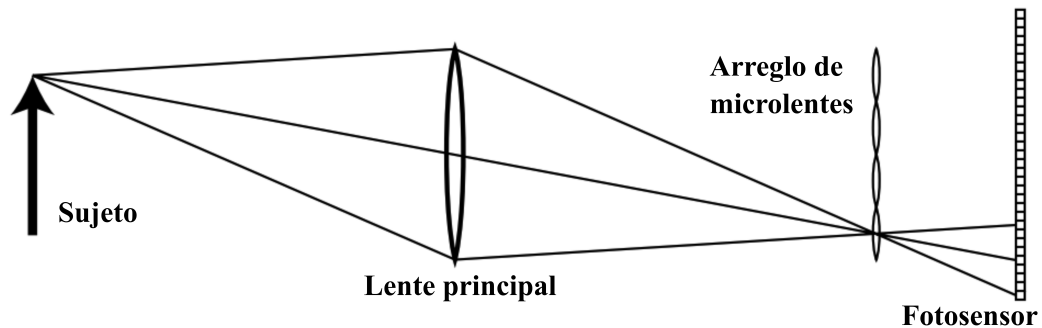
bloque, se obtiene la mitad del número de mapas de características utilizados por la capa convolucional para mantener la simetría y luego se conecta con el mapa de características original. En la capa final, el mapeo resultante pasa a través de una convolución 1x1 con el número de mapas de características igual al número de segmentos deseados.

## 2.4. DISEÑO ÓPTICO

**2.4.1. Configuración básica:** La configuración óptica sencilla comprende una lente principal fotográfica (como una lente de 50 mm en una cámara de formato de 35 mm), una matriz de microlentes y una matriz de fotosensores de paso más fino. La lente principal se puede trasladar a lo largo de su eje óptico, exactamente como en una cámara convencional, para enfocar un tema de interés a una profundidad deseada. Como se muestra en figura-8, los rayos de luz de un solo punto sobre

el sujeto se llevan a un solo punto de convergencia en el plano focal de la matriz de microlentes. La microlente en esa ubicación separa estos rayos de luz según la dirección, creando una imagen enfocada de la apertura de la lente principal en la matriz de píxeles debajo de la microlente.

**Figura 8.** La imagen ilustra los elementos conceptuales básicos que contiene una cámara. Adaptada de<sup>67</sup>



**2.4.2. Enfoque de micro lentes al infinito óptico:** La imagen bajo una microlente dicta la resolución direccional del sistema para esa ubicación en la película. Para maximizar la resolución direccional, se requieren las imágenes de microlente más nítidas posibles. Para lograr una mejor nitidez, se deben enfocar las microlentes en el plano principal de la lente principal. Dado que las microlentes son extremadamente pequeñas en comparación con la lente principal, la lente principal se fija de manera efectiva en el infinito óptico de las microlentes. Por lo tanto, para enfocar las microlentes se cementa el plano del fotosensor en la profundidad focal de las microlentes.

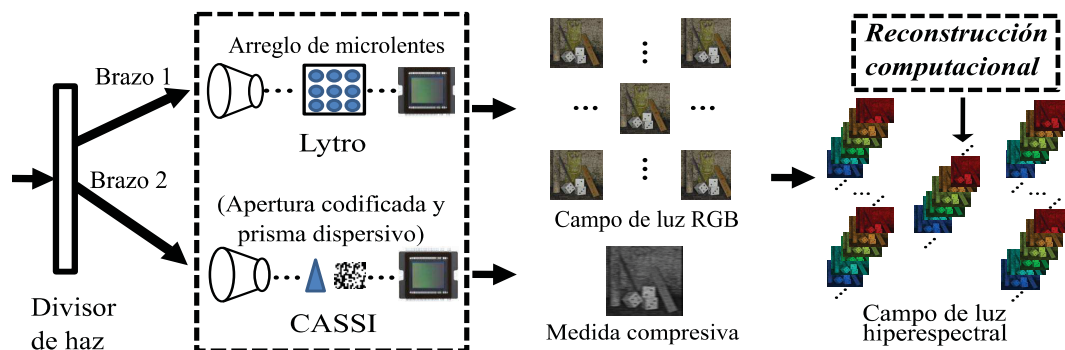
Las desviaciones de esta separación dan como resultado un desenfoque incorrecto en las subimágenes de la microlente. Para obtener una imagen nítida dentro de la profundidad de campo de las microlentes, necesitamos que la separación entre las microlentes y el fotosensor tenga una precisión de  $\Delta_{x(p)}, (f_m/\Delta_{x(m)})$ , donde  $\Delta_{x(p)}$  es el ancho de un píxel de sensor,  $f_m$  es la profundidad focal de las microlentes y

$\Delta_{x(m)}$  es el ancho de una microlente. Por ejemplo, en nuestro prototipo,  $\Delta_{x(p)} = 9$  micrones,  $f_m = 500$  micrones y  $\Delta_{x(m)} = 125$  micrones.

### 2.4.3. Sistemas tradicionales de campo de luz espectral de compresión:

**Sistemas de doble brazo** El uso de enfoques de información lateral requiere sistemas de doble brazo, que imponen un equilibrio entre la calidad de la reconstrucción, el índice de compresión, la complejidad de la implementación y los costos de implementación. Específicamente, la arquitectura Lytro-CASSI se simula como un sistema de instantáneas múltiples, donde el brazo Lytro adquiere una representación RGB de todos los ángulos de visión, mientras que el brazo CASSI adquiere K versiones compresivas del campo de visión completo del SLF<sup>68</sup>. Se tiene también en el sistema de brazo doble, un divisor de haz que divide igualmente la luz reflejada de la escena en el sensor de campo de luz y el generador de imágenes espectrales de compresión.

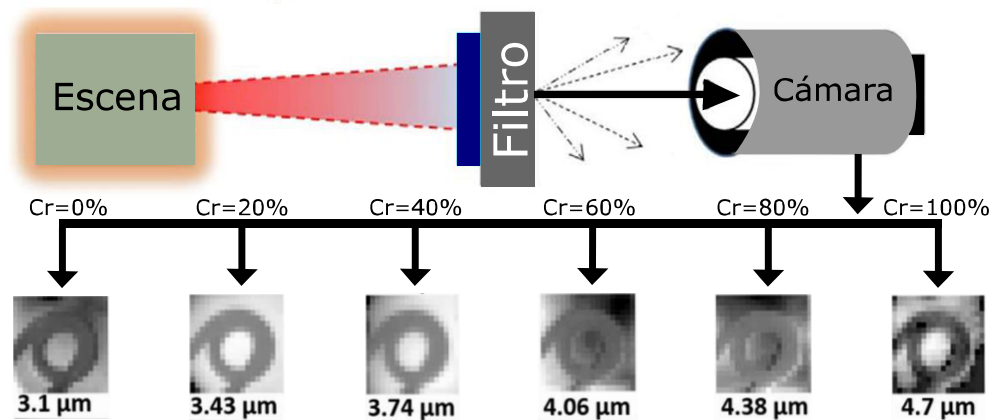
**Figura 9.** Flujo de datos en la arquitectura Lytro-CASSI. Tomada de<sup>69</sup>.



<sup>68</sup> MARQUEZ, Miguel; RUEDA-CHACON, Hoover y ARGUELLO, Henry. "Compressive spectral light field image reconstruction via online tensor representation". En: *IEEE Transactions on Image Processing* 29 (2020), págs. 3558-3568.

**Sistema de matriz de cámaras con filtros de paso de banda** Éste enfoque para la obtención de imágenes multispectrales históricamente se ha basado en tecnologías como cristales líquidos o prismas giratorios figura-10; sin embargo, estos tienen restricciones de longitud de onda y pueden ser bastante lentos.

**Figura 10.** Ejemplo del método de obtención de imágenes con múltiples capturas durante el proceso de cristalización. Adaptada de<sup>70</sup>.

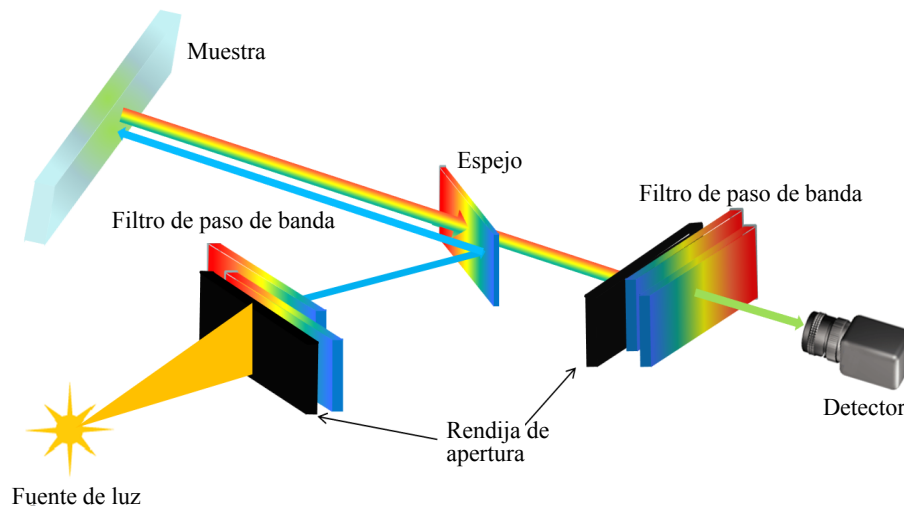


Se puede usar un filtro de paso de banda para tomar una imagen multispectral colocándolo en el camino óptico de una cámara de banda ancha y tomando una fotografía para cada estado del filtro, figura-11. La velocidad a la que se puede reajustar el filtro determina cuánto tiempo (cuántos fotogramas de la cámara) se necesitan para construir una imagen multispectral (en la mayoría de los casos es ventajoso que los anchos de banda del filtro se superpongan parcialmente, ya que esto se puede utilizar para mejorar la resolución espectral mediante procesamiento)<sup>71</sup>.

**2.4.4. Cámara plenótica:** La cámara plenótica tradicional se basa en una serie de microlentes en el plano de la imagen de la lente de la cámara principal, con el

<sup>71</sup> TRIMBY, Liam; BALDYCHEVA, Anna y WRIGHT, C David. "Phase-change band-pass filters for multispectral imaging". En: *Photonic and Phononic Properties of Engineered Nanostructures VIII*. Vol. 10541. International Society for Optics y Photonics. 2018, 105412B.

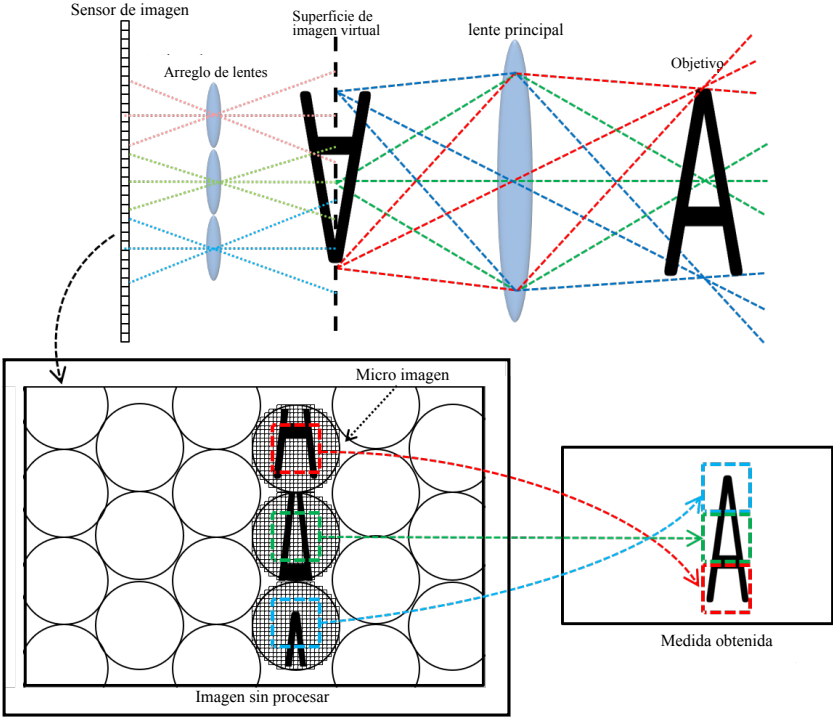
**Figura 11.** Arquitectura óptica usando filtros de paso de banda múltiple. Tomada de<sup>72</sup>.



sensor colocado a una distancia focal detrás de las microlentes, figura-12. La cámara muestrea el resplandor frente a las microlentes con un núcleo. Cada imagen de microlente es una pila vertical de muestras en el plano  $q, p$ , capturando estrictamente la distribución angular de la radiancia en el plano de la imagen.

Las imágenes se obtienen a partir del resplandor capturado por la cámara plenóptica tradicional integrando todas las muestras angulares en un punto espacial particular. Sin embargo, cada punto espacial es muestreado por una única microlente, por lo que el renderizado implica la integración de todos los píxeles en cada microimagen. El renderizado de la cámara plenóptica produce solo 1 píxel por microlente, lo que da como resultado una imagen renderizada con muy baja resolución.

Figura 12. Funcionamiento de una cámara plenoptica convencional. Adaptada de<sup>73</sup>.



### **3. METODOLOGÍA**

En el presente trabajo se desarrolla un algoritmo de aprendizaje profundo que basado en el uso de un arreglo de microlentes se pretende mejorar las tareas de clasificación a partir del uso de una única proyección comprimida de una imagen espectral. El algoritmo se desarrolla en el lenguaje de programación Python y Matlab, para luego unirse y ejecutarse en Google Colab.

**Objetivo específico No. 1. “Estudiar diferentes modelos tradicionales de adquisición y clasificación de imágenes espectrales, y sus requerimientos técnicos.”**

#### **3.1. MODELOS DE ADQUISICIÓN Y CLASIFICACIÓN DE SI**

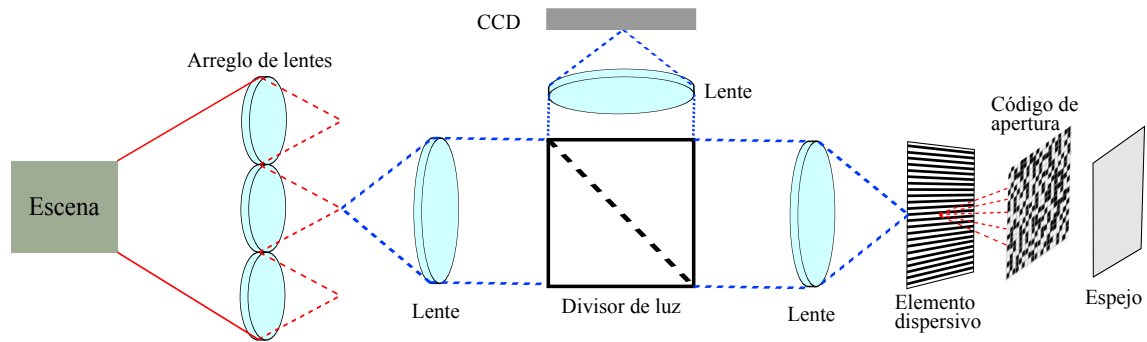
Para dar cumplimiento al objetivo específico No.1, la primera etapa de la metodología consiste en realizar una búsqueda de diferentes métodos ya establecidos para adquirir y clasificar imágenes espectrales, dichos métodos fueron consultados en diferentes bases de datos y avalados por revistas científicas para uso investigativo. Con ellos sentamos las bases para iniciar a plantear nuestra arquitectura.

Los diferentes métodos expuestos en esta fase de la metodología se exponen en el numeral 2.2 y 2.3 de la sección No. 2. Marco teórico.

**Objetivo específico No. 2. “Seleccionar y modelar una metodología de muestreo compresivo espectral en el cual sea viable la incorporación de un arreglo de lentes para la adquisición de múltiples versiones de la escena con diferente codificación.”**

De acuerdo con lo establecido en el objetivo específico No. 2, utilizamos el sistema óptico DD-CASSI, seleccionado como resultado de abordar los diferentes métodos de adquisición de SI explorados en la primera fase de la metodología. El modelo mencionado se expone a continuación:

**Figura 13.** Imagen representativa del campo de visión del sistema de lentes propuesto y su funcionamiento dentro de la arquitectura utilizada. Fuente: Autor.



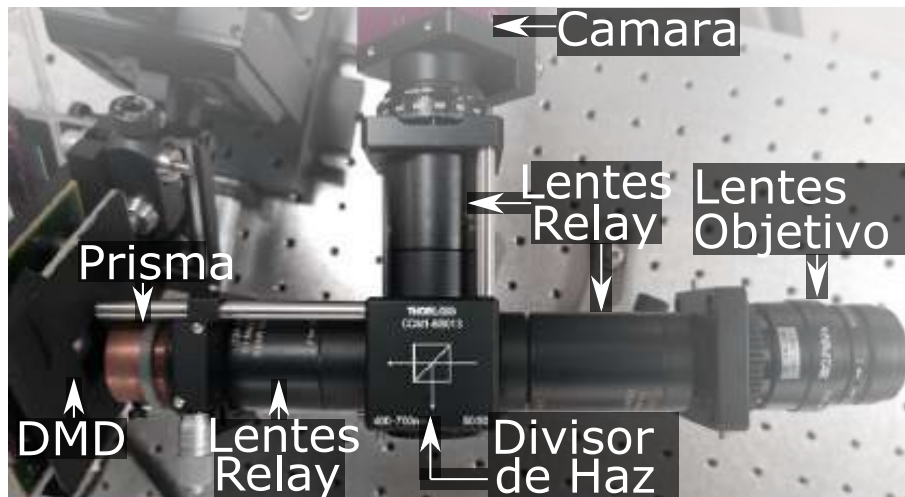
### 3.1.1. Dispensor dual de imágenes espectrales de instantáneas de apertura codificada (DD - CASSI)

A diferencia de un sensor pushbroom que tradicionalmente toma  $n$  instantáneas en total (una instantánea  $1 \times m$  por fila) para detectar una escena espacial  $N \times M \times L$ . El sensor CASSI-DD, por otro lado, toma medidas dentro de toda la escena espacial y espectral en una sola instantánea. Por cada instantánea tomada, una apertura codificada bloquea el 50% de los datos<sup>74</sup>.

DD-CASSI está compuesto por una lente de objetivo, un par de lentes de relé, un dispositivo de espejo digital (DMD), un divisor de haz, un prisma y un sensor de escala de grises, acomodados como se detalla en la figura-14. Esta arquitectura funciona de la siguiente manera: en primer lugar, el lente objetivo enfoca la escena en el plano de la imagen de una lente de relevo, que propaga el frente de onda a tra-

<sup>74</sup> BUSUIOCEANU, Maria, *et al.* "Evaluation of the CASSI-DD hyperspectral compressive sensing imaging system". En: *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XIX*. Vol. 8743. International Society for Optics y Photonics. 2013, pág. 87431V.

**Figura 14.** Esquema de la arquitectura DD-CASSI. Tomado de<sup>75</sup>.



vés de un divisor de haz hasta una segunda lente de relevo. Esta segunda lente de relé propaga el frente de onda entrante a través de un prisma para luego enfocar la versión espectral dispersa de la imagen en el DMD. En segundo lugar, aprovechando la superficie del espejo del DMD, el frente de onda espectral codificado disperso se devuelve al prisma y luego a la segunda lente de relé. Aquí, el prisma deshace la dispersión espectral, lo que da como resultado un frente de onda espectral disperso codificado disperso. Finalmente, la segunda lente de relé propaga el frente de onda entrante disperso-codificado-disperso a través del divisor de haz, redirigiéndolo a una tercera lente de relé que enfoca el frente de onda en el sensor CCD<sup>76</sup>.

**Objetivo específico No. 3. “Diseñar y modelar un sistema espectral compresivo para adquirir múltiples versiones comprimidas de la escena a partir de una única captura.”**

En cumplimiento con el objetivo específico No. 3, Tenemos un modelo con doble

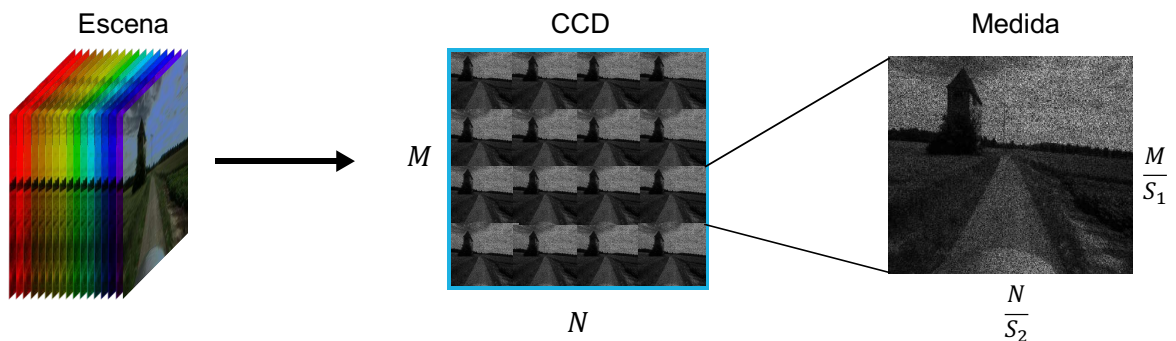
---

<sup>76</sup> MONSALVE, Jonathan, *et al.* “Compressive Covariance Matrix Estimation from a Dual-Dispersive Coded Aperture Spectral Imager”. En: *2021 IEEE International Conference on Image Processing (ICIP)*. IEEE. 2021, págs. 2823-2827.

codificación para capturar una versión comprimida del resultado del arreglo de microlentes, obteniendo múltiples vistas angulares.

Utilizamos arreglos de lentes de tal manera que la imagen que con ellos se genere ocupe el mismo tamaño del sensor. Tenemos que la resolución espacial del sensor CCD no cambia de tal manera que se deja fija. Por ejemplo: Se tiene una resolución  $M \times N$  e independientemente de las versiones que se generen en el CCD, se tiene que la versión espacial disminuye en tamaño, es decir, si tengo un solo microlente toda la imagen ocupa el sensor con la resolución inicial de  $M \times N$  pero, si tengo un arreglo de microlentes  $S_1 \times S_2$ , para este caso  $S_1 = S_2$  donde  $S_1$  y  $S_2$  son el número de capturas, obteniendo cada imagen con una resolución de  $\frac{M}{S_1} \times \frac{N}{S_2}$  y así sucesivamente.

**Figura 15.** Visualización de las medidas para un arreglo de lentes en el sensor CCD. Fuente: Autor.



Con lo anterior nos aseguramos de que con una sola captura podamos hacer la clasificación, simulando con la ayuda del arreglo de microlentes múltiples capturas en un solo disparo.

**3.1.2. Modelo de detección continuo y discreto** Sea  $f_0(x, y, \lambda)$  la función que modela la información espacio-espectral de la escena. El frente de onda entrante es modulada por un arreglo de lentes que genera múltiples versiones repetidas de baja

resolución espacial de la escena, donde  $x$  y  $y$  indexa las coordenadas espaciales y  $\lambda$  indexa la longitud de onda. Luego, un lente objetivo estándar forma una imagen en el plano de entrada de un sistema  $4f$ , que propaga el frente de onda entrante a través de un prisma y luego forma una versión de imagen dispersa sobre la apertura codificada. La versión modulada dispersa del objeto se puede expresar como

$$\tilde{g}(x, y, \lambda) = Q(x, y) \iint h(x - x' - S(\lambda), y - y', \lambda) \times f(x, y, \lambda) dx' dy', \quad (21)$$

con

$$Q(x, y) = \frac{1}{2} \sum_{i_1, i_2} C_{i_1, i_2, j} \text{rect} \left( \frac{y - i_1 \Delta_c}{\Delta_c} \right) \text{rect} \left( \frac{y - i_2 \Delta_c}{\Delta_c} \right), \quad (22)$$

donde  $\text{rect}(\cdot)$  representa la función rectangular,  $S(\lambda)$  es una función dependiente de la longitud de onda,  $h(x - x' - S(\lambda), y - y', \lambda)$  representa la respuesta al impulso del sistema óptico,  $\Delta_c$  denota el tamaño de píxel,  $\mathbf{C} \in \mathbb{R}^{M \times N \times L}$  representa la apertura codificada binaria con  $C_{i,j,k} \in \{0, 1\}$ ,  $i = \{0, \dots, M - 1\}$ ,  $j = \{0, \dots, N - 1\}$ ,  $M = N + L - 1$  y  $k = \{0, \dots, L - 1\}$  indexando las filas, columnas e instantáneas, respectivamente. Luego, el frente de onda modulado disperso se devuelve al prisma en el que se deshace el efecto de dispersión. El frente de onda disperso-codificado-disperso resultante se propaga y se enfoca en una matriz de detectores. Este proceso de detección se puede expresar como

$$g(x, y) = \iiint h(x - x' + S(\lambda), y - y', \lambda) \tilde{g}(x, y, \lambda) dx' dy' d\lambda, \quad (23)$$

donde  $g(\cdot)$  representa la medida comprimida. Con base en la ecuación-22 y 23 el

proceso de muestreo se puede representar matricialmente como,

$$G = \sum_{k=0}^{L-1} F_{:, :, k} \cdot C_{:, :, k} + \epsilon_{:, :, k}, \quad (24)$$

donde  $\mathbf{F} \in \mathbb{R}^{M \times N \times L}$  es la representación tensorial de la escena y  $\epsilon \in \mathbb{R}^{M \times N}$  representa el ruido agregado por el sistema. Con base en la ecuación-24, para el modelo de este trabajo tenemos en cuenta que el sistema óptico que proponemos se puede representar como un sistema lineal de la forma,

$$\mathbf{g} = \mathbf{H}\mathbf{f} + \epsilon, \quad (25)$$

donde  $\mathbf{f} = [\mathbf{f}_0^T, \dots, \mathbf{f}_n^T]^T \in \mathbb{R}^{n \times 1}$  es la representación vectorial de la imagen espectral,  $n = M \cdot N \cdot L$ ,  $\mathbf{g} \in \mathbb{R}^{m \times 1}$  representa las medidas comprimidas,  $m = M \cdot N \cdot L$ , y  $\mathbf{H} \in \mathbb{R}^{m \times n}$  es la matriz de detección. Específicamente,  $\mathbf{H}$  puede ser definido como

$$H_{i,j} = \begin{cases} f_n & \text{Si } i = \text{mod}(j, M \cdot N), \\ 0, & \text{de lo contrario} \end{cases} \quad (26)$$

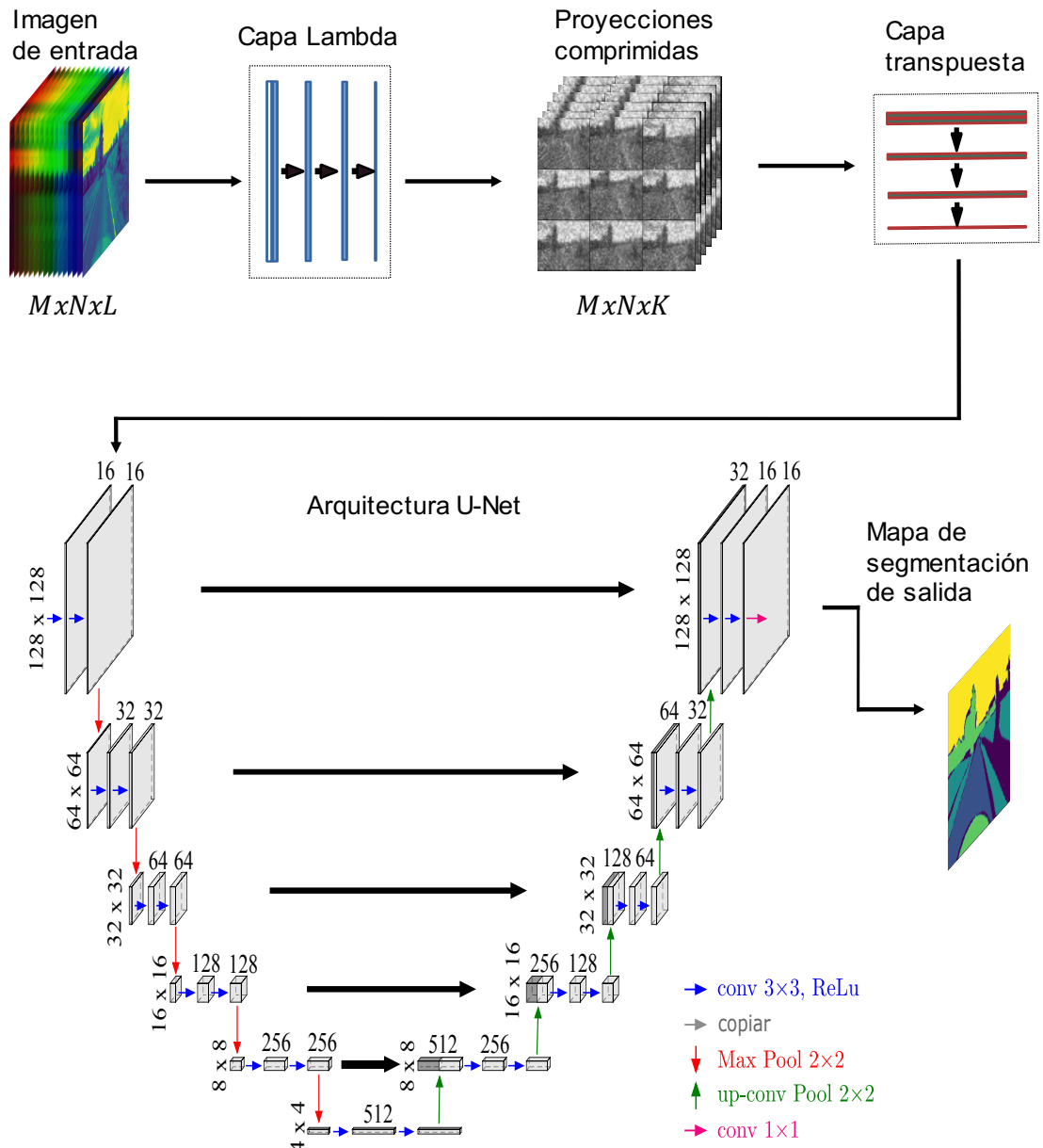
donde  $n = \text{mod}(j, M \cdot N) + N(\frac{j}{M \cdot N})$ . Para el caso del presente trabajo tendríamos que  $\mathbf{f}$  representa el plano imagen generado por nuestro sistema óptico.

**Objetivo específico No. 4. “Implementar una red neuronal para la clasificación de imágenes espectrales a partir del uso de las medidas comprimidas adquiridas por el sistema propuesto.”**

Para desarrollar el objetivo específico No. 4, se describe a continuación una arquitectura de CNNs para la segmentación de imágenes que tiene concatenado a su capa el componente de procesamiento óptico el cual permite hacer la simulación

de múltiples capturas en un objetivo, siendo esta un diseño plenamente de clasificación. Esta arquitectura es implementada en Python durante el desarrollo de esta cuarta fase de la metodología. El propósito es obtener el mejor resultado posible y así ver el alcance de esta propuesta.

**Figura 16.** Esquema ilustrativo del método propuesto. Fuente: Autor.

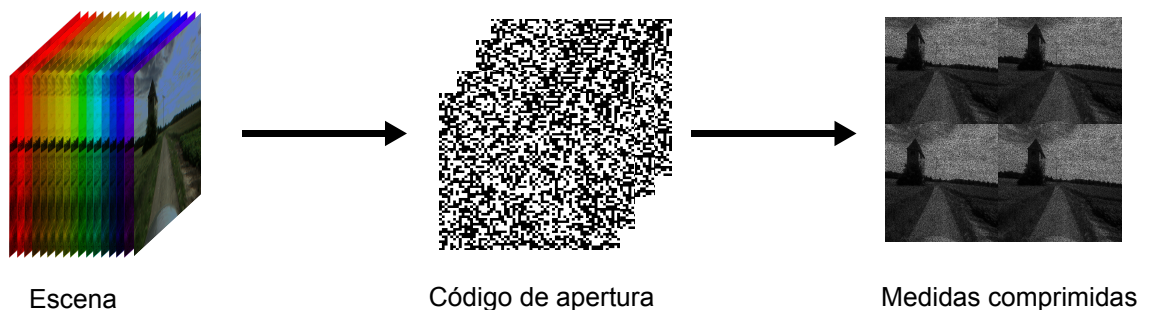


## 3.2. MODELO DE APRENDIZAJE PROFUNDO PARA LA SEGMENTACIÓN DE IMÁGENES

**3.2.1. Arquitectura Implementada** Con el objetivo de desarrollar un modelo de red convolucional para la tarea segmentación de imágenes espectrales, se implementa un modelo basado en una estructura U-Net.

La capa de entrada del modelo implementado recibe imágenes espectrales de tamaño  $M \times N \times L$ , donde  $M$  y  $N$  son las dimensiones espaciales con un tamaño de 128 píxeles cada una y  $L$  es la dimensión espectral con 15 bandas igualmente espaciadas entre 400 y 1000 nanómetros, desde el espectro visible hasta el infrarrojo cercano. Posteriormente, se conecta una capa Lambda personalizada que tiene como finalidad implementar una función de generación de medidas usando  $K$  disparos por cada imagen espectral. Para esto, se utiliza un código de apertura aleatorio binario de tamaño  $M \times N \times L \times K$  con 50 % transmitancia del cuál se toma un bloque  $M \times N \times L$  por cada medida para realizar una multiplicación elemento a elemento con el cubo de datos 3D. Luego, la matriz resultante es sumada a lo largo de su dimensión espectral obteniendo así una medida comprimida de tamaño  $M \times L$ . Este proceso se repite según el número de medidas definidas y se concatenan las medidas compresas en la tercera dimensión obteniendo así un cubo de datos de tamaño  $M \times N \times K$ .

**Figura 17.** Construcción del código de apertura de una escena para 4 microlentes con una transmitancia del 50 %. Fuente: Autor.



Una vez implementada la capa del modelo óptico propuesto, se alimenta una U-Net con la salida de la misma para generar los mapas de segmentación de las imágenes espectrales. Para este caso, además de las capas mencionadas anteriormente, se agregaron 4 bloques de dos capas convolucionales con tamaño  $3 \times 3$  con una capa de agrupación máxima de tamaño  $2 \times 2$ , correspondientes al camino de contracción. Luego, se añade un bloque convolucional de las mismas características para el camino intermedio sin capa de agrupación máxima. Finalmente, para el camino de expansión se agregan 4 bloques de convolución compuestos por una capa de convolución superior de  $2 \times 2$  seguida de dos capas de convolución de  $3 \times 3$ . La capa final consta de una capa de convolución  $1 \times 1$  con el número de mapas de características igual al número de segmentos deseados.

Adicionalmente, para ajustar el modelo neuronal durante su entrenamiento a la tarea de segmentación semántica, se emplea el índice de Jaccard, siendo este índice la *Loss* con la cual vamos a entrenar la red, también conocido como la pérdida de intersección sobre unión. Esta métrica puede expresarse matemáticamente como

$$Jaccard\_Index = \frac{TP}{TP + FN + FP}, \quad (27)$$

donde  $TP$  corresponde a los verdaderos positivos, es decir, los píxeles correctamente clasificados,  $FN$  representa los falsos negativos, es decir, los píxeles que debían ser predichos en el mapa y no lo fueron y  $FP$  se refiere a los falsos positivos, ya que son aquellos píxeles que fueron predichos y no pertenecían a la clase. Esta métrica es ampliamente utilizada para evaluar el rendimiento de los modelos de segmentación de imágenes debido a que es poco común que las predicciones de los modelos se ajusten a la fuente de verdad en las coordenadas exactas, por esto, el índice de Jaccard genera una medida cuantitativa teniendo en cuenta los píxeles que se superponen en gran medida a la fuente de verdad, haciendo útil su

calidad perceptiva e invariancia de escala, lo que mejora la convergencia<sup>77</sup>. Es importante resaltar que el objetivo final de la implementación de esta arquitectura de red y su configuración no es un cubo espectral si no un mapa de segmentación de alta fidelidad de la imagen espectral.

**Objetivo específico No. 5. “Evaluar el desempeño del sistema óptico-computacional y de la red neuronal, realizando la clasificación de varias escenas a partir del uso de base de datos disponibles y/o adquiridas en laboratorio.”**

Con el fin de evaluar el desempeño del sistema óptico-computacional propuesto, tal como se plantea en el objetivo específico No.5, durante el desarrollo de la quinta y última fase del proyecto se implementará la métrica descrita a continuación. Adicionalmente, se registrará la RAM consumida (GB), el almacenamiento (GB), la GPU usada (GB), y el tiempo.

**Exactitud (*Accuracy*).** La ecuación-28 se refiere a una medida de sesgo estadístico, error absoluto o relativo de una cantidad aproximada, que se utiliza específicamente, para describir la proximidad de una medida al valor real<sup>78</sup>. La exactitud en específico, se refiere a las predicciones correctas que ha hecho el modelo y el total de predicciones.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (28)$$

---

<sup>77</sup> DUQUE-ARIAS, David, *et al.* “On power Jaccard losses for semantic segmentation”. En: *VISAPP 2021: 16th International Conference on Computer Vision Theory and Applications*. 2021.

<sup>78</sup> HUANG, Jin y LING, Charles X. “Using AUC and accuracy in evaluating learning algorithms”. En: *IEEE Transactions on knowledge and Data Engineering* 17.3 (2005), págs. 299-310.

## 4. RESULTADOS

### 4.1. Selección de base de datos.

Para ayudar a dar resultado del presente trabajo se seleccionó el conjunto de datos **HyKo1** de **hyko-dataset**, este conjunto de datos contiene 280 archivos o imágenes MAT de MATLAB, con un tamaño espacial de  $256 \times 512$  y con un total de 15 bandas espectrales. La base de datos en mención es de acceso público. Utilizamos la base de datos completa (280 imágenes) distribuidas de la siguiente manera: 36,07 % para prueba y 63,93 % para entrenamiento, tabla-1. Cada una de las imágenes es capturada con cámaras de imágenes de mosaico de instantáneas (SSM)<sup>79</sup> compactas y de bajo costo, que pueden capturar un cubo espectral completo en una sola toma. Los datos cubren longitudes de onda de 400 a 1000 nm que abarcan los rangos espectrales visible e infrarrojo cercano<sup>80</sup>, en este conjunto de datos el autor introduce tres clases de anotación llamadas semántica, material y capacidad de conducción, como se muestra en la figura-18. En la figura-19 se ilustra 6 ejemplos de imágenes del conjunto de datos **HyKo1**, las cuales incluyen la imagen RGB y sus etiquetas en diferentes clases de anotación del conjunto de datos. Para las diferentes pruebas que se realizaron con las imágenes espectrales del conjunto de datos seleccionado, tomamos un total de 15 bandas espectrales a diferentes longitudes de onda de manera equitativa, un ejemplo de dichas bandas está representado en la figura-20.

---

<sup>79</sup> TZAGKARAKIS, George; CHARLE, Wouter y TSAKALIDES, Panagiotis. "Data compression for snapshot mosaic hyperspectral image sensors". En: *2016 24th European Signal Processing Conference (EUSIPCO)*. IEEE. 2016, págs. 1558-1562.

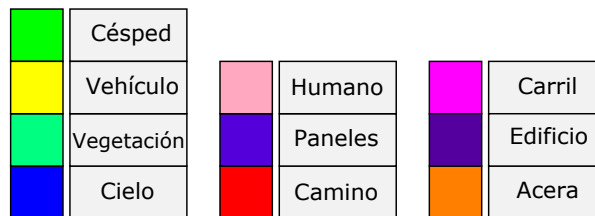
<sup>80</sup> WINKENS, Christian, *et al.* "HyKo: A Spectral Dataset for Scene Understanding". En: *Proceedings of the IEEE International Conference on Computer Vision Workshops*. 2017, págs. 254-261.

**Tabla 1.** Cantidad de imágenes y destino del conjunto de datos (*HyKo1*).

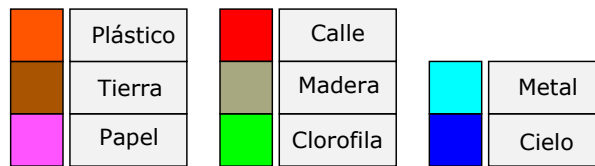
	%	Total de Imágenes
<b>Entrenamiento</b>	63,93	179
<b>Prueba</b>	36,07	101
<b>Total</b>	100	280

Para cada una de las 15 bandas, se eligieron coordenadas aleatorias para sus clases y así poder observar el comportamiento de sus firmas espectrales para dichas características, la figura-23 muestra el mapa de segmentación original con sus clases de acuerdo a la intensidad, a su vez la gráfica de 20 firmas espectrales para cada una de sus clases.

**Figura 18.** Clases de anotación introducidas. Tomada de<sup>81</sup>.



(a) Anotación basada en clases semánticas.



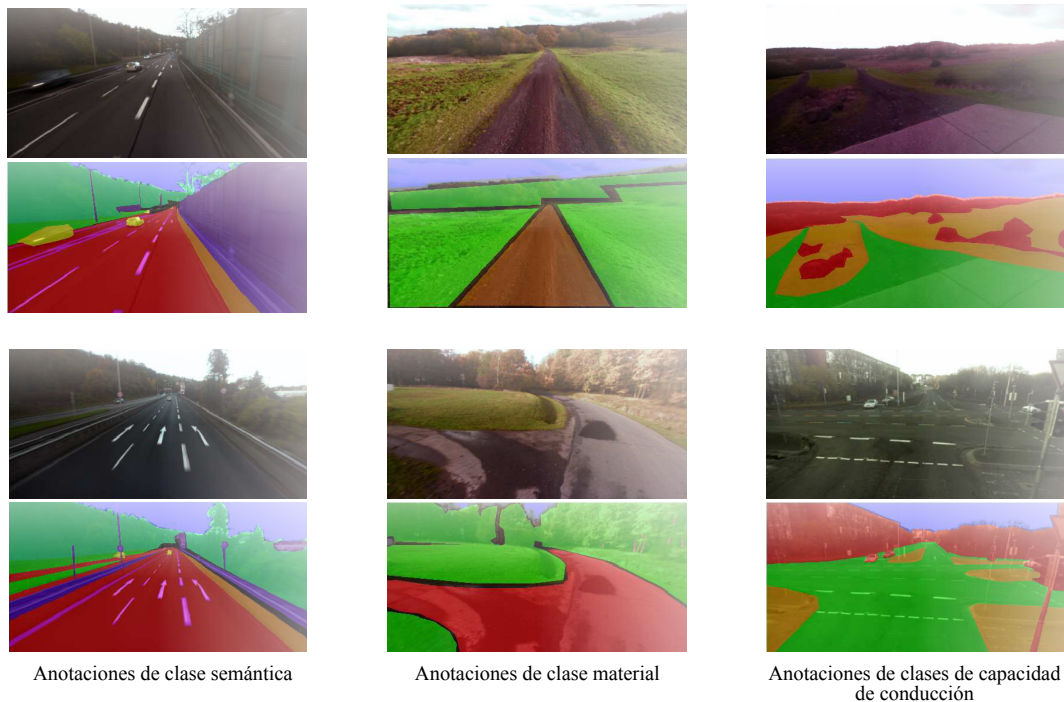
(b) Anotación basada en clases de material.



(c) Anotación basada en capacidad de conducción.

Por otra parte, el conjunto de datos **HyKo1** contiene la imagen preprocesada, la lista de longitudes de onda correspondientes a las bandas espectrales de la imagen preprocesada y sus etiquetas (una anotación por píxel de la imagen). La figura-21,

**Figura 19.** Algunos ejemplos de conjuntos de datos y anotaciones. Adaptada de<sup>82</sup>.

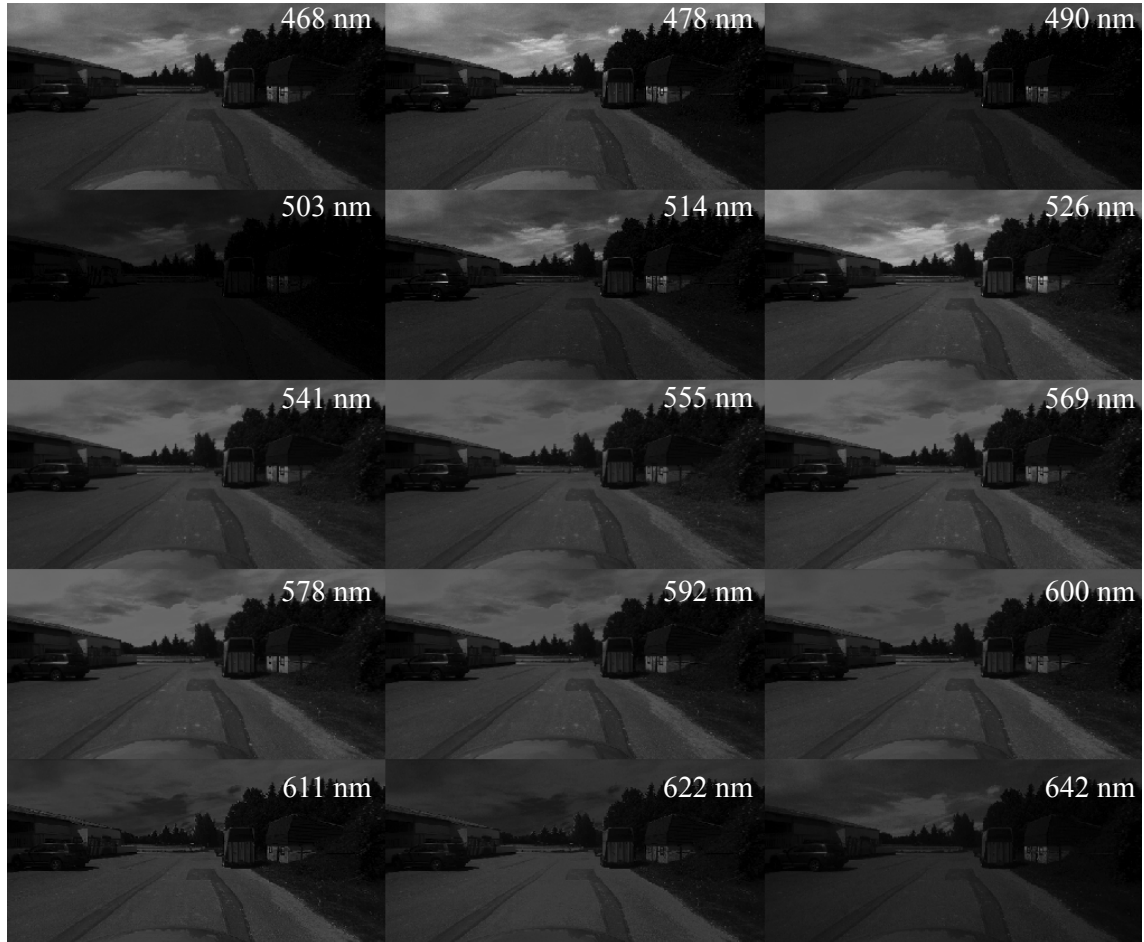


por su parte, presenta una captura de pantalla del sitio web en el cuál se encuentra disponible el conjunto de datos para su libre descarga.

## 4.2. Marco tecnológico

**4.2.1. Software** En términos de Software, el presente trabajo empleó el lenguaje de programación Python y el desarrollo del código, se escribió en Google Colab, también se usó MATLAB en la parte de procesamiento de imagen. GPU accedida virtualmente, proporcionada de manera gratuita por Google Colab. Además, se hizo uso de herramientas básicas de ofimática para el desarrollo de los reportes y presentación final de resultados, tales como licencia de Windows, Microsoft office, y Microsoft Power Point. Todas ellas son patrocinadas por la Universidad Industrial de Santander.

**Figura 20.** Visualización de bandas espectrales para un ejemplo de la base de datos. Fuente: Autor.



**4.2.2. Hardware** El autor del presente trabajo es integrante activo del Grupo de Investigación en diseño de algoritmos y procesamiento de datos multidimensionales (HDSP), adscrito a la Escuela de Ingeniería de Sistemas e Informática de la UIS. Por lo tanto, dispuso de los recursos de trabajo con que cuenta el grupo de investigación para llevar a cabo el desarrollo del proyecto. Específicamente, los equipos a los cuales se accedieron fueron:

- Computador de escritorio HP CORE 17 QUAD Core 3.60 Ghz. disco SSD 500GB Memoria RAM 8 GB DDR4, sistema operativo Windows 10 Profesio-

**Figura 21.** Captura de pantalla de la dirección web en la cual se encuentra disponible el conjunto de datos HyKo1. Disponible en línea: <https://hyko-proxy.uni-koblenz.de/hyko-dataset/HyKo1/>

Index of /hyko-dataset/HyKo1

## Index of /hyko-dataset/HyKo1

	Name	Last modified	Size	Description
⊕	<a href="#">Parent Directory</a>		-	
📁	<a href="#">calib/</a>	2020-07-15 12:52	-	
📁	<a href="#">labels/</a>	2017-10-22 23:10	-	
📁	<a href="#">nir_float_terrclass/</a>	2017-10-17 17:21	-	
📁	<a href="#">raw-data/</a>	2017-10-18 15:10	-	
📄	<a href="#">readme.txt</a>	2017-10-17 15:48	2.1K	
📁	<a href="#">vis_float_terrclass/</a>	2017-10-17 17:21	-	

---

## GENERAL INFORMATION

- Title of Dataset: HyKo1
- Author Information
 

Name: Christian Winkens  
 Institution: University of Koblenz-Landau  
 Address: Universitätsstraße 1, 56070 Koblenz  
 Email: cwinkens@uni-koblenz.de
- Date of data collection 19.06.2016
  - Geographic location of data collection (where was data collected?): Koblenz, Germany

nal, 64 With Windows Pro Licence 3-3-3, Monitor HP V221 21.5-IN.

- Laboratorio de óptica adscrito al grupo de investigación HDSP junto con sus elementos (lentes, cámaras, filtros. etc.)

**Tabla 2.** Desempeño del sistema óptico-computacional.

	<b>Epocas</b>	<b>RAM(GB)</b>	<b>GPU</b>
<b>1 Medida</b>	3000	1.09	NVIDIA-SMI 495.46
<b>4 Medidas</b>	3000	1.3	NVIDIA-SMI 495.46
<b>8 Medidas</b>	3000	1.7	NVIDIA-SMI 495.46

### 4.3. Desempeño de la arquitectura propuesta

En esta sección se analiza el comportamiento del método propuesto en cuanto a la capacidad de mejorar la clasificación de imágenes con determinada cantidad de medidas.

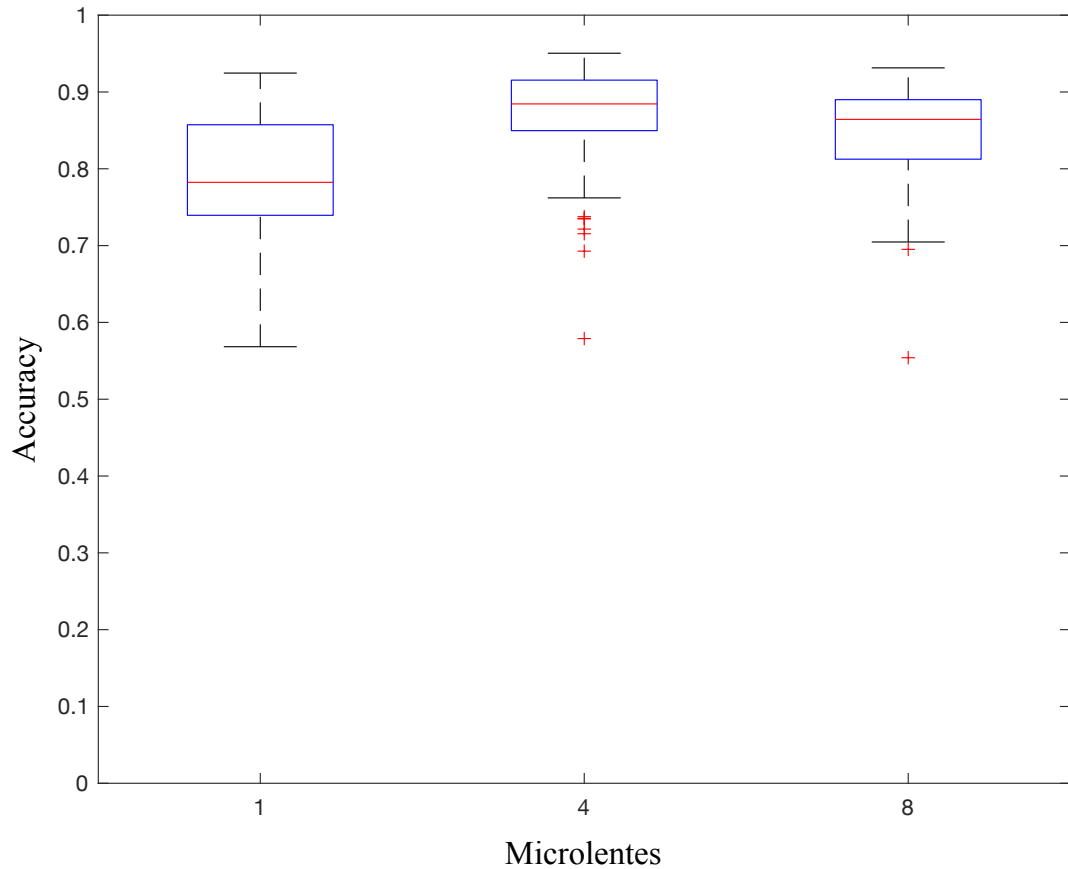
El ajuste final de los hiperparámetros quedó de la siguiente manera: se varió la tasa de aprendizaje de  $1e - 5$ ; el valor de lote o *batch-size* de 4; el número de medidas para hacer la comparación fue de 1,4 y 8, un tamaño de la imagen de  $256 \times 256$  y una tramitancia del 50% . Además, se mantuvo un valor constante respecto al número de épocas siendo esta de 3000.

Podemos evidenciar en figura-16 el funcionamiento de nuestro método propuesto, teniendo como entrada una imagen espectran con  $L$  bandas espectrales y como resultado obtenemos la clasificación de sus clases. Para entender mas a detalle el funcionamiento del método, se expone más a detalle en el numeral 3.2.1 de la sección No. 3. Metodología.

En la figura-23 notamos que cada clase tiene una respuesta espectral única, un ejemplo de lo anterior se puede ver en el rango 600-100 (nm) de la clase cielo, permitiendonos la clasificación de sus características usando el método propuesto.

En las figuras-24, 25 y 26 podemos evidenciar como mejora el etiquetado al aumentar hasta determinado valor el número de medidas, teniendo un aumento en

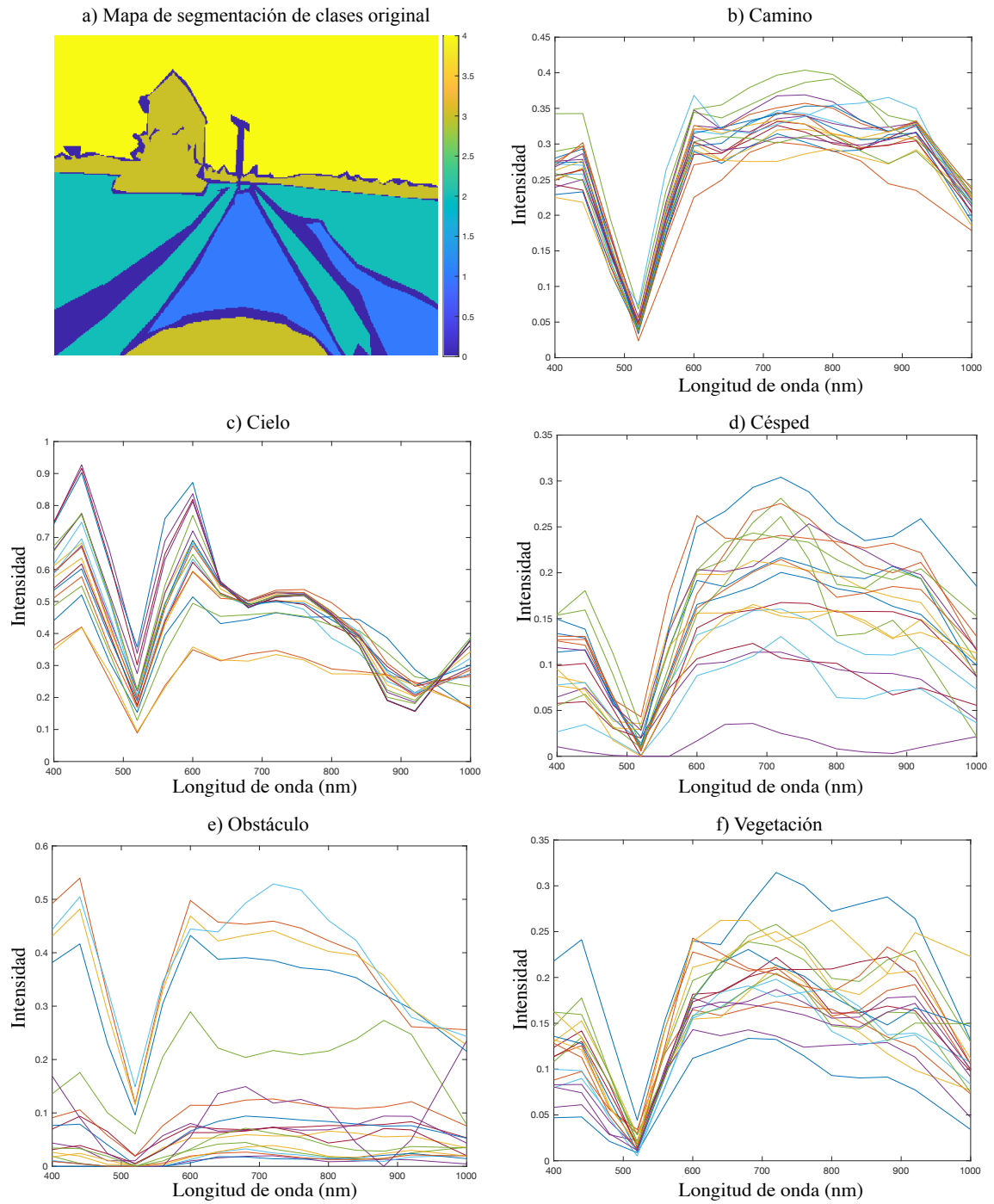
**Figura 22.** Resultado de exactitud de clasificación de imágenes para las diferentes pruebas realizadas con el método propuesto. Fuente: Autor.



la precisión en algunos escenarios, siendo nuestra red capaz de clasificar algunos componentes dependiendo de la escena como por ejemplo, el cielo, la vegetación, carretera, edificios, automóviles, entre otros, en cada una de estas imágenes se puede observar como con diferentes escenarios y múltiples características que varían dependiendo de la escena podemos obtener resultados más precisos que otros dependiendo también del número de medidas que utilizemos para cada imagen de nuestra base de datos. En la tabla-3, vemos el aumento para 4 medidas en la media para la precisión en la clasificación de clases, respecto a 1 sola medida.

Respecto a los tiempos de procesamiento, varía de acuerdo a los recursos compu-

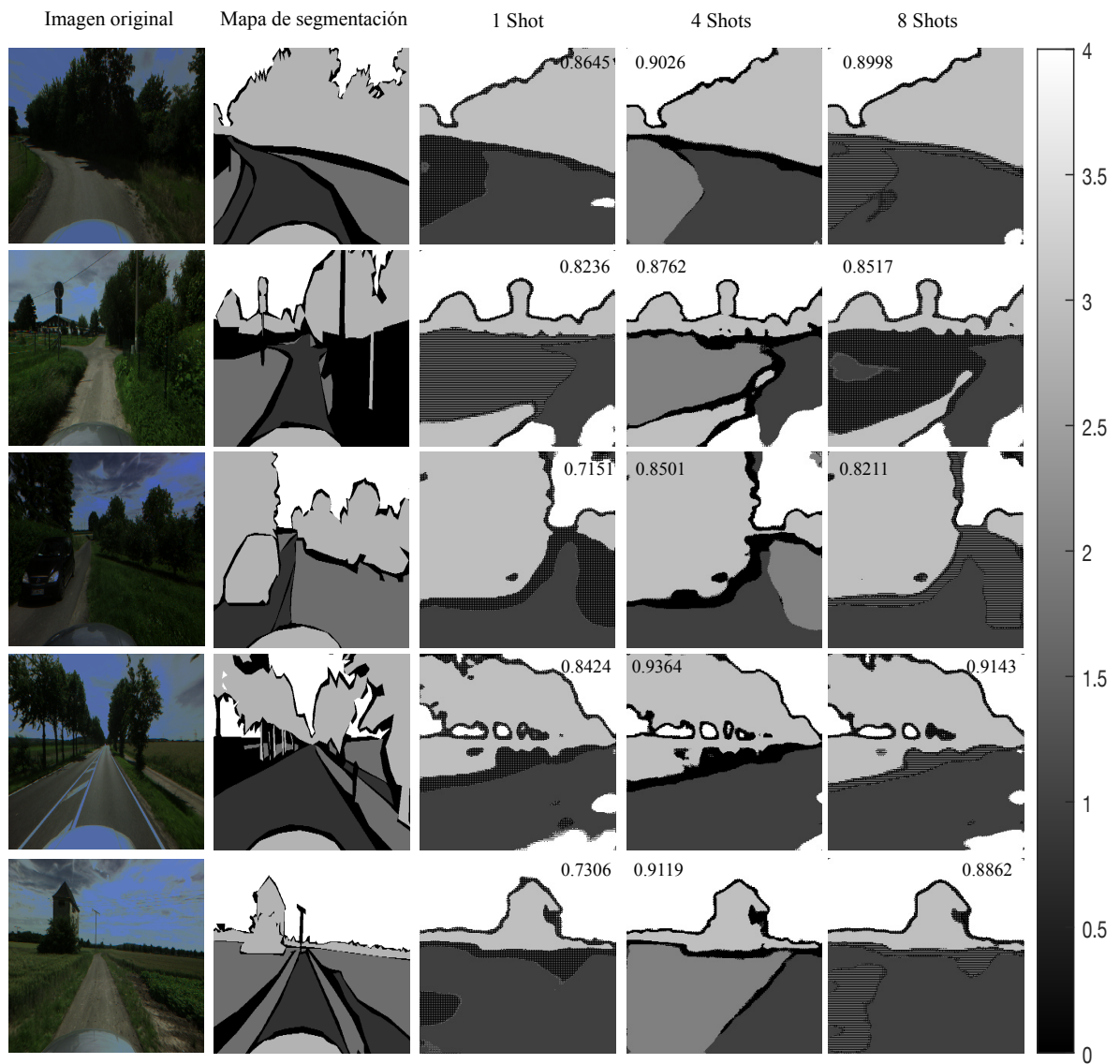
**Figura 23.** Se ilustra 20 ejemplos de firmas espectrales para las cinco clases, (b) camino, (c) cielo, (d) césped, (e) obstáculo, (f) vegetación. Fuente: Autor.



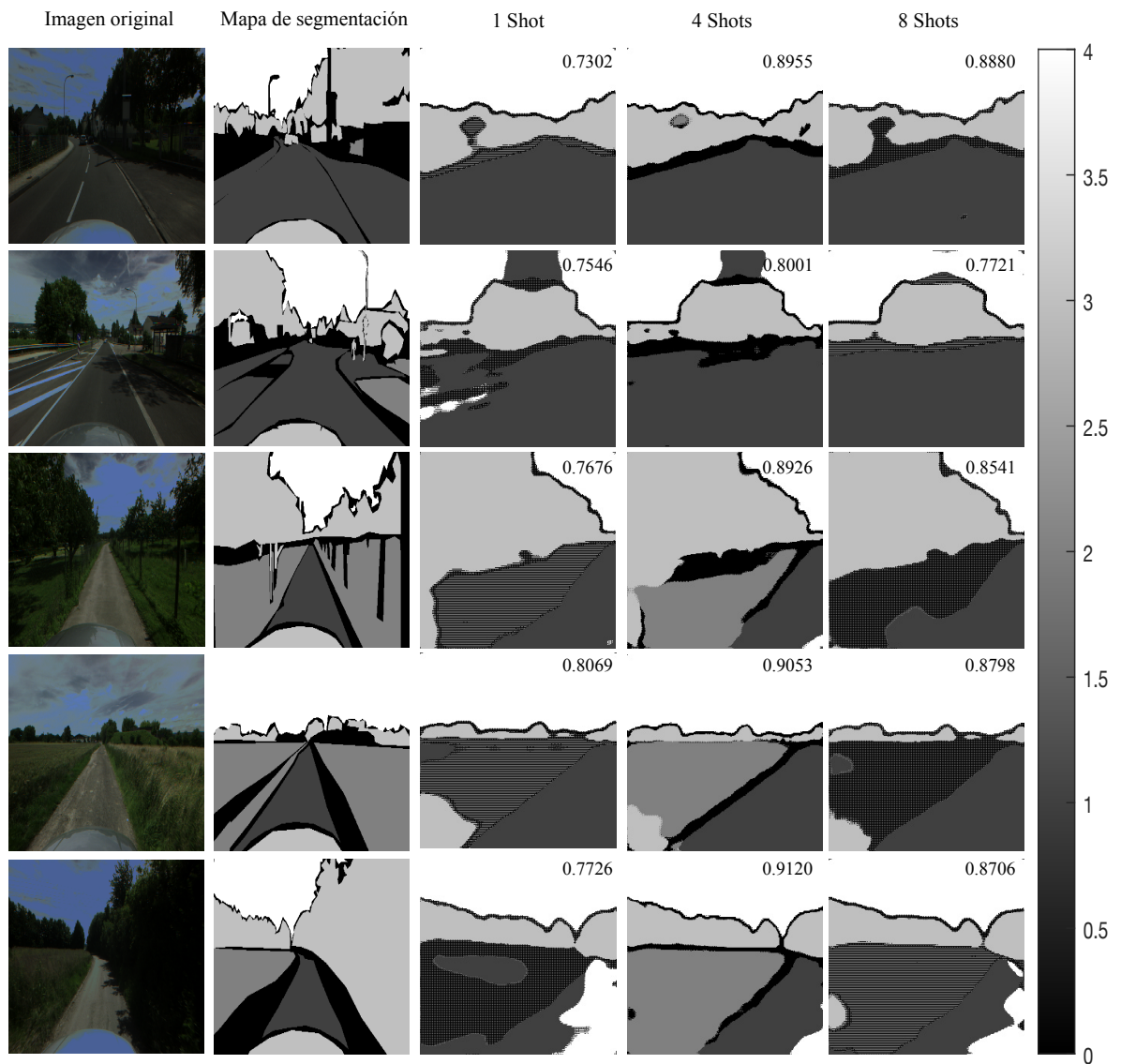
**Tabla 3.** Datos de los resultados de la figura-22.

	<b>Media</b>	<b>Máximo</b>	<b>Mínimo</b>	<b>Desviación Estándar</b>
<b>1 Medida</b>	0.7788	0.9247	0.5702	0.0693
<b>4 Medidas</b>	0.8853	0.9480	0.5779	0.0629
<b>8 Medidas</b>	0.8643	0.9297	0.5541	0.0642

**Figura 24.** Primer ejemplo de visualización de los resultados de clasificación para 1, 4 y 8 medidas, junto con su imagen y mapa de segmentación originales. Fuente: Autor.



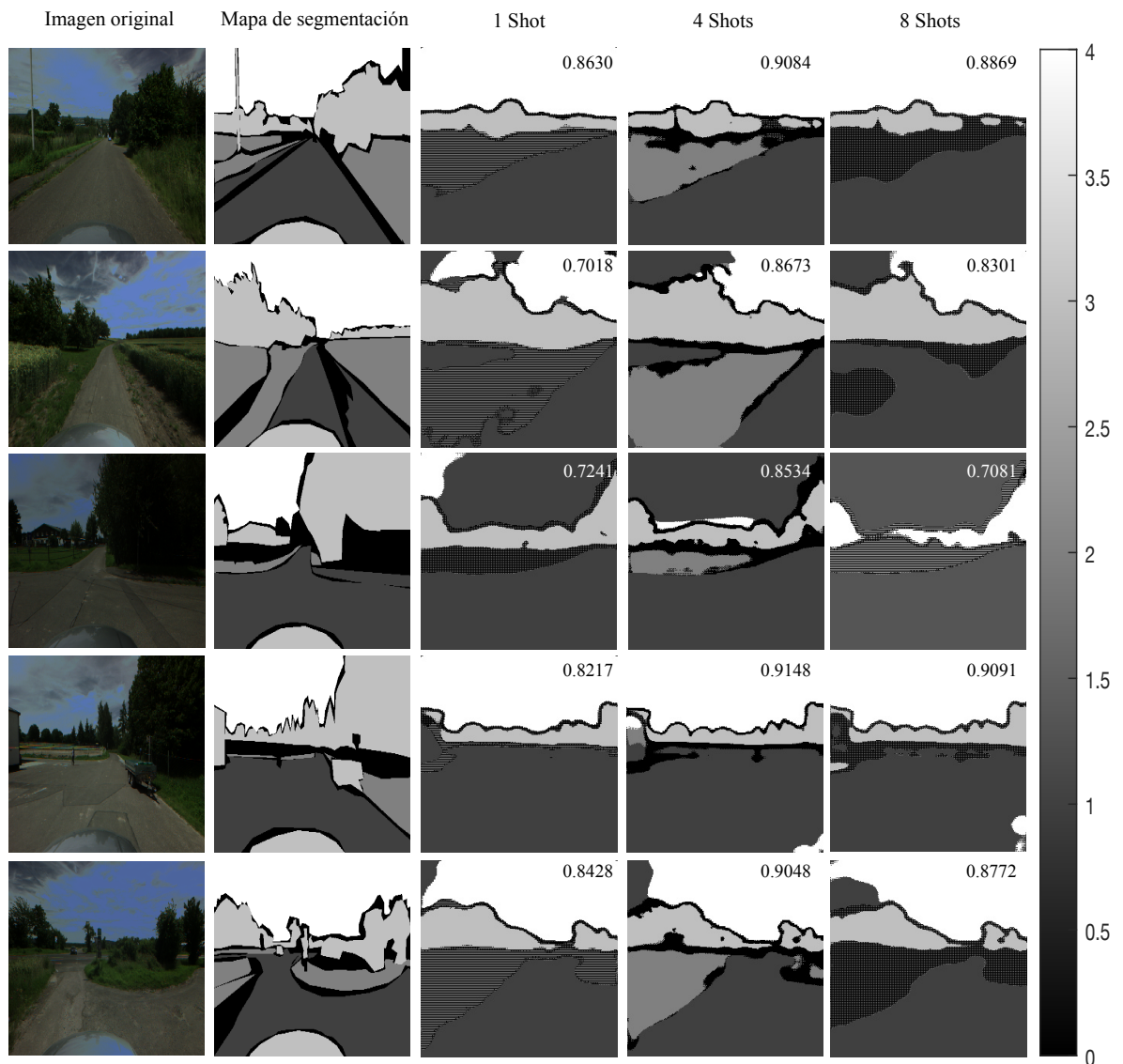
**Figura 25.** Segundo ejemplo de visualización de los resultados de clasificación para 1, 4 y 8 medidas, junto con su imagen y mapa de segmentación originales. Fuente: Autor.



tacionales donde se ejecuten, en la tabla-2 se puede ver algunas características al momento de las pruebas, teniendo en cuenta la magnitud de la base de datos y sus parámetros, los tiempos de ejecución de cada prueba fue relativamente alto.

Contamos con un total de 5 clases con diferentes características en nuestro conjunto

**Figura 26.** Tercer ejemplo de visualización de los resultados de clasificación para 1, 4 y 8 medidas, junto con su imagen y mapa de segmentación originales. Fuente: Autor.



de datos, las cuales en la mayoría de los casos nuestro método fue capaz de clasificar obteniendo una mejor aproximación al ejecutar las pruebas con 4 medidas. Al realizar las pruebas para 8 medidas con el método propuesto, obtenemos una caída respecto a la capacidad de clasificación de clases, dichos datos los podemos ver en la figura-22 y la tabla-3, mostrando también el resultado óptimo para 4 medi-

das, en las figuras-24, 25 y 26 podemos observar los resultados de este decaimiento de clasificación para 8 medidas esto debido porque a medida que aumentamos el número de medidas cae la resolución espacial, por ende concluimos que lo óptimo para este problema son 4 medidas, es decir, un arreglo de microlentes de  $2 \times 2$ .

## 5. CONCLUSIONES

En este trabajo de investigación se desarrolló un método para la clasificación de imágenes espectrales a partir del sistema DD-CASSI que hace uso de un arreglo de lentes. El uso de un arreglo de lentes permite obtener múltiples medidas comprimidas de una escena a partir de una única adquisición. Para este propósito se seleccionó el conjunto de datos HyKo1 de hyko-dataset, el cual contiene 280 imágenes en archivos MAT de MATLAB de diferentes escenarios, permitiendo identificar diferentes clases dependiendo de la escena haciendo posible su uso para la ejecución del método y ayudar con el entrenamiento del modelo neuronal de clasificación. La estrategia propuesta incorpora una arquitectura de red basada en U-Net, además de capas adicionales como lambda y transpuesta, que permiten recibir y procesar un cubo de datos espectral. Las capas de dicha arquitectura permiten identificar y extraer características de los cubos espectrales desde diferentes vistas angulares con ayuda del arreglo de lentes, obteniendo múltiples medidas en una sola toma. Con la finalidad de demostrar el funcionamiento del método propuesto, se simuló arreglos de lentes para tres opciones de captura con 1, 4 y 8 medidas, con cubos de 15 bandas espectrales a diferentes longitudes de onda que hacían su recorrido a través de la arquitectura propuesta, realizando fases de entrenamiento y pruebas dentro de la red neuronal. En el desarrollo de este proyecto se pudo evidenciar que al aumentar el número de lentes empieza a decaer la resolución espacial, resultando en un decaimiento en la exactitud para la clasificación de características. El método propuesto fue evaluado mediante diferentes parámetros que permitieron cuantificar el desempeño en cuanto a la capacidad de identificar y clasificar las características de una escena de acuerdo a sus clases de anotación. Por último y de acuerdo con el decaimiento espacial anteriormente mencionado, se encontró que el número óptimo de lentes en el arreglo de lentes es de cuatro, con el que mostró los mejores

resultados en cuanto a exactitud al momento de clasificar las características.

## BIBLIOGRAFÍA

ARGUELLO, Henry y ARCE, Gonzalo R. "Code aperture optimization for spectrally agile compressive imaging". En: *JOSA A* 28.11 (2011), págs. 2400-2413 (vid. pág. 13).

BALAS, Costas, *et al.* "A novel hyper-spectral imaging apparatus for the non-destructive analysis of objects of artistic and historic value". En: *Journal of Cultural Heritage* 4 (2003), págs. 330-337 (vid. pág. 12).

BHATTA, Suman. "Band to Band Calibration and Relative Gain Analysis of Satellite Sensors Using Deep Convective Clouds". En: (2015) (vid. págs. 19, 20).

BIOUCAS-DIAS, José M, *et al.* "Hyperspectral remote sensing data analysis and future challenges". En: *IEEE Geoscience and remote sensing magazine* 1.2 (2013), págs. 6-36 (vid. pág. 18).

BUSUIOCEANU, Maria, *et al.* "Evaluation of the CASSI-DD hyperspectral compressive sensing imaging system". En: *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XIX*. Vol. 8743. International Society for Optics y Photonics. 2013, pág. 87431V (vid. pág. 46).

CAMACHO-VELASCO, Ariolfo, *et al.* "Aplicaciones y retos del sensado remoto hiperespectral en la geología colombiana". En: *Facultad de Ingeniería* 24.40 (2015), págs. 17-29 (vid. pág. 11).

CANDÈS, Emmanuel J, *et al.* “Compressive sampling”. En: *Proceedings of the international congress of mathematicians*. Vol. 3. Madrid, Spain. 2006, págs. 1433-1452 (vid. págs. 21-23).

CHAN, Tsung-Han, *et al.* “PCANet: A simple deep learning baseline for image classification?” En: *IEEE transactions on image processing* 24.12 (2015), págs. 5017-5032 (vid. pág. 14).

CHUANG, Keh-Shih, *et al.* “Fuzzy c-means clustering with spatial information for image segmentation”. En: *computerized medical imaging and graphics* 30.1 (2006), págs. 9-15 (vid. pág. 12).

CICHY, Radoslaw Martin, *et al.* “Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence”. En: *Scientific reports* 6 (2016), pág. 27755 (vid. pág. 34).

CONSTANTIN, Dragos, *et al.* “Detection of crop properties by means of hyperspectral remote sensing from a micro UAV”. En: *Bornimer Agrartechnische Berichte* 88.CONF (2015), págs. 129-137 (vid. pág. 18).

CORREA, Claudia V; ARGUELLO, Henry y ARCE, Gonzalo R. “Snapshot colored compressive spectral imager”. En: *JOSA A* 32.10 (2015), págs. 1754-1763 (vid. pág. 29).

— “Spatiotemporal blue noise coded aperture design for multi-shot compressive spectral imaging”. En: *JOSA A* 33.12 (2016), págs. 2312-2322 (vid. pág. 13).

DENG, Li; HINTON, Geoffrey y KINGSBURY, Brian. “New types of deep neural network learning for speech recognition and related applications: An overview”. En:

2013 IEEE international conference on acoustics, speech and signal processing. IEEE. 2013, págs. 8599-8603 (vid. pág. 34).

DUQUE-ARIAS, David, *et al.* “On power Jaccard losses for semantic segmentation”. En: *VISAPP 2021: 16th International Conference on Computer Vision Theory and Applications*. 2021 (vid. pág. 54).

EWERLÖF, Maria; LARSSON, Marcus y SALERUD, E Göran. “Spatial and temporal skin blood volume and saturation estimation using a multispectral snapshot imaging camera”. En: *Imaging, Manipulation, and Analysis of Biomolecules, Cells, and Tissues XV*. Vol. 10068. International Society for Optics y Photonics. 2017, pág. 1006814 (vid. pág. 18).

FAUVEL, Mathieu, *et al.* “Advances in spectral-spatial classification of hyperspectral images”. En: *Proceedings of the IEEE* 101.3 (2012), págs. 652-675 (vid. pág. 11).

FOUCART, Simon y RAUHUT, Holger. “An invitation to compressive sensing”. En: *A mathematical introduction to compressive sensing*. Springer, 2013, págs. 1-39 (vid. págs. 22, 23).

GAO, Hao, *et al.* “Multi-energy CT based on a prior rank, intensity and sparsity model (PRISM)”. En: *Inverse problems* 27.11 (2011), pág. 115012 (vid. pág. 13).

GAT, Nahum. “Imaging spectroscopy using tunable filters: a review”. En: *Wavelet Applications VII*. Vol. 4056. International Society for Optics y Photonics. 2000, págs. 50-64 (vid. págs. 19, 20).

GAVHALE, Kiran R y GAWANDE, Ujwalla. “An overview of the research on plant leaves disease detection using image processing techniques”. En: *IOSR Journal of Computer Engineering (IOSR-JCE)* 16.1 (2014), págs. 10-16 (vid. pág. 11).

GAWEHN, Erik; HISS, Jan A y SCHNEIDER, Gisbert. “Deep learning in drug discovery”. En: *Molecular informatics* 35.1 (2016), págs. 3-14 (vid. pág. 35).

GEORGIEV, Todor G y LUMSDAINE, Andrew. “Focused plenoptic camera and rendering”. En: *Journal of electronic imaging* 19.2 (2010), pág. 021106 (vid. pág. 44).

GHAMISI, Pedram, *et al.* “Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art”. En: *IEEE Geoscience and Remote Sensing Magazine* 5.4 (2017), págs. 37-78 (vid. pág. 12).

GRAVES, Alex; MOHAMED, Abdel-rahman e HINTON, Geoffrey. “Speech recognition with deep recurrent neural networks”. En: *2013 IEEE international conference on acoustics, speech and signal processing*. IEEE. 2013, págs. 6645-6649 (vid. págs. 35, 37).

GUPTA, Rajiv y HARTLEY, Richard I. “Linear pushbroom cameras”. En: *IEEE Transactions on pattern analysis and machine intelligence* 19.9 (1997), págs. 963-975 (vid. pág. 19).

GUTIÉRREZ, Mauricio Arango. “Clasificación no supervisada de coberturas Vegetales sobre imágenes digitales de sensores Remotos: “landsat-etm+””. En: *Revista Facultad Nacional de Agronomía Medellín* 58.1 (2005), págs. 2611-2634 (vid. pág. 33).

HUANG, Jin y LING, Charles X. “Using AUC and accuracy in evaluating learning algorithms”. En: *IEEE Transactions on knowledge and Data Engineering* 17.3 (2005), págs. 299-310 (vid. pág. 54).

KÖHLER, Claas Henning. “Airborne imaging spectrometer hypex”. En: *Journal of large-scale research facilities JLSRF* 2.A93 (2016), págs. 1-6 (vid. págs. 19, 20).

KRIZHEVSKY, Alex; SUTSKEVER, Ilya e HINTON, Geoffrey E. "Imagenet classification with deep convolutional neural networks". En: *Advances in neural information processing systems* 25 (2012), págs. 1097-1105 (vid. pág. 36).

KUSSUL, Nataliia, *et al.* "Deep learning classification of land cover and crop types using remote sensing data". En: *IEEE Geoscience and Remote Sensing Letters* 14.5 (2017), págs. 778-782 (vid. pág. 14).

LAPRAY, Pierre-Jean, *et al.* "Multispectral filter arrays: Recent advances and practical implementation". En: *Sensors* 14.11 (2014), págs. 21626-21659 (vid. pág. 21).

LECUN, Yann; BENGIO, Yoshua e HINTON, Geoffrey. "Deep learning". En: *nature* 521.7553 (2015), págs. 436-444 (vid. pág. 35).

LEÓN-LÓPEZ, Kareth M; CARRENO, Laura V Galvis y FUENTES, Henry Arguello. "Temporal colored coded aperture design in compressive spectral video sensing". En: *IEEE Transactions on Image Processing* 28.1 (2018), págs. 253-264 (vid. pág. 13).

LEUNG, Michael KK, *et al.* "Deep learning of the tissue-regulated splicing code". En: *Bioinformatics* 30.12 (2014), págs. i121-i129 (vid. pág. 37).

LI, Shutao, *et al.* "Deep learning for hyperspectral image classification: An overview". En: *IEEE Transactions on Geoscience and Remote Sensing* 57.9 (2019), págs. 6690-6709 (vid. pág. 14).

LIU, Li, *et al.* "Deep learning for generic object detection: A survey". En: *International journal of computer vision* 128.2 (2020), págs. 261-318 (vid. pág. 34).

LOUIZOS, Christos; ULLRICH, Karen y WELLING, Max. “Bayesian compression for deep learning”. En: *Advances in neural information processing systems*. 2017, págs. 3288-3298 (vid. pág. 14).

LU, Pei, *et al.* “Digital image information encryption based on compressive sensing and double random-phase encoding technique”. En: *Optik* 124.16 (2013), págs. 2514-2518 (vid. pág. 22).

MAKANTASIS, Konstantinos, *et al.* “Deep supervised learning for hyperspectral data classification through convolutional neural networks”. En: *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE. 2015, págs. 4959-4962 (vid. pág. 15).

MARQUEZ, Miguel; RUEDA-CHACON, Hoover y ARGUELLO, Henry. “Compressive spectral light field image reconstruction via online tensor representation”. En: *IEEE Transactions on Image Processing* 29 (2020), págs. 3558-3568 (vid. pág. 41).

MATICH, Damián Jorge. “Redes Neuronales: Conceptos básicos y aplicaciones”. En: *Universidad Tecnológica Nacional, México* (2001) (vid. pág. 34).

MONSALVE, Jonathan, *et al.* “Compressive Covariance Matrix Estimation from a Dual-Dispersive Coded Aperture Spectral Imager”. En: *2021 IEEE International Conference on Image Processing (ICIP)*. IEEE. 2021, págs. 2823-2827 (vid. pág. 47).

MORENO, Rodolfo G, *et al.* “Desarrollo de algoritmos para muestreo compresivo aplicado a señales de audio”. En: *Información tecnológica* 25.5 (2014), págs. 137-146 (vid. pág. 21).

NG, Ren, *et al.* “Light field photography with a hand-held plenoptic camera”. Tesis doct. Stanford University, 2005 (vid. pág. 40).

PÉREZ, Aritz; LARRANAGA, Pedro e INZA, Inaki. “Modelos gráficos probabilísticos para la clasificación supervisada empleando la estimación basada en kernels Gausianos esféricos”. En: *III Taller Nacional de Minería de Datos y Aprendizaje* (2005), págs. 125-134 (vid. pág. 34).

POLANS, James, *et al.* “Compressed wavefront sensing”. En: *Optics letters* 39.5 (2014), págs. 1189-1192 (vid. pág. 13).

QU, Lele, *et al.* “Diffraction tomographic ground-penetrating radar multibistatic imaging algorithm with compressive frequency measurements”. En: *IEEE Geoscience and Remote Sensing Letters* 12.10 (2015), págs. 2011-2015 (vid. pág. 13).

RONNEBERGER, O.; P.FISCHER y BROX, T. “U-Net: Convolutional Networks for Biomedical Image Segmentation”. En: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Vol. 9351. LNCS. (available on arXiv:1505.04597 [cs.CV]). Springer, 2015, págs. 234-241 (vid. págs. 38, 39).

RUEDA-CHACON, Hoover F. “Higher order discretization model for coded aperture spectral imaging systems”. Tesis doct. University of Delaware, 2015 (vid. pág. 27).

RUEDA CHACÓN, Hoover Fabian y ARGUELLO FUENTES, Henry. “Spatial super-resolution in coded aperture-based optical compressive hyperspectral imaging systems”. En: *Revista Facultad de Ingeniería Universidad de Antioquia* 67 (2013), págs. 7-18 (vid. págs. 24, 25).

SCHMIDHUBER, Jürgen. “Deep learning in neural networks: An overview”. En: *Neural networks* 61 (2015), págs. 85-117 (vid. pág. 14).

SHAW, Gary A y BURKE, Hsiaohua K. “Spectral imaging for remote sensing”. En: *Lincoln laboratory journal* 14.1 (2003), págs. 3-28 (vid. pág. 17).

THEMELIS, George; YOO, Jung Sun y NTZIACHRISTOS, Vasilis. "Multispectral imaging using multiple-bandpass filters". En: *Optics letters* 33.9 (2008), págs. 1023-1025 (vid. pág. 43).

TOMPSON, Jonathan J, *et al.* "Joint training of a convolutional network and a graphical model for human pose estimation". En: *Advances in neural information processing systems*. 2014, págs. 1799-1807 (vid. pág. 37).

TRIMBY, Liam; BALDYCHEVA, Anna y WRIGHT, C David. "Phase-change bandpass filters for multispectral imaging". En: *Photonic and Phononic Properties of Engineered Nanostructures VIII*. Vol. 10541. International Society for Optics y Photonics. 2018, 105412B (vid. pág. 42).

TZAGKARAKIS, George; CHARLE, Wouter y TSAKALIDES, Panagiotis. "Data compression for snapshot mosaic hyperspectral image sensors". En: *2016 24th European Signal Processing Conference (EUSIPCO)*. IEEE. 2016, págs. 1558-1562 (vid. pág. 55).

VAN NGUYEN, Hien, *et al.* "Kernel dictionary learning". En: *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2012, págs. 2021-2024 (vid. pág. 38).

VIBHUTE, Anup y BODHE, Shrikant K. "Applications of image processing in agriculture: a survey". En: *International Journal of Computer Applications* 52.2 (2012) (vid. pág. 18).

VO-DINH, Tuan; CULLUM, Brian y KASILI, Paul. "Development of a multi-spectral imaging system for medical applications". En: *Journal of Physics D: Applied Physics* 36.14 (2003), pág. 1663 (vid. pág. 12).

WAGADARIKAR, Ashwin, *et al.* “Single disperser design for coded aperture snapshot spectral imaging”. En: *Applied optics* 47.10 (2008), B44-B51 (vid. pág. 27).

WAGADARIKAR, Ashwin A, *et al.* “Video rate spectral imaging using a coded aperture snapshot spectral imager”. En: *Optics express* 17.8 (2009), págs. 6368-6388 (vid. pág. 27).

WAGADARIKAR, Ashwin Ashok. *Compressive spectral and coherence imaging*. Duke University, 2010 (vid. pág. 27).

WINKENS, Christian, *et al.* “HyKo: A Spectral Dataset for Scene Understanding”. En: *Proceedings of the IEEE International Conference on Computer Vision Workshops*. 2017, págs. 254-261 (vid. págs. 55-57).

XIONG, Zhiwei, *et al.* “Snapshot hyperspectral light field imaging”. En: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017, págs. 3270-3278 (vid. pág. 41).

YANG, Zhi, *et al.* “Deep transfer learning for military object recognition under small training set condition”. En: *Neural Computing and Applications* 31.10 (2019), págs. 6469-6478 (vid. pág. 35).

YUAN, Xin, *et al.* “Low-cost compressive sensing for color video and depth”. En: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2014, págs. 3318-3325 (vid. pág. 13).

ZHANG, Naiqian; WANG, Maohua y WANG, Ning. “Precision agriculture—a worldwide overview”. En: *Computers and electronics in agriculture* 36.2-3 (2002), págs. 113-132 (vid. pág. 11).

ZOU, James, *et al.* "A primer on deep learning in genomics". En: *Nature genetics* 51.1 (2019), págs. 12-18 (vid. pág. 35).