Physics-based Iterative Projection Complex Neural Network for Phase Retrieval in Lensless Microscopy Imaging

Feilong Zhang
Harbin institute of technology
flzhang.hit@gmail.com

Xianming Liu*
Harbin institute of technology
csxm@hit.edu.cn

Cheng Guo
Harbin institute of technology
guocheng_27@163.com

shiyi lin
Harbin institute of technology
shiyi.lin.hit@outlook.com

Junjun Jiang
Harbin institute of technology
jiangjunjun@hit.edu.cn

Xiangyang Ji
Tsinghua University
xyji@tsinghua.edu.cn

Abstract

Phase retrieval from intensity-only measurements plays a central role in many real-world imaging tasks. In recent years, deep neural networks based methods emerge and show promising performance for phase retrieval. However, their interpretability and generalization still remain a major challenge. In this paper, we propose to combine the advantages of both model-based alternative projection method and deep neural network for phase retrieval, so as to achieve network interpretability and inference effectiveness simultaneously. Specifically, we unfold the iterative process of the alternative projection phase retrieval into a feed-forward neural network, whose layers mimic the processing flow. The physical model of the imaging process is then naturally embedded into the neural network structure. Moreover, a complex-valued U-Net is proposed for defining image priori for forward and backward projection in dual planes. Finally, we designate physics-based formulation as an untrained deep neural network, whose weights are enforced to fit to the given intensity measurements. In summary, our scheme for phase retrieval is effective, interpretable, physics-based and unsupervised. Experimental results demonstrate that our method achieves superior performance compared with the state-of-the-arts in a practical phase retrieval application—lensless microscopy imaging.

1. Introduction

Optical detectors in practical imaging systems, such as CCD and CMOS, can measure only the magnitude (brightness or intensity) of an electromagnetic wave. Phase retrieval (PR), also referred to as phase imaging, is to recover the lost phase information at the detection process from intensity-only measurements. It plays a crucial role in many applications in science and engineering, including computational microscopy [8, 27, 10], quantitative phase imaging [11], X-ray crystallography [16], computer-generated holography [20, 3] and many more.

Since the optical phase cannot be directly measured by an electronic detector, algorithmic phase retrieval comes into play. Over the years, many approaches have been suggested for solving the phase retrieval problem [22]. The most popular kind of phase retrieval methods are based on alternative projection, which was originated by Gerchberg and Saxton [6] and extended by Fienup [4, 5]. The alternative projection algorithm aims to recover a complex image from magnitude measurements at two different planes—the sample object plane and the camera sensor plane. It starts from a random initial estimate and proceeds by iteratively applying dual-plane projections: at each iteration, the current estimate is projected to the image plane so that the magnitude of its frequency spectrum matches the observations; the signal is then projected to the object plane to conform to some priori knowledge about its structure.

An alternative class of approaches are to develop algorithms of phase retrieval by using the tools of modern optimization theory. Phase retrieval is a NP-hard problem. From the perspective of combinatorial optimization, semidefinite programming (SDP) is a powerful tool to solve this problem. Candes *et al.* [2] formulate phase retrieval as a matrix completion problem, which is addressed by combining multiple structured illuminations together with ideas from convex programming. Waldspurger *et al.* [25] cast the phase retrieval problem as a non-convex quadratic program over a complex phase vector, and formulate a tractable relaxation similar to the classical MaxCut semidefinite program. However, the computational complexity of SDP based algorithms increases sharply with the signal dimen-

*Corresponding author

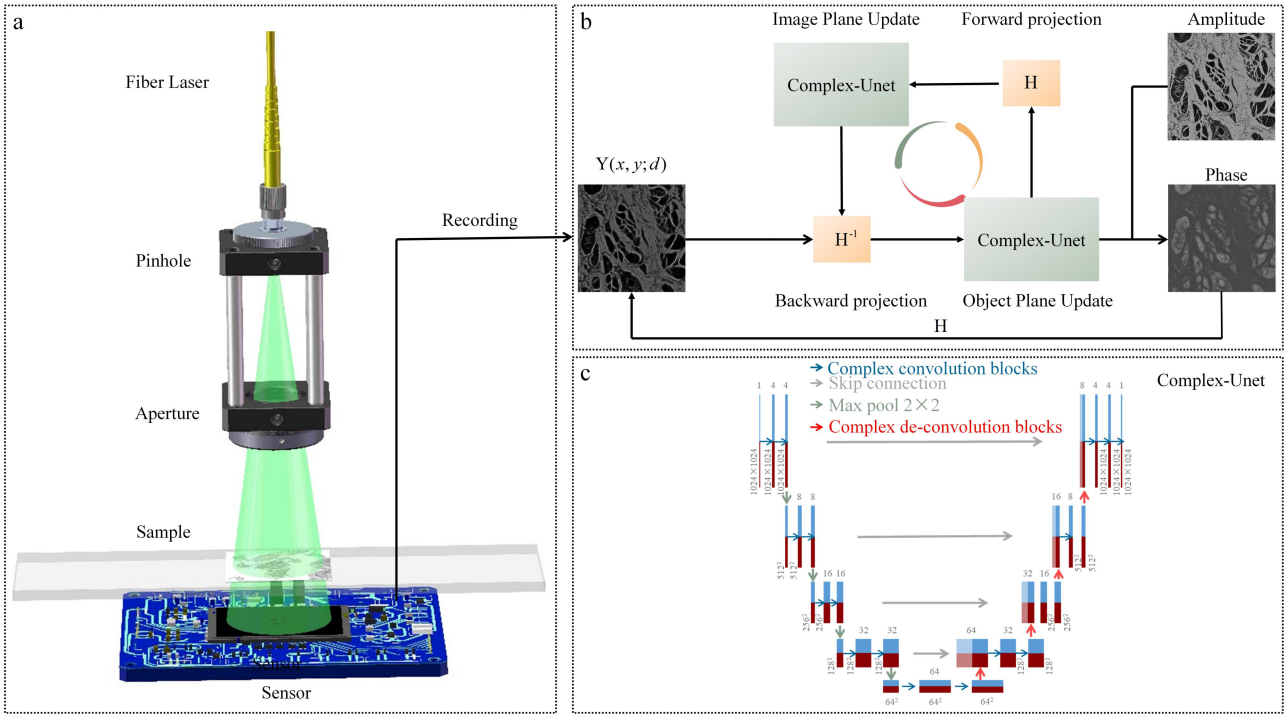

Figure 1. Schematic diagram of the imaging process. (a) The set-up of lensless microscopy imaging. (b) The structure diagram of the proposed neural network. (c) The architecture of the 2D complex-valued U-Net. The blue block represents the real part of the signal, and the red block represents the imaginary part of the signal.

sionality. Moreover, it requires a very large amount of memory for computation. Therefore, in practical applications, algorithms along this line are usually applied to one-dimensional signals.

In recent years, deep learning approaches for phase retrieval are becoming increasingly popular. The pioneer method to apply deep learning for phase retrieval was proposed by Rivenson *et al.* [21], which performs phase retrieval in an end-to-end fashion. In [14], Metzler *et al.* propose to leverage the regularization-by-denoising framework and a convolutional neural network based denoiser for phase retrieval. Hyder *et al.* [9] propose to learn a reference signal by using a small number of training images, which is further used to assist solving the Fourier phase retrieval problem by an unrolled network from gradient descent. The above methods all work in a supervised manner, in which the deep neural networks are trained by minimizing a loss function between the ground-truth and the observed measurements. However, in many imaging scenarios, it is difficult or even impossible to obtain sufficient numbers of ground-truth images for training. Moreover, these methods do not consider the imaging physics, making the networks required to learn both the physical measurement formation and the phase reconstruction process. This would cause difficulty in network training.

The emergence of deep image prior [24] yields new insights to develop unsupervised deep phase retrieval method. The deep phase decoder algorithm proposed in [1] reconstructs phase via minimizing the Euclidean distance between the measured intensity images and the hypothetical ones generated by deep decoder network, inspired by the idea of employing untrained generative DNNs as prior models for images. Similarly, Wang *et al.* [26] propose to recover phase with an untrained neural network, whose parameters are derived by minimizing the loss between the measured intensity and the image projected from the network output to the input plane through the imaging model. In these two methods, the network structures are directly borrowed from some popular ones, *e.g.*, deep decoder in [1] and U-Net in [26]. Both methods claim that the imaging model is incorporated, which however only works through the loss function but does not affect the network design.

Despite deep learning based methods have shown promising performance for the phase retrieval problem, their interpretability and generalization remain a major challenge, which is still needed for further investigation. In this paper, we propose an effective, interpretable, unsupervised, physics-based, complex-valued deep neural network for phase retrieval. Our method enjoys the following merits:

- Our network design is inspired by the well-known alternative projection phase retrieval algorithm. We unfold the iterative process of the alternative projection

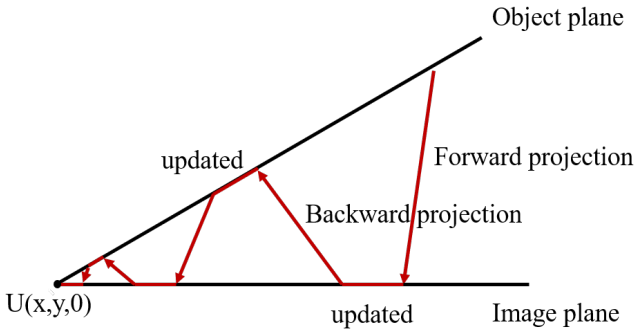

Figure 2. The principle of alternative projection method. The signal is projected back and forth between the object plane and the image plane. Before each projection, the signal will be updated according to the prior information in the domain to continuously approximate the ground truth.

method into a feed-forward neural network, whose layers mimic the processing flow. This makes our network interpretable and easily understood.

- The physical model of the imaging process is embedded into both the neural network structure and the loss function. The latter makes the network training without any labeled data. Compared with [26], our dual physics embedding strategy let network inference easier, and thus requires fewer amount of network parameters.
- A complex-valued U-Net is proposed for implicitly defining image priori in object and image plane update, without requiring additional hardware equipment or multiple measurements as done in original GS algorithm and its variants [6, 29, 18, 7].
- To the best of our knowledge, the proposed method is the first one in the literature that leverages the complex-valued neural network and dual physics embedding strategy to address phase retrieval problem. Our method achieves superior performance compared with the state-of-the-arts in a practical phase retrieval application—lensless microscopy imaging.

2. Preliminaries

In this section, we will first introduce the imaging physics for lensless microscopy imaging, and then revisit the classical phase retrieval (PR) by alternative projections, which inspires our deep unfolded neural network for phase retrieval.

2.1. Imaging Physics for Lensless Microscopy

The system of lensless microscopy imaging is shown in Fig. 1 (a). Define $\boldsymbol{U}(x, y; 0) \in \mathbb{C}^n$ as the biological sample illuminated by a quasi-monochromatic wave. The light field at a distance of d from the biological sample can be represented as

$$\begin{aligned}\boldsymbol{U}(x, y; d) &= \mathcal{F}^{-1}\big(\mathcal{F}(\boldsymbol{U}(x, y; 0)) \odot \boldsymbol{G}(f_x, f_y)\big) \\ &= \mathrm{H}\left(\boldsymbol{U}(x, y; 0)\right)\end{aligned} \tag{1}$$

where $\mathcal{F}$ represents the discrete Fourier transform and $\mathcal{F}^{-1}$ represents its inverse transform; $\odot$ represents the product of the corresponding elements of the matrix; (f_x, f_y) is the frequency domain coordinates. $\boldsymbol{G}(f_x, f_y)$ is the transmission matrix defined by the angular spectrum representation:

$$\boldsymbol{G}(f_x, f_y) = \exp\left[\boldsymbol{i}kd\sqrt{1 - (\lambda f_x)^2 - (\lambda f_y)^2}\right] \tag{2}$$

where $\boldsymbol{i}$ is the imaginary unit; d is the distance between the sample and the camera; λ is the wavelength of light; k is the wave number.

What the camera receives is the intensity information, and the propagation equation between light fields is established on the amplitude, so the image captured by the camera can be expressed as

$$\boldsymbol{Y}(x, y; d) = |\boldsymbol{U}(x, y; d)|^2 = |\mathrm{H}\left(\boldsymbol{U}(x, y; 0)\right)|^2 \tag{3}$$

where $\boldsymbol{Y}(x, y; d) \in \mathbb{R}$ is the intensity measurement. The reconstructed light field can be estimated by solving the following optimization problem

$$\hat{\boldsymbol{U}}(x, y, 0) = \underset{\boldsymbol{U}(x,y;0)}{\arg\min} \|\boldsymbol{Y}(x, y; d) - |\mathrm{H}(\boldsymbol{U}(x, y; 0))|^2\|_2^2 \tag{4}$$

After deriving the complex-valued $\hat{\boldsymbol{U}}(x, y; 0)$, the phase image is determined as its argument.

2.2. Classical Alternative Projection Method for PR

In the literature, the most famous method for PR is the alternative projection algorithm, which was proposed by Gerchberg and Saxton [6], later generalized by Fienup [4, 5]. In the following, we briefly introduce the process of alternative projection method. As illustrated in Fig. 2, it starts with a random initial estimate $\boldsymbol{U}^{(0)}(x, y, 0)$ and then performs the following steps iteratively. For instance, in the k-th iteration:

- **Forward Projection.** Project the signal to the image plane using angular spectrum representation:

$$\boldsymbol{U}^{(k)}(x, y; d) = \mathrm{H}(\boldsymbol{U}^{(k-1)}(x, y; 0)) \tag{5}$$

- **Image Plane Update.** It is generally considered that the intensity recorded by the image sensor is the true intensity of the light field of the image plane. Therefore, the captured intensity $\boldsymbol{Y}(x, y; d)$ is used to replace the amplitude of $\boldsymbol{U}^k(x, y; d)$ that is projected

onto the image plane:

$$\boldsymbol{I}^{(k)}(x,y;d) = \sqrt{\boldsymbol{Y}(x,y;d)}\cdot\exp[\boldsymbol{i}\cdot\text{angle}(\boldsymbol{U}^{(k)}(x,y;d))] \tag{6}$$

where $\text{angle}(\cdot)$ represents the operation of calculating the phase.

- **Backward Projection.** After the signal on the image plane is updated, it becomes closer to the true value of the light field actually transmitted to the image sensor. Then, the updated signal is projected onto the object plane:

$$\boldsymbol{O}^{(k)}(x,y;0) = \text{H}^{-1}(\boldsymbol{I}^{(k)}(x,y;d)) \tag{7}$$

where $\text{H}^{-1}(\cdot)$ represents the inverse transform of $\text{H}(\cdot)$, which is conducted by setting $d = -d$ in (2) since the propagation of light is reversible.

- **Object Plane Update.** Finally, we update the signal on object plane by using some priori knowledge, which can make it further closer to the ground truth. There have various strategies of obtaining the object plane priori been proposed in the literature. This is the main difference among various iteration projection methods. For instance, the Gerchberg-Saxton (GS) algorithm [6] requires to know the amplitude of the object plane, which is captured by an additional microscope; Fourier ptychographic microscopy (FPM) proposed by Zheng *et al.* [29] uses LED arrays to obtain multi-angle images as the spectrum prior; another popular approach for object prior is to employ multiple observation planes with various defocused distances [18, 7], which is referred to as multi-distance phase retrieval (MDPR). Here we take the GS algorithm as an example, in which the object plane update is formulated as

$$\boldsymbol{U}^{(k+1)}(x,y;0) = \boldsymbol{P}(x,y;0)\cdot\exp[\boldsymbol{i}\cdot\text{angle}(\boldsymbol{O}^{(k)}(x,y;0))] \tag{8}$$

where $\boldsymbol{P}(x,y;0)$ represents the captured object plane amplitude by another microscope.

The above process is iteratively performed until the stop condition is arrived. Finally, we can obtain the recovered complex signal $\boldsymbol{U}(x,y;0)$, which includes both amplitude and phase information.

From the above description, it can be found that the step of object plane update plays a crucial role in alternative projection based phase retrieval, while in which the core is the definition of object plane priori. Although the alternative projection based PR algorithm works well in practical applications, which requires additional hardware equipment to obtain reliable constraints, such as another microscope or LED arrays in [6, 29].

3. Proposed Method

In this work, we attempt to unfold the iterative process into a feed-forward neural network, whose layers mimic the processing flow of the alternative projection based PR method, while without introducing additional hardware or multiple measurements. To achieve this goal, three major problems should be carefully considered: 1) Existing DNN-based methods solve the PR problems by directly mapping observed amplitude to desirable phase [21, 14, 9, 13], in which the physics characterizing the imaging processes is ignored. 2) The imaging model involves Fourier transform, which produces complex values, making traditional real-valued deep network not applicable any more. 3) Current deep learning based approaches for PR rely on supervised learning with a large amount of training examples [21, 14, 9, 13]. However, for many practical optical imaging scenarios, the ground-truth is unknown and thus it is impossible to collect pairs of data for end-to-end learning.

As responses to the above challenges, we propose a complex-weighted deep neural network to unroll iterative projection for phase retrieval, encapsulating the image prior and imaging physics into network design. Our approach works in an unsupervised manner, which reconstructs the sought phase through fitting the weights of the network to the captured intensity measurements by incorporating imaging physics into the loss function. Corresponding to the classical iterative projection for PR, the proposed neural network is composed of four modules: backward projection, object plane updated, forward projection, and image plane updated. And the backward projection and object plane updated will be performed two times for one iteration. In the following, we will first introduce the imaging model represented by complex-valued neural networks, and then elaborate these four modules.

3.1. 2D Complex-valued U-Net

In the proposed method, we aim to combine the advantages of both model-based alternative projection method and deep neural network so as to achieve network interpretability and inference effectiveness simultaneously. In the iterative projection method, imaging physics knowledge is explicitly built into the projection constraints. However, as stated in Section 2, there are complex operations involved, making the commonly used real-valued neural network inapplicable. We propose to use complex-valued neural network [23] for our purpose, and extend the well-known real-valued U-Net to a complex-valued version for 2D images.

As illustrated in Fig. 1 (c), the complex-valued U-Net (CVU-Net) shares the same structure as the real-valued U-Net, which consists of four main components: convolution blocks (3×3 complex convolution, complex batch normalization, complex leaky ReLU), max pooling blocks (2×2),

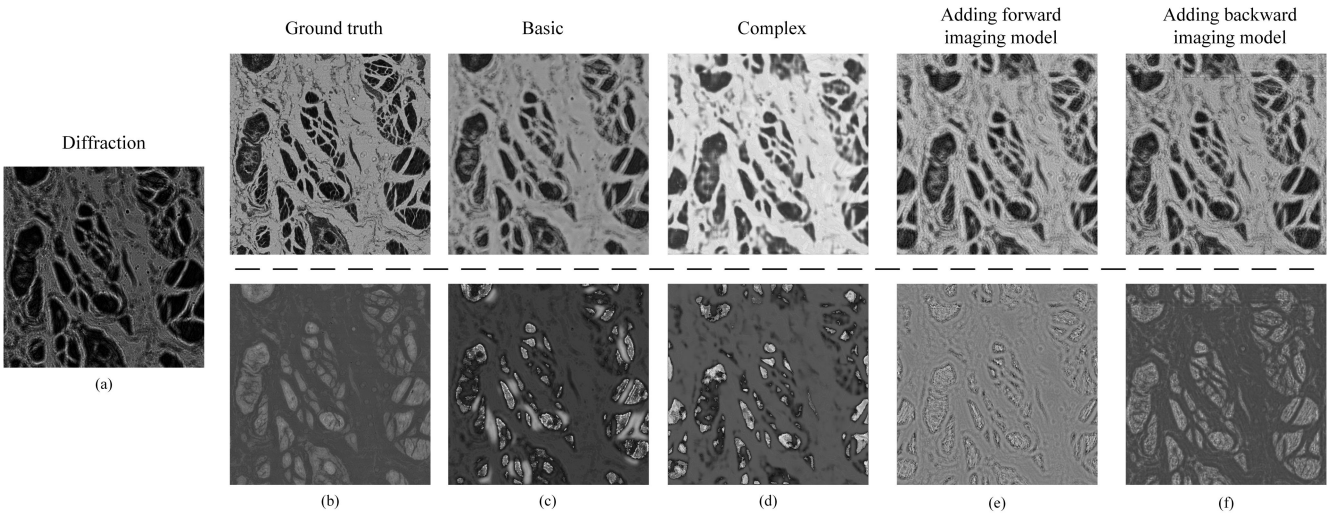

Figure 3. Ablation study. (a) show the diffraction image captured by image sensor. (b) is the ground truth. The first line shows the amplitude result, the second line shows the phase result.

up-convolution blocks (up-sampling, 3×3 complex convolution, complex batch normalization, complex leaky ReLU) and skip connection blocks.

In the following, we define these complex-valued building blocks. Given a complex-valued convolutional filter $\boldsymbol{F} = \boldsymbol{F}_r + \boldsymbol{i}\boldsymbol{F}_i$, where $\boldsymbol{F}_r$ and $\boldsymbol{F}_i$ are two real-valued matrices, the complex convolution operation on a complex input $\boldsymbol{h}$ and $\boldsymbol{F}$ can be done by

$$\begin{aligned} \boldsymbol{F} * \boldsymbol{h} &= (\boldsymbol{F}_r + \boldsymbol{i}\boldsymbol{F}_i) * (\boldsymbol{h}_r + \boldsymbol{i}\boldsymbol{h}_i) \\ &= (\boldsymbol{F}_r * \boldsymbol{h}_r - \boldsymbol{F}_i * \boldsymbol{h}_i) + \boldsymbol{i}(\boldsymbol{F}_r * \boldsymbol{h}_i + \boldsymbol{F}_i * \boldsymbol{h}_r) \end{aligned} \tag{9}$$

The complex convolutions can be conducted by two different real-valued convolution operations with shared weights [23]. In real-valued neural network, ReLU is widely used. However, for a complex signal, the case usually happens that its real or imaginary part is negative. Accordingly, we propose to leverage complex leaky ReLU. Moreover, for complex-valued activation function, it has been demonstrated that the separation of real and imaginary parts would bring benefits for some tasks, such as classification [23]. Therefore, in our scheme, the complex leaky ReLU is implemented as

$$\mathbb{C}LeReLU(\boldsymbol{z}) = LeReLU(\boldsymbol{z}_r) + \boldsymbol{i}LeReLU(\boldsymbol{z}_i) \tag{10}$$

where $\boldsymbol{z} = \boldsymbol{z}_r + \boldsymbol{i}\boldsymbol{z}_i$ represents the input signal to the activation layer; $LeReLU(\cdot)$ is the leaky ReLU operator. For the complex-valued batch normalization, similar to the complex activation function, we perform real-valued batch normalization on the real and imaginary parts respectively.

3.2. Deep Unfolded Neural Network

In the following, we introduce in detail the proposed method, which unfolds the inference process of the alternative projection method into a layer-wise structure analogous to a neural network. The resulting formula combines the representation ability of deep network with the internal structure of the model-based approach, allowing inference to be performed in a fixed number of four-layer units that can be optimized for best performance. Specifically, our alternative projection network starts from the image plane, setting $\boldsymbol{I}^{(0)}(x, y; d) = \sqrt{\boldsymbol{Y}(x, y; d)}$, several six-layers units are progressively performed. In the k-th unit, it contains the following six layers:

- **Backward Projection Layer.** The input signal of this layer is the estimate $\boldsymbol{I}^{(k-1)}(x, y; d)$ of the image plane in last unit. In this layer, we use the angular spectrum representation to project it onto the object plane. The output signal can be obtained by:

$$\begin{aligned} \boldsymbol{O}_r^{(k)}(x, y; 0) &= \mathrm{Re}(\mathrm{H}(I^{(k-1)}(x, y; d))) \\ \boldsymbol{O}_i^{(k)}(x, y; 0) &= \mathrm{Im}(\mathrm{H}(I^{(k-1)}(x, y; d))) \end{aligned} \tag{11}$$

  where $Re(\cdot)$ represents to get the real part of a complex-valued signal; $Im(\cdot)$ represents to get the imaginary part of a complex-valued signal. $\boldsymbol{O}_r^{(k)}(x, y; 0)$ and $\boldsymbol{O}_i^{(k)}(x, y; 0)$ are the real and imaginary parts of $\boldsymbol{O}^{(k)}(x, y; 0)$.

  It is worth noting that the signal of this layer is the real physics-based process of imaging, so there is no convolution operations involved. But it still participates in back propagation to update the network parameters of other layers.

- **Object Plane Update Layer.** The purpose of this layer is to update the estimate of $\boldsymbol{U}(x, y; 0)$ according to the

estimate $\boldsymbol{O}^{(k)}(x, y; 0)$ calculated at the backward projection layer. This is actually a signal restoration problem. Instead of explicitly defining the prior using additional hardware [6, 29], multiple-plane measurements [8] or hand-crafted heuristics [15, 19], we do it implicitly in a learning-based fashion through the developed CVU-Net in Section 3.1:

$$\boldsymbol{U}^{(k)}(x, y; 0) = \text{CVUNet}\left(\boldsymbol{O}^{(k)}(x, y; 0), \boldsymbol{\theta}\right) \tag{12}$$

where $\boldsymbol{\theta}$ is the network parameters. It means that the network itself acts as a priori model.

- **Forward Projection Layer.** The input signal of this layer is the updated estimate $\boldsymbol{U}^{(k)}(x, y; 0)$ in the object plane, which is further projected back to the image plane using angular spectrum representation. The output of this layer is defined as:

$$\begin{aligned} \boldsymbol{U}_r^{(k)}(x, y; d) &= \text{Re}(\text{H}(\boldsymbol{U}^{(k)}(x, y; 0))) \\ \boldsymbol{U}_i^{(k)}(x, y; d) &= \text{Im}(\text{H}(\boldsymbol{U}^{(k)}(x, y; 0))) \end{aligned} \tag{13}$$

  Similar to the backward projection layer, this layer does not perform any convolution operations, but just participates in calculating the gradient in back propagation.

- **Image Plane Update Layer.** The input of this layer the updated estimate $\boldsymbol{U}^{(k)}(x, y; d)$. Similar to the object plane update layer, the update process of this layer can be formulated as a signal recovery problem, which is addressed by the proposed CVU-Net:

$$\boldsymbol{I}^{(k)}(x, y; d) = \text{CVUNet}\left(\boldsymbol{U}^{(k)}(x, y; d), \boldsymbol{\phi}\right) \tag{14}$$

  where $\boldsymbol{\phi}$ is the network parameters. And the last two layers are the same as the first two layers, the overall framework of the network is shown in the Fig. 1 (b).

Loss function: To train the proposed deep unfolded complex-valued neural network, we no longer take the commonly used end-to-end training manner to directly map given intensity measurements to phase, considering the difficulty of collecting dense training data in lensless microscopy imaging. Instead, we propose to specify physics-based formulation as an untrained deep neural network, whose weights are enforced to fit to the given intensity measurements. Specifically, we propagate the network output back to the object plane through the angular spectrum representation, and use the difference of its intensity values to form the loss function. We define the loss function as:

$$\boldsymbol{W}^* = \arg\min_{\boldsymbol{W}} \left\| |\text{H}^{-1}(\mathcal{N}_{\boldsymbol{W}}(\boldsymbol{Y}))| - \boldsymbol{Y} \right\|^2 \tag{15}$$

where $\mathcal{N}_{\boldsymbol{W}}(\cdot)$ represents our proposed network and $\boldsymbol{W} = \{\boldsymbol{\theta}, \boldsymbol{\phi}\}$ are the network parameters. Since the ground-truth phase is not involved in the loss function, our method works in an unsupervised learning manner.

3.3. Implementation Details

The neural network is implemented based on the Pytorch version 1.7.0 platform using python 3.8. We adopt the Adam optimizer [12] with a learning rate of 0.01 to optimize the weights in the networks. And added uniformly distributed noise between 0 and 0.03 to the fixed input I in every optimization step to achieve better convergence. In our network, the image size is set as 1024 and the image sensor size is set as 1.34 μm. The number of iterations is 200. The computer we use is with a CPU E5-2620 V4, 64GB of RAM, and NVIDIA GTX 1080TI.

4. Experiments

In this section, we provide experimental results on real imaging task—lensless microscope imaging—to demonstrate the effectiveness of our method. In the following, we first introduce the experimental setup and then compare the proposed model with state-of-the-art algorithms, finally provide the ablation analysis.

4.1. Experimental Setup and Datasets

In order to validate the performance of the proposed algorithm, we build a typical lensless microscope as shown in Fig. 1 (a). In order to increase the numerical aperture (NA) of the device, the package of the camera is removed to reduce the distance between the sample and image sensor. A diverging spherical wave is emitted by a fiber laser and transformed into parallel light by a collimating lens. The shaped plane wave illuminates the specimen to produce the diffraction images with the distance d. The receiving device is a CMOS sensor (Sony, IMX206, $1.34\mu\text{m} \times 1.34\mu\text{m}$) on a precision linear stage under the laser wavelength of 532nm.

4.2. Compare with the State-of-the-arts

Since the network proposed in this article is unsupervised, for fair comparison, we include three representative unsupervised methods for comparison study:

- The GS algorithm with a support constraint [17]. The GS algorithm is the origin of the phase retrieval algorithm and the iterative projection strategy in this paper is inspired by this work.

- The compressed sensing based method proposed by Zhang *et al.* [28]. Compressed sensing is a useful tool for solving engineering optimization problems before the advent of deep learning.

- The deep learning based unsupervised method PhyseNet proposed by Wang *et al.* [26]. Because PhyseNet can only handle pure phase samples, for fair comparison, we replace the U-net used in their framework with a complex U-net. Although our network has

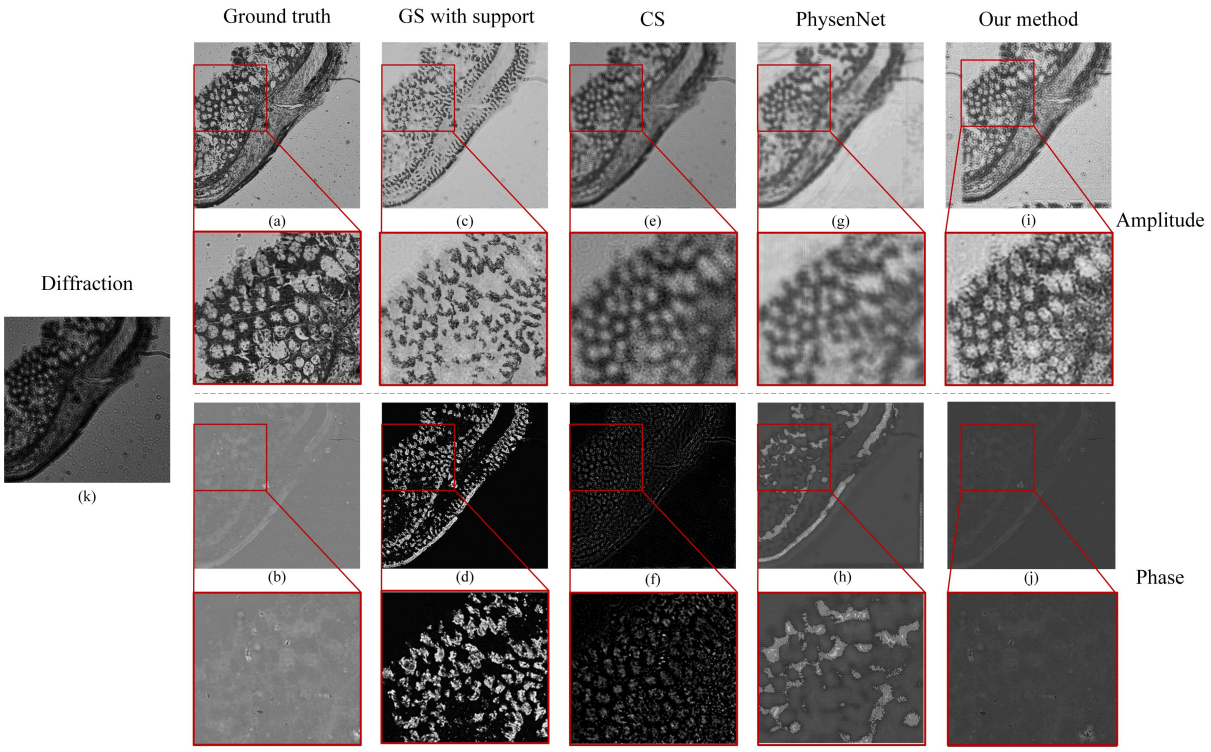

Figure 4. **Visual comparisons on the H&E stained pathological slides of rat intestine.** (a), (b) are the ground truth. (c)-(j) is the comparison result of our method and other methods. The first two lines are the amplitude comparison results, and the last two lines are the phase comparison results. The second row is an enlarged image of the position of the red box corresponding to the first row, and the fourth row is an enlarged image of the position of the red box corresponding to the third row. (k) is the diffraction pattern recorded by image sensor.

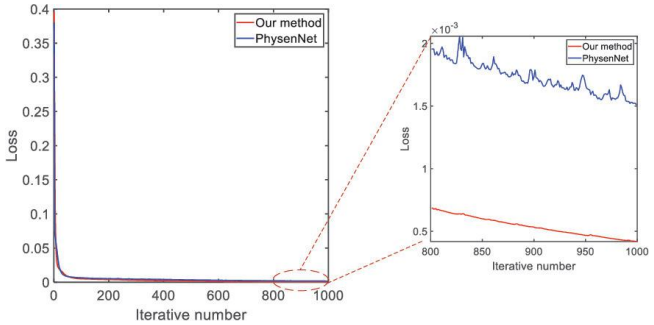

Figure 5. Convergence speed comparison.

three complex U-nets, the overall amount of parameters is fewer than that of PhyseNet, since the channel number of our U-net is only $1/4$ of PhyseNet.

The experimental results are shown in the Fig. 4. (k) is the collected diffraction pattern, and (a), (b) is the ground truth, which is calculated by collecting an $8\times5\times5$ volume data. (c) and (d) are the results of the GS algorithm with the support constraint [17]. (c) is the amplitude result, which likes a sketch of the ground truth, with a lot of details missing, while (d) is the phase result. Compared with (b), it looks like the result of binarization. (e) and (f) are the results based on the compressive sensing algorithm [28], (e) is the amplitude result, which has many artifacts. (g) and (h) are generated by the unsupervised deep learning method [26]. (g) is the amplitude result, judging from its enlarged image in the red box, and it looks like a denoised version of (e) and is much clearer than (e). However, (g) still loses a lot of details when compare with the phase result (h), which has a lot of overexposed points. (i) and (j) are the results of our method, the amplitude result (i) is the best among these methods. Its resolution is much higher than other methods. Nevertheless there is still a gap when compared with the ground truth (a), which is calculated by sampling $8\times5\times5$ volume data. Compared with the ground truth, the phase result (j) of the propped methods is only slightly darker. Other comparison results can be found in the supplementary materials. Similar visual results can be observed for other samples , as illustrated in Fig. 6.

For the convenience of analysis, we give the convergence curve of the proposed algorithm and PhysenNet. The number of iterations is 1000. It can be seen from the Fig. 5 that our algorithm is always better than PhysenNet in the iteration process.

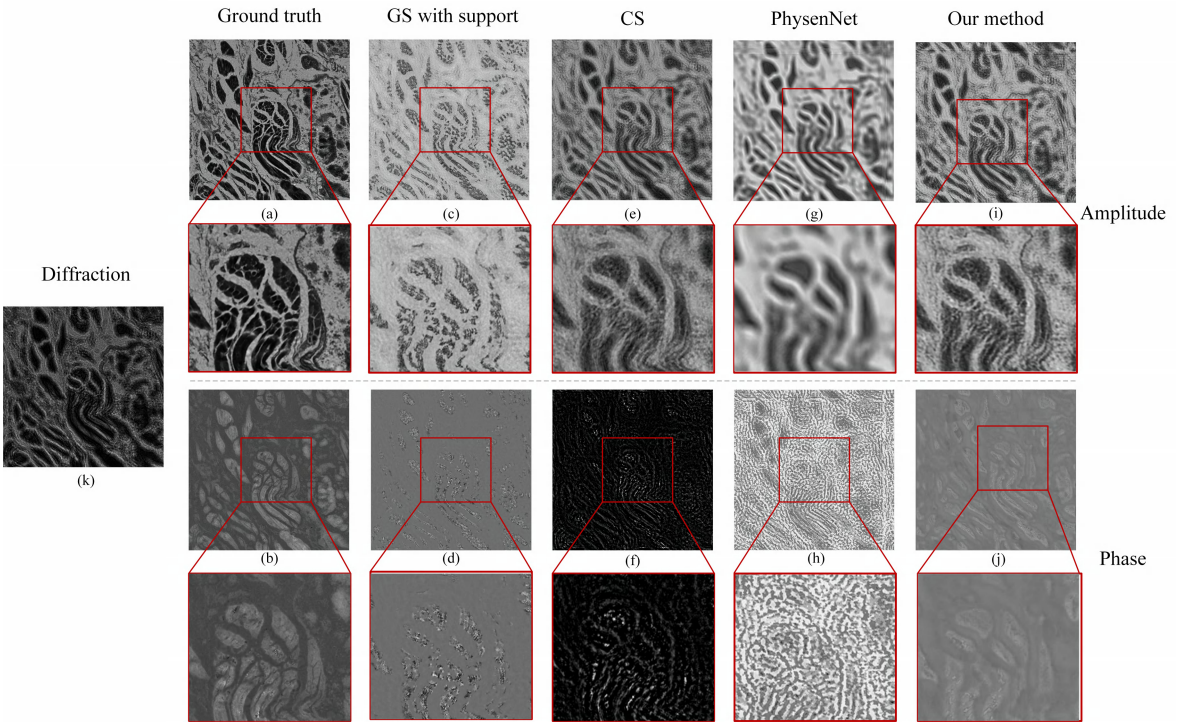

Figure 6. **Visual comparisons on the H&E stained pathological slides of human esophagus cancer cell.**

4.3. Ablation Study

In this section, we conduct ablation studies to analyze the effectiveness of each component in the proposed framework. Both quantitative and qualitative results on the lensless imaging microscopy system are reported for three variants of our methods where each component is gradually added: 1) only including the physics-based loss function; 2) complex-valued network; 3) adding forward imaging model in the network; 4) unfolding the network into the form of alternative projection.

The experiments are shown in the Fig. 3, (a) is the diffraction pattern captured by image sensors, (b) is the ground truth, calculated by collecting an $8\times5\times5$ three-dimensional volume data. The method of (c) used is the basic version of our network, with only the physics-based loss function. The method of (d) used is the complex-valued version of U-net. From (d), it can be seen that the result of amplitude reconstruction is a bit like overexposed, there are some speckles similar to water stains, and the phase is almost not retrieved. The algorithm of (e) used is based on (d). A physics-based forward imaging model is added to the network. From (e) it can be seen that after adding the forward imaging model, the reconstructed amplitude part has been significantly improved. But the phase result still has a big gap with the true value. The method of (f) used is based on (e), adding the backward imaging model, and the corresponding object plane update module and image plane update module. It can be seen that with the addition of the backward imaging model, the retrieved phase has been significantly improved. The amplitude image has also been improved, but the comparison is not so obvious. From these comparison results, we can see the advantages of combining traditional algorithms with deep learning methods.

5. Conclusion

In this paper, we presented an unsupervised complex-valued deep neural network for phase retrieval, which is interpretable and physics-based. Specially, a complex-weighted deep neural network is tailored to unroll iterative projection for phase retrieval, thus encapsulating the image prior and imaging physics into network design. Our approach works in an unsupervised manner, which reconstructs the sought phase through fitting the weights of the network to the captured intensity measurements by incorporating imaging physics into the loss function. Experimental results on lensless microscopy imaging demonstrate the effectiveness of our method.

6. Acknowledgement

This work was supported by National Natural Science Foundation of China under Grants 61922027 and 61827804, and National Key Research and Development Project under Grant 2019YFE0109600.

References

[1] Emrah Bostan, Reinhard Heckel, Michael Chen, Michael Kellman, and Laura Waller. Deep phase decoder: self-calibrating phase microscopy with an untrained deep neural network. *Optica*, 7(6):559–562, Jun 2020.

[2] Emmanuel J Candes, Thomas Strohmer, and Vladislav Voroninski. Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming. *Communications on Pure and Applied Mathematics*, 66(8):1241–1274, 2013.

[3] P. Chakravarthula, Y. Peng, J. S. Kollin, H Fuchs, and F. Heide. Wirtinger holography for near-eye displays. *ACM Transactions on Graphics (TOG)*, 2019.

[4] James R Fienup. Reconstruction of an object from the modulus of its fourier transform. *Optics letters*, 3(1):27–29, 1978.

[5] James R Fienup. Phase retrieval algorithms: a comparison. *Applied optics*, 21(15):2758–2769, 1982.

[6] Ralph W Gerchberg. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik*, 35:237–246, 1972.

[7] Alon Greenbaum and Aydogan Ozcan. Maskless imaging of dense samples using pixel super-resolution based multi-height lensfree on-chip microscopy. *Optics express*, 20(3):3129–3143, 2012.

[8] Cheng Guo, Xianming Liu, Xingchi Kan, Feilong Zhang, Jiubin Tan, Shutian Liu, and Zhengjun Liu. Lensfree on-chip microscopy based on dual-plane phase retrieval. *Opt. Express*, 27(24):35216–35229, Nov 2019.

[9] Rakib Hyder, Zikui Cai, and M. Salman Asif. Solving phase retrieval with a learned reference. In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm, editors, *Computer Vision – ECCV 2020*, pages 425–441, Cham, 2020. Springer International Publishing.

[10] Michael Kellman, Kevin Zhang, Eric Markley, Jon Tamir, Emrah Bostan, Michael Lustig, and Laura Waller. Memory-efficient learning for large-scale computational imaging. *IEEE Transactions on Computational Imaging*, 6:1403–1414, 2020.

[11] M. R. Kellman, E. Bostan, N. A. Repina, and L. Waller. Physics-based learned design: Optimized coded-illumination for quantitative phase imaging. *IEEE Transactions on Computational Imaging*, 5(3):344–353, 2019.

[12] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[13] Raunak Manekar, Kshitij Tayal, Vipin Kumar, and Ju Sun. End to end learning for phase retrieval. *ICML workshop on ML Interpretability for Scientific Discovery*.

[14] Christopher Metzler, Phillip Schniter, Ashok Veeraraghavan, and richard baraniuk. prDeep: Robust phase retrieval with a flexible deep network. volume 80 of *Proceedings of Machine Learning Research*, pages 3501–3510, Stockholmsmässan, Stockholm Sweden, 10–15 Jul 2018. PMLR.

[15] Artem Migukin, Vladimir Katkovnik, and Jaakko Astola. Wave field reconstruction from multiple plane intensity-only data: augmented lagrangian algorithm. *J. Opt. Soc. Am. A*, 28(6):993–1002, Jun 2011.

[16] R. P. Millane. Phase retrieval in crystallography and optics. *J. Opt. Soc. Am. A*, 7(3):394–411, Mar 1990.

[17] Onur Mudanyali, Derek Tseng, Chulwoo Oh, Serhan O Isikman, Ikbal Sencan, Waheb Bishara, Cetin Oztoprak, Sungkyu Seo, Bahar Khademhosseini, and Aydogan Ozcan. Compact, light-weight and cost-effective microscope based on lensless incoherent holography for telemedicine applications. *Lab on a Chip*, 10(11):1417–1428, 2010.

[18] Giancarlo Pedrini, Wolfgang Osten, and Yan Zhang. Wave-front reconstruction from a sequence of interferograms recorded at different planes. *Optics letters*, 30(8):833–835, 2005.

[19] A. Pein, S. Loock, G. Plonka, and T. Salditt. Using sparsity information for iterative phase retrieval in x-ray propagation imaging. *Opt. Express*, 24(8):8332–8343, Apr 2016.

[20] Y. Peng, S. Choi, N. Padmanaban, J. Kim, and G. Wetzstein. Neural holography. In *SIGGRAPH '20: Special Interest Group on Computer Graphics and Interactive Techniques Conference*, 2020.

[21] Yair Rivenson, Yibo Zhang, Harun Günaydin, Da Teng, and Aydogan Ozcan. Non-iterative holographic image reconstruction and phase retrieval using a deep convolutional neural network. In *CLEO: Science and Innovations*, pages STh1J–3. Optical Society of America, 2018.

[22] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev. Phase retrieval with application to optical imaging: A contemporary overview. *IEEE Signal Processing Magazine*, 32(3):87–109, 2015.

[23] Chiheb Trabelsi, Olexa Bilaniuk, Ying Zhang, Dmitriy Serdyuk, Sandeep Subramanian, Joao Felipe Santos, Soroush Mehri, Negar Rostamzadeh, Yoshua Bengio, and Christopher J Pal. Deep complex networks. In *International Conference on Learning Representations*, 2018.

[24] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9446–9454, 2018.

[25] Irène Waldspurger, Alexandre d'Aspremont, and Stéphane Mallat. Phase recovery, maxcut and complex semidefinite programming. *Mathematical Programming*, 149(1-2):47–81, 2015.

[26] Fei Wang, Yaoming Bian, Haichao Wang, Meng Lyu, Giancarlo Pedrini, Wolfgang Osten, George Barbastathis, and Guohai Situ. Phase imaging with an untrained neural network. *Light: Science & Applications*, 9(1):1–7, 2020.

[27] Feilong Zhang, Cheng Guo, Yulan Zhai, Jiubin Tan, Shutian Liu, Cuimei Tan, Hang Chen, and Zhengjun Liu. A noise-robust multi-intensity phase retrieval method based on structural patch decomposition. *Journal of Optics*, 2020.

[28] Wenhui Zhang, Liangcai Cao, David J Brady, Hua Zhang, Ji Cang, Hao Zhang, and Guofan Jin. Twin-image-free holography: a compressive sensing approach. *Physical review letters*, 121(9):093902, 2018.

[29] Guoan Zheng, Roarke Horstmeyer, and Changhuei Yang. Wide-field, high-resolution fourier ptychographic microscopy. *Nature photonics*, 7(9):739–745, 2013.