DOI: 10.58895/ksp/1000150865 -2 erschienen in:
Forum Bildverarbeitung 2022 | Image Pocessing Forum 2022
DOI: 10.5445/KSP/1000150865 | https://www.ksp.kit.edu/site/books/m/10.58895/ksp/1000150865/

Physics enhanced neural network for phase imaging using two axially displaced diffraction patterns

Rujia Li[1,2], Giancarlo Pedrini[1], Liangcai Cao[2], and Stephan Reichelt[1]

[1] Universität Stuttgart, Institut für Technische Optik,
Pfaffenwaldring 9, 70569 Stuttgart, Germany
[2] Tsinghua University, Department of Precision Instruments,
Qinghuayuan 1, 100084 Beijing, China

Abstract In this work, we propose a physics-enhanced two-to-one Y-neural network (two inputs and one output) for phase retrieval of complex wavefronts from two diffraction patterns. The learnable parameters of the Y-net are optimized by minimizing a hybrid loss function, which evaluates the root-mean-square error and normalized Pearson correlated coefficient on the two diffraction planes. An angular spectrum method network is designed for self-supervised training on the Y-net. Amplitudes and phases of wavefronts diffracted by a USAF-1951 resolution target, a phase grating of 200 lp/mm, and a skeletal muscle cell were retrieved using a Y-net with 100 learning iterations. Fast reconstructions could be realized without constraints or a priori knowledge of the samples.

Keywords Coherent diffraction imaging, phase retrieval, deep neural network

1 Introduction

Retrieving the phase from diffraction patterns is a long-standing problem. In the recorded intensity patterns, the object wavefront is superimposed with its a phase-conjugated and for reconstructing the wavefront without conjugation, the phase needs to be retrieved. Conventional methods used constraints to iteratively solve the phase retrieval

problem. A priori knowledge of the object plane [1] or modulations applied on the imaging path [2] [3] can be the constraints. Optimization iterations are needed.

Deep learning is a powerful approach for solving optimization problems. A convolutional neural network (CNN) is trained with a dataset for mapping input to output. CNNs are widely used in image processing, they have an end-to-end structure, which can be trained to retrieve a phase pattern from an intensity pattern [4] [5]. After training on a dataset, the reconstruction can be directly made by a CNN without further optimization. The phase retrieval problem has an explicit physical model and a CNN can be enhanced with the diffraction principle [5] in order to avoid training with thousands of patterns. However, the end-to-end structure of a CNN described in [6] limits the object to be phase-only. Splicing the phase and amplitude into one image seems to be a straightforward solution, but a CNN uses a convolution kernel for feature extraction. The connected edges of the amplitude and phase pattern may be convoluted with one kernel and generate data against the physical model.

In this work, we propose a physics-enhanced neural network for retrieving a complex wavefront from two axially displaced diffraction patterns. A two-to-one Y-net (two inputs and one output) is designed to retrieve the phase on the first plane. Then the complex wavefront is calculated with the retrieved phase and the square root of the recorded intensity pattern. An angular spectrum method (ASM) network is designed to calculate the wave propagation. The Y-net is trained with the diffraction between the two recording planes and produces a phase on the first plane, which can be used to generate two patterns on the two recording planes. The errors between generated and recorded patterns are evaluated with a hybrid loss function. The normalized Pearson correlation coefficient and root mean square error are used to build the hybrid loss function. The learnable parameters in the Y-net are optimized by gradient descent on the hybrid loss function. After training on a dataset, the Y-net can be generalized to retrieve complex wavefronts without optimization. Reconstruction can also be made using an untrained Y-net. An amplitude-only UASF-1951 resolution chart, a phase grating, and a skeletal muscle cell are experimentally reconstructed using an untrained Y-net.

2 Y-net for retrieving the complex wavefront

A schematic of the setup used for recording axially displaced diffraction patterns is shown in Fig 1 . The sample is illuminated by a plane wave and the diffraction patterns are recorded on two planes at distances of z'and z'+z. To reconstruct the complex-valued object, the phase on the two diffraction patterns is retrieved using a Y-net.

The proposed Y-net is a fusion of two U-nets. There are two down-sampling paths and one up-sampling path, which are composed of four down-sampling and corresponding up-sampling convolution blocks. In each convolution block, the information passes downstream along with two sets of batch normalization layers, rectified linear unit (ReLU) layer, and a convolution layer. The feature maps in each down-sampling block are extracted using a 3×3 convolution kernel with a stride of 2. In the bottleneck of the Y-net, the feature maps from the two down-sampling paths are connected as the input of the up-sampling path. Then the up sampling is made with transposed convolutions. There are residual layers and skip connections after the convolution blocks to make the deep Y-net easy to optimize by avoiding the vanishing gradients problem and mitigating the degradation problem.

The schematic for training the Y-net is shown in Fig.1 (b). In the first training loop, the learnable parameters are randomly initialized. This initialization helps keeping the signal from expanding to an extremely high value or vanishing to zero. Then the learnable parameters are optimized by minimizing a hybrid loss function, which is built by following the optical diffraction model.

The hybrid loss function for the Y-net is a linear combination of the loss function on two diffraction patterns. The output of the Y-net is set to be the phase on the first diffraction pattern. The complex wavefronts on the two planes follow the Rayleigh-Sommerfeld diffraction. By merging the phase $\varphi(x_1, y_1)$ with the first recorded intensity $I_1(x_1, y_1)$, we obtain the wavefront on the first plane $u_1(x_1, y_1) = \sqrt{I_1(x_1, y_1)} exp[i\varphi(x_1, y_1)]$. After propagating $u_1(x_1, y_1)$ to the second plane, we obtain the wavefront $u_2(x_2, y_2) = prop_z\{u_1(x_1, y_1)\}$, where z is the distance between the two planes. For evaluating the differences between $u_2(x_2, y_2)^2$ and $I_2(x_2, y_2)$, a loss function is built from the linear combination of the root-mean-square error (RMSE) and the

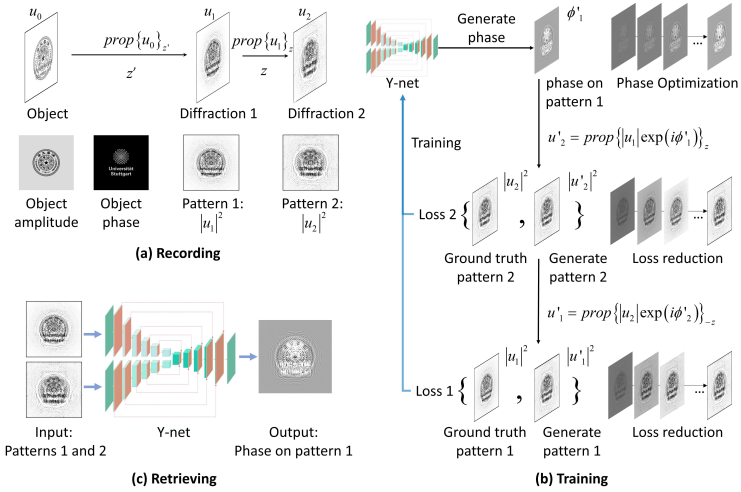

Figure 1: (a) Recording two patterns diffracted by a complex object; (b) Training the Y-net based on diffraction between the two planes; (c) Retrieving the phase on the first pattern.

normalized Pearson correlation coefficient (PCC),

$$Loss\{I, I'\} = l_{PCC} PCC\{I, I'\} + l_{RMSE} RMSE\{I, I'\} \tag{1}$$

$$PCC\{I, I'\} = \frac{1}{2}\{1 - \frac{\sum_{m,n}[I(m,n) - I_{ave}][I'(m,n) - I'_{ave}]}{\sqrt{\sum_{m,n}[I(m,n) - I_{ave}]^2 \sum_{m,n}[I'(m,n) - I'_{ave}]^2}}\} \tag{2}$$

$$RMSE\{I, I'\} = \sqrt{\frac{\sum_{m,n}[I(m,n) - I'(m,n)]^2}{MN}} \tag{3}$$

where l_{PCC} and l_{RMSE} are the relative weights of the normalized PCC and RMSE, m and n are integer numbers, M and N are the numbers of pixels in the patterns, I_{ave} and I'_{ave} are the average pixel values of the images. The PCC measures the linear similarity between the two patterns, which is evaluated by the ratio between the covariance of the pixel values and the product of their standard deviations. The PCC has a value between -1 and 1, where 1 represents two similar patterns. To perform gradient descent, the PCC operator is normalized as shown

in Eq.2. The normalized PCC has a value between 0 and 1, where 0 represents a high similarity. The RMSE is used together with PCC to obtain a better convergence in a training loop. The RMSE compares every pixel value on the generated intensity and the captured ground truth. The scaling effect of the PCC can be reduced by using the RMSE evaluation. When the RMSE value is 0, the generated intensity and the captured ground truth are the same on every pixel of the image.

In order to apply a sufficient constraint to the neural network, the loss function is also built on the second plane. The amplitude of the propagated wave $u_2(x_2, y_2)$ is replaced by $\sqrt{I_2(x_2, y_2)}$. Then the updated wavefront $u_2'(x_2, y_2)$ is propagated to the first plane, $u_1'(x_1, y_1) = prop_{-z}\{u_2'(x_2, y_2)\}$. The differences between $u_1'(x_1, y_1)^2$ and $I_1(x_1, y_1)$ are evaluated. The hybrid loss function for training the Y-net is $d_1 Loss_1\{I_1\} + d_2 Loss_2\{I_2\}$, where d_1 and d_2 are the weights of the loss on the two diffraction planes. Training the neural network is a process of optimizing the weights of each layer to minimize the prediction error between the outputs and ground truth. This is usually made by using gradient descent methods on the loss functions. In this work, the ADAM optimization is used for minimizing the hybrid loss function on the two planes. A well-trained Y-net retrieves a phase following the diffraction principles between the two planes.

3 Reconstructions in experiments

Experimental results were obtained by using an amplitude-only USAF-1951 resolution test target, a phase grating, and a skeletal cell sample. The diffraction patterns were recorded using the setup shown in Fig. 2(a). The samples were illuminated with a plane wave having wavelength 655 nm. The pixel size of the camera is 2 μm. After capturing the first diffraction pattern, the camera was shifted for capturing the second. The distance between the two planes was 400 μm. The size of the diffraction patterns was 512×512 pixel, this is a compromise for obtaining good resolution under fast training. Better results could be obtained using more pixels, but in this case a longer training time would be necessary.

Figs. 2(b) and (c) show the recorded patterns of the USAF-1951 resolution target. The first pattern was recorded at 4.4 mm distance

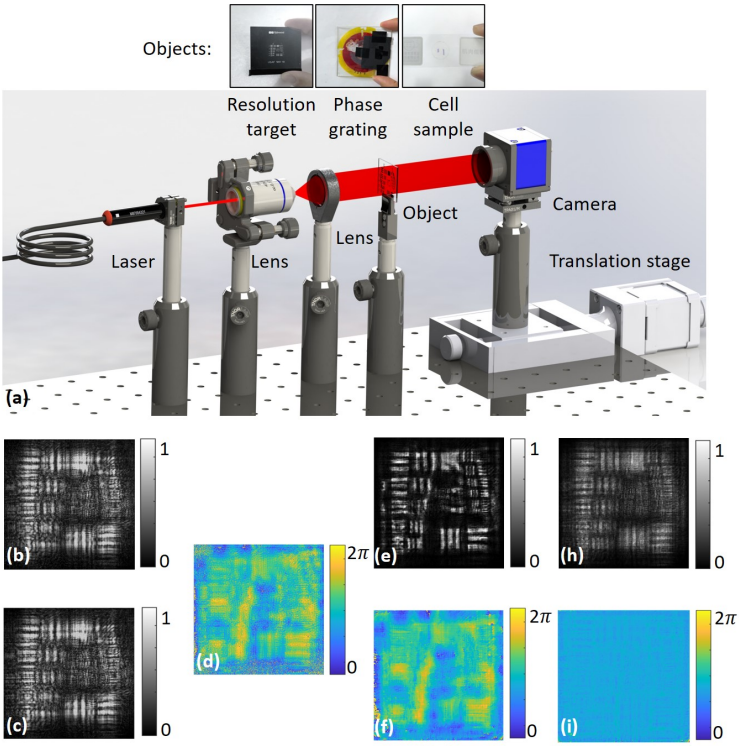

Figure 2: Experimental reconstruction for an amplitude-only USAF-1951 resolution target. (a) Schematic of the experimental setup; (b) and (c) The recorded diffraction patterns on the two planes; (d) Retrieved phase on the plane of (b); (e) and (f) Amplitude and phase of the reconstruction using the Y-net with 100 iterations; (h) and (i) Amplitude and phase of the reconstruction by propagating the first diffraction pattern.

from the object. An untrained Y-net was used for reconstruction. Self-supervised learning was performed by optimizing the hybrid loss function with 100 iterations. With the ASM network, the Y-net learns to retrieve a phase following the diffraction between the two recording planes. As shown in Fig. 2(d), the feature of the object can be distinguished from the retrieved phase. The complex wavefront on the first plane is calculated by multiplying the retrieved phase and the recorded amplitude. The phase and amplitude components are then reconstructed after propagating the calculated wavefront to the object plane. The intensity and the phase of the reconstruction are shown in Figs. 2(e) and (f). The sixth element of group five in the USAF-1951 target was resolved (line width of 8.77 μm). The reconstruction of the complex wavefront was made using the untrained Y-net without a priori knowledge. Figs. 2(h) and (i) shows the reconstruction of intensity and phase obtained by simply propagating the first diffraction pattern to the object plane. The intensity is not correctly reconstructed due to the presence of the conjugated wavefront.

A phase grating was also investigated with the same experimental setup shown in Fig. 2(a). The phase grating has a period of 5 μm (200 lp/mm). The first pattern was captured at a distance of 4.6 mm from the phase grating. Then the camera was shifted 400 μm for recording the second pattern. After self-supervised learning (100 iterations), the phase distributions of the gratings was reconstructed (see Figs. 3(b)). In this experiment, the phase grating cannot be reconstructed using simple propagation of the recorded diffraction pattern (Figs. 3(c), (d)).

A skeletal muscle cell was used in another experiment, to further demonstrate the capability of the Y-net. In this case the sample was illuminated with a plane wave having wavelength of 632.8 nm. The pixel size of the camera was 5.86 μm. The first diffraction pattern was captured 39.4 mm away from the specimen, this distance was numerically determined by back propagating the retrieved wave from the recording plane to the object plane. Then the camera was shifted 1 mm for capturing the second pattern. The phase at the first plane was retrieved after training the Y-net with 100 iterations. The reconstruction of the sample is obtained by propagating the retrieved wavefront. The reconstructed amplitude and phase of the skeletal muscle cell are shown in Figs. 3(e) and (f). The amplitude and phase show different structures of the skeletal muscle.

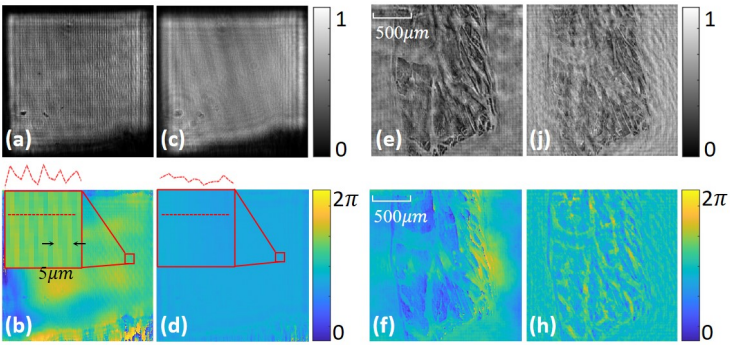

Figure 3: Experimental reconstruction for the phase grating and skeletal muscle cell sample. (a), (e) and (b), (f) Amplitude and phase of the reconstruction using the Y-net with 100 iterations; (c), (j) and (d), (h) Amplitude and phase of the reconstruction by propagating the first diffraction pattern.

4 Conclusion

Y-net is proposed to efficiently reconstruct complex wavefronts. With self-supervised training through an ASM network, the Y-net learns the diffraction between the two planes. Only two diffraction patterns are needed for the reconstruction. The two patterns may also be simultaneously captured using two cameras and one beam splitter. Then a well-trained Y-net may realize a quasi-real-time phase retrieval. The Y-net can be trained on a big dataset for the best generalization. The Y-net has a promising potential in the investigation of both timely and spatially varying physical processes. The large-scale complex wavefront can be rapidly retrieved using a well-trained Y-net. Besides the optical diffraction, this two-to-one Y-net may also be applied on learning other physical principles, such as the transmission of sound wave.

References

1. J. R. Fienup, "Phase retrieval algorithms: a comparison," *Appl. Opt.*, vol. 21, no. 15, pp. 2758–2769, Aug 1982. [Online]. Available: http://opg.optica.org/ao/abstract.cfm?URI=ao-21-15-2758

2. F. Zhang, G. Pedrini, and W. Osten, "Phase retrieval of arbitrary complex-valued fields through aperture-plane modulation," *Phys. Rev. A*, vol. 75, p. 043805, Apr 2007. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.75.043805

3. R. Li and L. Cao, "Complex wavefront sensing based on alternative structured phase modulation," *Appl. Opt.*, vol. 60, no. 4, pp. A48–A53, Feb 2021. [Online]. Available: https://opg.optica.org/ao/abstract.cfm?URI=ao-60-4-A48

4. A. Sinha, J. Lee, S. Li, and G. Barbastathis, "Lensless computational imaging through deep learning," *Optica*, vol. 4, no. 9, pp. 1117–1125, Sep 2017. [Online]. Available: https://opg.optica.org/optica/abstract.cfm?URI=optica-4-9-1117

5. F. Wang, C. Wang, C. Deng, S. Han, and G. Situ, "Single-pixel imaging using physics enhanced deep learning," *Photon. Res.*, vol. 10, no. 1, pp. 104–110, Jan 2022. [Online]. Available: https://opg.optica.org/prj/abstract.cfm?URI=prj-10-1-104

6. F. Wang, Y. Bian, H. Wang, M. Lyu, G. Pedrini, W. Osten, G. Barbastathis, and G. Situ, "Phase imaging with an untrained neural network," *Light: Science&Applications*, vol. 9, no. 77, May 2020. [Online]. Available: https://doi.org/10.1038/s41377-020-0302-3