PHASE MASK DESIGN FOR DIFFRACTIVE OPTICAL IMAGING

ANDRÉS FELIPE JEREZ ARIZA
M.S. in Applied Mathematics

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FISICOMECÁNICAS
ESCUELA DE INGENIERÍAS
ELÉCTRICA, ELECTRÓNICA Y TELECOMUNICACIONES
DOCTORADO EN INGENIERÍA, ÁREA INGENIERÍA ELECTRÓNICA
BUCARAMANGA
2024

PHASE MASK DESIGN FOR DIFFRACTIVE OPTICAL IMAGING

ANDRÉS FELIPE JEREZ ARIZA
M.S. in Applied Mathematics

Doctoral thesis to qualify for the title of
Doctor of Philosophy in Engineering

Advisor
Henry Arguello Fuentes
Ph.D. in Electrical and Computer Engineering

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FISICOMECÁNICAS
ESCUELA DE INGENIERÍAS
ELÉCTRICA, ELECTRÓNICA Y TELECOMUNICACIONES
DOCTORADO EN INGENIERÍA, ÁREA INGENIERÍA ELECTRÓNICA
BUCARAMANGA
2024

DEDICATION

Dedico con todo mi corazón este trabajo a las personas fundamentales en mi vida:

A mi amada madre, Luz Dary, quien ha sido el pilar esencial de mi existencia, proporcionando un apoyo incondicional, una fuente constante de motivación y un amor inquebrantable a lo largo de toda mi vida. Sus bendiciones han guiado cada uno de mis pasos. Mis logros siempre serán también suyos.

A mis adorados hermanos, Santi, Daniel, Julián, y Mile. A su lado, he aprendido a desempeñar el papel de hermano mayor y a comprender la responsabilidad que ello conlleva. Su devoción ha ayudado a fortalecer mi confianza y a reconocer el valor de mi opinión. Recuerden que nunca estarán solos.

A mis queridos amigos, Diego, Leidy, Angélica, Camilo, Mafe, Karen, Márgareth, y Kareth. Su permanente ánimo, compañía, y cariño han sido vitales en los desafíos que he enfrentado. Juntos, hemos compartido risas, lágrimas y experiencias que han consolidado nuestro vínculo. Siempre formarán parte integral de mi vida.

A mi compañero de vida, Leonardo, quien se ha convertido en mi refugio de felicidad y seguridad. Su amor ha iluminado mis momentos más oscuros, brindándome tranquilidad y estabilidad en esta etapa crucial de mi vida. Anhelo profundamente escribir el resto de mi historia a su lado.

A cada uno de ustedes, les agradezco infinitamente por estar a mi lado, celebrando mis triunfos y respaldando mis batallas. Su presencia ha enriquecido mi camino y ha otorgado significado a cada logro alcanzado y a cada desafío superado. Este trabajo no solo es un testimonio de mis esfuerzos, sino también un tributo a la influencia que cada uno de ustedes ha tenido en mi vida.

ACKNOWLEDGEMENTS

I am deeply grateful for the intellectual challenges and valuable insights shaping my research. This research would not have been possible without the contributions and encouragement of several people.

I express my gratitude to the Universidad Industrial de Santander and the High Dimensional Signal Processing research group for creating conducive academic and research environments and supplying essential resources for this study. I extend my appreciation to my advisor for his guidance throughout my doctoral journey. I also express my sincere gratitude to the dissertation committee members for their valuable feedback that enriched the quality of this work.

My heartfelt thanks also go to my family and friends, who have been my pillars of strength and resilience. Thank you for your unwavering encouragement, understanding, and endearment. To my dear Love, thank you for the joyous moments that provided a much-needed balance to the academic demands. Thank you for being my haven, accompanying me through every step, and celebrating this academic achievement alongside me.

This dissertation symbolizes the culmination of years of effort and collaboration. I will be eternally grateful for the invaluable contributions of everyone who played a key role in bringing this project to fruition.

Last but not least, I want to thank me for believing in me, doing all this hard work,
having no days off, never quitting, and just being me at all times.

CONTENTS

page

LIST OF FIGURES

LIST OF TABLES

LIST OF ALGORITHMS

RESUMEN

TÍTULO: DISEÑO DE MÁSCARAS DE FASE PARA IMÁGENES ÓPTICAS DIFRACTIVAS [*]

AUTOR: ANDRÉS FELIPE JEREZ ARIZA [**]

PALABRAS CLAVE: IMÁGENES DIFRACTIVAS, MÁSCARAS DE FASE MULTINIVEL, OPTIMIZACIÓN DE EXTREMO A EXTREMO, RECUPERACIÓN DE FASE Y CLASIFICACIÓN ESPECTRAL.

DESCRIPCIÓN: Los sistemas de imágenes ópticas difractivas (DOI), que se basan en máscaras de fase multinivel (MPMs), codifican el frente de onda del campo óptico antes de ser registrado por el sensor. Esto resulta en patrones de difracción codificados (CDPs) donde se pierde la información de fase. Así que, el problema de recuperación de fase (PR) implica aproximar el campo óptico utilizando los CDPs adquiridos. Este problema se ha abordado a través de dos aspectos principales: el diseño de máscaras de fase y el desarrollo de algoritmos de recuperación. Además de los sistemas DOI basados en luz coherente, esta tecnología se ha extendido a configuraciones ópticas basadas en luz incoherente, especialmente en sistemas de imágenes espectrales. La clasificación espectral es una tarea computacional clave en el análisis de imágenes espectrales. Esta disertación enfrenta tres desafíos relacionados con los sistemas DOI: el método de inicialización para el problema de PR en DOI coherente, el problema de PR basado en el diseño MPM a través de una red neuronal profunda (DNN) interpretable en DOI coherente, y la clasificación espectral basada en el diseño MPM utilizando una arquitectura óptica dual en DOI incoherente. Específicamente, esta tesis presenta tres metodologías de optimización de extremo a extremo (E2E) para abordar los desafíos mencionados anteriormente: un método de inicialización E2E utilizando una DNN de doble rama que entrena el paso de filtrado y aproxima el campo óptico; un enfoque de recuperación E2E a partir de una DNN interpretable basada en una formulación no convexa que aprende el MPM y recupera el campo óptico; y un enfoque de clasificación espectral E2E integrando una configuración óptica dual con una DNN que optimiza el MPM y etiqueta los materiales en imágenes espectrales. Los métodos E2E propuestos exhiben un mejor rendimiento que los enfoques convencionales. Finalmente, se implementan dos configuraciones ópticas para validar experimentalmente los métodos E2E propuestos.

---

[*] Tesis Doctoral

[**] Facultad de Ingenierías Fisicomecánicas. Escuela de Ingenierías Eléctrica, Electrónica y Telecomunicaciones. Director: Ph.D. Henry Arguello Fuentes.

ABSTRACT

TITLE: PHASE MASK DESIGN FOR DIFFRACTIVE OPTICAL IMAGING [*]

AUTHOR: ANDRÉS FELIPE JEREZ ARIZA [**]

KEYWORDS: DIFFRACTIVE IMAGING, MULTILEVEL PHASE MASKS, END-TO-END OPTIMIZATION, PHASE RETRIEVAL AND SPECTRAL CLASSIFICATION.

DESCRIPTION: Diffractive optical imaging (DOI) systems, which rely on multilevel phase masks (MPMs), encode the optical field wavefront before being recorded by the sensor. This results in coded diffraction patterns (CDPs) where the phase information is lost. Then, the phase retrieval (PR) problem involves approximating the optical field using the acquired CDPs. This problem has been tackled through two main aspects: the phase mask design and the recovery algorithm development. In addition to DOI setups based on coherent light, this technology has been extended to optical setups based on incoherent light, especially in spectral imaging systems. Spectral classification is a key computational task in spectral imaging analysis. This dissertation confronts three challenges related to DOI architectures: the initialization method for the PR problem in coherent DOI, the PR problem based on the MPM design through an interpretable deep neural network (DNN) in coherent DOI, and the spectral classification based on the MPM design using a dual optical architecture in incoherent DOI. Specifically, this thesis presents three end-to-end (E2E) optimization methodologies to tackle the challenges mentioned above: an E2E initialization method using a double-branch DNN that trains the filtering step and approximates the optical field; an E2E recovery approach from an interpretable DNN based on a non-convex formulation that learns the MPM and recovers the optical field; and an E2E spectral classification approach integrating a dual optical setup with a DNN that optimizes the MPM and labels materials into spectral images. The proposed E2E methods exhibit better performance than conventional approaches. Finally, two optical setups are implemented to experimentally validate the proposed E2E methods.

---

[*] Doctoral Thesis

[**] Facultad de Ingenierías Fisicomecánicas. Escuela de Ingenierías Eléctrica, Electrónica y Telecomunicaciones. Advisor: Ph.D. Henry Arguello Fuentes.

INTRODUCTION

Diffractive optical imaging (DOI) systems involve a coherent light source that illuminates an optical field [1]. This optical field is propagated to an image sensor, yielding diffraction patterns [2]. Diffraction patterns can be acquired along three different diffraction fields depending on the wave propagation distance, i.e., near-field, middle-field, and far-field, which can be modeled by using the angular spectrum method, Fresnel approximation, and Fraunhofer diffraction, respectively [3]. In particular, several applications have been developed based on each diffraction field's properties. For instance, some near-field applications are scanning near-field optical microscopy [4], near-field Raman imaging [5], and near-field optical spectroscopy [4]; some middle-field applications are Fresnel holography [6], and lensless imaging [7]; and some far-field

---

[1] Fucai Zhang et al. "Phase retrieval by coherent modulation imaging". In: *Nature communications* 7.1 (2016), p. 13367.

[2] Yoav Shechtman et al. "Phase retrieval with application to optical imaging: a contemporary overview". In: *IEEE signal processing magazine* 32.3 (2015), pp. 87–109.

[3] Joseph W Goodman. *Introduction to Fourier optics*. Roberts and Company Publishers, 2005.

[4] HF Hess et al. "Near-field spectroscopy of the quantum constituents of a luminescent system". In: *Science* 264.5166 (1994), pp. 1740–1745.

[5] CL Jahncke, MA Paesler, and HD Hallen. "Raman imaging with near-field scanning optical microscopy". In: *Applied physics letters* 67.17 (1995), pp. 2483–2485.

[6] Ting-Chung Poon and Jung-Ping Liu. *Introduction to modern digital holography: with MATLAB*. Cambridge University Press, 2014.

[7] Mayu Sao et al. "Lensless close-up imaging with Fresnel zone aperture". In: *Japanese Journal of Applied Physics* 57.9S1 (2018), 09SB05.

applications are X-ray crystallography [8], astronomy [9], and microscopy [10].

In practice, DOI setups cannot directly measure phase information, as these optical architectures can only capture the intensity of the incoming light from the underlying optical field. Recovering the phase information from the captured intensity involves solving an inverse problem known as phase retrieval (PR) [2,11]. Specifically, PR allows the optical field estimation from phaseless measurements [12,13]. Mathematically, for the ψ-th diffraction field, PR seeks to solve quadratic equations given by $(\mathbf{y}_{\psi})_k = |\mathbf{a}_{k,\psi}^{\mathcal{H}}\mathbf{x}|^2$, where $\mathbf{a}_{k,\psi} \in \mathbb{C}^n$ corresponds to the sampling vectors, $\mathbf{x} \in \mathbb{C}^n$ is the unknown optical field, $(\mathbf{y}_{\psi})_k$ represents the diffraction patterns collected by the sensor, $(\cdot)^{\mathcal{H}}$ denotes the conjugate transpose operator, $|\cdot|$ denotes the pointwise magnitude, and $\psi \in \{1, 2, 3\}$ indexes each diffraction field.

In the literature, the acquisition in DOI systems has been upgraded by assuring redundant information at each ℓ snapshot through an optical modulator [14] named mul-

---

[8] Samuel Pinilla, Juan Poveda, and Henry Arguello. “Coded diffraction system in X-ray crystallography using a boolean phase coded aperture approximation”. In: *Optics Communications* 410 (2018), pp. 707–716.

[9] Pierre Connes. “Astronomical fourier spectroscopy”. In: *Annual review of Astronomy and Astrophysics* 8 (1970), p. 209.

[10] Sheridan C Mayo et al. “X-ray phase-contrast microscopy and microtomography”. In: *Optics Express* 11.19 (2003), pp. 2289–2302.

[11] Jorge Bacca, Samuel Pinilla, and Henry Arguello. “Super-resolution phase retrieval from designed coded diffraction patterns”. In: *IEEE Transactions on Image Processing* 29 (2019), pp. 2598–2609.

[12] Andrés Guerrero, Samuel Pinilla, and Henry Arguello. “Phase Recovery Guarantees From Designed Coded Diffraction Patterns in Optical Imaging”. In: *IEEE Transactions on Image Processing* 29 (2020), pp. 5687–5697.

[13] Emmanuel J Candes, Xiaodong Li, and Mahdi Soltanolkotabi. “Phase retrieval from coded diffraction patterns”. In: *Applied and Computational Harmonic Analysis* 39.2 (2015), pp. 277–299.

[14] David Gross, Felix Krahmer, and Richard Kueng. “Improved recovery guarantees for phase retrieval from coded diffraction patterns”. In: *Applied and Computational Harmonic Analysis* 42.1

tilevel phase mask (MPM) $\mathbf{D} \in \mathbb{C}^{n\times n}$. This optical element modulates the wavefront, producing coded diffraction patterns (CDPs) $(\mathbf{y}_{\ell,\psi})_k = |\mathbf{a}_{k,\psi}^{\mathcal{H}}\mathbf{D}_\ell\mathbf{x}|^2$, for $\ell \in \{1,\cdots,L\}$ and $k \in \{0,\cdots,n-1\}$ across each diffraction field [13]. The PR problem based on CDP has been addressed from two main aspects: the sensing matrix design [12] and the recovery algorithm development [15].

The sensing matrix design in DOI setups based on CDP is directly related to the MPM structure, which is conventionally set from a random distribution of their coding elements, affecting the PR performance [12], and increasing the number of projections required for retrieving the optical field [13]. Some traditional phase mask design methodologies [12,11,16] provide recovery conditions in PR to establish that image reconstruction quality directly depends on the phase mask structure. Specifically, [12] proposes a phase mask design strategy using a greedy methodology based on the sensing matrix analysis over the three different diffraction fields, achieving phase masks that are physically implementable and improving the image reconstruction quality. Nevertheless, traditional phase mask designs in PR do not exploit the fact that optoelectronic devices, such as phase-only spatial light modulators [17,18], allow setting MPMs to aid the reconstruction quality, providing versatility, flexibility, and

---

(2017), pp. 37–64.

[15] Rohan Chandra, Tom Goldstein, and Christoph Studer. “Phasepack: A phase retrieval library”. In: *2019 13th International conference on Sampling Theory and Applications (SampTA)*. IEEE. 2019, pp. 1–5.

[16] Samuel Pinilla et al. “Coded aperture design for solving the phase retrieval problem in X-ray crystallography”. In: *Journal of Comput. and Applied Mathematics* 338 (2018), pp. 111–128.

[17] Uzi Efron. *Spatial light modulator technology: materials, devices, and applications*. Vol. 47. CRC press, 1994.

[18] Naim Konforti, Emanuel Marom, and S-T Wu. “Phase-only modulation with twisted nematic liquid-crystal spatial light modulators”. In: *Optics letters* 13.3 (1988), pp. 251–253.

diversity in the encoding process against binary modulation strategies [19].

Regarding the PR algorithms, one of the most popular approaches is based on the alternating projection method [20,21,22]. Nevertheless, this alternative heavily depends on the initial guess and lacks convergence guarantees. Thus, modern PR algorithms include optimization techniques, which can be categorized into convex and non-convex formulations. PR algorithms via convex formulations involving a convex relaxation in the PR problem, such as PhaseLift [23], PhaseMax [24]; and PR algorithms via non-convex formulations requiring a carefully statistical initialization, which is refined using a gradient-based method by the Wirtinger derivate, such as reweighted amplitude flow (RAF) [25], truncate amplitude flow (TAF) [26], and truncated Wirtinger

---

[19] David McGloin et al. "Applications of spatial light modulators in atom optics". In: *Optics Express* 11.2 (2003), pp. 158–166.

[20] Ralph W Gerchberg. "Phase determination from image and diffraction plane pictures in the electron microscope". In: *Optik* 34 (1971), pp. 275–284.

[21] Ralph W Gerchberg. "A practical algorithm for the determination of plane from image and diffraction pictures". In: *Optik* 35.2 (1972), pp. 237–246.

[22] James R Fienup. "Phase retrieval algorithms: a comparison". In: *Applied optics* 21.15 (1982), pp. 2758–2769.

[23] Emmanuel J Candes, Thomas Strohmer, and Vladislav Voroninski. "Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming". In: *Communications on Pure and Applied Mathematics* 66.8 (2013), pp. 1241–1274.

[24] Tom Goldstein and Christoph Studer. "Phasemax: Convex phase retrieval via basis pursuit". In: *IEEE Transactions on Information Theory* 64.4 (2018), pp. 2675–2689.

[25] Gang Wang et al. "Phase retrieval via reweighted amplitude flow". In: *IEEE Transactions on Signal Processing* 66.11 (2018), pp. 2818–2833.

[26] Gang Wang, Georgios B Giannakis, and Yonina C Eldar. "Solving systems of random quadratic equations via truncated amplitude flow". In: *IEEE Transactions on Information Theory* 64.2 (2018), pp. 773–794.

flow (TWF) [27].

Currently, deep neural networks (DNNs) have been implemented to solve the PR inverse problem, since these DNNs can be trained to learn an inverse mapping from the measurements to the reference optical field, exhibiting promising results [28,29,30]. DNN is commonly known as black box, which drawbacks the neural model interpretability and produces inadequate control, leading to longer training times [31]. Then, deep unrolling techniques have effectively facilitated the DNN architecture design based on well-established iterative approaches [32,33]. Deep unrolling methods establish iterative steps as layers in a deep model, resulting in interpretable DNN architectures that require less training data and achieve faster convergence rates [34]. Several

---

[27] Yuxin Chen and Emmanuel Candes. "Solving random quadratic systems of equations is nearly as easy as solving linear systems". In: *Advances in Neural Information Processing Systems*. 2015, pp. 739–747.

[28] Christopher Metzler, Phillip Schniter, Ashok Veeraraghavan, et al. "prDeep: robust phase retrieval with a flexible deep network". In: *International Conference on Machine Learning*. PMLR. 2018, pp. 3501–3510.

[29] Feilong Zhang et al. "Physics-based Iterative Projection Complex Neural Network for Phase Retrieval in Lensless Microscopy Imaging". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, pp. 10523–10531.

[30] Kyung-Su Kim and Sae-Young Chung. "Fourier phase retrieval with extended support estimation via deep neural network". In: *IEEE Signal Processing Letters* 26.10 (2019), pp. 1506–1510.

[31] Samuel Pinilla et al. "Unfolding-aided bootstrapped phase retrieval in optical imaging: Explainable AI reveals new imaging frontiers". In: *IEEE Signal Processing Magazine* 40.2 (2023), pp. 46–60.

[32] Brayan Monroy, Jorge Bacca, and Henry Arguello. "JR2net: a joint non-linear representation and recovery network for compressive spectral imaging". In: *Applied Optics* 61.26 (2022), pp. 7757–7766.

[33] Roman Jacome, Jorge Bacca, and Henry Arguello. "Deep-fusion: An end-to-end approach for compressive spectral image fusion". In: *2021 IEEE International Conference on Image Processing (ICIP)*. IEEE. 2021, pp. 2903–2907.

[34] Naveed Naimipour, Shahin Khobahi, and Mojtaba Soltanalian. "Upr: A model-driven architecture for deep phase retrieval". In: *2020 54th Asilomar Conference on Signals, Systems, and*

unrolling-based PR algorithms have been proposed based on alternating minimization techniques [34,29,35,36,37]. However, these PR methods are still sensitive to the initialization stage [34].

More recently, end-to-end (E2E) optimization approaches can jointly design the sensing matrix and the DNN parameters for addressing computational imaging tasks, such as recovery, detection, classification, segmentation, and image fusion, among others [38,33,39]. E2E has shown remarkable results by mixing the sensing design and reconstruction process in a single training stage [33,39]. For instance, the PR problem has been recently addressed by E2E schemes; specifically, [40] proposes an E2E approach in PR relying on the far-field model to jointly learn the parameters associated with the CDPs and the DNNs by using an unrolled gradient descent network method, which is based on alternating minimization [41]. To the best of our knowledge,

---

Computers. IEEE. 2020, pp. 205–209.

[35] Shi BaoShun and Lian QiuSheng. “DualPRNet: Deep shrinkage dual frame network for deep unrolled phase retrieval”. In: *IEEE Signal Processing Letters* 29 (2022), pp. 1177–1181.

[36] Chang-Jen Wang et al. “Phase retrieval with learning unfolded expectation consistent signal recovery algorithm”. In: *IEEE Signal Processing Letters* 27 (2020), pp. 780–784.

[37] Pierre-Hugo Vial et al. “Learning the proximity operator in unfolded admm for phase retrieval”. In: *IEEE Signal Processing Letters* 29 (2022), pp. 1619–1623.

[38] Henry Arguello et al. “Shift-variant color-coded diffractive spectral imaging system”. In: *Optica* 8.11 (2021), pp. 1424–1434.

[39] Jorge Bacca, Tatiana Gelvez-Barrera, and Henry Arguello. “Deep coded aperture design: An end-to-end approach for computational imaging tasks”. In: *IEEE Transactions on Computational Imaging* 7 (2021), pp. 1148–1160.

[40] Zikui Cai, Rakib Hyder, and M Salman Asif. “Data-driven illumination patterns for coded diffraction imaging”. In: *2021 IEEE International Conference on Image Processing (ICIP)*. IEEE. 2021, pp. 2818–2822.

[41] Praneeth Netrapalli, Prateek Jain, and Sujay Sanghavi. “Phase retrieval using alternating minimization”. In: *Advances in Neural Information Processing Systems* 26 (2013).

the PR literature does not inform E2E unrolling frameworks for the MPM design in DOI systems from a non-convex formulation, particularly employing the near-field propagation model.

Notice that traditional DOI setups based on MPMs have been devised for monochromatic (coherent) light using a single wavelength because of chromatic aberration. DOI methods have been extended for polychromatic (incoherent) illumination using multiple wavelengths [42]. For instance, spectral imaging systems collect multiple wavelengths from a target across the electromagnetic spectrum. These images have been used for numerous applications, such as environmental monitoring [43], precision agriculture [44], aerospace [45], defense [46], and biomedicine [47]. A notable DOI system based on incoherent light is the phase-coded spectral imaging system, which has been designed relying on a diffractive optical camera (DOC) [48]. This architecture incorporates a diffractive optical element (DOE), also known as MPM, which can be

---

[42] Daniel S Jeon et al. “Compact snapshot hyperspectral imaging with diffracted rotation”. In: (2019).

[43] Qihao Weng. “Thermal infrared remote sensing for urban climate and environmental studies: Methods, applications, and trends”. In: *ISPRS Journal of photogrammetry and remote sensing* 64.4 (2009), pp. 335–344.

[44] Muhammad Jaleed Khan et al. “Modern trends in hyperspectral image analysis: A review”. In: *Ieee Access* 6 (2018), pp. 14118–14129.

[45] Donatella Guzzi et al. “An atmospheric correction iterative method for high spectral resolution aerospace imaging spectrometers”. In: *2009 IEEE International Geoscience and Remote Sensing Symposium*. Vol. 2. IEEE. 2009, pp. II–73.

[46] Michal Shimoni, Rob Haelterman, and Christiaan Perneel. “Hypersectral imaging for military and security applications: Combining myriad processing and sensing techniques”. In: *IEEE Geoscience and Remote Sensing Magazine* 7.2 (2019), pp. 101–117.

[47] Minh H Tran and Baowei Fei. “Compact and ultracompact spectral imagers: technology and applications in biomedical imaging”. In: *Journal of biomedical optics* 28.4 (2023), pp. 040901–040901.

[48] Longqian Huang et al. “Spectral imaging with deep learning”. In: *Light: Science & Applications* 11.1 (2022), p. 61.

easily integrated into a camera to modulate the input light phase for each wavelength. Specifically, the phase-coded spectral imaging formulates the image formation as a convolution process between the wavelength-specified point spread function (PSF) and the monochrome object image at each wavelength. Here, the DOE manipulates the phase term of the PSF, distinguishing spectral signatures as light propagates. A suitable DOE design influences the PSF, thereby leading to more accurate spectral reconstruction results [48]. Compared with amplitude-coded spectral imaging, the phase-coded approach can greatly increase the light throughput.

The single-pixel camera (SPC) is widely used in amplitude-coded spectral imaging, providing reliable spectral images at lower-cost hardware compared to bidimensional spectral sensors [49]. The SPC integrates the spatial information into a single measurement, resulting in spectral images with high spectral resolution and low spatial resolution [50]. To address the spatial resolution limitation in SPC, the state-of-the-art has introduced dual optical architectures incorporating complementary monochromatic sensors. These architectures have shown significant improvements in spectral image recovery, offering both high spatial and spectral resolution [51,52]. In addition to the spectral reconstruction task, the spectral classification task represents one

---

[49] Ondřej Denk, Artem Musiienko, and Karel Žídek. “Differential single-pixel camera enabling low-cost microscopy in near-infrared spectral region”. In: *Optics express* 27.4 (2019), pp. 4562–4571.

[50] Hans Garcia, Claudia V Correa, and Henry Arguello. “Multi-resolution compressive spectral imaging reconstruction from single pixel measurements”. In: *IEEE Transactions on Image Processing* 27.12 (2018), pp. 6174–6184.

[51] Andrés Jerez, Hans Garcia, and Henry Arguello. “Single pixel spectral image fusion with side information from a grayscale sensor”. In: *2018 IEEE 1st Colombian Conference on Applications in Computational Intelligence (ColCACI)*. IEEE. 2018, pp. 1–6.

[52] Laura Galvis et al. “Coded aperture design in compressive spectral imaging based on side information”. In: *Applied optics* 56.22 (2017), pp. 6332–6340.

of the main applications in spectral imaging systems [53,54]. This computational task involves categorizing objects or materials based on their spectral signatures, which depict their interactions with electromagnetic radiation across different wavelengths [55]. In particular, the spectral signatures are acquired using spectral imaging techniques, enabling the differentiation and identification of materials based on their distinct spectral properties [55]. In traditional spectral imaging techniques, dual optical architectures integrating both DOCs and SPCs for spectral classification tasks have not been reported; combining phase encoding from DOC with other encoding architectures as SPC can boost classification performance.

This dissertation aims to develop and implement MPM design methodologies for reconstruction and classification tasks in DOI architectures via DNNs. More precisely, three main challenges in DOI architectures are addressed: the initialization method involved in the PR problem for coherent DOI, the PR problem based on the MPM design through an interpretable DNN for coherent DOI, and the spectral classification based on the MPM design from a dual optical architecture for incoherent DOI. Therefore, this thesis presents three different E2E optimization approaches that face the abovementioned challenges. The first E2E method for coherent DOI can approximate the optical field by learning the initialization process and the recovery DNN parameters at each diffraction field. The proposed learnable initialization approach requires fewer iterations compared to traditional methods. The reconstruction stage

---

[53] Maryam Imani and Hassan Ghassemian. “An overview on spectral and spatial information fusion for hyperspectral image classification: Current trends and challenges”. In: *Information fusion* 59 (2020), pp. 59–83.

[54] Jorge Bacca, Emmanuel Martinez, and Henry Arguello. “Computational spectral imaging: a contemporary overview”. In: *JOSA A* 40.4 (2023), pp. C115–C125.

[55] Yushi Chen et al. “Deep feature extraction and classification of hyperspectral images based on convolutional neural networks”. In: *IEEE transactions on geoscience and remote sensing* 54.10 (2016), pp. 6232–6251.

is performed by a double-branch DNN using only a single snapshot. The second E2E method for coherent DOI can recover the optical field by learning the MPM and the recovery network parameters at the near-field. The proposed recovery approach introduces an interpretable DNN by unrolling the spectral initialization and iterative reconstruction algorithms for the near-field propagation model based on a non-convex formulation. The unrolled-based recovery method achieves better recovery performance than the benchmark PR algorithms. The third E2E method for incoherent DOI can classify the materials within the land cover spectral images by learning the MPM and the classification network parameters at the middle-field. The proposed spectral classification approach incorporates a dual optical architecture that combines an SPC to record the spectral information and a DOC to acquire the spatial information. The proposed spectral classification based on the dual optical setup improves the classification performance compared to conventional optical architectures. Finally, two optical setups are assembled, aligned, and calibrated in the High Dimensional Signal Processing (HDSP) optical laboratory for experimentally validating the proposed E2E recovery and classification methods. Integrating DNNs into MPM design significantly enhances phase reconstruction and spectral classification performance in both simulation and experimentation results.

The document is organized as follows: Chapter 1 provides the fundamental theoretical background and literature within the diffractive imaging framework. Chapter 2 introduces the proposed E2E learnable PR initialization approach based on double-branch DNN from a single snapshot. Chapter 3 presents the proposed E2E deep unrolled phase recovery approach based on a non-convex formulation from a single snapshot. Chapter 4 describes the proposed E2E spectral classification approach based on a dual optical architecture involving the SPC and DOC. Chapter 5 summarizes the conclusions, discussion, and future work.

GENERAL OBJECTIVE

To design and implement multilevel phase masks for reconstruction and classification tasks in diffractive optical imaging.

SPECIFIC OBJECTIVES

1. To develop a mathematical sensing model, a reconstruction method, and a classification approach for a diffractive optical imaging system employing multilevel phase masks.
2. To implement computational algorithms that simulate a diffractive imaging system's acquisition, reconstruction, and classification process based on multilevel phase masks.
3. To design the sensing matrix from multilevel phase masks for improving the reconstruction and classification tasks in diffractive optical systems.
4. To implement an optical testbed that validates the reconstruction and classification methods using designed multilevel phase masks in optoelectronic devices.
5. To analyze and evaluate the proposed multilevel phase mask design methodology for reconstruction and classification tasks over experimental data acquired from an optical testbed implementation using optoelectronic phase modulators.

IMPACT OF THE RESEARCH AND CONTRIBUTIONS

The contributions of this dissertation are mainly related to the multilevel phase mask design using deep neural network approaches for reconstruction and classification tasks in diffractive optical systems.

Specifically, the main contributions are summarized as follows.

- **Chapter 2:** This chapter introduces a deep spectral initialization approach using a single snapshot based on an E2E neural network scheme designed to learn the kernel filter. The proposed E2E neural network contains three main stages: an optical stage that characterizes the CDPs at each diffraction field; an initialization stage that computes the optical field approximation; and a double-branch DNN recovery stage that improves the estimated optical field. The proposed initialization approach achieves better estimation performance from a single snapshot and requires fewer iterations than traditional methods.
- **Chapter 3:** This chapter presents a deep unrolled PR approach from a non-convex formulation using a single snapshot based on an E2E neural network scheme that learns the MPM. The proposed E2E includes three main stages: an optical stage that models the CDPs through the angular spectrum method; an unrolling-based initialization stage that estimates the optical field; and an unrolling-based recovery stage that further refines the optical field using gradient descent steps and convolutional layers. The proposed recovery method outperforms traditional benchmark PR algorithms using a single snapshot and a learned MPM.
- **Chapter 4:** This chapter describes a dual optical system for spectral classification based on an E2E neural network scheme designed to learn the MPM. The proposed deep E2E approach comprises three main stages: a propagation stage involving the SPC to acquire spectral information and the DOC to collect spatial information, which is modeled by the Fresnel approximation; a spectral image fusion stage that computes the spectral image; and a classification stage that labels the materials defined within the spectral images. The proposed classification method improves classification performance using the dual spectral architecture compared to conventional optical setups.

LIST OF PUBLICATIONS

Most of the material presented in this thesis appears in the following publications by the author.

Journal papers:

1. **Jerez, A.**, Estupiñán, J., Bacca, J., & Arguello, H. (2024). Deep unrolled single snapshot phase retrieval via non-convex formulation and phase mask design. IEEE Journal of Selected Topics in Signal Processing.
   DOI: 10.1109/JSTSP.2024.3395979
2. **Jerez, A.**, Blanco, G., Urrea, S., García, H., & Arguello, H. (2024). Spectral classification using a dual optical setup and deep neural networks. Revista UIS Ingenierías.
3. Morales, D., Márquez, M., **Jerez, A.**, Arguello, H., Vera, E., & Meza, P. (2024). Compressive spectral video by dynamic spatial-spectral-temporal windowed codification. Optics Express. (Ongoing).
4. Morales, D., **Jerez, A.**, & Arguello, H. (2022). Learning spectral initialization for phase retrieval via deep neural networks. Applied Optics, 61(9), F25-F33.
   DOI: 10.1364/AO.445085
5. **Jerez, A.**, Márquez, M., & Arguello, H. (2021). Adaptive coded aperture design for compressive computed tomography. Journal of Computational and Applied Mathematics, 384, 113174. DOI: 10.1016/j.cam.2020.113174
6. **Jerez, A.**, Pinilla, S., & Arguello, H. (2020). Fast target detection via template matching in compressive phase retrieval. IEEE Transactions on Computational Imaging, 6, 934-944. DOI: 10.1109/TCI.2020.2995999

Conference papers:

1. **Jerez, A.**, Blanco, G., Urrea, S., García, H., Castillo, S., & Arguello, H. (2023, November). Dual optical design for VIS-NIR classification based on single pixel and diffractive Cameras. In Colombian Caribbean Conference (C3) (pp. 1-5). IEEE. DOI: 10.1109/C358072.2023.10436309
2. Herrera, M., Morales, D., García, H., Castillo, S., **Jerez, A.**, & Arguello, H. (2023, August). Multi-domain discrete spectral image fusion via implicit neural representation. In Computational Optical Sensing and Imaging (COSI) (pp. 1-2). Optica Publishing Group. DOI: 10.1364/COSI.2023.CTh2A.3
3. Morales, D., Márquez, M., **Jerez, A.**, Arguello, H., Meza, P. (2023, June). Compressive spectral video via wavelength-dynamic coded aperture. In Photonics North (PN) (pp. 1-2). IEEE. DOI: 10.1109/PN58661.2023.10221940
4. Estupiñán, J., **Jerez, A.**, Bacca, J., & Arguello, H. (2021, September). Deep unrolled phase retrieval approach from coded diffraction patterns. In XXIII Symposium on Image, Signal Processing and Artificial Vision (STSIVA) (pp. 1-4). IEEE. DOI: 10.1109/STSIVA53688.2021.9591671
5. Morales, D., **Jerez, A.**, & Arguello, H. (2021, September). Object classification using deep neural networks from coded diffraction patterns. In XXIII Symposium on Image, Signal Processing and Artificial Vision (STSIVA) (pp. 1-5). IEEE. DOI: 10.1109/STSIVA53688.2021.9591996
6. Morales, D., **Jerez, A.**, & Arguello, H. (2021, July). Deep phase retrieval by a learnable filtered spectral initialization. In Computational Optical Sensing and Imaging (COSI) (pp. 1-2). Optica Publishing Group.
   DOI: 10.1364/COSI.2021.CTh4A.6

INTERNATIONAL INTERNSHIPS

The author participated in two international internships as a visiting Ph.D. student in the below institutions.

1. **Institut National de la Recherche Scientifique (INRS):** This internship was developed at the Laboratory of Applied Computational Imaging in Montreal, Canada, from January 2023 to July 2023 in a face-to-face format under the supervision of Professor Jinyang Liang. During this internship, the research project focused on designing and simulating new encoding masks for an all-optics object classification system through diffractive neural networks. This internship was sponsored by the Emerging Leaders in the Americas Program (ELAP).

2. **Universidad de la Frontera (UFRO):** This internship was developed at the Information Processing Laboratory in Temuco, Chile, from January 2022 to May 2022 in a hybrid format under the supervision of Professor Pablo Meza. During this internship, the research project focused on designing, assembling, and calibrating a coded optical system to acquire and recover spectral-depth information from a single projection. This internship was sponsored by the National Fund for Scientific and Technological Development (FONDECYT).

UNDERGRADUATE ADVISING

The author guided the following undergraduate students.

Student 1

Title: Deep learning-based spectral image classification algorithm by using a diffractive optical architecture
Advised student: Geison Alfredo Blanco Rodríguez
Advisor: Andrés Felipe Jerez Ariza
Co-advisor: Henry Arguello Fuentes
Status: Approved
Program: Systems Engineering
University: Universidad Industrial de Santander
Year: 2024

Student 2

Title: Object classification algorithm in diffractive imaging based on coded quadratic measurements using a deep learning approach
Advised student: David Santiago Morales Norato
Advisor: Andrés Felipe Jerez Ariza
Co-advisor: Henry Arguello Fuentes
Status: Approved
Award: Outstanding undergraduate thesis
Program: Systems Engineering
University: Universidad Industrial de Santander
Year: 2022

1. BACKGROUND, THEORETICAL FRAMEWORK AND CURRENT STATE

In this chapter, some theoretical concepts and state-of-the-art in diffractive optical imaging are described, including the diffraction theory, the phase retrieval problem, the coded diffraction acquisition systems, the traditional coding mask designs, the optical modulators, the phase retrieval algorithms, the deep unrolled networks, the end-to-end optimization, the diffractive spectral imaging, and the diffractive imaging current state.

1.1. DIFFRACTION THEORY

According to the Maxwell equation, the wave formula in free space for the electric field $\mathbf{E}$ is given by

$$\nabla^2\mathbf{E} - \frac{1}{\nu^2}\frac{\partial^2\mathbf{E}}{\partial t^2} = 0, \tag{1}$$

where $\nu = 1/\sqrt{\mu\varepsilon}$ is the wave speed in the medium, with μ and ε the permittivity and permeability of the medium, respectively [56]. The electric field is defined by unit vectors $\mathbf{a}_x$, $\mathbf{a}_y$, $\mathbf{a}_z$ in x, y, z directions, respectively, as

$$\mathbf{E} = E_x\mathbf{a}_x + E_y\mathbf{a}_y + E_z\mathbf{a}_z. \tag{2}$$

Considering that Laplacian operator is given by $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$, (1) can be rewritten based on (2) as

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)(E_x\mathbf{a}_x + E_y\mathbf{a}_y + E_z\mathbf{a}_z) = \frac{1}{\nu^2}\frac{\partial^2}{\partial t^2}(E_x\mathbf{a}_x + E_y\mathbf{a}_y + E_z\mathbf{a}_z). \tag{3}$$

---

[56] Julius Adams Stratton. *Electromagnetic theory*. Vol. 33. John Wiley & Sons, 2007.

Comparing each component from both sides in (3), we obtain the following compact version for their three components

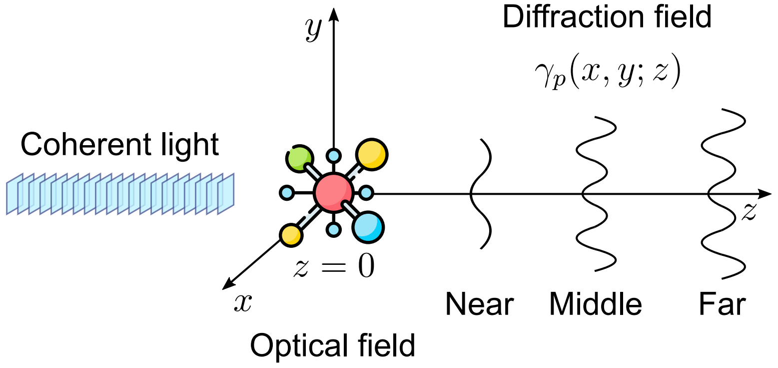
$$\frac{\partial^2\gamma}{\partial x^2} + \frac{\partial^2\gamma}{\partial y^2} + \frac{\partial^2\gamma}{\partial z^2} = \frac{1}{\nu^2}\frac{\partial^2\gamma}{\partial t^2}, \tag{4}$$

where γ represents a component E_x, E_y or E_z from the electric field $\mathbf{E}$. Here, (4) is known as a three-dimensional scalar wave equation [57]. Solving (4) in the z-direction, the field distribution for a monochromatic coherent wave can be estimated as

$$\gamma(x, y, z, t) = \gamma_p(x, y; z)e^{j\omega_0 t}, \tag{5}$$

where $j = \sqrt{-1}$ is the imaginary unit, ω_0 is the angular frequency, and $\gamma_p(x, y; z)$ is the unknown diffraction field along the z-axis from an object located in $z = 0$, i.e., $\gamma_p(x, y; z = 0) = \gamma_{p0}(x, y)$, as shown in Fig. 1.

Figure 1. Diffraction of an incident plane wave at each diffraction field. After being illuminated by a coherent light source, the optical field is propagated through three different diffraction fields.



---

[57] Julius Adams Stratton and LJ Chu. “Diffraction theory of electromagnetic waves”. In: *Physical Review* 56.1 (1939), p. 99.

Therefore, the diffracted field $\gamma_p(x, y; z)$ along the z-direction [6] can be modeled as

$$\gamma_p(x,y;z) \propto \begin{cases} \mathcal{F}^{-1}\left\{ \mathcal{F}\{\gamma_{p_0}(x,y)\} e^{-jk_0 z\sqrt{1-\frac{k_x^2}{k_0^2}-\frac{k_y^2}{k_0^2}}} \right\} & \text{if } F \gg 1 \rightarrow \text{Near-field} \\ \mathcal{F}\left\{ \gamma_{p_0}(x,y) e^{\frac{-jk_0}{2z}(x^2+y^2)} \right\} & \text{if } F \sim 1 \rightarrow \text{Middle-field} \\ \mathcal{F}\left\{ \gamma_{p_0}(x,y) \right\} & \text{if } F \ll 1 \rightarrow \text{Far-field} \end{cases}, \tag{6}$$

where $\mathcal{F}\{\cdot\}$ and $\mathcal{F}^{-1}\{\cdot\}$ correspond to the Fourier and inverse Fourier transform, respectively, F represents the Fresnel number [58], $k_0 = 2\pi/\lambda$ is the wavenumber with λ as the wavelength, k_x and k_y are the spatial frequencies.

Considering $\mathbf{x} \in \mathbb{C}^n$ as an unknown optical field, the discrete diffracted field intensity at the sensor plane [6] based on (6) can be expressed as

$$\mathbf{y}_\psi \propto \begin{cases} |\mathbf{F}\mathbf{T}\mathbf{F}^{\mathcal{H}}\mathbf{x}|^2 & \text{if } \psi = 1 \rightarrow \text{Near-field} \\ |\mathbf{F}^{\mathcal{H}}\mathbf{Q}\mathbf{x}|^2 & \text{if } \psi = 2 \rightarrow \text{Middle-field} \\ |\mathbf{F}\mathbf{x}|^2 & \text{if } \psi = 3 \rightarrow \text{Far-field} \end{cases}, \tag{7}$$

where $(\cdot)^{\mathcal{H}}$ is the conjugate transpose operation, $|\cdot|$ denotes the pointwise magnitude, $\mathbf{F} \in \mathbb{C}^{n\times n}$ is the discrete Fourier transform, $\mathbf{y}_\psi \in \mathbb{R}^n$ represents the measurements acquired by the sensor at the ψ-th diffraction field, and $\mathbf{T} \in \mathbb{C}^{n\times n}$ is the transfer function in the spatial frequency [6] given by

$$(\mathbf{T})_{r,s} = e^{-jk_0 z\sqrt{1-\frac{(r\delta_{kx})^2}{k_0^2}-\frac{(s\delta_{ky})^2}{k_0^2}}}, \tag{8}$$

where $[r, s]$ are the discrete indices of the samples in the Fourier domain. The terms δ_{kx} and δ_{ky} are the sampling periods in the frequency coordinates. Further, $\mathbf{Q} \in \mathbb{C}^{n\times n}$

---

[58] Yajun Li and Emil Wolf. "Three-dimensional intensity distribution near the focus in systems of different Fresnel numbers". In: *JOSA A* 1.8 (1984), pp. 801–808.

is the exponential function of the quadratic phase [6] given by

$$(\mathbf{Q})_{p,q} = e^{\frac{-jk_0}{2z}(p^2\delta_x^2 + q^2\delta_y^2)}, \tag{9}$$

where $[p, q]$ are the discrete indices of the samples in the spatial domain. The terms δ_x and δ_y are the sampling period in the spatial coordinates.

It is important to highlight that throughout this thesis, the variables x, y, and z are the Cartesian coordinates, while the bold variables $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ refer to the unknown object, the diffraction patterns, and the object approximation, respectively.

1.2. PHASE RETRIEVAL

Phase retrieval (PR) is an inverse problem that allows the optical field recovery using diffraction patterns along different diffraction fields [2]. The PR problem arises in electromagnetic theory applications in which phase information is lost or is impractical to measure from optical setups, being only able to capture the intensity of the incoming coherent light from the underlying optical field [2]. Mathematically, the PR problem is formulated as recovering a signal $\mathbf{x} \in \mathbb{C}^n$ from a set of m quadratic equations as follows

$$(\mathbf{y})_i = |\langle \mathbf{a}_i, \mathbf{x} \rangle|^2, \text{ for } i \in \{1, \cdots, m\}, \tag{10}$$

where $\langle \cdot \rangle$ denotes the inner product, $\mathbf{a}_i$ corresponds to the known sampling vectors, and $(\mathbf{y})_i$ represents the diffraction patterns. Specifically, PR is inherently an ill-posed problem; many signals may share the same magnitude measurements in their quadratic form. For instance, $\mathbf{a}_i$ is often assumed as random vectors to provide uniqueness guarantees (up to a global constant phase) [14], which are otherwise difficult to obtain. Specifically, any accessible solutions are given by $\hat{\mathbf{x}} = e^{j\beta}\mathbf{x}$, for some $\beta \in [0, 2\pi)$. In consequence, the Euclidean distance $\text{dist}(\cdot)$ between two complex vectors $\mathbf{w}_1, \mathbf{w}_2 \in \mathbb{C}^n$, on this PR problem is indifferent to a global constant expressed as

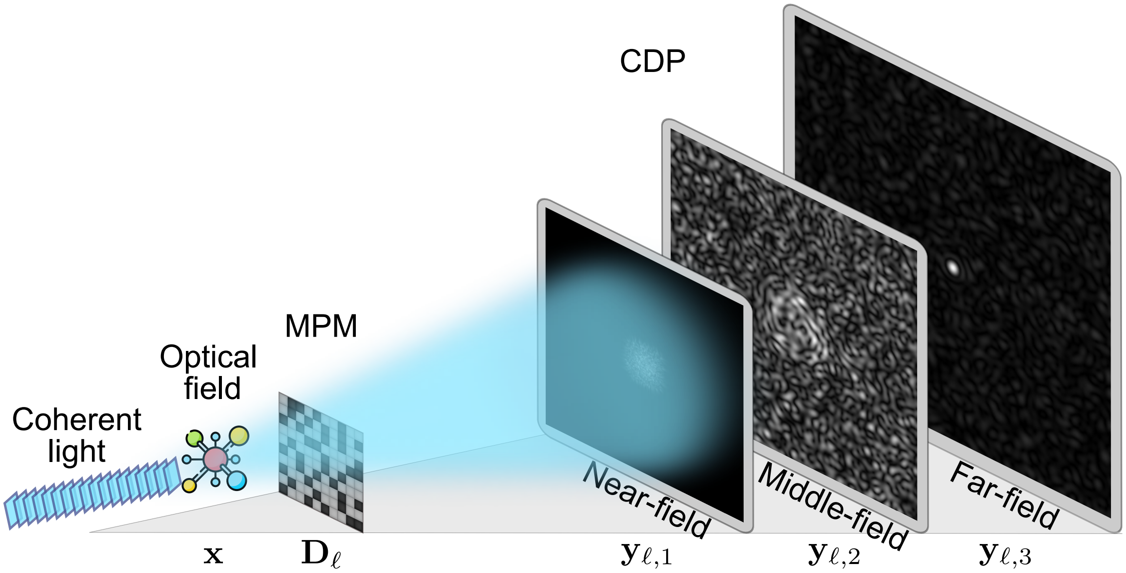
$$\operatorname{dist}(\mathbf{w}_1, \mathbf{w}_2) = \min_{\beta \in [0, 2\pi)} \left\| \mathbf{w}_1 e^{-j\beta} - \mathbf{w}_2 \right\|_2 . \tag{11}$$

In particular, this inverse problem has been solved by inducing redundancy in the measurement process from a multilevel phase mask (MPM), which allows optical field modulation, producing coded diffraction patterns (CDPs) at each diffraction field.

1.3. CODED DIFFRACTION PATTERNS

Here, the acquisition system that collects the CDPs of an optical field is described, specifically, at three diffraction fields, as illustrated in Fig. 2. The MPM is introduced in the object plane to module the optical field $\mathbf{x} \in \mathbb{C}^n$, and the CDPs are acquired by a sensor at each diffraction field. In fact, changing the spatial setup of the MPM allows the system to capture multiple projections of the scene.

Figure 2. Diffractive optical system for CDPs acquisition at each diffraction field. A coherent light source illuminates an optical field, then, an MPM modulates the incoming optical field, producing CDPs. These CDPs are collected by a sensor, which can be located at three different diffraction fields.



Mathematically, $\mathbf{D}_\ell \in \mathbb{C}^{n\times n}$ is a diagonal matrix that models the MPM at the ℓ-th projection where $\ell \in \{1, \cdots, L\}$ with L as the number of projections. Then, the acquired CDPs at three diffraction zones for the ℓ-th projection [12] are given by

$$\begin{aligned} \mathbf{y}_{\ell,1} &= |\mathbf{F}\mathbf{T}\mathbf{F}^{\mathcal{H}}\mathbf{D}_\ell\mathbf{x}|^2 + \boldsymbol{\omega}_{\ell,1} \ \rightarrow \text{Near-field}, \\ \mathbf{y}_{\ell,2} &= |\mathbf{F}^{\mathcal{H}}\mathbf{Q}\mathbf{D}_\ell\mathbf{x}|^2 + \boldsymbol{\omega}_{\ell,2} \ \rightarrow \text{Middle-field}, \\ \mathbf{y}_{\ell,3} &= |\mathbf{F}\mathbf{D}_\ell\mathbf{x}|^2 + \boldsymbol{\omega}_{\ell,3} \ \rightarrow \text{Far-field}, \end{aligned} \tag{12}$$

where $\boldsymbol{\omega}_{\ell,\psi} \in \mathbb{R}^n$ is the observed additive noise at each ψ-th diffraction field. Notice that the near and middle fields are described by using two auxiliary orthogonal diagonal matrices $\mathbf{T} \in \mathbb{C}^{n\times n}$, and $\mathbf{Q} \in \mathbb{C}^{n\times n}$ that model the discrete version of the transfer function in the discrete spatial frequency for these diffraction fields [3, 6], which depend on the propagation distance z and the wavelength of the coherent beam as shown in (9) and (8), respectively. Based on (12), the PR problem for the ψ-th diffraction field consists in estimating the optical field $\mathbf{x}$ from CDPs $\{\mathbf{y}_{\ell,\psi}\}_{\ell=1}^{L}$. Considering the CDPs in (12), the following section introduces the coding mask concept and several coding structures used to solve the PR problem.

1.4. CODING MASK DESIGN

This dissertation considers that MPMs $\mathbf{D}_\ell$ are random encoders in which the entries at each diagonal matrix $\mathbf{D}_\ell = \text{diag}(\mathbf{d})$ are *i.i.d.* copies of a random variable d with $\mathbf{d} \in \{d\}^n$, where $\text{diag}(\cdot)$ returns a square diagonal matrix from a given vector. Then, we assume that d satisfies the admissible condition in Definition 1.

Definition 1 *(Admissible random variable)* [12]*. A discrete random variable obeying* $|d| \leq 1$*, it is said to be admissible.*

Notice that the inequality in Definition 1 imposes that d cannot increase the light beam's energy during the modulation process. Thus, some examples of admissible random variables are shown in Table 1. These random variables have been recently used in [8,13,16,59,60] to solve the PR problem in computational imaging applications. Now, we will refer to each possible value of d as a coding element in the MPM. Further, the pattern type determines how the entries are distributed into the mask.

Table 1. Some admissible random variables.

Coding element	Coding probability
$d \in \{0, 1\}$	$\{\frac{1}{2}, \frac{1}{2}\}$
$d \in \{-1, 1\}$	$\{\frac{1}{2}, \frac{1}{2}\}$
$d \in \{-1, 1, -j, j\}$	$\{\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\}$

1.4.1. Boolean masks The boolean masks, also known as block-unblock masks, can be easily implemented in X-ray diffraction applications [61,62]. The entries of this type of mask can be modeled as a Bernoulli random variable with parameter t_r as the transmittance, i.e., $(\mathbf{D})_{k,i} \sim \mathrm{Bernoulli(t_r)}$, where $\mathbf{D} \in \{0, 1\}^{n \times n}$. Figure 3 shows an example of two random boolean masks with different transmittances.

---

[59] Emmanuel J Candes, Xiaodong Li, and Mahdi Soltanolkotabi. “Phase retrieval via Wirtinger flow: Theory and algorithms”. In: *IEEE Transactions on Information Theory* 61.4 (2015), pp. 1985–2007.

[60] Huishuai Zhang and Yingbin Liang. “Reshaped wirtinger flow for solving quadratic system of equations”. In: *Advances in Neural Information Processing Systems*. 2016, pp. 2622–2630.

[61] Kenneth P MacCabe et al. “Snapshot 2D tomography via coded aperture x-ray scatter imaging”. In: *Applied optics* 52.19 (2013), pp. 4582–4589.

[62] Joel A Greenberg et al. “Coding and sampling for compressive x-ray diffraction tomography”. In: *Wavelets and Sparsity XV*. vol. 8858. International Society for Optics and Photonics. 2013, p. 885813.

Figure 3. Sketch of boolean masks with transmittance $t_r \in \{0.25, 0.50\}$. The black and white pixels correspond to $d \in \{0,1\}$ coding elements, respectively.

1.4.2. Fourier masks Let $\mathbf{F} \in \mathbb{C}^{n\times n}$ be the discrete Fourier transform (DFT) matrix. Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ and $\mathbf{B} \neq \mathbf{0} \in \mathbb{R}^{n\times n}$ be the real and complex components of $\mathbf{F}$, respectively, i.e., $\mathbf{F} = \mathbf{A} + j\mathbf{B}$. Further, the phase $\boldsymbol{\theta} \in [0,1]^{n\times n}$ of the matrix $\mathbf{F}$ is given by

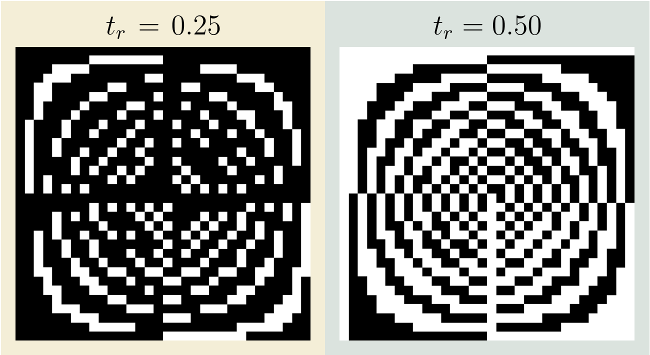
$$(\boldsymbol{\theta})_{u,v} = \frac{\left(\arctan\left(\frac{(\mathbf{A})_{u,v}}{(\mathbf{B})_{u,v}}\right) + \pi\right)}{2\pi}. \tag{13}$$

A mask designed from DFT is a matrix $\mathbf{D} \in \{0,1\}^{n\times n}$ whose entries are distributed with a Bernoulli random variable based on the phase $\boldsymbol{\theta}$ of the matrix $\mathbf{F}$ given a desired transmittance t_r [8,16], i.e.,

$$(\mathbf{D})_{k,i} = \begin{cases} 1, & \text{if } (\boldsymbol{\theta})_{k,i} \leq t_r \\ 0, & \text{otherwise} \end{cases}. \tag{14}$$

Figure 4 illustrates an example of two DFT masks with different transmittance.

Figure 4. Sketch of DFT masks with transmittance $t_r \in \{0.25, 0.50\}$. The black and white pixels correspond to $d \in \{0, 1\}$ coding elements, respectively.



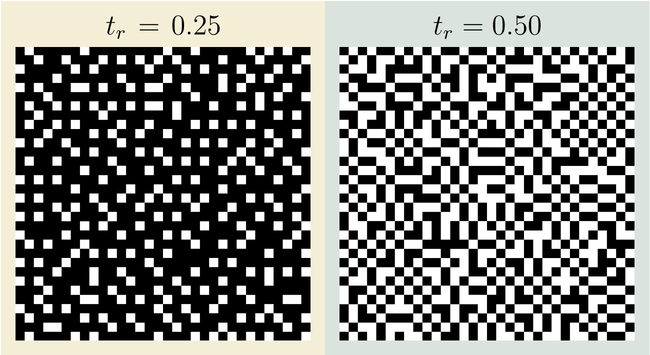
1.4.3. Uniform masks The uniform masks have been used in spectral imaging to reduce clusters and achieve uniform sensing [63]. These masks are designed by optimizing the spatial distribution, i.e., a distribution with the majority of the energy in the high spatial frequencies is used, this spectrum distribution is known as blue noise patterns [64]. Figure 5 shows an example of two uniform masks via blue noise patterns with different transmittance.

---

[63] Chen Fu et al. “Compressive spectral polarization imaging by a pixelized polarizer and colored patterned detector”. In: *JOSA A* 32.11 (2015), pp. 2178–2188.

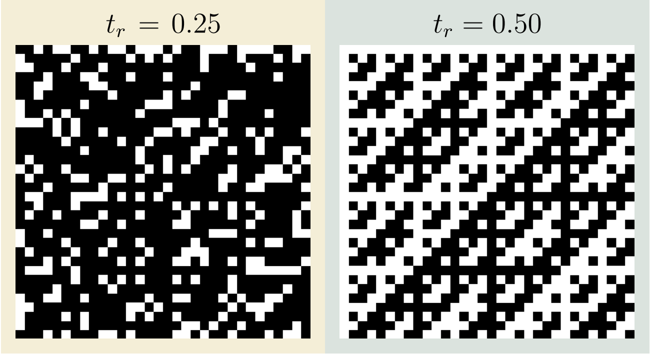
[64] J Bacca Rodriguez, Gonzalo R Arce, and Daniel Leo Lau. “Blue-noise multitone dithering”. In: *IEEE Transactions on Image Processing* 17.8 (2008), pp. 1368–1382.

Figure 5. Sketch of uniform masks with transmittance $t_r \in \{0.25, 0.50\}$. The black and white pixels correspond to $d \in \{0, 1\}$ coding elements, respectively.



1.4.4. Hadamard masks A Hadamard matrix exhibits two main characteristics: its entries are either 1 or -1, and its columns are orthogonal [65]. Figure 6 shows an example of two Hadamard masks with different transmittance.

Figure 6. Sketch of Hadamard masks with with transmittance $t_r \in \{0.25, 0.50\}$. The black and white pixels correspond to $d \in \{-1, 1\}$ coding elements, respectively.

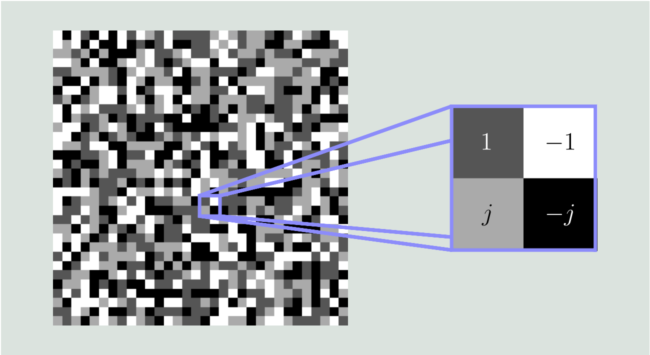


1.4.5. Octanary pattern The octanary pattern was introduced in [59]. The entries of this type of pattern are selected uniformly at random from $d \in \{-1, 1, -j, j\}$, where the elements $d \in \{-1, -j, j\}$ modify the phase of the diffracted beams. Figure 7 illustrates an example of an octanary pattern.

---

[65] David J Brady. *Optical imaging and spectroscopy*. John Wiley & Sons, 2009.

Figure 7. Sketch of octanary pattern. The grayscale pixels correspond to $d \in \{-1, 1, -j, j\}$ coding elements, respectively.



1.5. SPATIAL LIGHT MODULATORS

The coding process can be integrated into the optical systems through spatial light modulators (SLMs), which can be classified into two groups depending on their operating principle [66]: phase modulators, such as deformable mirror (DM) [67], liquid crystal on silicon device (LCOS) [68], diffractive optical element (DOE) [69]; and amplitude modulators, such as digital micromirror device (DMD) [70]. Figure 8 illustrates different spatial light modulators.

---

[66] Miguel Marquez et al. “Snapshot compressive spectral depth imaging from coded aberrations”. In: *Optics Express* 29.6 (2021), pp. 8142–8159.

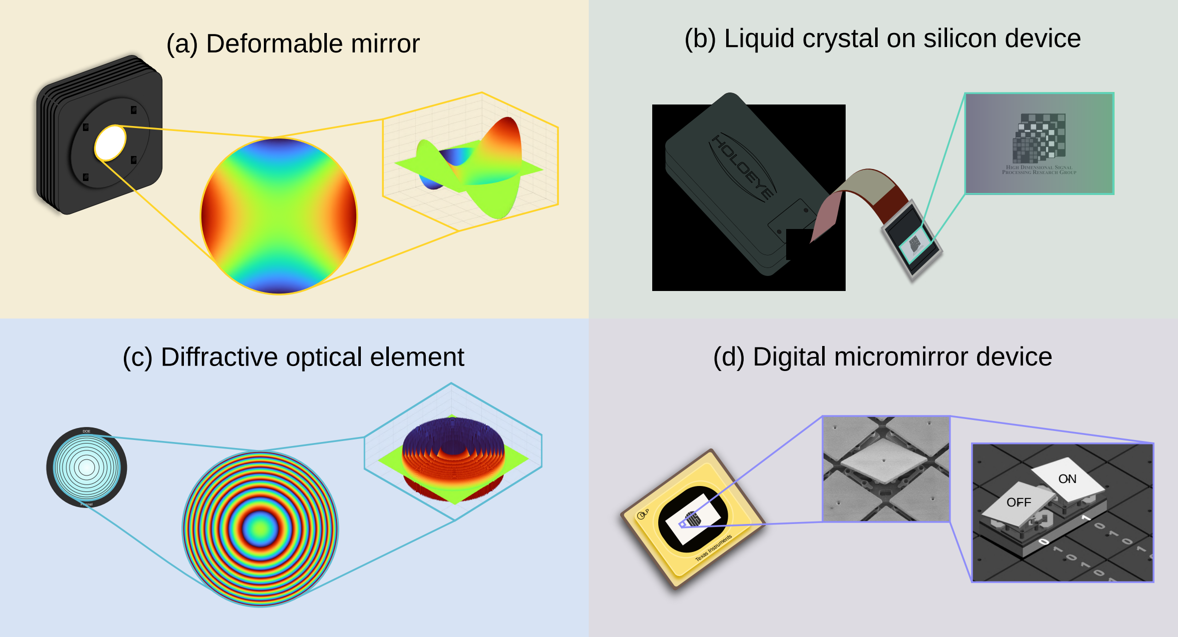
[67] Larry J Hornbeck. “Deformable-mirror spatial light modulators”. In: *Spatial Light Modulators and Applications III*. vol. 1150. International Society for Optics and Photonics. 1990, pp. 86–103.

[68] Zichen Zhang, Zheng You, and Daping Chu. “Fundamentals of phase-only liquid crystal on silicon (LCOS) devices”. In: *Light: Science & Applications* 3.10 (2014), e213–e213.

[69] Olivier Ripoll, Ville Kettunen, and Hans Peter Herzig. “Review of iterative Fourier-transform algorithms for beam shaping applications”. In: *Optical Engineering* 43.11 (2004), pp. 2549–2548.

[70] Yu-Xuan Ren, Rong-De Lu, and Lei Gong. “Tailoring light with a digital micromirror device”. In: *Annalen der physik* 527.7-8 (2015), pp. 447–470.

Figure 8. Some common SLMs implemented into optical setups for encoding the incoming light. Phase modulator: (a) DM, (b) LCOS, and (c) DOE. Amplitude modulator: (d) DMD.



(a) Deformable mirror

(b) Liquid crystal on silicon device

(c) Diffractive optical element

(d) Digital micromirror device

1.5.1. Deformable mirror This technology [67] consists of a metalized polymer material stretched over a spacer grid, i.e., an air gap formed between the membrane and the underlying address circuit. The membrane shape is deformed into the spacer grid to control the wavefront and correct the optical aberrations. Applications for the DM include optical correlation, spectrum analysis, crossbar switching, frequency excision, and projection displays, among others.

1.5.2. Liquid crystal on silicon device This technology [68] compounds the unique light-modulating properties of liquid crystal materials and the benefits of high performance silicon complementary metal-oxide-semiconductor technology by dedicated LCOS assembly processes. More precisely, the LCOS device can be either transmissive or reflective, besides, it can be used to modify the polarization or the phase of an incident light beam by the electrically modulated optical properties of liquid crystals. Applications for the LCOS include real-time holography, optical correlators, wavelength selective switches, and diffractive optical components, among others.

1.5.3. Diffractive optical elements This technology [69] simplifies the optical design and enhances the image quality of refractive optical systems. Specifically, DOEs are manufactured to have microstructure patterns that manipulate an incident light's phase to create the desired output pattern with special functionality. These elements are generally designed for a specific laser wavelength; their performance is wavelength-dependent. Applications for the DOEs include material processing, medical and aesthetic treatments, and semiconductor wafer inspection, among others.

1.5.4. Digital micromirror device This technology [70] contains millions of tiny switchable mirrors, which can be switched to either on or off states, e.g., each pixel rotates $+12^{\circ}$ or -12° concerning the surface normal. The on-state micromirror contributes to the reflected beam, while there is no contribution from the off-state micromirror. Applications for the DMD include laser beam shaping, correction of wavefront distortion, and high-resolution microscopy, among others.

1.6. PHASE RETRIEVAL ALGORITHMS

Here, some state-of-the-art iterative methods developed to solve the PR problem from CDPs are introduced. These algorithms can be classified according to the optimization problem statement, i.e., convex and non-convex formulations. Further, it is important to remark that these literature methods have been developed to solve the generalized PR problem $(\mathbf{y})_i = |\langle \mathbf{a}_i, \mathbf{x} \rangle|^2$ with $i \in \{1, \cdots, m\}$. These PR methods assume independent sampling vectors with complex-valued Gaussian distribution, i.e., $\mathbf{a}_i \sim C\mathcal{N}(0, \mathbf{I}_n) = \mathcal{N}(0, \frac{1}{2}\mathbf{I}_n) + j\mathcal{N}(0, \frac{1}{2}\mathbf{I}_n)$, where $\mathbf{I}_n \in \mathbb{R}^{n \times n}$ corresponds to the identity matrix. The PR algorithms described below have a computational complexity

$\mathcal{O}(mn\log(1/\epsilon))$ with $\epsilon > 0$.

1.6.1. Convex formulations These algorithms involve a convex relaxation that provides information about the PR problem.

- **PhaseLift** [23]**:** Casting the PR problem as a trace-minimization problem over an affine slice of the positive semidefinite cone, then, the PR problem based on the trace-norm relaxation is given by

  $$\begin{aligned} \underset{\mathbf{x}\in\mathbb{C}^n}{\text{minimize}} \quad & \text{Trace}\left(\mathbf{x}\mathbf{x}^{\mathcal{H}}\right) \\ \text{subject to} \quad & \mathcal{A}(\mathbf{x}\mathbf{x}^{\mathcal{H}}) = \mathbf{y} \\ & \mathbf{x}\mathbf{x}^{\mathcal{H}} \succeq 0, \end{aligned} \tag{15}$$

  where $\mathcal{A}(\cdot) : \mathbb{R}^n \to \mathbb{R}^m$ is a linear operator.

- **PhaseMax** [24]**:** Letting $\hat{\mathbf{x}} \in \mathbb{C}^n$ be an approximation vector to the true signal $\mathbf{x}$, the recovered signal is obtained by solving the following convex problem

  $$\begin{aligned} \underset{\mathbf{x}\in\mathbb{C}^n}{\text{maximize}} \quad & \langle \mathbf{x}, \hat{\mathbf{x}} \rangle_{\mathbb{R}} \\ \text{subject to} \quad & |\langle \mathbf{a}_i, \mathbf{x} \rangle| \leq (\boldsymbol{\phi})_i, \end{aligned} \tag{16}$$

  where $\langle \cdot \rangle_{\mathbb{R}}$ denotes the real part of the inner product and $(\boldsymbol{\phi})_i = \sqrt{(\mathbf{y})_i}$.

1.6.2. Non-convex formulations These algorithms require careful statistical initialization, which is refined based on a gradient-based method using the Wirtinger derivative.

- **Reweighted Amplitude Flow (RAF)** [25]**:** Following the least-squares criterion, which coincides with the maximum likelihood criterion when Gaussian white noise is assumed, thus, the PR problem can be formulated as

$$\underset{\mathbf{x}\in\mathbb{C}^n}{\text{minimize}}\, f(\mathbf{x}) := \frac{1}{2m}\sum_{i=1}^{m} g(\mathbf{x};(\boldsymbol{\phi})_i/(\mathbf{y})_i), \tag{17}$$

where $g(\mathbf{x};(\boldsymbol{\phi})_i/(\mathbf{y})_i)$ can be considered the amplitude-based loss function $g(\mathbf{x};(\boldsymbol{\phi})_i) := ((\boldsymbol{\phi})_i - |\langle \mathbf{a}_i, \mathbf{x}\rangle|)^2$ [26], or the intensity-based loss function $g(\mathbf{x};(\mathbf{y})_i) := ((\mathbf{y})_i - |\langle \mathbf{a}_i, \mathbf{x}\rangle|^2)^2$ [59, 71], or the Poisson likelihood $g(\mathbf{x};(\mathbf{y})_i) = -(\mathbf{y})_i \log(|\langle \mathbf{a}_i, \mathbf{x}\rangle|^2) + |\langle \mathbf{a}_i, \mathbf{x}\rangle|^2$ [27].

- **Truncated Amplitude Flow (TAF)** [26]**:** Adopting the least-squares criterion, the task of recovering an unknown signal from diffraction patterns with additive white Gaussian noise reduces to that of minimizing the amplitude-based loss function

$$\underset{\mathbf{x}\in\mathbb{C}^n}{\text{minimize}}\, f(\mathbf{x}) := \frac{1}{2m}\sum_{i=1}^{m}(|\langle \mathbf{a}_i, \mathbf{x}\rangle| - (\boldsymbol{\phi})_i)^2. \tag{18}$$

- **Truncated Wirtinger Flow (TWF)** [27]**:** Assuming $(\mathbf{y})_i = |\langle \mathbf{a}_i, \mathbf{x}\rangle|^2$, $i \in \{1, \cdots, m\}$ as a sequence of independent samples of the Poisson distribution, i.e., $(\mathbf{y})_i \sim \text{Poisson}(|\langle \mathbf{a}_i, \mathbf{x}\rangle|^2)$. Calculating the log-likelihood for independent samples, we have

$$g(\mathbf{x};(\mathbf{y})_i) = (\mathbf{y})_i \log(|\langle \mathbf{a}_i, \mathbf{x}\rangle|^2) - |\langle \mathbf{a}_i, \mathbf{x}\rangle|^2. \tag{19}$$

Then, the maximum likelihood can be expressed as

$$\underset{\mathbf{x}\in\mathbb{C}^n}{\text{minimize}}\, f(\mathbf{x}) := -\sum_{i=1}^{m} g(\mathbf{x};(\mathbf{y})_i). \tag{20}$$

---

[71] Yonina C Eldar and Shahar Mendelson. “Phase retrieval: Stability and recovery guarantees”. In: *Applied and Computational Harmonic Analysis* 36.3 (2014), pp. 473–494.

1.7. DEEP UNROLLED NETWORKS

Deep neural networks (DNNs) have been used to solve the PR problem [28,29,30]. The DNN architectures are often treated as black boxes, which limits their interpretability [31]. Deep unrolling or unfolding techniques address this issue by enabling model-aware design within DNN architectures, drawing from iterative signal processing algorithms [72]. In this approach, each unrolled network block learns the solution of one iteration in the optimization algorithm [31]. The unrolling methods leverage data to improve accuracy and performance, enhancing interpretability and leading to more trusted outcomes. As a result, these algorithms reduce the large dataset requirement and accelerate convergence rates.

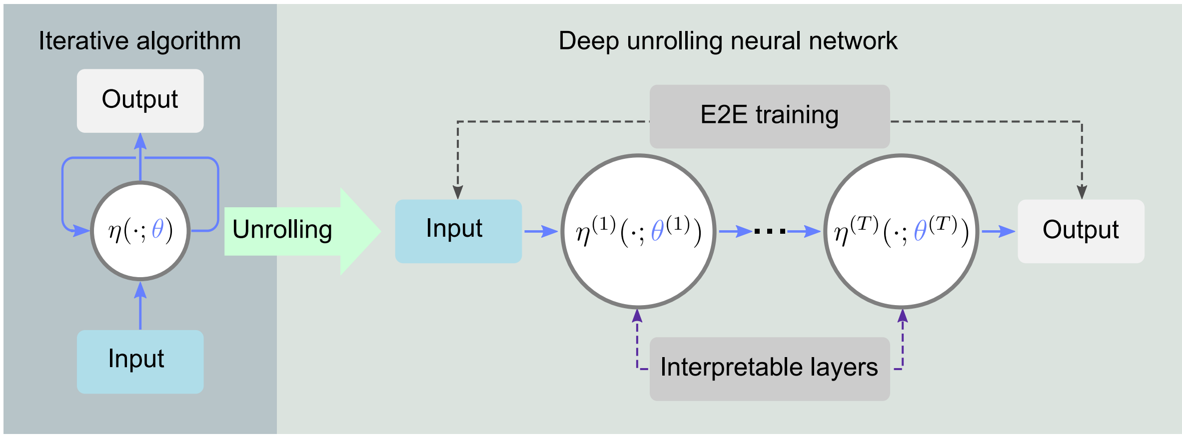
Figure 9 illustrates a high-level overview of the unrolling algorithm. In this process, a DNN can be described by cascading iterations η from a given iterative algorithm [73]. The iteration step η is executed T times, resulting in network layers $\{\eta^{(1)}, \ldots, \eta^{(T)}\}$. Each iteration η depends on algorithm parameters θ, which are transferred into network parameters $\{\theta^{(1)}, \cdots, \theta^{(T)}\}$. Instead of determining these parameters through cross-validation or analytical derivations, the unrolling approach learns them from training datasets through end-to-end (E2E) training. This approach enables the resulting network to perform better than the original iterative algorithm. Additionally, the network layers naturally inherit interpretability from the iteration procedure [73].

---

[72] Gregory Ongie et al. “Deep learning techniques for inverse problems in imaging”. In: *IEEE Journal on Selected Areas in Information Theory* 1.1 (2020), pp. 39–56.

[73] Vishal Monga, Yuelong Li, and Yonina C Eldar. “Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing”. In: *IEEE Signal Processing Magazine* 38.2 (2021), pp. 18–44.

Figure 9. Overview of a deep unrolling neural network via an iterative algorithm. (Left) Iterative algorithm. (Right) Unrolled DNN according to the given iterative algorithm. This DNN learns the parameters via E2E training.



Source: Modified from Monga, V., Li, Y., & Eldar, Y. C. (2021). Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing. IEEE Signal Processing Magazine, 38(2), 18-44.

1.8. END-TO-END OPTIMIZATION

The E2E training approach can be applied to optics design by jointly optimizing an optical encoder and a computational decoder [74]. In this context, E2E deep learning approach focuses on simultaneously learning the parameters of the sensing protocol and the deep network by designing a fully differentiable forward model of the optical system, which is coupled as a layer in the DNN to conduct any particular

---

[74] Lingen Li et al. “Quantization-aware deep optics for diffractive snapshot hyperspectral imaging”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, pp. 19780–19789.

computational imaging task, for instance, recovery [75], detection [76], classification [77] and segmentation [78] as illustrated in Fig. 10.

Recently, [39] proposes a customizable E2E approach that jointly designs the sensing matrix $\boldsymbol{\Psi}$ and learns the parameters $\boldsymbol{\theta}$ of a chosen DNN $\mathcal{M}(\cdot)$ for computational tasks $\mathcal{L}_{task}(\cdot)$ in optical coding systems based on incoherent light using a regularization function $\mathcal{R}(\cdot)$. From $\mathcal{S}$ scenes $\{\mathbf{x}_s\}_{s=1}^{\mathcal{S}}$, the E2E optimization problem in [39] is defined as

$$\begin{aligned} \{\boldsymbol{\Psi}^*, \boldsymbol{\theta}^*\} &\in \underset{\boldsymbol{\Psi},\boldsymbol{\theta}}{\operatorname{argmin}}\, \mathcal{L}\left(\boldsymbol{\Psi}, \boldsymbol{\theta}|\mathbf{x}, \mathbf{z}\right), \\ \mathcal{L}\left(\boldsymbol{\Psi}, \boldsymbol{\theta}|\mathbf{x}, \mathbf{z}\right) &= \frac{1}{\mathcal{S}} \sum_{s=1}^{\mathcal{S}} \mathcal{L}_{task}\left(\mathcal{M}_{\boldsymbol{\theta}}\left(\boldsymbol{\Psi}, \mathbf{x}_s\right), \mathbf{z}_s\right) + \rho \mathcal{R}\left(\boldsymbol{\Psi}\right), \end{aligned} \tag{21}$$

where $\rho > 0$ is a regularization parameter. Figure 10 shows the two-module procedure that models the encoder and decoder steps to solve (21). Here, the first stage corresponds to the sensing protocol, which consists of an optical layer that learns the sensing matrix; and the second stage contains a chosen trainable DNN with multiple hidden layers for achieving a specific computational task.

---

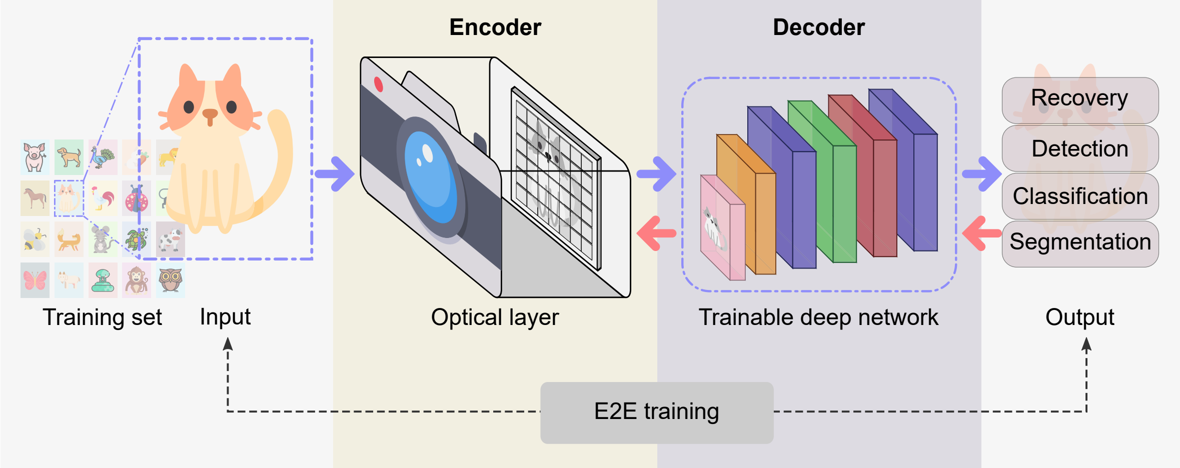
[75] Mário AT Figueiredo, Robert D Nowak, and Stephen J Wright. “Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems”. In: *IEEE Journal of selected topics in signal processing* 1.4 (2007), pp. 586–597.

[76] Suhas Lohit, Kuldeep Kulkarni, and Pavan Turaga. “Direct inference on compressive measurements using convolutional neural networks”. In: *2016 IEEE International Conference on Image Processing (ICIP)*. IEEE. 2016, pp. 1913–1917.

[77] Jorge Bacca, Laura Galvis, and Henry Arguello. “Coupled deep learning coded aperture design for compressive image classification”. In: *Optics Express* 28.6 (2020), pp. 8528–8540.

[78] Carlos Hinojosa, Jorge Bacca, and Henry Arguello. “Coded aperture design for compressive spectral subspace clustering”. In: *IEEE Journal of Selected Topics in Signal Processing* 12.6 (2018), pp. 1589–1600.

Figure 10. E2E approach for optical design in different computational tasks. First, each sample from the training set passes through the encoder, which describes the corresponding propagation model. Then, the resulting sample projections enter into the decoder, which defines the network specific to each computational task. Both encoder and decoder parameters are learned during the E2E training.



It is worthwhile to mention that the proposed method in [39] is limited to work on the real-valued domains, i.e., it cannot be applied to the PR problem because of the critical differences led by the complex nature of the sensing matrix. More recently, the PR problem has been addressed from E2E optimization [31,40,36,79].

1.9. DIFFRACTIVE SPECTRAL IMAGING

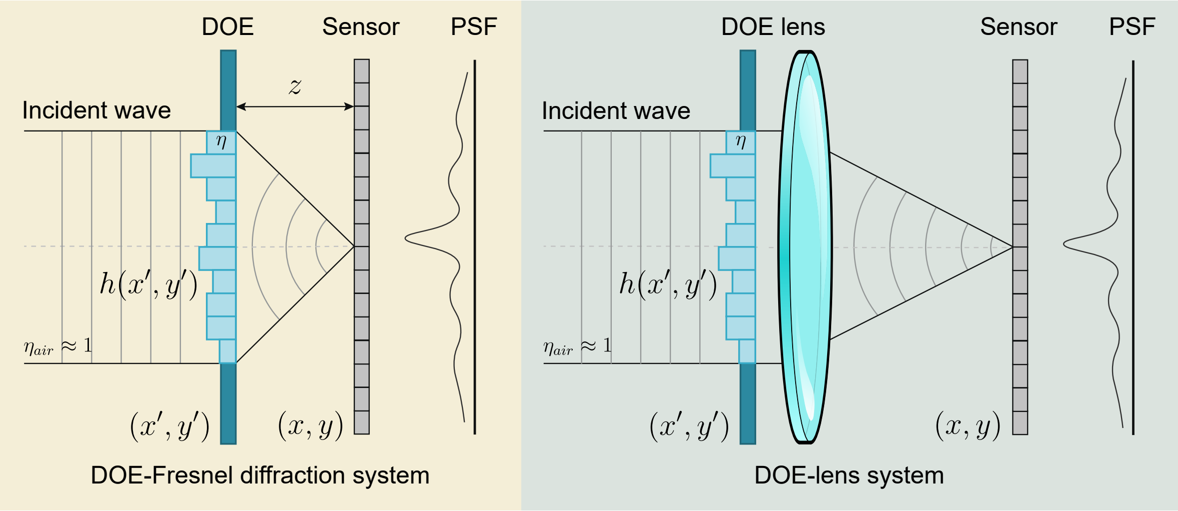
Based on chromatic aberration, traditional diffractive imaging has been devised for monochromatic (coherent) light using a single wavelength. Currently, diffractive imaging methods have been introduced even for polychromatic (incoherent) illumination using multiple wavelengths [42]. In particular, the diffractive model has been incorporated into spectral imaging, where multiple wavelengths are recorded from a target across the electromagnetic spectrum. This optical system contains a DOE

---

[79] Qiuliang Ye, Li-Wen Wang, and Daniel PK Lun. “SiSPRNet: end-to-end learning for single-shot phase retrieval”. In: *Optics Express* 30.18 (2022), pp. 31937–31958.

(transmissive or reflective lens) and a bare camera sensor, separated by a distance z. Figure 11 illustrates two phase-coded spectral imaging systems: the DOE-Fresnel diffraction system and the DOE-lens system.

Figure 11. Schematic representation of diffractive spectral imaging via a DOE. The left system uses a transmissive DOE and a sensor, where the incident wave passes a DOE and then propagates a distance z before being recorded by the sensor. The propagation can be modeled via the Fresnel approximation. The right system uses an imaging lens just behind the DOE. After passing the DOE, the incident wave converged on the sensor through the lens. DOE has a height profile that introduces the phase shift.



Source: Modified from Huang, L., Luo, R., Liu, X., & Hao, X. (2022). Spectral imaging with deep learning. Light: Science & Applications, 11(1), 61.

These systems can be modeled using the point spread function (PSF) $p_\lambda(x, y)$, which represents the system response to a point source at the plane image [48]. The wave field $A_\lambda(x', y')e^{j\varphi(x',y')}$ first experiences a phase shift $\phi(x', y')$ determined by the height profile $h(x', y')$ of the DOE

$$\gamma_\lambda(x', y'; 0) = A_\lambda(x', y')e^{j[\varphi(x',y')+\phi(x',y')]}, \tag{22}$$

where $\phi(x', y') = k_0 \Delta\eta h(x', y')$. Here, $\Delta\eta$ corresponds to the refractive index difference between air η_{air} and the DOE, and k_0 is the wavenumber.

For the DOE-lens system, the PSF can be defined as

$$p_\lambda(x, y) = \left|\mathcal{F}^{-1}\left\{\gamma_\lambda(x', y'; 0)\right\}\right|. \tag{23}$$

For the DOE-Fresnel diffraction system, the wave field propagates a distance z that can be modeled by the Fresnel diffraction law such that $\lambda \ll z$

$$\gamma_\lambda(x, y; z) = \frac{e^{jk_0 z}}{j\lambda z} \iint \gamma_\lambda(x', y'; 0) e^{\frac{jk_0}{2z}\left[(x'-x)^2+(y'-y)^2\right]} dx'dy'. \tag{24}$$

The resulting PSF can be formulated as

$$p_\lambda(x, y) \propto \left|\mathcal{F}\left\{A_\lambda(x', y') e^{j[\varphi(x',y')+\phi(x',y')]} e^{\frac{j\pi}{\lambda z}\left(x'^2+y'^2\right)}\right\}\right|^2. \tag{25}$$

Finally, the spectral image formation model based on the PSF can be expressed as

$$I(x, y) = \int_\Lambda \delta(\lambda)\left[I_\lambda(x, y) * p_\lambda(x, y)\right] d\lambda, \tag{26}$$

where $I_\lambda(x, y)$ represents the spectral image, $\delta(\lambda)$ corresponds to the sensor spectral sensitivity function, and $*$ is the convolution operator.

1.10. DIFFRACTIVE OPTICAL IMAGING CURRENT STATE

The following list includes the main state-of-the-art reports in the area that concerns this dissertation.

- **Phase smoothing for diffractive deep neural networks** [80] **(2024). Optics Communications.** This work proposes a diffractive network training framework

---

[80] Lin Wu. "Phase smoothing for diffractive deep neural networks". In: *Optics Communications* (2024), p. 130267.

to smooth the MPM while increasing the numerical modeling accuracy. The cascaded smooth MPMs can narrow the mismatch between the digital offline design and the practical optical deployment.

- **Phase unwrapping for phase imaging using the plug-and-play proximal algorithm [81] (2024). Applied Optics.** This work introduces an iterative method based on the plug-and-play proximal algorithm to unwrap two-dimensional phase values while simultaneously removing noise at each iteration.

- **On the use of deep learning for phase recovery [82] (2024). Light: Science & Applications.** This work examines deep learning techniques for PR and offers an outlook on how to apply deep learning more effectively to improve PR reliability and efficiency.

- **Amplitude/phase retrieval for terahertz holography with supervised and unsupervised physics-informed deep learning [83] (2024). IEEE Transactions on Terahertz Science and Technology.** This paper proposes novel supervised and unsupervised physics-informed deep learning methods for amplitude and phase recovery by incorporating angular spectrum diffraction theory as prior knowledge.

- **Unfolding-aided bootstrapped phase retrieval in optical imaging: Explainable AI reveals new imaging frontiers [31] (2023). IEEE Signal Processing**

---

[81] Jhon Ramirez, Henry Arguello, and Jorge Bacca. “Phase unwrapping for phase imaging using the plug-and-play proximal algorithm”. In: *Applied Optics* 63.2 (2024), pp. 535–542.

[82] Kaiqiang Wang et al. “On the use of deep learning for phase recovery”. In: *Light: Science & Applications* 13.1 (2024), p. 4.

[83] Mingjun Xiang et al. “Amplitude/phase retrieval for terahertz holography with supervised and unsupervised physics-informed deep learning”. In: *IEEE Transactions on Terahertz Science and Technology* (2024).

Magazine. This paper presents an overview of algorithms and applications related to deep unfolding networks for addressing the PR problem.

- **Phase retrieval with only a nonnegativity constraint** [84] **(2023). Electronic Imaging.** This letter demonstrates the potential for reconstructing a real-valued nonnegative two-dimensional object from the magnitude of its Fourier transform, even in the presence of significant noise in the Fourier intensity data, using only nonnegativity.

- **Hybrid diffractive optics (DOE & refractive lens) for broadband EDoF imaging** [85] **(2023). Electronic Imaging.** This paper proposes a fully differentiable image formation model that uses neural network techniques to maximize the imaging quality by optimizing MPMs; digital image reconstruction algorithms, refractive lens parameters (aperture size and focal length), and distance between the MPM and the sensor.

- **Phase retrieval: From computational imaging to machine learning: A tutorial** [86] **(2023). IEEE Signal Processing Magazine.** This tutorial reviews the PR problem under a unifying framework encompassing classical and machine learning methods, focusing on applications, an overview of recent reconstruction algorithms, and the latest theoretical results.

- **Unrolled wirtinger flow with deep decoding priors for phaseless imag-**

---

[84] Yaocheng Tian and James R Fienup. "Phase retrieval with only a nonnegativity constraint". In: *Optics Letters* 48.1 (2023), pp. 135–138.

[85] SeyyedReza MiriRostami et al. "Hybrid diffractive optics (DOE & refractive lens) for broadband EDoF imaging". In: *Electronic Imaging* 35 (2023), pp. 1–14.

[86] Jonathan Dong et al. "Phase retrieval: From computational imaging to machine learning: A tutorial". In: *IEEE Signal Processing Magazine* 40.1 (2023), pp. 45–57.

ing [87] (2022). IEEE Transactions on Computational Imaging. This work combines the Wirtinger flow algorithm and a deterministic theory for exact non-convex PR with a deep learning-based approach. Specifically, this work involves using deep decoding prior in conjunction with deep learning-based initialization and unrolling the Wirtinger flow algorithm into a recurrent DNN architecture, which enables E2E training.

- **DualPRNet: Deep shrinkage dual frame network for deep unrolled phase retrieval [35] (2022). IEEE Signal Processing Letters.** This work introduces a deep shrinkage network for supervised dual frame learning, dubbed DualNet. The authors develop a deep unrolled PR network based on the implemented dual frame architecture.

- **All-optical phase recovery: diffractive computing for quantitative phase imaging [88] (2022). Advanced Optical Materials.** This work proposes a diffractive quantitative phase imaging network to perform an all-optical PR.

- **Spectral imaging with deep learning [48] (2022). Light: Science & Applications.** This review paper summarizes deep learning techniques for computational spectral imaging. The spectral imaging systems are divided into amplitude-coded, phase-coded, and wavelength-coded methods, based on the different light properties used for encoding.

- **Diffractive deep neural networks at visible wavelengths [89] (2021). Engi-**

---

[87] Samia Kazemi, Bariscan Yonel, and Birsen Yazici. “Unrolled wirtinger flow with deep decoding priors for phaseless imaging”. In: *IEEE Transactions on Computational Imaging* 8 (2022), pp. 609–625.

[88] Deniz Mengu and Aydogan Ozcan. “All-optical phase recovery: diffractive computing for quantitative phase imaging”. In: *Advanced Optical Materials* 10.15 (2022), p. 2200281.

[89] Hang Chen et al. “Diffractive deep neural networks at visible wavelengths”. In: *Engineering* 7.10

neering. One landmark method is the diffractive networks based on three-dimensional printing technology, operating in the terahertz spectral range. This work proposes a general formula containing a detailed analysis of the diffractive network framework, including the required design parameters.

- **Phase coded speckle illumination for laser Fourier ptychographic microscopy** [90] **(2021). Optics Communications.** This work proposes a Fourier psychographic microscopy system based on phase-coded speckle illumination that employs a laser source with a speckle illumination control optical system. In particular, a phase-only spatial light modulator with a condenser lens is utilized in the proposed system to project the phase-coded speckle illumination patterns onto the sample.

- **Shift-variant color-coded diffractive spectral imaging system** [38] **(2021). Optica.** This work proposes an optical setup composed of a DOE and a color-coding mask for spectral modulation. These optical elements are jointly designed from a fully differentiable image formation model based on an E2E approach, which minimizes the deviation between the reference and reconstructed images over a large dataset.

- **PhaseGAN: a deep learning phase retrieval approach for unpaired datasets** [91] **(2021). Optics express.** This work proposes a deep learning approach based on generative adversarial networks, which allows unpaired dataset usage and includes image formation physics. Also, the proposed deep architec-

---

(2021), pp. 1483–1491.

[90] Xiao Tao et al. “Phase-coded speckle illumination for laser Fourier ptychographic microscopy”. In: *Optics Communications* 498 (2021), p. 127199.

[91] Yuhe Zhang et al. “PhaseGAN: a deep-learning phase-retrieval approach for unpaired datasets”. In: *Optics express* 29.13 (2021), pp. 19593–19604.

ture incorporates a novel Fourier loss for phase recovery.

- **Analysis of diffractive optical neural networks and their integration with electronic neural networks** [92] **(2020). IEEE Journal of Selected Topics in Quantum Electronics.** This paper proposes integrating diffractive networks with electronic neural networks to create hybrid classifiers from a five-layer phase-only network jointly optimized with a single fully connected electronic layer.

- **Phase recovery guarantees from designed coded diffraction patterns in optical imaging** [12] **(2020). IEEE Transactions on Image Processing.** This paper provides theoretical guarantees for recovering a scene from CDPs acquired at each diffraction field using admissible coding variables.

- **PhlatCam: Designed phase-mask-based thin lensless camera** [93] **(2020). IEEE Transactions on Pattern Analysis and Machine Intelligence.** This paper proposes a near-field PR framework to design phase masks that produce the target point-spread functions at the desired device thickness.

- **Convex phase retrieval with multiple structured illuminations design** [94] **(2020). Journal of Optics.** This paper proposes a physically realistic optical imaging system that can achieve the design of different masks using the optical field modulation characteristics of the spatial light modulator.

---

[92] Deniz Mengu et al. “Analysis of diffractive optical neural networks and their integration with electronic neural networks”. In: *IEEE Journal of Selected Topics in Quantum Electronics* 26.1 (2019), pp. 1–14.

[93] Vivek Boominathan et al. “PhlatCam: Designed phase-mask based thin lensless camera”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2020).

[94] Cheng Zhang et al. “Convex phase retrieval with multiple structured illuminations design”. In: *Journal of Optics* (2020), pp. 1–6.

"Research is formalized curiosity. It is poking and prying with a purpose."
- ***Zora Neale Hurston***

---

2. LEARNABLE SPECTRAL INITIALIZATION FOR PHASE RETRIEVAL

This chapter introduces an end-to-end (E2E) initialization framework for phase retrieval (PR). This E2E approach learns the filter kernel within the spectral initialization and optimizes the neural network parameters. The proposed deep learning initialization approach involves an optical stage that simulates the forward propagation model in diffractive optical architectures based on coherent light. This optical stage models the coded diffraction patterns (CDPs) acquired at each diffraction field: near-field, middle-field, and far-field. Subsequently, an initialization stage approximates the underlying optical field from the simulated CDPs using the filtered spectral initialization (FSI), which is implemented as an unrolled network approach, where each iteration corresponds to a network layer computing the optical field estimation, and training the kernel filter through a complex convolutional layer that shares weights among stages. Then, the obtained initialization is improved by using a double-branch U-net-based deep neural network (DNN) that performs the reconstruction process. Specifically, this double-branch DNN separately recovers the phase and magnitude information from the incoming optical field. Simulation results show that the learnable filtered initialization based on the proposed deep learning method outperforms the state-of-the-art initialization algorithms, thereby achieving comparable reconstruction quality using a single snapshot and requiring fewer iterations in the initialization step compared to previous PR works. In summary, the proposed E2E approach learns the initialization process for approximating the complex-valued optical field based on an unrolled initialization network, which reduces the number of iterations compared to traditional PR initializations. The reconstruction stage is performed via a double-branch DNN from a single snapshot. This chapter addresses the specific objectives 1 and 2.

The main contributions of this work are listed as follows:

- An E2E scheme is proposed for addressing the PR problem based on a learnable filtered spectral initialization and a double-branch recovery network.
- The kernel filter involved in the filtered spectral initialization is learned from the proposed E2E deep network. This filter improves the optical field approximation at each diffraction field while reducing the number of iterations and snapshots.
- The proposed deep network achieves optical field reconstruction from a single snapshot.

2.1. CODED DIFFRACTIVE ACQUISITION SYSTEM

The diffractive optical system that collects CDPs at each diffraction field is described in this section. Specifically, this system consists of a coherent light source that illuminates an unknown object $\mathbf{x} \in \mathbb{C}^n$ placed in the optical field; then, the optical field is modulated by a multilevel phase mask (MPM) $\mathbf{D}_\ell \in \mathbb{C}^{n\times n}$ introduced in the object plane; finally, the propagated CDPs are acquired by a sensor located at each diffraction field $\psi \in \{1,2,3\}$, as shown in Fig. 12. Mathematically, the CDPs $(\mathbf{y}_{\ell,\psi})_k \in \mathbb{R}^n$ at ℓ-th snapshot and ψ-th diffraction field can be modeled as

$$(\mathbf{y}_{\ell,\psi})_k = |\mathbf{a}_{k,\psi}^{\mathcal{H}}\mathbf{D}_\ell\mathbf{x}|^2 + (\boldsymbol{\omega}_{\ell,\psi})_k, \tag{27}$$

for $k \in \{0,\cdots,n-1\}$ and $\ell \in \{1,\dots,L\}$ with L as the total number of snapshots; where $(\cdot)^{\mathcal{H}}$ is the conjugate transpose operation, $|\cdot|$ denotes the pointwise magnitude, $(\boldsymbol{\omega}_{\ell,\psi})_k \in \mathbb{R}$ is the additive Gaussian noise, and $\mathbf{a}_{k,\psi}^{\mathcal{H}} \in \mathbb{C}^n$ is the k-th row of the matrix

$$\mathbf{A}_\psi = \begin{cases} \mathbf{F}\mathbf{T}\mathbf{F}^{\mathcal{H}} & \text{if } \psi = 1 \rightarrow \text{ Near-field} \\ \mathbf{F}^{\mathcal{H}}\mathbf{Q} & \text{if } \psi = 2 \rightarrow \text{ Middle-field}, \\ \mathbf{F} & \text{if } \psi = 3 \rightarrow \text{ Far-field} \end{cases} \tag{28}$$

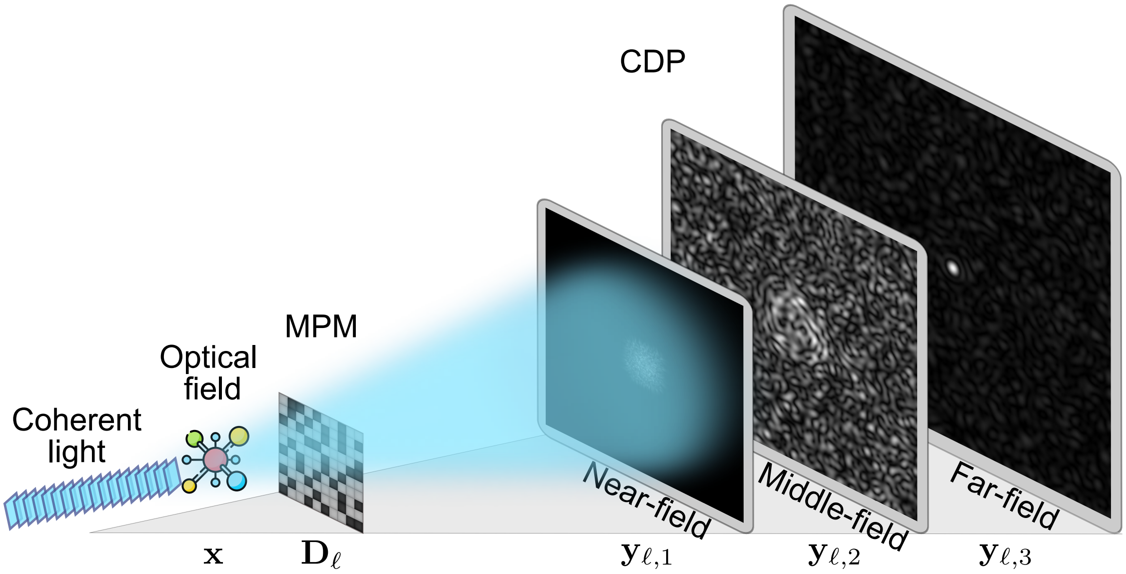
with $\mathbf{F} \in \mathbb{C}^{n\times n}$ as the discrete Fourier matrix, and $\mathbf{T} \in \mathbb{C}^{n\times n}$ and $\mathbf{Q} \in \mathbb{C}^{n\times n}$ as the transfer functions in Fourier optics [6] for near-field and middle-field, respectively, which are mathematically described as [12]

$$(\mathbf{T})_{r,s} = e^{-jk_0 z\sqrt{1-\frac{(r\delta_{kx})^2}{k_0^2}-\frac{(s\delta_{ky})^2}{k_0^2}}}, \text{ and} \tag{29}$$

$$(\mathbf{Q})_{p,q} = e^{\frac{-jk_0}{2z}(p^2\delta_x^2+q^2\delta_y^2)}, \tag{30}$$

where $j = \sqrt{-1}$, z is the propagation distance, $k_0 = 2\pi/\lambda$ is the wavenumber with λ the wavelength [6], and $[r, s]$ and $[p, q]$ are the discrete indices of the samples in the Fourier and spatial domains, respectively. Further, the terms δ_{kx} and δ_{ky} are the sampling periods in the frequency coordinates, and the terms δ_x and δ_y are the sampling period in the spatial coordinates.

Figure 12. Schematic representation of a diffractive optical system that acquires CDPs at each diffraction field. A coherent light source illuminates an object located in the optical field. Then, an MPM encodes the optical field, which is recorded by a sensor that can be placed at different propagation distances to capture CDPs at the near-field, middle-field, and far-field.



2.2. PROPOSED DEEP LEARNING APPROACH FOR PHASE RETRIEVAL

The proposed E2E approach, which is illustrated in Fig. 13, employs a set of $\mathcal{S}$ training scenes $\{\mathbf{x}^{(s)}\}_{s=1}^{\mathcal{S}}$ to learn the filter kernel $\mathbf{G} \in \mathbb{C}^{u\times v}$ included in the initialization stage to produce $\{\tilde{\mathbf{z}}^{(s)}\}_{s=1}^{\mathcal{S}}$. In this model, the filter kernel $\mathbf{G}$ and the network parameters $\boldsymbol{\theta}$ are jointly optimized from a DNN $\mathcal{M}_{\boldsymbol{\theta}}(\cdot)$ by minimizing the loss function defined as [95]

$$\underset{\mathbf{G},\boldsymbol{\theta}}{\operatorname{argmin}} \frac{1}{\mathcal{S}} \sum_{s=1}^{\mathcal{S}} \|\mathcal{M}_{\boldsymbol{\theta}}(\tilde{\mathbf{z}}^{(s)}) - \mathbf{x}^{(s)}\|_2^2 + \rho \mathcal{R}_{\Omega}(\mathbf{G}), \tag{31}$$

where $\rho > 0$ is a regularization parameter and $\mathcal{R}_{\Omega}(\cdot)$ is an indicator function applied to $\mathbf{G}$. Specifically, we constrain the magnitude values of the learned kernel $\mathbf{G}$ to be in the range $[0,1]$ to ensure that the filtering step does not increase the intensity of the recovered image. Mathematically, this kernel filter can be constrained as follows

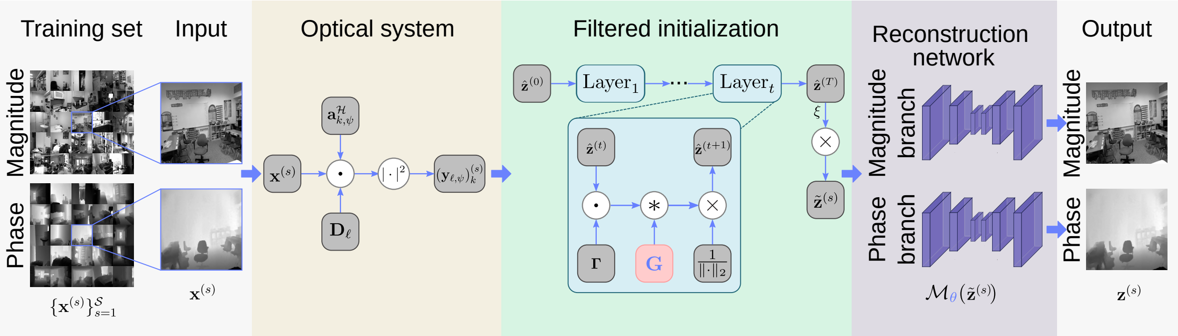
$$\mathcal{R}_{\Omega}(\mathbf{G}) = \begin{cases} 0 & \text{if } \mathbf{G} \in \Omega \\ +\infty & \text{if } \mathbf{G} \notin \Omega \end{cases}, \tag{32}$$

where $\Omega = \{\mathbf{G} \in \mathbb{C}^{u\times v} | |(\mathbf{G})_{p,q}| \leq 1\}$.

---

[95] David Morales, Andrés Jerez, and Henry Arguello. “Deep Phase Retrieval by a Learnable Filtered Spectral Initialization”. In: *Computational Optical Sensing and Imaging*. Optical Society of America. 2021, CTh4A–6.

Figure 13. Proposed E2E deed initialization approach based on a double-branch recovery network. Firstly, each complex-valued input image $\mathbf{x}^{(s)}$ passes through the optical acquisition stage that simulates the CDPs $(\mathbf{y}_{\ell,\psi})_k^{(s)}$. Secondly, the initialization stage approximates the optical field $\tilde{\mathbf{z}}^{(s)}$. Finally, the reconstruction network $\mathcal{M}_\theta(\tilde{\mathbf{z}}^{(s)})$ recovers the optical field from the obtained initialization.



In particular, the proposed E2E methodology comprises three main stages: an optical acquisition stage, a filtered initialization stage, and a reconstruction network. Here, each image input $\mathbf{x}^{(s)}$ of the training set is evaluated in the optical acquisition stage, which performs the measurement process for obtaining the CDPs $(\mathbf{y}_{\ell,\psi})_k^{(s)}$. Then, the FSI [96] algorithm is implemented as an unrolled network $\mathcal{I}_{\mathbf{G}}(\cdot)$ into the deep learning scheme, where the filter $\mathbf{G}$ is defined as a convolutional layer. The initialization stage approximates $\tilde{\mathbf{z}}$ based on the acquired CDPs from T number of iterations. Finally, the initial guess $\tilde{\mathbf{z}}$ is improved by a double-branch U-net architecture [97], returning $\mathbf{z}^{(s)}$. The Adam optimizer is used to minimize (31).

---

[96] Andrés Jerez, Samuel Pinilla, and Henry Arguello. "Fast target detection via template matching in compressive phase retrieval". In: *IEEE Transactions on Computational Imaging* 6 (2020), pp. 934–944.

[97] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical image computing and computer-assisted intervention*. Springer. 2015, pp. 234–241.

2.2.1. Filtered initialization This section introduces the spectral initialization method proposed in [96], which is summarized in Algorithm 1. FSI approximates an initial guess of the optical field $\tilde{\mathrm{z}}$ by computing a low-pass version of the leading eigenvector of the matrix

$$\mathbf{\Gamma}_{\psi} := \frac{1}{\mathrm{card}(\Xi)} \sum_{(\ell,k)\in\Xi} \frac{\overline{\mathbf{D}}_{\ell}\mathbf{a}_{k,\psi}\mathbf{a}_{k,\psi}^{\mathcal{H}}\mathbf{D}_{\ell}}{\|\mathbf{D}_{\ell}\mathbf{a}_{k,\psi}\|_2^2}, \tag{33}$$

scaled by the quantity $\xi := \sqrt{\frac{\sum_{\ell=1}^{L}\sum_{k=0}^{n-1}(\mathbf{y}_{\ell,\psi})_k}{nL}}$, i.e, $\tilde{\mathbf{z}} = \xi\hat{\mathbf{z}}$, where the set Ξ contains the indices associated to the $\lfloor\frac{nL}{2}\rfloor$ largest values of $\{(\mathbf{y}_{\ell,\psi})_k/\|\mathbf{a}_{k,\psi}\|_2\}$. The notation $\mathrm{card}(\Xi)$ represents the cardinality of the set Ξ, and $\overline{(\cdot)}$ is the complex conjugate operation. The low-pass constraint of FSI comes from the fact that an image is mostly composed of low frequencies [98]. Since FSI requires the estimation of the leading eigenvector of $\mathbf{\Gamma}_{\psi}$ across each diffraction field, the simplest way to achieve it is following a power iteration strategy [26]. This strategy involves recursively performing a matrix-vector multiplication between $\mathbf{\Gamma}_{\psi}$ and the current estimation of the scene. This process applies a low-pass filter $\mathbf{G} \in \mathbb{C}^{u\times v}$ to preserve the low-frequency information. Specifically, FSI facilitates solving the following optimization problem

$$\tilde{\mathbf{z}} = \underset{\|\mathbf{z}\|_2=1}{\arg\max}\, \mathbf{z}^{\mathcal{H}}\mathbf{\Gamma}_{\psi}\mathbf{z}. \tag{34}$$

Notice that the matrix $\mathbf{\Gamma}_{\psi}$ is built in line 4 of the Algorithm 1. In line 6, the filtering process over $\mathbf{\Gamma}_{\psi}\hat{\mathbf{z}}^{(t)}$ is computed by using the convolution operation $*$ based on the fixed kernel filter $\mathbf{G}$. Then, the filtered approximation is normalized in line 7. Finally, this algorithm returns the complex vector $\tilde{\mathrm{z}}$ in line 10, which is multiplied by the

---

[98] Paul A Lynn and Wolfgang Fuerst. *Introductory digital signal processing with computer applications*. John Wiley & Sons, 1998.

scaling factor ξ.

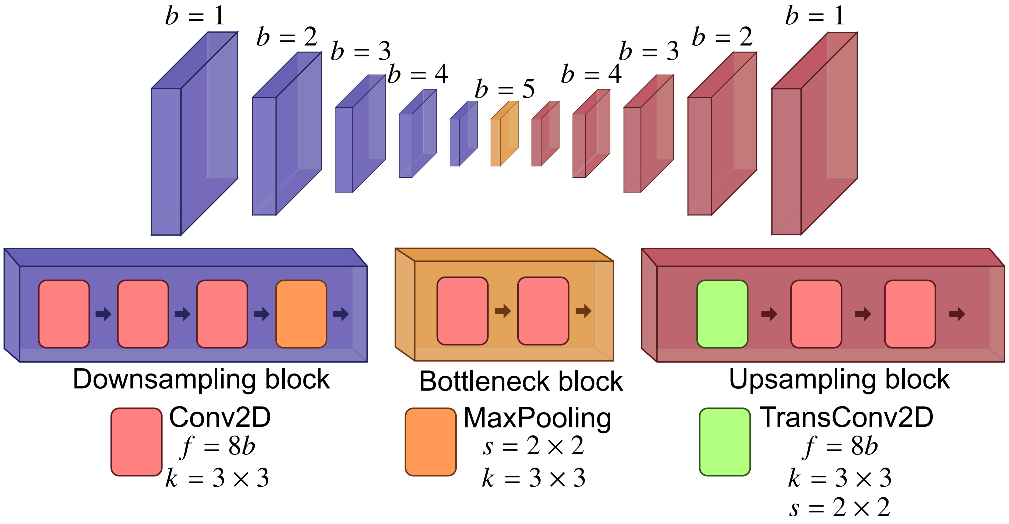
This initialization approach is implemented as an unrolled network in the proposed deep learning scheme as illustrated in Fig. 13, where the kernel filter $\mathbf{G}$ is a trainable parameter.

Algorithm 1 Filtered spectral initialization

```
Input: Acquired data $\{\overline{\mathbf{D}}_\ell \mathbf{a}_{k,\psi}; (\mathbf{y}_{\ell,\psi})_k\}$, maximum number of iterations $T$, and a low pass filter $\mathbf{G}$.
Initialize: $\hat{\mathbf{z}} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$.
Set $\Xi$ as the set of indices corresponding to the $\mathrm{card}(\Xi)$ largest values of $\{(\mathbf{y}_{\ell,\psi})_k / \|\mathbf{a}_{k,\psi}\|_2\}$.
$$\boldsymbol{\Gamma}_\psi := \frac{1}{\mathrm{card}(\Xi)} \sum_{(\ell,k)\in\Xi} \frac{\overline{\mathbf{D}}_\ell \mathbf{a}_{k,\psi} \mathbf{a}_{k,\psi}^{\mathcal{H}} \mathbf{D}_\ell}{\|\mathbf{D}_\ell \mathbf{a}_{k,\psi}\|_2^2}.$$
for $t = 0 : T-1$ do
    $\grave{\mathbf{z}}^{(t+1)} = \mathbf{G} * \left(\boldsymbol{\Gamma}_\psi \hat{\mathbf{z}}^{(t)}\right)$.    ▷ Filtering step
    $\hat{\mathbf{z}}^{(t+1)} = \frac{\grave{\mathbf{z}}^{(t+1)}}{\left\|\grave{\mathbf{z}}^{(t+1)}\right\|_2}$.    ▷ Normalizing step
end for
Compute $\tilde{\mathbf{z}} = \sqrt{\frac{\sum_{\ell=1}^{L}\sum_{k=0}^{n-1}(\mathbf{y}_{\ell,\psi})_k}{nL}}\hat{\mathbf{z}}^{(T)}$.    ▷ Scaling step
Return: Initial guess of the optical field $\tilde{\mathbf{z}}$.
```

2.2.2. Reconstruction network The reconstruction network $\mathcal{M}_{\boldsymbol{\theta}}(\cdot)$ chosen for the proposed deep learning approach is a double-branch neural network that separately recovers the magnitude information and the phase information provided by the FSI stage, as shown in Fig. 13. The initial guess of the optical field is split in their magnitude and phase information, then, each one passes through an independent branch of the reconstruction architecture. Specifically, each branch is a U-net-inspired architecture [97] depicted in Fig. 14. First, the branch's input passes through five downsampling blocks composed of three convolutional (Conv2D) layers and one max-pooling (MaxPooling) layer, performing feature extraction and decreasing spatial resolution, respectively. Next, one bottleneck block contains two Conv2D layers that learn the feature maps to reconstruct the optical field. Then, to recover the image from these feature maps, five upsampling blocks are applied to increase the spatial resolution using one transpose convolution (TransConv2D) layer and denoise the upsampling by two Conv2D layers. Considering the feature maps at high and low resolution, each input of the upsampling block is concatenated with each output of the downsampling block that contains the same number of filters. Here, both branches have the same number of trainable parameters. Finally, the Conv2D layers have a kernel size of $k = 3 \times 3$ and a number of filters of $f = 8b$ with b the number of the corresponding block; the MaxPooling and the TransConv2D layers have a kernel size of $k = 3 \times 3$ with stride $s = 2 \times 2$, and a number of filters of $f = 8b$.

Figure 14. Double-branch DNN architecture $\mathcal{M}_{\boldsymbol{\theta}}(\cdot)$. The DNN receives the optical field's initialization as input. Subsequently, the downsampling block with Conv2D and MaxPooling layers extracts feature maps, while the bottleneck block prepares these features for reconstruction. Finally, the upsampling block with TransConv2D layers increases the spatial resolution and recovers the optical field.



This double-branch recovery network is included in the proposed E2E scheme for optimizing its trainable parameters at each training epoch. Algorithm 2 summarizes the proposed E2E approach illustrated in Fig. 13, where the kernel filter used in the estimation stage is initialized by using a standard uniform distribution in line 2. In line 5, the propagation model (27) is computed across each diffraction field. Then, the spectral initialization $\mathcal{I}_{\mathbf{G}}(\cdot)$ is approximated in line 6 through Algorithm 1. The double-branch recovery network, as illustrated in Fig. 14, is computed in line 7. In line 8, the proposed loss function (31) is evaluated. In lines 9 and 10, the Adam optimizer $\mathcal{A}_{dam}(\cdot)$ is implemented to minimize this loss function over the kernel filter and the DNN parameters weighted by fixing β_1 and β_2, respectively. Finally, these optimal parameters $\mathbf{G}$ and $\boldsymbol{\theta}$ are returned in line 13.

Algorithm 2 Proposed E2E deep initialization learning

```
1:  Input: Training set $\{\mathbf{x}^{(s)}\}_{s=1}^{\mathcal{S}}$ with $\mathcal{S}$ images.
2:  Initialize: Set the kernel filter $\mathbf{G} \in \mathbb{C}^{u\times v}$ from a uniform distribution.
3:  for epoch $= 1 : \mathcal{E}$ do                                              ▷ $\mathcal{E}$ epochs
4:      for $s = 1 : \mathcal{S}$ do                                              ▷ $\mathcal{S}$ samples
5:          $(\mathbf{y}_{\ell,\psi})_k^{(s)} = |\mathbf{a}_{k,\psi}^{\mathcal{H}}\mathbf{D}_\ell\mathbf{x}^{(s)}|^2 + (\boldsymbol{\omega}_{\ell,\psi})_k$, $k \in \{1,\cdots,n\}$, $\ell \in \{1,\cdots,L\}$.   ▷ CDPs
6:          $\tilde{\mathbf{z}}^{(s)} \leftarrow \mathcal{I}_{\mathbf{G}}\left(\overline{\mathbf{D}}_\ell\mathbf{a}_{k,\psi}, (\mathbf{y}_{\ell,\psi})_k^{(s)}, \mathbf{G}\right)$.   ▷ Algorithm 1
7:          $\mathbf{z}^{(s)} \leftarrow \mathcal{M}_{\boldsymbol{\theta}}\left(\tilde{\mathbf{z}}^{(s)}\right)$.   ▷ Double-branch DNN
8:          $\mathcal{L}_{\mathbf{G},\boldsymbol{\theta}} = \frac{1}{\mathcal{S}}\sum_{s=1}^{\mathcal{S}} \|\mathbf{z}^{(s)} - \mathbf{x}^{(s)}\|_2^2 + \mathcal{R}_\Omega(\mathbf{G})$.   ▷ Loss function
9:          $\mathbf{G} \leftarrow \mathcal{A}_{dam}(\mathbf{G}, \beta_1\nabla_{\mathbf{G}}\mathcal{L}_{\mathbf{G},\boldsymbol{\theta}})$.   ▷ Optimize over $\mathbf{G}$
10:         $\boldsymbol{\theta} \leftarrow \mathcal{A}_{dam}(\boldsymbol{\theta}, \beta_2\nabla_{\boldsymbol{\theta}}\mathcal{L}_{\mathbf{G},\boldsymbol{\theta}})$.   ▷ Optimize over $\boldsymbol{\theta}$
11:     end for
12: end for
13: Return: Optimal kernel filter $\mathbf{G}$ and DNN parameters $\boldsymbol{\theta}$.
```

Once the proposed E2E deep initialization scheme is trained, the learned network infers the magnitude and phase images from a computational complexity $\mathcal{O}(Tn\log n)$, since the convolutional operator can be rapidly performed through the fast Fourier transform [99]. In contrast, reweighted amplitude flow (RAF) [25] and truncated amplitude flow (TAF) [26] algorithms have computational complexity $\mathcal{O}(TLn^2)$.

---

[99] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing (3rd Edition)*. USA: Prentice-Hall, Inc., 2006.

2.3. NUMERICAL RESULTS

This section compared the proposed method to traditional PR approaches by using CDPs acquired under noiseless and noisy scenarios at each diffraction field. Specifically, the noisy cases are fixed by different signal-to-noise (SNR) values, i.e., $(\boldsymbol{\omega}_{\ell,\psi})_k \sim \mathcal{N}(0,\sigma^2)$ was simulated, where σ^2 was set such that certain $\mathrm{SNR} = 10\log_{10}\left(\sum_{\ell=1}^{L}\sum_{k=0}^{n-1}\left((\mathbf{y}_{\ell,\psi})_k\right)^2 / \left(nL\sigma^2\right)\right)$ values were achieved. The proposed DNN approach was trained using Tensorflow on an Intel Xeon SILVER 4210R with 48 GB RAM and an NVIDIA RTX 3090 with 24 GB VRAM. All the experiments are performed by using MPMs $\mathbf{D}_\ell$ with entries $\{j,-j,1,-1\}$ in their main diagonal. Also, the evaluating metrics to measure the performance of the proposed network are the relative error, the peak signal-to-noise ratio (PSNR), and the structural similarity index measure (SSIM).

The resulting estimation accuracy in Algorithm 1 was calculated by using the relative error given by

$$\text{Relative error} := \frac{\mathrm{dist}(\tilde{\mathbf{z}},\mathbf{x})}{\|\mathbf{x}\|_2}, \tag{35}$$

where $\mathrm{dist}(\tilde{\mathbf{z}},\mathbf{x}) = \min_{\beta\in[0,2\pi)}\|\tilde{\mathbf{z}}e^{-j\beta}-\mathbf{x}\|_2$.

It is important to remark that the resulting phase images were improved by using a non-iterative simultaneous phase unwrapping and denoising (SPUD) algorithm [100], which allows the phase jump correction.

2.3.1. Image datasets Three different image datasets were used to train, validate, and test the proposed neural network approach. Figure 15 illustrates some images present in these datasets, which simulate complex scenarios $\mathbf{x} = |\mathbf{x}| \odot e^{j\varphi(\mathbf{x})}$. For all the experiments, the images were spatially adjusted to 128×128 pixels. In

---

[100] Jesus Pineda et al. “SPUD: simultaneous phase unwrapping and denoising algorithm for phase imaging”. In: *Applied optics* 59.13 (2020), pp. D81–D88.

Table 2, each dataset length according to the training, validation, and testing is presented.

- **NYU Depth:** The NYU (New York University) Depth dataset [101] contains $48,238$ RGB and depth map images. Specifically, the RGB images are converted to grayscale images for describing the magnitude information $|\mathbf{x}|$, and the depth maps are scaled in the range $[-\pi, \pi]$ for emulating the phase information $\varphi(\mathbf{x})$.
- **MNIST:** The MNIST (Modified National Institute of Standards and Technology) [102] dataset contains $70,000$ grayscale images used as phase information.
- **Fashion MNIST:** The Fashion MNIST dataset [103] contains $70,000$ grayscale images employed as phase information.

Table 2. Summary of the number of images selected for training, validation, and testing according to each dataset.

Dataset	Training ($\mathcal{S}$)	Validation	Testing	Total
NYU Depth	$42,826$	$4,758$	50	$47,634$
MNIST	$54,000$	$6,000$	50	$60,050$
Fashion MNIST	$54,000$	$6,000$	50	$60,050$

---

[101] Pushmeet Kohli Nathan Silberman Derek Hoiem and Rob Fergus. “Indoor Segmentation and Support Inference from RGBD Images”. In: *ECCV*. 2012.

[102] Li Deng. “The mnist database of handwritten digit images for machine learning research [best of the web]”. In: *IEEE Signal Processing Magazine* 29.6 (2012), pp. 141–142.

[103] Han Xiao, Kashif Rasul, and Roland Vollgraf. “Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms”. In: *arXiv preprint arXiv:1708.07747* (2017).

Figure 15. NYU Depth, MNIST, and Fashion MNIST datasets were used for training, validating, and testing the proposed deep initialization network.

2.3.2. Simulation setting The simulation parameters used for the propagation model, the filtered initialization stage, and the training neural network are presented in this section. Table 3 presents the optical parameters fixed to calculate the matrices $\mathbf{T}$ (29) and $\mathbf{Q}$ (30) of the sensing matrix $\mathbf{A}_{\psi}$ (28) corresponding to the near-field $\psi = 1$ and middle-field $\psi = 2$, respectively. These parameters allow simulating the CDPs acquisition (27) along each diffraction field. For CDPs at the far-field $\psi = 3$, the forward model can be calculated without requiring wavelength and propagation distance information, according to (28), as this diffraction model relies only on the discrete Fourier transform of the coded optical field. In the initialization stage, the filter $\mathbf{G} \in \mathbb{C}^{u\times v}$ was trained by fixing the kernel size $u = v = 5$ and the number of iterations $T = 10$. Finally, the E2E scheme has up to 2.7 million parameters optimized by using a batch size of 2 images and a learning rate of 1×10^{-3} with the Adam algorithm, setting $\mathcal{E} = 250$ epochs for the NYU Depth dataset, and $\mathcal{E} = 100$ epochs for the MNIST and Fashion MNIST.

Table 3. Optical parameters employed for simulating the propagation model (27) at each diffraction field ψ.

Parameter	Diffraction field		
	Near-field $\psi = 1$	Middle-field $\psi = 2$	Far-field $\psi = 3$
Wavelength (λ) $[nm]$	635	635	-
Propagation distance (z) $[cm]$	2.5	7.0	-

2.3.3. Initialization results Figure 16 compares the relative error obtained by the learned spectral initialization against traditional initializations such as FSI [96] based on a Gaussian filter $\mathbf{G}$, weighted maximal correlation initialization (WMCI) [25], and orthogonality-promoting initialization (OPI) [26]. This experiment demonstrates the relative error behavior over the iterations T across the testing dataset, considering each diffraction field with $\mathrm{SNR} = 10$ $[\mathrm{dB}]$.

Also, Fig. 16 shows that the proposed approach achieves a lower relative error using $\mathrm{SNR} = 10$ $[\mathrm{dB}]$ for each dataset across all diffraction fields compared to traditional alternatives. It is important to note that the FSI and the learned initialization only require $L = 1$ snapshots to properly estimate the optical field, while WMCI and OPI algorithms require at least $L = 4$ snapshots.

Figure 16. Comparison of the proposed learned spectral initialization to different initialization approaches in terms of the relative error for a different number of iterations with $\mathrm{SNR} = 10$ [dB]. Rows: Diffraction fields. Columns: Datasets.

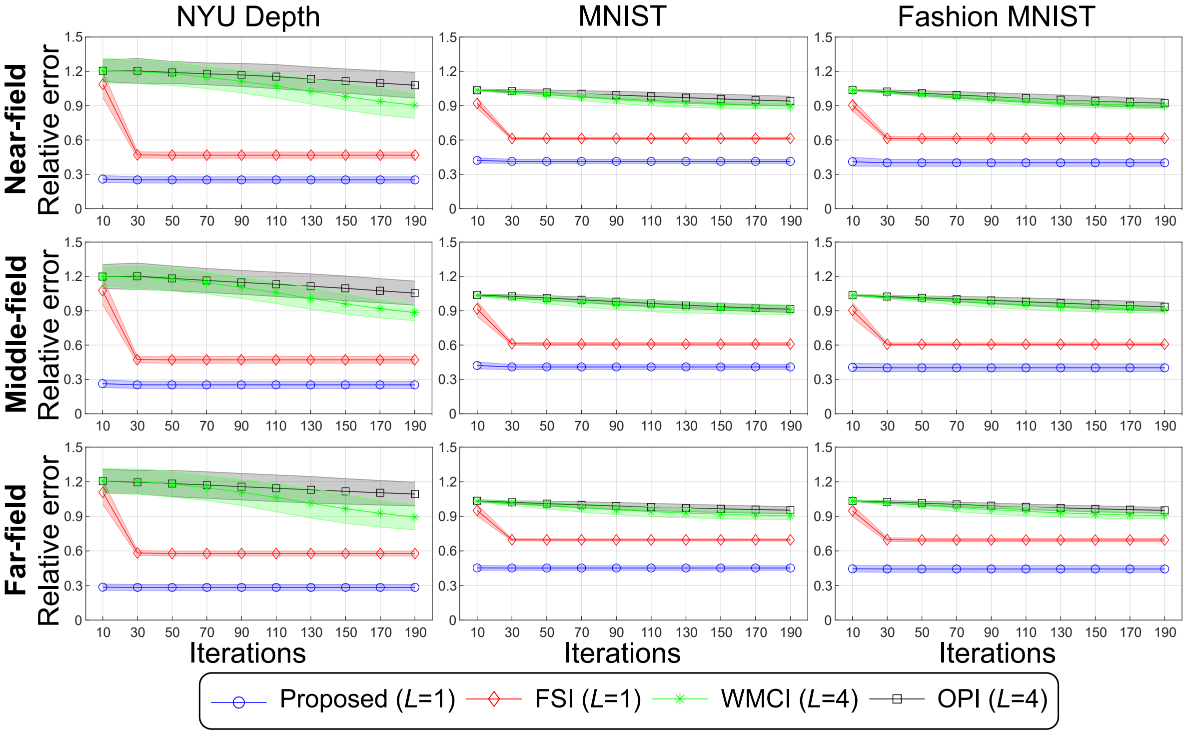
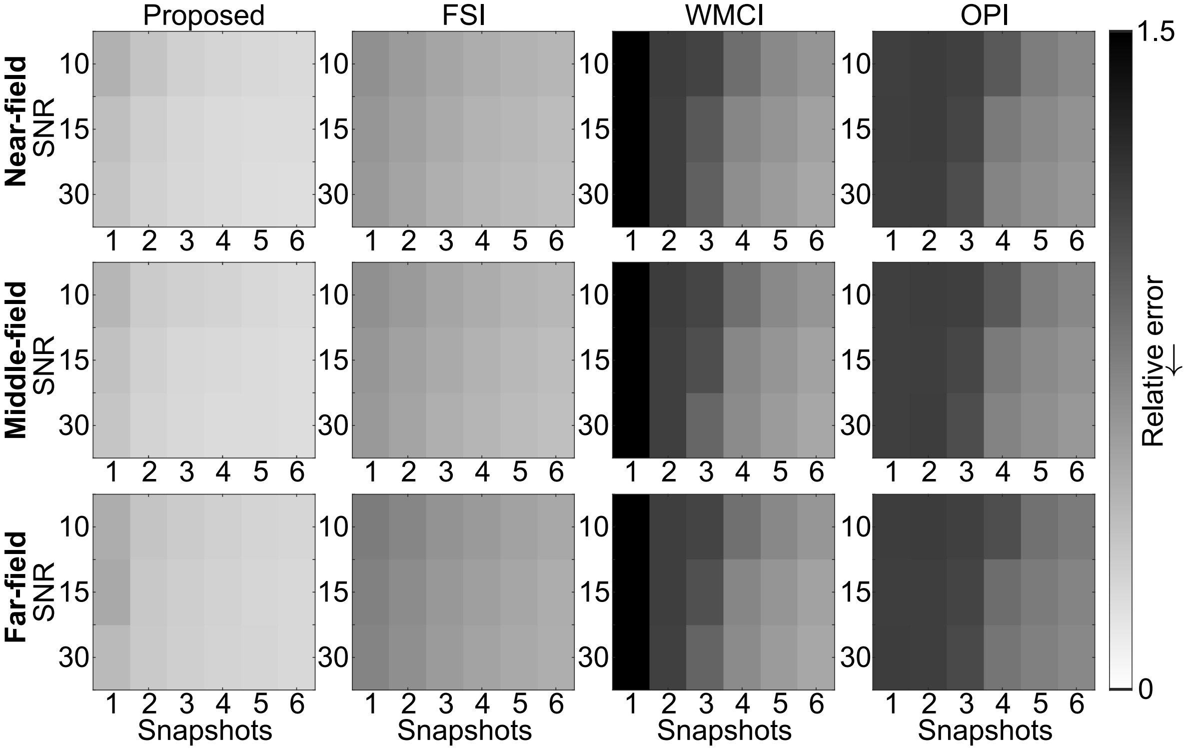


Figure 17 depicts the relative error performance comparing the initialization strategies under different noise levels $\mathrm{SNR} \in \{10, 15, 30\}$ [dB] by varying the number of snapshots L. In this figure, lighter colors represent lower relative error. The proposed learning-based initialization outperforms the traditional initialization algorithms at each noise level. Notice that the proposed method achieves better optical field approximations with fewer snapshots compared to traditional initialization methods. While the proposed method can accurately estimate the optical field with only $L = 1$ snapshots, traditional initialization methods require at least $L = 4$ to achieve comparable results.

Figure 17. Summary of the relative error using different initialization strategies, noise levels, and number of snapshots. Rows: Diffraction fields. Columns: Initializations. A lighter color indicates a closer optical field estimation.



Finally, Figs. 18, 19 and 20 illustrate some visual results of different initializations, including the proposed learned initialization and traditional algorithms across each dataset and each diffraction field. The PSNR and SSIM metrics were calculated for the magnitude and phase information to compare this initialization under a noiseless scenario. These figures show that the proposed learnable initialization obtains comparable results to the state-of-the-art initializations using $L = 1$ snapshot and $T = 10$ iterations. In contrast, traditional initializations such as WMCI and OPI require at least $L = 4$ snapshots and $T = 200$ iterations. Notice that the FSI improves image quality as the number of iterations increases from 10 to 200. This algorithm uses a Gaussian filter, whereas the proposed initialization learns a kernel filter during network training. This approach results in better estimations from fewer iterations.

Figure 18. Visual results of estimated magnitude and phase images from noiseless CDPs acquired at the **near-field** using the proposed learned spectral initialization and traditional initialization algorithms. Rows: Initializations. Columns: Datasets.

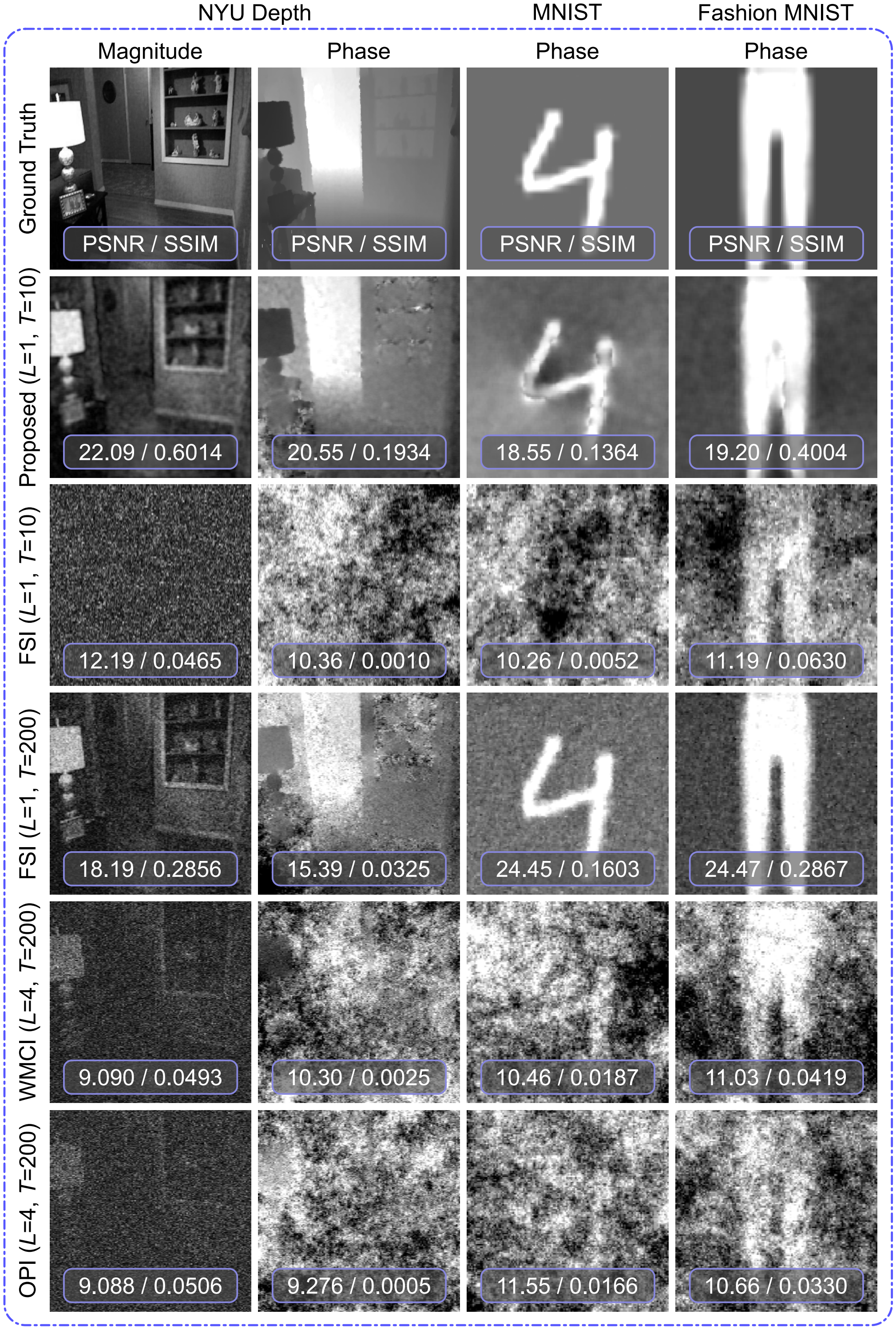


Figure 19. Visual results of estimated magnitude and phase images from noiseless CDPs acquired at the **middle-field** using the proposed learned spectral initialization and traditional initialization algorithms. Rows: Initializations. Columns: Datasets.

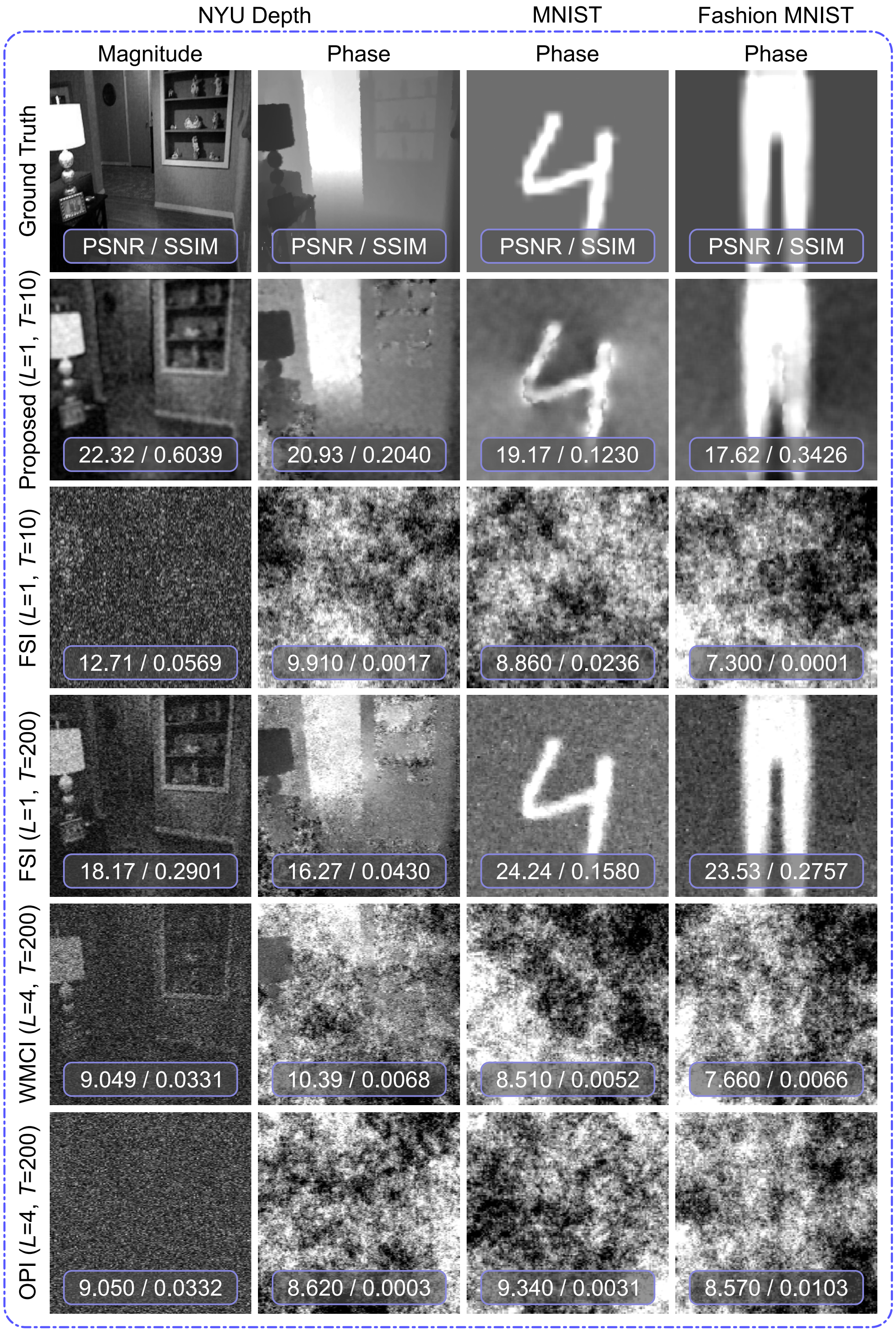
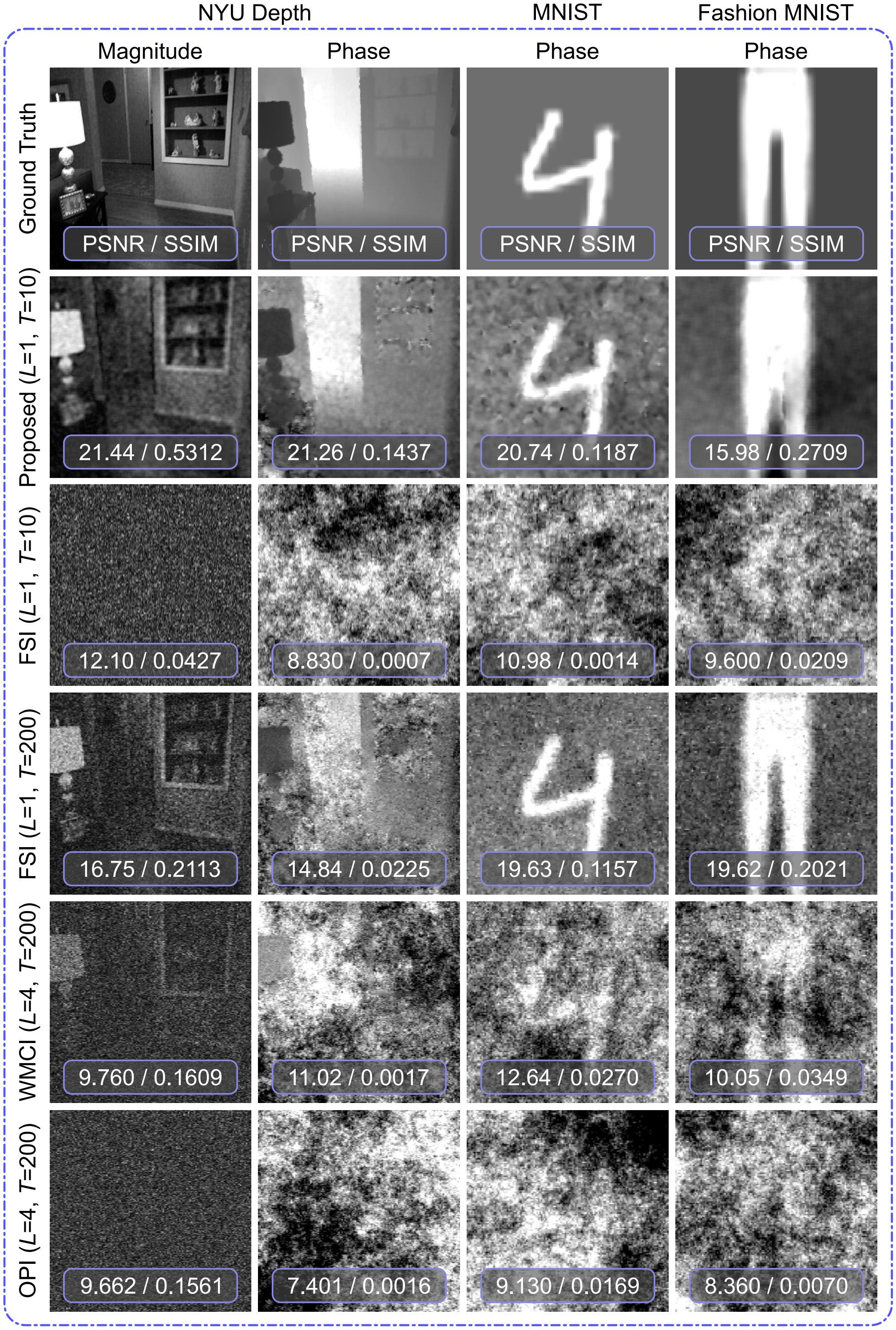


Figure 20. Visual results of estimated magnitude and phase images from noiseless CDPs acquired at the **far-field** using the proposed learned spectral initialization and traditional initialization algorithms. Rows: Initializations. Columns: Datasets.



2.3.4. Reconstruction results In Figs. 21, 22 and 23 the reconstruction performance is illustrated. The double-branch U-net architecture is compared to traditional PR algorithms such as RAF [25] and TAF [26] using the three datasets across each diffraction field. The PSNR and SSIM metrics were calculated over the reconstructed magnitude and phase images to quantitatively evaluate the reconstruction quality under a noiseless scenario.

The proposed method only requires $L = 1$ snapshot and $T = 10$ iterations to obtain suitable reconstructions. This method outperforms the traditional recovery algorithms, as these conventional methods cannot recover the optical field from a single snapshot $L = 1$. Specifically, RAF and TAF algorithms require at least $L = 4$ snapshots and $T = 200$ iterations.

It is important to mention that the double-branch recovery network particularly improves the phase reconstruction results for the MNIST and the Fashion MNIST. Although the magnitude information is also enhanced by the recovery network for the NYU dataset, the reconstructed phase image performance slightly decays. In fact, this behavior is attributed to the challenging nature of the NYU dataset, which contains complex images with both magnitude and phase information.

Figure 21. Visual results of reconstructed magnitude and phase images from noiseless CDPs acquired at the **near-field** using the proposed E2E approach and traditional PR algorithms. Rows: Reconstructions. Columns: Datasets.

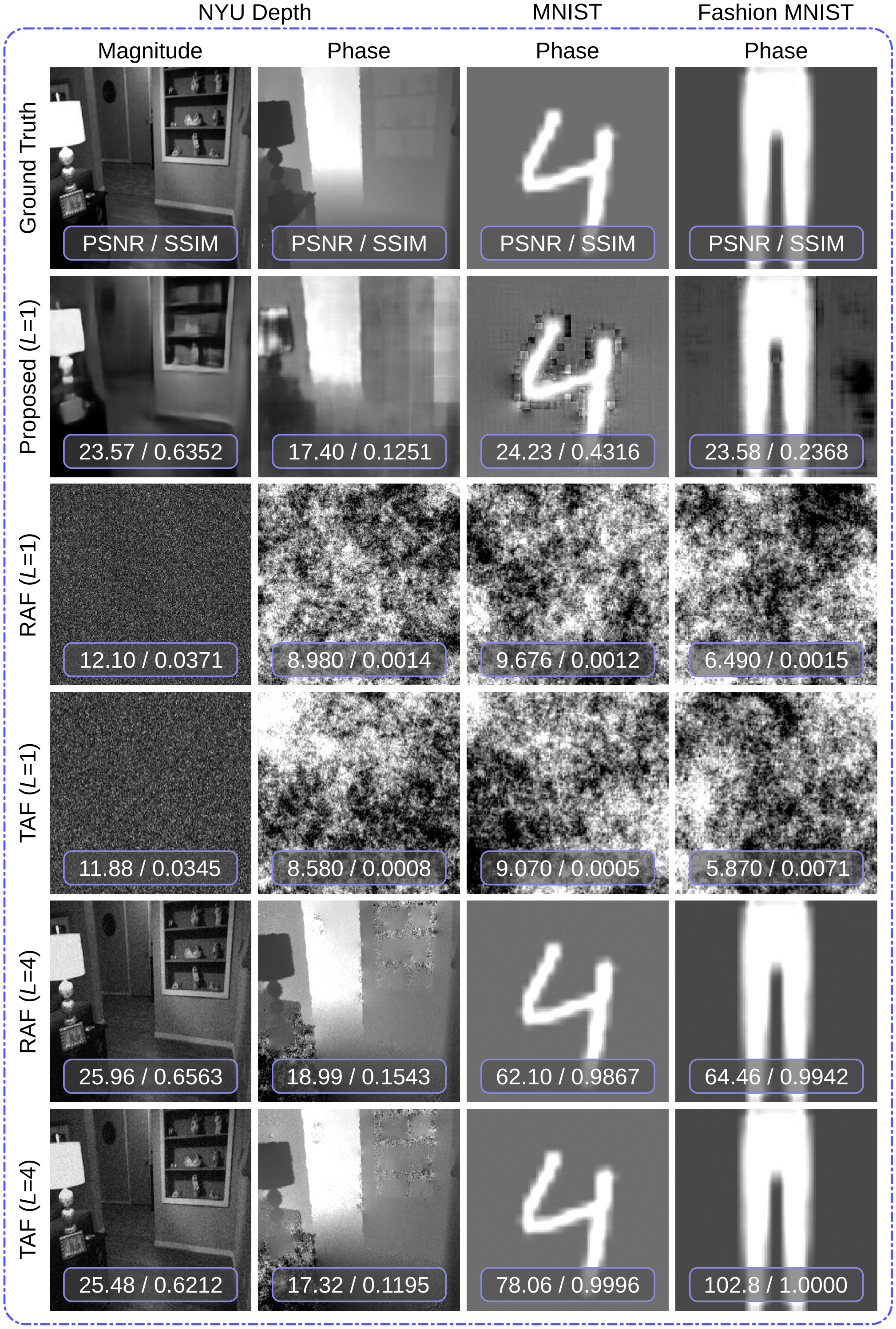


Figure 22. Visual results of reconstructed magnitude and phase images from noiseless CDPs acquired at the **middle-field** using the proposed E2E approach and traditional PR algorithms. Rows: Reconstructions. Columns: Datasets.

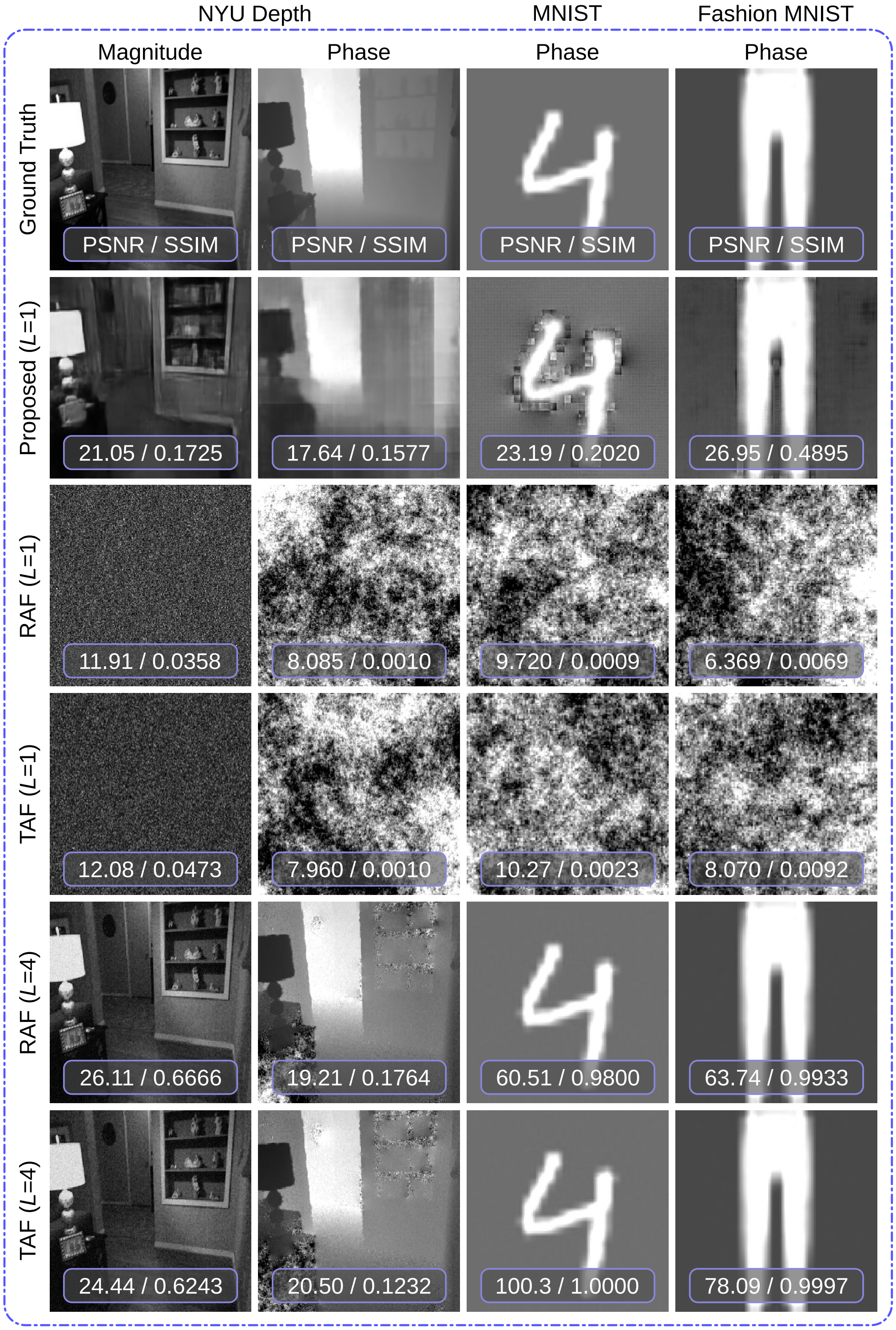
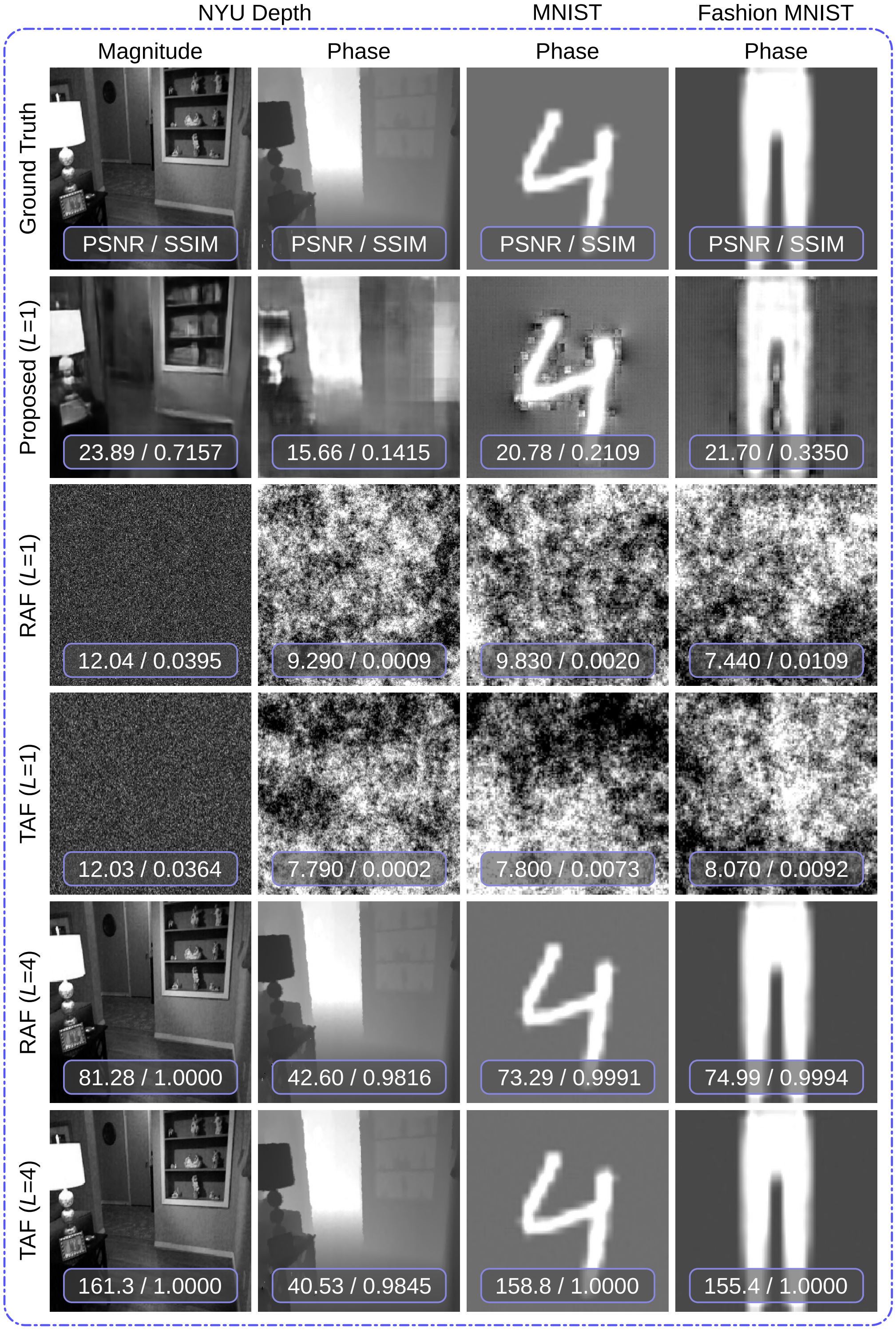


Figure 23. Visual results of reconstructed magnitude and phase images from noiseless CDPs acquired at the **far-field** using the proposed E2E approach and traditional PR algorithms. Rows: Reconstructions. Columns: Datasets.



2.3.5. Time Complexity Table 4 summarizes the run-time for different PR approaches over randomly selected testing images from the MNIST dataset. This table shows that the proposed E2E approach obtains image reconstruction up to 32 times faster than traditional PR schemes. Notice that the run-time for the proposed method corresponds only to the inference process, i.e., once the neural network is trained. All experiments were conducted on a desktop computer with a CPU Intel Xeon E5-2697 v3 @ 2.6 GHz and 192 GB RAM. The Tensorflow environment was also run over the CPU to achieve a fair comparison with traditional algorithms.

Table 4. Comparison of image reconstruction for different recovery approaches in terms of run-time ↓ [seconds]. The best result is highlighted in green and the second best is highlighted in yellow.

Diffraction field	PR method		
	Proposed ($L = 1$)	RAF ($L = 4$)	TAF ($L = 4$)
Near-field	0.2814	9.1429	9.0703
Middle-field	0.2065	6.0699	5.9895
Far-field	0.2236	4.2415	4.1753

2.4. CHAPTER CONCLUSION

An E2E deep network approach for learning filtered spectral initialization in PR was proposed at each diffraction field. The proposed method consists of an optical stage based on coherent light that collects the CDPs; an initialization stage that estimates the optical field; and a double-branch DNN that recovers the optical field. Numerical results over noiseless and noisy cases support that the learned kernel reduces the relative error of the filtered initialization with fewer iterations and snapshots than traditional methods. Despite relying on a single snapshot, the proposed double-branch reconstruction architecture achieves results comparable to conventional PR algorithms, which typically require at least four snapshots.

"In the fields of observation, chance favors only the prepared mind."
- ***Louis Pasteur***

---

3. DEEP UNROLLED APPROACH FOR PHASE RETRIEVAL

This chapter presents an end-to-end (E2E) deep unrolled phase recovery (PR) network. In particular, this E2E approach allows the obtainment of an interpretable deep neural network (DNN) by unrolling the spectral initialization and the iterative reconstruction algorithms based on a non-convex formulation and a multilevel phase mask (MPM) design. The proposed method consists of the integration of three main components into a single DNN: first, a fully differential layer that models the diffractive optical architecture based on coherent light by using the angular spectrum method (near-field propagation), allowing the acquisition of coded diffraction patterns (CDPs); second, an unrolling-based initialization stage that estimates an initial guess from the acquired CDPs; third, an unrolling-based recovery stage that further refines the obtained optical field approximation from the gradient descent step and convolutional layers. Specifically, the recovery approach employs a smoothness function and a deep regularization for solving the real and imaginary parts of the optical field. Simulation results show that the proposed E2E approach can reconstruct a high-fidelity phase image from a single snapshot in noiseless and noisy scenarios. The proposed initialization network outperforms traditional initialization algorithms by varying the number of snapshots and iterations. Furthermore, the proposed PR scheme achieves better recovery performance than the benchmark PR algorithms, relying on learning-based and non-learning-based approaches. Finally, an optical setup for CPD acquisition is implemented from a laser source that illuminates a liquid-crystal-on-silicon (LCOS), which sets the phase information over the underlying scene and the MPM. Subsequently, the propagated optical field is recorded by a CMOS camera. The experimental results validate the E2E deep recovery method based on a non-convex formulation from the CDPs collected by the implemented diffractive optical setup. This chapter addresses all the specific objectives.

The main contributions of this work are summarized as follows:

- An E2E deep unrolled scheme is proposed for addressing the PR problem using the near-field model through a smoothness function and a non-convex formulation.
- An MPM design for CDP acquisition at the near-field is introduced. The MPM distribution is learned from the proposed E2E deep unrolled network based on a non-convex formulation.
- The proposed deep unrolled network obtains optical field reconstruction from a single snapshot and a small number of iterations.

3.1. PHASE RETRIEVAL FORMULATION

Let $\mathbf{x} \in \mathbb{C}^n$ be a complex object that contains magnitude and phase components of the emanating electromagnetic field. In diffractive optical systems, an MPM $\mathbf{D}_\ell \in \mathbb{C}^{n\times n}$ is included just after the object. Mathematically, $\mathbf{D}_\ell \in \mathbb{C}^{n\times n}$ corresponds to the diagonal matrix that contains the ℓ-th modulation pattern. Then, the CDPs $(\mathbf{y}_\ell)_k$ for multiple snapshots at the same scene at the near-field $\psi = 1$ can be defined as

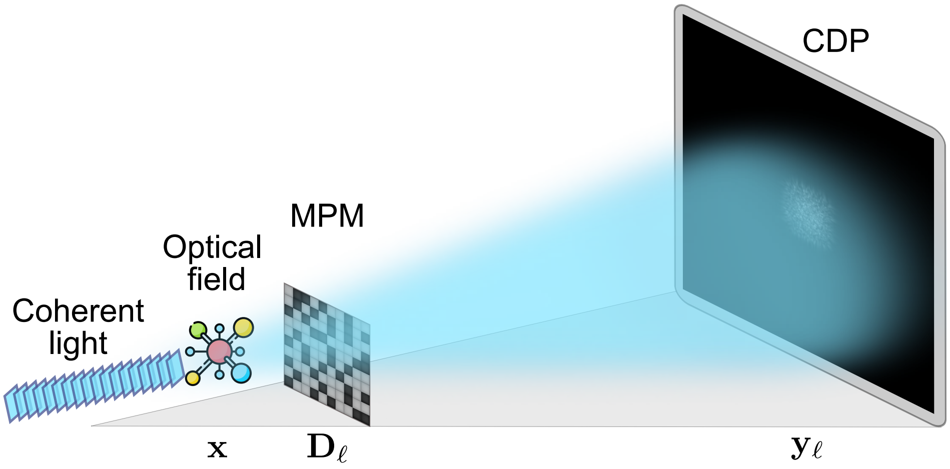
$$(\mathbf{y}_\ell)_k = |\mathbf{a}_k^{\mathcal{H}}\mathbf{D}_\ell\mathbf{x}|^2 + (\boldsymbol{\omega}_\ell)_k, \tag{36}$$

for $k \in \{0, \cdots, n-1\}$ and $\ell \in \{1, \cdots, L\}$ with L as the total number of snapshots. Specifically, $(\cdot)^{\mathcal{H}}$ is the conjugate transpose operation, $|\cdot|$ denotes the pointwise magnitude, $(\boldsymbol{\omega}_\ell)_k$ is the additive Gaussian noise, and $\mathbf{a}_k^{\mathcal{H}}$ represents the rows of the matrix $\mathbf{A} = \mathbf{F}\mathbf{T}\mathbf{F}^{\mathcal{H}}$, where $\mathbf{F} \in \mathbb{C}^{n\times n}$ is the discrete Fourier transform (DFT) matrix and $\mathbf{T} \in \mathbb{C}^{n\times n}$ is the transfer function in the spatial frequency [6]. The transfer function can be modeled as

$$(\mathbf{T})_{p,q} = e^{-jk_0 z\sqrt{1-\frac{(p\delta_{kx})^2}{k_0^2}-\frac{(q\delta_{ky})^2}{k_0^2}}}, \text{ with } j=\sqrt{-1}, \tag{37}$$

where δ_{kx} and δ_{ky} are the sampling periods in the frequency coordinates, z is the propagation distance, and $k_0 = 2\pi/\lambda$ is the wavenumber with λ as the wavelength [6]. Figure 24 illustrates a diffractive optical system at the near-field, where the CDPs acquisition procedure has been tackled by assuring redundant information from the inclusion of an optical modulator known as an MPM [59]. This coding mask is placed immediately after the scene of interest to modulate its optical field before being recorded by the sensor, which finally acquires the generated CDP [59].

Figure 24. Diffractive coded optical system for acquiring CDPs at the near-field. Initially, A coherent light source illuminates the optical field. Subsequently, the incoming light is modulated by an MPM, which results in CDPs. Finally, these CDPs are collected by a sensor.



For mathematical analysis, the entries of the MPMs $\mathbf{D}_\ell$ are treated as a random variable [59,14], to provide uniqueness guarantees (up to a global phase uncertainty)

[59,14,23,12]. From (36), the set of CDPs can be modeled via the MPM [104,12]. Specifically, any accessible solutions are given by $(\hat{\mathbf{x}})_t = e^{j\beta}(\mathbf{x})_t$, for some $\beta \in [0, 2\pi)$. In consequence, the Euclidean distance $\mathrm{dist}(\cdot)$ between two complex vectors $\mathbf{w}_1$, $\mathbf{w}_2$ $\in \mathbb{C}^n$, on this PR problem is indifferent to a global phase constant expressed as

$$\mathrm{dist}(\mathbf{w}_1, \mathbf{w}_2) = \min_{\beta \in [0,2\pi)} \|\mathbf{w}_1 e^{-j\beta} - \mathbf{w}_2\|_2. \tag{38}$$

3.1.1. Non-convex formulation Note that recovering the phase information from CDPs $(\mathbf{y}_\ell)_k$ poses an ill-posed problem, wherein the phase component of the complex wave is lost. Conventionally, the least-squares criterion has been adopted to obtain the optical field reconstruction $\mathbf{z} \in \mathbb{C}^n$, by minimizing the intensity-based empirical loss

$$\operatorname*{argmin}_{\mathbf{z} \in \mathbb{C}^n} \ h(\mathbf{z}) = \frac{1}{2nL} \sum_{\ell=1}^{L} \sum_{k=0}^{n-1} \left(|\mathbf{a}_k^{\mathcal{H}} \mathbf{D}_\ell \mathbf{z}|^2 - (\mathbf{y}_\ell)_k \right)^2. \tag{39}$$

Since the absolute value in (39) is highly non-convex, PR algorithms integrate statistical initializations that start near the convexity zone and then refine the estimation based on gradient-based methods using the Wirtinger derivative [105].

3.1.2. Filtered spectral initialization Given the crucial role of the initial guess in solving the PR problem, this section introduces the filtered spectral initialization (FSI) [96] for the near-field model. Essentially, FSI computes an estimation of the

---

[104] Jorge Bacca, Samuel Pinilla, and Henry Arguello. “Coded Aperture Design for Super-Resolution Phase Retrieval”. In: *2019 27th European Signal Processing Conference (EUSIPCO)*. IEEE. 2019, pp. 1–5.

[105] Samuel Pinilla et al. “A Smoothing Stochastic Phase Retrieval Algorithm for Solving Random Quadratic Systems”. In: *2018 IEEE Statistical Signal Processing Workshop (SSP)*. IEEE. 2018, pp. 278–282.

scene $\tilde{\mathbf{z}}$ by deriving a low-pass version of the leading eigenvector of the matrix

$$\mathbf{\Gamma} := \frac{1}{\mathrm{card}(\Xi)} \sum_{(\ell,k)\in\Xi} \frac{\overline{\mathbf{D}}_\ell \mathbf{a}_k \mathbf{a}_k^{\mathcal{H}} \mathbf{D}_\ell}{\|\mathbf{D}_\ell \mathbf{a}_k\|_2^2}. \tag{40}$$

scaled by the quantity $\xi := \sqrt{\frac{\sum_{\ell=1}^{L}\sum_{k=0}^{n-1}(\mathbf{y}_\ell)_k}{nL}}$, i.e, $\tilde{\mathbf{z}} = \xi\hat{\mathbf{z}}$, where the set Ξ contains the indices associated to the $\lfloor \frac{nL}{2} \rfloor$ largest values of $\{(\mathbf{y}_\ell)_k / \|\mathbf{D}_\ell \mathbf{a}_k\|_2\}$. The notation $\mathrm{card}(\Xi)$ represents the cardinality of the set Ξ, and $\overline{(\cdot)}$ is the complex conjugate operation. The low-pass filtering constraint in FSI arises from the fact that natural images are mostly composed of low-frequencies [98]. Since FSI requires the estimation of the leading eigenvector of $\mathbf{\Gamma}$, employing a power iteration strategy can efficiently solve this task [26]. This strategy recursively performs matrix-vector multiplication between the matrix $\mathbf{\Gamma}$ and the current estimation of the scene, preserving the low-frequency information by applying a low-pass filter.

Finally, the FSI for the near-field allows solving the following optimization problem

$$\tilde{\mathbf{z}} = \arg\max_{\|\mathbf{z}\|_2=1} \mathbf{z}^{\mathcal{H}} \mathbf{\Gamma} \mathbf{z}. \tag{41}$$

3.2. PROPOSED DEEP UNROLLED APPROACH FOR PHASE RETRIEVAL

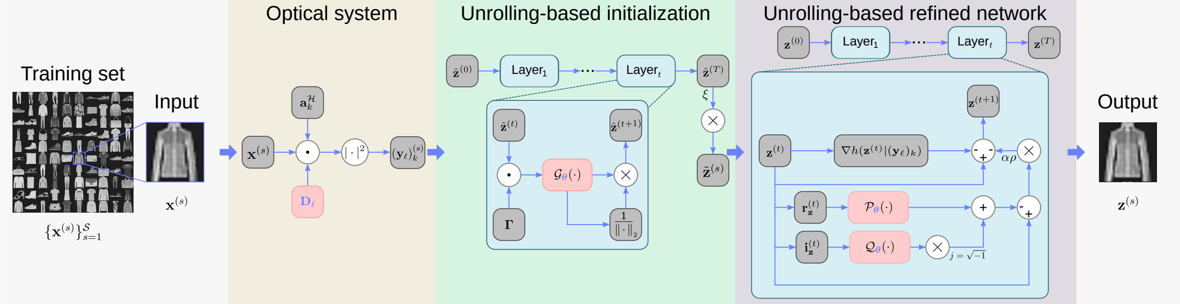
Although different methodologies have been proposed for the MPM design [12,104], and also recovery strategies for the PR problem [106,28,25]. To the best of our knowledge, there has not yet been a method that addresses the PR problem by leveraging the near-field model and adopting the non-convex formulation to simultaneously couple the phase mask design with the recovery step. Therefore, our proposed E2E

---

[106] Samuel Pinilla, Jorge Bacca, and Henry Arguello. “Phase retrieval algorithm via nonconvex minimization using a smoothing function”. In: *IEEE Transactions on Signal Processing* 66.17 (2018), pp. 4574–4584.

scheme, as illustrated in Fig. 25, aims at designing the MPMs and employing an inspired deep unrolled recovery network based on a non-convex formulation to obtain the desired optical field acquired at the near-field diffraction.

Figure 25. Proposed E2E unrolled PR approach. The proposed method simulates the propagation model by using the angular spectrum method. Then, an unrolling-based initialization stage estimates an initial guess of the optical field. Finally, the proposed unrolled recovery network produces the desired phase image.



3.2.1. Inspired deep unrolled recovery network architecture

The proposed deep recovery network is inspired by the non-convex formulation described in Section 3.1. Specifically, we propose unrolling the initialization steps as layers into a DNN model, denoted as $\mathcal{I}_{\boldsymbol{\theta}}(\cdot)$. Subsequently, this network is concatenated with an unrolled recovery network inspired by the gradient-descent method from the problem established in (39), denoted as $\mathcal{M}_{\boldsymbol{\theta}}(\cdot)$, which results in optical field recovery. Thus, the proposed methodology corresponds to a fully inspired E2E unrolled network.

Unrolling-based initialization The initial approximation of the optical field is computed using an unrolling approach over the FSI, where the iterative steps are summarized in Algorithm 3. Each iteration is modeled as a layer, where a double-branch neural network replaces the spatial filter step, denoted as $\mathcal{G}_{\boldsymbol{\theta}}(\cdot)$. In this path, $\mathcal{I}_{\boldsymbol{\theta}}(\cdot)$ denotes the steps in Algorithm 3, referred to as the unrolled initialization, where $\boldsymbol{\theta}$ indicates the learned parameters. Additional details about the network are described in the simulation section.

Algorithm 3 Unrolling-based spectral initialization

```
Input: Acquired data $\{\overline{\mathbf{D}}_\ell \mathbf{a}_k; (\mathbf{y}_\ell)_k\}$ and maximum number of iterations $T$.
Initialize: $\hat{\mathbf{z}} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$.
Set $\Xi$ as the set of indices corresponding to the $\mathrm{card}(\Xi)$ largest values of $\{(\mathbf{y}_\ell)_k / \|\mathbf{D}_\ell \mathbf{a}_k\|_2\}$.
$$\boldsymbol{\Gamma} := \frac{1}{\mathrm{card}(\Xi)} \sum_{(\ell,k)\in\Xi} \frac{\overline{\mathbf{D}}_\ell \mathbf{a}_k \mathbf{a}_k^{\mathcal{H}} \mathbf{D}_\ell}{\|\mathbf{D}_\ell \mathbf{a}_k\|_2^2}.$$
for $t = 0 : T-1$ do
    $\grave{\mathbf{z}}^{(t+1)} = \mathcal{G}_\theta\left(\boldsymbol{\Gamma}\hat{\mathbf{z}}^{(t)}\right)$.        ▷ Double-branch neural network
    $\hat{\mathbf{z}}^{(t+1)} = \frac{\grave{\mathbf{z}}^{(t+1)}}{\left\|\grave{\mathbf{z}}^{(t+1)}\right\|_2}$.        ▷ Normalizing step
end for
Compute $\tilde{\mathbf{z}} = \sqrt{\frac{\sum_{\ell=1}^{L}\sum_{k=0}^{n-1}(\mathbf{y}_\ell)_k}{nL}}\hat{\mathbf{z}}^{(T)}$.        ▷ Scaling step
Return: Initial guess $\tilde{\mathbf{z}}$.
```

Unrolling-based refined network The proposed unrolled decoder is inspired by the optimization formulation presented in [107], which introduces a smoothness function. Considering $\mathbf{x} = \mathbf{r} + j\mathbf{i}$, where $\mathbf{r} \in \mathbb{R}^n$, $\mathbf{i} \in \mathbb{R}^n$ are the real and imaginary part of $\mathbf{x}$, respectively, then, the refined network aims at solving the following optimization problem from $\mathbf{r}$ and $\mathbf{i}$

$$\min_{\mathbf{r},\mathbf{i},\mathbf{z}\in\mathbb{C}^n} \quad \frac{1}{2nL}\sum_{\ell=1}^{L}\sum_{k=0}^{n-1}\left(\vartheta_\mu(|\mathbf{a}_k^{\mathcal{H}}\mathbf{D}_\ell\mathbf{z}|) - \sqrt{(\mathbf{y}_\ell)_k}\right)^2 + \lambda_{\mathbf{r}}\mathcal{R}_{\mathbf{r}}(\mathbf{r}) + \lambda_{\mathbf{i}}\mathcal{R}_{\mathbf{i}}(\mathbf{i}) \qquad \text{subject to} \quad \mathbf{z} - \mathbf{r} - j\mathbf{i} = \mathbf{0}, \tag{42}$$

where the smoothness function $\vartheta_\mu : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is defined as $\vartheta_\mu(w) = \sqrt{w^2 + \mu^2}$ with $\mu \in \mathbb{R}_{++}$ and $\mathcal{R}_{\mathbf{r}}(\cdot)$ and $\mathcal{R}_{\mathbf{i}}(\cdot)$ denoted the regularization terms over the real and

---

[107] Jorge Bacca et al. “Super-Resolution Phase Retrieval Algorithm using a Smoothing Function”. In: *Mathematics in Imaging*. Optical Society of America. 2018, MW2D–3.

imaginary parts with its corresponding regularization terms $\lambda_{\mathbf{r}}$ and $\lambda_{\mathbf{i}}$, respectively. This work proposes to replace the effect of traditional regularization methods with a prior network, as explained below. In particular, the minimization respect to $\mathbf{z}, \mathbf{r}$ and $\mathbf{i}$ can be found by iteratively solving the following optimization problems

$$\mathbf{z}^{(t+1)} := \underset{\mathbf{z}\in\mathbb{C}^n}{\operatorname{argmin}} \frac{1}{2nL}\sum_{\ell=1}^{L}\sum_{k=0}^{n-1}\left(\vartheta_\mu(|\mathbf{a}_k^{\mathcal{H}}\mathbf{D}_\ell\mathbf{z}|) - \sqrt{(\mathbf{y}_\ell)_k}\right)^2 + \frac{\rho}{2}\|\mathbf{z} - \mathbf{r}^{(t)} - j\mathbf{i}^{(t)}\|_2^2, \quad (43)$$

$$\mathbf{r}^{(t+1)} := \underset{\mathbf{r}\in\mathbb{R}^n}{\operatorname{argmin}} \lambda_{\mathbf{r}}\mathcal{R}_{\mathbf{r}}\left(\mathbf{r}\right) + \frac{\rho}{2}\|\mathbf{r}_{\mathbf{z}}^{(t+1)} - \mathbf{r}\|_2^2, \quad (44)$$

$$\mathbf{i}^{(t+1)} := \underset{\mathbf{i}\in\mathbb{R}^n}{\operatorname{argmin}} \lambda_{\mathbf{i}}\mathcal{R}_{\mathbf{i}}\left(\mathbf{i}\right) + \frac{\rho}{2}\|\mathbf{i}_{\mathbf{z}}^{(t+1)} - \mathbf{i}\|_2^2, \quad (45)$$

where $\mathbf{r}_{\mathbf{z}}^{(t+1)}$ and $\mathbf{i}_{\mathbf{z}}^{(t+1)}$ are the real and imaginary part of $\mathbf{z}^{(t+1)}$ at the global iteration $t+1$. We solve (43) by using a gradient descent method based on the Wirtinger derivative as

$$\mathbf{z}^{(t+1)} := \mathbf{z}^{(t)} - \underbrace{\frac{\alpha}{Ln}\sum_{\ell=1}^{L}\sum_{k=0}^{n-1}\left(\mathbf{a}_k^{\mathcal{H}}\mathbf{D}_\ell\mathbf{z}^{(t)} - \frac{\sqrt{(\mathbf{y}_\ell)_k}\mathbf{a}_k^{\mathcal{H}}\mathbf{D}_\ell\mathbf{z}^{(t)}}{\vartheta_\mu(|\mathbf{a}_k^{\mathcal{H}}\mathbf{D}_\ell\mathbf{z}^{(t)}|)}\right)\overline{\mathbf{D}}_\ell\mathbf{a}_k}_{\nabla h(\mathbf{z}^{(t)}|(\mathbf{y}_\ell)_k)} - \alpha\rho\left(\mathbf{z}^{(t)} - \mathbf{r}^{(t)} - j\mathbf{i}^{(t)}\right), \quad (46)$$

where $\alpha > 0$ is the gradient descent step size and $\rho > 0$ is a regularization parameter. The minimizing of $\mathbf{r}^{(t)}$ and $\mathbf{i}^{(t)}$ can be seen as a proximal operator that can be addressed by applying a DNN at the real and imaginary part of $\mathbf{z}^{(t+1)}$ as

$$\mathbf{r}^{(t+1)} = \mathcal{P}_{\boldsymbol{\theta}}\left(\mathbf{r}_{\mathbf{z}}^{(t+1)}\right), \text{ and} \tag{47}$$

$$\mathbf{i}^{(t+1)} = \mathcal{Q}_{\boldsymbol{\theta}}\left(\mathbf{i}_{\mathbf{z}}^{(t+1)}\right), \tag{48}$$

where $\mathcal{P}_{\boldsymbol{\theta}}(\cdot)$ and $\mathcal{Q}_{\boldsymbol{\theta}}(\cdot)$ represent a DNN with trainable parameters $\boldsymbol{\theta}$ for the real and imaginary part, respectively. Finally, combing (46), (47) and (48) in a hold step, the image can be estimated as

$$\mathbf{z}^{(t+1)} := \mathbf{z}^{(t)} - \nabla h(\mathbf{z}^{(t)}|(\mathbf{y}_\ell)_k) - \alpha\rho\left(\mathbf{z}^{(t)} - \mathcal{P}_{\boldsymbol{\theta}}\left(\mathbf{r}_{\mathbf{z}}^{(t)}\right) - j\mathcal{Q}_{\boldsymbol{\theta}}\left(\mathbf{i}_{\mathbf{z}}^{(t)}\right)\right). \tag{49}$$

In order to find the optimal $\boldsymbol{\theta}$ parameter based on the gradient formulation (49), we propose to unroll the recursion iteration via DNN from $t \in \{1, \cdots, T\}$ step iterations, as described in Algorithm 4, which results in T layers of the proposed method denoted as $\mathcal{M}_{\boldsymbol{\theta}}(\cdot)$. The parameters α and ρ are also trainable in the unrolled recovery DNN scheme. Algorithm 4 summarizes the proposed unrolled recovery method for solving (42). In line 2, 3 and 4, the variables $\mathbf{z}^{(0)}$, $\mathbf{r}_{\mathbf{z}}^{(0)}$ and $\mathbf{i}_{\mathbf{z}}^{(0)}$ are initialized based on initial guess obtained in Algorithm 3. The gradient $\nabla h(\mathbf{z}^{(t)}|(\mathbf{y}_\ell)_k)$ is computed in line 6. Then, the iteration step $\mathbf{z}^{(t+1)}$ is evaluated in line 7. In lines 8 and 9, the real and imaginary parts of $\mathbf{z}^{(t+1)}$ are calculated, respectively. Finally, the reconstruction of the optical field $\mathbf{z}$ is returned in line 11.

3.2.2. End-to-end phase recovery methodology Once the unrolled initialization $\mathcal{I}_{\boldsymbol{\theta}}(\cdot)$ and the unrolled recovery network $\mathcal{M}_{\boldsymbol{\theta}}(\cdot)$ are defined, they can be coupled into a single DNN model $\mathcal{N}_{\boldsymbol{\theta}}(\cdot)$, to be jointly optimized. Furthermore, the PR sensing step can be modeled as an optical layer [39], which allows all the parameter training in an E2E manner. Specifically, the E2E framework seeks to jointly design the MPM

Algorithm 4 Unrolling-based refined network

```
Input: Acquired data $\{\overline{\mathbf{D}}_\ell\mathbf{a}_k; (\mathbf{y}_\ell)_k\}$, maximum number of iterations $T$, $\tilde{\mathbf{z}}$ initial guess, $\alpha$ and $\rho$.
Compute $\mathbf{z}^{(0)} \leftarrow \tilde{\mathbf{z}}$.
Compute $\mathbf{r}_{\mathbf{z}}^{(0)} \leftarrow \mathcal{R}_{eal}(\tilde{\mathbf{z}})$.
Compute $\mathbf{i}_{\mathbf{z}}^{(0)} \leftarrow \mathcal{I}_{imaginary}(\tilde{\mathbf{z}})$.
for $t = 0 : T-1$ do
    $$\nabla h(\mathbf{z}^{(t)}|(\mathbf{y}_\ell)_k) := \frac{\alpha}{Ln}\sum_{\ell=1}^{L}\sum_{k=0}^{n-1}\left(\mathbf{a}_k^{\mathcal{H}}\mathbf{D}_\ell\mathbf{z}^{(t)} - \frac{\sqrt{(\mathbf{y}_\ell)_k}\,\mathbf{a}_k^{\mathcal{H}}\mathbf{D}_\ell\mathbf{z}^{(t)}}{\vartheta_\mu(|\mathbf{a}_k^{\mathcal{H}}\mathbf{D}_\ell\mathbf{z}^{(t)}|)}\right)\overline{\mathbf{D}}_\ell\mathbf{a}_k.$$
    $$\mathbf{z}^{(t+1)} := \mathbf{z}^{(t)} - \nabla h(\mathbf{z}^{(t)}|(\mathbf{y}_\ell)_k) - \alpha\rho\left(\mathbf{z}^{(t)} - \mathcal{P}_{\boldsymbol{\theta}}\left(\mathbf{r}_{\mathbf{z}}^{(t)}\right) - j\mathcal{Q}_{\boldsymbol{\theta}}\left(\mathbf{i}_{\mathbf{z}}^{(t)}\right)\right).$$
    Compute $\mathbf{r}_{\mathbf{z}}^{(t+1)} \leftarrow \mathcal{R}_{eal}(\mathbf{z}^{t+1})$.
    Compute $\mathbf{i}_{\mathbf{z}}^{(t+1)} \leftarrow \mathcal{I}_{imaginary}(\mathbf{z}^{t+1})$.
end for
Return: Reconstruction $\mathbf{z}$.
```

$\{\mathbf{D}_\ell\}_{\ell=1}^{L}$ and the parameters of the unrolled network using an error back-propagation method. Therefore, it is necessary to consider the sensing process as a fully differentiable model coupled to the unrolled network. Formally, from a set of $\mathcal{S}$ scenes $\{\mathbf{x}^{(s)}\}_{s=1}^{\mathcal{S}}$, the proposed E2E optimization problem consist of solving the following optimization problem

$$\begin{gathered}\{\mathbf{D}_\ell^{*}, \boldsymbol{\theta}^{*}\} \in \underset{\mathbf{D}_\ell,\boldsymbol{\theta}}{\operatorname{argmin}}\, \mathcal{L}(\mathbf{D}_\ell, \boldsymbol{\theta}|\mathbf{x}^{(s)}),\\ \mathcal{L}(\mathbf{D}_\ell, \boldsymbol{\theta}|\mathbf{x}^{(s)}) = \frac{1}{\mathcal{S}}\sum_{s=1}^{\mathcal{S}}\|\varphi\left(\mathcal{N}_{\boldsymbol{\theta}}(|\mathbf{a}_k^{\mathcal{H}}\mathbf{D}_\ell\mathbf{x}^{(s)}|^2)\right) - \varphi\left(\mathbf{x}^{(s)}\right)\|_1,\end{gathered} \tag{50}$$

where $\varphi(\cdot) : \mathbb{C}^n \rightarrow \mathbb{R}^n$ is a function that calculates the phase information of a complex signal. Algorithm 5 summarizes the proposed E2E methodology to solve the optimization problem stated in (50). In line 5, the acquisition model is implemented using (36). Then, in line 6, the optical field is estimated through Algorithm 3. The designed recovery DNN refines the optical field approximation in line 7 by using Algorithm 4. The loss function is evaluated in line 8. Besides, the gradients of $\mathbf{D}_\ell$ and $\boldsymbol{\theta}$ are computed in lines 9 and 10, respectively, which are used in the Adam update $\mathcal{A}_{dam}(\cdot)$ [108] weighted by fixing β_1 and β_2, respectively. Finally, the optimal MPMs and parameters $\boldsymbol{\theta}$ of the unrolled recovery phase network are returned in line 13.

Algorithm 5 Proposed E2E deep unrolling PR

```
 1: Input: Training set $\{\mathbf{x}^{(s)}\}_{s=1}^{\mathcal{S}}$ with $\mathcal{S}$ images.
 2: Initialize: Set the MPMs $\{\mathbf{D}_\ell\}_{\ell=1}^{L}$ from a uniform distribution.
 3: for epoch $= 1:\mathcal{E}$ do                                              ▷ $\mathcal{E}$ epochs
 4:     for $s = 1:\mathcal{S}$ do                                              ▷ $\mathcal{S}$ samples
 5:         $(\mathbf{y}_\ell)_k^{(s)} = |\mathbf{a}_k^{\mathcal{H}}\mathbf{D}_\ell\mathbf{x}^{(s)}|^2 + (\boldsymbol{\omega}_\ell)_k$, $k \in \{1,\cdots,n\}$, $\ell \in \{1,\cdots,L\}$.   ▷ CDPs
 6:         $\tilde{\mathbf{z}}^{(s)} \leftarrow \mathcal{I}_{\boldsymbol{\theta}}\left(\overline{\mathbf{D}}_\ell\mathbf{a}_k, (\mathbf{y}_\ell)_k^{(s)}\right)$.          ▷ Algorithm 3
 7:         $\mathbf{z}^{(s)} \leftarrow \mathcal{M}_{\boldsymbol{\theta}}\left(\overline{\mathbf{D}}_\ell\mathbf{a}_k, (\mathbf{y}_\ell)_k^{(s)}, \tilde{\mathbf{z}}^{(s)}\right)$.    ▷ Algorithm 4
 8:         $\mathcal{L}_{\mathbf{D}_\ell,\boldsymbol{\theta}} = \frac{1}{S}\sum_{s=1}^{\mathcal{S}}\|\varphi\left(\mathbf{z}^{(s)}\right) - \varphi\left(\mathbf{x}^{(s)}\right)\|_1$.   ▷ Loss function
 9:         $\mathbf{D}_\ell \leftarrow \mathcal{A}_{dam}(\mathbf{D}_\ell, \beta_1\nabla_{\mathbf{D}_\ell}\mathcal{L}_{\mathbf{D}_\ell,\boldsymbol{\theta}})$.        ▷ Optimize over $\mathbf{D}_\ell$
10:         $\boldsymbol{\theta} \leftarrow \mathcal{A}_{dam}(\boldsymbol{\theta}, \beta_2\nabla_{\boldsymbol{\theta}}\mathcal{L}_{\mathbf{D}_\ell,\boldsymbol{\theta}})$.        ▷ Optimize over $\boldsymbol{\theta}$.
11:     end for
12: end for
13: Return: Optimal MPMs $\{\mathbf{D}_\ell\}_{\ell=1}^{L}$ and parameters $\boldsymbol{\theta}$.
```

Once the proposed E2E deep unrolling PR approach is trained, the learned network infers the phase image from a computational complexity $\mathcal{O}(Tn^2)$.

Physics-based learned multilevel phase mask design The MPM can be parameterized through the phase response into the LCOS. In practice, there is a finite discrete grayscale-phase response between $[0, 2\pi]$ for each LCOS pixel, because of the liquid crystal's nonlinear wavelength response. In particular, this work assumes a phase-only LCOS, then, the coding elements into the MPM $\mathbf{D}_\ell$ can be modeled as

---

[108] Diederik P Kingma and Jimmy Ba. "Adam: A method for stochastic optimization". In: *arXiv preprint arXiv:1412.6980* (2014).

an exponential function

$$\mathbf{D}_\ell = \mathrm{diag}(e^{j\boldsymbol{\phi}_\ell}), \tag{51}$$

where $\mathrm{diag}(\cdot)$ returns a square diagonal matrix from a given vector and $\boldsymbol{\phi}_\ell \in \mathbb{R}^n$ is the phase delay generated at the LCOS for the ℓ-th snapshot. During training, a clip function maintains the phase values between $[0, 2\pi]$. It is worth clarifying that the MPM in (51) trivially fulfills the admissible random variable definition [12], where their entries d are independent and identically distributed (*i.i.d.*) copies, obeying $|d| \leq 1$ [96], which imposes that d cannot increase the power of the optical field [12]. Once the E2E training process is completed, the learned MPM is quantized to be implemented into an LCOS device. Then, it is introduced in an optical testbed system to acquire designed measurements. Subsequently, it is integrated into an optical testbed system to capture the designed CDPs. The trained unrolled network can then be directly applied, without modification, to the acquired CDPs to obtain an estimation of the phase image.

3.3. SIMULATION RESULTS

The performance of the proposed E2E approach is evaluated by using CDPs acquired under noiseless and noisy scenarios. In particular, the noisy cases are computed by different signal-to-noise (SNR) values, i.e., $(\boldsymbol{\omega}_\ell)_k \sim \mathcal{N}(0, \sigma^2)$ was simulated, where σ^2 was set such that certain $\mathrm{SNR} = 10\log_{10}\left(\sum_{\ell=1}^{L}\sum_{k=0}^{n-1}\left((\mathbf{y}_\ell)_k\right)^2 / \left(nL\sigma^2\right)\right)$ values were achieved. The proposed DNN approach was trained using Tensorflow on an Intel Xeon SILVER 4210R with 64 GB RAM and an NVIDIA GEFORCE RTX 3090 with 24 GB VRAM. We conduct various simulation experiments to analyze the proposed E2E method's performance considering the unrolling-based initialization and the unrolling-based network. Specifically, the proposed E2E method performance is calculated using different evaluating metrics, such as the relative error, the peak signal-to-noise ratio (PSNR), and the structural similarity index measure (SSIM). The resulting approximation accuracy in Algorithm 3 can be calculated by using the relative error from (38) as follows

$$\text{Relative error} := \frac{\operatorname{dist}(\tilde{\mathbf{z}}, \mathbf{x})}{\|\mathbf{x}\|_2}. \tag{52}$$

Finally, the proposed E2E method is compared to learning-based and non-learning state-of-the-art methods.

3.3.1. Image datasets The implemented network was trained using the datasets detailed below. For all the experiments, the images are scaled into the range $[0, 1]$ and used as phase information. Figure 26 illustrates some images present in these datasets.

- **Fashion MNIST:** The Fashion MNIST (Modified National Institute of Standards and Technology) dataset [103] contains $70,000$ grayscale images with a spatial dimension of 28×28 pixels divided into 10 categories. For the experiments, $10,000$ images were used for training, and 30 images were selected for testing.
- **BSDS:** The BSDS (Berkeley segmentation) dataset [109] contains 400 grayscale natural images with a spatial dimension of 180×180 pixels obtained from different natural scenes. This dataset was used to train the proposed method and the benchmark PR algorithms based on deep learning.
- **Set12:** The Set12 dataset [110] contains 12 grayscale images with a spatial dimension of 180×180 pixels obtained from different natural scenes. This dataset

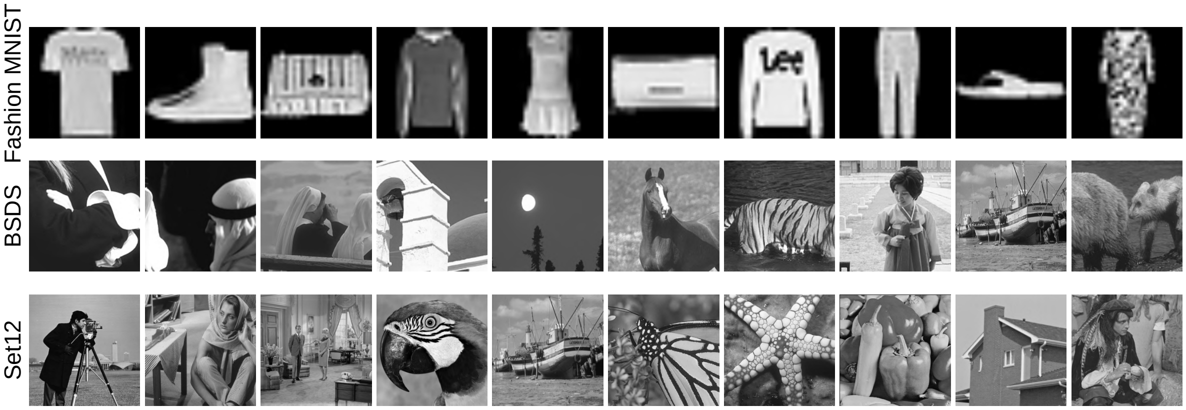
---

[109] Pablo Arbelaez et al. “Contour Detection and Hierarchical Image Segmentation”. In: *IEEE Trans. Pattern Anal. Mach. Intell.* 33.5 (May 2011), pp. 898–916. DOI: `10.1109/TPAMI.2010.161`.

[110] Kai Zhang et al. “Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising”. In: *IEEE transactions on image processing* 26.7 (2017), pp. 3142–3155.

was selected to test the proposed method and the benchmark PR algorithms.

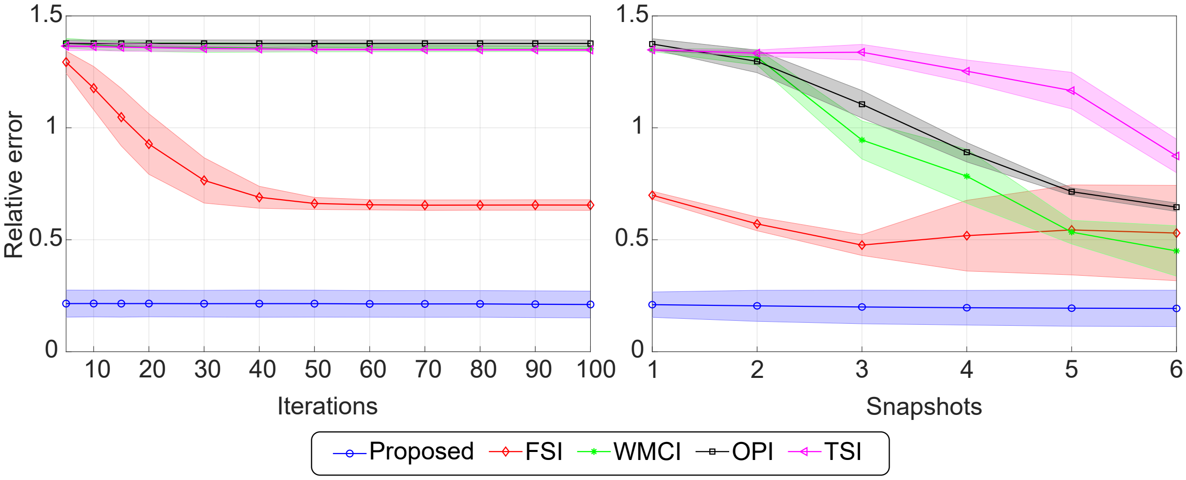
Figure 26. Fashion MNIST, BSDS, and Set12 datasets were used for performing the recovery task.



3.3.2. Initialization analysis The proposed unrolling-based initialization was compared to its alternatives under a noiseless scenario: FSI [96], orthogonality-promoting initialization (OPI) [26], weighted maximal correlation initialization (WMCI) [25], and truncated spectral initialization (TSI) [27]. For this test, the number of iterations T in Algorithm 3 varies from 5 to 100, and the number of snapshots from 1 to 6.

Here, the proposed initialization stage was trained for $\mathcal{E} = 100$ epochs using a batch size of 2 images with the Fashion MNIST dataset. Specifically, the filtering network architecture $\mathcal{G}_{\boldsymbol{\theta}}(\cdot)$ in the proposed unrolling-based initialization was fixed in three convolutional (Conv2D) layers with a kernel size of 3. The Adam optimizer was used for training the proposed initialization scheme by the loss function in (50) with a learning rate of 1×10^{-3}. To set the number of iterations in the proposed initialization and the number of snapshots, we fit the number of iterations $T \in \{5, 10, 15, 20, 40, 60, 80, 100\}$ and the number of snapshots $L \in \{1, 2, 3, 4, 5, 6\}$ in the algorithm 5 at each training. Figure 27 shows the relative error obtained by varying the iterations and snapshots, where the relative error slightly decreases as these parameters are increased. Note that the proposed method exhibits stable relative error behavior through the number of iterations and the number of snapshots, which allows the obtainment of initialization results with better performance than traditional approaches. A suitable scenario for the proposed initialization is to measure the optical field through a single snapshot and estimate it using $T = 5$ iterations.

Figure 27. Initialization results in terms of relative error under a noiseless scenario. (Left) Initialization performance by varying the number of iterations and fixing the number of snapshots $L = 1$. (Right) Initialization performance by varying the number of snapshots and fixing the number of iterations $T = 5$.



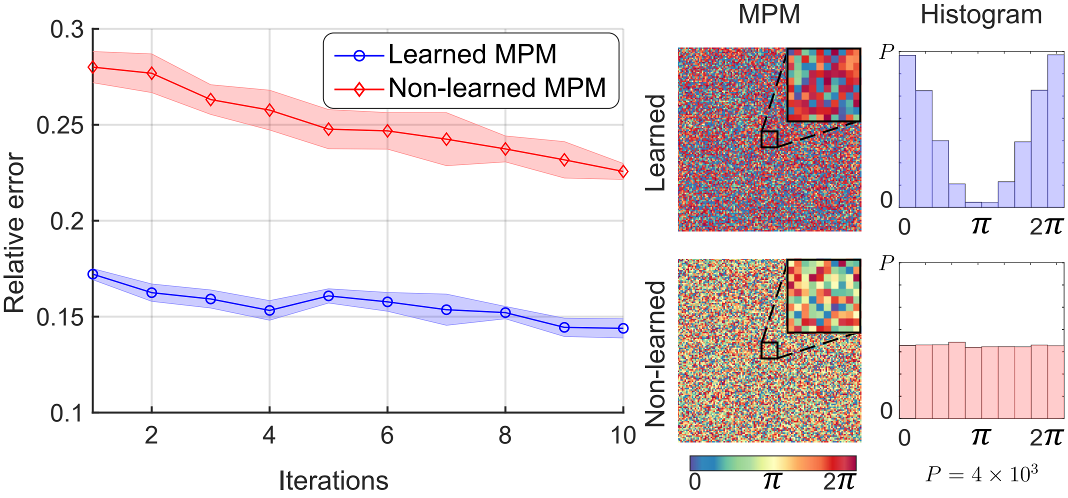
3.3.3. Recovery network analysis The initial setting of the proposed unrolled-based recovery scheme was trained using the Fashion MNIST dataset for $\mathcal{E} = 100$ epochs from a batch size of 2 with a learning rate of 1×10^{-3}. The proposed scheme was also trained from a single snapshot with $T = 5$ iterations on both the initialization and recovery stages. For this experiment, the number of layers and kernel size varied in the proposed network to obtain the best setting for recovery performance under a noiseless scenario. Table 5 shows the relative error across the testing dataset by varying the kernel size $\{3 \times 3, 5 \times 5, 7 \times 7\}$ and the number of Conv2D layers $\{3, 5, 7\}$. Note that the unrolling-based refined network achieves the best performance when the number of Conv2D layers is 3, and the kernel size is 3. Therefore, these parameters are fixed in the training stage of the proposed recovery network.

Table 5. Relative error analysis varying the number of layers and the kernel size in the proposed unrolling-based recovery network under a noiseless scenario. The best result is highlighted in green, and the second best is highlighted in yellow.

Layers	Kernel size		
	3×3	5×5	7×7
3	0.1673 ± 0.048	0.1732 ± 0.050	0.1701 ± 0.050
5	0.1728 ± 0.051	0.1728 ± 0.050	0.1723 ± 0.049
7	0.1681 ± 0.049	0.1729 ± 0.050	0.1717 ± 0.049

3.3.4. Multilevel phase mask analysis Figure 28 shows the relative error behavior for a learned and non-learned MPM by varying the number of iterations from 1 to 10 at each training of the proposed unrolling-based recovery method using a single snapshot across the Fashion MNIST dataset. Also, the learned and non-learned MPM distributions with their corresponding histogram are illustrated in Fig. 28. It is important to remark that the proposed PR approach improves the optical field estimation results compared to the initialization stage under a noiseless scenario. Also, this recovery approach shows a stable relative error performance along the iterations when the MPM is learned, which reduces the computational cost from a small number of iterations required by reconstructing the optical field. The proposed recovery network learns the MPM weights as a normal distribution, wherein the phase values are mostly 0 and 2π.

Figure 28. Recovery results for the learned and non-learned MPM in terms of relative error. The proposed unrolling-based refined network is tested by varying the number of iterations and fixing the number of snapshots $L = 1$ under a noiseless scenario.



3.3.5. Proposed recovery method against benchmark algorithms The proposed unrolled recovery network and traditional recovery methods are compared. Some PR algorithms were selected as the benchmark algorithms, such as DualPR-Net [35], prDeep [28], SPAR [111], BM3D-prGAMP [112], DOLPHIn [113], U-net [97], reweighted amplitude flow (RAF) [25], truncate amplitude flow (TAF) [26], and truncated Wirtinger flow (TWF) [27]. Notice that these recovery algorithms address the PR problem for the far-field model. To ensure an unbiased comparison, the proposed method was adapted to the far-field model, where $\mathbf{a}_k^{\mathcal{H}}$ in (36) now corresponds to the rows of the

---

[111] Vladimir Katkovnik and Karen Egiazarian. “Multi-frequency phase retrieval from noisy data”. In: *2018 26th European Signal Processing Conference (EUSIPCO)*. IEEE. 2018, pp. 2200–2204.

[112] Christopher A Metzler, Arian Maleki, and Richard G Baraniuk. “BM3D-PRGAMP: Compressive phase retrieval based on BM3D denoising”. In: *2016 IEEE International Conference on Image Processing (ICIP)*. IEEE. 2016, pp. 2504–2508.

[113] Andreas M Tillmann, Yonina C Eldar, and Julien Mairal. “DOLPHIn—dictionary learning for phase retrieval”. In: *IEEE Transactions on Signal Processing* 64.24 (2016), pp. 6485–6500.

DFT matrix $\mathbf{A} = \mathbf{F}$. It is important to highlight that both the proposed E2E method and PR benchmark algorithms were evaluated under the same scenario, guaranteeing the same MPM and noise distribution across the experiments.

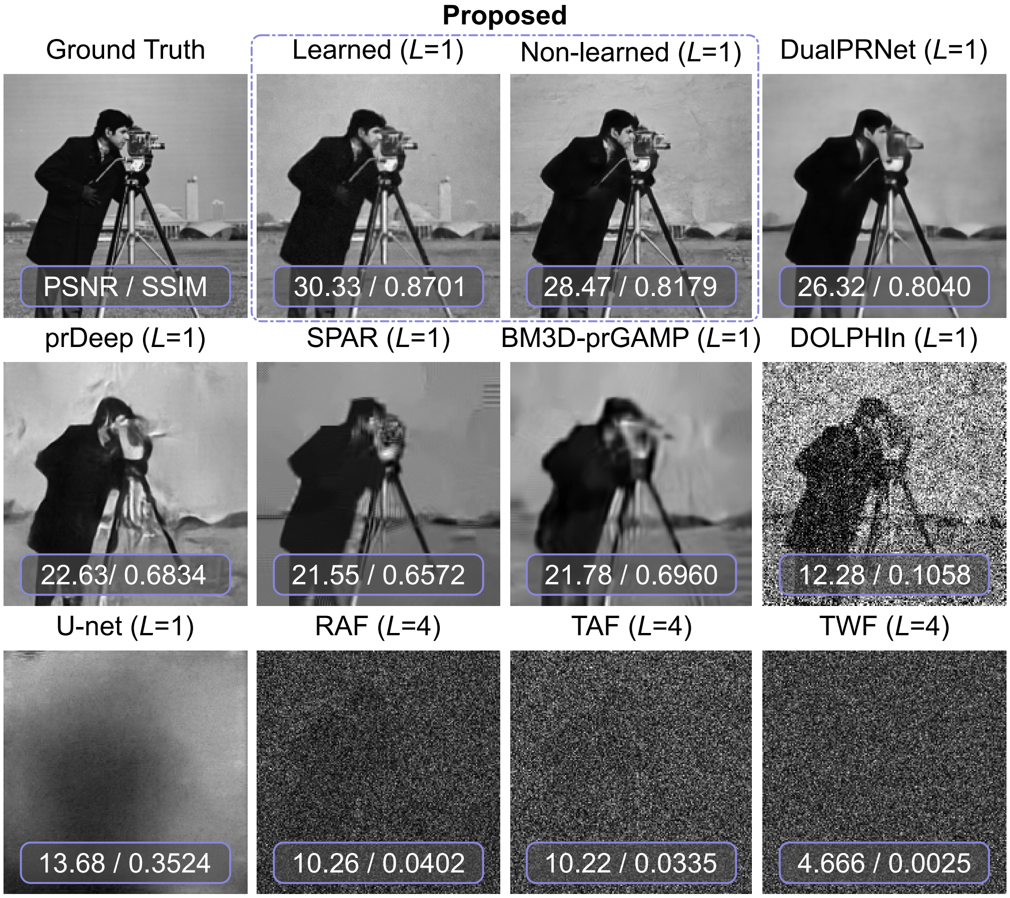
Table 6 summarizes the average PSNR and SSIM values for the benchmark PR algorithms and the proposed deep unrolled recovery network by using the Set12 dataset at different noise levels $\mathrm{SNR} \in \{5, 10, 15, 20\}$ $[\mathrm{dB}]$. The proposed method was evaluated by using a learned MPM and a non-learned MPM. In the latter case, the MPM was fixed to the same MPM implemented in the PR benchmark algorithms. The proposed method for the learned and non-learned MPM presents the best recovery performance by using a single snapshot $L = 1$ with $T = 5$ iterations for both the unrolling-based initialization and the unrolling-based reconstruction stages. It is important to mention that the proposed method was trained with $\mathcal{E} = 100$ epochs across the BSDS dataset for a fair comparison to the recovery results presented in [35]. Further, the recovery results for the algorithms DualPRNet, prDeep, SPAR, BM3D-prGAMP, and DOLPHIn were computed from a single snapshot $L = 1$. The U-net method was trained using the BSDS dataset with $\mathcal{E} = 100$ epochs by acquiring a single snapshot $L = 1$. Although RAF, TAF, and TWF algorithms reconstruct the acquired CDP from $L = 4$ snapshots and $T = 200$ for both the initialization and reconstruction stages, these approaches exhibit the lowest recovery performance compared to the other benchmark PR algorithms that employ only one snapshot $L = 1$. Overall, the proposed unrolling PR method improves the recovery performance up to 4 $[\mathrm{dB}]$ in terms of PSNR compared to the DualPRNet method under the most challenging noise scenario $\mathrm{SNR} = 5$ $[\mathrm{dB}]$. Although the DualPRNet framework incorporates an unrolled network inspired by the proximal gradient algorithm, this method lacks the initialization step and the learnable phase mask within the network structure, which are key components of the proposed E2E method.

Table 6. Quantitative results for the recovery task in terms of PSNR and SSIM by testing PR algorithms across the Set12 dataset at different noise levels. The best result is highlighted in green, and the second best is highlighted in yellow.

Algorithm	SNR							
	5 [dB]		10 [dB]		15 [dB]		20 [dB]	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
Proposed (Learned MPM)	28.18	0.8714	29.36	0.8846	30.36	0.8925	32.35	0.9012
Proposed (Non-learned MPM)	26.23	0.7100	28.17	0.8010	29.89	0.8611	31.96	0.8907
DualPRNet	24.46	0.6677	27.06	0.7530	29.67	0.8316	31.89	0.8805
prDeep	20.51	0.5471	24.83	0.7265	27.40	0.8088	30.55	0.8859
SPAR	21.72	0.6127	24.44	0.7236	27.03	0.8101	29.60	0.8735
BM3D-prGAMP	20.34	0.6356	23.13	0.7263	25.92	0.7933	28.40	0.8564
DOLPHIn	12.50	0.1233	17.62	0.2926	22.15	0.4987	24.89	0.6458
U-net	13.10	0.2476	13.67	0.2729	13.76	0.2975	13.84	0.2996
RAF	10.20	0.0454	12.76	0.1492	16.87	0.3055	21.34	0.4780
TAF	10.15	0.0377	13.41	0.1736	18.45	0.3647	23.29	0.5549
TWF	4.369	0.0039	5.451	0.0056	6.222	0.0059	6.631	0.0065

Figure 29 shows the cameraman image reconstruction using the proposed E2E non-convex unrolling PR method and the state-of-the-art PR algorithms with $\mathrm{SNR} = 5\ [\mathrm{dB}]$. Notice that the reconstructions obtained by TWF, TAF, RAF, U-net, and DOLPHIn display high noise levels. Despite the BM3D-prGAMP and SPAR algorithms considerably reducing the noise levels, the recovered images present many artifacts. In contrast, the prDeep and DualPRNet algorithms achieve reconstructed images with low noise levels and artifacts. However, these benchmark PR algorithms still produce recovered images with indistinguishable details. Visually, the proposed E2E method can obtain better reconstruction quality than traditional PR algorithms.

Figure 29. Visual results of recovered cameraman image from CDPs with $\mathrm{SNR} = 5$ $[\mathrm{dB}]$ by using the proposed E2E deep unrolling PR approach and benchmark PR algorithms.



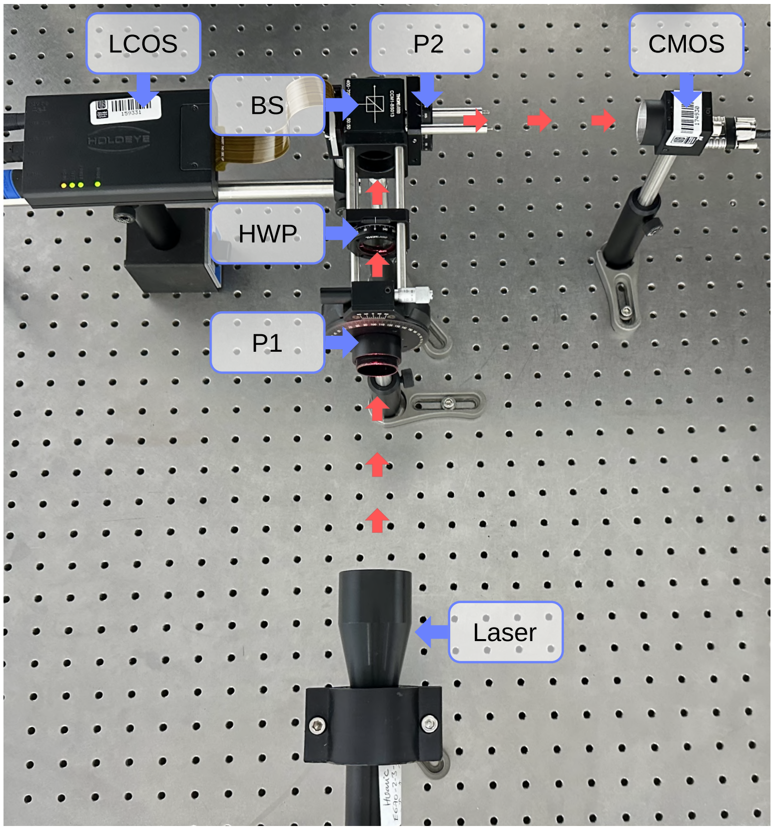
3.4. EXPERIMENTAL SETUP

This section describes the implemented optical setup, as illustrated in Fig 30. This experimental setup mainly contains a laser with a wavelength of $\lambda = 670$ $[nm]$; a CMOS camera FLIR BFS-U3-31S4M-C 2048×1536 with a pixel size of $\Delta_{\mathrm{CMOS}} = 3.45$ $[\mu m]$; and a phase-only reflective LCOS Holoeye PLUTO-2 NIR 011 1920×1080 with a pixel size of $\Delta_{\mathrm{LCOS}} = 8.0$ $[\mu m]$, and a fill factor of 93%.

In the implemented optical testbed, the light beam is polarized by a linear polarizer (P1), then, the polarization of the laser source is changed by a half-waveplate (HWP)

$\lambda/2$. This source passes through the non-polarizing beam splitter (BS) Thorlabs CCM1-BS013 to the LCOS, followed by a second linear polarizer (P2), obtaining the desired phase retardation. Finally, the wavefront is propagated until the CMOS camera. The optical distance between the LCOS and CMOS is $16\ [cm]$. According to [114], we can attach the sum of the object phase and the MPM on the LCOS screen. The proposed recovery algorithm was fine-tuned for estimating the underlying phase scene from the CDP acquired through the implemented optical testbed.

Figure 30. Optical setup for CDPs acquisition. Optical elements: Laser; P1 and P2, linear polarizers; HWP, half-waveplate; LCOS, phase modulator; BS, beam splitter; CMOS, registration camera.



---

[114] Vladimir Katkovnik et al. "Computational super-resolution phase retrieval from multiple phase-coded diffraction patterns: simulation study and experiments". In: *Optica* 4.7 (2017), pp. 786–794.

Figure 31. Real-world CDPs acquired from the implemented optical setup for both learned and non-learned MPM. The initializations and reconstructions were obtained from these CDPs using the proposed unrolling PR approach.

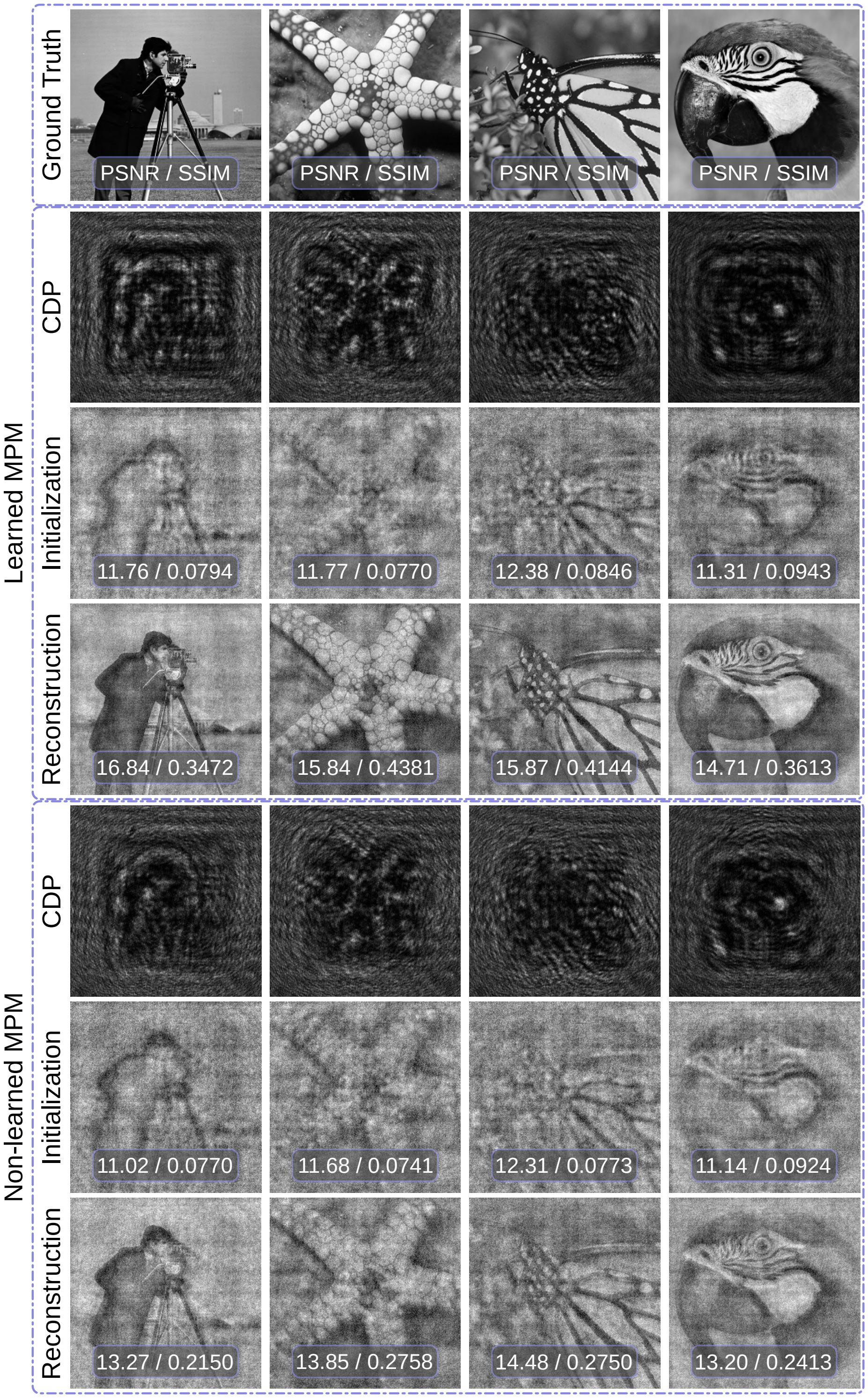


Figure 31 presents the real-world CDPs acquired for four scenes using the learned and non-learned MPMs. The coded scene is uploaded on the LCOS screen, where each grayscale value $[0, 255]$ is associated with the phase response $[0, 2\pi]$, corresponding to the optoelectronic phase modulator. Notice that the learned mask produces fewer noisy CDPs than the non-learned mask, which benefits the reconstruction process. In fact, the recovery results validate the mask design from the E2E optimization since the visual results for the learned mask are better than the non-learned mask. It is important to mention that the proposed non-convex unrolling PR method was fine-tuned from the acquired real-world CDPs for both learned and non-learned MPMs. This demonstrates that combining the MPM design and the non-convex formulation into an E2E unrolling approach for PR yields optimal results in real-world scenarios.

3.5. CHAPTER CONCLUSION

An unrolling-based network architecture for PR using a non-convex formulation is proposed. The proposed method comprises an optical stage based on coherent light that acquires the CDPs from the underlying scene using the near-field model; an unrolling-based initialization that estimates the optical field; and an unrolling-based refined network that further improves the initialization. The proposed E2E unrolled PR method jointly optimizes the deep unrolled network parameters and the MPM weights, enabling the approximation of the optical field from a single snapshot. Notice that the proposed method outperforms traditional benchmark PR algorithms across different noisy scenarios, as assessed by PSNR and SSIM metrics. Furthermore, experimental results demonstrate the proposed method's performance in practical scenarios, reconstructing real-world CDPs.

"The greatest challenge to any thinker is stating the problem in a way that will allow a solution."
*- **Bertrand Russell***

---

4. DUAL OPTICAL ARCHITECTURE FOR SPECTRAL CLASSIFICATION

This chapter describes a dual optical architecture for spectral classification. The proposed optical setup combines a single-pixel camera (SPC) that records the spectral information and a diffractive optical camera (DOC) that acquires the spatial information. Specifically, the DOC is modeled using the Fresnel approximation (middle-field propagation), wherein a multilevel phase mask (MPM) is incorporated. This MPM can be parameterized through the Zernike polynomials. An end-to-end (E2E) deep spectral classification framework is used to jointly learn the MPM and the network parameters to classify materials in land cover spectral datasets. The proposed E2E scheme is composed of three stages: an optical stage based on incoherent light that describes the propagation model tailored to the dual spectral setup leveraging the SPC and DOC; a spectral image fusion stage that computes the spectral image from the acquired measurements via the plug-and-play alternating direction method of multipliers (ADMM) algorithm; and a three-dimensional (3D) convolutional neural network (CNN) that labels the materials defined within the spectral datasets. The proposed E2E spectral classification based on the dual optical system achieves better classification performance than conventional optical setups. Finally, the proposed dual optical setup involves an incoherent source that illuminates the scene. The incoming wavefront is divided by a beam splitter (BS), where the first optical path encodes the scene through a deformable mirror (DM), then, the coded scene is integrated into a CCD camera; and the second optical path encodes the scene from a digital micromirror device (DMD), then, the coded scene is collected into a spectrometer. The implemented SPC leverages the spectrometer to acquire single-pixel spatial information across the spectrum. The experimental results validate the E2E deep spectral classification from the spatial-spectral measurements acquired by the implemented dual optical testbed. This chapter addresses all the specific objectives.

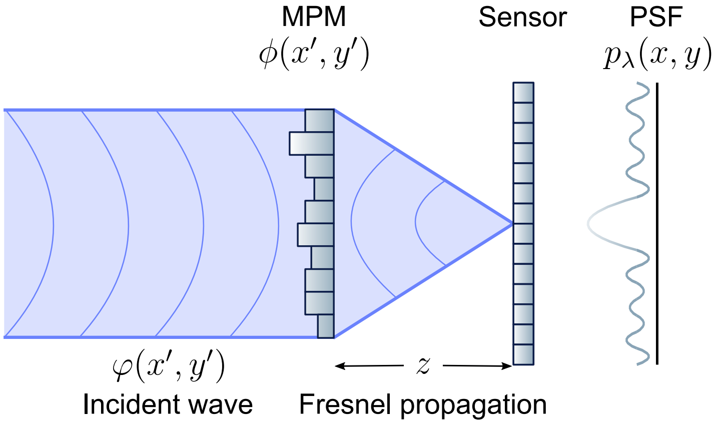
The main contributions of this work are summarized as follows:

- A dual optical architecture is proposed for addressing the spectral classification problem based on a single-pixel sensor and a diffractive system.
- An MPM design for spectral classification is presented. The height map distribution in the MPM is learned from the E2E deep spectral classification approach based on the Zernike polynomials.
- The proposed spectral classification methodology results in labeling materials across spectral datasets using a dual optical setup.

4.1. DIFFRACTIVE IMAGING MODEL

This section introduces the diffraction image formation model by assuming spatially incoherent light. The diffractive imaging system often consists of an MPM $\phi(x', y')$ and a sensor separated by a distance z, as shown in Fig. 32. Specifically, the MPM produces a phase delay in the incident wave $\varphi(x', y')$ before reaching the sensor.

Figure 32. Diffractive imaging system based on an MPM. A light wave $\varphi(x', y')$ with a wavelength λ incidents on the aperture plane containing an MPM $\phi(x', y')$. This MPM shifts the phase information present on the incident wavefront. The resulting wavefront is propagated to a distance z until the sensor. The collected intensities into the sensor describe the PSF $p_\lambda(x, y)$.



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

$$\phi(x', y') = \frac{2\pi\Delta\eta}{\lambda} h(x', y'), \tag{53}$$

where $\Delta\eta$ is the refractive index difference between air and the MPM material [74], λ is the wavelength, and $h(x', y')$ is the height map. A wave field with amplitude $A_\lambda(x', y')$ and phase $\varphi(x', y')$ incident on the MPM will be affected as

$$\gamma_\lambda(x', y'; z = 0) = A_\lambda(x', y') e^{j[\varphi(x', y') + \phi(x', y')]}, \tag{54}$$

where $j = \sqrt{-1}$, such that $\gamma_\lambda(x', y'; 0)$ is the wave field right after it passes through the MPM. As illustrated in Fig. 32, the optical field is propagated a distance z in free space, which can be modeled by the Fresnel approximation such that $\lambda \ll z$

$$\gamma_\lambda(x, y; z) = \frac{e^{jk_0 z}}{j\lambda z} \iint \gamma_\lambda(x', y'; 0) e^{\frac{jk_0}{2z}\left[(x - x')^2 + (y - y')^2\right]} dx' dy', \tag{55}$$

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

From (55), the point spread function (PSF) p_λ is formulated as

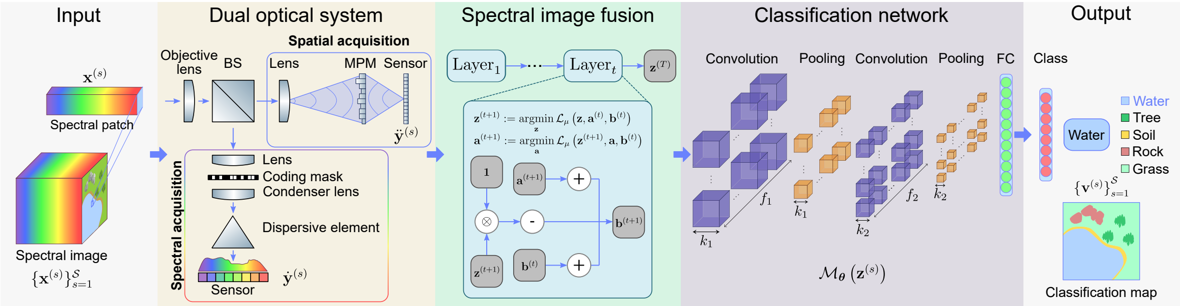
$$p_\lambda(x, y) \propto \left| \mathcal{F}\left\{ A_\lambda(x', y') e^{j[\varphi(x', y') + \phi(x', y')]} e^{\frac{j\pi}{\lambda z}\left(x'^2 + y'^2\right)} \right\} \right|^2, \tag{56}$$

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

4.2. PROPOSED SPECTRAL CLASSIFICATION METHODOLOGY

The proposed methodology mainly comprises three stages, as shown in Fig. 33, where an optical acquisition stage models the forward propagation model according to the DOC and SPC; a spectral image fusion stage computes the spectral image; and a classification network labels the elements within the spectral image [55]. For the optical stage, a spectral patch $\mathbf{x}^{(s)}$ is equally divided by a BS, where two optical branches separately acquire the spatial-spectral measurements by using the DOC and SPC, respectively. For the spectral fusion, an unrolling network based on the plug-and-play ADMM algorithm $\mathcal{I}_{\boldsymbol{\theta}}(\cdot)$ is implemented, where the proximal operator is learned. Last, for the classification network, a 3D CNN $\mathcal{M}_{\boldsymbol{\theta}}(\cdot)$ is composed of 3D convolutions, 3D MaxPooling, and fully connected (FC) layers.

Figure 33. Proposed E2E spectral classification approach based on the SPC and DOC. The dual optical system propagates the spectral patch $\mathbf{x}^{(s)}$. Then, the plug-and-play ADMM algorithm $\mathcal{I}_{\boldsymbol{\theta}}(\cdot)$ estimates the spectral patch $\mathbf{z}^{(s)}$ from the spatial-spectral measurements $\dot{\mathbf{y}}^{(s)}$ and $\ddot{\mathbf{y}}^{(s)}$. Finally, the classification network $\mathcal{M}_{\boldsymbol{\theta}}(\cdot)$ returns the corresponding class $\mathbf{v}^{(s)}$ based on the fused spectral patch.



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

Diffractive optical camera After formulating the PSF in (56), the DOC image formation model can be modeled as [115]

$$I(x,y) = \int_{\Lambda} \delta(\lambda)[I_{\lambda}(x,y) * p_{\lambda}(x,y)]d\lambda, \tag{57}$$

where $I_{\lambda}(x,y)$ represents the spectral image, $\delta(\lambda)$ denotes the sensor spectral sensitivity function, and $*$ is the convolution operator.

The image formation model (57) can be written in a discrete vector-matrix version. Let $\mathbf{x} \in \mathbb{R}^{nK}$ be the original spectral image vector with K as the spectral bands. Now, we can represent the sensor sensitivity $\delta(\lambda)$ and the convolution by the PSF $p_{\lambda}(x,y)$ as matrices $\boldsymbol{\Delta} \in \mathbb{R}^{n \times nK}$ and $\mathbf{P} \in \mathbb{R}^{nK \times nK}$, respectively. Then, the continuous image formation model (57) can be represented in a discrete matrix form as

$$\ddot{\mathbf{y}} = \ddot{\boldsymbol{\Psi}}\mathbf{x} + \ddot{\boldsymbol{\omega}}, \tag{58}$$

where $\ddot{\mathbf{y}} \in \mathbb{R}^{n}$ is the spatial measurements, $\ddot{\boldsymbol{\Psi}} \in \mathbb{R}^{n \times nK}$ is the DOC sensing matrix obtained from the product of $\boldsymbol{\Delta}$ and $\mathbf{P}$, and $\ddot{\boldsymbol{\omega}} \in \mathbb{R}^{n}$ is the spatial sensor noise.

Single pixel camera From the κ-th spectral band $\mathbf{x}_{\kappa} \in \mathbb{R}^{n}$, the SPC image formation at the ℓ-th snapshot can be modeled as [116]

$$(\mathbf{y}_{\ell})_{\kappa} = \boldsymbol{\psi}_{\ell}^{\mathcal{T}} \mathbf{x}_{\kappa} + (\boldsymbol{\omega}_{\ell})_{\kappa}, \tag{59}$$

---

[115] Vincent Sitzmann et al. “End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging”. In: *ACM Transactions on Graphics (TOG)* 37.4 (2018), pp. 1–13.

[116] Hans Garcia, Claudia V Correa, and Henry Arguello. “Optimized sensing matrix for single pixel multi-resolution compressive spectral imaging”. In: *IEEE Transactions on Image Processing* 29 (2020), pp. 4243–4253.

for $\kappa \in \{1, \cdots, K\}$ and $\ell \in \{1, \cdots, L\}$ with L as the total number of snapshots, where $(\mathbf{y}_\ell)_\kappa \in \mathbb{R}$ is the spectral measurement, $\boldsymbol{\psi} \in \mathbb{R}^n$ is the ℓ-th Hadamard pattern vector, $(\boldsymbol{\omega}_\ell)_\kappa \in \mathbb{R}$ is the spectral sensor noise, and $(\cdot)^{\mathcal{T}}$ is the transpose operator. The sensing model, including all snapshots L, is defined as

$$\mathbf{y}_\kappa = \boldsymbol{\Psi}\mathbf{x}_\kappa + \boldsymbol{\omega}_\kappa, \tag{60}$$

with $\mathbf{y}_\kappa = [(\mathbf{y}_1)_\kappa, \cdots, (\mathbf{y}_L)_\kappa]^{\mathcal{T}}$ and $\boldsymbol{\omega}_\kappa = [(\boldsymbol{\omega}_1)_\kappa, \cdots, (\boldsymbol{\omega}_L)_\kappa]^{\mathcal{T}}$, where $\boldsymbol{\Psi} \in \mathbb{R}^{L \times n}$, specifically, $\boldsymbol{\Psi} = [\boldsymbol{\psi}_1, \cdots, \boldsymbol{\psi}_L]^{\mathcal{T}}$. Finally, staking the measurements and the spectral noise from all spectral bands in a single vector $\dot{\mathbf{y}} = \left[\mathbf{y}_1^{\mathcal{T}}, \cdots, \mathbf{y}_K^{\mathcal{T}}\right]^{\mathcal{T}}$ and $\dot{\boldsymbol{\omega}} = \left[\boldsymbol{\omega}_1^{\mathcal{T}}, \cdots, \boldsymbol{\omega}_K^{\mathcal{T}}\right]^{\mathcal{T}}$, respectively, the sensing model for all spectral bands K and snapshots L can be expressed as

$$\dot{\mathbf{y}} = \dot{\boldsymbol{\Psi}}\mathbf{x} + \dot{\boldsymbol{\omega}}, \tag{61}$$

where $\dot{\boldsymbol{\Psi}} \in \mathbb{R}^{LK \times nK}$ is the SPC sensing matrix.

4.2.2. Multilevel phase mask parameterization To train the MPM values (53) using the proposed E2E optimization, the height map in the MPM can be parameterized by using Zernike polynomials [117]

$$\mathbf{H} = \sum_{o=1}^{O} \alpha_o \Xi_o \tag{62}$$

where $\mathbf{H} \in \mathbb{R}^{n' \times n'}$ corresponds to the discretized version of the height map in (53), Ξ_o denotes the o-th Zernike polynomial in Noll notation, and α_o is the corresponding

---

[117] Carlos Hinojosa, Juan Carlos Niebles, and Henry Arguello. “Learning privacy-preserving optics for human pose estimation”. In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2021, pp. 2573–2582.

coefficient [118]. Each Zernike polynomial describes a wavefront aberration [115], thus, the MPM is formed by the linear combination of O aberrations.

4.2.3. Spectral image fusion Once the spectral information (61) and spatial information (58) are acquired, the spectral image $\mathbf{z} \in \mathbb{R}^{nK}$ can be calculated from the following optimization problem

$$\underset{\mathbf{z}\in\mathbb{R}^{n^2K}}{\operatorname{argmin}} \frac{1}{2}\|\dot{\mathbf{y}} - \dot{\mathbf{\Psi}}\mathbf{z}\|_2^2 + \frac{1}{2}\|\ddot{\mathbf{y}} - \ddot{\mathbf{\Psi}}\mathbf{z}\|_2^2 + \tau\|\mathbf{z}\|_{TV}, \tag{63}$$

where $\tau > 0$ is a regularization parameter and $\|\cdot\|_{TV}$ represents the total variation norm. The minimization problem in (63) can be solved by using the ADMM algorithm [119], where the objective function and the constraint are split from auxiliary variables as below

$$\begin{aligned} &\underset{\mathbf{z},\mathbf{a}_\psi}{\operatorname{argmin}} \frac{1}{2}\|\dot{\mathbf{y}} - \dot{\mathbf{\Psi}}\mathbf{a}_1\|_2^2 + \frac{1}{2}\|\ddot{\mathbf{y}} - \ddot{\mathbf{\Psi}}\mathbf{a}_2\|_2^2 + \tau\|\mathbf{a}_3\|_{TV} \\ &\qquad\qquad \text{subject to } \mathbf{a}_\psi = \mathbf{z}, \end{aligned} \tag{64}$$

where $\mathbf{a}_\psi \in \mathbb{R}^{nK}$ is the split variable with $\psi \in \{1, 2, 3\}$. For notational convenience, we introduce $\mathbf{a} = \left[\mathbf{a}_1^{\mathcal{T}}, \mathbf{a}_2^{\mathcal{T}}, \mathbf{a}_3^{\mathcal{T}}\right]^{\mathcal{T}}$. Then, the objective function is defined as

$$f(\mathbf{a}) = \frac{1}{2}\|\dot{\mathbf{y}} - \dot{\mathbf{\Psi}}\mathbf{a}_1\|_2^2 + \frac{1}{2}\|\ddot{\mathbf{y}} - \ddot{\mathbf{\Psi}}\mathbf{a}_2\|_2^2 + \tau\|\mathbf{a}_3\|_{TV}. \tag{65}$$

---

[118] Max Born and Emil Wolf. *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light*. Elsevier, 2013.

[119] Edwin Vargas, Henry Arguello, and Jean-Yves Tourneret. “Spectral image fusion from compressive measurements using spectral unmixing and a sparse representation of abundance maps”. In: *IEEE Transactions on Geoscience and Remote Sensing* 57.7 (2019), pp. 5043–5053.

As a consequence, (64) can be reduced to

$$\underset{\mathbf{z},\mathbf{a}}{\operatorname{argmin}}\, f(\mathbf{a}) \;\text{ subject to } \mathbf{a} = \mathbf{1} \otimes \mathbf{z}, \tag{66}$$

where $\mathbf{1} \in \mathbb{R}^3$ is an all-ones vector and $\otimes$ represents the Kronecker product. The augmented Lagrangian associated with (66) is given by

$$\mathcal{L}_\mu(\mathbf{z}, \mathbf{a}, \mathbf{b}) = f(\mathbf{a}) + \frac{\mu}{2}\|\mathbf{a} - \mathbf{1} \otimes \mathbf{z} + \mathbf{b}\|_2^2, \tag{67}$$

where $\mathbf{b} \in \mathbb{R}^{3nK}$ is the dual variable, and $\mu \geq 0$ corresponds to the dual regularizer parameter. The plug-and-play ADMM [120] procedure is summarized in Algorithm 6. In line 2, the split $\mathbf{a}^{(0)}$ and dual $\mathbf{b}^{(0)}$ variables are initialized as all-zeros vectors. The plug-and-play ADMM iterations are computed in lines 4, 5, and 6. It is important to highlight that the sub-problem associated with the split variable $\mathbf{a}_3$ in (65) can be solved using a proximal operator of the TV regularization, then, this variable is computed by using a deep proximal operator at each iteration [121]. Finally, the spectral image $\mathbf{z}$ is returned in line 8.

4.2.4. Classification network The spectral classification network, denoted as $\mathcal{M}_{\boldsymbol{\theta}}(\cdot)$, receives as input a spectral patch $\mathbf{z}$. Specifically, this neural network corresponds to a 3D CNN that integrates spatial and spectral features into a joint spatial-spectral classification framework [55]. This CNN contains two 3D convolutions, two 3D MaxPooling operators, and a fully connected layer to conduct the labeling step. This

---

[120] Stanley H Chan, Xiran Wang, and Omar A Elgendy. “Plug-and-play ADMM for image restoration: Fixed-point convergence and applications”. In: *IEEE Transactions on Computational Imaging* 3.1 (2016), pp. 84–98.

[121] Roman Jacome, Jorge Bacca, and Henry Arguello. “D 2 uf: Deep coded aperture design and unrolling algorithm for compressive spectral image fusion”. In: *IEEE Journal of Selected Topics in Signal Processing* (2022).

Algorithm 6 Plug-and-play ADMM spectral fusion

```
Input: Acquired spectral data for SPC $\{\dot{\boldsymbol{\Psi}}; \dot{\mathbf{y}}\}$ and DOC $\{\ddot{\boldsymbol{\Psi}}; \ddot{\mathbf{y}}\}$, and maximum number of iterations $T$.
Initialize: $\mathbf{a}^{(0)} = \mathbf{0}$ and $\mathbf{b}^{(0)} = \mathbf{0}$.
for $t = 1 : T-1$ do
    $\mathbf{z}^{(t+1)} := \underset{\mathbf{z}}{\operatorname{argmin}}\, \mathcal{L}_{\mu}\left(\mathbf{z}, \mathbf{a}^{(t)}, \mathbf{b}^{(t)}\right).$    ▷ $\mathbf{z}$-minimization
    $\mathbf{a}^{(t+1)} := \underset{\mathbf{a}}{\operatorname{argmin}}\, \mathcal{L}_{\mu}\left(\mathbf{z}^{(t+1)}, \mathbf{a}, \mathbf{b}^{(t)}\right).$    ▷ $\mathbf{a}$-minimization
    $\mathbf{b}^{(t+1)} := \mathbf{a}^{(t+1)} + \mathbf{b}^{(t)} - \mathbf{1} \otimes \mathbf{z}^{(t+1)}.$    ▷ Dual update
end for
Return: Fused spectral image $\mathbf{z}$.
```

neural network learns the parameters $\boldsymbol{\theta}$ mapping the input (spectral patches) and the output (corresponding classes). The classification loss function can be expressed by the generalized cross-entropy (CE) over multiple classes as

$$\mathcal{L}_{\mathrm{CE}} := -\sum_{c=1}^{C} (\mathbf{u})_c \log\left((\mathbf{v})_c\right), \tag{68}$$

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

4.2.5. End-to-end spectral classification approach The proposed spectral classification scheme, illustrated in Fig. 33, jointly learns the height map values $\mathbf{H}$ in the MPM and the parameters $\boldsymbol{\theta}$ in the classification network. From $\mathcal{S}$ patches extracted from a spectral image, the E2E optimization problem can be formulated as

$$\begin{aligned} \{\mathbf{H}^*, \boldsymbol{\theta}^*\} &\in \underset{\mathbf{H}, \boldsymbol{\theta}}{\operatorname{argmin}}\, \mathcal{L}\left(\mathbf{H}, \boldsymbol{\theta} | \mathbf{u}^{(s)}, \mathbf{v}^{(s)}\right), \\ \mathcal{L}\left(\mathbf{H}, \boldsymbol{\theta} | \mathbf{u}^{(s)}, \mathbf{v}^{(s)}\right) &= \frac{1}{\mathcal{S}} \sum_{s=1}^{\mathcal{S}} \mathcal{L}_{\mathrm{CE}}\left(\mathbf{u}^{(s)}, \mathbf{v}^{(s)}\right) + \rho \mathcal{R}\left(\mathbf{H}\right), \end{aligned} \tag{69}$$

where $\rho > 0$ is a regularization parameter, and $\mathcal{R}(\cdot)$ is a regularization function based on the Zernike polynomials (62). Algorithm 7 summarizes the proposed E2E optimization for spectral classification. This algorithm receives as input the spectral image $\mathbf{x}^{(s)}$ and the corresponding class $\mathbf{u}^{(s)}$ in line 1. Then, the height map $\mathbf{H}$ is initialized as an all-zeros matrix in line 2. In lines 5 and 6, the spatial-spectral measurements $\ddot{\mathbf{y}}^{(s)}$ and $\dot{\mathbf{y}}^{(s)}$ are computed by using (58) and (61), respectively. The spectral image fusion $\mathbf{z}^{(s)}$ is calculated in line 7 by using the plug-and-play ADMM algorithm $\mathcal{I}_{\boldsymbol{\theta}}(\cdot)$ described in the Algorithm 6. The labeling $\mathbf{v}^{(s)}$ is obtained in line 8. Line 9 evaluates the loss function. In addition, the gradients of $\mathbf{H}$ and $\boldsymbol{\theta}$ are estimated in lines 10 and 11, respectively, which are used in the Adam update $\mathcal{A}_{dam}(\cdot)$ weighted by fixing β_1 and β_2, respectively. Finally, the optimal height map and network parameters of the spectral classification network are returned in line 14.

Algorithm 7 Proposed E2E deep spectral classification

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

Once the proposed E2E deep spectral classification architecture is trained, the learned network infers the classification map from a computational complexity $\mathcal{O}(Tn^2K^2)$.

4.3. NUMERICAL RESULTS

This section presents the classification results from spatial-spectral measurements acquired by the proposed dual architecture under noiseless and noisy scenarios. Specifically, the noisy cases in spatial-spectral measurements are fixed by different signal-to-noise (SNR) values, i.e., both $\dot{\boldsymbol{\omega}} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_{LK})$ and $\ddot{\boldsymbol{\omega}} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_n)$ were simulated, where σ^2 was set such that certain

$$\mathrm{SNR} = \begin{cases} 10\log_{10}\left(\|\dot{\mathbf{y}}\|_2^2/(LK\sigma^2)\right) & \text{for SPC} \\ 10\log_{10}\left(\|\ddot{\mathbf{y}}\|_2^2/(n\sigma^2)\right) & \text{for DOC} \end{cases} \tag{70}$$

values were achieved. The proposed E2E approach was trained using Tensorflow on a GPU Tesla T4 with 15 GB VRAM and 12.7 GB RAM. Several experiments were conducted to analyze the proposed methodology's performance through different evaluating metrics, such as accuracy, precision, recall, and F1 score.

4.3.1. Spectral image datasets Three public spectral datasets [122] were selected to train, validate and test the proposed E2E spectral classification methodology. From each spectral dataset, random patches were uniformly extracted. In particular, the training, validating, and testing samples correspond to 80%, 10%, and 10%, respectively. For instance, Table 7 shows the number of spectral patches extracted at each dataset with patch size $W = 8 \times 8$.

- **Indian Pines:** The Indian Pines dataset contains a spatial resolution of 145×145 pixels and a spectral resolution of 200 bands distributed along the spectral range $400 - 2500$ $[nm]$ for 16 different classes.

---

[122] The hyperspectral datasets are available on Grupo de Inteligencia Computacional (GIC) Web Page

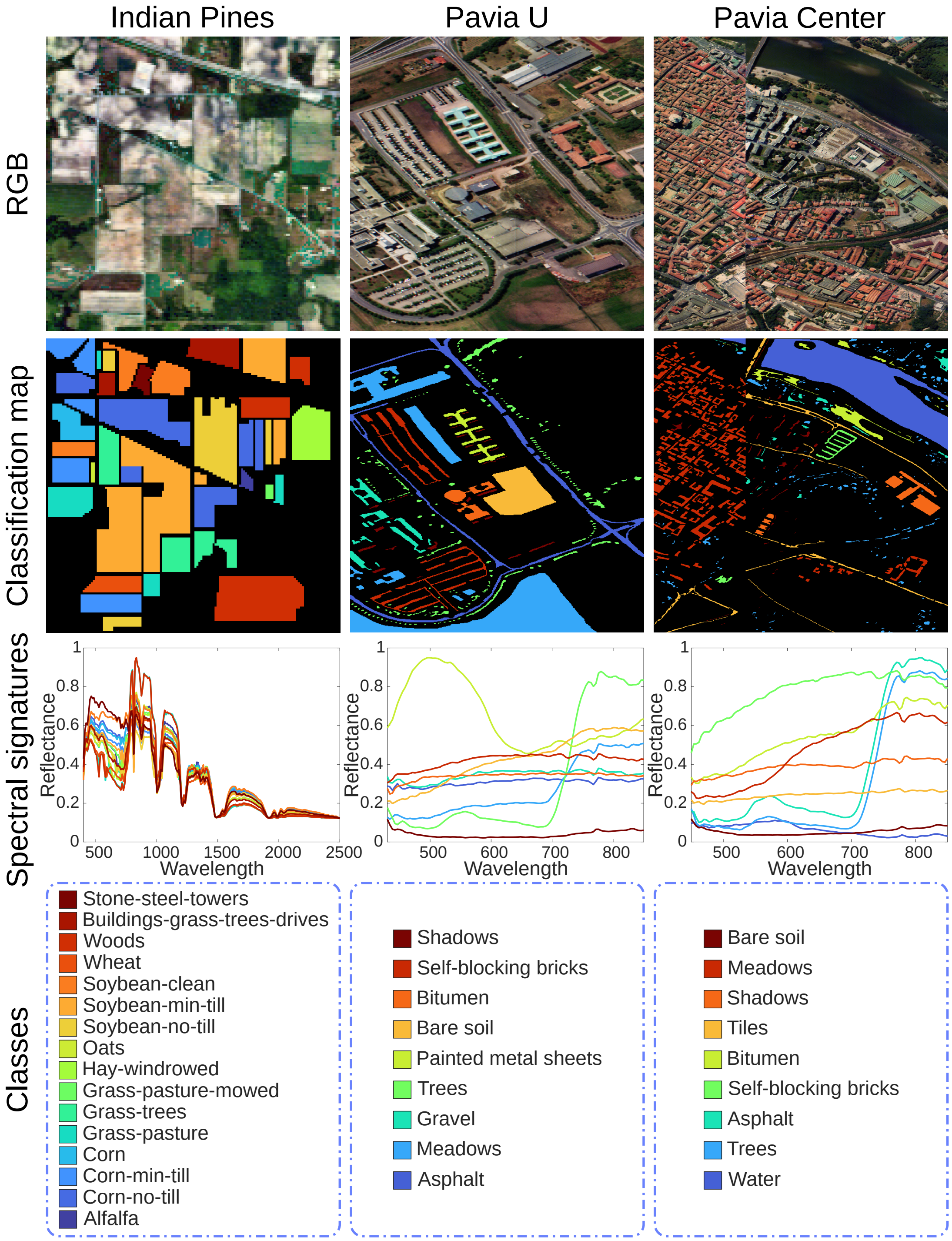
- **Pavia University:** The Pavia U dataset contains a spatial resolution of 610×340 pixels and a spectral resolution of 103 bands distributed along the spectral range $430 - 850$ $[nm]$ for 9 different classes.
- **Pavia Center:** The Pavia Center dataset contains a spatial resolution of 1096×715 pixels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

Table 7. Summary of the number of patches selected for training, validation, and testing according to each spectral dataset.

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

Figure 34 illustrates the spectral datasets: Indian Pines, Pavia University, and Pavia Center. The false RGB representation, the classification map, the spectral signatures for each land cover class, and the classes are shown for each spectral dataset. In the Indian Pines dataset, the classes are stone-steel-towers, buildings-grass-trees-drives, woods, what, soybean-clean, soybean-min-till, soybean-no-till, oats, hay-windrowed, grass-pasture-mowed, grass-trees, grass-pasture, corn, corn-min-till, corn-no-till, and alfalfa. In the Pavia University dataset, the classes are shadows, self-blocking bricks, bitumen, bare soil, painted metal sheets, trees, gravel, meadows, and asphalt. Finally, in the Pavia Center dataset, the classes are bare soil, meadows, shadows, tiles, bitumen, self-blocking bricks, asphalt, trees, and water.

Figure 34. Indian Pines, Pavia U, and Pavia Center spectral datasets were used for performing the classification task. Columns: Spectral datasets. Row 1: False RGB representation. Row 2: Classification map. Row 3: Spectral signatures for each land cover class. Row 4: Land cover classes.



4.3.2. Evaluation metrics The metrics for evaluating the proposed object classification method are presented. Specifically, we use accuracy, precision, recall, and F1 score that depend on the true positive (TP), true negative (TN), false positive (FP), and false negative (FN) associated with each class. These evaluation metrics are mathematically defined below.

- **Accuracy** is the ratio of correctly predicted instances to the total number of instances in the dataset. This metric is expressed as

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}. \tag{71}$$

- **Precision** is the ratio of true positives (instances correctly predicted as belonging to the positive class) to the sum of true positives and false positives (instances predicted as positive but actually belonging to a different class). This metric is defined as

$$\text{Precision} = \frac{TP}{TP + FP}. \tag{72}$$

- **Recall** is the ratio of true positives to the sum of true positives and false negatives (instances belonging to the positive class that were not correctly predicted). This metric is calculated as

$$\text{Recall} = \frac{TP}{TP + FN}. \tag{73}$$

- **F1 score** is the harmonic mean of precision (72) and recall (73), combining both aspects of a classifier's ability to correctly identify positive instances while minimizing false positives and false negatives. This metric is defined as

$$\text{F1 score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{74}$$

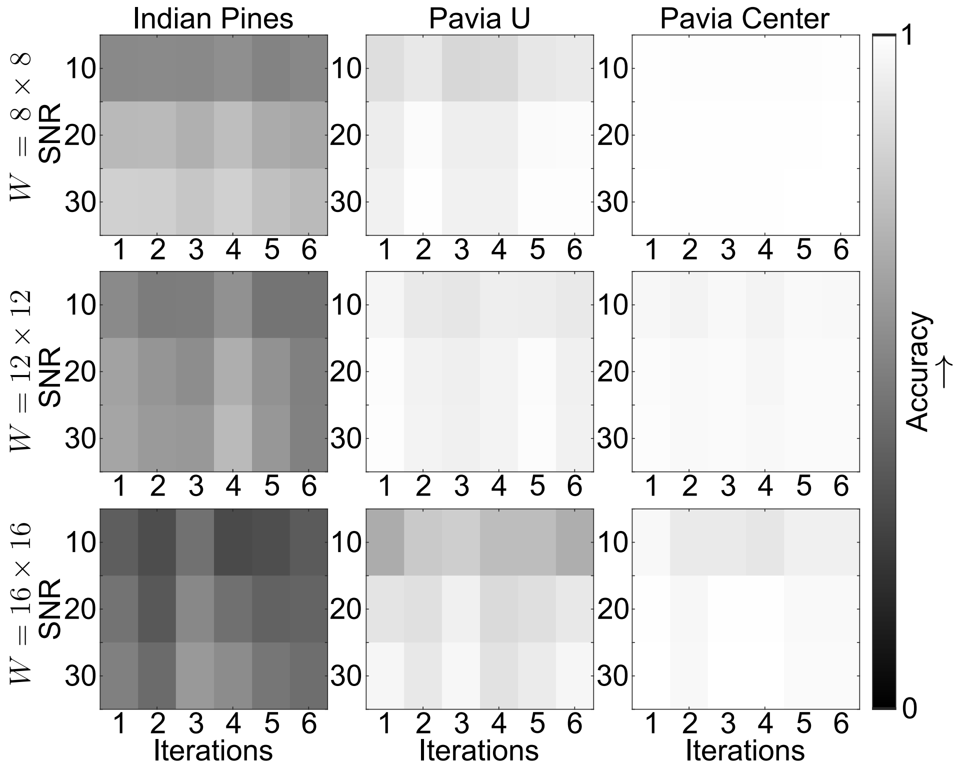
4.3.3. Classification network analysis The proposed E2E classification approach was trained for each spectral dataset during $\mathcal{E} = 100$ epochs from a batch size of 64 spectral patches by using the Adam optimizer with a learning rate of 1×10^{-3}. To analyze the classification performance, the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

Figure 35 presents the classification performance behavior by varying the noise level $\mathrm{SNR} \in \{10, 20, 30\}$ [dB], the number of iterations $T \in \{1, 2, 3, 4, 5, 6\}$, and the patch size $W \in \{8 \times 8, 12 \times 12, 16 \times 16\}$ pixels in terms of accuracy across each testing spectral dataset. For this experiment, the number of spectral bands was fixed at $K = 12$ across all spectral datasets, with uniform selection.

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

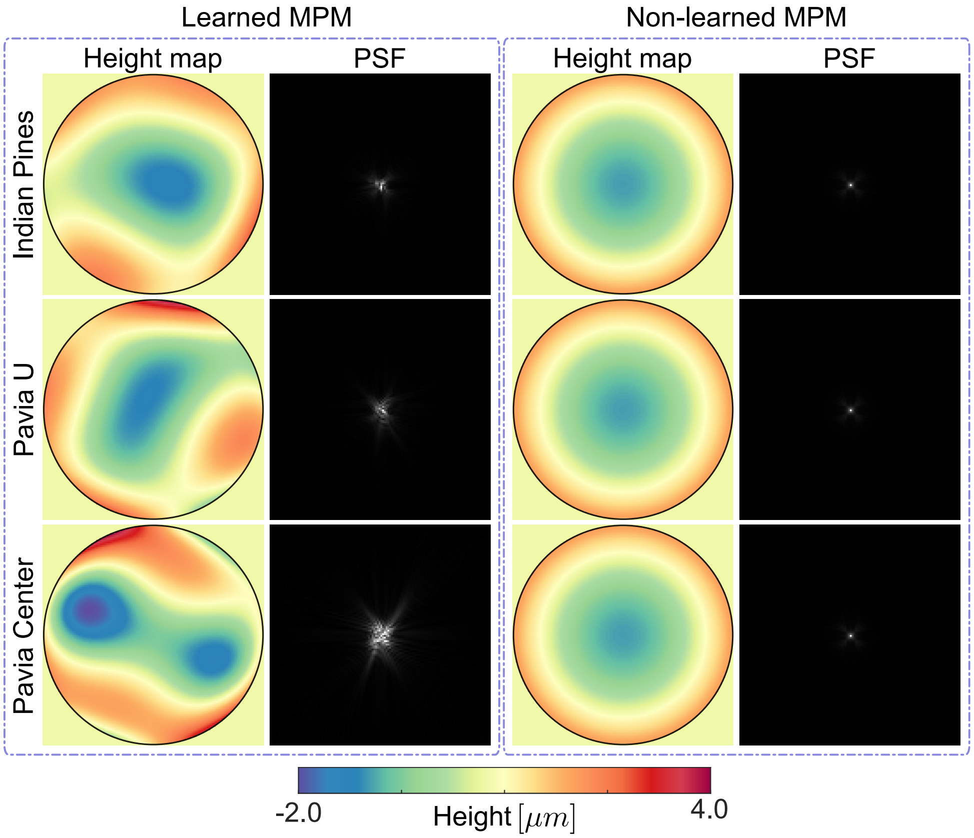
Based on these results, we fixed the patch size at $W = 8 \times 8$ and the number of iterations at $T = 2$ during the training stage of the proposed classification network for the experiments discussed in Sections 4.3.4 and 4.3.5.

Figure 35. Classification network analysis in terms of accuracy by varying the noise level, the number of iterations, and patch size at each spectral dataset. Rows: Patch size. Columns: Spectral dataset. A lighter color indicates a better spectral classification.



4.3.4. Multilevel phase mask analysis Figure 36 illustrates the learned and non-learned MPMs with the resulting PSFs for spectral classification across different spectral datasets using the proposed dual optical setup under a noiseless scenario. These MPMs were obtained by fixing $O = 15$ Zernike optical aberrations in the height map during the proposed E2E method training. It is important to note that the height map in the learned MPM encompasses different optical aberrations and adjusts depending on the spectral dataset. In contrast, the height map in the non-learned MPM was configured to consistently simulate the Fresnel lens across each spectral dataset, with the defocus coefficient set to one and the remaining coefficients set to zero.

Figure 36. The resulting learned and non-learned MPM, along with their corresponding PSF, by training the proposed E2E classification method over the dual optical setup across different spectral datasets under a noiseless scenario.

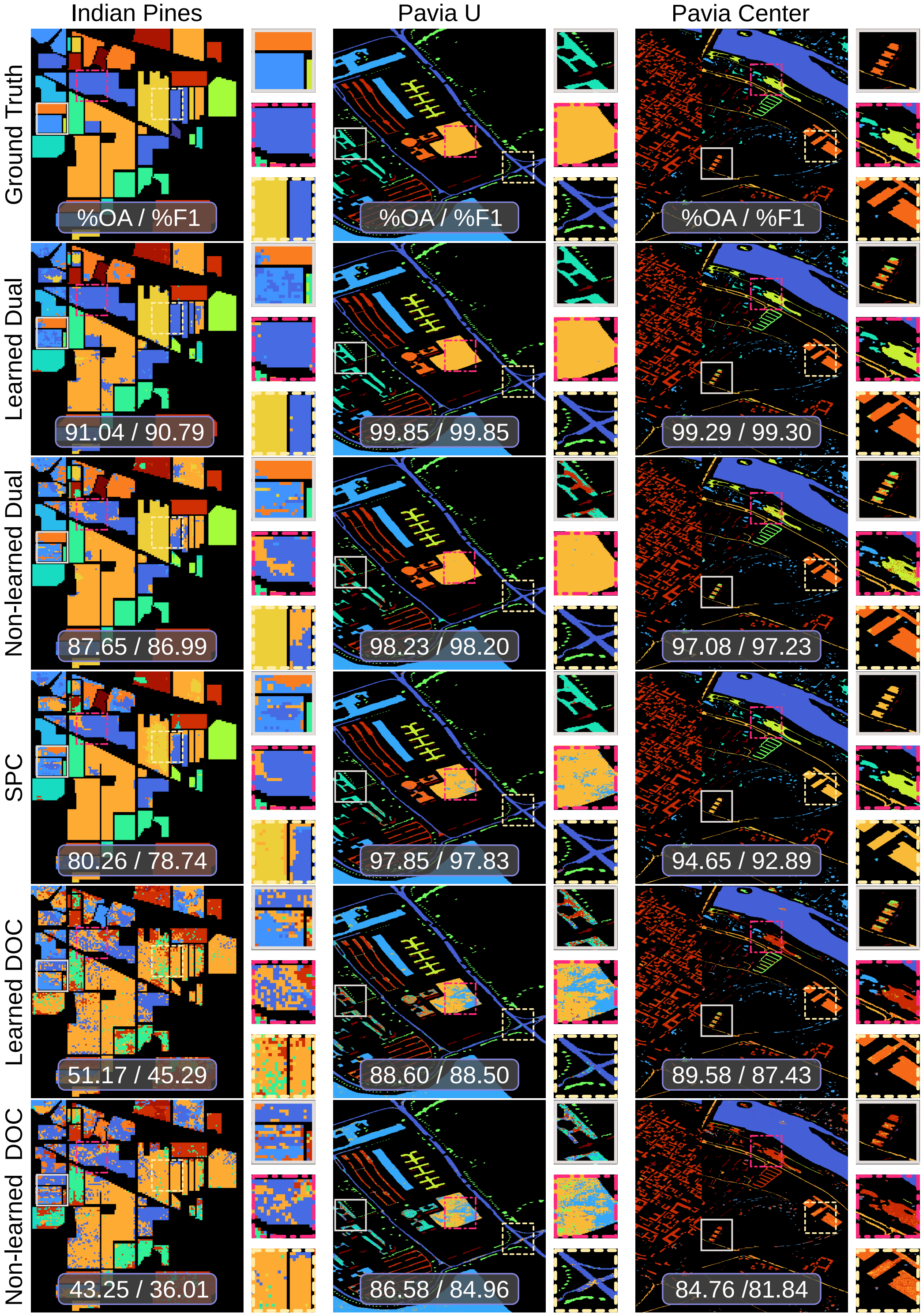


4.3.5. Optical setup impact on classification performance The proposed E2E classification method was assessed across various optical setups. Specifically, the proposed dual optical architecture was simulated using both learned and non-learned MPM. Furthermore, the spectral classification was evaluated separately using the SPC and the DOC, where the DOC involved both learned and non-learned MPM. Table 8 summarizes the classification quantitative results across the testing dataset with $\mathrm{SNR} = 30$ $[\mathrm{dB}]$. The best results are highlighted in the table with the green color and the second-best with yellow. The accuracy, precision, recall, and F1 score metrics were computed to evaluate the classification performance. Note that the proposed E2E method using the dual optical architecture with the learned MPM improves up to 3% in the evaluation metrics compared to the non-learned MPM. Moreover, Figure 37 shows the classification maps for the proposed spectral classification methodology using the dual optical setup by learning the MPM (Learned Dual) and non-learning the MPM (Non-learned Dual), the SPC, and the DOC by learning the MPM (Learned DOC) and non-learning the MPM (Non-learned DOC). The overall accuracy (OA) and F1 score were assessed across all spectral patches within each dataset with $\mathrm{SNR} = 30$ $[\mathrm{dB}]$. In particular, the proposed dual system with the learned MPM achieves better classification performance than the other configurations.

Table 8. Quantitative results for the classification task in terms of accuracy, precision, recall, and F1 score through the proposed E2E method using different optical setups with $\mathrm{SNR} = 30$ $[\mathrm{dB}]$. The best result is highlighted in green, and the second best is highlighted in yellow.

Dataset	Metric	Optical Setup				
		Dual		SPC	DOC	
		Learned	Non-learned		Learned	Non-learned
Indian Pines	Accuracy ↑	0.8839	0.8566	0.7834	0.4985	0.4234
	Precision ↑	0.8919	0.8550	0.7883	0.4509	0.3417
	Recall ↑	0.8839	0.8566	0.7834	0.4985	0.4234
	F1 score ↑	0.8833	0.8497	0.7640	0.4355	0.3481
Pavia U	Accuracy ↑	0.9974	0.9822	0.9778	0.8817	0.8476
	Precision ↑	0.9974	0.9824	0.9783	0.8793	0.8166
	Recall ↑	0.9974	0.9818	0.9778	0.8717	0.8476
	F1 score ↑	0.9974	0.9819	0.9776	0.8779	0.8303
Pavia Center	Accuracy ↑	0.9931	0.9713	0.9469	0.8896	0.8431
	Precision ↑	0.9934	0.9767	0.9193	0.8703	0.8250
	Recall ↑	0.9931	0.9713	0.9469	0.8896	0.8431
	F1 score ↑	0.9932	0.9629	0.9293	0.8684	0.8144

Figure 37. Classification maps for Indian Pines, Pavia U, and Pavia Center spectral datasets were obtained from the proposed E2E methodology using different optical setups with $\mathrm{SNR} = 30$ [dB].



4.4. EXPERIMENTAL SETUP

This section presents the implemented dual optical architecture, as illustrated in Fig. 38. This experimental setup mainly involves a halogen light source 3900e Illumination Technology with a power of $150\ [Watt]$; a CCD camera Stingray F-080B 1032×778 with a pixel size of $\Delta_{\text{CCD}} = 4.65\ [\mu m]$; a piezoelectric DM Thorlabs DMP40-P01 with an aperture of $10\ [mm]$; a DMD Texas Instruments D4120 1024×768 with a pixel size of $\Delta_{\text{DMD}} = 13.6\ [\mu m]$; and an Ocean Optics Flame S-VIS-NIR-ES spectrometer.

In the implemented optical testbed, the scene is illuminated by an incoherent light source, then, an objective lens guarantees the image formation in front of a lens with a focal distance of $f = 50\ [mm]$ (L1). The incoming light is 50:50 divided by a non-polarizing BS Thorlabs CCM1-BS013, resulting in two optical paths, each of which is composed of a lens with a focal distance of $f = 100\ [mm]$ (L2 and L3). Notice that the two optical paths (corresponding to the DOC and the SPC) are based on a $4f$ system. The first optical path incorporates a DM that emulates the phase mask, then, the modulated image is recorded by a CCD sensor. The second optical path employs a DMD that emulates the magnitude mask, then, a lens with a focal distance of $50\ [mm]$ (L4) ensures the light convergence over a condenser lens F220SMA-A (L5) connected through an optical fiber VIS-NIR with a core diameter of $1\ [mm]$ until a spectrometer. In this experiment, the scene is rotated by 45° to align with the DMD. The CCD sensor is also rotated by the same angle to obtain spatial measurements (from the DOC) with the same orientation as the spectral measurements (from the SPC). Finally, the spectral classification algorithm was tuned for labeling the spectral scene from the coded spatial-spectral measurements acquired through the implemented optical testbed.

Figure 38. Dual optical setup for spectral classification. Optical elements: Light source; OL, objective lens; L1, L2, L3, L4, and L5, lenses; DM, phase modulator; DMD, magnitude modulator; BS, beam splitter; CCD, registration camera; spectrometer; and optical fiber.

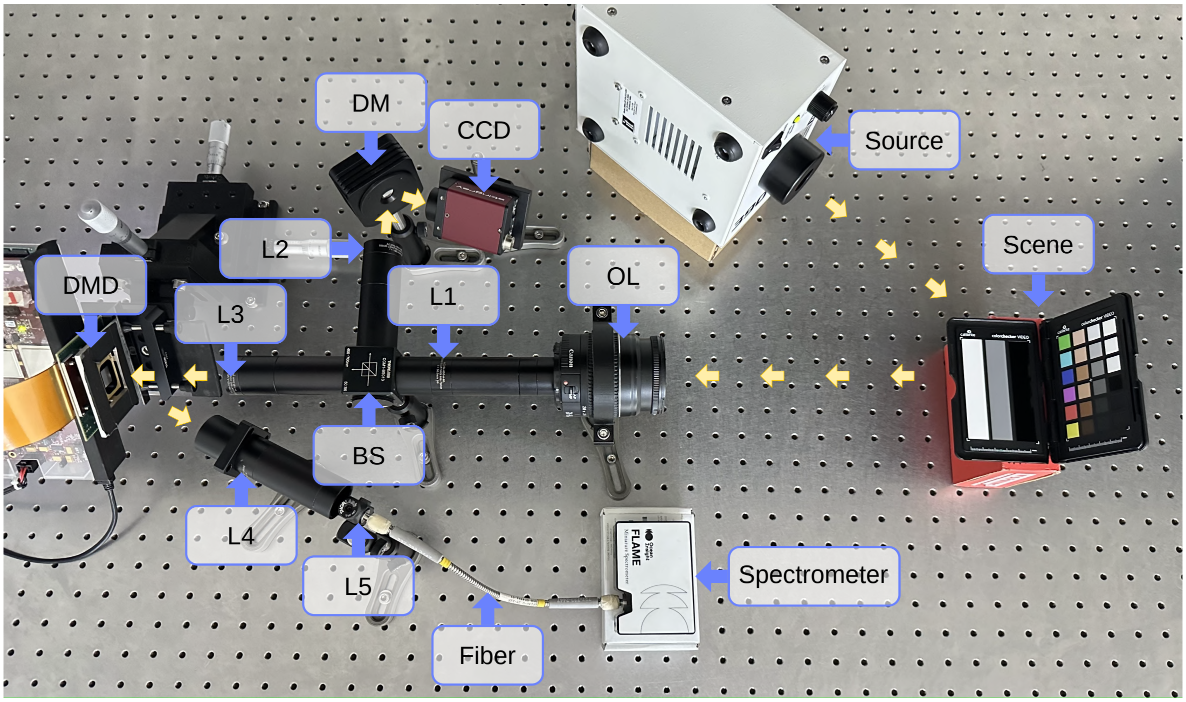
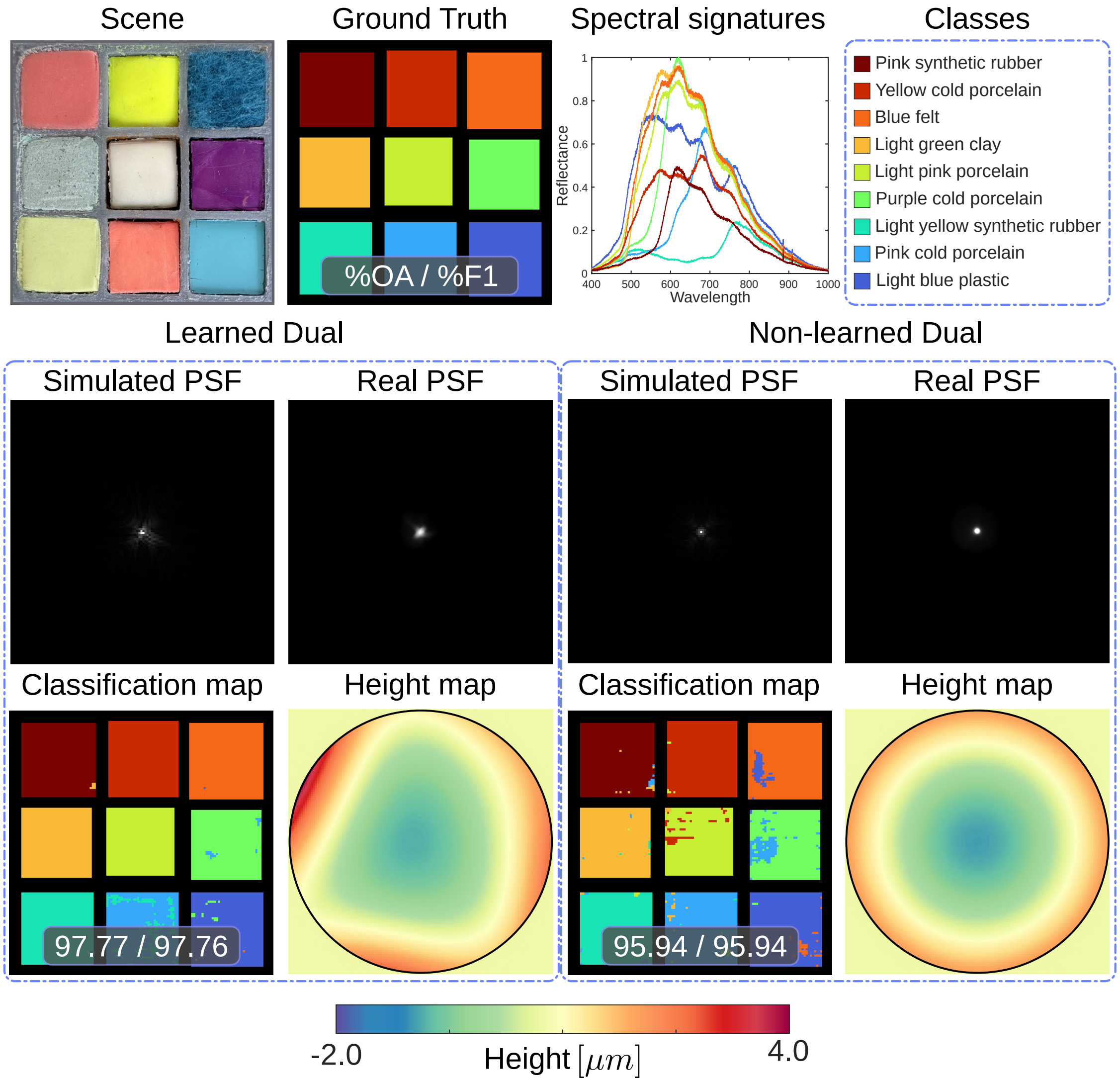


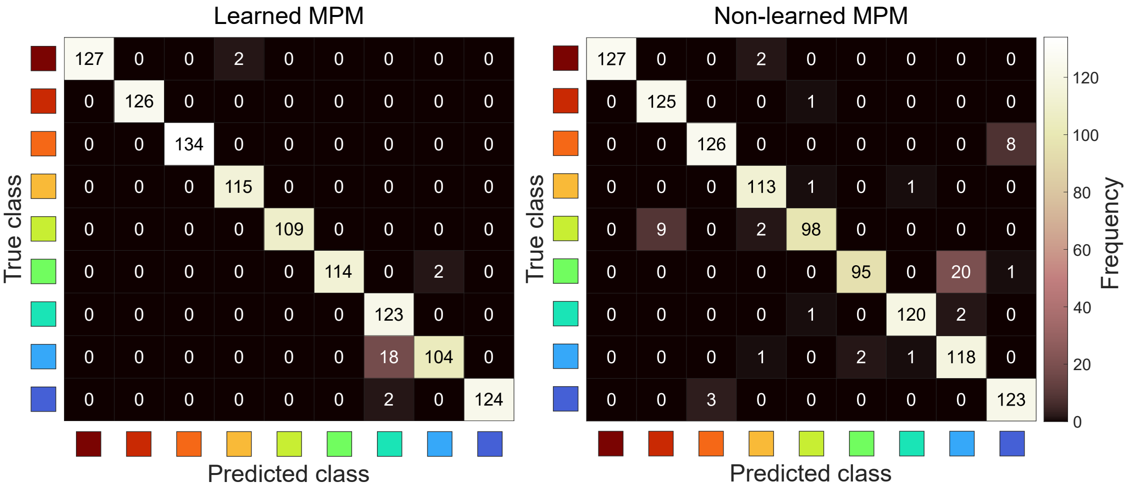
Figure 39 presents the classification results using the implemented dual optical testbed. The acquired dataset contains 9 different classes: pink synthetic rubber, yellow cold porcelain, blue felt, light green clay, light pink porcelain, light yellow synthetic rubber, pink cold porcelain, and light blue plastic. From this dataset, $10,997$ spectral patches with $K = 12$ bands and $W = 8 \times 8$ pixels were extracted, where $8,797$, $1,100$, and $1,100$ patches were selected for training, testing, and validating, respectively. Note that the proposed dual architecture with the learned MPM achieves a more accurate classification map than the non-learned MPM.

Figure 39. Classification results using the implemented dual optical setup. Row 1: RGB scene, classification map, spectral signatures for each material, and classes. Row 2: Simulated and real PSFs for the learned and non-learned MPM. Row 3: Classification results and height map for the learned and non-learned MPM.



Finally, Figure 40 presents the confusion matrix across the testing dataset for the proposed E2E spectral classification method based on the implemented optical architecture using the learned and non-learned MPM. The proposed dual system using the learned MPM obtains better performance in terms of the evaluated metrics compared to the non-learned MPM.

Figure 40. Confusion matrix for the dual optical setup using the learned and non-learned MPM. A lighter color on the main diagonal and a darker color off the diagonal indicate a better classification model.



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

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

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

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

4.5. CHAPTER CONCLUSION

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

"The future belongs to those who believe in the beauty of their dreams."
- ***Eleanor Roosevel***

5. CONCLUSIONS, DISCUSSION, AND FUTURE WORK

This dissertation demonstrates the feasibility of designing and implementing MPM strategies based on deep learning approaches for phase reconstruction and spectral classification tasks within DOI architectures. In this context, three works were conducted to showcase that designing the MPM structure in DOI systems from E2E optimization is achievable while boosting the recovery and classification tasks' performance. Throughout this doctoral study, three main challenges in DOI architectures have been extensively explored: the initialization method involved in the PR problem, the PR problem based on the MPM design through an interpretable DNN, and the spectral classification based on the MPM design from a dual optical architecture.

More precisely, the proposed E2E approaches consist of an encoder module that models the DOI architecture relying on MPMs and a decoder module that describes the implemented DNNs for initialization, recovery, and classification. Chapter 2 introduces the E2E learning initialization approach for coherent DOI at each diffraction field based on a filtering step, where the filter kernel involved in the approximation process and the DNN parameters presented in the recovery stage are simultaneously learned. Chapter 3 presents the E2E unrolling recovery approach for coherent DOI at the near-field based on a non-convex formulation, where the MPM weights in the optical encoder and DNN parameters in the interpretable network are jointly learned. Specifically, this phase recovery network includes a smoothness function to solve the non-convex problem. Chapter 4 describes the E2E spectral classification approach for incoherent DOI at the middle-field based on a dual diffractive spectral architecture, where the height map in the MPM and the DNN parameters in the spectral classification network are simultaneously learned. Particularly, the decoder stage performs the spectral image fusion from the plug-and-play ADMM algorithm. Last, two optical architectures were implemented to validate the proposed E2E methods

ods experimentally based on real-world measurements. The first assembled optical setup allows the CDPs acquisition using a coherent light that illuminates an LCOS. In contrast, the second assembled optical testbed allows the spatial-spectral measurement registration through an incoherent light that illuminates a DMD and a DM.

On the one hand, the proposed E2E strategies for approximating (Chapter 2) and recovering (Chapter 3) the optical field in DOI setups based on coherent light require a single snapshot and involve fewer iterations than traditional initialization methods and benchmark PR algorithms. On the other hand, the E2E proposed approach for classifying materials (Chapter 4) within spectral images involving a dual optical architecture based on incoherent light outperforms conventional optical setups.

Future work includes introducing different regularization terms for sensing matrix design in DOI architectures that promote different computational tasks' evaluation, such as segmentation, detection, superresolution, and denoising. It would also be interesting to analyze the effect of minor uncertainties affecting the sensing matrix in DOI systems on the computational tasks' performance. Furthermore, DOEs are an emerging technology that could alleviate MPM implementation in computational tasks, thereby compacting challenging optical systems. Currently, no Colombian companies are dedicated to manufacturing this type of technology; hence, there could be great potential for the country's development in designing and manufacturing diffractive lenses based on neural networks that provide solutions to various challenges across diverse economic sectors.

In retrospect, this dissertation successfully achieved the planned objectives and provided a foundation for potential further studies in the DOI area using both coherent and incoherent light. The methodologies discussed within this document attempt to provide a stepping stone for their application in several disciplines, including microscopy, holography, astronomy, crystallography, and spectroscopy, contributing to ongoing efforts to integrate deep learning into optical design.

BIBLIOGRAPHY

Arbelaez, Pablo et al. “Contour Detection and Hierarchical Image Segmentation”. In: *IEEE Trans. Pattern Anal. Mach. Intell.* 33.5 (May 2011), pp. 898–916. DOI: `10.1109/TPAMI.2010.161` (cit. on p. 91).

Arguello, Henry et al. “Shift-variant color-coded diffractive spectral imaging system”. In: *Optica* 8.11 (2021), pp. 1424–1434 (cit. on pp. 19, 54).

Bacca, Jorge, Laura Galvis, and Henry Arguello. “Coupled deep learning coded aperture design for compressive image classification”. In: *Optics Express* 28.6 (2020), pp. 8528–8540 (cit. on p. 47).

Bacca, Jorge, Tatiana Gelvez-Barrera, and Henry Arguello. “Deep coded aperture design: An end-to-end approach for computational imaging tasks”. In: *IEEE Transactions on Computational Imaging* 7 (2021), pp. 1148–1160 (cit. on pp. 19, 47, 48, 87).

Bacca, Jorge, Emmanuel Martinez, and Henry Arguello. “Computational spectral imaging: a contemporary overview”. In: *JOSA A* 40.4 (2023), pp. C115–C125 (cit. on p. 22).

Bacca, Jorge, Samuel Pinilla, and Henry Arguello. “Coded Aperture Design for Super-Resolution Phase Retrieval”. In: *2019 27th European Signal Processing Conference (EUSIPCO)*. IEEE. 2019, pp. 1–5 (cit. on pp. 82, 83).

Bacca, Jorge, Samuel Pinilla, and Henry Arguello. “Super-resolution phase retrieval from designed coded diffraction patterns”. In: *IEEE Transactions on Image Processing* 29 (2019), pp. 2598–2609 (cit. on pp. 15, 16).

Bacca, Jorge et al. “Super-Resolution Phase Retrieval Algorithm using a Smoothing Function”. In: *Mathematics in Imaging*. Optical Society of America. 2018, MW2D–3 (cit. on p. 85).

BaoShun, Shi and Lian QiuSheng. “DualPRNet: Deep shrinkage dual frame network for deep unrolled phase retrieval”. In: *IEEE Signal Processing Letters* 29 (2022), pp. 1177–1181 (cit. on pp. 19, 53, 95, 96).

Boominathan, Vivek et al. “PhlatCam: Designed phase-mask based thin lensless camera”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2020) (cit. on p. 55).

Born, Max and Emil Wolf. *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light*. Elsevier, 2013 (cit. on p. 108).

Brady, David J. *Optical imaging and spectroscopy*. John Wiley & Sons, 2009 (cit. on p. 39).

Cai, Zikui, Rakib Hyder, and M Salman Asif. “Data-driven illumination patterns for coded diffraction imaging”. In: *2021 IEEE International Conference on Image Processing (ICIP)*. IEEE. 2021, pp. 2818–2822 (cit. on pp. 19, 48).

Candes, Emmanuel J, Xiaodong Li, and Mahdi Soltanolkotabi. “Phase retrieval from coded diffraction patterns”. In: *Applied and Computational Harmonic Analysis* 39.2 (2015), pp. 277–299 (cit. on pp. 15, 16, 36).

Candes, Emmanuel J, Xiaodong Li, and Mahdi Soltanolkotabi. “Phase retrieval via Wirtinger flow: Theory and algorithms”. In: *IEEE Transactions on Information Theory* 61.4 (2015), pp. 1985–2007 (cit. on pp. 36, 39, 44, 81, 82).

Candes, Emmanuel J, Thomas Strohmer, and Vladislav Voroninski. “Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming”. In: *Communications on Pure and Applied Mathematics* 66.8 (2013), pp. 1241–1274 (cit. on pp. 17, 43, 82).

Chan, Stanley H, Xiran Wang, and Omar A Elgendy. “Plug-and-play ADMM for image restoration: Fixed-point convergence and applications”. In: *IEEE Transactions on Computational Imaging* 3.1 (2016), pp. 84–98 (cit. on p. 109).

Chandra, Rohan, Tom Goldstein, and Christoph Studer. “Phasepack: A phase retrieval library”. In: *2019 13th International conference on Sampling Theory and Applications (SampTA)*. IEEE. 2019, pp. 1–5 (cit. on p. 16).

Chen, Hang et al. “Diffractive deep neural networks at visible wavelengths”. In: *Engineering* 7.10 (2021), pp. 1483–1491 (cit. on p. 53).

Chen, Yushi et al. “Deep feature extraction and classification of hyperspectral images based on convolutional neural networks”. In: *IEEE transactions on geoscience and remote sensing* 54.10 (2016), pp. 6232–6251 (cit. on pp. 22, 104, 109).

Chen, Yuxin and Emmanuel Candes. “Solving random quadratic systems of equations is nearly as easy as solving linear systems”. In: *Advances in Neural Information Processing Systems*. 2015, pp. 739–747 (cit. on pp. 18, 44, 92, 95).

Connes, Pierre. “Astronomical fourier spectroscopy”. In: *Annual review of Astronomy and Astrophysics* 8 (1970), p. 209 (cit. on p. 15).

Deng, Li. “The mnist database of handwritten digit images for machine learning research [best of the web]”. In: *IEEE Signal Processing Magazine* 29.6 (2012), pp. 141–142 (cit. on p. 66).

Denk, Ondřej, Artem Musiienko, and Karel Žídek. “Differential single-pixel camera enabling low-cost microscopy in near-infrared spectral region”. In: *Optics express* 27.4 (2019), pp. 4562–4571 (cit. on p. 21).

Dong, Jonathan et al. “Phase retrieval: From computational imaging to machine learning: A tutorial”. In: *IEEE Signal Processing Magazine* 40.1 (2023), pp. 45–57 (cit. on p. 52).

Efron, Uzi. *Spatial light modulator technology: materials, devices, and applications*. Vol. 47. CRC press, 1994 (cit. on p. 16).

Eldar, Yonina C and Shahar Mendelson. “Phase retrieval: Stability and recovery guarantees”. In: *Applied and Computational Harmonic Analysis* 36.3 (2014), pp. 473–494 (cit. on p. 44).

Fienup, James R. “Phase retrieval algorithms: a comparison”. In: *Applied optics* 21.15 (1982), pp. 2758–2769 (cit. on p. 17).

Figueiredo, Mário AT, Robert D Nowak, and Stephen J Wright. “Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems”. In: *IEEE Journal of selected topics in signal processing* 1.4 (2007), pp. 586–597 (cit. on p. 47).

Fu, Chen et al. “Compressive spectral polarization imaging by a pixelized polarizer and colored patterned detector”. In: *JOSA A* 32.11 (2015), pp. 2178–2188 (cit. on p. 38).

Galvis, Laura et al. “Coded aperture design in compressive spectral imaging based on side information”. In: *Applied optics* 56.22 (2017), pp. 6332–6340 (cit. on p. 21).

Garcia, Hans, Claudia V Correa, and Henry Arguello. “Multi-resolution compressive spectral imaging reconstruction from single pixel measurements”. In: *IEEE Transactions on Image Processing* 27.12 (2018), pp. 6174–6184 (cit. on p. 21).

— “Optimized sensing matrix for single pixel multi-resolution compressive spectral imaging”. In: *IEEE Transactions on Image Processing* 29 (2020), pp. 4243–4253 (cit. on p. 106).

Gerchberg, Ralph W. “A practical algorithm for the determination of plane from image and diffraction pictures”. In: *Optik* 35.2 (1972), pp. 237–246 (cit. on p. 17).

— “Phase determination from image and diffraction plane pictures in the electron microscope”. In: *Optik* 34 (1971), pp. 275–284 (cit. on p. 17).

Goldstein, Tom and Christoph Studer. “Phasemax: Convex phase retrieval via basis pursuit”. In: *IEEE Transactions on Information Theory* 64.4 (2018), pp. 2675–2689 (cit. on pp. 17, 43).

Gonzalez, Rafael C. and Richard E. Woods. *Digital Image Processing (3rd Edition)*. USA: Prentice-Hall, Inc., 2006 (cit. on p. 64).

Goodman, Joseph W. *Introduction to Fourier optics*. Roberts and Company Publishers, 2005 (cit. on pp. 14, 35).

Greenberg, Joel A et al. “Coding and sampling for compressive x-ray diffraction tomography”. In: *Wavelets and Sparsity XV*. Vol. 8858. International Society for Optics and Photonics. 2013, p. 885813 (cit. on p. 36).

Gross, David, Felix Krahmer, and Richard Kueng. “Improved recovery guarantees for phase retrieval from coded diffraction patterns”. In: *Applied and Computational Harmonic Analysis* 42.1 (2017), pp. 37–64 (cit. on pp. 15, 33, 81, 82).

Guerrero, Andrés, Samuel Pinilla, and Henry Arguello. “Phase Recovery Guarantees From Designed Coded Diffraction Patterns in Optical Imaging”. In: *IEEE Transactions on Image Processing* 29 (2020), pp. 5687–5697 (cit. on pp. 15, 16, 35, 55, 58, 82, 83, 90).

Guzzi, Donatella et al. “An atmospheric correction iterative method for high spectral resolution aerospace imaging spectrometers”. In: *2009 IEEE International Geoscience and Remote Sensing Symposium*. Vol. 2. IEEE. 2009, pp. II–73 (cit. on p. 20).

Hess, HF et al. “Near-field spectroscopy of the quantum constituents of a luminescent system”. In: *Science* 264.5166 (1994), pp. 1740–1745 (cit. on p. 14).

Hinojosa, Carlos, Jorge Bacca, and Henry Arguello. “Coded aperture design for compressive spectral subspace clustering”. In: *IEEE Journal of Selected Topics in Signal Processing* 12.6 (2018), pp. 1589–1600 (cit. on p. 47).

Hinojosa, Carlos, Juan Carlos Niebles, and Henry Arguello. “Learning privacy-preserving optics for human pose estimation”. In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2021, pp. 2573–2582 (cit. on p. 107).

Hornbeck, Larry J. “Deformable-mirror spatial light modulators”. In: *Spatial Light Modulators and Applications III*. Vol. 1150. International Society for Optics and Photonics. 1990, pp. 86–103 (cit. on pp. 40, 41).

Huang, Longqian et al. “Spectral imaging with deep learning”. In: *Light: Science & Applications* 11.1 (2022), p. 61 (cit. on pp. 20, 21, 49, 53).

Imani, Maryam and Hassan Ghassemian. “An overview on spectral and spatial information fusion for hyperspectral image classification: Current trends and challenges”. In: *Information fusion* 59 (2020), pp. 59–83 (cit. on p. 22).

Jacome, Roman, Jorge Bacca, and Henry Arguello. “D 2 uf: Deep coded aperture design and unrolling algorithm for compressive spectral image fusion”. In: *IEEE Journal of Selected Topics in Signal Processing* (2022) (cit. on p. 109).

— “Deep-fusion: An end-to-end approach for compressive spectral image fusion”. In: *2021 IEEE International Conference on Image Processing (ICIP)*. IEEE. 2021, pp. 2903–2907 (cit. on pp. 18, 19).

Jahncke, CL, MA Paesler, and HD Hallen. “Raman imaging with near-field scanning optical microscopy”. In: *Applied physics letters* 67.17 (1995), pp. 2483–2485 (cit. on p. 14).

Jeon, Daniel S et al. “Compact snapshot hyperspectral imaging with diffracted rotation”. In: (2019) (cit. on pp. 20, 48).

Jerez, Andrés, Hans Garcia, and Henry Arguello. “Single pixel spectral image fusion with side information from a grayscale sensor”. In: *2018 IEEE 1st Colombian Conference on Applications in Computational Intelligence (ColCACI)*. IEEE. 2018, pp. 1–6 (cit. on p. 21).

Jerez, Andrés, Samuel Pinilla, and Henry Arguello. “Fast target detection via template matching in compressive phase retrieval”. In: *IEEE Transactions on Computational Imaging* 6 (2020), pp. 934–944 (cit. on pp. 60, 61, 68, 82, 90, 92).

Katkovnik, Vladimir and Karen Egiazarian. “Multi-frequency phase retrieval from noisy data”. In: *2018 26th European Signal Processing Conference (EUSIPCO)*. IEEE. 2018, pp. 2200–2204 (cit. on p. 95).

Katkovnik, Vladimir et al. “Computational super-resolution phase retrieval from multiple phase-coded diffraction patterns: simulation study and experiments”. In: *Optica* 4.7 (2017), pp. 786–794 (cit. on p. 99).

Kazemi, Samia, Bariscan Yonel, and Birsen Yazici. “Unrolled wirtinger flow with deep decoding priors for phaseless imaging”. In: *IEEE Transactions on Computational Imaging* 8 (2022), pp. 609–625 (cit. on p. 53).

Khan, Muhammad Jaleed et al. “Modern trends in hyperspectral image analysis: A review”. In: *Ieee Access* 6 (2018), pp. 14118–14129 (cit. on p. 20).

Kim, Kyung-Su and Sae-Young Chung. “Fourier phase retrieval with extended support estimation via deep neural network”. In: *IEEE Signal Processing Letters* 26.10 (2019), pp. 1506–1510 (cit. on pp. 18, 45).

Kingma, Diederik P and Jimmy Ba. “Adam: A method for stochastic optimization”. In: *arXiv preprint arXiv:1412.6980* (2014) (cit. on p. 89).

Konforti, Naim, Emanuel Marom, and S-T Wu. “Phase-only modulation with twisted nematic liquid-crystal spatial light modulators”. In: *Optics letters* 13.3 (1988), pp. 251–253 (cit. on p. 16).

Li, Lingen et al. “Quantization-aware deep optics for diffractive snapshot hyperspectral imaging”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, pp. 19780–19789 (cit. on pp. 46, 104).

Li, Yajun and Emil Wolf. “Three-dimensional intensity distribution near the focus in systems of different Fresnel numbers”. In: *JOSA A* 1.8 (1984), pp. 801–808 (cit. on p. 32).

Lohit, Suhas, Kuldeep Kulkarni, and Pavan Turaga. “Direct inference on compressive measurements using convolutional neural networks”. In: *2016 IEEE International Conference on Image Processing (ICIP)*. IEEE. 2016, pp. 1913–1917 (cit. on p. 47).

Lynn, Paul A and Wolfgang Fuerst. *Introductory digital signal processing with computer applications*. John Wiley & Sons, 1998 (cit. on pp. 61, 83).

MacCabe, Kenneth P et al. “Snapshot 2D tomography via coded aperture x-ray scatter imaging”. In: *Applied optics* 52.19 (2013), pp. 4582–4589 (cit. on p. 36).

Marquez, Miguel et al. “Snapshot compressive spectral depth imaging from coded aberrations”. In: *Optics Express* 29.6 (2021), pp. 8142–8159 (cit. on p. 40).

Mayo, Sheridan C et al. “X-ray phase-contrast microscopy and microtomography”. In: *Optics Express* 11.19 (2003), pp. 2289–2302 (cit. on p. 15).

McGloin, David et al. “Applications of spatial light modulators in atom optics”. In: *Optics Express* 11.2 (2003), pp. 158–166 (cit. on p. 17).

Mengu, Deniz and Aydogan Ozcan. “All-optical phase recovery: diffractive computing for quantitative phase imaging”. In: *Advanced Optical Materials* 10.15 (2022), p. 2200281 (cit. on p. 53).

Mengu, Deniz et al. “Analysis of diffractive optical neural networks and their integration with electronic neural networks”. In: *IEEE Journal of Selected Topics in Quantum Electronics* 26.1 (2019), pp. 1–14 (cit. on p. 55).

Metzler, Christopher, Phillip Schniter, Ashok Veeraraghavan, et al. “prDeep: robust phase retrieval with a flexible deep network”. In: *International Conference on Machine Learning*. PMLR. 2018, pp. 3501–3510 (cit. on pp. 18, 45, 83, 95).

Metzler, Christopher A, Arian Maleki, and Richard G Baraniuk. “BM3D-PRGAMP: Compressive phase retrieval based on BM3D denoising”. In: *2016 IEEE International Conference on Image Processing (ICIP)*. IEEE. 2016, pp. 2504–2508 (cit. on p. 95).

MiriRostami, SeyyedReza et al. “Hybrid diffractive optics (DOE & refractive lens) for broadband EDoF imaging”. In: *Electronic Imaging* 35 (2023), pp. 1–14 (cit. on p. 52).

Monga, Vishal, Yuelong Li, and Yonina C Eldar. “Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing”. In: *IEEE Signal Processing Magazine* 38.2 (2021), pp. 18–44 (cit. on p. 45).

Monroy, Brayan, Jorge Bacca, and Henry Arguello. “JR2net: a joint non-linear representation and recovery network for compressive spectral imaging”. In: *Applied Optics* 61.26 (2022), pp. 7757–7766 (cit. on p. 18).

Morales, David, Andrés Jerez, and Henry Arguello. “Deep Phase Retrieval by a Learnable Filtered Spectral Initialization”. In: *Computational Optical Sensing and Imaging*. Optical Society of America. 2021, CTh4A–6 (cit. on p. 59).

Naimipour, Naveed, Shahin Khobahi, and Mojtaba Soltanalian. “Upr: A model-driven architecture for deep phase retrieval”. In: *2020 54th Asilomar Conference on Signals, Systems, and Computers*. IEEE. 2020, pp. 205–209 (cit. on pp. 18, 19).

Nathan Silberman Derek Hoiem, Pushmeet Kohli and Rob Fergus. “Indoor Segmentation and Support Inference from RGBD Images”. In: *ECCV*. 2012 (cit. on p. 66).

Netrapalli, Praneeth, Prateek Jain, and Sujay Sanghavi. “Phase retrieval using alternating minimization”. In: *Advances in Neural Information Processing Systems* 26 (2013) (cit. on p. 19).

Ongie, Gregory et al. “Deep learning techniques for inverse problems in imaging”. In: *IEEE Journal on Selected Areas in Information Theory* 1.1 (2020), pp. 39–56 (cit. on p. 45).

Pineda, Jesus et al. “SPUD: simultaneous phase unwrapping and denoising algorithm for phase imaging”. In: *Applied optics* 59.13 (2020), pp. D81–D88 (cit. on p. 65).

Pinilla, Samuel, Jorge Bacca, and Henry Arguello. “Phase retrieval algorithm via nonconvex minimization using a smoothing function”. In: *IEEE Transactions on Signal Processing* 66.17 (2018), pp. 4574–4584 (cit. on p. 83).

Pinilla, Samuel, Juan Poveda, and Henry Arguello. “Coded diffraction system in X-ray crystallography using a boolean phase coded aperture approximation”. In: *Optics Communications* 410 (2018), pp. 707–716 (cit. on pp. 15, 36, 37).

Pinilla, Samuel et al. “A Smoothing Stochastic Phase Retrieval Algorithm for Solving Random Quadratic Systems”. In: *2018 IEEE Statistical Signal Processing Workshop (SSP)*. IEEE. 2018, pp. 278–282 (cit. on p. 82).

Pinilla, Samuel et al. “Coded aperture design for solving the phase retrieval problem in X-ray crystallography”. In: *Journal of Comput. and Applied Mathematics* 338 (2018), pp. 111–128 (cit. on pp. 16, 36, 37).

Pinilla, Samuel et al. “Unfolding-aided bootstrapped phase retrieval in optical imaging: Explainable AI reveals new imaging frontiers”. In: *IEEE Signal Processing Magazine* 40.2 (2023), pp. 46–60 (cit. on pp. 18, 45, 48, 51).

Poon, Ting-Chung and Jung-Ping Liu. *Introduction to modern digital holography: with MATLAB*. Cambridge University Press, 2014 (cit. on pp. 14, 32, 33, 35, 58, 80, 81).

Ramirez, Jhon, Henry Arguello, and Jorge Bacca. “Phase unwrapping for phase imaging using the plug-and-play proximal algorithm”. In: *Applied Optics* 63.2 (2024), pp. 535–542 (cit. on p. 51).

Ren, Yu-Xuan, Rong-De Lu, and Lei Gong. “Tailoring light with a digital micromirror device”. In: *Annalen der physik* 527.7-8 (2015), pp. 447–470 (cit. on pp. 40, 42).

Ripoll, Olivier, Ville Kettunen, and Hans Peter Herzig. “Review of iterative Fourier-transform algorithms for beam shaping applications”. In: *Optical Engineering* 43.11 (2004), pp. 2549–2548 (cit. on pp. 40, 42).

Rodriguez, J Bacca, Gonzalo R Arce, and Daniel Leo Lau. “Blue-noise multitone dithering”. In: *IEEE Transactions on Image Processing* 17.8 (2008), pp. 1368–1382 (cit. on p. 38).

Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. “U-net: Convolutional networks for biomedical image segmentation”. In: *International Conference on Medical image computing and computer-assisted intervention*. Springer. 2015, pp. 234–241 (cit. on pp. 60, 62, 95).

Sao, Mayu et al. “Lensless close-up imaging with Fresnel zone aperture”. In: *Japanese Journal of Applied Physics* 57.9S1 (2018), 09SB05 (cit. on p. 14).

Shechtman, Yoav et al. “Phase retrieval with application to optical imaging: a contemporary overview”. In: *IEEE signal processing magazine* 32.3 (2015), pp. 87–109 (cit. on pp. 14, 15, 33).

Shimoni, Michal, Rob Haelterman, and Christiaan Perneel. “Hypersectral imaging for military and security applications: Combining myriad processing and sensing techniques”. In: *IEEE Geoscience and Remote Sensing Magazine* 7.2 (2019), pp. 101–117 (cit. on p. 20).

Sitzmann, Vincent et al. “End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging”. In: *ACM Transactions on Graphics (TOG)* 37.4 (2018), pp. 1–13 (cit. on pp. 106, 108).

Stratton, Julius Adams. *Electromagnetic theory*. Vol. 33. John Wiley & Sons, 2007 (cit. on p. 30).

Stratton, Julius Adams and LJ Chu. “Diffraction theory of electromagnetic waves”. In: *Physical Review* 56.1 (1939), p. 99 (cit. on p. 31).

Tao, Xiao et al. “Phase-coded speckle illumination for laser Fourier ptychographic microscopy”. In: *Optics Communications* 498 (2021), p. 127199 (cit. on p. 54).

Tian, Yaocheng and James R Fienup. “Phase retrieval with only a nonnegativity constraint”. In: *Optics Letters* 48.1 (2023), pp. 135–138 (cit. on p. 52).

Tillmann, Andreas M, Yonina C Eldar, and Julien Mairal. “DOLPHIn—dictionary learning for phase retrieval”. In: *IEEE Transactions on Signal Processing* 64.24 (2016), pp. 6485–6500 (cit. on p. 95).

Tran, Minh H and Baowei Fei. “Compact and ultracompact spectral imagers: technology and applications in biomedical imaging”. In: *Journal of biomedical optics* 28.4 (2023), pp. 040901–040901 (cit. on p. 20).

Vargas, Edwin, Henry Arguello, and Jean-Yves Tourneret. “Spectral image fusion from compressive measurements using spectral unmixing and a sparse representation of abundance maps”. In: *IEEE Transactions on Geoscience and Remote Sensing* 57.7 (2019), pp. 5043–5053 (cit. on p. 108).

Vial, Pierre-Hugo et al. “Learning the proximity operator in unfolded admm for phase retrieval”. In: *IEEE Signal Processing Letters* 29 (2022), pp. 1619–1623 (cit. on p. 19).

Wang, Chang-Jen et al. “Phase retrieval with learning unfolded expectation consistent signal recovery algorithm”. In: *IEEE Signal Processing Letters* 27 (2020), pp. 780–784 (cit. on pp. 19, 48).

Wang, Gang, Georgios B Giannakis, and Yonina C Eldar. “Solving systems of random quadratic equations via truncated amplitude flow”. In: *IEEE Transactions on Information Theory* 64.2 (2018), pp. 773–794 (cit. on pp. 17, 44, 61, 64, 68, 74, 83, 92, 95).

Wang, Gang et al. “Phase retrieval via reweighted amplitude flow”. In: *IEEE Transactions on Signal Processing* 66.11 (2018), pp. 2818–2833 (cit. on pp. 17, 43, 64, 68, 74, 83, 92, 95).

Wang, Kaiqiang et al. “On the use of deep learning for phase recovery”. In: *Light: Science & Applications* 13.1 (2024), p. 4 (cit. on p. 51).

Weng, Qihao. “Thermal infrared remote sensing for urban climate and environmental studies: Methods, applications, and trends”. In: *ISPRS Journal of photogrammetry and remote sensing* 64.4 (2009), pp. 335–344 (cit. on p. 20).

Wu, Lin. “Phase smoothing for diffractive deep neural networks”. In: *Optics Communications* (2024), p. 130267 (cit. on p. 50).

Xiang, Mingjun et al. “Amplitude/phase retrieval for terahertz holography with supervised and unsupervised physics-informed deep learning”. In: *IEEE Transactions on Terahertz Science and Technology* (2024) (cit. on p. 51).

Xiao, Han, Kashif Rasul, and Roland Vollgraf. “Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms”. In: *arXiv preprint arXiv:1708.07747* (2017) (cit. on pp. 66, 91).

Ye, Qiuliang, Li-Wen Wang, and Daniel PK Lun. “SiSPRNet: end-to-end learning for single-shot phase retrieval”. In: *Optics Express* 30.18 (2022), pp. 31937–31958 (cit. on p. 48).

Zhang, Cheng et al. “Convex phase retrieval with multiple structured illuminations design”. In: *Journal of Optics* (2020), pp. 1–6 (cit. on p. 55).

Zhang, Feilong et al. “Physics-based Iterative Projection Complex Neural Network for Phase Retrieval in Lensless Microscopy Imaging”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, pp. 10523–10531 (cit. on pp. 18, 19, 45).

Zhang, Fucai et al. “Phase retrieval by coherent modulation imaging”. In: *Nature communications* 7.1 (2016), p. 13367 (cit. on p. 14).

Zhang, Huishuai and Yingbin Liang. “Reshaped wirtinger flow for solving quadratic system of equations”. In: *Advances in Neural Information Processing Systems*. 2016, pp. 2622–2630 (cit. on p. 36).

Zhang, Kai et al. “Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising”. In: *IEEE transactions on image processing* 26.7 (2017), pp. 3142–3155 (cit. on p. 91).

Zhang, Yuhe et al. “PhaseGAN: a deep-learning phase-retrieval approach for unpaired datasets”. In: *Optics express* 29.13 (2021), pp. 19593–19604 (cit. on p. 54).

Zhang, Zichen, Zheng You, and Daping Chu. “Fundamentals of phase-only liquid crystal on silicon (LCOS) devices”. In: *Light: Science & Applications* 3.10 (2014), e213–e213 (cit. on pp. 40, 41).