

ALGORITMO DE CLASIFICACIÓN DE OBJETOS EN IMÁGENES DIFRACTIVAS  
BASADO EN MEDIDAS CUADRÁTICAS CODIFICADAS USANDO UN ENFOQUE  
DE APRENDIZAJE PROFUNDO

DAVID SANTIAGO MORALES NORATO

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

ALGORITMO DE CLASIFICACIÓN DE OBJETOS EN IMÁGENES DIFRACTIVAS  
BASADO EN MEDIDAS CUADRÁTICAS CODIFICADAS USANDO UN ENFOQUE  
DE APRENDIZAJE PROFUNDO

DAVID SANTIAGO MORALES NORATO

Trabajo de Grado para optar al título de  
Ingeniero de Sistemas

Director:

Andrés Felipe Jerez Ariza

*Magíster en Matemática Aplicada*

Codirector:

Henry Arguello Fuentes

*Ph.D. Electrical and Computer Engineering*

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

2022

## DEDICATORIA

*Para mi mamá Mariana, mi papá Hanssel,  
mis hermanas María Silvana, Katherin y mi hermano Enrrique.*

*ellos son el círculo de personas que más amo.*

*Este trabajo de grado es el resultado de todo el esfuerzo y apoyo que mi familia me  
ha brindado, todo lo que soy y lo que seré, se lo debo a ellos.*

## **AGRADECIMIENTOS**

Agradezco a mi director Andrés Jerez, una gran persona que me ha acompañado en mi formación tanto en lo académico como en lo personal, en especial le agradezco por su dedicación y paciencia como director de tesis.

Al grupo de investigación HDSP, por ser una comunidad de excelentes personas, las cuales me han brindado un gran apoyo tanto en lo académico como en lo personal desde mi vinculación al grupo.

A la Universidad Industrial de Santander por ser la institución que me brindó lo necesario para crecer y formalizar mi curiosidad científica.

A Santiago y Ángela, su compañía, consejo y amistad ha marcado estos años.

## CONTENIDO

	pág.
<b>INTRODUCCIÓN</b>	<b>13</b>
<b>1 OBJETIVOS</b>	<b>19</b>
<b>2 ADQUISICIÓN DE LA INFORMACIÓN DE FASE</b>	<b>20</b>
2.1 SISTEMA ÓPTICO DE DIFRACCIÓN	20
2.2 PROBLEMA DE RECUPERACIÓN DE FASE	22
2.3 MEDIDAS CUADRÁTICAS CODIFICADAS	23
<b>3 ALGORITMOS DE RECUPERACIÓN DE FASE</b>	<b>25</b>
3.1 FORMULACIONES CONVEXAS	25
3.2 FORMULACIONES NO CONVEXAS	26
<b>4 SISTEMAS DE CLASIFICACIÓN</b>	<b>28</b>
4.1 MÁQUINAS DE SOPORTE VECTORIAL	28
4.2 K VECINOS MÁS CERCANOS	30
4.3 REDES NEURONALES	31
4.4 CLASIFICACIÓN DE IMÁGENES USANDO REDES NEURONALES ÓPTICAS	32
4.5 CLASIFICACIÓN USANDO MEDIDAS CUADRÁTICAS	34
<b>5 METODOLOGÍA PROPUESTA</b>	<b>35</b>
5.1 ETAPA DE INICIALIZACIÓN	35
5.2 ETAPA DE CLASIFICACIÓN	37
5.3 ESQUEMA PROPUESTO DE CLASIFICACIÓN DE OBJETOS	38

<b>6 SIMULACIONES Y RESULTADOS</b>	<b>40</b>
6.1 CONJUNTO DE DATOS	40
6.2 MÉTRICAS	41
6.2.1 Métricas para evaluar la inicialización.	41
6.2.2 Métricas para evaluar la clasificación.	42
6.3 EXPERIMENTOS	43
6.3.1 Configuración experimental.	43
6.3.2 Resultados de la inicialización.	45
6.3.3 Resultados de la clasificación de objetos.	46
<b>7 CONCLUSIONES</b>	<b>52</b>
<b>8 TRABAJO FUTURO</b>	<b>53</b>
<b>BIBLIOGRAFÍA</b>	<b>54</b>

## LISTA DE FIGURAS

	<b>pág.</b>	
Figura 1	Sistema óptico de difracción convencional.	21
Figura 2	Sistema óptico de difracción codificado.	24
Figura 3	Representación de una SVM en $\mathbb{R}^2$ . La recta $wx - b = 0$ en rojo representa el plano óptimo que soluciona el problema de optimización (17). Las rectas $wx - b = 1$ y $wx - b = -1$ representan los extremos de separación de las clases de puntos en rojo y en azul.	29
Figura 4	Representación de KNN en $\mathbb{R}^2$ con $K = 5$ .	30
Figura 5	Esquema de una arquitectura de red neuronal $f : \mathbb{R}^{d_1} \rightarrow \mathbb{R}^{d_2}$ .	32
Figura 6	Implementación de redes neuronales ópticas implementadas físicamente mediante elementos difractivos.	33
Figura 7	Esquema propuesto de red neuronal profunda de tres etapas para la clasificación de objetos, a partir de medidas cuadráticas codificadas.	35
Figura 8	Ejemplo de las imágenes presentes en los conjuntos de datos MNIST y Fashion-MNIST.	40
Figura 9	Resumen del desempeño de la inicialización propuesta comparando con los métodos FSI, OPI, WMCI sobre la métrica de RE, variando el número de iteraciones del algoritmo para estimar imágenes de los conjuntos MNIST y Fashion-MNIST en los campos de difracción cercano, medio y lejano.	46

- Figura 10 Resumen del desempeño de la inicialización comparando con diferentes estrategias del estado del arte. Se varió el número de medidas  $T$  para realizar la estimación y diferentes niveles de ruido afectando las medidas descritos por el SNR. 47
- Figura 11 Resultados visuales de la estimación del campo óptico inicial comparando el método propuesto contra el algoritmo FSI con  $L = 1, T = 10$ , FSI usando  $L = 1, T = 200$ , OPI  $L = 4, T = 200$  y WMCI  $L = 4, T = 200$ . 48
- Figura 12 Resultados en la métrica de exactitud, precisión, exhaustividad y F1 para evaluar la clasificación de medidas cuadráticas codificadas simuladas sobre el campo cercano, medio y lejano. Se muestra la comparación de la clasificación Tradicional, el uso del propagador inverso propuesto (BPM) y el método de inicialización aprendida (LFSI) usando la arquitectura MobilNetV2, sobre la división de prueba de los conjuntos de datos MNIST y Fashion-MNIST. 49
- Figura 13 Resultados en la métrica de exactitud, precisión, exhaustividad y F1 para evaluar la clasificación de medidas cuadráticas codificadas simuladas sobre el campo cercano, medio y lejano. Se muestra la comparación de la clasificación Tradicional, el uso del propagador inverso propuesto (BPM) y el método de inicialización aprendida (LFSI) usando la arquitectura InveptionV3, sobre la división de prueba de los conjuntos de datos MNIST y Fashion-MNIST. 50

Figura 14 Resultados en la métrica de exactitud, precisión, exhaustividad y F1 para evaluar la clasificación de medidas cuadráticas codificadas simuladas sobre el campo cercano, medio y lejano. Se muestra la comparación de la clasificación Tradicional, el uso del propagador inverso propuesto (BPM) y el método de inicialización aprendida (LFSI) usando la arquitectura Xception, sobre la división de prueba de los conjuntos de datos MNIST y Fashion-MNIST.

51

## LISTA DE CUADROS

	<b>pág.</b>
Cuadro 1 Ejemplos de codificaciones aleatorias admisibles según la Definición 1.	24
Cuadro 2 Resumen de las arquitecturas de redes neuronales de clasificación utilizadas.	38
Cuadro 3 Resumen de la división de los conjuntos de datos usados para evaluar el método propuesto.	41
Cuadro 4 Parámetros de propagación usados para simular el modelo de propagación (7) para cada campo de difracción.	45

## RESUMEN

**TÍTULO:** ALGORITMO DE CLASIFICACIÓN DE OBJETOS EN IMÁGENES DIFRACTIVAS BASADO EN MEDIDAS CUADRÁTICAS CODIFICADAS USANDO UN ENFOQUE DE APRENDIZAJE PROFUNDO \*

**AUTOR:** DAVID SANTIAGO MORALES NORATO \*\*

**PALABRAS CLAVE:** Medidas cuadráticas codificadas, recuperación de fase, máscaras de codificación, aprendizaje profundo, clasificación de objetos.

### **DESCRIPCIÓN:**

El aprendizaje profundo se ha usado ampliamente en el área de visión por computador, así como en la clasificación de objetos. Sin embargo, los enfoques tradicionales se basan principalmente en imágenes de intensidad adquiridas por sistemas de propagación lineal. Dado que los detectores ópticos solo pueden medir la intensidad del campo óptico complejo subyacente, estos sistemas lineales solo registran la información de magnitud, mientras que la información de fase no se puede medir. La recuperación de fase se ha abordado a través de sistemas cuadráticos que modulan el campo óptico usando máscaras de codificación, produciendo medidas cuadráticas codificadas. La clasificación de medidas cuadráticas se ha propuesto recientemente sin el paso de reconstrucción debido a su alto costo computacional. No obstante, la clasificación de medidas cuadráticas codificadas no ha sido estudiada en el estado del arte. Este trabajo propone un esquema de clasificación de objetos sobre medidas cuadráticas codificadas usando aprendizaje profundo. El esquema propuesto consta de tres etapas principales: una capa de adquisición, que simula el proceso de adquisición; un paso de inicialización que estima el campo óptico; y una arquitectura de red neuronal que realiza la tarea de clasificación sobre la estimación inicial. Los resultados de la simulación muestran que el método de clasificación propuesto supera el enfoque tradicional al evaluar diferentes métricas de clasificación sobre los conjuntos de datos MNIST y Fashion-MNIST a partir de los campos cercano, medio y lejano

---

\* Trabajo de grado

\*\* Facultad de Ingenierías Fisicomecánicas. Escuela de Ingeniería de Sistemas e Informática. Director: Andrés Felipe Jerez Ariza. Codirector: Henry Arguello Fuentes.

## ABSTRACT

**TITLE:** OBJECT CLASSIFICATION ALGORITHM IN DIFFRACTIVE IMAGING BASED ON CODED QUADRATIC MEASUREMENTS USING A DEEP LEARNING APPROACH \*

**AUTHOR:** DAVID SANTIAGO MORALES NORATO \*\*

**KEYWORDS:** Coded quadratic measurements, phase retrieval, coding masks, deep learning, object classification.

### DESCRIPTION:

Deep learning has been broadly used in computer vision such as in object classification. However, traditional object classification approaches are mostly based on intensity images acquired by linear propagation systems. Since optical detectors can only measure the intensity underlying complex optical field, these linear systems capture only magnitude information, while the phase information cannot be directly measured. Then, the phase recovery has been addressed by using quadratic systems that modulate the optical field from coding masks, producing coded quadratic measurements. Quadratic measurement classification has been recently proposed without the reconstruction step because of its high computational cost. Nevertheless, coded quadratic measurement classification has not been studied in the state-of-the-art. This work proposes an object classification scheme over coded quadratic measurements through deep learning. The proposed scheme consists of three main stages: an acquisition layer that simulates the acquisition process; an initialization step that estimates the optical field; and a neural network architecture that performs the classification task over the initial guess. Simulation results show that the proposed classification method overcomes traditional approach by evaluating different classification metrics over MNIST and Fashion-MNIST datasets from the near, middle, and far-fields.

---

\* Bachelor Thesis

\*\* Facultad de Ingenierías Fisicomecánicas. Escuela de Ingeniería de Sistemas e Informática. Advisor: Andrés Felipe Jerez Ariza. Co-advisor: Henry Arguello Fuentes

## INTRODUCCIÓN

Los algoritmos computacionales basados en aprendizaje profundo han sido ampliamente estudiados en la literatura, especialmente, la clasificación de objetos en imágenes ha sido una de las tareas computacionales más abordadas en ese tópico <sup>1,2,3</sup>. Los enfoques de aprendizaje profundo utilizan arquitecturas de redes neuronales que consisten en la concatenación de múltiples capas compuestas de unidades mínimas llamadas neuronas. Cada neurona realiza una combinación lineal entre las entradas, para posteriormente, usar una función no lineal en la salida. Las salidas de cada neurona en una capa funcionan como entrada de las neuronas ubicadas en la siguiente capa, creando así, una arquitectura de red neuronal profunda <sup>4</sup>.

En general, la clasificación de objetos se realiza sobre imágenes en escala de grises <sup>5</sup>, RGB <sup>6</sup>, o más recientemente, imágenes espectrales <sup>1</sup>. Sin embargo, los enfoques de clasificación de objetos basados en aprendizaje profundo incorporan como entrada de la arquitectura de red neuronal, la información de la intensidad de la luz

- 
- <sup>1</sup> Shutao Li y col. "Deep learning for hyperspectral image classification: An overview". En: *IEEE Transactions on Geoscience and Remote Sensing* 57.9 (2019), págs. 6690-6709.
  - <sup>2</sup> Ying Li y col. "Deep learning for remote sensing image classification: A survey". En: *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 8.6 (2018), e1264.
  - <sup>3</sup> Wei Wang y col. "Development of convolutional neural network and its application in image classification: a survey". En: *Optical Engineering* 58.4 (2019), pág. 040901.
  - <sup>4</sup> Jianqing Fan, Cong Ma y Yiqiao Zhong. "A selective overview of deep learning". En: *Statistical Science* 36.2 (2021), págs. 264-290.
  - <sup>5</sup> Hieu Minh Bui y col. "Using grayscale images for object recognition with convolutional-recursive neural network". En: *2016 IEEE Sixth International Conference on Communications and Electronics (ICCE)*. IEEE. 2016, págs. 321-325.
  - <sup>6</sup> Alex Krizhevsky, Ilya Sutskever y Geoffrey E Hinton. "ImageNet classification with deep convolutional neural networks". En: *Communications of the ACM* 60.6 (2017), págs. 84-90.

incidente sobre el sensor, omitiendo la información de fase, que resulta fundamental en aplicaciones como cristalografía de rayos-x <sup>7</sup>, astronomía <sup>8</sup>, holografía <sup>9</sup>, entre otras. Esta limitación se atañe a los sistemas ópticos que dependen de la conversión de fotones a electrones, puesto que no permiten una adquisición directa de la información de fase <sup>10</sup>. Por lo tanto, la obtención de esta información de fase requiere la implementación de algoritmos computacionales que logren recuperar los datos perdidos.

Los algoritmos de recuperación de fase permiten reconstruir la información de un campo óptico inicial con base en la adquisición de medidas de intensidad que siguen un modelo cuadrático de propagación según diferentes campos de difracción, tales como campo cercano, medio y lejano <sup>11</sup>. Dentro de estos sistemas ópticos de difracción se han incorporado máscaras de fase para la modulación del campo óptico inicial, puesto que, la literatura ha demostrado que la inclusión de este tipo de elementos ópticos durante el proceso de adquisición, genera redundancia en las medidas captadas, garantizando la recuperación de fase en hasta una constante unimodular <sup>12</sup>. Estas medidas adquiridas a través de sistemas ópticos de difracción,

---

<sup>7</sup> Samuel Pinilla, Juan Poveda y Henry Arguello. "Coded diffraction system in X-ray crystallography using a boolean phase coded aperture approximation". En: *Optics Communications* 410 (2018), págs. 707-716.

<sup>8</sup> C Fienup y J Dainty. "Phase retrieval and image reconstruction for astronomy". En: *Image recovery: theory and application* 231 (1987), pág. 275.

<sup>9</sup> Yair Rivenson y col. "Phase recovery and holographic image reconstruction using deep learning in neural networks". En: *Light: Science & Applications* 7.2 (2018), págs. 17141-17141.

<sup>10</sup> Yoav Shechtman y col. "Phase retrieval with application to optical imaging: a contemporary overview". En: *IEEE signal processing magazine* 32.3 (2015), págs. 87-109.

<sup>11</sup> Joseph W Goodman. *Introduction to Fourier optics*. Roberts y Company Publishers, 2005.

<sup>12</sup> Emmanuel J Candes, Xiaodong Li y Mahdi Soltanolkotabi. "Phase retrieval from coded diffraction patterns". En: *Applied and Computational Harmonic Analysis* 39.2 (2015), págs. 277-299.

que incluyen máscaras de fase, se denominan medidas cuadráticas codificadas. Las imágenes recuperadas usando medidas cuadráticas codificadas se conocen como imágenes ópticas difractivas. Diversos algoritmos iterativos han sido propuestos para resolver la reconstrucción de imágenes difractivas a partir del problema de recuperación de fase. Tradicionalmente, estos algoritmos se construyen bajo formulaciones convexas, tales como el *PhaseLift*<sup>13</sup> y *PhaseMax*<sup>14</sup>; o formulaciones no convexas, tales como *Wirtinger Flow*<sup>15</sup>, *Truncated Wirtinger Flow*<sup>16</sup>, *Truncated Amplitude Flow*<sup>17</sup> y *Reweighted Amplitude Flow*<sup>18</sup>.

Actualmente, se han involucrado arquitecturas de redes neuronales profundas en el campo de formación de imágenes ópticas, puesto que, permite incorporar el modelo de adquisición como una capa de la misma arquitectura de red neuronal para promover una mejor reconstrucción de las medidas adquiridas. La inclusión del modelo de adquisición ha permitido el diseño de sistemas ópticos de adquisición a través de

- 
- <sup>13</sup> Emmanuel J. Candès, Thomas Strohmer y Vladislav Voroninski. "PhaseLift: Exact and Stable Signal Recovery from Magnitude Measurements via Convex Programming". En: *Communications on Pure and Applied Mathematics* 66.8 (nov. de 2012), págs. 1241-1274. DOI: 10.1002/cpa.21432.
- <sup>14</sup> Tom Goldstein y Christoph Studer. "Phasemax: Convex phase retrieval via basis pursuit". En: *IEEE Transactions on Information Theory* 64.4 (2018), págs. 2675-2689.
- <sup>15</sup> Emmanuel J. Candès, Xiaodong Li y Mahdi Soltanolkotabi. "Phase Retrieval via Wirtinger Flow: Theory and Algorithms". En: *IEEE Transactions on Information Theory* 61.4 (2015), págs. 1985-2007. DOI: 10.1109/TIT.2015.2399924.
- <sup>16</sup> Yuxin Chen y Emmanuel J Candès. "Solving random quadratic systems of equations is nearly as easy as solving linear systems". En: *Communications on pure and applied mathematics* 70.5 (2017), págs. 822-883.
- <sup>17</sup> Gang Wang, Georgios B Giannakis y Yonina C Eldar. "Solving systems of random quadratic equations via truncated amplitude flow". En: *IEEE Transactions on Information Theory* 64.2 (2017), págs. 773-794.
- <sup>18</sup> Gang Wang y col. "Phase retrieval via reweighted amplitude flow". En: *IEEE Transactions on Signal Processing* 66.11 (2018), págs. 2818-2833.

su implementación en configuraciones de red neuronal donde los pesos entrenables representan las variables optimizables del modelo de propagación <sup>19</sup>.

Por otra parte, algunos trabajos han estudiado la clasificación de objetos usando únicamente medidas de intensidad captadas mediante sistemas ópticos, definidos matemáticamente a través de operadores lineales, por ejemplo, sistemas de tomografía computarizada <sup>20</sup>, imágenes de un solo píxel <sup>21</sup>, entre otros. De manera que, los sistemas de clasificación que omiten el proceso de reconstrucción de las medidas obtenidas disminuyen el tiempo de inferencia, debido a que, se elimina la etapa de reconstrucción de la imagen que, generalmente, implica un alto costo computacional en los sistemas de clasificación. A pesar de que se han desarrollado arquitecturas de redes neuronales para la clasificación usando medidas de intensidad bajo un modelo cuadrático <sup>22,23</sup>, en la literatura de imágenes difractivas que resultan del problema de recuperación de fase no se han abordado modelos de redes neuronales para la detección de objetos basados en medidas cuadráticas codificadas. Así que, surge el interés de estudiar sistemas de clasificación que permitan la discriminación de objetos con base en medidas cuadráticas codificadas.

Este trabajo de investigación propone el diseño de un algoritmo de clasificación de

---

<sup>19</sup> Jorge Bacca, Tatiana Gelvez y Henry Arguello. "Transmittance Regularizer for Binary coded Aperture Design in a Computational Imaging end-to-end Approach". En: *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2021, págs. 1470-1474.

<sup>20</sup> Clément Douarre y col. "On the value of CTIS imagery for neural-network-based classification: a simulation perspective". En: *Appl. Opt.* 59.28 (2020), págs. 8697-8710. DOI: 10.1364/AO.394868.

<sup>21</sup> Jorge Bacca, Laura Galvis y Henry Arguello. "Coupled deep learning coded aperture design for compressive image classification". En: *Optics express* 28.6 (2020), págs. 8528-8540.

<sup>22</sup> Sung-Jin Kim y col. "Deep transfer learning-based hologram classification for molecular diagnostics". En: *Scientific reports* 8.1 (2018), págs. 1-12.

<sup>23</sup> Angelo Ziletti y col. "Insightful classification of crystal structures using deep learning". En: *Nature communications* 9.1 (2018), págs. 1-10.

objetos en imágenes difractivas sobre medidas cuadráticas codificadas mediante un esquema de aprendizaje profundo de tres etapas. Primero, una capa de adquisición simula el proceso de medición. Esta capa contiene un modelo de propagación del campo óptico codificado. Segundo, un procedimiento de inicialización aproxima el campo óptico inicial. Por último, la red de clasificación infiere la clase correspondiente a cada medida cuadrática codificada según la estimación inicial. Específicamente, el esquema propuesto incluye la simulación de medidas cuadráticas codificadas bajo los modelos de propagación de difracción del campo cercano, medio y lejano. Además, en la etapa de clasificación se usaron tres redes neuronales del estado del arte, tales como MobilNetV2 <sup>24</sup>, InceptionV3 <sup>25</sup> y Xception <sup>26</sup> para clasificar medidas cuadráticas codificadas simuladas sobre los conjuntos de datos Fashion-MNIST <sup>27</sup> y MNIST <sup>28</sup>. Los resultados de la clasificación muestran que el esquema propuesto logra superar el esquema de clasificación tradicional para cada una de las zonas de difracción y redes de clasificación. Concretamente, el método propuesto supera al método tradicional en hasta 0.24, 0.2, 0.25 y 0.22 en términos de exactitud, precisión, exhaustividad y métrica F1, respectivamente. En particular, estas ganancias se reportan para la clasificación de objetos utilizando medidas simuladas del campo lejano a partir del conjunto de datos de Fashion-MNIST.

---

<sup>24</sup> Mark Sandler y col. "MobileNetV2: Inverted Residuals and Linear Bottlenecks". En: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2018.

<sup>25</sup> Christian Szegedy y col. "Rethinking the Inception Architecture for Computer Vision". En: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016.

<sup>26</sup> Francois Chollet. "Xception: Deep Learning With Depthwise Separable Convolutions". En: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017.

<sup>27</sup> Han Xiao, Kashif Rasul y Roland Vollgraf. "Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms". En: *arXiv preprint arXiv:1708.07747* (2017).

<sup>28</sup> Li Deng. "The mnist database of handwritten digit images for machine learning research [best of the web]". En: *IEEE Signal Processing Magazine* 29.6 (2012), págs. 141-142.

Este documento se encuentra organizado de la siguiente manera: en la sección 1 se enumeran los objetivos formulados en este trabajo de investigación. Las secciones 2, 3, y 4 describen la adquisición de medidas cuadráticas codificadas, los algoritmos de recuperación y los sistemas de clasificación del estado del arte, respectivamente. La sección 5 describe la metodología propuesta para la clasificación de objetos utilizando medidas cuadráticas codificadas basado en aprendizaje profundo. Los resultados de la evaluación del método propuesto comparado con técnicas del estado del arte se presenta en la sección 6. Finalmente, las secciones 7 y 8 muestran las conclusiones y el trabajo futuro, respectivamente.

## 1. OBJETIVOS

### Objetivo general

- Desarrollar un algoritmo de clasificación de objetos en imágenes difractivas basado en medidas cuadráticas codificadas usando un enfoque de aprendizaje profundo.

### Objetivos específicos

1. Modelar matemáticamente el proceso de adquisición de medidas cuadráticas codificadas utilizando máscaras de fase.
2. Diseñar e implementar un algoritmo de clasificación de objetos en medidas cuadráticas codificadas a partir de aprendizaje profundo que incorpore el modelo de adquisición.
3. Simular una configuración óptica difractiva para la adquisición de medidas cuadráticas codificadas usando máscaras de fase.
4. Evaluar el algoritmo de clasificación de medidas cuadráticas codificadas basado en aprendizaje profundo frente a otras técnicas del estado del arte usando medidas adquiridas con la configuración óptica simulada.

## 2. ADQUISICIÓN DE LA INFORMACIÓN DE FASE

Los sistemas ópticos tradicionales de adquisición captan únicamente la información de la intensidad de la luz incidente sobre el sensor, perdiendo la información de fase que resulta fundamental en aplicaciones, tales como cristalografía de rayos-x <sup>7</sup>, astronomía <sup>8</sup>, holografía <sup>9</sup>, entre otras. Por lo tanto, la obtención de la información de fase requiere la implementación de sistemas ópticos que logren codificar la fase del campo óptico, preservándola implícita en las medidas de intensidad adquiridas.

### 2.1. SISTEMA ÓPTICO DE DIFRACCIÓN

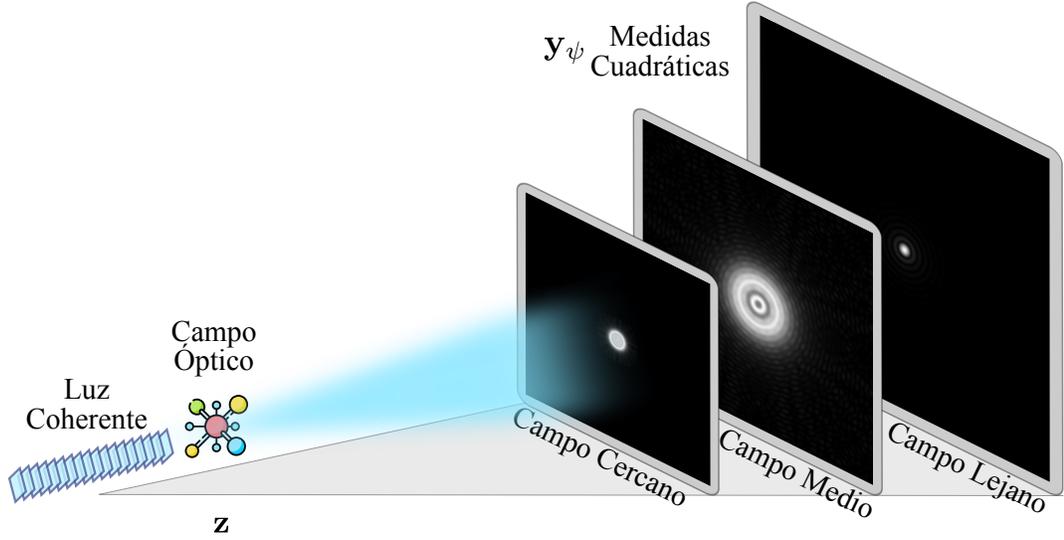
En múltiples áreas de la ciencia e ingeniería se presenta la adquisición de medidas cuadráticas de intensidad, con base en sistemas ópticos de difracción <sup>8,7,9</sup>. La Figura 1 muestra un esquema convencional de los sistemas ópticos de difracción a lo largo de diferentes campos de propagación. Específicamente, este sistema de adquisición está compuesto por un objeto  $\mathbf{z} \in \mathbb{C}^n$  ubicado en el campo óptico, que es iluminado por una fuente de luz coherente posteriormente, este campo óptico inicial se propaga una distancia  $z$ , produciendo medidas cuadráticas o patrones de difracción que, finalmente, se adquieren en el sensor. Cabe señalar que, se pueden definir tres diferentes campos de difracción  $\psi$  según la distancia de propagación  $z$ , comúnmente conocidos como campo cercano, medio y lejano.

Matemáticamente, la adquisición de medidas cuadráticas  $\mathbf{y}_\psi \in \mathbb{R}^n$  en cada campo de difracción  $\psi$  se puede modelar como

$$\mathbf{y}_\psi = |\mathbf{A}_\psi \mathbf{z}|^2, \quad (1)$$

donde  $|\cdot|$  denota la magnitud, y  $\mathbf{A}_\psi \in \mathbb{C}^{n \times n}$  es la matriz que modela el sistema de adquisición según cada campo de difracción. Particularmente, el objeto de interés

Figura 1. Sistema óptico de difracción convencional.



contiene información de magnitud y fase  $\mathbf{z} = |\mathbf{z}| \odot e^{j\text{ang}(\mathbf{z})}$  con  $j = \sqrt{-1}$ , donde  $\text{ang}(\cdot)$  devuelve la fase de un vector complejo. La matriz  $\mathbf{A}_\psi$  se define como

$$\mathbf{A}_\psi = \begin{cases} \mathbf{F}\mathbf{T}\mathbf{F}^H & \text{si } \psi = 1 \rightarrow \text{Campo cercano,} \\ \mathbf{F}^H\mathbf{Q} & \text{si } \psi = 2 \rightarrow \text{Campo medio,} \\ \mathbf{F} & \text{si } \psi = 3 \rightarrow \text{Campo lejano,} \end{cases} \quad (2)$$

donde  $\mathbf{F} \in \mathbb{C}^{n \times n}$  corresponde a la transformada discreta de Fourier,  $\mathbf{Q} \in \mathbb{C}^{n \times n}$  y  $\mathbf{T} \in \mathbb{C}^{n \times n}$  representan matrices que describen las funciones de transferencia usadas en óptica de Fourier<sup>29</sup> para modelar la difracción en el campo cercano y medio, respectivamente, dadas por

$$(\mathbf{Q})_{p,q} = e^{\frac{-jk_0}{2z}(p^2\delta_x^2 + q^2\delta_y^2)}, \quad (3)$$

<sup>29</sup> Ting-Chung Poon y Jung-Ping Liu. *Introduction to modern digital holography: with MATLAB*. Cambridge University Press, 2014.

$$(\mathbf{T})_{r,s} = e^{-jk_0z} \sqrt{1 - \frac{(r\delta_{k_x})^2}{k_0^2} - \frac{(s\delta_{k_y})^2}{k_0^2}}. \quad (4)$$

Aquí,  $k_0 = 2\pi/\lambda$  representa el número de onda,  $[p, q]$  y  $[r, s]$  son los índices discretos de las muestras en el dominio espacial y el dominio de Fourier, respectivamente. Los términos  $\delta_x$  y  $\delta_y$  son el periodo de muestreo en las coordenadas espaciales, y  $\delta_{k_x}$  y  $\delta_{k_y}$  son el periodo de muestreo en las coordenadas de las frecuencias.

## 2.2. PROBLEMA DE RECUPERACIÓN DE FASE

La recuperación del campo óptico inicial  $\mathbf{z}$  usando medidas cuadráticas es un problema inverso comúnmente denominado recuperación de fase. En general, el problema de recuperación de fase basado en (1) se puede definir como

$$\underset{\mathbf{z} \in \mathbb{C}^n}{\text{minimizar}} \|\mathbf{y}_\psi - |\mathbf{A}_\psi \mathbf{z}|^2\|_2^2, \quad (5)$$

donde  $\|\cdot\|_2$  es la norma euclidiana. Usualmente, se puede garantizar una recuperación que conserva un error de fase global (7)<sup>30,31,32</sup>, es decir que,  $\hat{\mathbf{z}} = \hat{\mathbf{z}}e^{j\theta}$  con  $\theta \in [0, 2\pi)$ . En consecuencia, la distancia euclidiana entre la escena  $\mathbf{z}$  y su estimación  $\hat{\mathbf{z}}$  está dada por

$$\text{dist}(\mathbf{z}, \hat{\mathbf{z}}) = \min_{\theta \in [0, 2\pi)} \|\mathbf{z}e^{-j\theta} - \hat{\mathbf{z}}\|_2. \quad (6)$$

---

<sup>30</sup> Yonina C Eldar y Shahar Mendelson. "Phase retrieval: Stability and recovery guarantees". En: *Applied and Computational Harmonic Analysis* 36.3 (2014), págs. 473-494.

<sup>31</sup> David Gross, Felix Kraemer y Richard Kueng. "Improved recovery guarantees for phase retrieval from coded diffraction patterns". En: *Applied and Computational Harmonic Analysis* 42.1 (2017), págs. 37-64.

<sup>32</sup> Andrés Guerrero, Samuel Pinilla y Henry Arguello. "Phase Recovery Guarantees From Designed Coded Diffraction Patterns in Optical Imaging". En: *IEEE Transactions on Image Processing* 29 (2020), págs. 5687-5697. DOI: 10.1109/TIP.2020.2985208.

Adicionalmente, el problema inverso (7) se ha resuelto creando redundancia en el proceso de medición al incluir una máscara de fase, que permite la modulación del campo óptico, produciendo medidas cuadráticas codificadas según cada campo de difracción.

### 2.3. MEDIDAS CUADRÁTICAS CODIFICADAS

En esta subsección, se describe el sistema de adquisición que capta medidas cuadráticas codificadas de un campo óptico a lo largo de cada campo de difracción. La Figura 2 ilustra el esquema de adquisición de medidas cuadráticas codificadas que incluye una máscara de codificación de fase  $\mathbf{D}_\ell \in \mathbb{C}^{n \times n}$  para modular el campo óptico inicial. De hecho, al modificar la configuración espacial de la máscara de fase es posible adquirir múltiples proyecciones de la escena. De modo que, las proyecciones cuadráticas codificadas en cada campo de difracción para la  $\ell$ -ésima proyección está dada por

$$\mathbf{y}_{\ell,\psi} = |\mathbf{A}_{\ell,\psi}\mathbf{z}|^2, \ell = \{1, \dots, L\}, \quad (7)$$

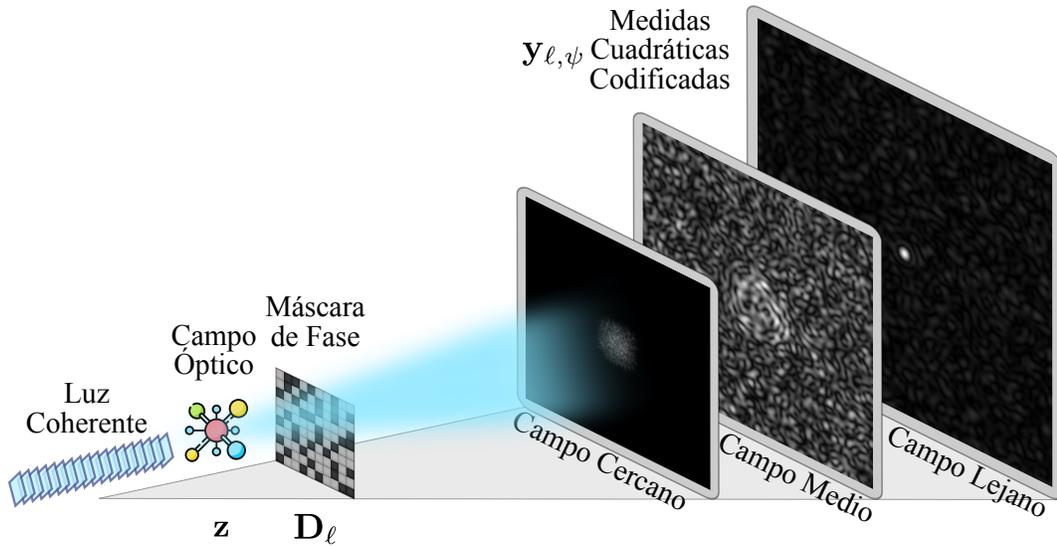
donde  $\mathbf{A}_{\ell,\psi} \in \mathbb{C}^{n \times n}$  es la matriz que describe la  $\ell$ -ésima proyección según la propagación de cada campo de difracción  $\psi$ , siendo  $L$  el número total de proyecciones.

Específicamente, la matriz del sistema de adquisición  $\mathbf{A}_{\ell,\psi}$  introduciendo una máscara de fase, se puede expresar como

$$\mathbf{A}_{\ell,\psi} = \begin{cases} \mathbf{F}^T \mathbf{F}^H \mathbf{D}_\ell & \text{si } \psi = 1 \rightarrow \text{Campo cercano,} \\ \mathbf{F}^H \mathbf{Q} \mathbf{D}_\ell & \text{si } \psi = 2 \rightarrow \text{Campo medio,} \\ \mathbf{F} \mathbf{D}_\ell & \text{si } \psi = 3 \rightarrow \text{Campo lejano.} \end{cases} \quad (8)$$

En general, las entradas aleatorias de la matriz  $\mathbf{D}_\ell = \text{diag}(\mathbf{d})$  son *i.i.d* copias de una variable aleatoria  $d \in \mathbb{C}$  con  $\mathbf{d} \in \{d\}^n$ , donde  $\text{diag}(\cdot)$  devuelve una matriz diagonal cuadrada de un vector dado. Aquí,  $d$  representa una modulación pasiva en la fase,

Figura 2. Sistema óptico de difracción codificado.



por lo tanto, se impone que  $d$  no aumente la energía de la escena durante el proceso de modulación. Así que, se establece la Definición 1 para garantizar la admisibilidad de la variable aleatoria.

**Definición 1.** (*Variable aleatoria admisible*).<sup>32</sup> Una variable aleatoria que cumple con  $|d| \leq 1$ , se considera admisible como elemento de modulación en fase.

Algunos ejemplos de variables aleatorias que satisfacen la Definición 1 se listan en el Cuadro 1. Estas variables aleatorias han sido usados en<sup>32</sup> como elementos de codificación para resolver el problema de recuperación de fase.

Cuadro 1. Ejemplos de codificaciones aleatorias admisibles según la Definición 1.

Elemento de codificación	Probabilidad
$d \in \{0, 1\}$	$\{\frac{1}{2}, \frac{1}{2}\}$
$d \in \{-1, 1\}$	$\{\frac{1}{2}, \frac{1}{2}\}$
$d \in \{-1, 1, -j, j\}$	$\{\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\}$

### 3. ALGORITMOS DE RECUPERACIÓN DE FASE

En esta sección, se presentan los algoritmos comúnmente utilizados para recuperar el campo óptico inicial a partir de medidas cuadráticas codificadas. La literatura ha planteado diferentes algoritmos basados en formulaciones convexas y no convexas como se expone a continuación.

#### 3.1. FORMULACIONES CONVEXAS

Este tipo de algoritmos relajan el problema de recuperación de fase a un problema convexo equivalente.

- **PhaseLift**<sup>13</sup>: este algoritmo plantea el problema de recuperación de fase como una minimización de la traza de la siguiente forma

$$\begin{aligned} & \underset{\mathbf{z} \in \mathbb{C}^n}{\text{minimizar}} && \text{Traza}(\mathbf{z}\mathbf{z}^H), \\ & \text{sujeto a} && \mathcal{A}(\mathbf{z}\mathbf{z}^H) = \mathbf{b}, \\ & && \mathbf{z}\mathbf{z}^H \succeq 0, \end{aligned} \tag{9}$$

donde  $\mathcal{A}(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}^m$  es un operador lineal y  $\text{Traza}(\cdot)$  representa la traza de una matriz.

- **PhaseMax**<sup>14</sup>: sea  $\hat{\mathbf{z}} \in \mathbb{C}^n$  un vector aproximación de la señal original  $\mathbf{z}$ , de modo que, la señal reconstruida se obtiene solucionando el siguiente problema convexo

$$\begin{aligned} & \underset{\mathbf{z} \in \mathbb{C}^n}{\text{maximizar}} && \langle \mathbf{z}, \hat{\mathbf{z}} \rangle_{\mathbb{R}}, \\ & \text{sujeto a} && |\langle \mathbf{a}_i, \mathbf{z} \rangle| \leq (\phi)_i, \end{aligned} \tag{10}$$

donde  $\langle \cdot, \cdot \rangle_{\mathbb{R}}$  denota la parte real del producto interno y  $(\phi)_i = \sqrt{(\mathbf{y})_i}$ .

### 3.2. FORMULACIONES NO CONVEXAS

Las formulaciones no convexas requieren una estimación inicial de la escena o inicialización, además, calculan el gradiente siguiendo la diferenciación de Wirtinger.

- **TRUNCATED WIRTINGER FLOW (TWF)** <sup>16</sup>: este algoritmo se basa en el modelo de muestreo según unas variables aleatorias que siguen una distribución de Poisson, de la forma

$$(\mathbf{y})_i \sim \text{Poisson}(|\langle \mathbf{a}_i, \mathbf{z} \rangle|^2), \quad i = 1, \dots, m. \quad (11)$$

TWF busca minimizar la máxima estimación de probabilidad

$$\underset{\mathbf{z} \in \mathbb{C}^n}{\text{minimizar}} - \sum_{i=1}^m \ell(\mathbf{z}; \mathbf{y}_i), \quad (12)$$

donde  $\ell(\mathbf{z}; \mathbf{y}_i) = \mathbf{y}_i \log(|\mathbf{a}_i^H \mathbf{z}|^2) - |\mathbf{a}_i^H \mathbf{z}|^2$  con  $(\cdot)^H$  el operador de conjugada transpuesta

- **TRUNCATED AMPLITUDE FLOW (TAF)** <sup>17</sup>: este algoritmo adopta un criterio de mínimos cuadrados para recuperar  $\mathbf{z}$  basado en las medidas sin fase  $\mathbf{y}$

$$\underset{\mathbf{z} \in \mathbb{C}^n}{\text{minimizar}} \frac{1}{2m} \sum_{i=1}^m (|\langle \mathbf{a}_i, \mathbf{z} \rangle| - (\phi)_i)^2. \quad (13)$$

Este algoritmo asume que las medidas  $\mathbf{y}_i$  provienen de un sistema gaussiano de la forma  $\mathbf{y}_i \sim \mathcal{N}(|\langle \mathbf{a}_i, \mathbf{z} \rangle|^2, 1)$

- **REWEIGHTED AMPLITUDE FLOW (RAF)** <sup>18</sup>: este algoritmo sigue el criterio de maximizar la estimación de probabilidad de la forma

$$\underset{\mathbf{z} \in \mathbb{C}^n}{\text{minimizar}} - \sum_{i=1}^m \ell(\mathbf{z}; (\boldsymbol{\phi})_i / (\mathbf{y})_i), \quad (14)$$

donde  $\ell(\cdot)$  se puede definir a través de un muestreo con ruido gaussiano donde basado en la amplitud  $\ell(\mathbf{z}; \mathbf{y}_i) = (|\langle \mathbf{a}_i, \mathbf{z} \rangle| - (\boldsymbol{\phi})_i)^2$  o basado en la intensidad  $\ell(\mathbf{z}; (\mathbf{y})_i) = (|\langle \mathbf{a}_i, \mathbf{z} \rangle|^2 - (\mathbf{y})_i)^2$ . por último parte, basado en un muestreo con distribución de Poisson,  $\ell(\mathbf{z}; (\mathbf{y})_i) = (\mathbf{y})_i \log(|\mathbf{a}_i^H \mathbf{z}|^2) - |\mathbf{a}_i^H \mathbf{z}|^2$

## 4. SISTEMAS DE CLASIFICACIÓN

La clasificación ha sido una de las tareas computacionales más abordadas en el estado del arte. Específicamente, dentro de los algoritmos de clasificación tradicionales se encuentran las máquinas de soporte vectorial, K vecinos más cercanos <sup>33</sup> y los enfoques basados en redes neuronales <sup>1,2,3</sup>, los cuales se describen a continuación.

### 4.1. MÁQUINAS DE SOPORTE VECTORIAL

Las máquinas de soporte vectorial (SVM, de su sigla en inglés, *support vector machine*) <sup>34</sup> permiten la clasificación binaria, donde cada punto  $n$ -dimensional  $\mathbf{x}_i$  tiene asociada una etiqueta de clase  $c_i \in \{1, -1\}$ .

Suponiendo que los datos de ambas clases son separables linealmente, este método propone separar los datos usando los hiperplanos descritos en (15). En la Figura 3 se muestra una representación de una SVM en  $\mathbb{R}^2$

$$\begin{aligned} \mathbf{w}\mathbf{x}_i + b &\geq 1 & \text{si } c_i = 1, \\ \mathbf{w}\mathbf{x}_i + b &\leq -1 & \text{si } c_i = -1. \end{aligned} \tag{15}$$

Cabe resaltar que, para todos los elementos del conjunto de datos se cumple que

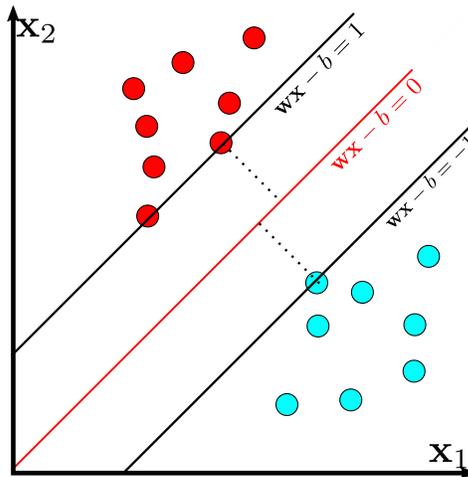
---

<sup>33</sup> JINHO Kim<sup>1</sup>, BS Kim y Silvio Savarese. "Comparing image classification methods: K-nearest-neighbor and support-vector-machines". En: *Proceedings of the 6th WSEAS international conference on Computer Engineering and Applications, and Proceedings of the 2012 American conference on Applied Mathematics*. Vol. 1001. 2012, págs. 48109-2122.

<sup>34</sup> Shan Suthaharan. "Support vector machine". En: *Machine learning models and algorithms for big data classification*. Springer, 2016, págs. 207-235.

$$c_i(\mathbf{w}\mathbf{x}_i + b) \geq 1, \quad i = 1, \dots, m. \quad (16)$$

Figura 3. Representación de una SVM en  $\mathbb{R}^2$ . La recta  $\mathbf{w}\mathbf{x} - b = 0$  en rojo representa el plano óptimo que soluciona el problema de optimización (17). Las rectas  $\mathbf{w}\mathbf{x} - b = 1$  y  $\mathbf{w}\mathbf{x} - b = -1$  representan los extremos de separación de las clases de puntos en rojo y en azul.



El problema de optimización se plantea de la siguiente forma

$$\begin{aligned} & \underset{\mathbf{w}}{\text{minimizar}} \quad \|\mathbf{w}\|, \\ & \text{sujeto a} \quad c_i(\mathbf{w}\mathbf{x}_i + b) \geq 1, \\ & \quad \quad \quad i=1, \dots, m \end{aligned} \quad (17)$$

Usualmente, no es posible separar los datos de forma lineal, por esta razón, se puede incluir una función no lineal  $\phi$  que transforme los datos a un conjunto de características donde las clases sean separables linealmente. El problema de optimización para una SVM usando un kernel  $\phi$  se formula como

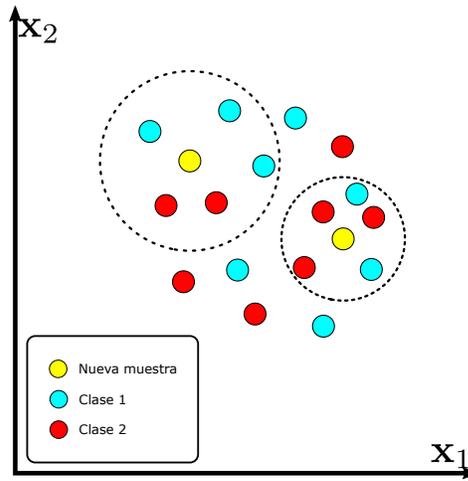
$$\begin{aligned} & \underset{\mathbf{w}}{\text{minimizar}} \quad \|\mathbf{w}\|, \\ & \text{sujeto a} \quad c_i(\mathbf{w}\phi(\mathbf{x}_i) + b) \geq 1, \\ & \quad \quad \quad i=1, \dots, m \end{aligned} \quad (18)$$

## 4.2. K VECINOS MÁS CERCANOS

K vecinos más cercanos (KNN, por sus siglas en inglés, *k-nearest neighbors*.) propone que un conjunto  $D = \{(\mathbf{x}_i, c_i)\}_1^n$ , siendo  $\mathbf{x}_i$  el vector de características y  $c_i$  la clase correspondiente. Para un nuevo vector a clasificar  $\hat{\mathbf{x}}$ , el algoritmo KNN encuentra los  $K$  puntos más cercanos del conjunto de datos. La Figura 4 muestra una representación de la clasificación de dos nuevas muestras en  $\mathbb{R}^2$  con  $K = 5$ . Tradicionalmente, se emplea la distancia euclidiana

$$d(\mathbf{x}_p, \mathbf{x}_q) = \|\mathbf{x}_p - \mathbf{x}_q\|_2. \quad (19)$$

Figura 4. Representación de KNN en  $\mathbb{R}^2$  con  $K = 5$ .



Posteriormente, con base en las clases de los  $K$  puntos encontrados, de tal forma que,  $C \subset D$  y  $|C| = K$ , se cuantifica la cantidad de veces que aparece cada clase y se clasifica la nueva muestra  $\hat{\mathbf{x}}$  con la clase que más veces aparezca de la forma

$$\text{clase}(\hat{\mathbf{x}}) = \arg \max_{\hat{c}} \left\{ \sum_{\hat{c} \in C} \delta(C, \hat{c}) \right\}, \quad (20)$$

donde  $\delta(\cdot, \cdot)$  corresponde a la función delta de Kronecker, dada por

$$\delta(a, b) = \begin{cases} 1, & a = b, \\ 0, & a \neq b, \end{cases} \quad (21)$$

### 4.3. REDES NEURONALES

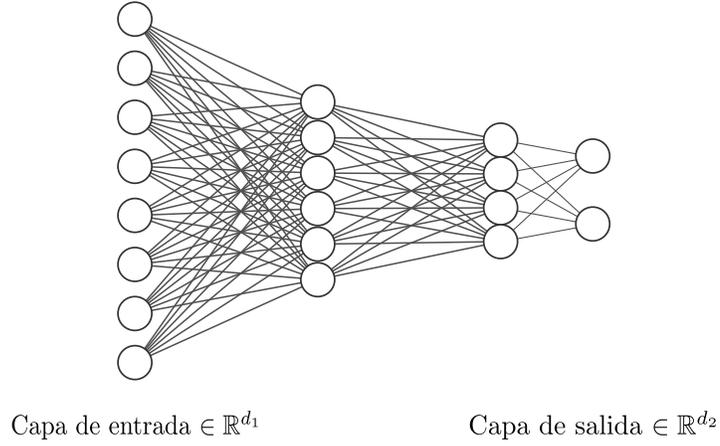
Los enfoques de aprendizaje profundo han generado un gran progreso en el desarrollo de problemas muy complejos en los últimos años <sup>35,1,2,3,36</sup>. El aprendizaje profundo busca encontrar una función  $f : \mathbb{R}^{d_1} \rightarrow \mathbb{R}^{d_2}$  que relacione un dato de entrada con una respectiva salida. La función  $f$  comúnmente se denomina arquitectura de red neuronal profunda, puesto que, consiste en la concatenación de múltiples capas compuestas por unidades mínimas llamadas neuronas. Cada neurona realiza una combinación lineal entre las entradas, posteriormente, se emplea una función no lineal en la salida. Las salidas de cada neurona en una capa funcionan como entrada de las neuronas ubicadas en la siguiente capa, por lo tanto, se construye una arquitectura de red neuronal profunda <sup>4</sup>. En la Figura 5, se muestra una arquitectura de red neuronal  $f : \mathbb{R}^{d_1} \rightarrow \mathbb{R}^{d_2}$ .

---

<sup>35</sup> Kaiming He y col. "Deep residual learning for image recognition". En: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, págs. 770-778.

<sup>36</sup> Ashish Vaswani y col. "Attention is all you need". En: *Advances in neural information processing systems*. 2017, págs. 5998-6008.

Figura 5. Esquema de una arquitectura de red neuronal  $f : \mathbb{R}^{d_1} \rightarrow \mathbb{R}^{d_2}$ .



Por otra parte, el modelo matemático de una red neuronal estándar se puede definir como

$$\{f_{\theta}(\mathbf{x}) = \sigma_J(\mathbf{W}_J \sigma_{J-1}(\mathbf{W}_{J-1}(\dots \sigma_2(\mathbf{W}_2 \sigma_1(\mathbf{W}_1 \mathbf{x} + \mathbf{b}_1) + \mathbf{b}_2)) + \mathbf{b}_{J-1}) + \mathbf{b}_J) \mid \theta = \{\mathbf{W}_j, \mathbf{b}_j\}\}, \quad (22)$$

donde  $1 \leq j \leq J$  representa cada una de las capas que compone la red neuronal,  $\sigma_j$  corresponde a una función no lineal en dicha capa,  $\mathbf{W}_j$  es la matriz de pesos y  $\mathbf{b}_j$  es el bias de la capa. Con el fin de entrenar los pesos  $\theta$  bajo un enfoque de aprendizaje supervisado, se requiere de un conjunto de entrenamiento  $\{(\mathbf{x}_k, c_k)\}_{k=1}^N$  y una función de costo  $\mathcal{L}(c_k, f_{\theta}(\mathbf{x}_k))$  para plantear el problema de optimización

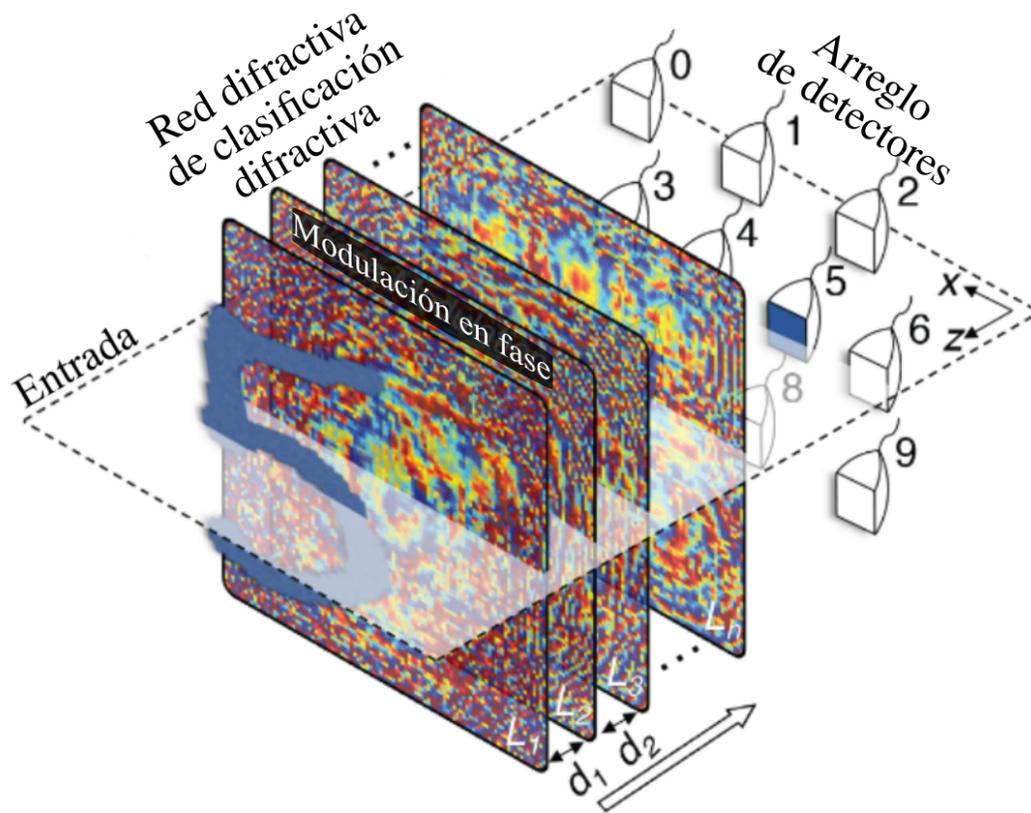
$$\underset{\theta}{\text{minimizar}} \frac{1}{N} \sum_{k=1}^N \mathcal{L}(c_k, f_{\theta}(\mathbf{x}_k)). \quad (23)$$

#### 4.4. CLASIFICACIÓN DE IMÁGENES USANDO REDES NEURONALES ÓPTICAS

Recientemente, se han publicado avances en el campo de redes neuronales implementadas físicamente haciendo uso de propiedades ópticas. En este tipo de redes se destaca la velocidad de la inferencia, debido a que, esta se realiza de manera óptica, es decir, la velocidad de cálculo está limitada por la velocidad de la luz en

el medio y no requiere de una fuente de energía activa para realizar los cálculos. Específicamente, en <sup>37</sup> Xing et al, implementa las capas de una red neuronal haciendo uso de materiales difractivos, realizando la modulación de la luz en fase y aprovechando la propagación del campo óptico para realizar tareas de clasificación.

Figura 6. Implementación de redes neuronales ópticas implementadas físicamente mediante elementos difractivos.



Fuente: Modificado de *All-optical machine learning using diffractive deep neural networks*, 2018.

<sup>37</sup> Xing Lin y col. "All-optical machine learning using diffractive deep neural networks". En: *Science* 361.6406 (2018), págs. 1004-1008.

#### **4.5. CLASIFICACIÓN USANDO MEDIDAS CUADRÁTICAS**

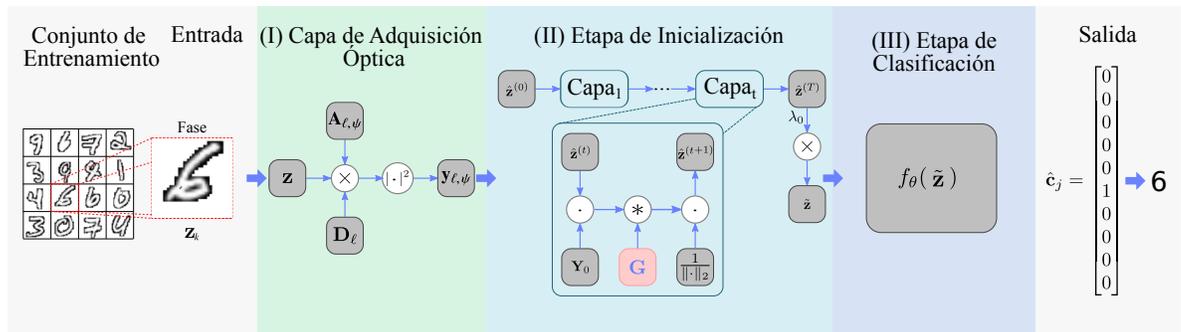
En el campo de tomografía computarizada <sup>20</sup> e imágenes de un solo píxel <sup>21</sup>, se han propuesto sistemas de clasificación que usan únicamente medidas obtenidas a través de sistemas lineales de adquisición, debido al enfoque basado en el aprendizaje profundo en el que la arquitectura incluye una capa con el sistema óptico de adquisición. Asimismo, este enfoque ha sido estudiado en holografía <sup>22</sup> y cristalografía de rayos-x <sup>23</sup>, donde las medidas obtenidas son difícilmente reconocidas por el ojo humano.

Sin embargo, enfoques tradicionales de clasificación no han sido abordados en imágenes de medidas cuadráticas codificadas. Por lo tanto, este trabajo propone el desarrollo de un algoritmo de clasificación de objetos usando medidas cuadráticas codificadas bajo un enfoque de aprendizaje profundo.

## 5. METODOLOGÍA PROPUESTA

El enfoque propuesto de red neuronal profunda para la clasificación de objetos, a partir de medidas cuadráticas codificadas, incluye tres etapas principales: (i) capa de adquisición, (ii) enfoque de inicialización, (iii) y red de clasificación. La Figura 7 ilustra el esquema de red neuronal profunda propuesto. Primero, una capa de adquisición simula el proceso de medición (7), a través del modelado de la propagación del campo óptico. Luego, un procedimiento de inicialización aproxima el campo óptico  $z$ . Finalmente, la red de clasificación infiere la clase correspondiente a cada medida cuadrática codificada.

Figura 7. Esquema propuesto de red neuronal profunda de tres etapas para la clasificación de objetos, a partir de medidas cuadráticas codificadas.



### 5.1. ETAPA DE INICIALIZACIÓN

En la etapa de inicialización del campo óptico, este trabajo aprovecha la matriz de adquisición  $A_{\ell, \psi}$  para crear un mapeo entre las medidas  $y_{\ell, \psi}$  y una aproximación más cercana del campo óptico  $z$ , dada por el desenvolvimiento del algoritmo tra-

dicional de inicialización de filtrado espectral <sup>38</sup> (FSI, de sus siglas en inglés *filtered spectral initialization*), que incluye el aprendizaje de un filtro convolucional <sup>39</sup> (LFSI, de sus siglas en inglés *learning filtered spectral initialization*), es decir que,  $\hat{\mathbf{z}} = \text{LFSI}(\mathbf{A}_{\ell,\psi}, \mathbf{y}_{\ell,\psi})$ .

En concreto, este enfoque aplica un método iterativo para aproximar el campo óptico a partir de medidas cuadráticas codificadas, el cual se resume en el Algoritmo 1. La aproximación dada por este algoritmo corresponde al cálculo de una versión filtrada de los mayores valores propios de la matriz

$$\mathbf{Y}_0 := \frac{1}{\text{card}(\Xi)} \sum_{i \in \Xi} \frac{\mathbf{a}_{i,\psi} \mathbf{a}_{i,\psi}^H}{\|\mathbf{a}_{i,\psi}\|_2^2}, \quad (24)$$

donde  $\mathbf{a}_{i,\psi}$  es la  $i$ -ésima columna de la matriz  $\mathbf{A} = [\mathbf{A}_{1,\psi}, \dots, \mathbf{A}_{L,\psi}]^T$ ,  $\text{card}(\cdot)$  es la cardinalidad del conjunto  $\Xi$  correspondiente al conjunto de índices asignados a los valores más grandes de  $\{(\mathbf{y})_i / \|\mathbf{a}_{i,\psi}\|_2\}$  con  $\mathbf{y} = [\mathbf{y}_{1,\psi}^T, \dots, \mathbf{y}_{L,\psi}^T]^T$ . Encontrar  $\tilde{\mathbf{z}}$ , teniendo en cuenta  $\mathbf{Y}_0$  es posible resolviendo el siguiente problema de optimización

$$\tilde{\mathbf{z}} = \arg \max_{\|\mathbf{z}\|_2=1} \mathbf{z}^H \mathbf{Y}_0 \mathbf{z}. \quad (25)$$

Tenga en cuenta que, la matriz  $\mathbf{Y}_0$  se calcula en la línea 4 del Algoritmo 1. En la línea 6, se aplica el filtrado mediante la operación de convolución  $*$  entre  $\mathbf{Y}_0 \hat{\mathbf{z}}^{(t)}$  y en el kernel  $\mathbf{G}$ . Luego, la aproximación filtrada se normaliza en la línea 7. Finalmente, este algoritmo devuelve el vector complejo  $\tilde{\mathbf{z}}$  en la línea 10, que se multiplica por el factor de escala  $\lambda_0 = \sqrt{\sum_{\ell=1}^L \frac{\|\mathbf{y}_{\ell,\psi}\|_2^2}{nL}}$ .

---

<sup>38</sup> Andrés Jerez, Samuel Pinilla y Henry Arguello. “Fast Target Detection via Template Matching in Compressive Phase Retrieval”. En: *IEEE Transactions on Computational Imaging* 6 (2020), págs. 934-944.

<sup>39</sup> David Morales, Andrés Jerez y Henry Arguello. “Learning spectral initialization for phase retrieval via deep neural networks”. En: *Appl. Opt.* 61.9 (2022), F25-F33. DOI: 10.1364/AO.445085.

Este trabajo propone la inicialización aprendida realizando un desenvolvimiento del algoritmo FSI en un esquema de red neuronal, como se muestra en la Figura 7. Adicionalmente, se incluye el filtro  $\mathbf{G}$  como un parámetro entrenable de la red neuronal bajo una regularización definida a través de la siguiente función indicadora

$$\mathcal{I}_\Omega(\mathbf{G}) = \begin{cases} 0 & \text{si } \mathbf{G} \in \Omega, \\ +\infty & \text{si } \mathbf{G} \notin \Omega, \end{cases} \quad (26)$$

donde  $\Omega = \{\mathbf{G} \in \mathbb{C}^{g \times g} \mid \|(\mathbf{G})_{p,q}\| \leq 1\}$ .

---

### Algoritmo 1 Algoritmo LFSI

---

- 1: **Entrada:** Matriz de sensado y muestras  $\{(\mathbf{A}_{\ell,\psi}; \mathbf{y}_{\ell,\psi})\}_{\ell=1}^L$ , máximo número de iteraciones  $T$ , filtro  $\mathbf{G} \in \mathbb{C}^{g \times g}$ .
  - 2:  $\hat{\mathbf{z}} \leftarrow$  Escogido aleatoriamente.
  - 3: **Asignar**  $\Xi$  como el conjunto de índices correspondientes a los valores más grandes de  $\{(\mathbf{y})_i / \|\mathbf{a}_{i,\psi}\|_2\}$ .
  - 4:
$$\mathbf{Y}_0 := \frac{1}{\text{card}(\Xi_0)} \sum_{i \in \Xi} \frac{\mathbf{a}_{i,\psi} \mathbf{a}_{i,\psi}^H}{\|\mathbf{a}_{i,\psi}\|_2^2}$$
  - 5: **para**  $t = 0 : T - 1$  **hacer**
  - 6:      $\dot{\mathbf{z}}^{(t+1)} = \mathbf{G} * (\mathbf{Y}_0 \hat{\mathbf{z}}^{(t)})$ . ▷ Filtrado
  - 7:      $\hat{\mathbf{z}}^{(t+1)} = \frac{\dot{\mathbf{z}}^{(t+1)}}{\|\dot{\mathbf{z}}^{(t+1)}\|_2}$ . ▷ Normalización
  - 8: **fin para**
  - 9: Compute  $\tilde{\mathbf{z}} = \lambda_0 \hat{\mathbf{z}}^{(T)}$ . ▷ Escalado
  - 10: **salida:** Estimación inicial del campo óptico  $\tilde{\mathbf{z}}$ .
- 

## 5.2. ETAPA DE CLASIFICACIÓN

En este trabajo se propone un esquema de clasificación de medidas cuadráticas codificadas, que incluye una etapa de estimación inicial del campo óptico. De hecho, dentro del esquema de clasificación es posible emplear diferentes arquitecturas de clasificación previamente validadas en la literatura. Específicamente, se utilizaron tres arquitecturas de clasificación del estado del arte: MobilNetV2 <sup>24</sup>, InceptionV3 <sup>25</sup>

y Xception <sup>26</sup>. El Cuadro 2 resume el número de parámetros, la cantidad de capas de las redes neuronales de clasificación empleadas.

Cuadro 2. Resumen de las arquitecturas de redes neuronales de clasificación utilizadas.

<b>Modelo</b>	<b>Número de parámetros</b>	<b>Número de capas</b>
MobileNetV2	3,538,984	53
InceptionV3	23,851,784	48
Xception	22,910,480	71

El entrenamiento de los pesos  $\theta$  de la red de clasificación  $f_\theta(\cdot)$  se realizó mediante el siguiente problema de optimización

$$\theta^* \in \arg \min_{\theta} \frac{1}{\mathcal{K}} \sum_{k=1}^{\mathcal{K}} \mathcal{L}(c^{(k)}, f_\theta(\text{LFSI}(\mathbf{A}_{\ell, \psi}, \mathbf{y}_{\ell, \psi}))). \quad (27)$$

Este problema minimiza la función de pérdida, usualmente conocida como entropía categórica cruzada  $\mathcal{L}(\cdot, \cdot)$ , entre las etiquetas del conjunto de datos  $c^{(k)}$  y las clases predichas  $\hat{c}^{(k)} = f_\theta(\text{LFSI}(\mathbf{A}_{\ell, \psi}, \mathbf{y}_{\ell, \psi}))$ ,  $\mathbf{y}_{\ell, \psi}^{(k)}$  es la medida cuadrática codificada del  $k$ -ésimo ejemplo y  $\mathcal{K}$  es el número total de ejemplos del conjunto de datos.

### 5.3. ESQUEMA PROPUESTO DE CLASIFICACIÓN DE OBJETOS

El Algoritmo 2 resume el proceso de entrenamiento del esquema mostrado en la Figura 7. En primer lugar, en la línea 2 el filtro es inicializado usando una distribución uniforme  $\mathcal{U}(0, 1)$ . Luego, el modelo de propagación descrito por (7) es simulado por la capa de adquisición en la línea 5. Posteriormente, en línea 6 se realiza el proceso de inicialización que aproxima el campo óptico inicial con base en las medidas cuadráticas codificadas obtenidas. En las líneas 7-9, el optimizador Adam  $\mathcal{A}_{dam}(\cdot)$  se implementa para minimizar la función de pérdida (27) entre la clase etiquetada en el conjunto de datos  $c^{(k)}$  y la predicción  $f_\theta(\tilde{z}^{(k)})$ . Esta función se minimiza sobre el filtro de la inicialización y los parámetros de la red neuronal de clasificación pon-

derados por  $\beta_1$  y  $\beta_2$ , respectivamente. Finalmente, estos parámetros óptimos  $G$  y  $\theta$  se devuelven en la línea 12.

---

**Algoritmo 2** Enfoque de clasificación de objetos.

---

- 1: **Entrada:** Conjunto de entrenamiento  $\{\mathbf{z}^{(k)}, \mathbf{c}^{(k)}\}_{k=1}^{\mathcal{K}}$  con  $\mathcal{K}$  imágenes.
  - 2: **Inicialización filtro:**  $G \in \mathcal{U}(0, 1)^{5 \times 5}$
  - 3: **para** época = 1: $\mathcal{E}$  **hacer** ▷  $\mathcal{E}$  épocas
  - 4:     **para**  $k = 1:\mathcal{K}$  **hacer** ▷  $\mathcal{K}$  ejemplos
  - 5:          $\mathbf{y}_{\ell, \psi} = |\mathbf{A}_{\ell, \psi} \mathbf{z}^{(k)}|^2, \quad \ell \in \{1, \dots, L\}$  ▷ Métricas cuadráticas codificadas
  - 6:          $\tilde{\mathbf{z}}^{(k)} \leftarrow \text{LFSI}(\mathbf{A}_{\ell, \psi}, \mathbf{y}_{\ell, \psi})$  ▷ Algoritmo 1
  - 7:          $\mathcal{L}_{G, \theta} = \frac{1}{\mathcal{K}} \sum_{k=1}^{\mathcal{K}} \mathcal{L}(c^{(k)}, f_{\theta}(\tilde{\mathbf{z}}^{(k)}))$  ▷ Función de costo
  - 8:          $G \leftarrow \mathcal{A}_{dam}(G, \beta_1 \nabla_G \mathcal{L}_{G, \theta})$  ▷ Optimización sobre  $G$
  - 9:          $\theta \leftarrow \mathcal{A}_{dam}(\theta, \beta_2 \nabla_{\theta} \mathcal{L}_{G, \theta})$  ▷ Optimización sobre  $\theta$
  - 10:     **fin para**
  - 11: **fin para**
  - 12: **Salida:** Kernel óptimo  $G$  y parámetros de la red neuronal  $\theta$ .
-

## 6. SIMULACIONES Y RESULTADOS

En esta sección se presentan los resultados obtenidos a partir del método propuesto para la clasificación de objetos basado en medidas cuadráticas codificadas y su comparación con esquemas de clasificación del estado del arte.

### 6.1. CONJUNTO DE DATOS

Durante el desarrollo de este trabajo, no se encontraron conjuntos de datos públicos enfocados en la clasificación de medidas cuadráticas. Debido a esto, se decidió simular la propagación de las medidas usando conjuntos de datos tradicionales. Para este fin se usaron imágenes de los conjuntos de datos MNIST <sup>28</sup> y Fashion-MNIST <sup>27</sup>. Por un lado, la Figura 8 muestra un ejemplo de las imágenes presentes en los conjuntos de datos usados. Por otro lado, el Cuadro 3 muestra la división de datos en entrenamiento, validación y prueba. Cada imagen fue escalada en el rango de  $[-\pi, \pi]$  y usada como información de fase de la forma  $\mathbf{z} = e^{j\text{ang}(\mathbf{z})}$ .

Figura 8. Ejemplo de las imágenes presentes en los conjuntos de datos MNIST y Fashion-MNIST.



Cuadro 3. Resumen de la división de los conjuntos de datos usados para evaluar el método propuesto.

Conjunto de datos	Entrenamiento	Validación	Prueba	Total
MNIST	54000	6000	10000	70000
Fashion-MNIST	54000	6000	10000	70000

## 6.2. MÉTRICAS

La calidad de la inicialización y la precisión de la clasificación en sistemas de difracción basados en medidas cuadráticas codificadas, se mide a través de las siguientes métricas.

**6.2.1. Métricas para evaluar la inicialización.** A continuación, se describen las métricas usadas para evaluar la estimación inicial  $\tilde{\mathbf{z}}$  respecto a la imagen de referencia  $\mathbf{z}$ .

- Medida del índice de similitud estructural (SSIM):

$$\text{SSIM}(\mathbf{z}, \tilde{\mathbf{z}}) = \frac{2(\mu_{\mathbf{z}}\mu_{\tilde{\mathbf{z}}} + C_1) + (2\sigma_{\mathbf{z}\tilde{\mathbf{z}}} + C_2)}{(\mu_{\mathbf{z}} + \mu_{\tilde{\mathbf{z}}} + C_1)(\sigma_{\mathbf{z}}\sigma_{\tilde{\mathbf{z}}} + C_1)}, \quad (28)$$

donde  $(\mu_{\mathbf{z}}, \sigma_{\mathbf{z}})$  y  $(\mu_{\tilde{\mathbf{z}}}, \sigma_{\tilde{\mathbf{z}}})$  representan la media y la varianza de  $\mathbf{z}$  y  $\tilde{\mathbf{z}}$  respectivamente, además  $\sigma_{\mathbf{z}\tilde{\mathbf{z}}}$  es la covarianza entre ellos. Por último,  $C_1 = (k_1L)^2$  y  $C_2 = (k_2L)^2$  son dos constantes para prevenir división por cero y  $k_1 = 0,01$  y  $k_2 = 0,03$ . Esta métrica tiene valores en el rango de  $[0, 1]$ , es decir, a mayor valor, mejor es la estimación.

- Relación de señal a ruido máxima (PSNR):

$$\text{PSNR}(\mathbf{z}, \tilde{\mathbf{z}}) = 10 \cdot \log_{10} \left( \frac{\text{MAX}^2}{\text{MSE}(\mathbf{z}, \tilde{\mathbf{z}})} \right), \quad (29)$$

donde  $\text{MAX}(\cdot)$  es el rango dinámico de la imagen y  $\text{MSE}(\cdot)$  es el error cuadrá-

tico medio entre la imagen original y la estimación. Esta métrica tiene valores en el rango de  $[0, \infty)$ , es decir, a mayor valor, mejor es la estimación.

- Error cuadrático medio (MSE):

$$\text{MSE}(\mathbf{z}, \tilde{\mathbf{z}}) = \frac{1}{n} \|\mathbf{z} - \tilde{\mathbf{z}}\|_2^2. \quad (30)$$

Esta métrica tiene valores en el rango de  $[0, \infty)$ , es decir, a menor valor, mejor es la estimación.

- Error relativo (RE):

$$\text{RE}(\mathbf{z}, \hat{\mathbf{z}}) = \frac{\text{dist}(\mathbf{z}, \hat{\mathbf{z}})}{\|\mathbf{z}\|_2}. \quad (31)$$

Esta métrica tiene valores en el rango de  $[0, \infty)$ , donde a menor valor, mejor es la estimación.

**6.2.2. Métricas para evaluar la clasificación.** Las métricas usadas para evaluar la clasificación se presentan a continuación, donde los valores a calcular dependen de los verdaderos positivos  $TP$ , verdaderos negativos  $TN$ , falsos positivos  $FP$ , y falsos negativos  $FN$ , asociados a la clase de cada ejemplo del conjunto de datos.

- Exactitud: esta medida corresponde a la proporción de predicciones predichas correctamente,

$$\text{Exactitud} = \frac{TP + TN}{TP + TN + FP + FN}. \quad (32)$$

- Precisión: esta medida es la proporción de ejemplos pertenecientes a una clase que fueron correctamente predichos en dicha clase,

$$\text{Precisión} = \frac{TP}{TP + FP}. \quad (33)$$

- Exhaustividad: esta medida es la proporción entre las predicciones que fueron correctamente clasificadas en una clase sobre los ejemplos que pertenecían a esa clase,

$$\text{Exhaustividad} = \frac{TP}{TP + FN}. \quad (34)$$

- Métrica-F1: esta medida está diseñada para ponderar de manera equilibrada entre las métricas de precisión (33) y exhaustividad (34),

$$F1 = \frac{2 \cdot \text{Precisión} \cdot \text{Exhaustividad}}{\text{Precisión} + \text{Exhaustividad}}. \quad (35)$$

### 6.3. EXPERIMENTOS

**6.3.1. Configuración experimental.** En esta subsección, se presentan los parámetros de simulación utilizados para el modelo de propagación, la etapa de inicialización y la red neuronal de clasificación. Todos los experimentos realizados en este trabajo fueron implementados en *Python 3.9* usando la herramienta *Tensorflow 2.4.1*<sup>40</sup>. Los experimentos se ejecutaron sobre un computador con GPU Nvidia 3090 RTX con 64 Gb de memoria RAM y CPU Intel(R) Xeon(R) W-3223 CPU @ 3.50GHz. El código usado para los experimentos de este trabajo puede ser consultado de manera pública en<sup>41</sup>. En relación con los parámetros de propagación, el Cuadro 4 presenta los parámetros ópticos fijados para calcular la matriz  $A_{\ell, \psi}$  definida en (7) según cada campo de difracción. Estos parámetros permiten simular la adquisición de las medidas cuadráticas codificadas a lo largo de cada campo de

---

<sup>40</sup> Martín Abadi y col. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015.

<sup>41</sup> [https://github.com/David-Morales-Norato/code\\_undergrad\\_thesis](https://github.com/David-Morales-Norato/code_undergrad_thesis)

difracción usando la capa de adquisición óptica. En la etapa de inicialización, el filtro  $\mathbf{G} \in \mathbb{C}^{g \times g}$  entrenable fue fijado a un tamaño de kernel  $g = 5$  y un número de iteraciones  $T = 10$ . La estimación del campo óptico inicial es una tarea ampliamente estudiada en el estado del arte de recuperación de la fase. Específicamente, se destacan métodos de inicialización, tales como *orthogonality-promoting initialization* (OPI) <sup>17</sup>, *weighted maximal correlation initialization* (WMCI) <sup>18</sup>, y FSI <sup>38</sup>. Cabe señalar que, estos métodos de inicialización fueron usados para comparar la inicialización propuesta LFSI, donde se comparó la robustez al ruido, el número de iteraciones  $T$  y el número de proyecciones  $L$  requeridas para lograr una estimación apropiada. Finalmente, para realizar la clasificación se usaron las arquitecturas MobilNetV2 <sup>24</sup>, InceptionV3 <sup>25</sup> y Xception <sup>26</sup>, las cuales fueron entrenadas usando una tasa de aprendizaje de  $1 \times 10^{-3}$  con el algoritmo de optimización Adam a partir de  $\mathcal{E} = 100$  épocas para entrenar los conjuntos de datos MNIST y Fashion-MNIST. Para comparar el esquema de clasificación propuesto se implementó el esquema de clasificación tradicional <sup>22,23</sup>, donde se realiza únicamente el proceso de simulación de las medidas cuadráticas sin el proceso de estimación del campo óptico. Asimismo, se empleó el método propuesto de filtrado espectral aprendido en las Figuras 12, 13 y 14. Por otra parte, se reemplazó en la etapa de inicialización del esquema de clasificación propuesto, el algoritmo de inicialización LFSI por el modelo de propagación inverso (BPM, de sus siglas en inglés *back-propagation matrix*)  $\mathbf{A}_{\ell, \psi}^{-1}$  para obtener una estimación rápida del campo óptico inicial, como se presenta en las Figuras 12, 13 y 14. Esta estimación rápida está dada por

$$\hat{\mathbf{z}} = \frac{1}{L} \sum_{\ell=1}^L \mathbf{A}_{\ell, \psi}^{-1} \mathbf{y}_{\ell, \psi}. \quad (36)$$

Es importante señalar que,  $\mathbf{z} \neq \mathbf{A}_{\ell, \psi}^{-1} \mathbf{y}_{\ell}$ , debido a la pérdida de la información de fase que induce el operador de magnitud  $|\cdot|$  presente en el modelo de propagación (7).

Cuadro 4. Parámetros de propagación usados para simular el modelo de propagación (7) para cada campo de difracción.

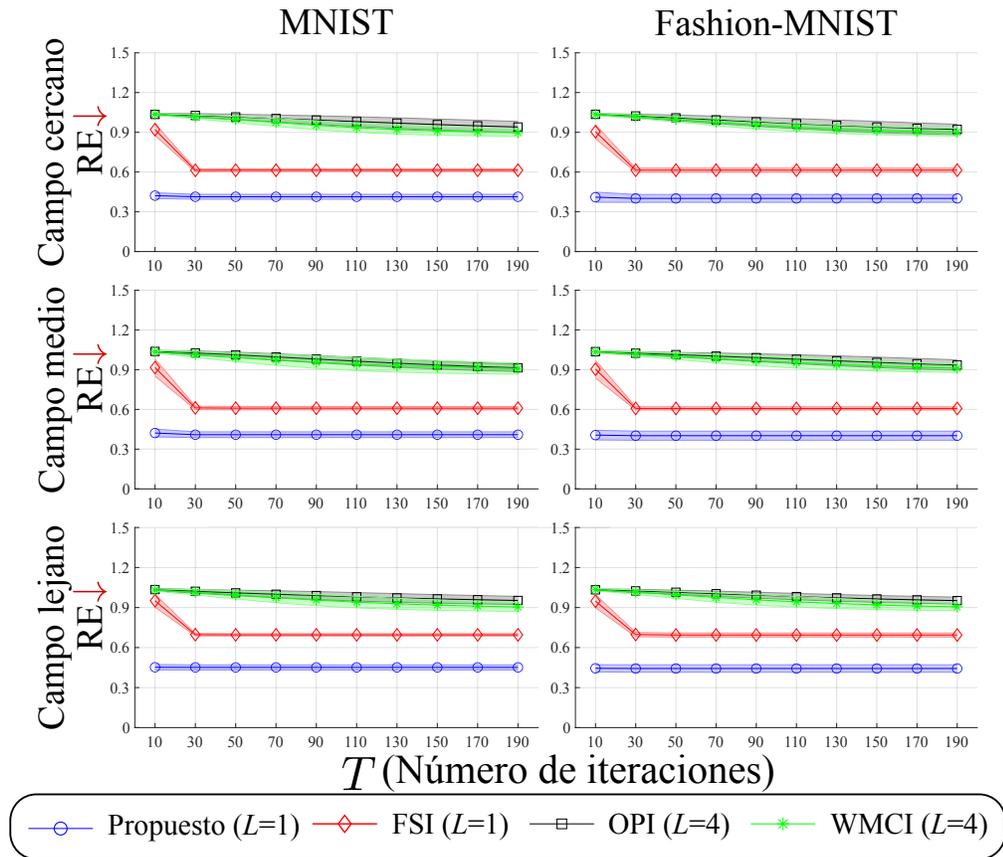
Parámetros ópticos	Campo cercano	Campo medio	Campo lejano
Longitud de onda ( $\lambda$ ) [nm]	635	635	-
Distancia de propagación ( $z$ ) [cm]	2.5	7	-

**6.3.2. Resultados de la inicialización.** La Figura 9 muestra el desempeño en la métrica de RE del método de estimación inicial propuesto comparando con los algoritmos FSI, OPI y WMCI variando el número de iteraciones  $T$  para los conjuntos de datos MNIST y Fashion-MNIST en los tres campos de difracción cercano, medio y lejano, donde se puede observar que el método propuesto logra superar los métodos tradicionales obteniendo una estimación inicial apropiada a partir de un número menor de iteraciones.

La Figura 10 muestra la evaluación de la robustez del método frente al ruido. En este caso, se adicionó ruido aditivo gaussiano con  $\text{SNR} = 10, 15$  y  $30$  dB. Cabe resaltar que, el método propuesto muestra una mayor estabilidad al ruido comparando con los métodos de inicialización del estado del arte.

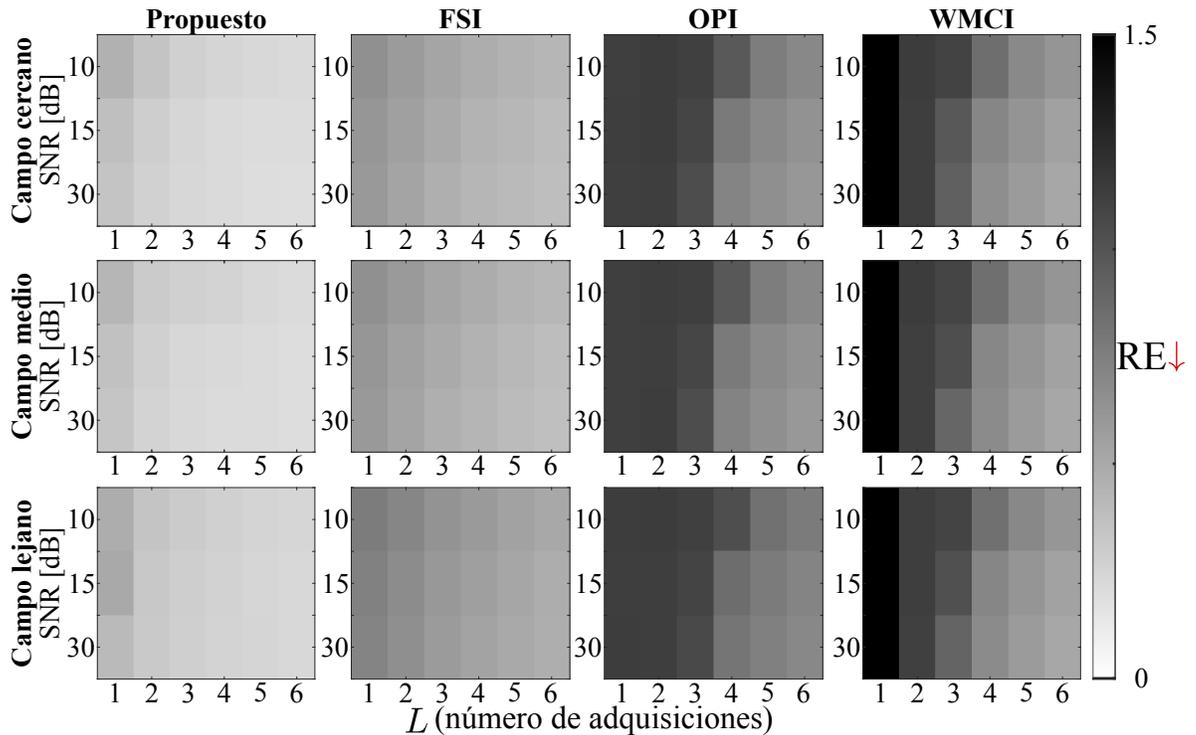
Por último, la Figura 11 presenta los resultados visuales de la estimación del campo inicial, donde se observa en escala de grises la fase de cada estimación para el conjunto de datos MNIST y Fashion-MNIST. Para comparar el método propuesto se implementaron los algoritmos FSI con  $L = 1, T = 10$ , FSI usando  $L = 1, T = 200$ , OPI  $L = 4, T = 200$  y WMCI  $L = 4, T = 200$ . Se puede observar que, el método propuesto logra aproximar el campo óptico en los tres campos de difracción simulados a un bajo número de iteraciones utilizando una única proyección  $L = 1$  y  $T = 10$  número de iteraciones.

Figura 9. Resumen del desempeño de la inicialización propuesta comparando con los métodos FSI, OPI, WMCI sobre la métrica de RE, variando el número de iteraciones del algoritmo para estimar imágenes de los conjuntos MNIST y Fashion-MNIST en los campos de difracción cercano, medio y lejano.



**6.3.3. Resultados de la clasificación de objetos.** Los resultados de los experimentos de clasificación se resumen a continuación. Las Figuras 12, 13 y 14 exhiben los resultados del desempeño de la clasificación de medidas cuadráticas codificadas en términos de las métricas de exactitud, precisión, exhaustividad y F1 utilizando las redes de clasificación MobilNetV2 en la Figura 12, InceptionV3 en la Figura 13 y Xception en la Figura 14. Cada figura representa la comparación de los esquemas de clasificación tradicional, y los métodos propuestos BPM y LFSI. La clasificación se realizó sobre medidas simuladas en los campos de difracción cercano, medio

Figura 10. Resumen del desempeño de la inicialización comparando con diferentes estrategias del estado del arte. Se varió el número de medidas  $T$  para realizar la estimación y diferentes niveles de ruido afectando las medidas descritos por el SNR.



y lejano. Es importante resaltar que, el método tradicional muestra un desempeño menor en todas las métricas comparando con los métodos propuestos BPM y LFSI para cada uno de los campos de difracción y redes de clasificación.

Específicamente, el método propuesto supera al método tradicional en hasta 0.24, 0.2, 0.25 y 0.22 en términos de exactitud, precisión, exhaustividad y métrica F1, respectivamente. En particular, estas ganancias se reportan para la clasificación de objetos utilizando medidas simuladas del campo lejano a partir del conjunto de datos de Fashion-MNIST. En el caso del campo medio, se supera en 0.21, 0.13, 0.21, y 0.17 y para el campo cercano, se supera en 0.1, 0.14, 0.13, y 0.14 en exactitud, en precisión, en exhaustividad y en F1, respectivamente.

Figura 11. Resultados visuales de la estimación del campo óptico inicial comparando el método propuesto contra el algoritmo FSI con  $L = 1, T = 10$ , FSI usando  $L = 1, T = 200$ , OPI  $L = 4, T = 200$  y WMCI  $L = 4, T = 200$ .

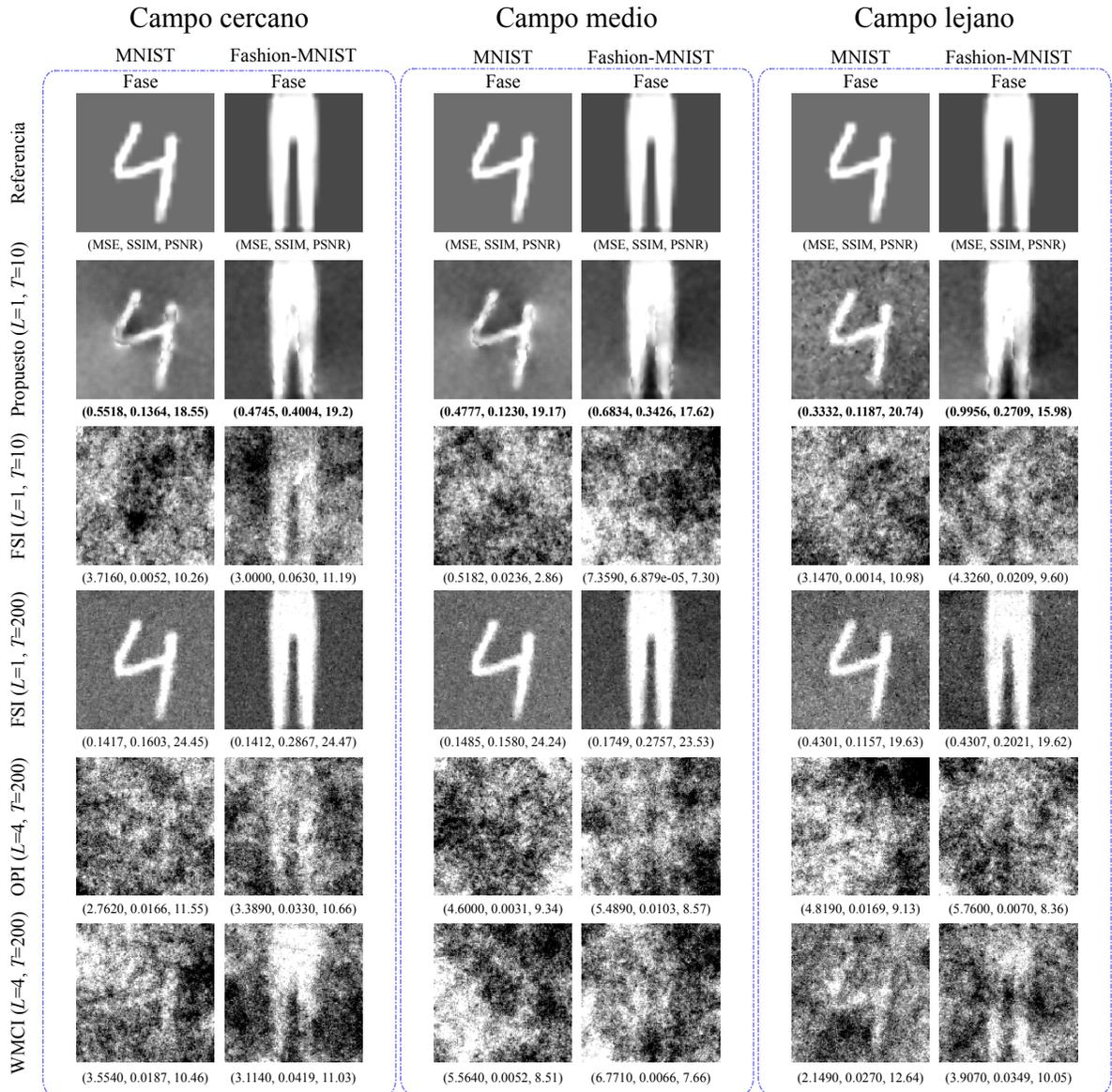


Figura 12. Resultados en la métrica de exactitud, precisión, exhaustividad y F1 para evaluar la clasificación de medidas cuadráticas codificadas simuladas sobre el campo cercano, medio y lejano. Se muestra la comparación de la clasificación Tradicional, el uso del propagador inverso propuesto (BPM) y el método de inicialización aprendida (LFSI) usando la arquitectura MobilNetV2, sobre la división de prueba de los conjuntos de datos MNIST y Fashion-MNIST.

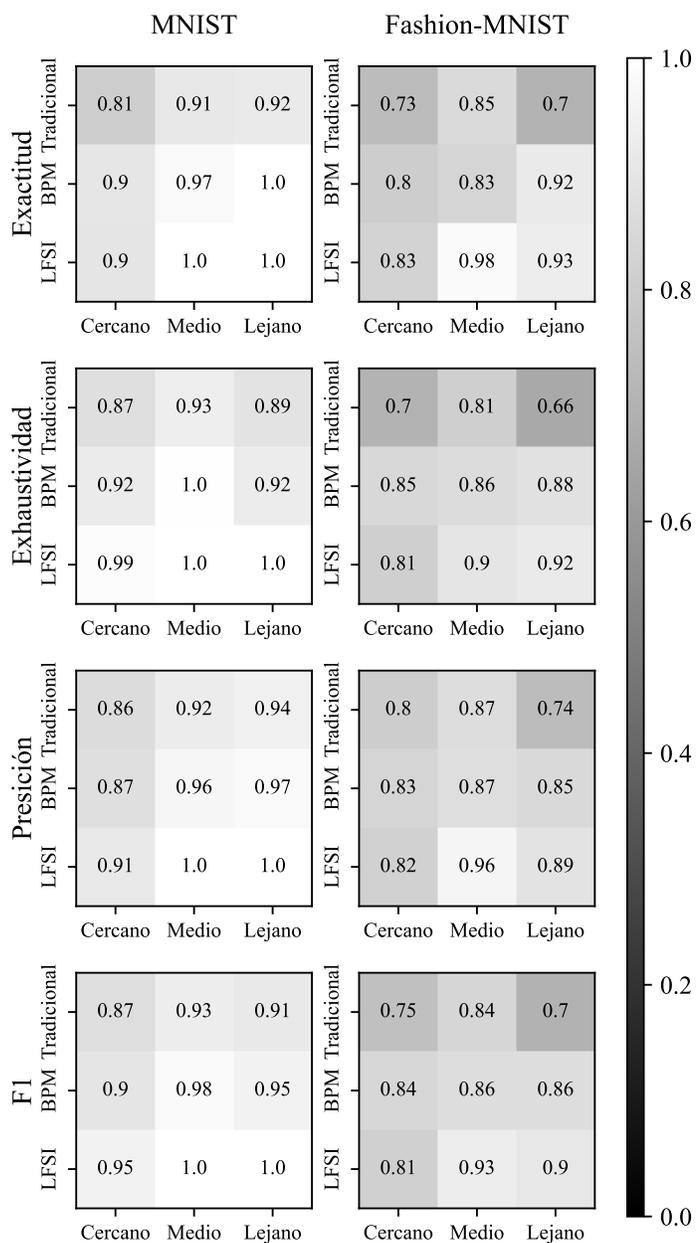


Figura 13. Resultados en la métrica de exactitud, precisión, exhaustividad y F1 para evaluar la clasificación de medidas cuadráticas codificadas simuladas sobre el campo cercano, medio y lejano. Se muestra la comparación de la clasificación Tradicional, el uso del propagador inverso propuesto (BPM) y el método de inicialización aprendida (LFSI) usando la arquitectura InveptionV3, sobre la división de prueba de los conjuntos de datos MNIST y Fashion-MNIST.

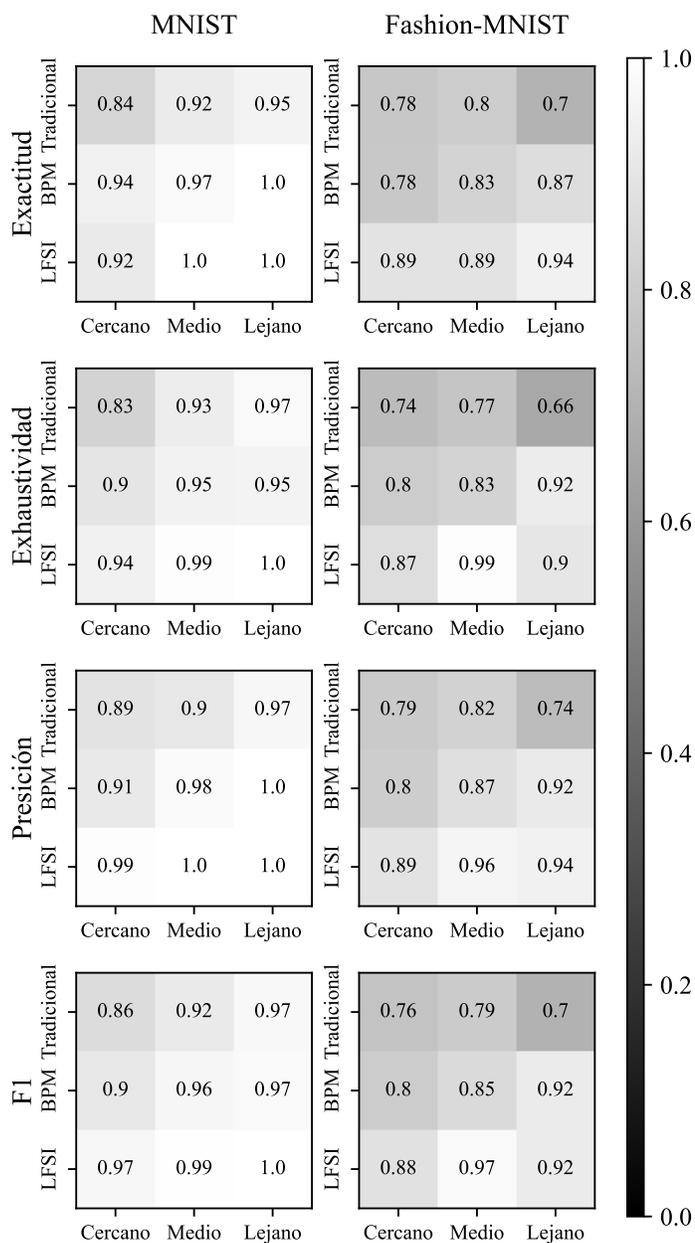
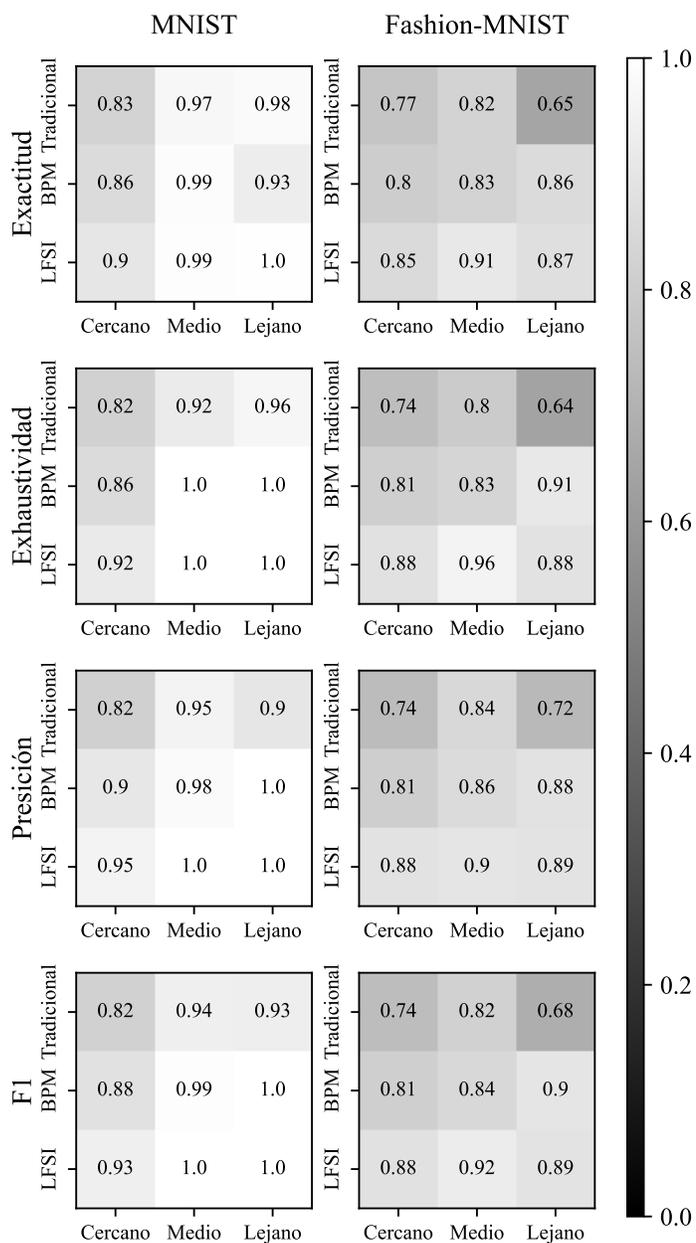


Figura 14. Resultados en la métrica de exactitud, precisión, exhaustividad y F1 para evaluar la clasificación de medidas cuadráticas codificadas simuladas sobre el campo cercano, medio y lejano. Se muestra la comparación de la clasificación Tradicional, el uso del propagador inverso propuesto (BPM) y el método de inicialización aprendida (LFSI) usando la arquitectura Xception, sobre la división de prueba de los conjuntos de datos MNIST y Fashion-MNIST.



## 7. CONCLUSIONES

En este trabajo se propuso un esquema de tres etapas para la clasificación de medidas cuadráticas codificadas utilizando aprendizaje profundo. Primero, una capa simula el proceso de adquisición. Teniendo en cuenta el modelo matemático del proceso físico, esta capa contiene un modelo de propagación del campo óptico codificado siguiendo tres diferentes campos de difracción, dando cumplimiento a los objetivos específicos 1 y 3. Luego, un procedimiento de inicialización aproxima el campo óptico inicial mediante el desenvolvimiento de un algoritmo tradicional y el aprendizaje de un kernel convolucional. Por último, la red de clasificación infiere la clase correspondiente a cada medida cuadrática codificada según la estimación inicial, desarrollando con éxito el objetivo general y el objetivo específico 2 de este trabajo. Para lograr el objetivo específico 4 y validar el método propuesto, se realizaron experimentos sobre diferentes campos de difracción y redes de clasificación del estado del arte, tales como MobilNetV2, InceptionV3 y Xception. Adicionalmente, se emplearon los conjuntos de datos MNIST y Fashion-MNIST para evaluar los diferentes esquemas de aprendizaje profundo.

Los experimentos realizados exhiben un mejor rendimiento de clasificación usando el método propuesto comparado con el enfoque de clasificación tradicional que evalúa directamente las medidas adquiridas sin tener en cuenta el modelo físico. En particular, el método propuesto supera al método tradicional en hasta 0.24, 0.2, 0.25 y 0.22 en términos de exactitud, precisión, exhaustividad y métrica F1, respectivamente. Estas ganancias se reportan para la clasificación de objetos utilizando medidas simuladas del campo lejano a partir del conjunto de datos de Fashion-MNIST.

## 8. TRABAJO FUTURO

El método propuesto puede llegar a ser validado experimentalmente mediante la implementación de un sistema óptico de adquisición de medidas cuadráticas codificadas. Adicionalmente, en este trabajo se realizó la inclusión de una capa que simula el proceso de adquisición de las medidas, por lo tanto, se abre la posibilidad de introducir una máscara de codificación  $D_\ell$  entrenable dentro del esquema de red neuronal, con el objetivo de determinar una codificación óptima que mejore el proceso de clasificación.

## BIBLIOGRAFÍA

- Abadi, Martín y col. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015 (vid. pág. 43).
- Bacca, Jorge, Laura Galvis y Henry Arguello. “Coupled deep learning coded aperture design for compressive image classification”. En: *Optics express* 28.6 (2020), págs. 8528-8540 (vid. págs. 16, 34).
- Bacca, Jorge, Tatiana Gelvez y Henry Arguello. “Transmittance Regularizer for Binary coded Aperture Design in a Computational Imaging end-to-end Approach”. En: *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2021, págs. 1470-1474 (vid. pág. 16).
- Bui, Hieu Minh y col. “Using grayscale images for object recognition with convolutional-recursive neural network”. En: *2016 IEEE Sixth International Conference on Communications and Electronics (ICCE)*. IEEE. 2016, págs. 321-325 (vid. pág. 13).
- Candes, Emmanuel J, Xiaodong Li y Mahdi Soltanolkotabi. “Phase retrieval from coded diffraction patterns”. En: *Applied and Computational Harmonic Analysis* 39.2 (2015), págs. 277-299 (vid. pág. 14).
- Candès, Emmanuel J., Thomas Strohmer y Vladislav Voroninski. “PhaseLift: Exact and Stable Signal Recovery from Magnitude Measurements via Convex Programming”. En: *Communications on Pure and Applied Mathematics* 66.8 (nov. de 2012), págs. 1241-1274. DOI: 10.1002/cpa.21432 (vid. págs. 15, 25).

- Candès, Emmanuel J., Xiaodong Li y Mahdi Soltanolkotabi. "Phase Retrieval via Wirtinger Flow: Theory and Algorithms". En: *IEEE Transactions on Information Theory* 61.4 (2015), págs. 1985-2007. DOI: 10.1109/TIT.2015.2399924 (vid. pág. 15).
- Chen, Yuxin y Emmanuel J Candès. "Solving random quadratic systems of equations is nearly as easy as solving linear systems". En: *Communications on pure and applied mathematics* 70.5 (2017), págs. 822-883 (vid. págs. 15, 26).
- Chollet, Francois. "Xception: Deep Learning With Depthwise Separable Convolutions". En: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017 (vid. págs. 17, 38, 44).
- Deng, Li. "The mnist database of handwritten digit images for machine learning research [best of the web]". En: *IEEE Signal Processing Magazine* 29.6 (2012), págs. 141-142 (vid. págs. 17, 40).
- Douarre, Clément y col. "On the value of CTIS imagery for neural-network-based classification: a simulation perspective". En: *Appl. Opt.* 59.28 (2020), págs. 8697-8710. DOI: 10.1364/AO.394868 (vid. págs. 16, 34).
- Eldar, Yonina C y Shahar Mendelson. "Phase retrieval: Stability and recovery guarantees". En: *Applied and Computational Harmonic Analysis* 36.3 (2014), págs. 473-494 (vid. pág. 22).
- Fan, Jianqing, Cong Ma y Yiqiao Zhong. "A selective overview of deep learning". En: *Statistical Science* 36.2 (2021), págs. 264-290 (vid. págs. 13, 31).
- Fienup, C y J Dainty. "Phase retrieval and image reconstruction for astronomy". En: *Image recovery: theory and application* 231 (1987), pág. 275 (vid. págs. 14, 20).

- Goldstein, Tom y Christoph Studer. "Phasemax: Convex phase retrieval via basis pursuit". En: *IEEE Transactions on Information Theory* 64.4 (2018), págs. 2675-2689 (vid. págs. 15, 25).
- Goodman, Joseph W. *Introduction to Fourier optics*. Roberts y Company Publishers, 2005 (vid. pág. 14).
- Gross, David, Felix Kraemer y Richard Kueng. "Improved recovery guarantees for phase retrieval from coded diffraction patterns". En: *Applied and Computational Harmonic Analysis* 42.1 (2017), págs. 37-64 (vid. pág. 22).
- Guerrero, Andrés, Samuel Pinilla y Henry Arguello. "Phase Recovery Guarantees From Designed Coded Diffraction Patterns in Optical Imaging". En: *IEEE Transactions on Image Processing* 29 (2020), págs. 5687-5697. DOI: 10.1109/TIP.2020.2985208 (vid. págs. 22, 24).
- He, Kaiming y col. "Deep residual learning for image recognition". En: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, págs. 770-778 (vid. pág. 31).
- Jerez, Andrés, Samuel Pinilla y Henry Arguello. "Fast Target Detection via Template Matching in Compressive Phase Retrieval". En: *IEEE Transactions on Computational Imaging* 6 (2020), págs. 934-944 (vid. págs. 36, 44).
- Kim, Sung-Jin y col. "Deep transfer learning-based hologram classification for molecular diagnostics". En: *Scientific reports* 8.1 (2018), págs. 1-12 (vid. págs. 16, 34, 44).
- Kim<sup>1</sup>, JINHO, BS Kim y Silvio Savarese. "Comparing image classification methods: K-nearest-neighbor and support-vector-machines". En: *Proceedings of the 6th*

- WSEAS international conference on Computer Engineering and Applications, and Proceedings of the 2012 American conference on Applied Mathematics*. Vol. 1001. 2012, págs. 48109-2122 (vid. pág. 28).
- Krizhevsky, Alex, Ilya Sutskever y Geoffrey E Hinton. "ImageNet classification with deep convolutional neural networks". En: *Communications of the ACM* 60.6 (2017), págs. 84-90 (vid. pág. 13).
- Li, Shutao y col. "Deep learning for hyperspectral image classification: An overview". En: *IEEE Transactions on Geoscience and Remote Sensing* 57.9 (2019), págs. 6690-6709 (vid. págs. 13, 28, 31).
- Li, Ying y col. "Deep learning for remote sensing image classification: A survey". En: *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 8.6 (2018), e1264 (vid. págs. 13, 28, 31).
- Lin, Xing y col. "All-optical machine learning using diffractive deep neural networks". En: *Science* 361.6406 (2018), págs. 1004-1008 (vid. pág. 33).
- Morales, David, Andrés Jerez y Henry Arguello. "Learning spectral initialization for phase retrieval via deep neural networks". En: *Appl. Opt.* 61.9 (2022), F25-F33. DOI: 10.1364/AO.445085 (vid. pág. 36).
- Pinilla, Samuel, Juan Poveda y Henry Arguello. "Coded diffraction system in X-ray crystallography using a boolean phase coded aperture approximation". En: *Optics Communications* 410 (2018), págs. 707-716 (vid. págs. 14, 20).
- Poon, Ting-Chung y Jung-Ping Liu. *Introduction to modern digital holography: with MATLAB*. Cambridge University Press, 2014 (vid. pág. 21).

- Rivenson, Yair y col. "Phase recovery and holographic image reconstruction using deep learning in neural networks". En: *Light: Science & Applications* 7.2 (2018), págs. 17141-17141 (vid. págs. 14, 20).
- Sandler, Mark y col. "MobileNetV2: Inverted Residuals and Linear Bottlenecks". En: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2018 (vid. págs. 17, 37, 44).
- Shechtman, Yoav y col. "Phase retrieval with application to optical imaging: a contemporary overview". En: *IEEE signal processing magazine* 32.3 (2015), págs. 87-109 (vid. pág. 14).
- Suthaharan, Shan. "Support vector machine". En: *Machine learning models and algorithms for big data classification*. Springer, 2016, págs. 207-235 (vid. pág. 28).
- Szegedy, Christian y col. "Rethinking the Inception Architecture for Computer Vision". En: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016 (vid. págs. 17, 37, 44).
- Vaswani, Ashish y col. "Attention is all you need". En: *Advances in neural information processing systems*. 2017, págs. 5998-6008 (vid. pág. 31).
- Wang, Gang, Georgios B Giannakis y Yonina C Eldar. "Solving systems of random quadratic equations via truncated amplitude flow". En: *IEEE Transactions on Information Theory* 64.2 (2017), págs. 773-794 (vid. págs. 15, 26, 44).
- Wang, Gang y col. "Phase retrieval via reweighted amplitude flow". En: *IEEE Transactions on Signal Processing* 66.11 (2018), págs. 2818-2833 (vid. págs. 15, 26, 44).

Wang, Wei y col. "Development of convolutional neural network and its application in image classification: a survey". En: *Optical Engineering* 58.4 (2019), pág. 040901 (vid. págs. 13, 28, 31).

Xiao, Han, Kashif Rasul y Roland Vollgraf. "Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms". En: *arXiv preprint arXiv:1708.07747* (2017) (vid. págs. 17, 40).

Ziletti, Angelo y col. "Insightful classification of crystal structures using deep learning". En: *Nature communications* 9.1 (2018), págs. 1-10 (vid. págs. 16, 34, 44).