Unfolding-Aided Bootstrapped Phase Retrieval in Optical Imaging

*Explainable AI reveals new imaging frontiers*

Phase retrieval in optical imaging refers to the recovery of a complex signal from phaseless data acquired in the form of its diffraction patterns. These patterns are acquired through a system with a coherent light source that employs a diffractive optical element (DOE) to modulate the scene, resulting in coded diffraction patterns (CDPs) at the sensor. Recently, the hybrid approach of a model-driven network or deep unfolding has emerged as an effective alternative to conventional model- and learning-based phase-retrieval techniques because it allows for bounding the complexity of algorithms while also retaining their efficacy. Additionally, such hybrid approaches have shown promise in improving the design of DOEs that follow theoretical uniqueness conditions. There are opportunities to exploit novel experimental setups and resolve even more complex DOE phase-retrieval applications. This article presents an overview of algorithms and applications of deep unfolding for bootstrapped—regardless whether near, middle, or far zones—phase retrieval.

**Introduction**

Phase retrieval in diffractive optical imaging (DOI) deals with the estimation of a complex domain signal from magnitude-only data in the form of its diffraction patterns [1]. This problem has been at the center of exciting developments in DOI [2], [3], [4] during the last decade to improve achievable spatial resolution [5], [6], [7], lighter optical setups [6], and better detection [2]. When the diffraction patterns (phaseless data) are acquired through a setup that employs coherent light and a DOE (also as a coded aperture) to modulate the scene, phase CDPs are obtained. Typically, this setup allows for acquiring several snapshots of the scene by changing the spatial configuration of the DOE. The inclusion of the DOE in the phase-retrieval setting paves the way for uniqueness guarantees for all 2D signals up to a unimodular constant [3], [8], [9], overcoming the classical result in [10] that holds for only nonreducible signals.

The CDP is experimentally acquired in three diffraction zones: near, middle, and far [3], [11]. Mathematically, assuming a diagonal matrix $D_\ell \in \mathbb{C}^{n \times n}$ modeling the DOE for the $\ell$-th snapshot, $\ell = 1, \ldots, L$, the CDP consists of quadratic
equations of the form $y_{k,i} = a_{k,i}^H D_i x$ where $a_{k,i} \in \mathbb{C}^n$ are the known wavefront propagation vectors associated with the $k$th diffraction for $i = 1,\ldots,n$. $x \in \mathbb{C}^n$ is the unknown scene of interest; and $k = 1, 2, 3$ indexes the near, middle, and far zones, respectively. By harnessing specific properties of each diffraction zone, several advances in imaging applications have been made. The near zone is exploited in scanning near-field optical microscopy [12], Raman imaging [13], and spectroscopy [14]. Applications such as Fresnel holography [11] and lensless imaging [15] take advantage of the middle diffraction zone to develop new acquisition imaging devices [15], and optical elements such as Fresnel lenses [16]. The far or Fraunhofer diffraction zone has been instrumental in the development of applications such as crystallography, astronomical imaging, and microscopy [17].

Motivated by the aforementioned applications and guided by the analytical models of the physical setups for a given diffraction zone, [3] has showed that the DOE and the phase-retrieval algorithm need to be jointly designed to guarantee unique recovery; these theoretical results were not possible a decade ago. The entries of each matrix $D_i$ (a DOE for the $k$th snapshot) were assumed to be independent identically distributed from a discrete random variable $d$ obeying $|d| \leq 1$. Interestingly, [3] also shows that the design criteria for DOEs is independent of the diffraction zone. This result provides a bootstrapped (regardless of the diffraction zone) phase-retrieval in optical imaging. Table 1 summarizes common DOE phase-retrieval applications. The role of DOEs in recovering the phase across all diffraction zones has been studied in [2], [7], [18], [19], [20], and [21], where either the DOE was modeled by specific measurement setups or machine learning (ML) was deployed to obtain data-driven model-free recovery. Although very flexible and powerful, model-based methods also yield significant reconstruction artefacts arising from imperfect system models, mismatch, calibration errors, hand-tuned parameters, and hand-picked regularizers. Furthermore, these methods take hundreds to thousands of iterations to converge, which is unacceptable for real-time imaging. These shortcomings are overcome by employing deep neural networks, whose expressive power allows training such that the resulting network acts as an estimator of the true signal. For example, [22] proposed a deep neural network for object classification from CDPs, achieving high stability in the classification performance when all CDPs are not employed. However, neural networks are treated as black boxes, leading to limited interpretability and, therefore, inadequate control. Further, the training times could be prohibitively longer.

In this context, a hybrid model-based and data-driven technique has the potential to further improve data acquisition by effectively mitigating the adverse effects of both techniques [23], [24]. This so-called deep unfolding technique effectively facilitates the design of model-aware deep architectures based on well-established iterative signal processing techniques [25], [26], [27]. Unfolding exploits the data to enhance accuracy and performance so that the resulting interpretability leads to trusted outcomes, requiring fewer data, and a faster convergence rate. Another related work in [26] viewed the model-based iterative shrinkage thresholding algorithm as a recurrent neural network, where its activation functions are the shrinkage operator. In the context of inverse problems in imaging, [27] provides further insights into the theory and implementations of unrolling. The related terms include unrolling [28] and unrectifying [29]. Whereas the former is used for the algorithms employed by unfolding networks (or simply synonymously with unfolding), the latter refers to the process of transforming nonlinear activations in neural networks into data-dependent linear equations and constraints. Very recently, unfolded networks were proposed for both conventional and sparse phase-retrieval problems [30]. In this article, we present a bootstrapped overview of deep unfolding methods for imaging tasks in a CDP phase-retrieval setting to improve the design of DOEs that can follow theoretical conditions to meet uniqueness. We also summarize the use of designed DOEs together with their unfolded phase-retrieval (UPR) algorithms in terms of robustness against noise [31], with/without priors on the scene [30], and smoothness [32]. Finally, we review progress in application areas that enable experimental setups to implement these techniques to address even more ill-posed problems.

Optimal DOE design

The design principles [3] of DOEs require that the set of matrices $D_i$ satisfies $\mathbb{E} \left[ \sum_{l=1}^{L} D_i D_i^H \right] = r I$, $i = 1,\ldots,L$ to guarantee uniqueness in recovery for all 2D scenes with high probability (regardless of the zone), with $0 < r \leq L$ where $L \geq c_0 n$ for some large constant $c_0 > 0$, with $I$ as the identity matrix, and $\mathbb{E} \left[ \cdot \right]$ is the statistical expectation. The implication is that the variance of the entries of matrices $D_i$ needs to be designed along all snapshots to guarantee unique recovery. More information is collected from the scene when this variance is maximized along the snapshots [3], [34]. In the following, we summarize the design principles for two common scenarios: without any prior assumption on the scene and when the scene is known to be sparse in some basis. As explained later, unfolding is applicable to both scenarios.

Design without priors on the scene

This is accomplished by either by targeting the design of the DOEs in the image-registration process or obtaining the sensing matrix $A_k$ to acquire CDPs at the $k$th diffraction zone co-designed jointly with phase-retrieval algorithm. It is the latter where unfolding is commonly employed (see “A Comparison Between Diffraction Optical Element Design Methodologies”).

Greedy DOE methodology

For ease of exposition, consider $C^N \in \mathbb{C}^{N \times N}$ to be the 2D version of the diagonal matrix $D_1$, whose entries are $(D_1)_{q,q} = (C^N)_{q-q_{N+1}}$ for $v = \lfloor q - 1/N \rfloor$, $q = 1,\ldots,n$ and $n = N^2$. The design principles are based on considering an admissible random variable $d = \{e_1, e_2, e_3, e_4\}$ with probability $\{1/4, 1/4, 1/4, 1/4\}$, respectively, assuming that $L = c_0 b$ for some integer $b > 0$ where $C_{n} = 4$ is the number of coding elements.
In particular, the DOE design takes temporal correlation and spatial separation into account. For temporal correlation, the condition on the 2D version $C^t$ of $D_y$ can be satisfied if matrices $C^t$ are constrained to have all the coding elements of $d$ along the $L$ projections at each particular spatial position of the ensemble [34] to directly maximize the variance of the entries along snapshots. It means that $\{(C^t)_{s,u}|t=1,\ldots,L\}$ contains $b$ times each possible value of $d$ for any $s,u \in \{1,\ldots,N\}$. For spatial separation, it has been shown that a set of DOEs with an equispaced distribution of the coding elements improves the ability of data $y_{k,\ell}$ to uniquely identify the scene $x$ and increases image reconstruction quality [34], [35].

| Application and Setup | Description |
|-----------------------|-------------|
| Phase Imaging         | Reconstruction of the 3D shape of an object via phase retrieval. Far-zone scenarios consist of estimating the optical phase of the object by low-pass filtering the leading eigenvector of a carefully constructed matrix [17]. |
| Source                | The optical phase used to detect objects within a scene; a near-zone CDP for rapid detection using cross-correlation analysis to detect the target using its optical phase as a discriminant [2]. |
| Coherent Light        | Reconstruction of the object wavefront; lensless, single-shot phase retrieval for pixel superresolution phase imaging in the middle zone [6]. Noise is suppressed by a combination of sparse and deep learning-based filters. Single-shot allows for recording of dynamic scenes (frame rate limited only by the camera). |
| 3D Object             | Computational imaging with DOEs. Phase retrieval appears in the design of the DOEs, instead of the wavefront propagation model; an effectively designed DOE for all-in-focus intensity imaging in the middle zone [33]. |
| DOE                  | |
| Far Zone             | |
| Phase Retrieval      | |
| Unwrapped Phase      | |
| Object Detection     | |
| Source               | |
| Coherent Light       | |
| 3D Object            | |
| DOE                  | |
| Near Zone            | |
| Fast Phase Retrieval | |
| Correlation Analysis | |
| Decision Process     | |
| Microscopy           | |
| Source               | |
| Coherent Light       | |
| Sample               | |
| DOE Sensor           | |
| CDP                  | |
| Phase Retrieval      | |
| Intensity Lensless Imaging | |
| Scene                | |
| Incoherent Light     | |
| Middle Zone          | |
| $d_1$                | |
| $d_2$                | |

Table 1. Common DOE-based optical imaging applications.
The overall design strategy comprises the following steps:

1) **Step 1: Split into several cells.** The matrix $C^i$ is divided into squared cells (the green squares highlighted in Figure S1). The positions of all $e_i$ of $d$ in the first cell of $C^i$ are chosen uniformly at random as $(C^i)_{s,u} \sim \mathcal{U} \times [(e_1, \ldots, e_d) - ((C^i)_{s,u}, \ldots, (C^{i-1})_{s,u})]$ for $s, u$ in the cell. Note that set subtraction $(e_1, \ldots, e_d) - ((C^i)_{s,u}, \ldots, (C^{i-1})_{s,u})$ guarantees that a different coding element of $d$ is chosen for each of the first $\ell - 1$ coded apertures.

2) **Step 2: Determine the coding element location.** Move to the cell on the right to determine at random a proper position for each coding element. These positions are determined by maximizing the Manhattan distance between pixels $(C^i)_{s_1,u_1}$ and $(C^i)_{s_2,u_2}$ in the cell on the left and right, respectively. The positions maximizing this distance for the $e_j$ element are one of the two highlighted yellow squares.

3) **Step 3: Maximal spatial separation.** Define $\Omega_{e_i}$ to be the set of distances between the pixel, with value $e_i$ in the first cell and the positions $(s_2, u_2)$ of the next right cell. The positions $\mathcal{R}_{e_i}$ of maximal distance (as in the previous step) are obtained as $\mathcal{R}_{e_i} = \arg\max \Omega_{e_i}$. Define the set $\mathcal{B}_{e_i} = \{(s, u) \in \mathcal{R}_{e_i} | (C^i)_{s,u} = e_i, \forall 1 \leq k \leq \ell - 1\}$. From $\mathcal{F}_{e_i} = \mathcal{R}_{e_i} - \mathcal{B}_{e_i}$, we can randomly determine the position of the coding element $e_i$ in the next cell. The positions in $\mathcal{F}_{e_i}$ maximize the distance between pixels and guarantee choosing a different coding element than the first $\ell - 1$ projections. For $e_2$, it would be the blue square highlighted in the next right cell.

4) **Step 4: Optimize across the $\ell$-th DOE.** Steps 2 and 3 are repeated for all coding elements $e_i$ of the admissible variable $d$. Thus, the spatial distribution of the $\ell$-th DOE is optimized.

5) **Step 5: Optimize across all DOEs.** The set of DOEs maximizes the variances along snapshots until $\ell = C_d$. If $\ell = C_d + 1$, a new DOE dedicated to optimizing the spatial distribution must be generated. Thus, from $\ell = C_d + 2$ to $\ell = 2C_d$, the temporal and spatial correlation must be exploited considering steps 1–4 until the $L$ projections are completed.

---

**A Comparison Between Diffractive Optical Element Design Methodologies**

Greedy methods design the diffractive optical element itself (instead of the sensing matrix). The experimental performance of this strategy for phase retrieval is typically not as expected because of imperfect system modeling; reconstruction artefacts such as mismatch, calibration errors, and hand-tuned parameters; and even poorer performance for a single snapshot.

In unfolding-based techniques, the sensing matrix is designed across all diffraction zones. The performance of the unfolding design method overcomes the greedy methodology because the sensing matrix and the phase-retrieval algorithm are jointly designed; and the reconstruction algorithm is adjusted for the single-snapshot extreme case.

![FIGURE S1. A diffractive optical element design strategy using admissible random variable $d = \{e_i, e_2, e_3, e_4\}$ with probability $d = \{1/4, 1/4, 1/4, 1/4\}$, respectively, and $C_2 = 4$.](image)

![FIGURE S2. A high-level overview of the unfolding method, resulting in the network layers $f^1, \ldots, f^\ell$. The first layer $f^1(x; \Theta_1) = |A_1x|$ to model the coded diffraction pattern for the $1$st diffraction zone. Each iteration for phase retrieval is transferred into network parameters $\Theta^1, \ldots, \Theta^\ell$. These parameters are learned from training data sets through end-to-end training.](image)
End-to-end unfolding co-design of sensing matrix

The most important advantage of unfolding co-design is the ability to simultaneously design the entire or end-to-end sensing process and the phase-retrieval algorithm. Define the global noiseless measurement vector for the $k$th diffraction zone as $y_k = [y_{k,1}, \ldots, y_{k,n}]^T \in \mathbb{R}^{m \times n}$ and consider the set of sensing matrices $\{A_k\}_{k=1}^K$ with

$$
A_1 = [\tilde{D}_1 F T F^H, \ldots, \tilde{D}_L F T F^H]^H \quad \text{(near zone)}
$$

$$
A_2 = [\tilde{D}_1 Q F, \ldots, \tilde{D}_L Q F]^H \quad \text{(middle zone)}
$$

$$
A_3 = [\tilde{D}_1 F, \ldots, \tilde{D}_L F]^H \quad \text{ext (far zone)} \quad (1)
$$

where $F \in \mathbb{C}^{n \times n}$ is the discrete Fourier transform matrix, and $T \in \mathbb{C}^{n \times n}$ ($Q \in \mathbb{C}^{n \times n}$) is the auxiliary orthogonal diagonal matrix that depends on the propagation distance (wavelength of the coherent source) to model the near (middle) zone (see [11, Ch. 4] for further details). The sensing process for the $k$th diffraction zone is $y_k = |A_k x|$ which corresponds to the first layer $f^i(x; A_i) = |A_k x|$ of the deep network (Figure S2) to model the CDP for the $k$th diffraction zone.

We seek to recover the signal of interest $x$ from the embedded phaseless data $y_k$ given the knowledge of $A_k$ by solving the following optimization problem:

$$
\mathcal{P}_0: \text{Find } x \text{ subject to } y_k = |A_k x| \quad (2)
$$

where the nonconvex constraint models the measurement process. Additionally, if $x^i$ is a feasible point of the problem $\mathcal{P}_0$ then $e^{-j\varphi} x^i$ is also a solution to the problem for an arbitrary phase constant $\varphi$. Hence, it is only possible to recover the underlying signal of interest up to a constant global phase $\varphi \in \mathbb{R}$. This ambiguity arising from the global phase naturally leads to the following meaningful quantifying metric of closeness of the recovered signal $x^i$ to the true signal $x$:

$$
\text{dist}(x, x^i) = \min_{\varphi \in [0, 2\pi]} \| x - e^{-j\varphi} x^i \|_2. \quad (3)
$$

If $\text{dist}(x, x^i) = 0$, then $x$ and $x^i$ are equal up to some global phase.

The end-to-end unfolding co-design seeks to jointly design the sensing matrix (the encoder module or first layer in Figure S2) and the reconstruction algorithm (decoder module from layers 2 to $T$), in contrast to the existing methodologies that consider either the development of the reconstruction algorithm or the design of only DOEs. As a result, it is a unification of both approaches. Denote the class of decoder functions parameterized on a set of parameters as $\Phi$ by $\mathcal{S}_\Phi: \mathbb{R}^n \rightarrow \mathbb{R}^n$. Assume that for a given measurement matrix $A_k$ and the corresponding measurement vector $y_k$, an estimation $\hat{x}$ of the signal of interest is provided by a certain characterization of the decoder function, i.e.,

$$
\hat{x} = \mathcal{S}_\theta(y_k; A_k) \equiv \mathcal{S}_\theta(f^i(x; A_k); A_k) \quad (4)
$$

where $\theta \in \Phi$ denotes a characterization of the decoder module from the original class. In addition, as $A_k$ models a physical image formation model, it is desired to define the set of feasible sensing matrices for the optical setup as $\mathcal{T} \subset \mathbb{C}^{m \times n}$. This set contains the physical constraints related to implementation of the DOE encapsulated in matrices $D_k$ as in (1) (see the “Physical Constraints” section for examples of these constraints). Consequently, this means that a realistic $A_k$ depends on the feasibility of the DOE (see, for instance, Figure 1 for a practical implementation). Then, for a fixed characterization of the decoder function $\mathcal{S}_\theta \in \Phi$, the sensing matrix $A_k \in \mathcal{T}$ for the $k$th diffraction zone is obtained by solving the optimization problem

$$
\minimize \mathbb{E}[\text{dist}(x; \mathcal{S}_\theta(y_k \equiv f^i(x; A_k); A_k))] \quad (5)
$$

where the expectation is over the distribution of $x$ through the chosen image data set for end-to-end training. Note that this approach considers signal reconstruction accuracy as a criterion with respect to a particular realization of the decoder functions. Thus, $A_k$ obtained in this technique is a task-specific encoding matrix that considers not only the underlying distribution of the signal but also the class of decoder functions, resulting in a superior performance. After solving the aforementioned optimization problem, the sensing matrix is utilized for data-acquisition purposes while the fixed decoder module $\mathcal{S}_\theta$ is used for the reconstruction of $x$.

DOE design with sparsity prior on the scene

In optical applications such as radio astronomy [36] and microscopy [6], the scene $x$ is naturally sparse in some orthogonal domain $\Psi$, i.e., $\Psi^\dagger \Psi = I$; common examples include wavelet or discrete cosine transform. Then, there exists an $s$ sparse $s < n$ representation $z \in \mathbb{C}^n$ such that $\| z \|_0 = s$ where $x = \Psi^\dagger z$ and $\| \cdot \|_0$ are the $\ell_0$ pseudonorm. It was shown in [3] that very low estimation errors in the sparse representation $z$ of $x$ were obtained when the admissible random variable $d$ that models the DOEs satisfies $\mathbb{E}[d] \neq 0$. This result suggests the type of coding elements of $d$ needed for the highest performance without any restriction on design strategy. Consequently, the unfolding co-design method is also valid for the sparsity prior scenarios; only the first layer needs to be changed to incorporate the aforementioned result by choosing an appropriate $d$. Note that $d$ plays a crucial role in uniquely identifying $x$ from the CDP and the flexibility of the unfolding framework to combine model-based theoretical results and neural networks. Thus, it is desirable to consider and replicate [3] for the analogous analysis of a given application to achieve optimal imaging quality.

Physical constraints

The optimization problem in (5) is intended to design the measurement matrix $A_k$ for a given diffraction zone $k$. More physical constraints may be incorporated into the problem. For example, given the block-matrix structure in (1), the DOE may be modeled as a piecewise continuous function [37] wherein the constant intervals of the function are characterized by the random variable $d$. Similarly, the Fresnel order or DOE thickness
may be described using a smoothing function [33]. In each one of these cases, the physical parameters are learnable.

Fresnel order [thickness]
Thickness of the DOE is defined as \( Q = 2\pi m_Q \) (in radians) where \( m_Q \) is the “Fresnel order” of the mask, which is not necessarily an integer. Denote the wavelength of the source by \( \lambda_0 \) and the resulting phase profile of the DOE by \( \phi_{\lambda_0} \). Then, the phase profile of the DOE considering the thickness is given by

\[
(\phi_{\lambda_0})_{p,q} = \text{mod}\left((\phi_{\lambda_0})_{p,q} + \frac{Q}{2}, Q\right) - \frac{Q}{2} \tag{6}
\]

resulting in \( \phi_{\lambda_0} \) taking values in the interval \([-Q/2, Q/2)\).

Number of levels
The DOE is defined on a discretized 2D grid. We obtain a piecewise-invariant surface for the DOE after a nonlinear transformation of its absolute phase. The uniform grid discretization of the wrapped phase profile \( \phi_{\lambda_0} \) up to \( N \) levels is

\[
(\beta_{\lambda_0})_{p,q} = \left[(\phi_{\lambda_0})_{p,q}/N\right] \cdot N \tag{7}
\]

where \( \lfloor w \rfloor \) denotes a floor operation that yields the largest integer less than or equal to \( w \). The values of \( \beta_{\lambda_0} \) are restricted to \([-Q/2, Q/2)\) where each step of the piecewise function is a coding element of \( d \). Here \( Q \) is an upper bound of the thickness phase of \( \beta_{\lambda_0} \).

FIGURE 1. (a) The physical implementation of a compact microscope using a DOE manufactured in Tampere’s Photonics Laboratory with an electron beam lithography system on a fused silica glass and its fragments. The CDPs are acquired in the near zone \( \mathcal{A}_1 \) in (1). When compared with a theoretically expected design, note that the fabricated DOE has imperfections arising from physical artefacts such as material imperfections and printing inaccuracies. The top and side views of scanning electron microscope images of the DOE are shown in the right half of the image and magnified further. (b) Frames from a complex-valued dynamic object video with amplitudes [in or airy units (a.u.) top row] and phases (in radians or rad; bottom row) of a moving single-celled eukaryote reconstructed using sparsity-based filtering method. The video footage consisted of 287 frames through 10 s (see visualization), of which frames 50, 100, 150, and 200 frames are reproduced (left to right) here.
The floor and modulo functions are not differentiable. Therefore, a smoothing approximation is used for optimizing the DOE’s thickness and number of levels (see [33] for details).

Pixel resolution and size of the DOE
These parameters are also considered while designing the measurement matrix, because in practice, all sizes and resolutions may not be implementable [38]. Further details on the impact of these physical DOE constraints in a practical imaging application are available in [39].

Unfolding for optical imaging
Compared to some popular neural network architectures (Figure 2), the unfolded network is a directed acyclic graph that is trained like any standard neural network. However, the input to this unfolded network is not a single vector from the sequence, but rather it is the entire sequence used all at once. The output used to compute the gradients is also the entire sequence of output values that a typical neural network would have produced for each step in the input sequence. As such, the unfolded network has multiple inputs and outputs with identical weight matrices for each step.

A preliminary step in unfolding is to mathematically formulate a proper parameterized class of decoder and encoder modules that facilitate incorporating the domain knowledge. To this end, the decoder is obtained by first formulating a phase-retrieval problem as a minimization program and then resorting it to first-order optimization techniques that lay the ground for creating a rich class of parameterized decoder functions. Once such a class is formalized, a connection between deep neural networks and first-order optimization techniques is established.

Underlying principles
Assume $S_{θ,υ} ∈ Φ$ and $S_{θ,υ} ∈ Φ$ represent two realizations of the decoder module from the same class. Denote the fixed sensing matrix that will be the solution to (5) for $θ = θ_i$ by $A_k$ for the $k$th diffraction zone. Then, for a signal of interest $x$ and corresponding phaseless measurements $y_i$, it is possible that

$$E[\text{dist}(x; S_{θ_i}(y_i; A_i))] \leq E[\text{dist}(x; S_{θ_i}(y_i; A_i))]. \quad (8)$$

As a result, it is more appropriate to consider the design of the sensing matrix $A_k$ with respect to $S_{θ_i}$ and find an optimal $θ^* ∈ Φ$ such that

$$E[\text{dist}(x; S_{θ^*}(y_i; A_i))] \leq E[\text{dist}(x; S_{θ}(y_i; A_i))] \quad \forall \theta ∈ Φ. \quad (9)$$

If such a characterization $θ^*$ is known, the sensing matrix is easily designed. The aforementioned observation is further evidence of the paramount importance of developing a unified framework that allows for a joint optimization of the reconstruction algorithm (i.e., finding an optimal or suboptimal $θ^*$) and the sensing matrix. Then, the problem of jointly designing $A_k$ and the reconstruction algorithm is equivalent to the optimization program

$$\min_{A_k ∈ Φ, \theta ∈ Φ} E[\text{dist}(x; S_{θ}(f^1(x; A_k); A_i))]. \quad (10)$$

Next, we describe tackling this optimization problem using the available data.

Encoder architecture
Define the hypothesis class $H_e$ of the possible encoder functions (module) whose computation dynamics mimics the behavior of the data-acquisition system in a phase-retrieval model (e.g., $E[d] ≠ 0$, a temporal or spatial correlation of $A_k$). Then,

$$H_e = \{f^1(\cdot; A_k) : x → |A_kx| : A_k ∈ C^{m × n}\}. \quad (11)$$

The superscript in $f^1$ denotes that this hypothesis class corresponds to the first layer of the unfolded network. It can be interpreted as a one-layer neural network with $n$ input neurons and $m$ output neurons, where the activation function is $|\cdot|$ with the trainable weight matrix $A_k$. The encoder (matrix $A_k$) for a given realization of a decoder function $S_{θ}$ with respect to the hypothesis class $H_e$ is designed by solving the following learning problem.
minimize \( \mathbb{E} [\text{dist}(x; S_{\theta} \cdot f^i(x; A_i))] \) \quad (12)

Where \( \ast \) stands for the composition of functions. This problem seeks to find an encoder function \( f^i(\cdot; A_i) \) such that the resulting functions maximize the reconstruction accuracy according to the decoder function \( S_{\theta} \).

**Decoder architecture**

The goal here is to obtain an estimate of the underlying signal of interest from the phaseless measurements of the form \( f^i(x; A_i) = |A_i x| \) for a given sensing matrix \( A_i \) of the \( k \)th diffraction zone. Equivalently, a decoder function seeks to tackle the following problem:

\[
\min_{z \in C^n} \frac{1}{2m} \| f^i(z; A_i) - f^i(x; A_i) \|_2^2 \quad \text{subject to} \quad z \in \Gamma
\]

(13)

where \( \Gamma \) represents the search space of the underlying signal of interest. For instance, if \( x \) is a sparse, this information is encoded in \( \Gamma = \{ z \in C^n : \| z \|_0 = s \} \).

**Application to phase retrieval**

Deep unfolded networks have repeatedly displayed great promise in various signal processing applications [22], [31] by exploiting copious amounts of data along with the domain knowledge gleaned from the underlying problem. This is suitable for deploying phase retrieval in optical imaging, particularly in nonconvex settings where bounding the complexity of signal processing algorithms is often traded off for performance. In general, iterative-optimization techniques are a popular choice for nonconvex phase-retrieval problems. Specifically, first-order methods are widely used iterative-optimization techniques because of their low per-iteration complexity and efficiency in complex scenarios. One such prominent approach that is also suitable for training the decoder is the projected gradient descent (PGD) algorithm. Here, given an initial point \( z^{(0)} \), the \( r \)th iteration of PGD is

\[
z^{(r+1)} = \mathcal{P}_{\Gamma} \left( z^{(r)} - \frac{\alpha^{(r)}}{m} \nabla_{z} \ell(z^{(r)}) \right)
\]

(14)

where \( \mathcal{P}_{\Gamma} : C^n \rightarrow \Gamma \subseteq C^n \) denotes a mapping function of the vector argument to the feasible set \( \Gamma \), \( \nabla_{z} \ell(z^{(r)}) \) is the gradient of the objective function obtained at the point \( z^{(r)} \), and \( \alpha^{(r)} \) represents the step size of the PGD algorithm at the \( r \)th iteration. To bound the computational cost of first-order methods, a sensible approach is to fix the total number of iterations of such algorithms (e.g., \( T \)) and to follow up with a proper choice of the parameters for each iteration, which results in the best improvement in the objective function while allowing only \( T \) iterations.

Define a parameterized mapping operator \( f^i(\cdot; \rho^i) : C^n \rightarrow C^n \) for \( t \geq 2 \) (see Figure S2) as

\[
f^i(z; \rho^i) = \mathcal{P}_{\Gamma} \left( z - \frac{1}{m} G^{(i)} \nabla_{z} \ell(z) \right)
\]

(15)

where \( \rho^i = \{ G^{(i)} \} \) is the set of parameters of the mapping function \( f^i(\cdot; \rho^i) \) and \( G^{(i)} \) is a positive-semidefinite matrix.

This mapping is then utilized to model various first-order optimization techniques for a given problem. In particular, for the PGD algorithm, performing \( T - 1 \) iterations of the aforementioned form (Figure S2) is modeled as

\[
N_{T-1}^{-1}(x) = f^i(\cdot; \rho^{T}) \circ f^{T-1}(\cdot; \rho^{T-1}) \circ \cdots \circ f^2(\cdot; \rho^{2})
\]

(16)

where the aforementioned function corresponds to the PGD in all scenarios, including for a fixed step size \( G^{(i)} = aI \), time-varying step sizes \( G^{(i)} = \alpha^{(i)} I \) or the general preconditioned PGD algorithm with an arbitrary choice of \( G^{(i)} \geq 0 \).

The specific preconditioning-based update rule in (18) is known to be advantageous in model-based algorithms to speed up the convergence rate [41, Ch. 5] and yield robustness when the sensing matrix is ill-conditioned (i.e., the condition number is far greater than one) [42], [43], [44]. This latter advantage is critical in practical imaging applications (listed in Table 1) to overcome imperfect system modeling, reconstruction artefacts such as mismatch, calibration errors, and even poorer performance for the single snapshot of an experimental sensing matrix [45], [46]. In the context of unfolding [47], the benefits of preconditioning appear indirectly because it was learned as a matrix that encapsulates both sensing (via its transpose) and preconditioning matrices.

Finally, for a fixed \( A_i \), learning the deep decoder module with respect to a realization of the encoder function corresponds to the problem

\[
\min_{T = (G^{(0)} \geq 0)} \mathbb{E} [\text{dist}(x; N_{T-1}^{-1} \circ f^i(x; A_i))].
\]

(17)

This formulation is akin to training a \( T \)-layer deep neural network \( N_{T-1}^{-1} \circ f^i(x; A_i) \) with a task-specific architecture whose computation dynamics at each layer mimic the behavior of one iteration of a first-order optimization algorithm, and the input to such a model-based deep scheme is given by a realization of the encoder module. As domain knowledge is included in the design of such a deep unfolded network, it is more interpretable and has fewer training parameters than its black-box counterpart.

**Unfolding-aided techniques for coded DOI**

The prevailing usage of unfolding algorithms in phase retrieval is centered on the sparsity assumption, or lack thereof, of sparsity in the target scenario in a nonconvex setting. More recently, optical applications such as 3D object detection and 3D phase imaging have also exploited unfolded networks.

**UPR without priors**

The computational dynamics and network structure of a UPR architecture (see “Unfolded Phase Retrieval Without Priors”) for this scenario are depicted in Figure S3. rom the empirical success rate, it follows that the unfolded network significantly outperforms the model-based algorithm in all scenarios.

When the measurement matrix is fixed (by the physics of the system) and known a priori, learning the preconditioning matrices \( G^{(i)} \) significantly improves the performance...
of the underlying signal-recovery algorithm. This supports the proposition that a judicious design of the preconditioning matrices, when the number of layers (iterations) is fixed, indeed results in an accelerated convergence. The unfolding approach is better, even when only the measurement matrix is learned while employing a fixed-scalar step size. With both a learned $A_k$ and $G^{(l)}$, the unfolded network is the most successful.

Figure 3 demonstrates the relative error as in (3) as a function of iterations (layers). It follows that UPR methodology benefits from accelerated convergence, and an optimal point of the objective function is reached with as few as $T = 30$ layers.

Without sparsity prior, the optimization problem for phase retrieval follows directly as

$$\min_{z \in \mathbb{C}^n} \mathcal{L}(z; f(x; A_k)) \equiv \frac{1}{2m} \| f(x; A_k) - f(z; A_k) \|_2^2 \quad \text{(S1)}$$

where $m = n L$. $L$ is the number of snapshots of the scene, and signal length $n$ is a constant parameter that controls the step size of each iteration.

The iterations for finding the critical points of this nonconvex problem are derived as follows. Starting from a proper initial point $z^{(0)}$, a sequence of points $\{z^{(0)}, z^{(1)}, \ldots, z^{(t)}\}$ is generated according to the following update rule:

$$z^{(t+1)} = z^{(t)} - \frac{1}{m} G^{(1)} \nabla \mathcal{L}(z^{(t)}; f(x; A_k)) \quad \text{(S2)}$$

This update rule was shown to be highly effective in solving the phase-retrieval problem [40], [S1]. Therefore, it is instructive to further explore its potential in the unfolding context. The gradient of the objective function is

$$\nabla \mathcal{L}(z^{(t)}; f(x; A_k)) = A_k^T z^{(t)} - f(x; A_k) \odot \text{Ph} (A_k z^{(t)}), \quad \text{(S3)}$$

where $\odot$ is the inner product, and the function $\text{Ph}(z)$ is applied elementwise and captures the phase of the vector argument; e.g., for real-valued signals $\text{Ph}(z) = \text{sign}(z)$.

Reference

[S1] G. Wang, L. Zhang, G. B. Giannakis, M. Akçakaya, and J. Chen, "Sparse phase retrieval via truncated amplitude flow," IEEE Trans. Signal Process., vol. 66, no. 2, pp. 479–491, 2017, doi: 10.1109/TSP.2017.2771733.

**Unfolded Phase Retrieval Without Priors**

**Computation Dynamics**

Initialize

$$z^{(0)} \leftarrow A_k^T (A_k z - f^1(x; A_k) \odot \text{Ph} (A_k z))$$

Every layer $t \in \{2, \ldots, T\}$ is tuned to compute

$$f^t(z^{(t-1)}; G^{(t)}) \leftarrow z^{(t)} - \frac{1}{m} G^{(t)} \nabla \mathcal{L}(z^{(t)}; f^t(x; A_k))$$

where $z^{(t)}$ is the input to the $r$th layer.

**FIGURE S3.** (a) A single iteration of the unfolded phase-retrieval (UPR) method executes both linear and nonlinear operations and is thus recast into (b) a network layer. Stacking the layers together, (c) a deep unfolded network is formed. The network is subsequently trained using paired inputs and outputs by backpropagation to optimize the parameters $G^{(l)}$ and $A_k$. The learning rate was set to $10^{-4}$ for 100 epochs using an Adam stochastic optimizer. (d) The empirical success rate, obtained by averaging more than 100 Monte Carlo trials, is shown with respect to the measurement-to-signal-length ratio $(m/n)$ when the signal length is fixed and set to $n = 100$. The size of the training data set is 2,048, where each data point is an $n$-dimensional vector following a standard Gaussian distribution. In the model-based case, the sensing matrix $A_k$ follows a standard Gaussian distribution with a variance equal to one.
This reveals that the proposed architecture may be truncated to even fewer iterations, thereby saving computations in the overall algorithm.

**UPR with sparsity prior**

In sparse phase retrieval (see “UPR With Sparsity Prior”), decoding signal $x$ from the measurement vector $f^t(x; A_k)$ is formulated as a nonconvex optimization problem (see the “Underlying Principles” section). Both the objective function $\mathcal{L}(z; f^t(x; A_k))$ and the constraint for this sparse problem are nonconvex, making it NP-hard in its general form. Figure S4 shows how the relevant UPR architecture, which significantly outperforms the standard state-of-the-art, sparse model-based algorithm under various learned parameters. When the measurement matrix is imposed by the physical system, then learning the preconditioning matrices significantly increases the recovery performance. Considering learning only the matrix $A_k$ and employing fixed step sizes further indicates the effectiveness of learning task-specific measurement matrices tailored to the application at hand.

The objective of learning the preconditioning matrices is to accelerate convergence. This may be validated through a per-layer analysis of the sparse UPR framework in terms of achieved relative error [as in (3)]. The model-based nature of the proposed approach facilitates analysis of the resulting interpretable deep architectures as opposed to the conventional black-box data-driven approaches. Figure 4(a) demonstrates the relative error versus the number of iterations/layers for the case of recovering a five-sparse signal $x \in \mathbb{R}^n$, $n = m = 300$. It follows that the UPR yields both a lower relative error and faster convergence than the standard model-based algorithm [40]. The success rate versus the sparsity plot in Figure 4(b) further highlights the superior performance of UPR.

**Unfolded object detection**

For 3D object detection [2] based on phase-only information from the CDP, a rapid estimation of the scene from the CDP is carried out before detecting the target using its optical phase through a cross-correlation analysis (template matching). For the estimation, filtered spectral initialization approximates the phase of the 3D object from the acquired CDP. This requires low-pass filtering of the leading eigenvector of matrix $A = (1/|I_0|) \sum_{\epsilon \in I_0} (a_\epsilon a_\epsilon^T)/\|a_\epsilon\|^2$. Here, unfolding (Figure 5) is employed to learn the low-pass filter [31], where $I_0$ is as in the “UPR With Sparsity Prior” section. The low-pass filter, which facilitates accurate estimation of the phase of the object, is the key aspect of initialization.

The filtered spectral initialization follows a power-iteration method (see Figure 5), wherein $z^{(r+1)} = W * (\Lambda z^{(r)})$ where $z^{(0)}$ is chosen at random, $W$ is the low-pass filter, and $*$ is the convolution operation. This product is then normalized. Iteratively applying $W$ over the estimation of the image selects those low frequencies that sparsely represent the image in the Fourier domain. As the filtering process is a convolution, this selection is rapidly performed.
In the next step, a circular harmonic filter (CHF) (in Fourier domain polar coordinates) is built based on a reference pattern of the scene [2]. This system of coordinates is preferred to make CHF invariant to rotations so that the filter detects the object, regardless of any of its rotated versions. The correlation matrix $\hat{C}$ between the CHF and the Fourier transform of the approximated optical field $\hat{Z}$ is then computed. The target is detected using a thresholding approach by building the decision matrix

$$(R)_{u,v} = \begin{cases} 1, & \text{if } |(\hat{C})_{u,v}| \geq \varepsilon \cdot \max(\hat{C}) \\ 0, & \text{otherwise} \end{cases} \quad (18)$$

where $\varepsilon \in (0, 1]$ is the tolerance parameter. Figure 5 shows the unfolded detection performance using an experimental CDP in the near zone for a single snapshot. Here, the experimental data were acquired following the optical setup in [7].

Unfolded phase imaging

Phase imaging is of high interest in autonomous vehicles, the detection of moving objects, and precision agriculture. Here the 3D shape of an object is recovered by solving the non-convex phase retrieval in a setup that records CDPs. Usually, the acquisition of several snapshots from the scene is required. Recently, a single-shot technique was proposed in [17]; it accurately estimates the optical phase $\phi(x, y)$ across the spatial domain $(x, y)$ of the object following the pipeline of Figure 5. The estimated phase is then used to infer the 3D shape by performing an unwrapping over $\phi(x, y)$ to obtain

**Unfolded Phase Retrieval With Sparsity Prior**

Starting from an initial guess $z^0$, the algorithm estimates the signal in an iterative manner via update equations as

$$z^{n+1} = \mathcal{P}_I(z^n - \frac{\alpha}{m} \nabla_z \mathcal{L}_m(z^n; f(x; A_i))). \quad (S4)$$

These iterations are viewed as performing projected gradient descent with a per-iteration step size $\alpha$ on the truncated loss function objective $\mathcal{L}_m$. The gradient is

$$\nabla_i \mathcal{L}_m(z^n; f(x; A_i)) = \sum_{i \in I} \left( a_i^m z^n - f(x; A_i) \right) \frac{a_i^m z^n}{|a_i^m z^n|} a_i,$$

where $I = \{1 \leq i \leq m : |a_i^m z^n| \geq |f(x; A_i)| / (1 + \tau) \}$. The constant $\tau$ represents the truncation threshold and $\mathcal{P}_I(u) : C^n \to C^n$ sets all entries of $u$ to zero except for the $s$ entries with the largest magnitudes. This truncation addresses the noncontinuity of the gradient by eliminating the terms that could lead to unbounded descent direction [S2]. This provides strong motivation to investigate this procedure for unfolding.

The vector $z^m$ is initialized as $\sqrt{\Sigma_{\nu=1}^{n-1} \left| f'(x; A_i) \right| / m} z^0$ where $z^0 \in C^n$ is created by placing zeros in $z^0$ where the indices are not in $\mathcal{S}$ (the indices with the s-largest values of $z^0$). The principal eigenvector $z^\text{opt} \in C^n$ is determined by performing power-method iterations on the matrix

$$\Lambda = \frac{1}{|x|} \sum_{i \in I} a_i z_i / |a_i z_i|^2.$$

The condition $\mathbb{E}[\Sigma_i \Sigma_i \cdot D_i D_i] = r I$ imposed over the set of diffractive optical elements to guarantee uniqueness is needed for this initialization to return an accurate estimation of $x$ [3]. This result connects the initialization procedure with the recovery conditions of coded diffraction pattern phase retrieval.

**Reference**

[S2] Y. Chen and E. Candès, “Solving random quadratic systems of equations is nearly as easy as solving linear systems,” in Proc. Adv. Neural Inf. Process. Syst., 2015, pp. 739–747.

![Figure S4](image_url)

**Figure S4.** (a) The unfolded network is trained using paired inputs and outputs by backpropagation to optimize the parameters $G^{\text{opt}}$ and $A_k$. Here, the signal length $n$ is a constant that controls the step size (equal to one) of each iteration. The learning rate was set to $10^{-4}$ for 100 epochs, using an Adam stochastic optimizer. (b) The empirical success rate obtained by averaging more than 100 Monte Carlo trials versus the measurement-to-signal-length ratio $(m/n)$ when the signal length is set to $n = 100$ and $s = 5$. The size of the training data set is 2,048, where each data point is an $n$-dimensional vector following a standard Gaussian distribution, whose $(n - s)$ entries set to zero are chosen uniformly at random. In the model-based case, the sensing matrix $A_k$ follows a standard Gaussian distribution with a variance equal to one.
\( \varphi_{\text{unwrap}}(x, y) = \varphi(x, y) + 2k\pi \) where \( k \) is an integer that represents the projection period [17]. The depth coordinate is based on the difference between the measured phase \( \varphi_{\text{unwrap}} \) and the phase value from a reference plane. This reference plane is obtained during calibration of the acquisition system [17]. Thus, the depth coordinate of the 3D object is

\[
\text{depth}(x, y) = c_0 + c_1 \cdot (\varphi_{\text{unwrap}}(x, y) - \varphi_0(x, y))
\]

where \( \varphi_0(x, y) \) is the reference phase, and \( \{c_i\}_{i=0} \) are tunable constants.

### Enabling DOE experimental setups

The advantage of the DOE lies in its imaging capability to successfully recover the phase without additional optical elements (such as lenses), leading to even more compact imaging devices [6]. The eschewing of the lens, which is the case in phase retrieval from CDPs, makes the system not only light and cost-effective but also lens-aberration free and with a larger field of view. Here we summarize a few recent developments in experimental setups that employ DOEs and are of interest to enable the use of unfolding-aided phase-retrieval techniques.

### Lensless superresolution microscopy

In case of lensless superresolution microscopy for middle-zone CDP observations (Figure 1 and Table 1), the propagation operator \( P_z(w) \) at a distance \( z \) is precisely modeled through convolution of a wavefront \( w \) with propagation kernel \( K_z \) as \( P_z(w) = F^{-1}\{K_z \cdot F\{w\}\} \) where

\[
K_z = \begin{cases} 
  e^{2\pi i \phi_{02}(1 - \frac{x^2 + y^2}{\Lambda^2})}, & \frac{x^2 + y^2}{\Lambda^2} \leq 1, \\
  0, & \text{otherwise}.
\end{cases}
\]

Considering the spatial distribution of the DOE, the observation model [6] for the scene \( x \) and the DOE \( D \) is \( y = |P_z(D \cdot P_z(x))|^2 \) where \( z_1 \) and \( z_2 \) are object-mask and mask-sensor distances, respectively (see Table 1). In a phase retrieval with a DOE setting, the object, mask, and sensor planes may have different sampling intervals, \( \Delta_o \), \( \Delta_m \), and \( \Delta_s \), respectively. If the resulting pattern \( y \) has computational pixels smaller than the physical sensor \( (\Delta_s < \Delta_o) \) and the reconstruction of the object is produced with \( \Delta_s \), it leads to a subpixel resolution or superresolution scenario. The ratio \( r_s = \Delta_s / \Delta_o \geq 1 \) is the superresolution factor.

With the objective of manufacturing the DOE, the matrix \( D \) is modeled as \( D_s = \exp(j\phi_{s,u}) \) for \( u, s = 1, \ldots, n \) where \( \phi_{s,u} \) takes value from a uniform random variable \( d \) with two possible equiprobable coding elements zero and 2.62 rad i.e., \( d = \{\exp(0), \exp(2.62\pi \text{rad})\} \). These values of \( d \) are chosen because they lead to the best performance in terms of reconstruction quality [6] (recall \( E[d] \neq 0 \)). Figure 1 shows the resulting DOE (manufactured in Tampere’s Photonics Laboratory) with the pixel size of \( \Delta_m = 1.73 \mu m \). This size is half of the sensor, and hence a better fit for the superresolution factor \( r_s \) in powers of two. The compact microscope in Figure 1 enables a full wavefront reconstruction of dynamic scenes.
Contrary to existing phase-retrieval setups, this unique system is capable of computational, superresolved video reconstruction via phase retrieval, with a high frame rate limited only by the camera.

To reconstruct the signal of interest as reported in the bottom of Figure 1(b), [6] studied the impact of a special multiplicative calibration term, which allows for building a good propagation model to compensate for the errors of theoretical models. This corrective term was shown to improve the imaging quality and resolution. However, it is constant across iterations. Therefore, it may benefit from a reconditioning unfolded method, wherein the correction is varied on a per-iteration basis.

**Hyperspectral complex-domain imaging**

Hyperspectral complex-domain imaging is a comparatively new development that deals with the phase delay of coherent light in transparent or reflective objects [49]. The hyperspectral broadband-phase imaging is more informative than the monochromatic technique (see the near-zone setup and results in Figure 6). Conventionally, for the processing of hyperspectral images, 2D spectral narrow-band images are stacked together and represented as 3D cubes with two spatial coordinates, \((x, y)\), and a third longitudinal spectral coordinate. In hyperspectral phase imaging, for a given wavelength \(\lambda\), the data \(x_{\lambda} \in \mathbb{C}^n\) in these 3D cubes are complex valued with spatially and spectrally varying amplitudes and phases. Denote the set of wavelengths under study by \(\mathcal{L}\). From (1), the measurement matrix for diffraction zone \(k\) and wavelength \(\lambda\) is \(A_{k,\lambda}\). Then, the noisy phaseless measurements are [49]

\[
y_{\lambda} = \sum_{k \in \mathcal{L}} |A_{k,\lambda}x_{\lambda}|^2 + \epsilon.
\]  

The reconstruction of \(x_{\lambda}\) here is more ill-conditioned than the classical phase-retrieval problem because the acquired data are summed across wavelengths of phaseless measurements. This sum does not appear in the classical phase retrieval, which employs a single wavelength. This also makes phase image processing more complex than the hyperspectral intensity imaging, where the corresponding 3D cubes are real valued.

Model-based algorithms require significant amounts of measurements to efficiently recover the complex-valued \(x_{\lambda}\) [49]. As suggested by the aforementioned numerical results, the preconditioning unfolding is useful in the hyperspectral, complex-domain phase-retrieval problem because it facilitates reduction of the condition number, admitting fewer data samples and converging quickly.

**Blind deconvolution superresolution phase retrieval**

In certain CDP applications, the combination of blind deconvolution, superresolution, and phase retrieval naturally manifests. Although this is a severely ill-posed problem, it has been shown [52] that image of interest could be estimated in polynomial time. The approach relies on previous results, which established the DOE design to achieve high-quality images [5] and partially analyzed the combined problem by solving the superresolution phase-retrieval problem [7] from a CDP (Figure 7). These designs are obtained by exploiting the model of

![Diagram](image-url)
the physical setup using ML methods, where the DOE is modeled as a layer of a neural network (a data-driven model or unrolled) that is trained to act as an estimator of the true image [33]. This data-driven design has shown outstanding imaging quality, using a single snapshot as well as robustness against noise.

Summary and future outlook

The plethora of phase-retrieval algorithms and their relative advantages make it difficult to conclude which of them will always be best. The only foregone conclusion has been that the computational imaging community needs to continue basing these algorithms on models devoid of several practical measurement errors. In this context, deep unfolding methods are gaining salience in optical phase-retrieval problems to offer interpretability, state-of-the-art performance, and high computational efficiency. By learning priors from the trained data, these empirically designed architectures have been exploited to build globally converged learning networks for a specific phase-retrieval setting. A particularly desired outcome of this development is improvement in the design of DOEs that satisfies theoretical conditions for the uniqueness of signal recovery.

We provided examples of several recent novel setups, which indicate that unfolding has the potential to enable a new generation of optical devices as a consequence of co-designing optics and recovery algorithms. Several applications of lensless imaging, synthetic apertures, and highly ill-posed variants of phase retrieval appear within reach in the near future using this technique, including in areas such as optical metrology, digital holography, and medicine.

Authors

Samuel Pinilla (samuel.pinilla@stfc.ac.uk) received his Ph.D. degree in electrical engineering. He is with the University of Manchester at Harwell Science and Innovation campus, Oxfordshire, Oxon OX11 0FA, U.K. He is a Member of IEEE.

Kumar Vijay Mishra (kvm@ieee.org) received his Ph.D. degree in electrical and computer engineering. He is with the United States DEVCOM Army Research Laboratory, Adelphi, MD 20783 USA. He is a Senior Fellow of IEEE.

Igor Shevkunov (igor.shevkunov@tuni.fi) received his Ph.D. degree in optics. He is with the Faculty of Information Technology and Communication Sciences, Tampere University, 33720, Finland.

Mojtaba Soltanalian (msol@uic.edu) received his Ph.D. degree in electrical engineering. He is with the University of Illinois Chicago, Chicago, 60607 USA. He is a Senior Member of IEEE.

Vladimir Katkovnik (vladimir.katkovnik@tuni.fi) received his Ph.D. degree in technical cybernetics. He is with the Faculty of Information Technology and Communication Sciences, Tampere University, 33720, Finland and Noiseless Imaging Ltd., Tampere, 33720, Finland.

Karen Egiazarian (karen.egiazarian@tuni.fi) received his Ph.D. degree in physics and mathematics. Faculty of Information Technology and Communication Sciences, Tampere University, 33720, Finland and Noiseless Imaging Ltd., Tampere, 33720, Finland. He is a Fellow of IEEE.
Phase retrieval guarantees from [22].

Artif. Vis. Lett.

[21] Phase recovery using four rotated versions of a single binary amplitude modulating [20] vol. [19] volution patterns [18], pp. [17], doi: 10.1016/j.cam.2018.02.002.

J. Baca, S. Pinilla, and H. Arguello, “Coded aperture design for super-resolu-

tion phase retrieval,” in Proc. Eur. Signal Process. Conf., 2019, pp. 1–5, doi: 10.23919/EUSIPCO.2019.8803020.

P. Kocius, I. Shevkunov, V. Katkovnik, H. Rekola, and K. Egiazarian, “Single-

shot pixel super-resolution phase image by wavefront separation approach,” Opt.

Exp., vol. 29, no. 26, pp. 43,662–43,678, 2021, doi: 10.1364/OE.445218.

V. Katkovnik, I. Shevkunov, N. V. Petrov, and K. Egiazarian, “Computational super-resolution phase retrieval from multiple phase-coded diffraction patterns: Simulation study and experiments,” Optica, vol. 4, no. 7, pp. 786–794, 2017, doi: 10.1364/OPTICA.4.000786.

E. I. Candés, X. Li, and M. Soltanolkotabi, “Phase retrieval from coded diffraction patterns,” Appl. Opt. Commun., vol. 39, no. 2, pp. 277–299, 2015, doi: 10.1016/j.aopca.2014.09.004.

D. Gross, F. Kraehmer, and R. Kueng, “Improved recovery guarantees for phase retrieval from coded diffraction patterns,” Appl. Opt. Commun., vol. 42, no. 1, pp. 37–64, 2017, doi: 10.1016/j.aopca.2015.05.004.

M. Hayes, “The reconstruction of a multidimensional sequence from the phase or magnitude of its Fourier transform,” IEEE Trans. Acoust. Speech, Signal Process., vol. 30, no. 2, pp. 140–154, Apr. 1982, doi: 10.1109/TASSP.1982.1163863.

T.-C. Poon and J.-P. Liu, Introduction to Modern Digital Holography: With MATLAB. Cambridge, U.K.: Cambridge Univ. Press, 2014.

U. Dürig, D. W. Pohl, and F. Rohner, “Near-field optical-scanning microscopy,” J. Appl. Phys., vol. 59, no. 10, pp. 3318–3327, 1986, doi: 10.1063/1.336848.

C. Jahncke, M. Paesler, and H. Hallen, “Raman imaging with near-field scan-

ning optical microscopy,” Appl. Phys. Lett., vol. 67, no. 27, pp. 2483–2485, 1995, doi: 10.1063/1.114165.

H. Hess, E. Betzig, T. Harris, L. Pfeiffer, and K. West, “Near-field spectroscop-

y of the quantum constituents of a luminescent system,” Science, vol. 264, no. 5166, pp. 1740–1745, 1994, doi: 10.1126/science.264.5166.1740.

T. Shimano, Y. Nakamura, K. Tajima, M. Sao, and T. Hoshizawa, “Lensless light-field imaging with Fresnel zone aperture: Quasi-coherent coding,” Opt. Lett., vol. 57, no. 11, pp. 2841–2850, 2018, doi: 10.1364/AO.57.028281.

M. Sao, Y. Shimano, K. Tajima, and T. Shimano, “Lensless close-up imaging with Fresnel zone aperture,” Japanese J. Appl. Phys., vol. 57, no. 981, p. 095805, 2018, doi: 10.1109/JAP.2018.095805.

S. Pinilla, L. Galvis, K. Egiazarian, and H. Arguello, “Single-shot coded dif-

fraction system for 3D object shape estimation,” Electron. Imag., vol. 2020, no. 14, pp. 59–69, 2020, doi: 10.2352/ISSN.2470-1173.2020.14.CIMG-059.

H. Arguello, S. Pinilla, Y. Peng, H. Ikoma, J. Baca, and G. Wetzstein, “Shift-

variant color-coded differential spectral imaging system,” Optica, vol. 8, no. 12, pp. 1424–1434, 2021, doi: 10.1364/OPTICA.439142.

K. Monakhova, J. Yurtshev, G. Kuo, N. Antipa, K. Kanny, and L. Waller, “Learned reconstructions for practical mask-based lensless imaging,” Opt. Exp., vol. 27, no. 20, pp. 28,075–28,090, 2019, doi: 10.1364/OE.27.028075.

V. Karintas, E. Nitiss, A. Tokmakov, M. Oszlinski, and S. Logina, “Optical phase retrieval using four rotated versions of a single binary amplitude modulating mask,” Opt. Express, vol. 26, no. 3, pp. 30904–30904, 2018, doi: 10.1117/1.JOE.1.3.30904.

Z. Tong, X. Ren, Q. Ye, D. Xiao, J. Huang, and G. Meng, “Quantitative reconstruction of the complex-valued object based on complementary phase modula-

tions,” Ultramicroscopy, vol. 228, Sep. 2021, Art. no. 113343, doi: 10.1016/j.ultramic.2021.113343.

D. Morales, A. Jerez, and H. Arguello, “Object classification using deep neural networks from coded diffraction patterns,” in Proc. Symp. Image, Signal Process. Artif. Vis., 2021, pp. 1–5, doi: 10.1109/SSPA.2021.959199.

C.-J. Wang, C.-K. Wen, S.-H. L. Tsai, and S. Jin, “Phase retrieval with learning unfolded expectation consistent signal recovery algorithm,” IEEE Signal Process. Lett., vol. 27, pp. 780–784, Apr. 2020, doi: 10.1109/LSP.2020.2990767.

C.-J. Wang, C.-K. Wen, S.-H. Tsai, S. Jin, and G. Y. Li, “Phase retrieval using expectation consistent signal recovery algorithm based on hypernetwork,” IEEE Trans. Signal Process., vol. 69, pp. 5770–5783, Oct. 2021, doi: 10.1109/TSP.2021.3118494.

D. Morales, A. Jerez, and H. Arguello, “Object classification using deep neural networks from coded diffraction patterns,” in Proc. Symp. Image, Signal Process. Artif. Vis., 2021, pp. 1–5, doi: 10.1109/SSPA.2021.959199.

C. J. Wang, C.-K. Wen, S.-H. L. Tsai, and S. Jin, “Phase retrieval with learning unfolded expectation consistent signal recovery algorithm,” IEEE Signal Process. Lett., vol. 27, pp. 780–784, Apr. 2020, doi: 10.1109/LSP.2020.2990767.

C.-J. Wang, C.-K. Wen, S.-H. Tsai, S. Jin, and G. Y. Li, “Phase retrieval using expectation consistent signal recovery algorithm based on hypernetwork,” IEEE Trans. Signal Process., vol. 69, pp. 5770–5783, Oct. 2021, doi: 10.1109/TSP.2021.3118494.