

Aprendizaje de un modelo de propagación no lineal para el sistema óptico de adquisición de imágenes espectrales basado en apertura codificada de única captura (CASSI) a través de redes neuronales profundas.

Jessica Paola Escobar Pérez

Trabajo de Grado para Optar al Título de Ingeniera de Sistemas

Director

Ph.D Jorge Luis Bacca Quintero

Codirector

Ph.D Henry Arguello Fuentes

Universidad Industrial de Santander

Facultad de Ingenierías Fisicomecánicas

Escuela de Ingeniería de Sistemas e Informática

Bucaramanga

2022

Dedicatoria

A mis padres, quienes con su amor, apoyo incondicional y esfuerzo me permitieron culminar esta meta. Por enseñarme que con disciplina y dedicación es posible lograr grandes cosas.

A mi hermano, por estar siempre presente, por creer en mí y ofrecerme su cariño.

A Camilo, quien me alentó y apoyó en cada paso. Por darme su amor y motivación durante todo este proceso.

Agradecimientos

A mi director Jorge Bacca por brindarme su acompañamiento en el desarrollo de este proyecto. Por su dedicación, paciencia y significativos aportes que contribuyeron a mi formación profesional.

Al profesor Henry Arguello por sus consejos y enseñanzas y por permitirme formar parte de su grupo de investigación.

A los integrantes del grupo de investigación HDSP que contribuyeron en mi formación académica e investigativa.

Tabla de Contenido

Introducción	11
1 Objetivos	15
2 Imagen Espectral	16
2.1 Muestreo Compresivo	18
2.2 Imagen Espectral Compresiva	18
3 Imágenes Espectrales basadas en Apertura Codificada de Única Captura	20
3.1 Modelo de Orden Superior CASSI	21
4 Algoritmos de Recuperación	25
4.1 Plug and Play - ADMM	26
5 Redes Neuronales para la Representación	29
6 Método Propuesto	31
6.1 Conjunto de Datos	34
6.2 Metodología de Simulación	35
6.3 Red Neuronal Profunda	37
6.4 Adquisición de Datos Reales	39

6.4.1	Elementos Ópticos	39
6.4.2	Medidas Adquiridas	41
7	Resultados	45
7.1	Métricas	45
7.2	Resultados Simulados	47
7.3	Resultados Experimentales	53
8	Conclusiones	57
	Referencias Bibliograficas	58

Lista de Figuras

Figura 1	Representación gráfica de una imagen espectral de ocho bandas y una firma espectral generada por la respuesta de un vóxel.	16
Figura 2	Métodos de adquisición de imágenes espectrales. (A) Escaneo de puntos. (B) Escaneo de líneas. (C) Escaneo de longitud de onda. (D) Única captura. Adaptada de Wang et al. (2017)	17
Figura 3	Arquitectura del sistema óptico CASSI para la adquisición de imágenes espectrales compresivas. Adaptada de Arce et al. (2013)	20
Figura 4	Modelo de integración CASSI para un segmento S de la fuente de entrada en donde un solo vóxel incide en tres píxeles en el detector. Adaptada de Arguello et al. (2013)	22
Figura 5	Modelo de discretización de primer orden y de orden superior. Adaptada de Arguello et al. (2013)	23
Figura 6	Estructura de la matriz $\hat{\mathbf{H}}$ para el modelo tradicional CASSI (a) y para el modelo CASSI de orden superior (b). Adaptada de Arguello et al. (2013)	23
Figura 7	Flujo de trabajo del método propuesto en el cual se aprende la medida corrupta en la fase de entrenamiento y se recupera la imagen espectral en la fase de prueba.	34
Figura 8	Representación del proceso de sensado para un entorno ideal y real.	37

- Figura 9 Arquitectura de red neuronal propuesta basada en una estructura U-Net para el aprendizaje del modelo de propagación no lineal del sistema óptico CASSI. 38
- Figura 10 Estructura del monocromador. Contiene un banco de filtros que se ajustan a las longitudes de onda requeridas, una fuente de luz y dos rendijas que limitan el ancho de banda de la luz emitida. 40
- Figura 11 Montaje experimental del sistema óptico CASSI. 41
- Figura 12 Curva de caracterización del prisma no lineal. 42
- Figura 13 Escenas capturadas en el laboratorio donde se muestra la escena del montaje, la representación RGB de la captura mediante la cámara CCD y las medidas adquiridas a través del sensor CASSI. 43
- Figura 14 Resultados de reconstrucción para tres imágenes espectrales del conjunto de datos ARAD en los que se muestra la imagen original y sus respectivas estimaciones utilizando el PnP tradicional con una matriz de sensado exacta y una calibrada, además del método propuesto. 48
- Figura 15 Respuestas espectrales de tres puntos seleccionados en cada una de las tres escenas tomadas del conjunto de datos ARAD junto a sus reconstrucciones, mostradas en la figura 14. 51

Figura 16 Resultados obtenidos para tres escenas capturadas experimentalmente en el laboratorio usando el sistema CASSI. En la primera columna se encuentra la imagen original. La segunda columna representa la reconstrucción obtenida usando el algoritmo PnP tradicional. La tercer columna muestra las reconstrucciones del método propuesto y la última columna muestra la firma espectral de un punto de la escena.

Resumen

Título: Aprendizaje de un modelo de propagación no lineal para el sistema óptico de adquisición de imágenes espectrales basado en apertura codificada de única captura (CASSI) a través de redes neuronales profundas. *

Autor: Jessica Paola Escobar Pérez **

Palabras Clave: Imagen espectral compresiva, Sistema óptico, Calibración, Redes neuronales profundas

Descripción: Las imágenes espectrales detectan la información espacial de una escena a lo largo de múltiples longitudes de onda; las imágenes multiespectrales tienen una alta resolución espacial y baja resolución espectral, por el contrario, las imágenes hiperespectrales se caracterizan por tener alta resolución espectral y baja resolución espacial. La adquisición de dichas imágenes a través de la imagen espectral compresiva (CSI, por sus siglas en inglés compressive spectral imaging) reduce significativamente la velocidad de detección ya que adquiere la información a partir de una cantidad reducida de muestras. El sistema de adquisición de imágenes espectrales basado en apertura codificada de única captura (CASSI, por sus siglas en inglés coded aperture snapshot spectral imagers) es una de las arquitecturas más representativas de CSI en la cual se asume un modelo de propagación lineal. Sin embargo, al realizar la adquisición se presentan diferentes fenómenos de dispersión óptica que provocan que el modelo de propagación no sea lineal, por lo que se desconoce el modelo real de sensado. Por este motivo, se propone aprender un modelo de propagación no lineal a través de redes neuronales profundas (DNN, por sus siglas en inglés deep neural networks), el cual se usará en algoritmos de recuperación para obtener una mejor calidad en la estimación.

* Trabajo de grado

** Facultad de Ingenierías Fisicomecánicas. Escuela de Ingeniería de Sistemas e Informática. Director: Ph.D Jorge Luis Bacca Quintero.

Abstract

Title: Non-linear propagation model learning for coded aperture snapshot spectral imaging (CASSI) optical system through deep neural networks. *

Author: Jessica Paola Escobar Pérez **

Keywords: Compressive spectral imaging, Optical system, Calibration, Deep neural network

Description: Spectral images detect spatial information in a scene across multiple wavelengths; multispectral images have high spatial resolution and low spectral resolution, by contrast, hyperspectral images are characterized by high spectral resolution and low spatial resolution. Acquiring these images through compressive spectral imaging (CSI) significantly reduces the speed of detection since it acquires the information from a reduced number of samples. The coded aperture snapshot spectral imagers (CASSI) based spectral image acquisition system is one of the most representative architectures of the CSI in which a linear propagation model is assumed. However, when performing the acquisition, different optical dispersion phenomena occur that cause the propagation model to be non-linear, so the real sensing model is unknown. For this reason, it is proposed to learn a model of non-linear propagation through deep neural networks (DNN), which will be used in recovery algorithms to obtain a better quality in the estimation.

* Degree Work

** Facultad de Ingenierías Fisicomecánicas. Escuela de Ingeniería de Sistemas e Informática. Director: Ph.D Jorge Luis Bacca Quintero.

Introducción

Las imágenes espectrales recopilan información espacial a lo largo del espectro electromagnético y varían según su resolución espacio-espectral. Éstas se dividen en dos categorías, las multispectrales que tienen una alta resolución espacial y baja resolución espectral, y las hiperespectrales que se caracterizan por tener alta resolución espectral y baja resolución espacial Garini et al. (2006). Las imágenes espectrales dieron paso a numerosas aplicaciones debido a que estiman características únicas que permiten conocer la composición morfológica y molecular de un objeto Huang et al. (2020), Devassy and George (2021), Salmi et al. (2021).

Existen varias técnicas para la adquisición de este tipo de imágenes, como el escaneo de puntos Vane et al. (1993), de líneas Gupta and Hartley (1997) o de longitudes de onda Gat (2000), los cuales se consideran de baja velocidad ya que escanean sólo una parte de la imagen, lo que conlleva a realizar varias tomas para poder capturar todo el cubo de datos. Por este motivo, el proceso de adquisición de imágenes espectrales es un área de investigación de gran interés en el que se han propuesto diferentes arquitecturas basadas en el principio de imagen espectral compresiva, en donde de la escena se adquiere a partir de una menor cantidad de muestras lo que conduce a una mayor velocidad de detección. Una de las arquitecturas de imagen espectral compresiva más representativa es el sistema de adquisición de imágenes espectrales basado en apertura codificada de única captura (CASSI, por sus siglas en inglés coded aperture snapshot spectral imagers) que permite obtener la información espacio-espectral de la escena en una sola captura Arce et al. (2013).

Al hacer la adquisición de las imágenes mediante esta arquitectura, se asume un modelo

de propagación lineal, sin embargo, en la práctica se presentan diversas aberraciones ópticas Sa-sián (2013) como la distorsión, que hace que las líneas rectas de la escena aparezcan curvas en la imagen, la distancia focal y las superficies de proyección, que pueden producir desenfoque, la reducción de luz que ocasiona un oscurecimiento en las esquinas de la imagen y las aberraciones cromáticas que producen diferentes tipos de efectos en el color a causa de la dispersión del prisma Brunner (1997). Estos factores hacen que el modelo de propagación no sea lineal, razón por la cual se desconoce el modelo exacto de sensado; es por esto que la calibración juega un papel muy importante en este tipo de problemas. Por ejemplo, Kasper et al. (2004) presentan un método de calibración de sistemas de óptica adaptativa para corregir en tiempo real las aberraciones ópticas temporalmente introducidas por la atmósfera turbulenta de la tierra con el objetivo de mejorar la calidad de la imagen estimada. También se han propuesto trabajos en el estado del arte para la calibración de imágenes microscópicas en donde se utilizan cámaras telecéntricas debido a su aumento constante y su pequeña distorsión en la profundidad de campo. Los parámetros precisos de la cámara se obtienen mediante optimización no lineal; sin embargo, la aleatoriedad del algoritmo de optimización sin las restricciones adecuadas hace que los resultados sean inconsistentes con la realidad, es por esto que, Xiao et al. (2022) propusieron un enfoque de optimización en donde los parámetros de la cámara sin distorsiones se resuelven mediante una solución de forma cerrada en conjunto con un algoritmo de optimización no lineal que restringe las coordenadas del objetivo de calibración preciso para refinar los parámetros globales, lo que hace que los resultados de la calibración sean más precisos. De la misma manera, se han propuesto otros trabajos para la calibración usando regularización por eliminación de ruido (RED, por sus siglas en inglés regularization by

denoising) la cual es un marco ampliamente aplicable para resolver problemas inversos mediante el uso de información a priori especificada. Sin embargo, los algoritmos RED existentes requieren un conocimiento exacto del operador de medición que caracteriza el sistema de imágenes, lo que limita su aplicabilidad en problemas en los que el operador de medición tiene incertidumbres paramétricas. Por este motivo, los autores Xie et al. (2020) proponen un nuevo método llamado Calibrated RED (Cal-RED), que permite la calibración conjunta del operador de medición junto con la reconstrucción de la imagen desconocida. Este método se ha validado sobre el problema de reconstrucción de imágenes en tomografía computarizada bajo ángulos de proyección perturbados (Xie et al. (2021)) en el cual ha demostrado una gran efectividad para la calibración y reconstrucción. Recientemente se propuso un trabajo para solucionar el problema de mala calibración del hardware al implementar modelos de imágenes hiperespectrales profundos en sistemas CASSI reales; para esto, Wang et al. (2021) plantearon la calibración de una sola red de reconstrucción a través de la incertidumbre de la máscara de modelado con un tratamiento de aprendizaje bayesiano variacional completo sobre una posible descomposición de máscaras inspirada en observaciones sobre máscaras reales. Con esto, presentaron e implementaron una nueva red de auto ajuste basada en gráficos que continúa con la reconstrucción de las imágenes hiperespectrales y el razonamiento de incertidumbre bajo un marco de optimización de dos niveles para equilibrar la reconstrucción y la estimación de la incertidumbre, teniendo en cuenta la propiedad de hiperparámetro de las máscaras.

El objetivo central de este proyecto es aprender el modelo de propagación no lineal a través de una DNN, simulando como una capa el modelo de propagación de la luz de cada elemento

óptico de la arquitectura CASSI. Las DNN han demostrado ser una herramienta exitosa debido a su capacidad para manejar una gran cantidad de datos y aprender representaciones de alto nivel Albawi et al. (2017). Recientemente, se han destacado por sus aplicaciones en procesos no lineales Zhou et al. (2020), Fukami et al. (2020), Chuang et al. (2018). Sin embargo, para obtener un buen resultado en la estimación del modelo de propagación no lineal es necesario contar con una base de datos para el entrenamiento, por lo que se requieren datos reales en el laboratorio.

Una vez obtenidas las medidas compresas, se utiliza un algoritmo de optimización para recuperar la escena a partir de las mediciones escasamente muestreadas utilizando como entradas las medidas y el modelo de propagación del sistema. Por lo tanto, conocer el modelo exacto de sensado define la calidad de la recuperación. Se han propuesto múltiples algoritmos para la recuperación de imágenes espectrales en el estado del arte, como el algoritmo de proyección de gradiente para reconstrucción de señales escasas Figueiredo et al. (2007), el algoritmo de seguimiento de coincidencia ortogonal Tropp and Gilbert (2007), el umbral duro iterativo Blumensath and Davies (2009), el método de multiplicadores de dirección alterna (ADMM, por sus siglas en inglés alternating direction method of multipliers) y su variante, el Plug and Play (PnP) Zheng et al. (2021), Zhou and Rodrigues (2021), Liu et al. (2021b).

1. Objetivos

Objetivo General

- Realizar un modelamiento no lineal para la arquitectura óptica CASSI por medio de redes neuronales convolucionales profundas para aprender una aproximación del modelo de propagación y realizar la recuperación de imágenes espectrales a partir de medidas comprimidas.

Objetivos Específicos

- Modelar cada elemento óptico de la arquitectura CASSI mediante una red neuronal, identificando los parámetros entrenables del sistema óptico y las implicaciones físicas presentes en un entorno real.
- Realizar la adquisición de una base de datos, compuesta por imágenes espectrales y medidas compresivas de la arquitectura óptica CASSI.
- Entrenar la red neuronal convolucional profunda con la base de datos adquirida en el laboratorio para aprender el modelo de propagación no lineal.
- Validar el modelo de propagación obtenido utilizando algoritmos de recuperación del estado del arte y compararlo con el modelo de propagación obtenido en el laboratorio.

2. Imagen Espectral

Las imágenes espectrales (SI, por sus siglas en inglés spectral imaging) están compuestas por dos dimensiones espaciales (x,y) y una dimensión espectral λ , la cual muestrea las imágenes en múltiples longitudes de onda del espectro electromagnético. Las imágenes hiperespectrales (HS, por sus siglas en inglés hyperspectral imaging) tienen la capacidad de cubrir cientos de longitudes de onda, por otra parte, las multispectrales (MS, por sus siglas en inglés multispectral imaging) comprenden un menor número de bandas espectrales. Por esta razón, las HS poseen una mayor resolución espectral y menor resolución espacial, y las MS una mayor resolución espacial y menor resolución espectral.

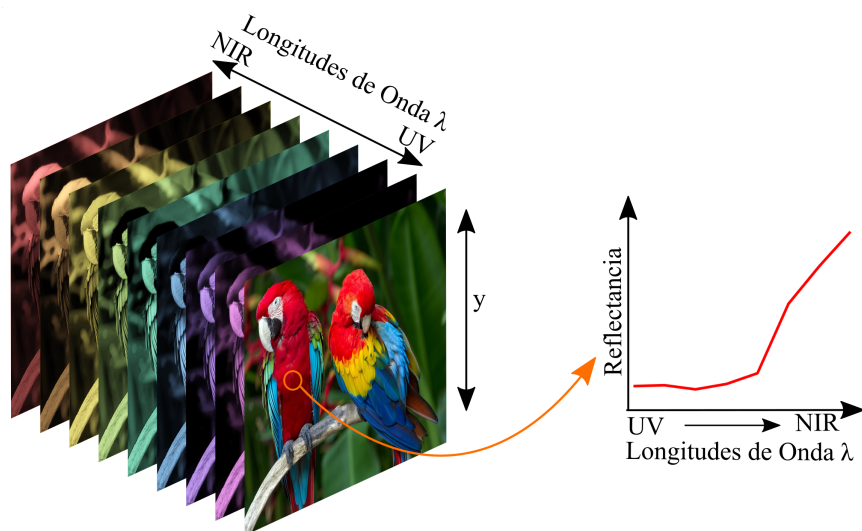


Figura 1. Representación gráfica de una imagen espectral de ocho bandas y una firma espectral generada por la respuesta de un vóxel.

Adquirir imágenes que abarquen un rango espectral desde el ultravioleta hasta el infrarrojo resulta ser ventajoso para múltiples aplicaciones en medicina Li et al. (2021a), agricultura Yu et al.

(2021), arte Vermeulen et al. (2021), entre otras. Existen algunos métodos de adquisición como el escaneo de puntos Vane et al. (1993), el cual consiste en recorrer la escena espacialmente y capturar el espectro de cada punto. Otro método es el escaneo de línea por línea Gupta and Hartley (1997), en donde se recorre la escena en alguna de las dos dimensiones espaciales para capturar el cubo de datos y el escaneo banda por banda Gat (2000) que recorre la escena en cada longitud de onda. Debido a que estos métodos requieren de un escaneo para poder capturar todo el cubo de datos, se presentan tiempos de exposición más prolongados, lo cual ralentiza el proceso de adquisición. Por este motivo, el método más eficiente para capturar este tipo de imágenes es el de única captura Vervandier and Gioux (2013), el cual adquiere la imagen con alta resolución espectral en un instante.

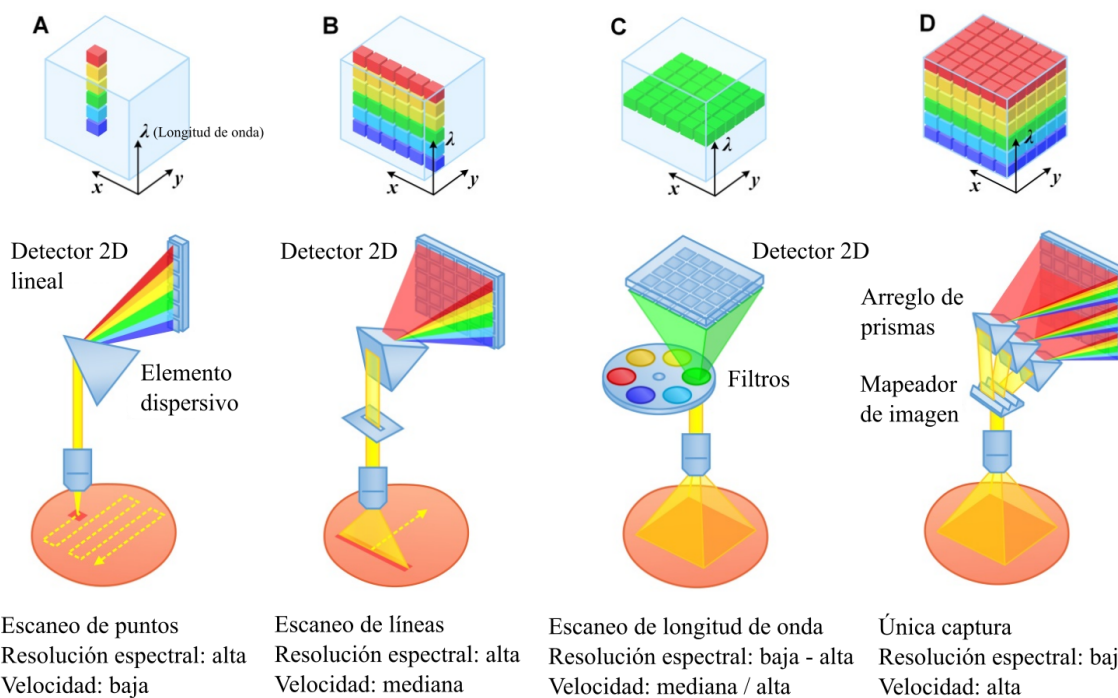


Figura 2. Métodos de adquisición de imágenes espectrales. (A) Escaneo de puntos. (B) Escaneo de líneas. (C) Escaneo de longitud de onda. (D) Única captura. Adaptada de Wang et al. (2017)

2.1. Muestreo Compresivo

El muestreo compresivo (CS, por sus siglas en inglés *compressive sensing*) Baraniuk (2007) es una técnica para adquirir y reconstruir una señal de manera eficiente a partir de muchas menos muestras que las requeridas por sistemas tradicionales basados en el teorema de muestreo de Nyquist-Shannon Jerri (1977), el cual afirma que una señal se puede reconstruir si se muestrea mediante una tasa igual al doble del ancho de banda de la señal. El CS se basa en dos principios, la escasez, que refiere a una representación escasa de la señal y la incoherencia, la cual establece que la matriz de medición y la base de representación escasa deben ser incoherentes entre sí, es decir, ningún elemento de la matriz de medición puede ser obtenido mediante una combinación lineal de elementos de la base de representación escasa (y viceversa).

Esta técnica de compresión reduce el costo computacional ya que las medidas escasas representan un menor tamaño de almacenamiento, lo que a su vez aumenta la transmisión de datos. Se han desarrollado múltiples trabajos basados en CS en áreas como radar Wu et al. (2021), medicina Upadhyaya and Salim (2021), procesamiento de señales Liu et al. (2021a) y holografía de Souza et al. (2021).

2.2. Imagen Espectral Compresiva

La imagen espectral compresiva (CSI, por sus siglas en inglés *compressive spectral imaging*) utiliza la teoría de CS para adquirir la información espacio-espectral de una imagen. El proceso de sensado en CSI se modela como:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w} \quad (1)$$

donde \mathbf{H} representa la matriz de sensado del sistema, $\mathbf{x} \in \mathbb{R}^{MNL}$ es la imagen vectorizada donde M y N son las dimensiones espaciales y L el número de bandas espectrales, \mathbf{y} es el la imagen muestreada y comprimida y \mathbf{w} es el ruido del sensor.

3. Imágenes Espectrales basadas en Apertura Codificada de Única Captura

La arquitectura óptica CASSI está compuesta por un lente objetivo, un conjunto de lentes de retransmisión, una apertura codificada, un elemento dispersivo y un detector de matriz de plano focal (FPA, por sus siglas en inglés focal plane array), tal como se muestra en la Figura 3. En esencia, la escena es capturada por el lente objetivo, seguidamente, se codifica con una apertura codificada que bloquea o deja pasar ciertos píxeles en la escena a lo largo de la dimensión espectral utilizando un dispositivo de microespejos digitales (DMD, por sus siglas en inglés digital micro-mirror device). Luego, atraviesa el prisma, en donde el campo codificado se desplaza en el eje x . Finalmente, el campo codificado y disperso se suma en la dimensión espectral, y se integran los elementos en el detector FPA.

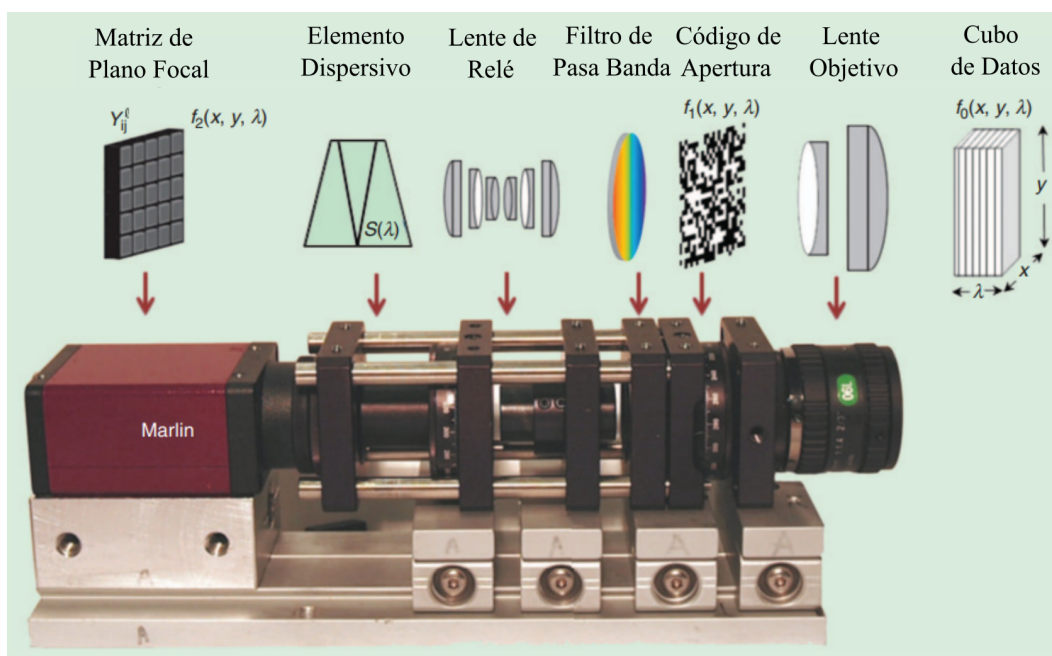


Figura 3. Arquitectura del sistema óptico CASSI para la adquisición de imágenes espectrales compresivas. Adaptada de Arce et al. (2013)

3.1. Modelo de Orden Superior CASSI

Este modelo propone un enfoque para la detección óptica de la arquitectura CASSI, el cual conduce a reconstrucciones de alta calidad con una menor dependencia a procesos de calibración Arguello et al. (2013). La principal contribución de ese trabajo consiste en realizar una discretización más precisa de las señales subyacentes que mitigue la interferencia entre vóxeles, utilizando un conjunto de aperturas codificadas calibradas que disminuyan los efectos de desenfoque y los fenómenos de dispersión que surgen cuando el prisma dispersa un vóxel de la escena. Esto se debe a que los vóxeles inciden en la escena en forma oblicua y no cúbica como supone el modelamiento tradicional del CASSI, el cual se expresa como:

$$g_{mn} = \sum_{k=0}^{L-1} f_{(m-k)jk} T_{(m-k)j} \quad (2)$$

donde f_{ijk} es la discretización de la densidad de la fuente de energía espacio-espectral subyacente y T_{ij} es la apertura codificada discretizada. En un caso ideal, el vóxel cúbico f_{ijk} incidiría en un elemento detector de un solo píxel g_{mn} , sin embargo, cuando el prisma dispersa un único vóxel de una escena, incide en varios elementos detectores a la vez, lo que provoca un efecto borroso que perjudica la reconstrucción. En la figura 4 se observa dicho comportamiento ya que cada vóxel cúbico que se corta por los efectos del prisma, se convierte en un vóxel oblicuo. El vóxel oblicuo se estira a lo largo del eje x , de modo que cuando se proyecta sobre la rejilla del detector, incide sobre varios elementos del detector a la vez. Por lo tanto, varios vóxeles en la fuente afectarán a cada uno de los píxeles en el FPA.

Para dar solución a ese problema, se utiliza un conjunto de aperturas codificadas calibradas $\{T_{ij}^k\}_{k=0}^{L-1}$ en lugar de una T_{ij} , y se utiliza en el proceso de reconstrucción para tener en cuenta el desenfoque óptico no ideal y dispersión no lineal. En ese enfoque se presentan dos inconvenientes. El primero es respecto a la calibración ya que se parte de la suposición de que es un vóxel cubico codificado el que incide en el detector, cuando en realidad es un vóxel oblicuo. El segundo inconveniente se debe a que la calibración de las aperturas codificadas es difícil para los sistemas CASSI de multitrama donde se utiliza una secuencia de aperturas codificadas.

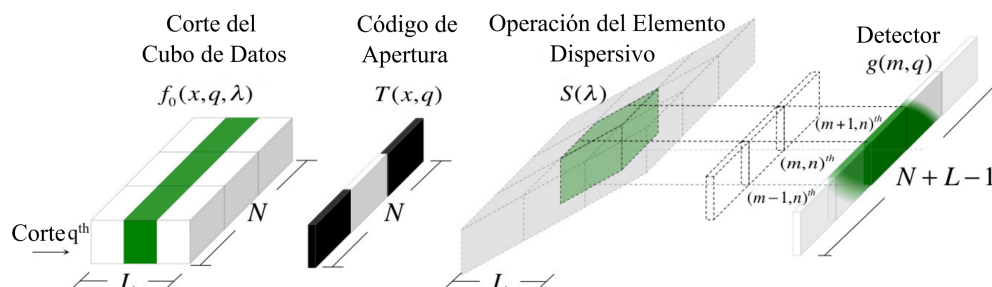


Figura 4. Modelo de integración CASSI para un segmento S de la fuente de entrada en donde un solo vóxel incide en tres píxeles en el detector. Adaptada de Arguello et al. (2013)

En la figura 5 se observa la interferencia entre vóxeles de dos modelos de discretización, el modelo tradicional de primer orden (Figura 5 (a)) y el modelo de orden superior (Figura 5 (b)) en donde el detector FPA captura la energía de varios vóxeles simultáneamente. Se puede observar que en el modelo tradicional se toma solo una región espacio-espectral del cubo de datos integrado en el detector mientras que en el modelo de orden superior cada vóxel en la fuente se puede dividir en tres regiones diferentes, lo que resulta en una mayor precisión. Cabe aclarar que dependiendo de la naturaleza de la dispersión del prisma, un vóxel puede afectar a más de 3 elementos detectores.

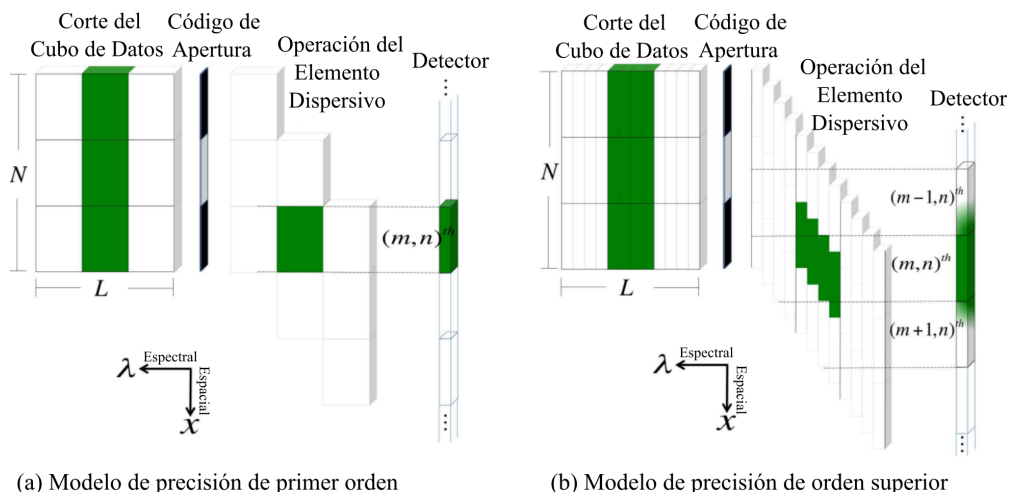


Figura 5. Modelo de discretización de primer orden y de orden superior. Adaptada de Arguello et al. (2013)

La figura 6 muestra la estructura de la matriz \hat{H} para el modelo de discretización de primer orden (Figura 6 (a)) y el modelo de discretización de orden superior (Figura 6 (b)), cuando se utilizan tres disparos FPA para capturar un cubo de datos de dimensiones espaciales $N = M = 6$, dimensión espectral $L = 5$ y número de disparos $K = 3$.

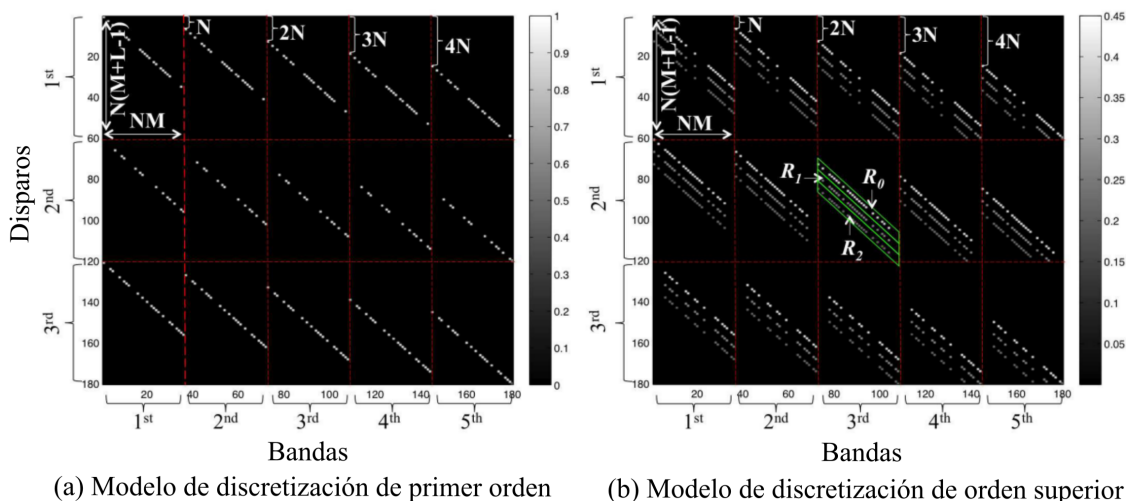


Figura 6. Estructura de la matriz \hat{H} para el modelo tradicional CASSI (a) y para el modelo CASSI de orden superior (b). Adaptada de Arguello et al. (2013)

El modelo de orden superior se puede expresar como:

$$g_{mn} = \sum_{k=0}^{L-1} \sum_{u=0}^d w_{mnku} t_{(m-k-u)n} f_{(m-k-u)nk} \quad (3)$$

donde la distribución de energía se modela con los pesos w en el que u representa la región del modelo, k el rango de las bandas espectrales, t el código de apertura y f el vóxel.

4. Algoritmos de Recuperación

Estos algoritmos tienen como objetivo recuperar el cubo de datos original a partir de un conjunto de medidas sub-muestreadas mediante un proceso de optimización, el cual se plantea a través de un problema inverso. Este problema implica resolver un sistema lineal indeterminado ya que la dimensión de la medida \mathbf{y} , es menor que la dimensión de \mathbf{x} . El problema de optimización para dicho problema está dado por:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda R(\mathbf{x}), \quad (4)$$

donde $\lambda \in \mathbb{R}^+$ es un parámetro de regularización, y $R(\cdot)$ representa un regularizador que aprovecha ciertas propiedades de las imágenes espectrales, como la escasez Beck and Teboulle (2009), bajo rango Gelvez et al. (2017), Bacca et al. (2021), variación total Yuan (2016), entre otros.

Se han propuesto múltiples algoritmos de recuperación. Uno de los más conocidos es el método de dirección alterna del multiplicador ADMM (por sus siglas en inglés, alternating direction method of multiplier) con su variante Plug-and-Play (PnP) Chan et al. (2016). El enfoque principal de este algoritmo es poder resolver un problema de optimización convexa transformándolo en subproblemas fáciles de manejar. El PnP se aprovecha de la estructura del ADMM reemplazando uno de sus pasos de optimización por un eliminador de ruido. Estas características hacen del PnP-ADMM una herramienta poderosa para resolver problemas de restauración de imágenes

Wang and Chan (2017), super resolución Chang et al. (2021), imágenes de resonancia magnética Ahmad et al. (2020), entre otros.

4.1. Plug and Play - ADMM

Siendo $\mathbf{x} \in \mathbb{R}^n$ una imagen ruidosa, es decir, una imagen que contiene ciertas variaciones que alteran su forma ideal, es posible encontrar una imagen subyacente $\hat{\mathbf{x}} \in \mathbb{R}^n$ que realice una estimación de la imagen eliminando el ruido. Esto es equivalente a resolver el problema de optimización planteado en la ecuación (4), el cual es un problema genérico sin restricciones. La idea principal del algoritmo ADMM Boyd et al. (2011) es convertirlo en un problema restringido, tal como se plantea a continuación.

Siendo $f(\mathbf{x}) = \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2$ y $g(\mathbf{x}) = R(\mathbf{x})$,

$$(\hat{\mathbf{x}}, \hat{\mathbf{v}}) = \arg \min_{\mathbf{x}, \mathbf{v}} f(\mathbf{x}) + \lambda g(\mathbf{v}), \quad \text{sujeto a } \mathbf{x} = \mathbf{v}, \quad (5)$$

y considerando su función Lagrangiana aumentada:

$$\mathcal{L}(\mathbf{x}, \mathbf{v}, \mathbf{u}) = f(\mathbf{x}) + \lambda g(\mathbf{v}) + \mathbf{u}^T (\mathbf{x} - \mathbf{v}) + \frac{\rho}{2} \|\mathbf{x} - \mathbf{v}\|^2. \quad (6)$$

La solución de (5) se puede encontrar minimizando el Lagrangiano (6), resolviendo para cada una de las variables, una secuencia de subproblemas más sencillos.

$$\mathbf{x}^{(k+1)} = \arg \min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}) + \frac{\rho}{2} \|\mathbf{x} - \tilde{\mathbf{x}}^{(k)}\|^2, \quad (7)$$

$$\mathbf{v}^{(k+1)} = \underset{\mathbf{v} \in \mathbb{R}^n}{\operatorname{arg\,min}} \quad \lambda g(\mathbf{v}) + \frac{\rho}{2} \|\mathbf{v} - \tilde{\mathbf{v}}^{(k)}\|^2, \quad (8)$$

$$\bar{\mathbf{u}}^{(k+1)} = \bar{\mathbf{u}}^{(k)} + (\mathbf{x}^{(k+1)} - \mathbf{v}^{(k+1)}), \quad (9)$$

donde $\tilde{\mathbf{x}}^{(k)} = \mathbf{v}^{(k)} - \bar{\mathbf{u}}^{(k)}$, $\tilde{\mathbf{v}}^{(k)} = \mathbf{x}^{(k+1)} + \bar{\mathbf{u}}^{(k)}$ y $\bar{\mathbf{u}}^{(k)} = (1/\rho)\mathbf{u}^{(k)}$ es llamado el multiplicador de Lagrange. En este conjunto de ecuaciones, el subproblema (7) es un paso de inversión que minimiza f usando una regularización cuadrática, mientras que el subproblema (8) es un paso de eliminación de ruido que elimina $\tilde{\mathbf{v}}^{(k)}$ usando una función de regularización g , y por último, el subproblema (9) hace referencia al paso de actualización. Los autores Venkatakrishnan et al. (2013) propusieron una variante del algoritmo ADMM, llamado Plug-and-Play ADMM, al sugerir que no es necesario especificar g antes de ejecutarlo, sino que en su lugar, la ecuación (8) se puede reemplazar usando un algoritmo de eliminación de ruido denotado como \mathcal{D}_σ , donde $\sigma = \sqrt{\lambda/\rho}$.

$$\mathbf{v}^{(k+1)} = \mathcal{D}_\sigma \left(\tilde{\mathbf{v}}^{(k)} \right). \quad (10)$$

El principal desafío del enfoque PnP-ADMM es el eliminador de ruido \mathcal{D}_σ , ya que a menudo no es lineal y no tiene expresiones de forma cerrada. Por este motivo, se deben tener en cuenta algunos aspectos. El primero se basa en la convergencia del algoritmo debido a que hay pocos trabajos del estado del arte que aborden el ADMM para problemas no convexos y con eliminadores de ruido. Además, encontrar un ρ óptimo es esencial para la convergencia del algoritmo

ya que éste controla la intensidad de la regularización. Si ρ se establece demasiado grande, la regularización cuadrática en (8) domina y, por lo tanto, el eliminador de ruido es débil. Si ρ se establece demasiado pequeño, el eliminador de ruido es fuerte pero el subproblema (7) se vuelve mal condicionado.

5. Redes Neuronales para la Representación

El aprendizaje profundo LeCun et al. (2015) es una herramienta poderosa que cada vez adquiere más popularidad debido a su capacidad para aprender representaciones de datos en diferentes niveles de abstracción. Las redes neuronales convolucionales (CNN, por sus siglas en inglés convolutional neural networks) Albawi et al. (2017) son una subcategoría del aprendizaje profundo y constan de múltiples capas, como la capa convolucional, la capa completamente conectada y la capa de agrupación; esta última se encarga de reducir los parámetros calculados, lo cual evita el sobreajuste y mejora la robustez del modelo. Por este motivo, una ventaja importante de las CNN respecto a las redes neuronales consiste en que al reducir el número de parámetros se pueden explorar tareas complejas que involucren modelos más grandes. Además, se destacan por la operación de convolución, que permite extraer características a gran escala. También existen las redes neuronales profundas (DNN, por sus siglas en inglés deep neural networks) Liu et al. (2017) en donde el sistema debe procesar capas de datos entre la entrada y la salida para resolver una tarea específica; cuantas más capas tenga que procesar para obtener el resultado, más profunda se considera la red.

Las redes neuronales han sido muy útiles a lo largo del tiempo en diferentes tareas como clasificación, detección, reconstrucción y representación. Por ejemplo, Zhang et al. (2021) plantearon una tarea de clasificación referente al cáncer de mama. Por otra parte, Li et al. (2021b) propusieron un método de detección de rostros a través de un marco totalmente convolucional de doble rama; los autores Chandankhede et al. (2021) realizaron una revisión que muestra una

técnica de reconocimiento de voz única, basada en el análisis planificado, la utilización de la red neuronal y la API de Google empleando las características del habla. Un área muy explorada últimamente en el estado del arte es la reconstrucción de la fase mediante redes neuronales profundas, Işıl et al. (2019) desarrollaron un algoritmo de recuperación de fase que utiliza dos DNN junto con el método híbrido de entrada-salida basado en modelos. En primer lugar, entrenaron una DNN para eliminar los artefactos y lo utilizaron de forma iterativa con el método híbrido de entrada-salida para mejorar las reconstrucciones. Después de esta fase iterativa, entrenaron un segundo DNN para eliminar los artefactos restantes. Por último, se han utilizado las redes neuronales para la tarea de representación, los autores Rippel et al. (2015) demostraron que el dominio espectral proporciona una poderosa representación modelando y entrenando una CNN. Principalmente, propusieron una agrupación espectral que realiza una reducción de la dimensionalidad mientras retiene significativamente más información que otros enfoques de agrupación. Además, dicha representación también permite una nueva forma de regularización estocástica mediante la modificación aleatoria de la resolución con la cual se demuestra la efectividad de la parametrización espectral, lo que resulta en una convergencia más rápida del procedimiento de optimización. Esto demuestra que mediante redes neuronales es posible aprender diferentes modelos de representación, como el modelo exacto de sentido, el cual es el objetivo de este trabajo.

6. Método Propuesto

El uso de las imágenes espectrales abrió un campo diverso de aplicaciones en la ciencia debido a la capacidad de analizar las firmas o respuestas espectrales de los objetos, no solo en el espectro electromagnético visible, sino en otras longitudes de onda como los rayos ultravioleta, el infrarrojo cercano, el infrarrojo lejano, entre otros. Sin embargo, realizar la captura de estas imágenes resulta ser una tarea tediosa, puesto que se deben tener en cuenta múltiples aspectos físicos y ambientales a los que debe estar sujeto el entorno al momento de realizar la captura. Aun teniendo en cuenta estas consideraciones, resulta difícil garantizar la obtención de una imagen completamente ideal, es decir, sin presencia de aberraciones. Esto se debe a que existen diferentes factores en el mismo montaje de adquisición que añaden ruido a la imagen; un ejemplo de esto es la iluminación, la cual no es totalmente uniforme. De la misma manera, algunos lentes poseen pequeños defectos que no son perceptibles a simple vista y que afectan la captura de la información ya que ocasionan que las ondas de luz no lleguen correctamente al detector. Otro factor de ruido se debe al DMD ya que no es perfectamente cuadrado y, en consecuencia, produce una pérdida de píxeles durante la captura. Es por esto que en el laboratorio no se obtienen imágenes ideales, sino imágenes corruptas sujetas a entornos reales y, por ende, el modelo de propagación no es lineal lo cual provoca que se desconozca el modelo exacto de sensado. Por otra parte, el sistema óptico CASSI es de gran interés ya que le ha brindado a la comunidad científica una manera diferente de adquirir la información espacio espectral de la escena con una menor cantidad de medidas. Sin embargo, este sistema tiene una particularidad y es que no se conoce el modelo exacto de sensado

lo cual induce a un problema de no linealidad. Para ello, se puede hacer uso de redes neuronales convolucionales profundas ya que han demostrado que estiman correctamente el aprendizaje de modelos sin supervisión.

El objetivo de este trabajo es aprender el modelo de propagación no lineal del sistema óptico CASSI a través de una DNN, simulando las circunstancias reales expuestas anteriormente. Para esto, se plantea el siguiente problema de optimización para la fase de entrenamiento:

$$\hat{\theta} = \arg \min_{\theta} \|\mathbf{y}_r - \mu_{\theta}(\mathbf{H}\mathbf{x})\|_1, \quad (11)$$

donde $\mathbf{y}_r \in \mathbb{R}^{M(N+L-1)}$ representa las medidas reales realizadas en simulación, μ_{θ} denota la DNN, $\mathbf{H} \in \mathbb{R}^{M(N+L-1) \times MNL}$ representa el conocimiento del sistema como la matriz de sensado usualmente obtenida después de la calibración y $\mathbf{x} \in \mathbb{R}^{MNL}$ es la imagen espectral vectorizada. μ_{θ} pretende aprender los efectos no lineales del sistema real.

Una vez obtenido un modelo de propagación aprendido, se resuelve un problema de optimización tradicional para la recuperación del cubo de datos a través del PnP-ADMM. Partiendo de la ecuación (7) y teniendo en cuenta que $f(\mathbf{x}) = \|\mathbf{y} - \mu_{\theta}(\mathbf{H}\mathbf{x})\|_2^2$, el nuevo problema a resolver es:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{y} - \mu_{\theta}(\mathbf{H}\mathbf{x})\|_2^2 + g(\mathbf{x}), \quad (12)$$

donde $g(\mathbf{x})$ es un regularizador que limita el conjunto solución del problema de optimización.

Agregando una nueva variable auxiliar \mathbf{v} , tenemos que el lagrangiano del nuevo problema queda

como:

$$\mathcal{L}(\mathbf{x}, \mathbf{v}, \mathbf{u}) = \|\mathbf{y} - \mu_{\theta}(\mathbf{H}\mathbf{x})\|_2^2 + \rho \|\mathbf{x} - \mathbf{v} + \mathbf{u}\|_2^2 + g(\mathbf{v}). \quad (13)$$

La ecuación (13) puede resolverse de manera alternada en una serie subproblemas más sencillos:

$$\mathbf{x}^{(k+1)} = \arg \min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{y} - \mu_{\theta}(\mathbf{H}\mathbf{x})\|_2^2 + \frac{\rho}{2} \|\mathbf{x} - \mathbf{v}^{(k)} + \mathbf{u}^{(k)}\|_2^2, \quad (14)$$

$$\mathbf{v}^{(k+1)} = \mathcal{D}_{\sigma} \left(\mathbf{x}^{(k+1)} + \mathbf{u}^{(k)} \right). \quad (15)$$

$$\mathbf{u}^{(k+1)} = \mathbf{u}^{(k)} + (\mathbf{x}^{(k+1)} - \mathbf{v}^{(k+1)}), \quad (16)$$

donde la ecuación (14) representa el paso de inversión, la ecuación (15) representa el operador proximal del regularizador explícito que es reemplazado por un eliminador de ruido, de filtro recursivo Gastal and Oliveira (2011) y la ecuación (16) denota el paso de actualización. Sin embargo, para adaptar el PnP tradicional al método propuesto, se resuelve la ecuación (14) mediante el gradiente descendiente:

$$\mathbf{x}^{(t+1)} = \mathbf{x}^{(t)} - \tau (\mathbf{H}^T \nabla \mu_{\theta}(\mathbf{H}\mathbf{x}) \cdot (\mathbf{y} - \mu_{\theta}(\mathbf{H}\mathbf{x}^{(t)})) + \rho (\mathbf{x}^{(t)} - \mathbf{v}^{(k)} + \mathbf{u}^{(k)})), \quad (17)$$

donde $\mathbf{x}^{(k+1)}$ es el $\mathbf{x}^{(t=T)}$ donde T es un número fijo de iteraciones y $\tau = 0,01$.

En la figura 7 se observa el flujo de trabajo de la metodología propuesta, la cual consiste

en dos fases. En la fase de entrenamiento la imagen espectral $\mathbf{x} \in \mathbb{R}^{M \times N \times L}$ pasa por el proceso de sensado en el que se adquiere la medida ideal $\mathbf{y}_i \in \mathbb{R}^{M \times (N+L-1)}$. Seguidamente, la red neuronal recibe como entrada esta medida y en su salida se obtiene la medida corrupta $\mathbf{y}_r \in \mathbb{R}^{M \times (N+L-1)}$, de esta manera es como la red aprende el modelo de propagación no lineal. En la fase de prueba, la medida obtenida en el laboratorio ingresa a la red neuronal, la cual devuelve la medida corrupta que es recibida por el PnP para realizar la tarea de recuperación de la imagen. Los detalles de cada paso se presentan en las siguientes secciones.

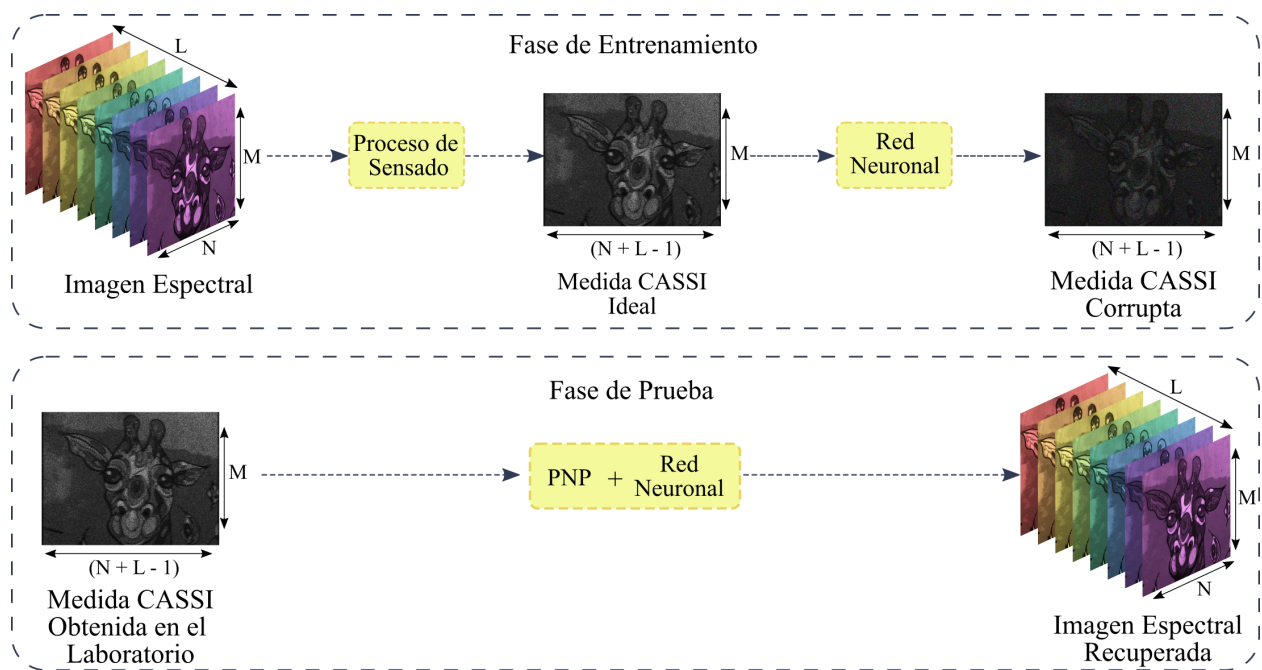


Figura 7. Flujo de trabajo del método propuesto en el cual se aprende la medida corrupta en la fase de entrenamiento y se recupera la imagen espectral en la fase de prueba.

6.1. Conjunto de Datos

Para validar el modelo de propagación, se realizó una simulación mediante una red neuronal profunda, la cual fue entrenada con el conjunto de datos ARAD Arad et al. (2020). Este conjunto

de datos surgió del desafío NTIRE 2020 de reconstrucción espectral y contiene 460 imágenes hiperespectrales en un rango espectral de 400nm a 700nm con 31 bandas espectrales y 482×512 píxeles.

6.2. Metodología de Simulación

Debido a que en el laboratorio se presentan múltiples aberraciones ópticas que corrompen la escena, se aplicó el siguiente método para simular la captura de una medida en un entorno real. Inicialmente se crea un código de apertura aleatorio binario de 512×512 píxeles, representado como $CA_i \in \{0, 1\}^{M \times N}$, siendo M y N las dimensiones espaciales. Seguidamente, se aplica una convolución 2D con un núcleo de filtro pasa bajo (denotado como k), el cual permite el paso de las frecuencias más bajas y atenúa las frecuencias más altas lo cual añade ruido ya que desenfoca la imagen.

$$CA_r = CA_i * k \quad (18)$$

Luego, se modifican las dimensiones espaciales de la imagen espectral representada como $X_i \in \mathbb{R}^{M \times N \times L}$ donde M y N son las dimensiones espaciales con un tamaño de 512 píxeles cada una y L corresponde a la dimensión espectral con 31 bandas igualmente espaciadas en el rango de 400nm a 700nm. $X_{i_\ell} \in \mathbb{R}^{M \times N}$ denota la ℓ -ésima banda del cubo de datos espectral X_i , a la cual se le agrega ruido con una distribución normal gaussiana.

$$X_{r_\ell} = X_{i_\ell} + \mathcal{n}(\mu, \sigma^2), \quad (19)$$

donde μ es la media y σ^2 la desviación estándar.

Siendo CA_i el código de apertura binario, CA_r el código de apertura corrupto, X_{i_ℓ} la ℓ -ésima banda de la imagen espectral y X_{r_ℓ} la ℓ -ésima banda de la imagen espectral corrupta, se calcula \hat{X}_{i_ℓ} y \hat{X}_{r_ℓ} como:

$$\hat{X}_{i_\ell} = X_{i_\ell} \odot CA_i \quad \hat{X}_{r_\ell} = X_{r_\ell} \odot CA_r \quad (20)$$

Posteriormente, se realiza el desplazamiento horizontal a lo largo de las columnas y se suman en la dimensión espectral para obtener una medida de dimensiones $\mathbf{Y} \in \mathbb{R}^{M \times (N+L-1)}$. Para esto, se introduce la función de desplazamiento $\text{shift}_\ell(\cdot) : \mathbb{R}^{M \times N} \rightarrow \mathbb{R}^{M \times (N+L-1)}$.

$$\text{shift}_\ell(\mathbf{X}) := \begin{cases} \mathbf{X}_{(m,n-\ell)}, & \text{si } 1 \leq n - \ell \leq N \\ 0, & \text{de lo contrario.} \end{cases} \quad (21)$$

Finalmente, la salida del proceso de detección se puede expresar matemáticamente como:

$$Y_i = \sum_{\ell=1}^L \text{shift}_\ell(\hat{X}_{i_\ell}) \quad Y_r = \sum_{\ell=1}^L \text{shift}_\ell(\hat{X}_{r_\ell}), \quad (22)$$

donde Y_i es la medida obtenida en entornos ideales y Y_r la medida corrupta que se presenta en entornos reales, tal como se observa en la figura 8. Con esto se cumple el primer objetivo específico de este trabajo que consistía en modelar cada elemento óptico de la arquitectura CASSI mediante una red neuronal, identificando los parámetros entrenables del sistema y las implicaciones físicas presentes en un entorno real.

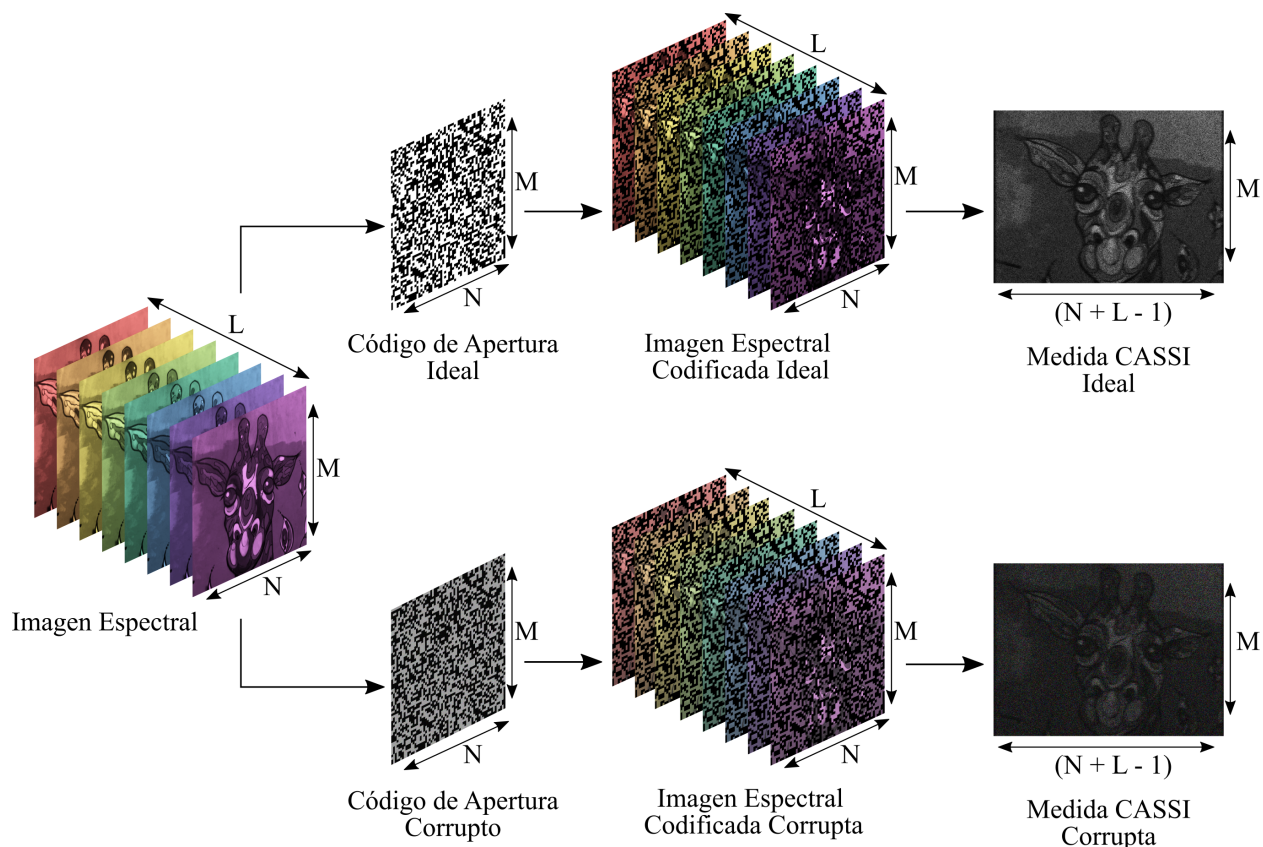


Figura 8. Representación del proceso de sensado para un entorno ideal y real.

6.3. Red Neuronal Profunda

Con el objetivo de aprender el modelo de propagación no lineal del sistema óptico CASSI, se implementó una red neuronal basada en una arquitectura U-Net Ronneberger et al. (2015). Esta red neuronal se propuso como un modelo de segmentación de imágenes biomédicas para aplicaciones en las que la clasificación de la enfermedad no fuera suficiente y se necesitara determinar el área de una anomalía. De esta manera, la salida de la red no corresponde a una única etiqueta que identifique una clase, sino a una clasificación basada en píxeles que agrupen diferentes características de la imagen de entrada. La U-Net tiene una forma de U y consiste en un camino de

contracción compuesto principalmente por convoluciones y capas de submuestreo, seguido de un camino de expansión que contiene capas convolucionales transpuestas y de sobremuestreo, realizando así una segmentación semántica.

Sin embargo, esta red neuronal puede ser usada para aplicaciones diferentes a la medicina como la extracción de carreteras a partir de imágenes de teledetección Chen et al. (2021b), detección de techos de edificios Chen et al. (2021a), estimación de datos sísmicos Park et al. (2019) y microscopía de fluorescencia de dos fotones para la captura de imágenes Lee et al. (2020). Por esto, es posible usar un enfoque de red neuronal U-Net para el aprendizaje del modelo de propagación no lineal del sistema CASSI, permitiendo estimar medidas reales de la arquitectura.

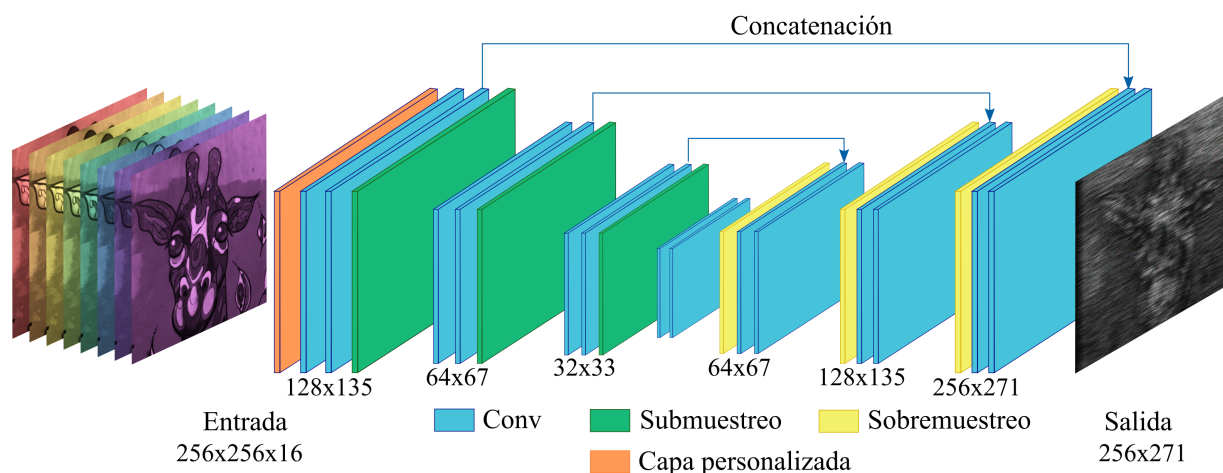


Figura 9. Arquitectura de red neuronal propuesta basada en una estructura U-Net para el aprendizaje del modelo de propagación no lineal del sistema óptico CASSI.

La arquitectura de la red neuronal propuesta se muestra en la figura 9, la cual se compone principalmente por una capa personalizada en donde se realiza el proceso de sensado utilizando la metodología descrita en la sección 6.2, que recibe el cubo de datos espectral y genera la medida

comprimida 2D. Posteriormente, el proceso de submuestreo se realiza a través de tres bloques de convolución compuestos por dos capas convolucionales con tamaños de núcleo 3×3 y función de activación ReLU, las cuales le añaden la no linealidad a la red, seguidas de una capa de agrupación máxima. Después del camino de contracción, se encuentran dos capas convolucionales con la misma configuración que las del bloque convolucional, que permiten obtener un mayor número de mapas de características. Seguidamente, el camino de expansión se compone de tres bloques convolucionales con tamaños de núcleo 2×2 conformados por capas de convolución transpuesta con tamaños de núcleo 1×2 que se utilizaron con el objetivo de ajustar las dimensiones de la red; función de activación ReLU y una capa de sobremuestreo. Por último, la salida de la red da como resultado una imagen bidimensional con el mismo tamaño que la entrada, la cual representa la medida real. Para su configuración se utilizó el optimizador Adam con una tasa de entrenamiento de 1×10^{-3} . De función de pérdida se empleó el error absoluto medio (MAE, por sus siglas en inglés mean absolute error) debido a que pondera todos los errores en la misma escala.

6.4. Adquisición de Datos Reales

6.4.1. Elementos Ópticos. Como se mencionó en la sección 3, la arquitectura CASSI consta de ciertos elementos ópticos que son empleados en el montaje para realizar la captura de las medidas. Principalmente, la escena es iluminada por una fuente de luz, la cual es modulada por un monocromador (ver figura 10) que contiene un conjunto de rejillas difractivas que descomponen la luz blanca de la fuente y selecciona una longitud de onda en el rango espectral de 430nm a 650nm. Dicha luz monocromática se propaga a través de una fibra óptica y su luz reflejada es capturada por el lente objetivo, el cual enfoca la luz en el plano del DMD. Este elemento desem-

peña el papel de apertura codificada ya que consta de una serie de microespejos que se encargan de bloquear o no bloquear el paso de información de píxeles. Posteriormente, los lentes de relé enfocan la luz reflejada del DMD en un divisor de haz, el cual se utiliza con el objetivo de dividir el rayo de luz en dos, debido a que el sistema consta de dos cámaras, una que captura la imagen en escala de grises y otra que captura la medida con la dispersión. De este modo, el divisor de haz refleja el rayo de luz en el plano de imagen del prisma, el cual dispersa la luz filtrada en el sensor de la cámara de dispositivo de carga acoplada (CCD, por sus siglas en inglés charge-coupled device) en escala de grises, que mediante el sensor del CASSI integra la imagen hiperespectral 3D subyacente en el FPA 2D. De la misma manera, el divisor de haz refleja el rayo de luz en la otra cámara CCD, la cual captura la imagen en escala de grises.

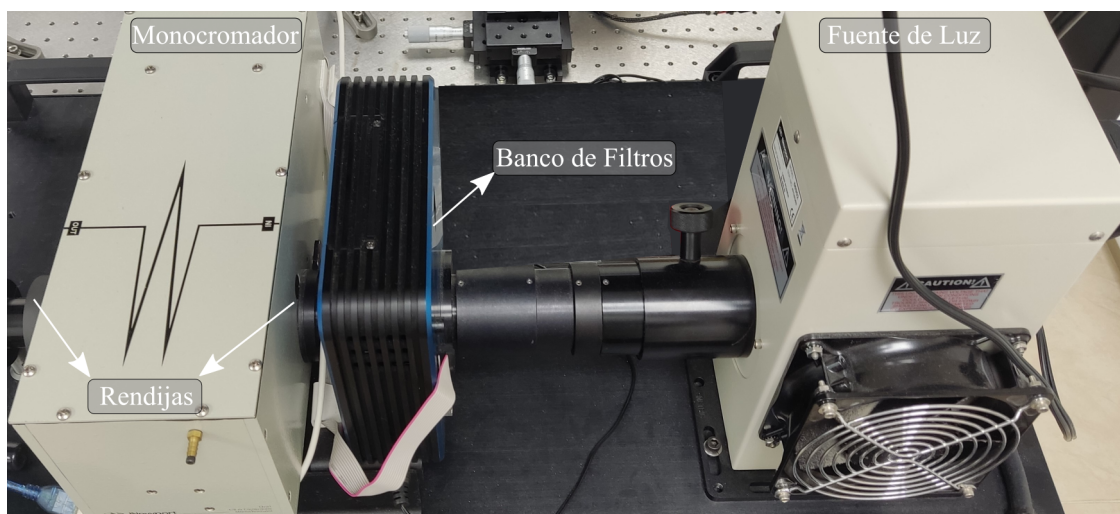


Figura 10. Estructura del monocromador. Contiene un banco de filtros que se ajustan a las longitudes de onda requeridas, una fuente de luz y dos rendijas que limitan el ancho de banda de la luz emitida.

El ensamble y la configuración del sistema (ver figura 11) se realizó en el laboratorio del grupo de investigación, diseño de algoritmos y procesamiento de datos multidimensionales HDSP.

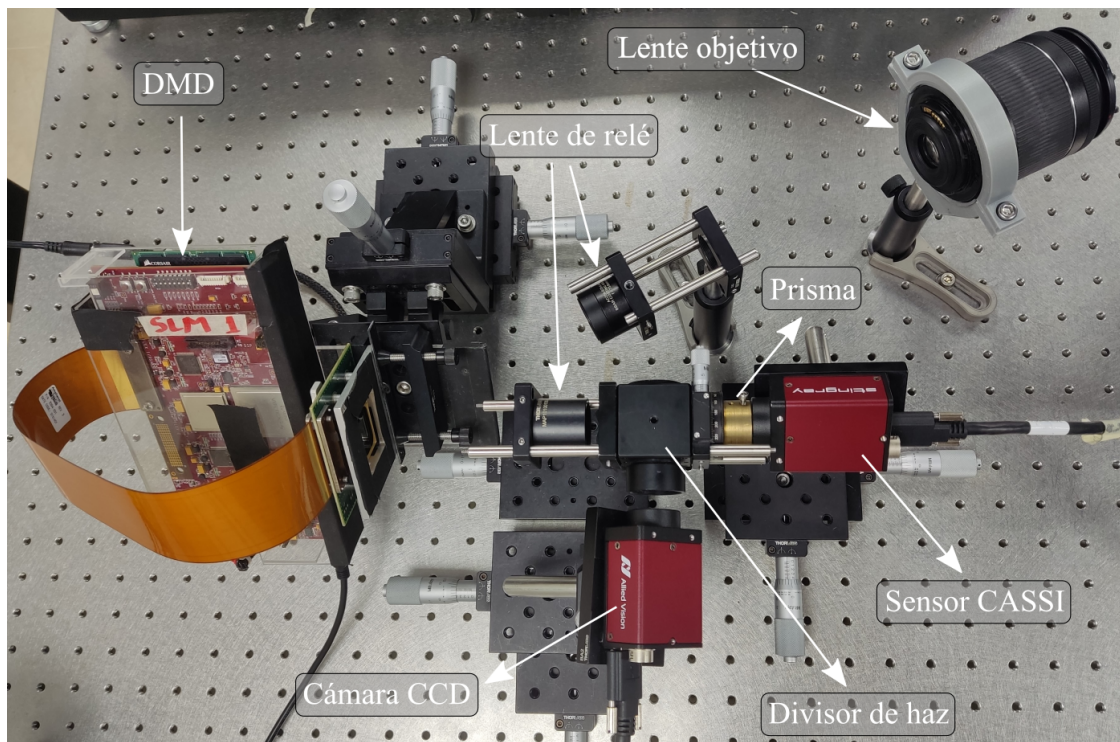


Figura 11. Montaje experimental del sistema óptico CASSI.

6.4.2. Medidas Adquiridas. Con el montaje expuesto en la sección 6.4.1, se adquirieron las medidas experimentales. Inicialmente, se realizó la calibración del prisma usando como código de apertura una línea y como escena una hoja en blanco, recorriendo el rango espectral de 430nm a 650nm. Esto con el objetivo de determinar cuántas bandas se deben usar ya que la curva discretiza la dispersión del prisma con los píxeles del sensor. La curva resultante se muestra en la figura 12 en donde se puede observar que se obtuvieron 27 bandas ya que cada cambio en la curva establece una nueva banda espectral. Sin embargo, en la etapa de posprocesamiento se realizó una decimación, es decir, se disminuyó la frecuencia de muestreo de la señal por un factor de submuestreo de dos. Por este motivo, se recuperaron 13 bandas espectrales.

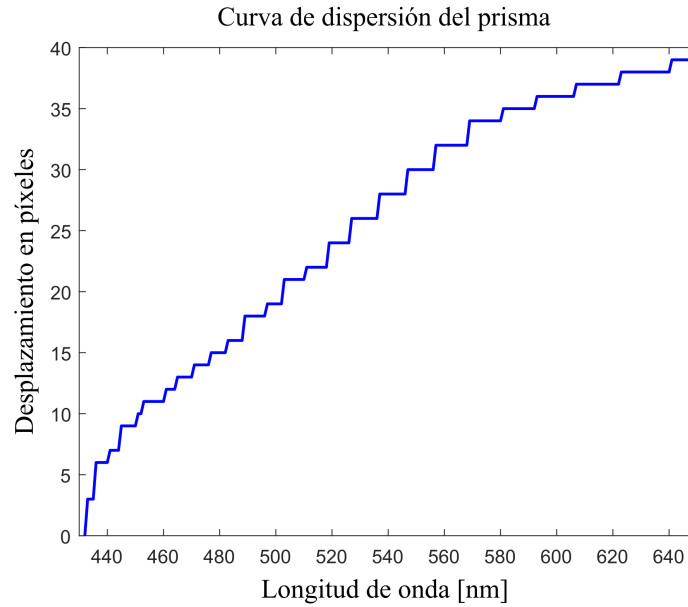


Figura 12. Curva de caracterización del prisma no lineal.

Seguidamente, se generó un código de apertura aleatorio binario de tamaño 128×128 y se empleó un producto Kronecker entre el código de apertura y una matriz de 2×2 llena de unos para replicar la apertura codificada de 128×128 en una de 256×256 . El producto Kronecker se define como:

Si \mathbf{A} es una matriz $m \times n$ y \mathbf{B} es una matriz $p \times q$ entonces el producto Kronecker de $\mathbf{A} \otimes \mathbf{B}$ es la matriz $mp \times nq$.

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{1,1}\mathbf{B} & \cdots & a_{1,n}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{m,1}\mathbf{B} & \cdots & a_{m,n}\mathbf{B} \end{bmatrix} \quad (23)$$

Esto se hizo debido a que el sistema CASSI tiene una dimensión espacial de 128×128 ,

pero el DMD tiene una resolución espacial fija de 256×256 , entonces, cada píxel del código de apertura se replica en un área de 2×2 píxeles en el DMD. De la misma manera, para mantener la relación del tamaño de paso entre el DMD y el sensor, el sistema se calibró para obtener una imagen de resolución espacial 512×512 en la cámara. Una vez obtenidas las medidas, se realizó un ajuste en la dimensión espacial para reducirla a 256×256 . Siendo así, las medidas adquiridas por el sistema óptico CASSI poseen unas dimensiones de $256 \times 256 \times 13$.

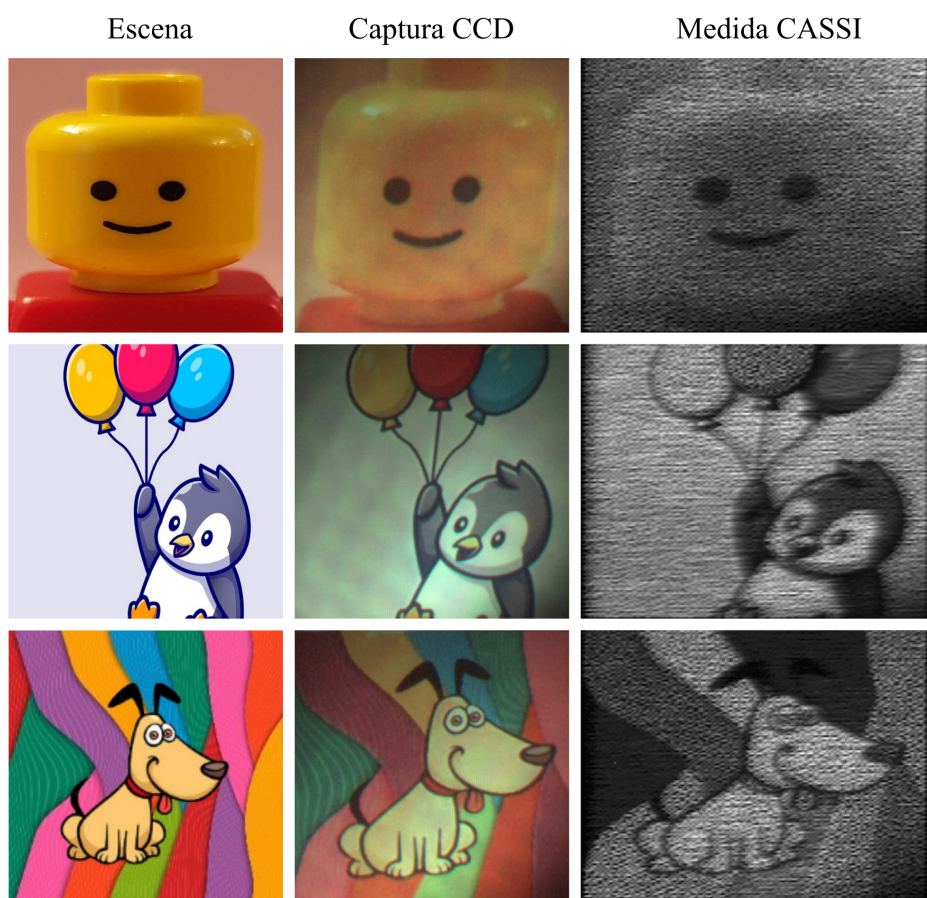


Figura 13. Escenas capturadas en el laboratorio donde se muestra la escena del montaje, la representación RGB de la captura mediante la cámara CCD y las medidas adquiridas a través del sensor CASSI.

En la figura 13 se muestran 3 de las 20 escenas capturadas en el laboratorio. En la primera columna se observa la escena puesta en el montaje. En la segunda columna se encuentra la captura obtenida por la cámara CCD, la cual se obtuvo con dimensiones $256 \times 256 \times 13$. Sin embargo, en la figura se muestran 3 de sus bandas espectrales para obtener su representación RGB. Finalmente, en la tercera columna se encuentra la medida obtenida por el sensor del CASSI, con dimensiones $256 \times 256 \times 13$, pero, al realizar el corrimiento horizontal y la suma de las bandas espectrales, se obtiene la medida de dimensiones 256×268 . Con esto se cumple el segundo objetivo específico de este trabajo el cual consistía en realizar la adquisición de una base de datos, compuesta por imágenes espectrales y medidas compresivas de la arquitectura optica CASSI.

7. Resultados

A continuación, se presentan los resultados simulados y experimentales para la reconstrucción de imágenes espectrales. En la parte simulada se muestran sus respectivas métricas de evaluación que permiten tener un análisis cuantitativo del rendimiento del método en la estimación espacial y espectral del cubo de datos. Es importante resaltar que debido a que en el estado del arte no se encuentra la implementación del algoritmo de recuperación Plug-and-Play en el lenguaje de programación Python, durante el desarrollo de este trabajo se implementó dicho algoritmo con el objetivo de incluir en su marco de trabajo la red neuronal propuesta. Por último, con fines de visualización, la estimación del cubo de datos se pasa por un filtro no lineal de mediana que permite reducir el ruido de la imagen y conservar los bordes.

7.1. Métricas

Para medir la calidad de la reconstrucción se utilizaron las siguientes métricas que evalúan la estimación de la señales, similaridad estructural o espectral, entre otras:

1. PSNR: Se define como la proporción máxima de señal a ruido o PSNR (por sus siglas en inglés, Peak Signal-to-Noise Ratio) y es la relación entre el valor máximo posible de una señal y la potencia del ruido distorsionador que afecta la calidad de su representación. El PSNR se expresa en términos de la escala logarítmica de decibelios (dB), un mayor PSNR

indica una mayor calidad en la estimación de la imagen.

$$PSNR = 20 \cdot \log_{10} \left(\frac{MAX_{\mathbf{f}}}{\sqrt{MSE(\mathbf{f}, \hat{\mathbf{f}})}} \right), \quad (24)$$

donde $\mathbf{f} \in \mathbb{R}^n$ representa la forma vectorial de la imagen original y $\hat{\mathbf{f}} \in \mathbb{R}^n$ la forma vectorial de la estimación. El MSE (por sus siglas en inglés, mean squared error) mide el promedio de los cuadrados de los errores, es decir, la diferencia cuadrática promedio entre los valores estimados y el valor real y se expresa como:

$$MSE(\mathbf{f}, \hat{\mathbf{f}}) = \frac{1}{n} \sum_{i=0}^{n-1} (\mathbf{f}_i - \hat{\mathbf{f}}_i)^2, \quad (25)$$

donde $n = M \times N \times L$.

2. SSIM: La medida del índice de similitud estructural o SSIM (por sus siglas en inglés, structural similarity index measure) se utiliza como una métrica para medir la similitud estructural entre dos imágenes dadas.

$$SSIM(\mathbf{f}, \hat{\mathbf{f}}) = \frac{(2\mu_{\mathbf{f}}\mu_{\hat{\mathbf{f}}} + C_1) + (2\sigma_{\mathbf{f}\hat{\mathbf{f}}} + C_2)}{(\mu_{\mathbf{f}}^2 + \mu_{\hat{\mathbf{f}}}^2 + C_1) (\sigma_{\mathbf{f}}^2 + \sigma_{\hat{\mathbf{f}}}^2 + C_2)}, \quad (26)$$

donde $\mu_{\mathbf{f}}$ y $\mu_{\hat{\mathbf{f}}}$ son el promedio de \mathbf{f} y $\hat{\mathbf{f}}$ respectivamente; $\sigma_{\mathbf{f}}^2$ y $\sigma_{\hat{\mathbf{f}}}^2$ son la varianza de \mathbf{f} y $\hat{\mathbf{f}}$ respectivamente, y por último, $\sigma_{\mathbf{f}\hat{\mathbf{f}}}$ corresponde a la covarianza entre \mathbf{f} y $\hat{\mathbf{f}}$. Además, $C_1 = (k_1P)^2$ y $C_2 = (k_2P)^2$ son dos variables para estabilizar la división con denominador débil, y

P es un rango dinámico de los valores de los píxeles con $k_1 = 0,01$ y $k_2 = 0,03$ por defecto. Esta métrica se mide de 0 a 1, en donde un resultado más cercano al 1 representa un mejor rendimiento.

3. SAM: El mapeador de ángulo espectral o SAM (por sus siglas en inglés, spectral angle mapper) determina la similitud espectral entre la imagen original y su estimación, calculando el ángulo entre los espectros reconstruidos con los espectros reales. Esta métrica se mide en grados, en donde un menor grado indica un mejor resultado en la estimación.

$$SAM = \cos^{-1} \left(\frac{\sum_{l=0}^{L-1} \mathbf{F}_l \hat{\mathbf{F}}_l}{\sqrt{(\sum_{l=0}^{L-1} \mathbf{F}_l)} \sqrt{(\sum_{l=0}^{L-1} \hat{\mathbf{F}}_l)}} \right) \quad (27)$$

7.2. Resultados Simulados

Para los resultados simulados se utilizó el conjunto de datos ARAD, del cual se tomaron 450 imágenes para el entrenamiento y 10 imágenes para la prueba. Además, sus dimensiones se redimensionaron a un tamaño de $256 \times 256 \times 13$ debido a limitaciones de cómputo. El entrenamiento de la red neuronal se realizó durante 20000 épocas, obteniendo como resultado un PSNR de 44 dB. El tiempo de entrenamiento fue de 8 horas en un computador con tarjeta gráfica NVIDIA Quadro K5200 8GB y procesador Intel Xeon E5-1603 128GB.

Con el objetivo de evaluar el rendimiento del método propuesto, se utilizaron 3 imágenes espectrales del conjunto de datos de prueba del ARAD. Para esto, las medidas fueron generadas siguiendo la metodología de simulación descrita en la sección 6.2 y se reconstruyeron mediante en

el algoritmo PnP tradicional y el método propuesto.

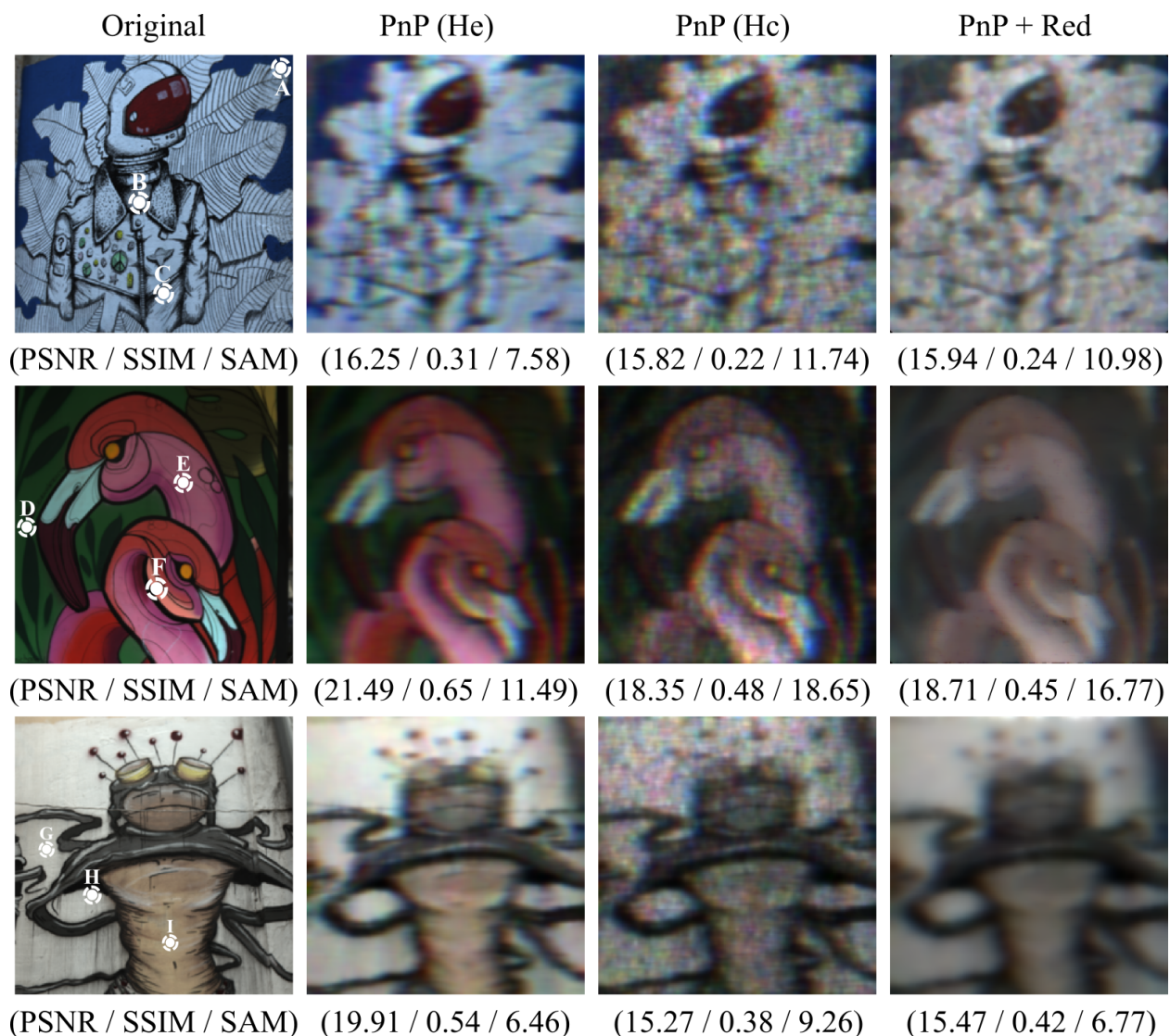


Figura 14. Resultados de reconstrucción para tres imágenes espectrales del conjunto de datos ARAD en los que se muestra la imagen original y sus respectivas estimaciones utilizando el PnP tradicional con una matriz de sensado exacta y una calibrada, además del método propuesto.

En la figura 14 se observa el resultado de la reconstrucción de las escenas de un astronauta, unos flamencos y una ilustración de arte surrealista. En la primera columna se encuentra el cubo de datos original, en la segunda se presenta la reconstrucción que se obtendría en un escenario exacto,

es decir, utilizando una matriz de sensado exacta \mathbf{H}_e , la cual este trabajo supone que no se puede obtener en un entorno real. Sin embargo, representa la estimación a la que se espera llegar. En la tercera columna se encuentra la reconstrucción obtenida con el PnP tradicional utilizando la matriz de sensado calibrada \mathbf{H}_c , es decir, la que se obtiene en el laboratorio. Finalmente, en la última columna se encuentra la reconstrucción del método propuesto, la cual usa un PnP tradicional mas una red neuronal. Todas las imágenes presentadas en la figura 14 se encuentran en su representación RGB.

Para la escena del astronauta puede notarse que, como era de esperarse, la reconstrucción a partir de la matriz de sensado exacta \mathbf{H}_e es la que obtiene un mejor desempeño obteniendo 16.25 dB en PSNR, 0.31 en SSIM y 7.58 en SAM. Sin embargo, es necesario resaltar que este es un escenario ideal y no es lo que se obtiene en la realidad, por esto, en el caso de la matriz de sensado obtenida en el laboratorio, el rendimiento disminuye a 15.82 dB en PSNR, 0.22 en SSIM y 11.74 en SAM. Por último, el método propuesto logra mejorar la calidad de la reconstrucción para el escenario real obteniendo 15.94 dB, 0.24 y 10.98 en PSNR, SSIM y SAM, respectivamente, acercándose más al escenario ideal y mejorando visualmente respecto a la estimación de la matriz de sensado calibrada \mathbf{H}_c .

En la escena de los flamencos se obtienen resultados similares. Por ejemplo, para la reconstrucción en un escenario exacto se alcanzan 21.49 dB en PSNR, 0.65 para SSIM y 11.49 en SAM, con un rendimiento visual bastante notorio. Sin embargo, utilizando la matriz de sensado capturada en el laboratorio se observa una disminución en las métricas alcanzando 18.35 dB en PSNR, 0.48 en SSIM y 18.65 en SAM. Además, se aprecia la presencia de ruido en la imagen. Finalmente, el

método propuesto logra corregir estos defectos e incrementar la calidad de la reconstrucción alcanzando 18.71 dB en PSNR y 16.77 en SAM pero obteniendo una pequeña disminución en SSIM con 0.45.

Para la escena del arte surrealista, se obtienen resultados satisfactorios en el desempeño del método propuesto. En primer lugar, la reconstrucción en el escenario exacto \mathbf{H}_e obtiene en las métricas de PSNR, SSIM y SAM 19.91 dB, 0.54 y 6.46, respectivamente. Luego, para la estimación en el escenario calibrado \mathbf{H}_c la imagen presenta niveles de ruido altos y un rendimiento de reconstrucción mucho menor, llegando a 15.27 dB en PSNR, 0.38 en SSIM y 9.26 en SAM, donde nuevamente se observa ruido en la imagen reconstruida. Por último, el método propuesto logra obtener una imagen con mejor rendimiento y menos ruido, alcanzando 15.47 dB para PSNR, 0.42 en SSIM y una reducción notable en el SAM y cercana al escenario ideal con 6.77.

Con el objetivo de complementar los análisis de los resultados presentados anteriormente, se realiza un análisis espectral de las reconstrucciones obtenidas. Para esto, se seleccionan aleatoriamente 3 puntos espaciales en cada imagen y se obtienen las intensidades en cada una de sus bandas de manera que puede realizarse un estudio de su respuesta espectral. Los resultados se muestran en la figura 15 y los puntos escogidos se denotan en cada escena de la figura 14 comparando las firmas espectrales originales y las reconstrucciones del PnP tradicional y el método propuesto.

Para los puntos A, B y C correspondientes a la escena del astronauta, se puede observar que en A la respuesta espectral de un escenario ideal \mathbf{H}_e obtiene el mejor SAM con 4.29. Sin embargo, como se ha mencionado anteriormente, no es posible llegar a este entorno ideal. Por esto, es nece-

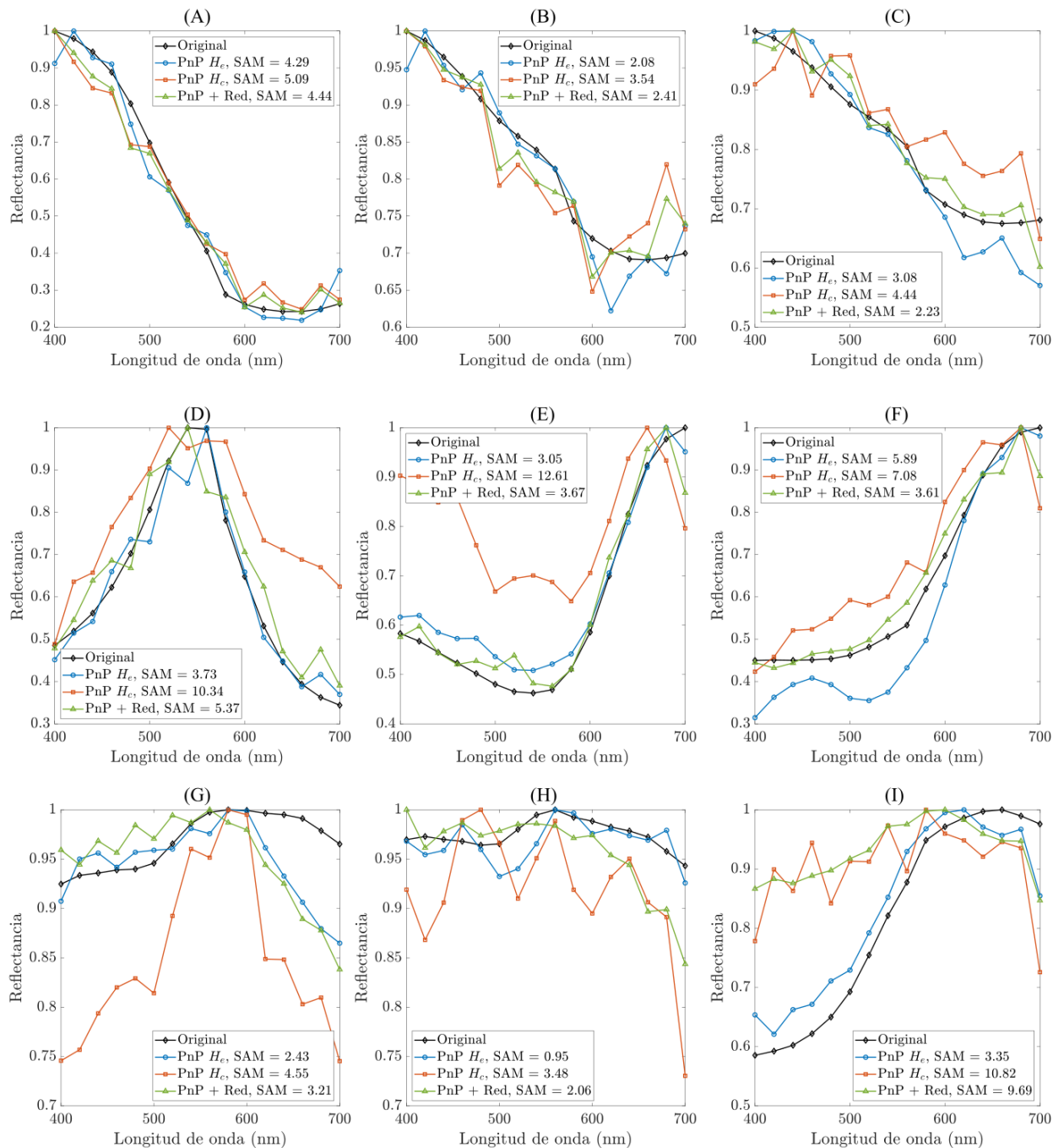


Figura 15. Respuestas espectrales de tres puntos seleccionados en cada una de las tres escenas tomadas del conjunto de datos ARAD junto a sus reconstrucciones, mostradas en la figura 14.

sario resaltar que el método propuesto se acerca considerablemente a la respuesta espectral de la firma original y del escenario ideal con un SAM de 4.44, mientras que la reconstrucción obtenida a partir de una matriz de sensado calibrada \mathbf{H}_c obtiene un SAM mayor, de 5.09, lo cual indica que la imagen original y su estimación tiene menor similitud. Para el punto B se obtienen resultados similares. La reconstrucción de la matriz exacta logra un SAM de 2.08, el método propuesto alcanza una buena estimación con un SAM de 2.41 y la firma obtenida a partir de la matriz calibrada es la que consigue el valor de SAM más alto de 3.54. Finalmente, para el punto C es interesante notar que el método propuesto obtiene un mejor resultado espectral que la matriz de sensado exacta, con 2.23 de SAM frente a 3.08 con \mathbf{H}_e .

En los puntos D, E y F que corresponden a la escena de los flamencos se pueden observar resultados destacables. Por ejemplo, en el punto D mientras que la matriz calibrada \mathbf{H}_c obtiene un SAM de 10.34 y una gráfica bastante alejada de la firma original, el método propuesto alcanza un SAM mucho menor, con 5.37 y una respuesta espectral mucho más cercana a la firma original y a la matriz exacta \mathbf{H}_e que obtiene 3.73 de SAM. A su vez, el punto E presenta aún mejores resultados debido a que la reconstrucción calibrada obtiene un SAM de 12.61. En cambio, el método propuesto alcanza un valor mucho más bajo, con 3.67, que se acerca bastante a la firma original y al valor de SAM para un escenario ideal, con 3.05. Por último, para el punto F nuevamente ocurre que el método propuesto obtiene un mejor SAM que el escenario con matriz de sensado exacta, llegando a 3.61 frente a 5.89, respectivamente, mucho más bajo que la reconstrucción con matriz de sensado calibrada que obtiene 7.08.

Respecto a la escena con el arte surrealista que tiene los puntos G, H e I, se puede observar

el mismo comportamiento en donde la firma generada a partir de la reconstrucción con la matriz de sensado exacta obtiene el menor SAM de 2.43 para el punto G, 0.95 para el punto H y 3.35 para el punto I. De la misma manera, el método propuesto muestra el mejor comportamiento, después de la matriz de sensado exacta, obteniendo valores de SAM de 3.21 en el punto G, 2.06 en el punto H y 9.69 en el punto I. Por último, los peores resultados se encuentran con las reconstrucciones generadas a partir de la matriz de sensado calibrada en el laboratorio con valores de SAM de 4.55 para el punto G, 3.48 para el punto H y 10.82 para el punto I. Con esto se demuestra tanto visual como cuantitativamente que las firmas obtenidas por la matriz de sensado exacta y el método propuesto son las que más se aproximan a la firma original, lo cual indica un buen resultado.

7.3. Resultados Experimentales

En esta sección se presentan los resultados de reconstrucción obtenidos para 3 escenas de las 20 capturadas experimentalmente en el laboratorio de óptica del grupo de investigación HDSP usando la arquitectura CASSI. Como se explicó en la sección 6.4.2, las medidas poseen un tamaño de $256 \times 256 \times 13$. Teniendo en cuenta que se tomó un solo disparo (K), la compresión (C) obtenida por el sistema CASSI fue del 8,05 %, la cual se calcula de la siguiente manera:

$$C = \frac{K(M * (N + L - 1))}{MNL} * 100 = \frac{1(256 * (256 + 13 - 1))}{256 * 256 * 13} * 100 = 8,05 \% \quad (28)$$

Las medidas de la escena se capturaron usando el montaje descrito en la sección 6.4.1 y se reconstruyeron usando el algoritmo PnP tradicional y el método propuesto que implementa este algoritmo junto con una red neuronal, con lo cual se usaron los pesos pre entrenados del

aprendizaje obtenido con el conjunto de datos ARAD y posteriormente se hizo un re entrenamiento de la red neuronal utilizando 17 de las medidas experimentales capturadas en el laboratorio, ya que 3 de las medidas se tomaron para la prueba. En la figura 16 se presentan los resultados. La primera columna corresponde a la escena ubicada en el campo visual del lente objetivo del montaje. La segunda presenta la representación RGB del cubo de datos recuperado usando el algoritmo PnP tradicional, la tercera columna muestra la representación RGB de la reconstrucción obtenida usando el método propuesto y la última columna muestra la firma espectral de un punto de la escena escogido aleatoriamente con el objetivo de analizar la respuesta espectral de la reconstrucción.

En esta figura se pueden observar características importantes en los resultados de las reconstrucciones. Por ejemplo, para la primera escena correspondiente a un lego se puede notar que en la estimación obtenida con el PnP tradicional se presenta una tonalidad verde a lo largo de toda la figura, lo que indica una incorrecta estimación de color. Esto causa que múltiples zonas de la imagen difieran del color original. Sin embargo, para la reconstrucción a partir del PnP y la red neuronal, es posible notar que los colores estimados son mas similares a los de la imagen original. En la escena del pingüino y para la escena del perro se puede apreciar el mismo comportamiento ya que las reconstrucciones obtenidas con el PnP tradicional toman una tonalidad verdosa, mientras que las reconstrucciones obtenidas con el método propuesto demuestran un mejor desempeño ya que estiman de una mejor manera la escena tanto espacial como espectralmente.

Dicho resultado se ve reflejado en las firmas espectrales de los puntos escogidos, en donde se puede observar que en todas, el método propuesto, mostrado de color verde, se aproxima más a la firma espectral del punto original en color negro. Por otro lado se observa como las firmas

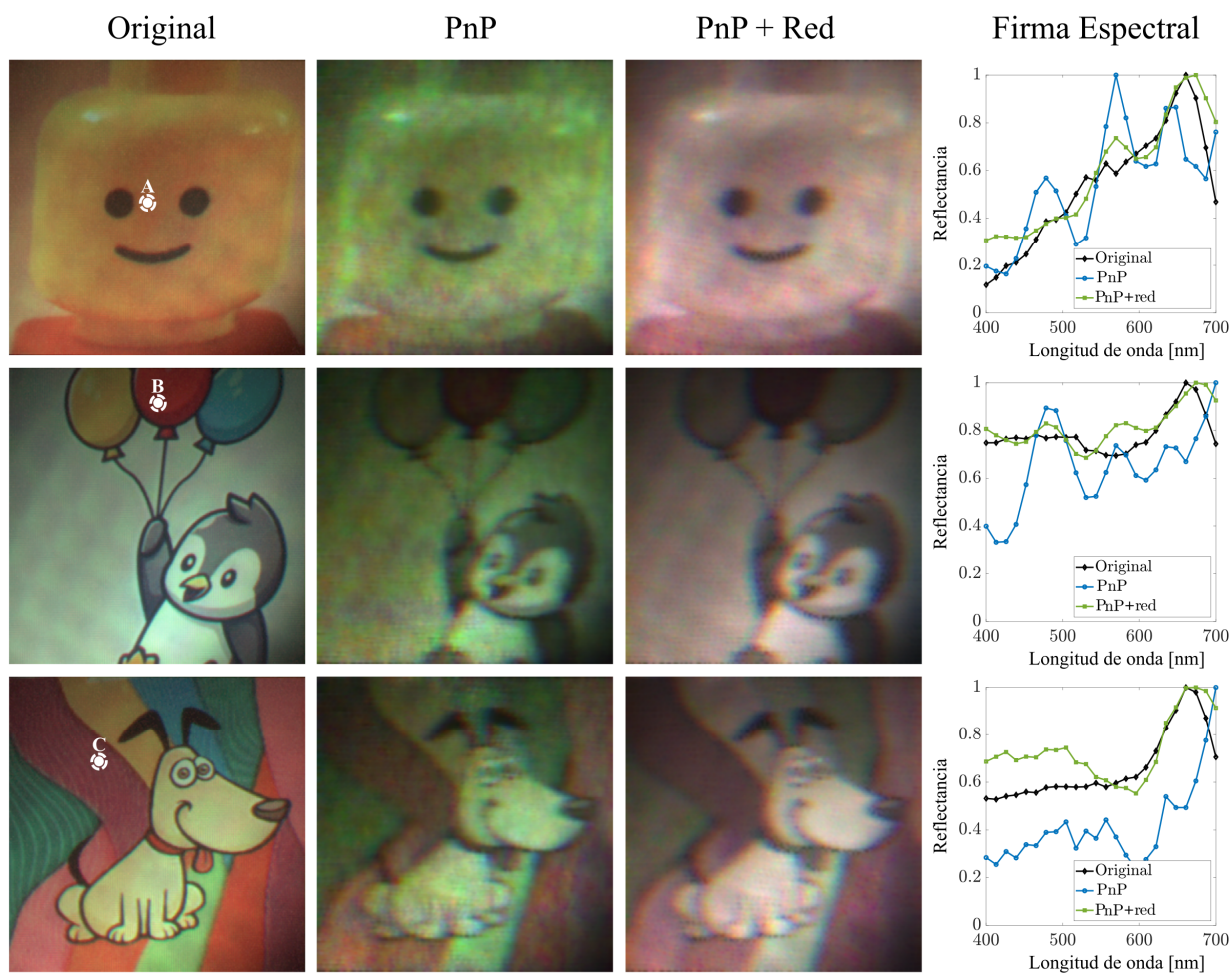


Figura 16. Resultados obtenidos para tres escenas capturadas experimentalmente en el laboratorio usando el sistema CASSI. En la primera columna se encuentra la imagen original. La segunda columna representa la reconstrucción obtenida usando el algoritmo PnP tradicional. La tercer columna muestra las reconstrucciones del método propuesto y la última columna muestra la firma espectral de un punto de la escena.

obtenidas con el PnP tradicional, de color azul son las que más se alejan de la firma original, lo cual indica un mal resultado. Específicamente, para el punto A se observa que las intensidades más altas se encuentran en las longitudes de onda largas correspondientes al color amarillo y para el método propuesto se aprecia un ajuste cercano en las primeras longitudes de onda con respecto a la respuesta original. Además, a pesar de que en algunos puntos tiende a distanciarse en intensidad, posteriormente se acerca de nuevo en los últimos puntos y por otro lado, presenta mejor comportamiento que el método tradicional. Adicionalmente, en el punto B se observa una firma con intensidades altas en las longitudes de onda más largas, ya que representa al globo rojo de la escena. Para este caso, la firma espectral del método propuesto se mantiene casi en la totalidad del espectro superando al método tradicional en la similitud de las intensidades a excepción de algunos puntos. Para el punto C se obtienen notables resultados. Si se observan las respuestas espectrales para las longitudes de onda más cortas, a pesar de que el método no se ajusta correctamente a la firma original, es más similar que la firma obtenida por el PnP tradicional. Además, para las longitudes de onda mayores de 550nm se puede notar cómo la firma resulta bastante similar a la original, mostrando un buen rendimiento de reconstrucción espectral.

Con esto se cumple el tercer y cuarto objetivo específico de este trabajo, los cuales consistían en entrenar la red neuronal con el conjunto de datos adquirido en el laboratorio para aprender el modelo de propagación no lineal y seguidamente, validar el modelo utilizando un algoritmo de recuperación para comparar la reconstrucción obtenida por el método propuesto y por la matriz de calibración obtenida en el laboratorio.

8. Conclusiones

Se propuso una metodología para el aprendizaje del modelo de propagación no lineal del sistema óptico CASSI. En particular, se implementó una red neuronal profunda basada en la arquitectura U-Net para la fase de entrenamiento, con la cual se aprendieron las medidas corruptas del sistema. Esta red fue entrenada con el conjunto de datos ARAD y fue probada con un conjunto de datos obtenido en el laboratorio del grupo de investigación, diseño de algoritmos y procesamiento de datos multidimensionales HDSP. La recuperación del cubo de datos espectral se realizó mediante una adaptación del algoritmo PnP-ADMM, el cual se alimentaba con las predicciones de la red neuronal propuesta en su paso de actualización con el objetivo de mejorar el rendimiento de las estimaciones.

Los resultados obtenidos muestran que el método propuesto estima imágenes espectrales con un rendimiento de reconstrucción mayor al que se obtendría usando un algoritmo PnP tradicional. La parte simulada con las imágenes espectrales del conjunto de datos ARAD, presenta una mejoría en las métricas de PSNR, SSIM y SAM, así como en la calidad de la reconstrucción a través del análisis de la respuesta espectral de múltiples puntos. Los resultados experimentales obtenidos a partir de las medidas adquiridas en el laboratorio con el sistema óptico CASSI muestran un buen rendimiento ya que se obtienen mejorías visuales, las cuales reducen las estimaciones incorrectas de color que dificultan la reconstrucción del cubo de datos. Además, el análisis de la respuesta espectral demuestra que la reconstrucción se ajusta de forma muy cercana a la firma espectral del cubo de datos original.

Referencias Bibliográficas

- Ahmad, R., Bouman, C. A., Buzzard, G. T., Chan, S., Liu, S., Reehorst, E. T., and Schniter, P. (2020). Plug-and-play methods for magnetic resonance imaging: Using denoisers for image recovery. *IEEE signal processing magazine*, 37(1):105–116.
- Albawi, S., Mohammed, T. A., and Al-Zawi, S. (2017). Understanding of a convolutional neural network. In *2017 International Conference on Engineering and Technology (ICET)*, pages 1–6. Ieee.
- Arad, B., Timofte, R., Ben-Shahar, O., Lin, Y.-T., and Finlayson, G. D. (2020). Ntire 2020 challenge on spectral reconstruction from an rgb image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 446–447.
- Arce, G. R., Brady, D. J., Carin, L., Arguello, H., and Kittle, D. S. (2013). Compressive coded aperture spectral imaging: An introduction. *IEEE Signal Processing Magazine*, 31(1):105–115.
- Arguello, H., Rueda, H., Wu, Y., Prather, D. W., and Arce, G. R. (2013). Higher-order computational model for coded aperture spectral imaging. *Applied optics*, 52(10):D12–D21.
- Bacca, J., Fonseca, Y., and Arguello, H. (2021). Compressive spectral image reconstruction using deep prior and low-rank tensor representation. *Applied optics*, 60(14):4197–4207.
- Baraniuk, R. G. (2007). Compressive sensing [lecture notes]. *IEEE signal processing magazine*, 24(4):118–121.

- Beck, A. and Teboulle, M. (2009). A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202.
- Blumensath, T. and Davies, M. E. (2009). Iterative hard thresholding for compressed sensing. *Applied and computational harmonic analysis*, 27(3):265–274.
- Boyd, S., Parikh, N., and Chu, E. (2011). *Distributed optimization and statistical learning via the alternating direction method of multipliers*. Now Publishers Inc.
- Brunner, T. A. (1997). Impact of lens aberrations on optical lithography. *IBM Journal of Research and Development*, 41(1.2):57–67.
- Chan, S. H., Wang, X., and Elgendy, O. A. (2016). Plug-and-play admm for image restoration: Fixed-point convergence and applications. *IEEE Transactions on Computational Imaging*, 3(1):84–98.
- Chandankhede, P. H., Titarmare, A. S., and Chauhvan, S. (2021). Voice recognition based security system using convolutional neural network. In *2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, pages 738–743. IEEE.
- Chang, X., Bian, L., Jiang, S., Zheng, G., and Zhang, J. (2021). Plug-and-play optimization for pixel super-resolution phase retrieval. *arXiv preprint arXiv:2105.14746*.
- Chen, Z., Li, D., Fan, W., Guan, H., Wang, C., and Li, J. (2021a). Self-attention in reconstruction bias u-net for semantic segmentation of building rooftops in optical remote sensing images. *Remote Sensing*, 13(13):2524.

- Chen, Z., Wang, C., Li, J., Xie, N., Han, Y., and Du, J. (2021b). Reconstruction bias u-net for road extraction from optical remote sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:2284–2294.
- Chuang, C.-Y., Liu, L.-C., Wei, C.-C., Liu, J.-J., Henrickson, L., Huang, W.-J., Wang, C.-L., Chen, Y.-K., and Chen, J. (2018). Convolutional neural network based nonlinear classifier for 112-gbps high speed optical link. In *Optical Fiber Communication Conference*, pages W2A–43. Optical Society of America.
- de Souza, J. C., Freire, R., and dos Santos, P. A. M. (2021). Subsampled digital holographic image reconstruction by a compressive sensing approach. *Applied Optics*, 60(1):1–9.
- Devassy, B. M. and George, S. (2021). Forensic analysis of beverage stains using hyperspectral imaging. *Scientific Reports*, 11(1):1–13.
- Figueiredo, M. A., Nowak, R. D., and Wright, S. J. (2007). Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE Journal of selected topics in signal processing*, 1(4):586–597.
- Fukami, K., Nakamura, T., and Fukagata, K. (2020). Convolutional neural network based hierarchical autoencoder for nonlinear mode decomposition of fluid field data. *Physics of Fluids*, 32(9):095110.
- Garini, Y., Young, I. T., and McNamara, G. (2006). Spectral imaging: principles and applications.

Cytometry Part A: The Journal of the International Society for Analytical Cytology, 69(8):735–747.

Gastal, E. S. and Oliveira, M. M. (2011). Domain transform for edge-aware image and video processing. In *ACM SIGGRAPH 2011 papers*, pages 1–12. Association for Computing Machinery.

Gat, N. (2000). Imaging spectroscopy using tunable filters: a review. In *Wavelet Applications VII*, volume 4056, pages 50–64. International Society for Optics and Photonics.

Gelvez, T., Rueda, H., and Arguello, H. (2017). Joint sparse and low rank recovery algorithm for compressive hyperspectral imaging. *Applied optics*, 56(24):6785–6795.

Gupta, R. and Hartley, R. I. (1997). Linear pushbroom cameras. *IEEE Transactions on pattern analysis and machine intelligence*, 19(9):963–975.

Huang, C., Tanaka, T., Kagami, S., Ninomiya, Y., Kakuda, M., Watanabe, K., Inoue, S., Nanba, K., Igarashi, Y., Yamamoto, T., et al. (2020). Multispectral imaging of mineral samples by infrared quantum dot focal plane array sensors. *Measurement*, 159:107775.

Işıl, Ç., Oktem, F. S., and Koç, A. (2019). Deep iterative reconstruction for phase retrieval. *Applied optics*, 58(20):5422–5431.

Jerri, A. J. (1977). The shannon sampling theorem—its various extensions and applications: A tutorial review. *Proceedings of the IEEE*, 65(11):1565–1596.

- Kasper, M., Fedrigo, E., Looze, D. P., Bonnet, H., Ivanescu, L., and Oberti, S. (2004). Fast calibration of high-order adaptive optics systems. *JOSA A*, 21(6):1004–1008.
- LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *nature*, 521(7553):436–444.
- Lee, S., Negishi, M., Urakubo, H., Kasai, H., and Ishii, S. (2020). Mu-net: Multi-scale u-net for two-photon microscopy image denoising and restoration. *Neural Networks*, 125:92–103.
- Li, W.-X., Miao, F., Xu, X.-Q., Zhang, J., Wu, Z.-Y., Chen, K.-M., Yan, F.-H., and Lin, X.-Z. (2021a). Pancreatic neuroendocrine neoplasms: Ct spectral imaging in grading. *Academic radiology*, 28(2):208–216.
- Li, X., Lai, S., and Qian, X. (2021b). Dbcface: Towards pure convolutional neural network face detection. *IEEE Transactions on Circuits and Systems for Video Technology*.
- Liu, J., Wu, Q., and Amin, M. G. (2021a). Multi-task bayesian compressive sensing exploiting signal structures. *Signal Processing*, 178:107804.
- Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., and Alsaadi, F. E. (2017). A survey of deep neural network architectures and their applications. *Neurocomputing*, 234:11–26.
- Liu, Y.-Y., Zhao, X.-L., Zheng, Y.-B., Ma, T.-H., and Zhang, H. (2021b). Hyperspectral image restoration by tensor fibered rank constrained optimization and plug-and-play regularization. *IEEE Transactions on Geoscience and Remote Sensing*.
- Park, J., Yoon, D., Seol, S. J., and Byun, J. (2019). Reconstruction of seismic field data with

convolutional u-net considering the optimal training input data. In *SEG International Exposition and Annual Meeting*. OnePetro.

Rippel, O., Snoek, J., and Adams, R. P. (2015). Spectral representations for convolutional neural networks. *arXiv preprint arXiv:1506.03767*.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer.

Salmi, P., Eskelinen, M. A., Leppänen, M. T., and Pölönen, I. (2021). Rapid quantification of microalgae growth with hyperspectral camera and vegetation indices. *Plants*, 10(2):341.

Sasián, J. (2013). *Introduction to aberrations in optical imaging systems*. Cambridge University Press.

Tropp, J. A. and Gilbert, A. C. (2007). Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on information theory*, 53(12):4655–4666.

Upadhyaya, V. and Salim, M. (2021). To analyse the effect of relaxation type on magnetic resonance image compression using compressive sensing. *International Journal of Online & Biomedical Engineering*, 17(4).

Vane, G., Green, R. O., Chrien, T. G., Enmark, H. T., Hansen, E. G., and Porter, W. M. (1993). The airborne visible/infrared imaging spectrometer (aviris). *Remote sensing of environment*, 44(2-3):127–143.

- Venkatakrishnan, S. V., Bouman, C. A., and Wohlberg, B. (2013). Plug-and-play priors for model based reconstruction. In *2013 IEEE Global Conference on Signal and Information Processing*, pages 945–948. IEEE.
- Vermeulen, M., Smith, K., Eremin, K., Rayner, G., and Walton, M. (2021). Application of uniform manifold approximation and projection (umap) in spectral imaging of artworks. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 252:119547.
- Vervandier, J. and Gioux, S. (2013). Single snapshot imaging of optical properties. *Biomedical optics express*, 4(12):2938–2944.
- Wang, J., Zhang, Y., Yuan, X., Meng, Z., and Tao, Z. (2021). Calibrated hyperspectral image reconstruction via graph-based self-tuning network. *arXiv preprint arXiv:2112.15362*.
- Wang, X. and Chan, S. H. (2017). Parameter-free plug-and-play admm for image restoration. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1323–1327. IEEE.
- Wang, Y. W., Reder, N. P., Kang, S., Glaser, A. K., and Liu, J. T. (2017). Multiplexed optical imaging of tumor-directed nanoparticles: a review of imaging systems and approaches. *Nanotheranostics*, 1(4):369.
- Wu, Q., Lai, Z., and Amin, M. G. (2021). Through-the-wall radar imaging based on bayesian compressive sensing exploiting multipath and target structure. *IEEE Transactions on Computational Imaging*, 7:422–435.

- Xiao, W., Zhou, P., An, S., and Zhu, J. (2022). Constrained nonlinear optimization method for accurate calibration of a bi-telecentric camera in a three-dimensional microtopography system. *Applied Optics*, 61(1):157–166.
- Xie, M., Liu, J., Sun, Y., Gan, W., Wohlberg, B., and Kamilov, U. S. (2021). Joint reconstruction and calibration using regularization by denoising with application to computed tomography. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4028–4037.
- Xie, M., Sun, Y., Liu, J., Wohlberg, B., and Kamilov, U. S. (2020). Joint reconstruction and calibration using regularization by denoising. *arXiv preprint arXiv:2011.13391*.
- Yu, X., Her, Y., Chang, A., Song, J.-H., Campoverde, E. V., and Schaffer, B. (2021). Assessing the effects of irrigation water salinity on two ornamental crops by remote spectral imaging. *Agronomy*, 11(2):375.
- Yuan, X. (2016). Generalized alternating projection based total variation minimization for compressive sensing. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 2539–2543. IEEE.
- Zhang, Y.-D., Satapathy, S. C., Guttery, D. S., Górriz, J. M., and Wang, S.-H. (2021). Improved breast cancer classification through combining graph convolutional network and convolutional neural network. *Information Processing & Management*, 58(2):102439.
- Zheng, S., Liu, Y., Meng, Z., Qiao, M., Tong, Z., Yang, X., Han, S., and Yuan, X. (2021). Deep

plug-and-play priors for spectral snapshot compressive imaging. *Photonics Research*, 9(2):B18–B29.

Zhou, C. and Rodrigues, M. R. (2021). An admm based network for hyperspectral unmixing tasks. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1870–1874. IEEE.

Zhou, Q., Li, Y., Tian, Y., and Jiang, L. (2020). A novel method based on nonlinear auto-regression neural network and convolutional neural network for imbalanced fault diagnosis of rotating machinery. *Measurement*, 161:107880.