Optics Letters

Fourier phase retrieval using physics-enhanced deep learning

Zike Zhang,[1,2,3] Fei Wang,[1,2,*] Qixuan Min,[1,2] Ying Jin,[1,2] and Guohai Situ[1,2,3,4,5]

[1]Wangzhijiang Innovation Center for Laser, Aerospace Laser Technology and System Department, Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Shanghai 201800, China
[2]Key Laboratory of Space Laser Communication and Detection Technology, Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Shanghai 201800, China
[3]Center of Materials Science and Optoelectronics Engineering, University of Chinese Academy of Sciences, Beijing 100049, China
[4]Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou 310024, China
[5]ghsitu@siom.ac.cn
*wangfei@siom.ac.cn

Received 26 July 2024; revised 15 September 2024; accepted 30 September 2024; posted 30 September 2024; published 23 October 2024

Fourier phase retrieval (FPR) aims to reconstruct an object image from the magnitude of its Fourier transform. Despite its widespread utility in various fields of engineering and science, the inherent ill-posed nature of the FPR problem poses a significant challenge. Here we propose a learning-based approach that incorporates the physical model of the FPR imaging system with a deep neural network. Our method includes two steps: First, we leverage the image formation model of the FPR to guide the generation of data for network training in a self-supervised manner. Second, we exploit the physical model to fine-tune the pre-trained model to impose the physics-consistency constraint on the network prediction. This allows us to integrate both implicit prior from training data and explicit prior from the physics of the imaging system to address the FPR problem. Simulation and experiments demonstrate that the proposed method is accurate and stable, showcasing its potential for wide application in fields utilizing the FPR. We have made our source code available for non-commercial use.

https://doi.org/10.1364/OL.537792

The task of recovering an object image from the (squared) magnitude of its Fourier transform, known as Fourier phase retrieval (FPR), is prevalent in significant fields such as materials science, biology, and astronomy [1,2]. The FPR problem is challenging due to the absence of phase information, which results in non-unique mapping between the object image and its phaseless Fourier transform measurements, rendering the problem ill-posed [3].

Taking multiple acquisitions, each with a different mask, to enhance the diversity of measurements is one of the primary methods used to mitigate the ill-posed nature of the inverse reconstruction problem in the FPR [4,5]. However, this method suffers from long data acquisition times and a complex imaging system. Incorporating constraints such as nonnegative, support, and sparse into the FPR solving process can reconstruct the object information from a single magnitude measurement. These algorithms include alternating projections [6] and compressive sensing [7], which, however, do not have guaranteed convergence. Although the semi-definite programming approach [8] offers strong convergence guarantees, it comes with significantly high computational complexity and often requires a random Gaussian sampling.

Recently, deep learning (DL) has also been used in the FPR. Several DL-based algorithms have been developed as FPR solvers [9,10]. These methods have garnered significant attention from researchers in the FPR field because they enable single-shot, high-quality, and non-iterative reconstructions. However, such data-driven image reconstruction algorithms suffer from significant limitations in terms of training data acquisition, generalization, and interpretability [11].

To address the limitations of traditional data-driven DL methods mentioned above, we have developed a physics-enhanced DL approach called PhysenNet [12]. This method integrates the physical model of the imaging system with neural networks and leverages physics-consistency loss to guide weight optimization, making it an untrained, scalable, and interpretable inverse solver. Moreover, by employing data-driven pre-training followed by physics-driven fine-tuning, this approach serves as a bridge between traditional model-based optimization algorithms and data-driven DL, capable of providing the best overall performance [13]. This method has been widely applied [14,15], including in the phase retrieval (FPR) problem in the Bragg coherent diffraction imaging for 3D reconstruction of single-particle structures [16].

In this work, we extend the concept of physics-driven fine-tuning to the field of the FPR, with a particular focus on the forward-scattering coherent diffraction imaging system that can be used for various sample types and applications [17]. Specifically, as shown in Fig. 1, our methodology involves initially training a neural network model in a self-supervised manner. Subsequently, we leverage a physics-consistency loss to refine the pre-trained model. Additionally, we implement a Bartlett window to minimize spectral leakage, thereby enhancing imaging performance. To validate our approach, we conducted

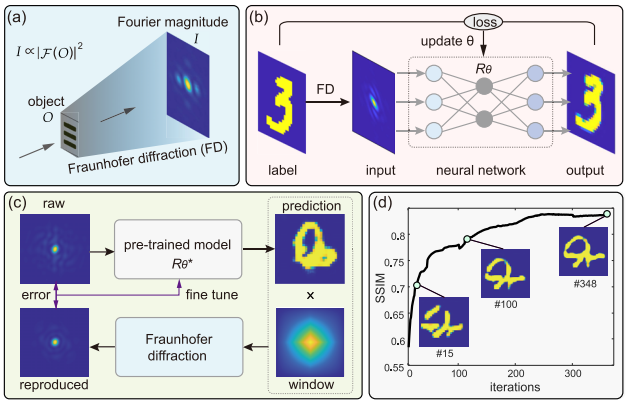

Fig. 1. Overview of the proposed physics-enhanced deep learning approach for the Fourier phase retrieval. (a) Schematic diagram of the Fourier phase retrieval. (b) Self-supervised pre-training process. (c) Physics-driven fine-tuning process. (d) Reconstruction results of out-of-domain test data at different fine-tuning steps.

experiments utilizing both simulations and real-world coherent diffraction imaging experiments. Our findings underscore the efficacy of our physics-enhanced DL approach, demonstrating its ability to accurately reconstruct images from a single Fourier magnitude pattern, even for instances that obey a different statistical distribution of that training set.

Now, let us delve into the technical intricacies of the proposed approach. According to the diffraction theory [18], as plotted in Fig. 1(a), when an object $O(\mathbf{x}')$ is illuminated by a light field $U_0(\mathbf{x}')$, the resulting field $U_d(x)$ after propagating a sufficiently long diffraction distance can be described by the following equation:

$$\begin{aligned} U_d(\mathbf{x}) &\propto \int O(\mathbf{x}')U_0(\mathbf{x}')P(\mathbf{x}')\exp\left(-i\frac{2\pi}{\lambda}\frac{\mathbf{x}'\mathbf{x}}{Z}\right)d\mathbf{x}' \\ &\propto \mathcal{F}\left\{O\left(\mathbf{x}'\right)U_0\left(\mathbf{x}'\right)P\left(\mathbf{x}'\right)\right\}_{\frac{2\pi}{\lambda z}\mathbf{x}'}, \end{aligned} \tag{1}$$

in which $U_0(x')$ can be uniformed when it is a plane wave. $P(\mathbf{x}')$ denotes the imaging field of view (FOV) at the object plane. All $\mathbf{x}'$ is the spatial coordinate at the original object plane, and $\mathbf{x}$ describes the Fourier measurement plane. As existing cameras are insensitive to phase information of the light wave, one can only detect the intensity of $U_d(x)$:

$$\begin{aligned} I\left(\mathbf{x}\right) &= H(O(\mathbf{x}')) = \left|U_d\left(\mathbf{x}\right)\right|^2 \\ &= C\cdot\left|\mathcal{F}\left\{O(\mathbf{x}')U_0\left(\mathbf{x}'\right)P\left(\mathbf{x}'\right)\right\}\right|^2, \end{aligned} \tag{2}$$

where $I(x)$ is the raw measured intensity image, $H(\cdot)$ denotes the image formation model, and C is a constant, with $\mathcal{F}$ representing the Fourier transform computation process. The FPR problem is to restore the object $O(\mathbf{x}')$ from the Fourier magnitude pattern $I(\mathbf{x})$ provided that the illumination $U_0(x')$ and the FOV $P(\mathbf{x}')$ are known.

Here we proposed a two-step approach to address the FPR problem. In the first step, as shown in Fig. 1(b), we train a neural network $\mathcal{R}_\theta$ to restore the object image O from its Fourier magnitude pattern I. We exploit a self-supervised approach to finish the training process. The objective function is as follows:

$$\theta_p^* = \arg\min_\theta ||\mathcal{R}_\theta(H(O^k)) - O^k||^2,\ \forall\ O^k \in S_T, \tag{3}$$

where the training set $S_T = \{O^k | k = 1, 2, 3, \ldots, K\}$ contains K digital images O^k that can be obtained from an open database. Here as a preliminary study, we simply take 10,000 images from the widely used MNIST for training. We exploit a U-Net-like network architecture to implement $\mathcal{R}_\theta$ (see Supplement 1). The network was trained with a batch size of 16 and a learning rate of 0.01. The training process lasted for 50 epochs. The Adam optimizer was used, with hyperparameters $\beta_1 = 0.8$ and $\beta_2 = 0.999$. After training, we have the trained model $\mathcal{R}_{\theta_p^*}$ to establish a mapping from the Fourier magnitude pattern I and the associated object image O, i.e., $\mathcal{R}_{\theta_p^*}$:I $\rightarrow O$. Thus, for an unknown test object x, we can use the pre-trained model $\mathcal{R}_{\theta_p^*}$ to provide a prediction $\hat{x}$ given the Fourier magnitude pattern $I = H(O)$, i.e., $\hat{O} = \mathcal{R}_{\theta_p^*}(I)$.

As the pre-trained model $\mathcal{R}_{\theta_p^*}$ contains implicit prior learned from the training set S_T that mitigate the ill-posedness of the inverse problem, one can hopefully obtain a good estimation of the image of the object from its Fourier magnitude pattern. However, the trained model $\mathcal{R}_{\theta_p^*}$ is inherently biased toward the training set S_T, and the prediction is not physically constrained, leading to the lack of generalization and interpretability.

Here, in the second step, as shown in Fig. 1(c), we address these limitations through a physics-driven fine-tuning approach. Specifically, we exploit the physics-consistency loss to refine the weights in the $\mathcal{R}_{\theta_p^*}$ by the following:

$$\theta_f^* = \arg\min_\theta ||H(\mathcal{R}_\theta(I)\odot W) - I||^2, \tag{4}$$

where W is a Bartlett window and the weights in $\mathcal{R}_\theta$ are initialized with θ_p^*. Along with the fine-tuning process, the network prediction gradually converges to the ground truth even for out-of-domain test data (see Fig. 1(d)). It is worth noting that the physics-driven fine-tuning method has better scalability and

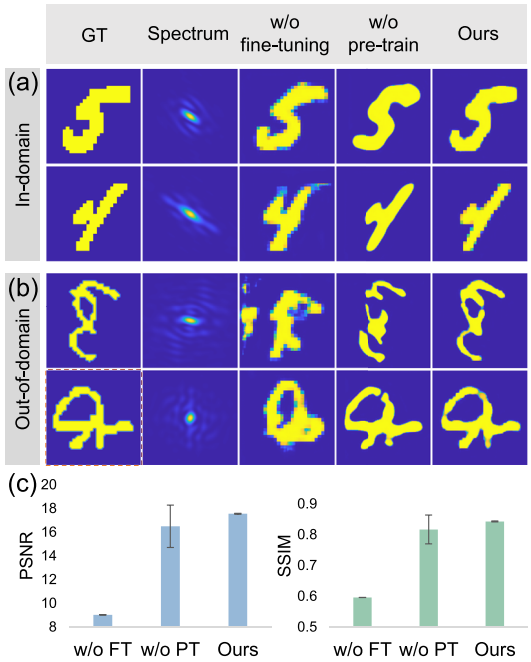

Fig. 2. Recovered images from the ablation study. (a) Results for in-domain (MNIST) test data. (b) Results for out-of-domain (EMNIST) test data. (c) Quantitative results from nine repeated runs using the dashed box target in (b). Error bars represent the standard deviation of the evaluation metrics of results from nine runs.

interpretability compared to traditional transfer learning methods [19]. This is because it does not require any data from the target domain for retraining and ensures that the output meets physical constraints.

Next, we present simulation results to analyze the effectiveness of the two key components of the proposed approach, i.e., fine-tuning and pre-train. We removed fine-tuning and the pre-training process from the algorithm separately as controls. Note that when fine-tuning is removed, the approach becomes a typical data-driven end-to-end DL method, and when the pre-training process is removed, the approach becomes an emerging untrained method. We randomly selected 10 unseen objects from MNIST to act as in-domain test data and 10 objects from EMNIST with different structures to act as out-of-domain test data. Figure 2 provides a summary of the results.

One can see that the results without fine-tuning are far from satisfactory, especially for out-of-domain test data, indicating that the fine-tuning process is indispensable (see Fig. 2(b), third column). When only exploiting the fine-tuning process without using pre-training, such a physics-driven untrained approach is capable of providing good results for images with different types of structure. However, since the network parameters are randomly initialized, the untrained approach has two limitations. First, even for in-domain test targets, the network output does not reflect any meaningful information before the iterations begin; second, the results from multiple runs vary significantly, indicating poor stability (see Fig. 2(c)). Our algorithm simultaneously uses data-driven pre-training to enhance the stability of the algorithm and employs a physics-driven fine-tuning process to improve scalability. By synergistically utilizing pre-training and fine-tuning processes, we can introduce implicit priors from the training data and explicit physical priors from the knowledge of the imaging system. This allows us to quickly generate preview results and further enhances them iteratively, yielding stable and good outcomes on both in-domain and out-of-domain test data (see Visualization 1 for more detail). In Supplement 1, we provide reconstruction results of our method on various test images and also discuss the algorithm's time efficiency and noise robustness.

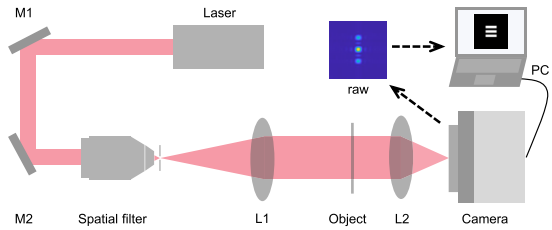

Fig. 3. Fourier coherent diffraction imaging experiment setup. M1 and M2, mirrors; L1, L2, lens.

Now, we will present the experimental demonstration. As plotted in Fig. 3, we built a typical Fourier coherent diffraction imaging experimental setup. A 632.8 nm He–Ne laser (Thorlabs HNL210L) was employed, and the optical path was adjusted using two reflecting mirrors. The light was focused through an objective lens (Olympus, 4×, NA = 0.2), and spatial filtering was performed using a 10 μm diameter pinhole. Next, a lens (L1) with a focal length of 200 mm transformed the light into parallel rays. The collimated light was then directed through an aperture stop (not shown) to illuminate an appropriate area of the object. The object, part of a resolution chart (Thorlabs, USAF1951), was positioned at the front focal plane of a Fourier transform lens (L2, focal length 180 mm), forming the Fourier magnitude pattern on its rear focal plane. After passing through an objective lens (Olympus, 10×, NA = 0.25) followed with a tube lens (not shown), the image was magnified 10 times and finally measured by a camera (PCO edge 4.2, 6.5 μm pixel pitch). We take four different parts of the resolution chart as the imaging target and acquire the corresponding Fourier magnitude pattern for further evaluations (see Figs. 4(a)–4(b)).

As many algorithms can be used to restore the object image from its Fourier spectrum, here, for the sake of comparison, we take the classical iterative phase retrieval method that combines hybrid input–output [6], shrinkwrap [20], and error reduction [21], referred to as HSE. We also take the emerging deep neural network-aided hybrid input–output (DNN-HIO) [22] for comparison. Besides, we show the results of the "w/o pre-train" method and the "w/o fine-tuning" method, which can be seen as representative results of untrained and end-to-end approaches, respectively. We also provide the result of our algorithm when removing the window, i.e., using a rectangle window instead of the Bartlett window. The results of the four selected imaging scenarios reconstructed by different FPR algorithms are summarized in Fig. 4.

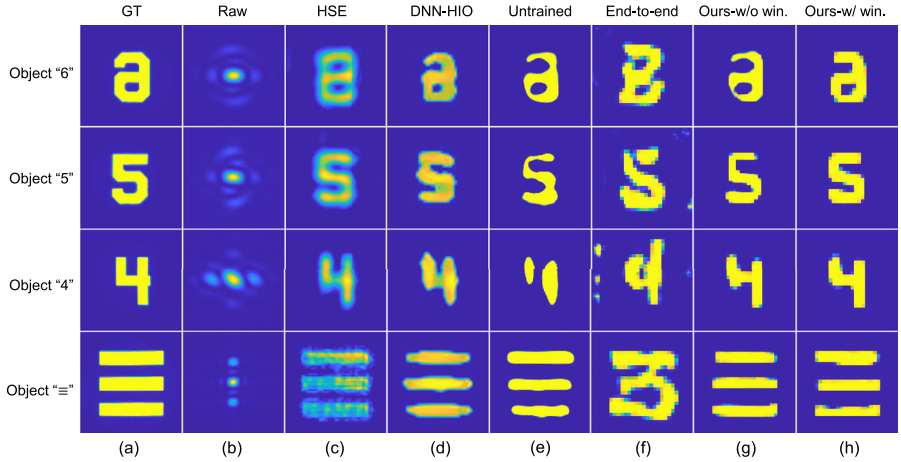

Fig. 4. Recovered images from the coherent diffraction imaging experiment. (a) Ground truth images captured by a camera. (b) Raw measured Fourier magnitude patterns (scaled). We exploit different reconstruction algorithms to restore the object image from the patterns shown in (b), resulting in results of HSE (hybrid input–output + shrinkwrap + error reduction) (c), DNN-HIO (deep neural network-aided hybrid input-output) (d), untrained (e), end-to-end (f), our method without (g), and with (h) Bartlett window.

The results obtained using the HSE approach demonstrate that while the object's contour can be reconstructed, the outcome is blurry and the grayscale restoration is inaccurate (see Fig. 4(c)). The DNN-HIO method provides more detailed information about the object and more accurate grayscale information, but the reconstruction results lack realism (see Fig. 4(d)). As we demonstrated before, the results of the untrained method fluctuate significantly due to random initialization. Here, we provide typical results for reference. These results indicate that the untrained approach recovers accurate grayscale information, but the object's structure is distorted (see Fig. 4(e)). The end-to-end data-driven deep learning method produced the worst results, especially for reconstructing triple-slit objects (non-digits-like) (see Fig. 4(f)). This is mainly because our model was trained only with simulated MNIST data, indicating the poor generalization ability of the neural network. Our algorithm without the Bartlett window (see Fig. 4(g)) provides satisfactory results, and we achieve the best results when using the Bartlett window (see Fig. 4(h)). This is because the Bartlett window can alleviate spectral leakage, thus providing a more accurate physical model to guide the fine-tuning process, resulting in better results. We also provide quantitative results for comparison (see Supplement 1). Consistent with visual perception, our method achieves the best image quality evaluation metrics, suggesting the excellent performance of our algorithm.

In summary, we have proposed a physics-enhanced DL approach for the FPR. This method integrates the physical model of the FPR with deep neural networks for data-driven pre-training and physics-driven fine-tuning. This approach has three advantages. First, our method can overcome the generalization limitations of traditional end-to-end data-driven DL methods, enabling the reconstruction of out-of-domain data without the need for target domain training datasets as in transfer learning. Second, our method balances reconstruction speed and quality by providing an initial result using the pre-trained model, followed by iterative refinement for further enhancement. Third, by combining traditional data-driven DL and physics-driven optimization algorithms, our method leverages both implicit priors from data and physical priors from the FPR imaging system, achieving high-quality results from a single Fourier magnitude pattern. However, as a preliminary proof-of-concept study, we have only validated the effectiveness of our algorithm through a simple coherent diffraction imaging experiment. In the future, we plan to apply our approach to more practical phase retrieval problems, such as lensless x-ray imaging.

Funding. National Natural Science Foundation of China (61991452, 62061136005, 62101537, 62305357, 62325508); Program of Shanghai Academic Research Leader (22XD1403900); Shanghai Municipal Science and Technology Major Project; Shanghai Sailing Program (22YF1455200, 23YF1454000, 23YF1454200).

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request. The source code is available in Ref. [23].

Supplemental document. See Supplement 1 for supporting content.

REFERENCES

1. F. Pfeiffer, Nat. Photonics **12**, 9 (2018).
2. H. N. Chapman and K. A. Nugent, Nat. Photonics **4**, 833 (2010).
3. M. Bertero and P. Boccacci, *Introduction to Inverse Problems in Imaging* (Institute of Physics, 1998).
4. F. Zhang, G. Pedrini, and W. Osten, Phys. Rev. A **75**, 043805 (2007).
5. L. Song and E. Y. Lam, Photonics Res. **10**, 758 (2022).
6. J. R. Fienup, Appl. Opt. **21**, 2758 (1982).
7. M. L. Moravec, J. K. Romberg, and R. G. Baraniuk, *Wavelets XII*, Vol. 6701 (SPIE, 2007), pp. 712–722.
8. E. J. Candes, T. Strohmer, and V. Voroninski, Comm. Pure Appl. Math. **66**, 1241 (2013).
9. E. Cha, C. Lee, M. Jang, *et al.*, IEEE Trans. Pattern Anal. Mach. Intell. **44**, 9931 (2022).
10. Q. Ye, L.-W. Wang, and D. P. Lun, Opt. Express **30**, 31937 (2022).
11. B. Neyshabur, S. Bhojanapalli, D. McAllester, *et al.*, Adv. Neural Inf. Process Syst. **30**, 1 (2017).
12. F. Wang, Y. Bian, H. Wang, *et al.*, Light: Sci. Appl. **9**, 77 (2020).
13. F. Wang, C. Wang, S. Han, *et al.*, Photonics Res. **10**, 104 (2022).
14. Z. Tang, F. Wang, Z. Fu, *et al.*, Opt. Lett. **48**, 2285 (2023).
15. C. O. Quero, D. Leykam, and I. R. Ojeda, J. Opt. Soc. Am. A **41**, 766 (2024).
16. L. Wu, S. Yoo, A. F. Suzana, *et al.*, npj Comput. Mater. **7**, 175 (2021).
17. J. Miao, P. Charalambous, J. Kirz, *et al.*, Nature **400**, 342 (1999).
18. J. W. Goodman, *Introduction to Fourier Optics* (Roberts and Company Publishers, 2005).
19. F. Zhuang, Z. Qi, K. Duan, *et al.*, Proc. IEEE **109**, 43 (2021).
20. S. Marchesini, H. He, H. N. Chapman, *et al.*, Phys. Rev. B **68**, 140101 (2003).
21. H. H. Bauschke, P. L. Combettes, and D. R. Luke, J. Opt. Soc. Am. A **19**, 1334 (2002).
22. Z. Chen, S. Zheng, Z. Tong, *et al.*, Optica **9**, 677 (2022).
23. Z. Zhang, F. Wang, Q. Min, *et al.*, "Demo code for physics-enhanced deep learning in Fourier phase retrieval using Jupyter notebook," GitHub (2024) [accessed 22 October 2024], https://github.com/SituLab/Fourier-Phase-Retrieval.