Wang et al. *Light: Science & Applications* (2020)9:77
https://doi.org/10.1038/s41377-020-0302-3

Official journal of the CIOMP 2047-7538
www.nature.com/lsa

LETTER Open Access

Phase imaging with an untrained neural network

Fei Wang[1,2], Yaoming Bian[1,2], Haichao Wang[1,2], Meng Lyu[1,2], Giancarlo Pedrini[3], Wolfgang Osten[3], George Barbastathis[4] and Guohai Situ[1,2,5]

Abstract

Most of the neural networks proposed so far for computational imaging (CI) in optics employ a supervised training strategy, and thus need a large training set to optimize their weights and biases. Setting aside the requirements of environmental and system stability during many hours of data acquisition, in many practical applications, it is unlikely to be possible to obtain sufficient numbers of ground-truth images for training. Here, we propose to overcome this limitation by incorporating into a conventional deep neural network a complete physical model that represents the process of image formation. The most significant advantage of the resulting physics-enhanced deep neural network (PhysenNet) is that it can be used without training beforehand, thus eliminating the need for tens of thousands of labeled data. We take single-beam phase imaging as an example for demonstration. We experimentally show that one needs only to feed PhysenNet a single diffraction pattern of a phase object, and it can automatically optimize the network and eventually produce the object phase through the interplay between the neural network and the physical model. This opens up a new paradigm of neural network design, in which the concept of incorporating a physical model into a neural network can be generalized to solve many other CI problems.

Recently, deep learning (DL) has shown great potential for solving inverse problems in computational imaging (CI)[1]. Pioneering studies have demonstrated the applicability of DL in optical tomography[2], computational ghost imaging[3,4], digital holography[5–7], imaging through scattering media[8–10], fluorescence lifetime imaging[11] imaging under low-light conditions[12], phase imaging[13–15], unwrapping[16], and fringe analysis[17]. Generally, an artificial neural network used in CI requires a large set of labeled data to optimize its weight and bias parameters (training) so that it can represent a universal function that maps the data in the object space into the image space[1]. Depending on the network architecture and the amount of data used for training, the network training process can take several hours or even several days, although the reconstruction process is very quick in most cases. Thus, the acquisition of a sufficiently large set of training data is crucial for the training of a good neural network. However, in many applications, one is usually required to image something that has never been seen before. It is thus impossible to acquire sufficient ground-truth images for network training, resulting in limited generalization ability[9,18].

We demonstrate in this letter that it is possible to experimentally recover an image with an untrained neural network that is built by combining a conventional artificial network such as U-Net[19] with a real-world physical model that represents the image formation physics; we call the resulting model PhysenNet. Thus, one does not need thousands of labeled data to train PhysenNet before it can be used. Instead, one needs only to feed a single image to be processed into a PhysenNet model with a suitable handcrafted structure, and the network weight and bias factors will be optimized through the interplay between the neural network and the physical model, eventually resulting in a feasible solution that satisfies the imposed physical constraints. The idea of enforcing implicit priors by means of the handcrafted network structure in PhysenNet is inspired by the concept of the deep image prior (DIP)[20]. We note that the DIP alone has

Correspondence: Guohai Situ (ghsitu@siom.ac.cn)
[1]Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, 201800 Shanghai, China
[2]Center of Materials Science and Optoelectronics Engineering, University of Chinese Academy of Sciences, 100049 Beijing, China
Full list of author information is available at the end of the article

© The Author(s) 2020
Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

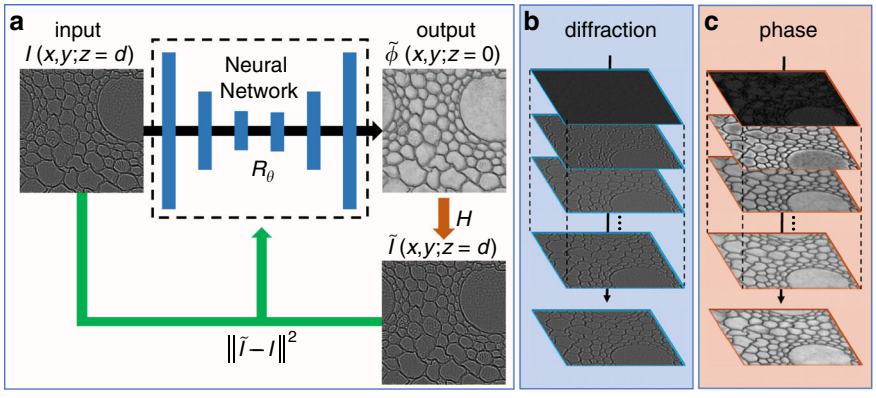

Fig. 1 Schematic illustration of the pipeline of PhysenNet. a A measured diffraction pattern I of a phase object ϕ is the input to the neural network. The output of the neural network is taken as the estimated phase $\tilde{\phi}$, which is then numerically propagated to simulate the diffraction and measurement processes H to generate $\tilde{I}$. We measure the mean square error (MSE) between I and $\tilde{I}$ as the loss value to adjust the neural network parameters θ. **b** $\tilde{I}$ during the optimization process. **c** $\tilde{\phi}$ during the optimization process

been used in some CI applications[20–23], but all these studies have been largely limited to simulations. The incorporation of the DIP with a task-specific physical model for optical imaging and its demonstration for coherent imaging experiments are the main contributions of this work. Here, we take phase imaging as a typical example to explain the principle more explicitly.

Phase problems are encountered in many applications, ranging from astronomy to industrial inspection. However, phase imaging is a highly ill-posed problem[24] when relying on intensity-only measurements[25,26], and sometimes requires a separate reference beam to encode the phase into fringe patterns[27]. The PhysenNet approach proposed here requires only one intensity $I(x, y; z = d)$, which is a diffraction pattern of a phase-only object $\phi(x, y; z = 0)$ located at $z = 0$ over a distance $z = d$, acquired using a single-beam set-up, i.e., without a separate reference. The basic concept is schematically outlined in Fig. 1a. The diffraction pattern $I(x, y; d)$ is the only input to PhysenNet, which has a handcrafted structure that is designed to generate an estimate of the phase object, $\tilde{\phi}(x, y; 0)$. In a conventional neural network, the ground-truth phase object $\phi(x, y; 0)$ in the training set must be known, and one can calculate the error between $\phi(x, y; 0)$ and $\tilde{\phi}(x, y; 0)$ to optimize the weights and biases[1,13–15]. By contrast, PhysenNet does not need the ground-truth phase $\phi(x, y; 0)$. Instead, it uses a physical model H to calculate a diffraction pattern $\tilde{I}(x, y; d)$ from $\tilde{\phi}(x, y; 0)$ according to, for example, the Huygens–Fresnel principle[28] and then uses the error between $\tilde{I}(x, y; d)$ and the measured $I(x, y; d)$ to optimize the weights and biases via gradient descent. This will force the calculated diffraction pattern $\tilde{I}$ to converge to the measured pattern I as the iterative process proceeds, as schematically shown in Fig. 1b. Throughout this iterative process, the search for the phase converges to a feasible solution, as shown by the simulation results presented in Fig. 1c.

Now, let us take a closer look at the technical details of PhysenNet. For a phase object, $\phi(x, y; 0)$, illuminated by a coherent plane wave, the complex amplitude immediately behind it can be written as

$$U_0(x, y; 0) = \exp[i\phi(x, y; 0)] \tag{1}$$

The diffraction of U_0 over a propagation distance $z = d$ is given by[28]

$$U_d(x, y; d) = \int\int \hat{U}_0(f_x, f_y) G \exp\left[i2\pi(f_x x + f_y y)\right] \mathrm{d}f_x \mathrm{d}f_y \tag{2}$$

where $G = \exp\left[ikd\sqrt{1 - \lambda^2 f_x^2 - \lambda^2 f_y^2}\right]$ is the transfer function, $\hat{U}_0$ is the Fourier transform of U_0, and f_x and f_y are the spatial frequencies in the x and y directions, respectively. The diffraction pattern recorded by an image sensor can be expressed as

$$I(x, y; d) = |U_d(x, y; d)|^2 = H(\phi) \tag{3}$$

where $H(\cdot)$ represents the mapping function that relates the phase object ϕ to the measured diffraction pattern I. The objective of the phase imaging problem is then to formulate an inverse mapping $H^{-1}(\cdot)$ such that

$$\phi(x, y; 0) = H^{-1}(I(x, y; d)) \tag{4}$$

One typical method is to solve the minimization problem $\tilde{\phi}(x, y) = \arg\min_\phi \|H(\phi) - I\|^2 + \rho(\phi)$, where $\rho(\phi)$ is a handcrafted or dictionary prior[29,30] that captures the generic regularity of the object, for $\tilde{\phi}$.

A typical DL-based method is to attempt to learn a mapping function R from a large number of labeled data (ϕ_k, I_k), $k = 1, \ldots, K$, that form the training set $S_T = \{(\phi_k, I_k); k = 1, \ldots, K\}$ by solving

$$R_{\theta^*} = \arg\min_{\theta \in \Theta} \|R_\theta(I_k) - \phi_k\|^2, \qquad \forall(\phi_k, I_k) \in S_T \tag{5}$$

where R_θ is the mapping function of the neural network defined by a set of weights and biases $\theta \in \Theta$. The training process results in a feasible mapping function R_{θ^*} that can map a diffraction pattern I that is not in S_T back to the corresponding phase $\tilde{\phi}$, i.e., $\tilde{\phi} = R_{\theta^*}(I)$. The size K of the training set S_T can be a few thousand or even tens of thousands in a typical CI application[1–17]. Experimentally collecting such a large set of diffraction patterns I_k and their corresponding ground-truth phases ϕ_k is time consuming and usually requires mechanical and environmental stability during the many hours of data acquisition. Although a training set can be created through numerical modeling of the image formation physics[4], the mapping function learned in such a case works well only for test images that are similar to those in the training set, resulting in good generalization only within the set of objects with the same priors used during training.

Instead, in the PhysenNet model proposed here, the retrieval of the phase is formulated as

$$R_{\theta^*} = \arg\min_{\theta \in \Theta} \|H(R_\theta(I)) - I\|^2 \tag{6}$$

where $H(\cdot)$ is defined through the physical model of diffraction described by Eqs. (1)–(3). The ground-truth phase ϕ explicitly does not appear in objective function (6), meaning that PhysenNet does not require the ground-truth phase for training. Instead, it is the interplay between H and R_θ that causes the prior of I to be captured by the handcrafted neural network. When the optimization is complete, the resulting mapping function R_{θ^*} can then be used to reconstruct the phase:

$$\tilde{\phi} = R_{\theta^*}(I) \tag{7}$$

It is worth pointing out that there is no limitation on the network architecture that can be chosen to implement R_θ. In our study, we simply adopt U-Net[19], which has been widely used for CI[1,4,6,7]. Typically, this network structure consists of an encoder path that takes the diffraction pattern as its input, a decoder path that outputs a predicted phase map, and skip paths in the middle. We use four main types of modules to connect the input to the output: convolution blocks (3×3 convolution + batch normalization + leaky ReLU), max pooling blocks (2×2), up-convolution blocks (3×3 de-convolution + batch normalization + leaky ReLU), and skip connection blocks. We use ReLU as the activation function in the output layer. (See Fig. S1 in the Supplementary Information for more details about the architecture.)

The neural network was implemented based on the TensorFlow version 1.9.0 platform using Python 3.6.5. We adopted the Adam optimizer[31] with a learning rate of 0.01 to optimize the weights and biases, and added uniformly distributed noise between 0 and $\frac{1}{30}$ to the fixed input I in every optimization step to achieve better convergence[20]. When the training process was complete, we removed the noise and obtained the reconstructed phase in accordance with Eq. (7). In our study, the size of the input image I was 256×256 pixels. The network usually needed 10,000 epochs to find a very good estimate. This took ~10 min on a computer with an Intel Xeon CPU E5-2696 V3, 64 GB of RAM, and an NVIDIA Quadro P6000 GPU.

We demonstrated the performance of the proposed PhysenNet method through both simulation and experiment. In the simulations, we first compared the proposed method with typical phase retrieval methods, i.e., the Gerchberg–Saxton (GS) algorithm[24,25] and the transport-of-intensity (TIE) equation[26]. Simulations were conducted using the aforementioned weight parameters. The results are illustrated in Fig. 2. We used the mean square error (MSE) to measure the quality of the reconstructed phase image in comparison to the ground truth shown in Fig. 2a. For a quantitative performance evaluation, we rescaled the reconstructed phases to the same range. The MSE value between the phase reconstructed using PhysenNet (Fig. 2f) and the ground truth is 0.01 rad, whereas the corresponding values associated with the GS algorithm (Fig. 2d) and the TIE equation (Fig. 2e) are 0.03 and 0.06 rad, respectively. In this simulation, PhysenNet used only one diffraction pattern to retrieve the phase, whereas the GS and TIE methods used multiple measurements along the z axis as inputs to enhance the quality of the reconstructed phase. In principle, the GS algorithm can retrieve a phase from a single measurement, provided that additional knowledge such as the support of the object is known. However, greater diversity is always preferable[24].

Next, we numerically analyzed the effect of the diffraction distance d on the quality of the reconstructed image. We take three diffraction distances, i.e., $d = 10$ mm, $d = 95$ mm, and $d = 180$ mm, as examples to examine the performance. The results are presented in Fig. 3. One can clearly see from Fig. 3d, e, f, that in all these cases, the phase can be successfully reconstructed from the corresponding diffraction patterns plotted in Fig. 3a–c. This observation is consistent with the reduction in the MSE with an increasing number of epochs that can be seen from the plot in Fig. 3g. Indeed, the MSE values associated with the reconstructed phase

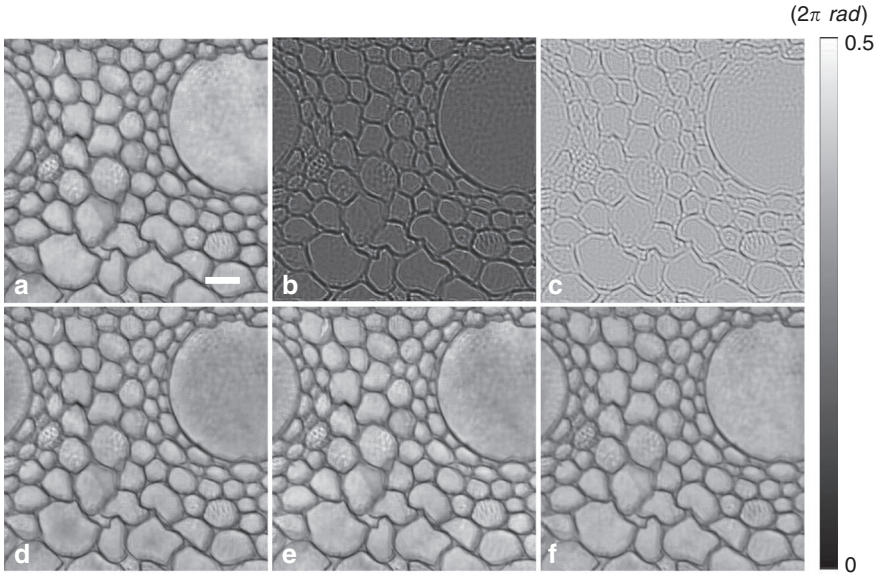

Fig. 2 Comparison of different phase imaging methods. a The ground-truth phase image: a plant cell slice obtained in our previous work[35]. **b** The diffraction pattern calculated from **a** at $d = 10$ mm. **c**–**f** The phases reconstructed by means of **c** direct reconstruction from **b** via back propagation, **d** the GS algorithm, **e** the TIE equation, and **f** PhysenNet. Scale bar: 256 μm

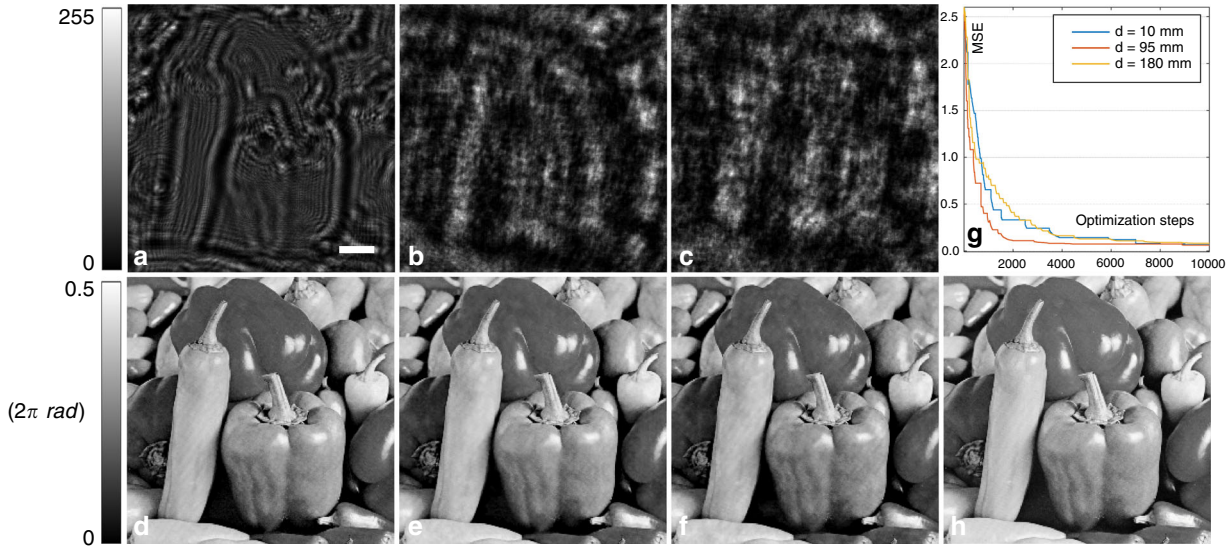

Fig. 3 Effect of *d* on the reconstructed phase. The first three images in the top row show the diffraction patterns at *d* values of **a** 10 mm, **b** 95 mm, and **c** 180 mm, and the phase maps reconstructed from these patterns are shown in **d**, **e**, and **f**, respectively. **g** The evolution of the MSE with an increasing number of epochs, and **h** the ground-truth object phase image. Scale bar: 256 μm

maps in Fig. 3d–f with respect to the ground-truth phase image in Fig. 3h are 0.067, 0.061, and 0.076 rad, respectively.

We also conducted a direct comparison of PhysenNet and conventional end-to-end approaches for phase imaging. We employed the same neural network structure (without the physical model) to fit the training set (10,000 human face images from Faces-LFW[32]) to obtain a trained model for mapping intensity patterns to phase images (see Table S1 in the Supplementary Information for more details). The results are illustrated in Fig. 4. Again, we used the MSE to measure the quality of the reconstructed phase image in comparison to the ground truth shown in Fig. 4b, which is one of the test images. The MSE value between the phase reconstructed from the diffraction pattern (Fig. 4a) using the pure end-to-end deep learning approach (Fig. 4c) and the ground truth is 0.038 rad,

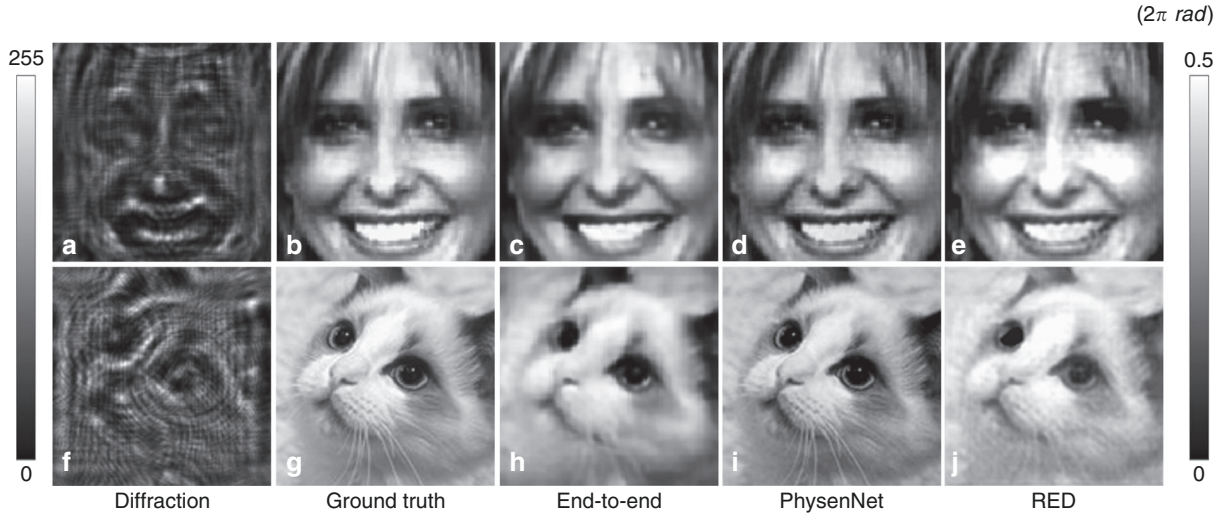

Fig. 4 Comparison of PhysenNet, conventional end-to-end DL and RED. a, **f** The diffraction patterns at a distance of 10 mm corresponding to the real phase distributions shown in **b** and **g**. **c**, **h** The phase images reconstructed using the conventional end-to-end strategy, **d**, **i** the phase images reconstructed by PhysenNet, and **e**, **j** the phase images reconstructed by RED

whereas the corresponding value associated with PhysenNet (Fig. 4d) is 0.033 rad. However, we observe that when the phase image is from another set, such as the cat face shown in Fig. 4g, the MSE between the phase reconstructed using the conventional end-to-end approach (Fig. 4h) and the ground truth is 0.127 rad, whereas the corresponding error associated with PhysenNet (Fig. 4i) is 0.025 rad, which is tenfold better. As expected, for the conventional end-to-end deep learning approach, the recovery quality decreases as the similarity between the test object and the training objects decreases. However, the performance of PhysenNet is not similarly affected.

We also performed simulations to compare PhysenNet with Regularization by Denoising (RED). We generated the training dataset by adding AWGN (std = 30 dB) to 10,000 images from Faces-LFW[32]. We employed DnCNN[33] to fit the training dataset to obtain the denoiser for RED. Following[34], we again used Adam[31] to minimize the objective as follows:

$$\tilde{\phi}(x,y) = \arg\min_{\phi,\theta} \|H(R_\theta^1(I)) - I\|^2 + \frac{\lambda}{2}\phi^T(\phi - R_{\theta^*}^2(\phi)) \tag{8}$$

where R_θ^1 is the deep neural network we used to generate the phase ϕ from the diffraction pattern I, λ is the RED regularization strength, and $R_{\theta^*}^2$ is the pre-trained denoising model. The results are illustrated in Fig. 4e, j. The MSE values between these results and the ground-truth images are 0.039 and 0.068 rad, respectively.

Now, we will present the experimental demonstration. The experimental apparatus is schematically shown in Fig. 5a. One can see that this is actually a single-beam lens-less imaging geometry. A laser beam emitted from a He–Ne laser at a wavelength of 632.8 nm (NEC Electronics Inc. GLG5002) was first spatially filtered by a pinhole with an aperture of 10 μm and then collimated by a lens with a focal length of $f = 200$ mm. The plane wave was guided to illuminate a phase object, producing intensity images as shown in Fig. 5b. To acquire the diffraction pattern, we placed the camera (SensiCam EM, pixel pitch: 8 μm) at a distance $d = 22.3$ mm from the phase object. The recorded diffraction pattern is shown in Fig. 5c. The proposed PhysenNet takes this diffraction pattern as its only input and generates an output phase map, as shown in Fig. 5d. Off-axis digital holography (DH)[27] was used to retrieve the object phase image shown in Fig. 5e. As there was only one diffraction pattern available, it was not possible to retrieve the phase by using the TIE equation; however, we did reconstruct the phase from the single diffraction pattern shown in Fig. 5c using the GS algorithm with the phase-only constraint on the object plane, and the result is plotted in Fig. 5f. Note that a separate carrier beam in DH encodes the phase into an intensity pattern, essentially making the phase problem a well-posed one. Here, by taking the DH reconstruction result as the ground truth, we can calculate the MSE between Fig. 5d, e to be 0.084 rad. The cross section highlighted by the dashed line indicates that the phase map reconstructed by the proposed PhysenNet is relatively smooth. In contrast, the MSE between the images reconstructed using the GS and DH methods is 1.926 rad, as clearly evidenced by the noise present in Fig. 5f. Similar observations hold for Fig. 5g–k, which show the results of retrieving the phase for another

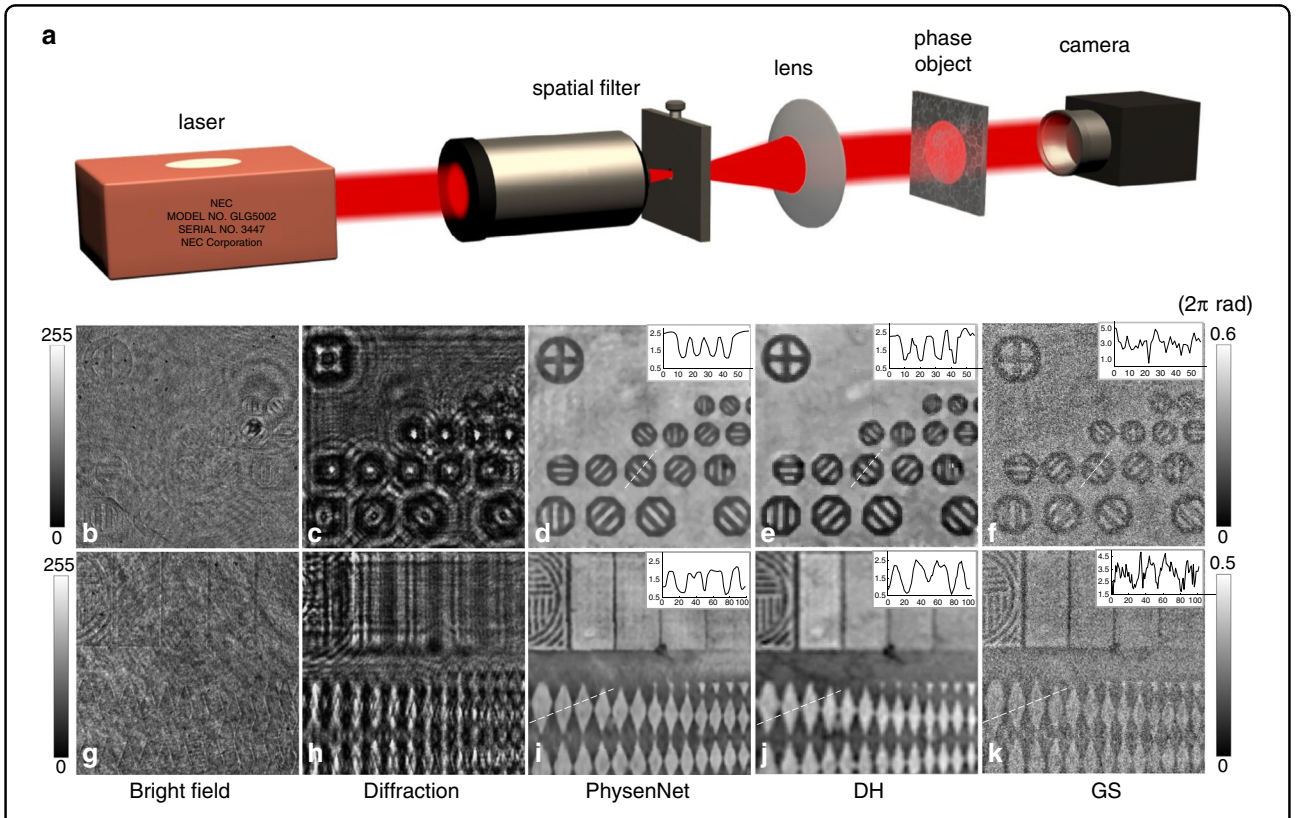

Fig. 5 Experimental results. a Experimental set-up. **b**, **g** Two different parts of the phase object, **c**, **h** the diffraction patterns, **d**, **i** the phase images reconstructed by PhysenNet, **e**, **j** the phase images reconstructed via off-axis digital holography, and **f**, **k** the phase images reconstructed with the GS algorithm

part of the sample. The MSE value between Fig. 5i and j is 0.093 rad, whereas a value of 2.981 rad is associated with Fig. 5k.

In all the above investigations, we imposed no assumptions on the profile or support of the phase object, in contrast to almost all other phase retrieval algorithms[24,25]. However, we found that PhysenNet does not work well for phase modulation ranges larger than 2π. Resolving this limitation is beyond the scope of the present study.

PhysenNet requires precise modeling of the image formation mechanism [e.g., Eqs. (1)–(3) in our study], and the incorporation of the resulting physical model into a conventional deep neural network (U-Net in our case). It is the interplay between the physical model and the neural network that allows the object phase to be reconstructed with a single intensity measurement. The advantages of PhysenNet in comparison to the pure end-to-end approaches for CI[1] are straightforward. First, pure end-to-end approaches usually require many labeled data to train a neural network. In physical experiments, such labeled data can be generated by using an SLM, or they can be numerically synthesized using a rigorous image formation model[4]. PhysenNet, on the other hand, does not require any labeled data for training. Instead, all it needs as input is the image to be processed. Second, pure end-to-end approaches learn a mapping function from the statistics of a large set of training data, represented by the weights of the network. When test data are fitted with the same set of weights, test error will inevitably emerge, resulting in artefacts and noise in the reconstructed images, particularly in cases where the test data are far from the training data in terms of their statistics. PhysenNet, inspired by the DIP, does not learn a mapping function from the statistics of the training data but rather is based on the interplay between a handcrafted network structure and a physical image formation model. As a result, the network in PhysenNet is more specifically tuned to perform well in reconstruction from the given input, at the cost of some generalization ability. Although we have demonstrated PhysenNet only for a use case of 2D phase retrieval, it is, in principle, also applicable for 3D objects provided that a multi-projection technique such as tomography is used to collect the data. In these cases, there should be multiple mapping functions H_i, where $i = 1,2, \ldots, N$ denotes the number of projections, that relate the measured intensity I_i to the 3D object function in the ith view. These functions H_i should be

implemented to represent the associated physics, and objective function (6) should accordingly be generalized to $R_{\theta^*} = \arg\min_{\theta\in\Theta} \sum_i \|H_i(R_\theta(I_i)) - I_i\|^2$.

In comparison to conventional DL approaches for CI, the only extra ingredient that PhysenNet needs is a known forward mapping function H, as described in Eq. (6). This means that, given an estimate $\tilde{U}$ of an object function, PhysenNet requires the calculability of the forward transform of $\tilde{U}$ through an imaging system specified by H, which is required to evaluate the cost function. No additional requirements are imposed on either the method of data acquisition or the illumination conditions. As a result, PhysenNet should be applicable for diverse imaging modalities, provided that the forward mapping function is known.

Acknowledgements
This work was supported by the Key Research Program of Frontier Sciences of the Chinese Academy of Sciences (QYZDB-SSW-JSC002), the Sino-German Center (GZ1391), and the National Natural Science Foundation of China (61991452).

Author details
[1]Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, 201800 Shanghai, China. [2]Center of Materials Science and Optoelectronics Engineering, University of Chinese Academy of Sciences, 100049 Beijing, China. [3]Institut für Technische Optik, Universität Stuttgart, Pfaffenwaldring 9, 70569 Stuttgart, Germany. [4]Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139-4301, USA. [5]Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, 310024 Hangzhou, China

Author contributions
G.S., F.W. and M.L. conceived the idea; F.W. performed the numerical simulations and conducted the experiments, guided by discussions with Y.B., H.W. and G.P.; F.W., G.P. and G.S. analyzed the results; F.W., G.S., G.P., W.O. and G. B. wrote the manuscript; and G.S. supervised the project.

Conflict of interest
The authors declare that they have no conflict of interest.

Supplementary information is available for this paper at https://doi.org/10.1038/s41377-020-0302-3.

Received: 15 November 2019 Revised: 17 March 2020 Accepted: 23 March 2020
Published online: 06 May 2020

References

1. Barbastathis, G., Ozcan, A. & Situ, G. On the use of deep learning for computational imaging. *Optica* **6**, 921–943 (2019).
2. Kamilov, U. S. et al. Learning approach to optical tomography. *Optica* **2**, 517–522 (2015).
3. Lyu, M. et al. Deep-learning-based ghost imaging. *Sci. Rep.* **7**, 17865 (2017).
4. Wang, F. et al. Learning from simulation: an end-to-end deep-learning approach for computational ghost imaging. *Opt. Express* **27**, 25560–25572 (2019).
5. Ren, Z. B., Xu, Z. M. & Lam, E. Y. Learning-based nonparametric autofocusing for digital holography. *Optica* **5**, 337–344 (2018).
6. Wang, H., Lyu, M. & Situ, G. eHoloNet: a learning-based end-to-end approach for in-line digital holographic reconstruction. *Opt. Express* **26**, 22603–22614 (2018).
7. Rivenson, Y. et al. Phase recovery and holographic image reconstruction using deep learning in neural networks. *Light Sci. Appl.* **7**, 17141 (2018).
8. Lyu, M. et al. Learning-based lensless imaging through optically thick scattering media. *Adv. Photonics* **1**, 036002 (2019).
9. Li, Y. Z., Xue, Y. J. & Tian, L. Deep speckle correlation: a deep learning approach toward scalable imaging through scattering media. *Optica* **5**, 1181–1190 (2018).
10. Li, S. et al. Imaging through glass diffusers using densely connected convolutional networks. *Optica* **5**, 803–813 (2018).
11. Wu, G. et al. Artificial neural network approaches for fluorescence lifetime imaging techniques. *Opt. Lett.* **41**, 2561–2564 (2016).
12. Goy, A. et al. Low photon count phase retrieval using deep learning. *Phys. Rev. Lett.* **121**, 243902 (2018).
13. Sinha, A. et al. Lensless computational imaging through deep learning. *Optica* **4**, 1117–1125 (2017).
14. Li, X. et al. Quantitative phase imaging via a cGAN network with dual intensity images captured under centrosymmetric illumination. *Opt. Lett.* **44**, 2879–2882 (2019).
15. Xue, Y. J. et al. Reliable deep-learning-based phase imaging with uncertainty quantification. *Optica* **6**, 618–629 (2019).
16. Wang, K. Q. et al. One-step robust deep learning phase unwrapping. *Opt. Express* **27**, 15100–15115 (2019).
17. Feng, S. J. et al. Fringe pattern analysis using deep learning. *Adv. Photonics* **1**, 025001 (2019).
18. Goodfellow, I., Bengio, Y. & Courville, A. *Deep Learning* 775 (MIT Press, Cambridge, 2016).
19. Ronneberger, O., Fischer, P. & Brox, T. U-Net: convolutional networks for biomedical image segmentation. In *Proc. 18th International Conference on Medical Image Computing and Computer-Assisted Intervention* 234–241 (Springer, Munich, 2015).
20. Lempitsky, V., Vedaldi, A. & Ulyanov, D. Deep image prior. In *Proc. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* 9446–9454 (IEEE, Salt Lake City, 2018).
21. Anirudh, R et al. An unsupervised approach to solving inverse problems using generative adversarial networks. Preprint at https://arxiv.org/pdf/1805.07281.pdf (2018).
22. Liu, J. M. et al. Image restoration using total variation regularized deep image prior. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* 7715–7719 (IEEE, Brighton, 2019).
23. Jagatap, G. & Hegde, C. Phase retrieval using untrained neural network priors. In *NeurIPS 2019 Workshop on Solving Inverse Problems with Deep Networks*. Vancouver (2019).
24. Shechtman, Y. et al. Phase retrieval with application to optical imaging: a contemporary overview. *IEEE Signal Process. Mag.* **32**, 87–109 (2015).
25. Fienup, J. R. Phase retrieval algorithms: a comparison. *Appl. Opt.* **21**, 2758–2769 (1982).
26. Teague, M. R. Deterministic phase retrieval: a Green's function solution. *J. Opt. Soc. Am.* **73**, 1434–1441 (1983).
27. Osten, W. et al. Recent advances in digital holography [Invited]. *Appl. Opt.* **53**, G44–G63 (2014).
28. Goodman, J. W. *Introduction to Fourier Optics* 3rd edn (Roberts and Company Publishers, Greenwoood Village, 2005).
29. Aharon, M., Elad, M. & Bruckstein, A. K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Process.* **54**, 4311–4322 (2006).
30. Rubinstein, R., Bruckstein, A. M. & Elad, M. Dictionaries for sparse representation modeling. *Proc. IEEE* **98**, 1045–1057 (2010).
31. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. Preprint at https://arxiv.org/abs/1412.6980 (2014).
32. Huang, G. B. et al. *Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments* (University of Massachusetts, 2007).
33. Zhang, K. et al. Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising. *IEEE Trans. Image Process.* **26**, 3142–3155 (2017).
34. Mataev, G., Elad, M. & Milanfar, P. DeepRED: deep image prior powered by RED. Preprint at https://arxiv.org/abs/1903.10176 (2019).
35. Zhou, A. et al. Fast and robust misalignment correction of Fourier ptychographic microscopy for full field of view reconstruction. *Opt. Express* **26**, 23661–23674 (2018).