Quantitative phase microscopy using deep neural networks

Citation: Li, Shuai, et al. "Quantitative Phase Microscopy Using Deep Neural Networks." Quantitative Phase Imaging IV, 27 January, - February 1, 2018, San Francisco, California, edited by Gabriel Popescu and YongKeun Park, SPIE, 2018, p. 84. © SPIE.

As Published: http://dx.doi.org/10.1117/12.2289056

Publisher: SPIE

Persistent URL: http://hdl.handle.net/1721.1/119144

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Quantitative phase microscopy using deep neural networks

Shuai Li, Ayan Sinha, Justin Lee, George Barbastathis

Shuai Li, Ayan Sinha, Justin Lee, George Barbastathis, "Quantitative phase microscopy using deep neural networks," Proc. SPIE 10503, Quantitative Phase Imaging IV, 105032D (23 February 2018); doi: 10.1117/12.2289056

SPIE.

Event: SPIE BiOS, 2018, San Francisco, California, United States

Quantitative phase microscopy using deep neural networks

Shuai Li[a], Ayan Sinha[a,¶], Justin Lee[b], and George Barbastathis[a,c]

[a]Department of Mechanical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139
[b]Institute for Medical Engineering Science, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139
[c]Singapore-MIT Alliance for Research and Technology (SMART) Centre, One Create Way, Singapore 117543, Singapore
[¶]Current address: Magic Leap Inc,1376 Bordeaux Drive, Sunnyvale, CA 94089

ABSTRACT

Deep learning has been proven to achieve ground-breaking accuracy in various tasks. In this paper, we implemented a deep neural network (DNN) to achieve phase retrieval in a wide-field microscope. Our DNN utilized the residual neural network (ResNet) architecture and was trained using the data generated by a phase SLM. The results showed that our DNN was able to reconstruct the profile of the phase target qualitatively. In the meantime, large error still existed, which indicated that our approach still need to be improved.

Keywords: phase retrieval, deep neural network, wide-field microscopy

1. INTRODUCTION

Phase is an important characteristic of light waves. The phase of the light that is transmitted through an object holds crucial information about the object's shape and refractive index.[1] Especially in biological and biomedical research, without staining, the specimens are rarely visible under conventional microscope.[2] Only phase shifts are introduced when the light passes through these samples. Therefore, quantitative phase retrieval becomes crucial in studying the behaviors of cells and organisms.

Apart from those conventional phase retrieval approaches such as phase measurement interferometry,[3,4] Transport of Intensity Equation method (TIE)[5] and Digital Holography (DH),[6] deep neural networks (DNNs) have been recently proposed by Sinha et al.[7] as a novel end to end approach for quantitative phase retrieval. DNN is a well-known machine learning architecture inspired by the natural visual perception mechanism of the living creatures.[8] In 2012, Krizhevsky et al.[9] proposed a multi-layer convolutional neural network (CNN) architecture, i.e. AlexNet and showed significant improvements upon previous methods on the image classification task. With the success of AlexNet, more and more novel structures have been proposed since then and they have been applied to various research fields like image classification,[10,11] object detection,[12,13] object tracking,[14] pose estimation,[15] imaging[16–19] and so on.

In Sinha's work, a deep residual network architecture was designed to quantitatively reconstruct the profile of unknown phase objects. Compared with conventional phase retrieval approaches, the main advantage of using DNN is that the DNN can learn the forward operator and the priors implicitly through training. Therefore, there is no need to explicitly calibrate the forward model of the imaging system. Nevertheless, Sinha's work used a lensless imaging system, i.e. no lenses existed between the object and the detector. As a result, the numerical aperture (NA) of the system was small, thus restricting the attainable spatial resolution.

In this paper, we modify the optical system to achieve higher spatial resolution. An additional path containing a wide-field microscopy module is used to capture test images. A DNN is trained using the data generated by a transmissive SLM and its performance is evaluated using the test data.

Further author information: (Send correspondence to S.L.)
S.L.: E-mail: shuaili@mit.edu

Quantitative Phase Imaging IV, edited by Gabriel Popescu, YongKeun Park,
Proc. of SPIE Vol. 10503, 105032D · © 2018 SPIE · CCC code:
1605-7422/18/$18 · doi: 10.1117/12.2289056

The structure of the paper is as follows: In Section 2, we describe the architecture of our system, including the optical setup and the DNN architecture. In Section 3, we describe the calibration procedure for the transmissive SLM being used in our system. The DNN training and validation results are shown in Section 4. The reconstruction results on a focus star phase target is presented in Section 5 and a summary is given in Section 6.

2. SYSTEM ARCHITECTURE

In this section, we describe the detailed architecture of our phase microscopy system, including the optical setup for capturing the diffraction images and the deep neural network (DNN) for reconstructing the phase objects.

2.1 Optical setup

Our optical setup is shown in Figure 1. Light from a He-Ne laser source (Research Electro-Optics, Model 30995, 633nm) first transmits through a spatial filter, which consists of a microscope objective (Newport, M-60X, 0.85NA) and a pinhole aperture (D = $5\mu m$), to remove spatial noise. After being collimated by the lens (f = 200mm), the light hits on the first beam splitter (BS1) and is split into two arms: training arm and testing arm. The light in the training arm then passes through a linear polarizer (Thorlabs, LPVISC100-MP2), a transmissive spatial light modulator (Holoeye, LC2012) and a linear polarization analyzer (Thorlabs, LPVISC100-MP2) consecutively. The light in the testing arm is first reflected by a mirror (M1) and then illuminates on the phase object located at the sample plane. The light transmits through the object is collected by a microscope objective (Nikon, 20X, NA=0.4) and then passes through a tube lens (Thorlabs, TTL200-A), before being reflected by another mirror (M2). The beams from the two arms are combined together by the second beam splitter (BS2) and collected by a CMOS camera (Basler, A504k). The distance (optical path length) between the SLM and the CMOS is set to be $z = 495$mm.

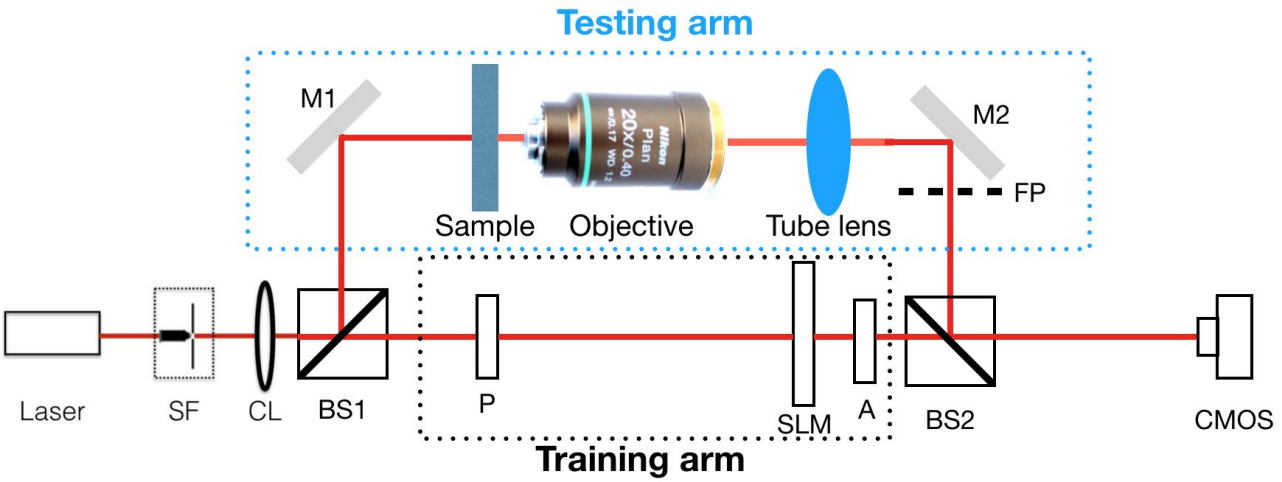

Figure 1. Optical setup of our system. SF: spatial filter; CL: collimating lens; M: mirror; P: linear polarizer; A: linear polarization analyzer; BS: beam splitter; SLM: spatial light modulator; FP: focal plane of the tube lens.

In the training arm, the spatial light modulator (SLM) acts as a phase object. According to its user manual, the LC2012 SLM can realize (approximate) pure-phase modulation if we modulate the light polarization properly. Specifically, for He-Ne laser light, if we set the incident beam to be linearly polarized at 310° with respect to the vertical direction and also set the linear polarization analyzer to be oriented at 5° with respect to the vertical direction, then the amplitude modulation of the SLM will become almost independent of the assigned (8-bit gray-level) input. In this arrangement, the phase modulation of the SLM follows a monotonic relationship with the assigned pixel value (with maximum phase depth: $\sim 1.2\pi$). We experimentally evaluated the correspondence between 8-bit grayscale input images projected onto the SLM and phase values in the range $[0, 1.2\pi]$. The calibration results are shown in Section 3. The SLM pixel size is $36\mu m \times 36\mu m$ and the number of pixels is

1024×768. In our experiment, only the central 160×160 portion is used. The CMOS pixel size is $12\mu m \times 12\mu m$ and the number of pixels is 1280×1024, which is cropped to a 512×512 square for processing.

In the testing arm, the objective and the tube lens actually constitute a wide-field microscopy module, which will generate a magnified image of the phase object at the focal plane of the tube lens (denoted as FP in Figure 1). While aligning the system, we make the optical path length between the FP and the CMOS to be equal to that between the SLM and the CMOS. As a result, the underlying model for the light propagation from FP to the CMOS should be the same as that for the propagation from the SLM to the CMOS. Therefore, being trained with the phase objects generated by the SLM and their respective diffraction patterns captured by the CMOS, the DNN should be able to reconstruct the magnified phase image at FP given its corresponding diffraction pattern.

2.2 Deep neural network

The same as Sinha's work, the DNN used in this paper utilizes a convolutional residual neural network (ResNet) architecture.[20] A diagram of our specific DNN architecture is shown in Figure 2. The input layer is the diffraction patterns captured by the CMOS. It is then successively decimated by 7 residual blocks of convolution + downsampling followed by 5 residual blocks of deconvolution + upsampling, 1 resizing layer and finally 2 standard residual blocks. Three of the down-residual blocks are comprised of dilated convolutions[21] so as to increase the receptive field of the convolution filters. The resizing layer performs bilinear interpolation of the input signal. We use skip connections to pass high frequency information learnt in the initial layers down the network towards the output reconstruction. At the very last layer of our DNN, the values represent an estimate of the phase object that we want to recover. The detailed structures of the residual blocks used in this paper are the same as those reported before.[7]

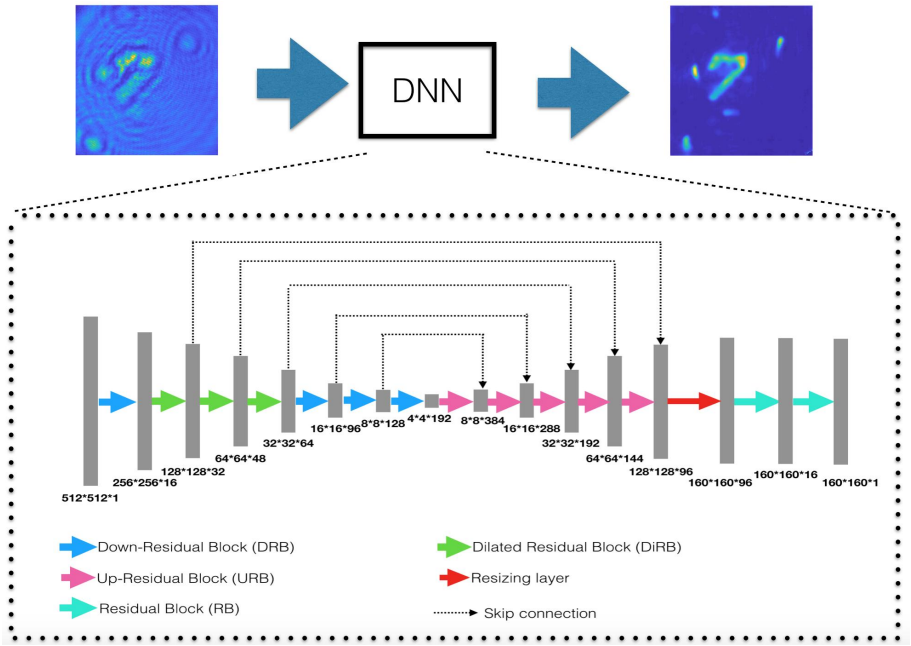

Figure 2. Detailed schematic of our DNN architecture.

While training the DNN, we use mean absolute error (MAE) as the loss function, which is defined as:

$$\mathrm{MAE} = \frac{1}{wh} \sum_{i=1}^{w} \sum_{j=1}^{h} |Y(i,j) - G(i,j)| \tag{1}$$

Here, w, h are the width and height of the output, Y is the output of the last layer, and G is the ground truth.

The DNN is trained using a GPU card (Nvidia, GTX1080). We used weight decay of $1E-4$ in all convolutional filters initialized with random numbers from a Gaussian distribution. We set a small dropout rate of 0.05 between all layers to prevent overfitting. We used the ADAM optimizer in Tensorflow to minimize the MAE over the training samples with batch size 4. We start the training with a learning rate of 0.001 and drop it by a factor of 2 after every 5 epochs. Additionally, we clip the gradients at value 1 to stabilize the training. We trained the neural network for 20 epochs shuffling the training samples at every epoch.

3. SLM CALIBRATION

In this section, we describe the approaches that we use to calibrate the intensity modulation and phase modulation curves of the SLM used in the system.

3.1 Intensity modulation

The optical setup for calibrating the intensity modulation of the SLM is shown in Figure 3(a). We use a photon diode sensor (Newport, 818-SL) as the detector. All the pixels of the SLM is driven by the same value V (uniform) and the corresponding intensity values measured by the photon diode is recorded. By changing V from 0 to 255 and repeating the measurement, we obtain the intensity modulation curve as shown in Figure 4(a), which is normalized to the intensity values at $V = 0$. The maximum intensity modulation ratio measured was ~ 2.7 for our configuration of polarizer and analyzer. In contrast, per manual specification, the SLM's maximum intensity modulation ratio is ~ 1300 for other polarizer/analyzer configurations. Therefore, we consider this intensity modulation is negligible and the SLM can be approximated as a pure phase object.

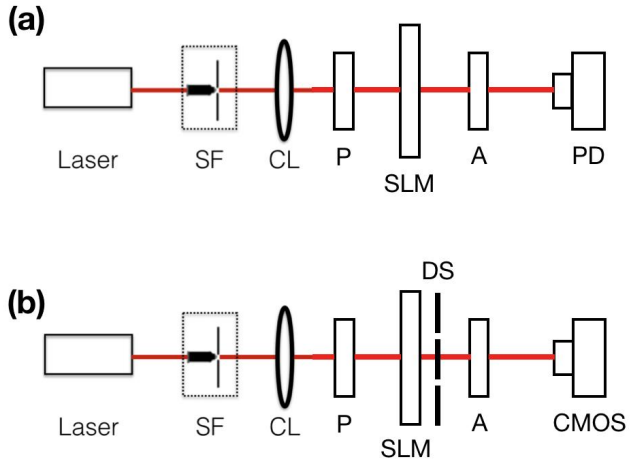

Figure 3. The optical setup for calibrating the (a) intensity and (b) phase modulation of SLM. SF: spatial filter; CL: collimating lens; P: linear polarizer; A: linear polarization analyzer; SLM: spatial light modulator; DS: double slits; PD: photon diode sensor.

3.2 Phase modulation

To calibrate the phase modulation of the SLM, we build the optical setup as shown in Figure 3(b). A double-slits mask is placed immediately after the SLM, making the system a Young's interferometer. As a result, a set of interference fringes can be observed at the CMOS plane. The SLM is driven in the half-half mode: the left-hand half pixels is driven by a constant value zero and the right-hand half is driven by a different value V. Under paraxial approximation, the location of the central peak of the interference pattern (constructive interference) can be determined as:

$$x = \frac{\phi(V)\lambda L}{2\pi d} \tag{2}$$

where λ is the light wavelength, d is the spacing between the two slits, L is the distance between the mask and the CMOS, $\phi(V)$ is the phase modulation of SLM when it is driven by value V [Assuming $\phi(0) = 0$rad].

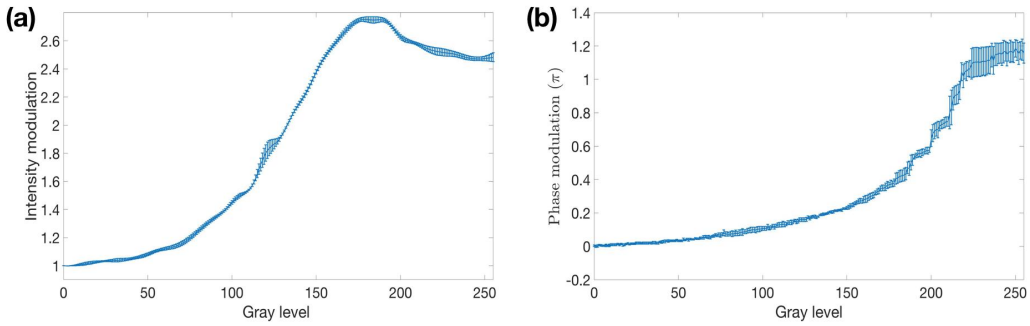

Figure 4. Experimentally calibrated (a) intensity modulation and (b) phase modulation curve with error bounds in the grayscale range of [0,255] for the SLM.

Therefore, by gradually increasing V from 0 to 255 and finding the central peak locations of the corresponding interference patterns, we can determine $\phi(V)$.

In our experiment, we use a commercial double-slits mask (3B Scientific) with spacing $d = 250\mu m$ and slit width $d = 150\mu m$. The distance between the mask and the CMOS is set to be $L = 1000$mm. The evolution of 1D profile of the fringes as V increases from 0 to 255 is shown in Figure 5, and the calibrated phase modulation curve is shown in Figure 4(b).

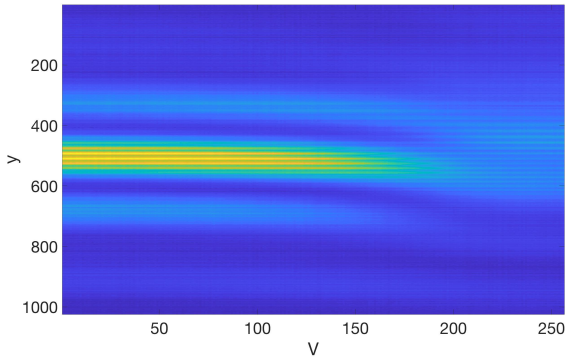

Figure 5. Evolution of 1D profile of the fringes as V increases from 0 to 255.

4. NETWORK TRAINING AND VALIDATION

To collect the data for training the DNN, we block the testing arm of the system. As a result, only the light from the SLM is captured by the CMOS. Our experiment consists two phases: training and validation. During the training phase, we modulate the phase SLM according to samples randomly selected from the ImageNet database. Each image is resized to 160×160 and then zero-padded to 1024×768 before being displayed on the SLM. The corresponding diffraction patterns are captured by the CMOS. As mentioned in Section 2, we only crop the central 512×512 square for processing. The captured diffraction patterns are normalized before being feed into the DNN. Our training set consisted of 10000 such images - diffraction pattern pairs. In the validation phase, we sample disjoint examples from the ImageNet database[22] and other databases such as Faces-LFW,[23] Characters, MNIST,[24] CIFAR[25] and Faces-ATT.[26] We upload these validation examples to the SLM and capture their corresponding diffraction patterns. We then feed these diffraction patterns into our trained DNN and compare the output to the ground truth.

The quantitative results of our trained DNN are shown in Figure 6. We can find that low MAEs are achieved not only for the validation images coming from the ImageNet database, but also for the images coming from other databases. This result indicates that our DNN generalizes well across different domains. As shown in

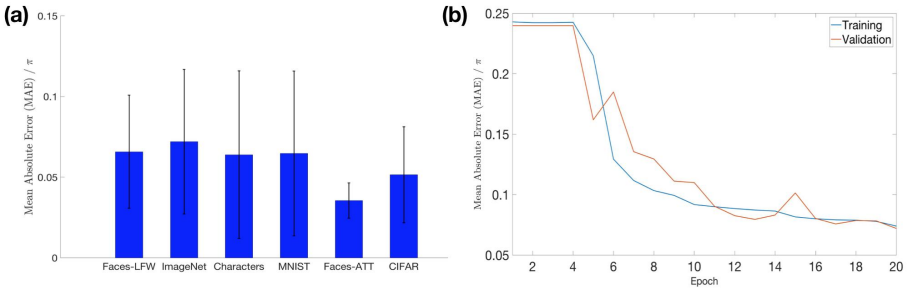

Figure 6. Quantitative analysis of our trained deep neural networks. (a) Validation errors in 6 different databases. (b) The training and validation error curves over 20 epochs.

Figure 6 (b), the training and validation curves converge to almost the same level. This also suggests that our DNN is not overfitted to the training data.

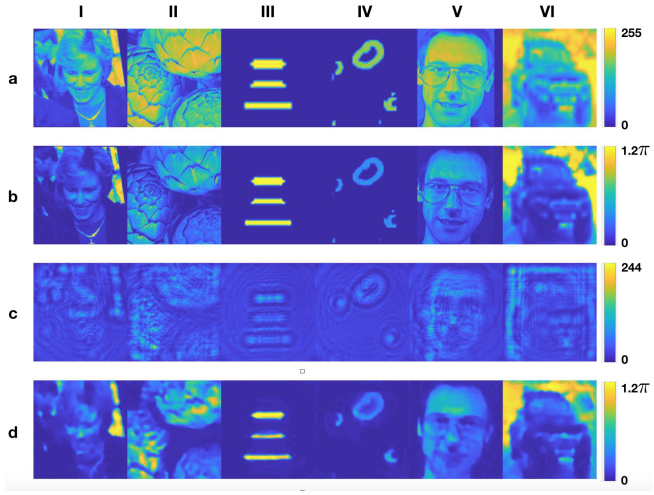

Figure 7. Qualitative analysis of our trained deep neural networks. (a) Ground truth pixel value inputs to the SLM. (b) Corresponding phase images calibrated by SLM response curve shown in Figure 4. (c) Raw intensity images captured by CMOS. (d) DNN reconstruction. Columns (i-vi) correspond to the dataset from which the validation image is drawn: (i) Faces-LFW, (ii) ImageNet, (iii) Characters, (iv) MNIST Digits, (v) Faces-ATT, or (vi) CIFAR, respectively.

The qualitative results of our trained DNN are shown in Figure 7. As expected, our DNN is able to reconstruct the phase profiles of images from different classes, despite the fact that it was trained exclusively on images from the ImageNet database. Therefore, our DNN has indeed learned a model of the underlying physics of the imaging system or at the very least a generalizable mapping of low-level textures between our output diffraction patterns and input images.

5. PHASE TARGET EXPERIMENT

Now, we open the testing arm and block the training arm to capture the diffraction patterns of actual phase objects through the wide-field microscopy module. Similar to the procedure in collecting the training and validation data, the captured diffraction patterns are cropped and normalized, before being feed into the trained DNN for phase retrieval.

The experimental results are shown in Figure 8. As shown in subplot (a), the object that we use is a commercial phase target of the focus star shape (Benchmark Technologies). Specifically, this phase target is made of an acrylate polymer ($n = 1.52$) on Corning Eagle XG Glass. The height of the target is 335nm, which results in a phase delay of $\phi = 0.55\pi$ for the light with wavelength $\lambda = 633$nm. Due to the limited field of view (FOV) of the imaging system, only the object within the region indicated by the red box in (a) is captured

by the CMOS and the resulting diffraction pattern in shown in (b). Using the DNN trained in Section 4, we obtain the retrieved phase profile for the phase target, as shown in (c). We can find that the reconstruction is qualitatively similar to the ground truth. However, in terms of quantitative phase retrieval, the error is still large. The reconstructed phase values are close to 0.55π in some regions, as shown in (d), but are very different from the ground truth in some other regions.

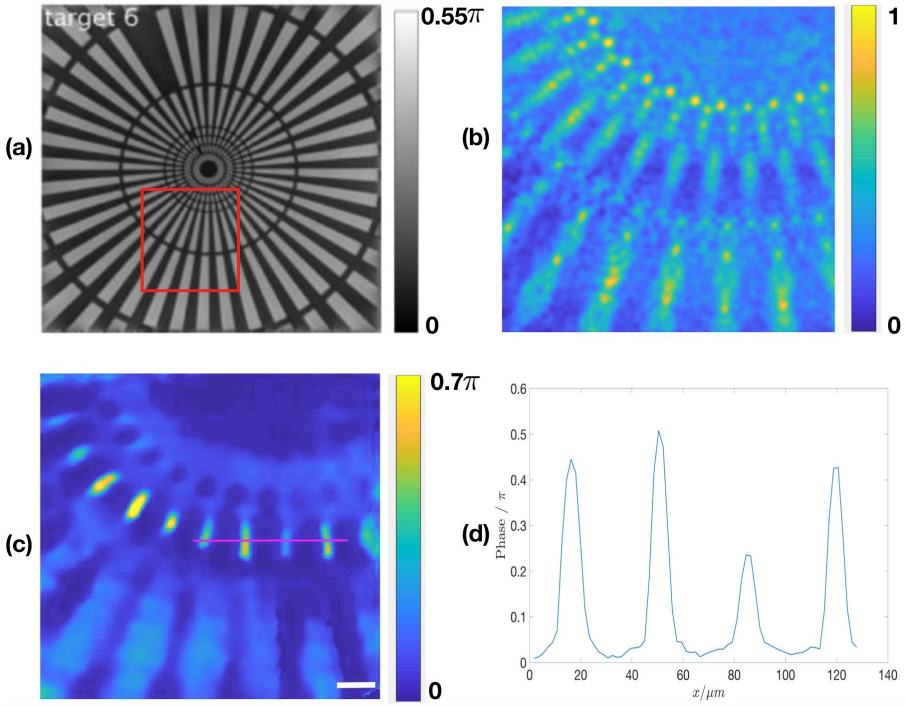

Figure 8. Phase target experiment. (a) The focus star phase object. (b) The diffraction pattern captured by the CMOS. (c) The recovered phase profile by the DNN. Scale bar: $30\mu m$. (d) 1D cross-section along the red line in (c).

The unsatisfactory performance of our DNN in reconstructing this phase target can be explained in the following aspects. First, due to the wide-field microscopy module, the light intensity in the testing arm is much lower than that in the training arm. As a result, the signal to noise ratio (SNR) of the captured diffraction pattern is small, making it difficult for the DNN to do the reconstruction. Second, what the DNN tends to reconstruct is actually the phase image at FP. However, due to the limited numerical aperture of the microscopy module, the phase image at FP is not simply a magnified version of the phase target, but a low-pass filtered version. As a result, a binary phase target at the sample plane will no longer generate a binary image at FP. In addition, the aberrations caused by the microscopy module will also make the FP phase image distorted. Third, the training data are generated by the transmissive SLM, which is not a perfect phase object. It has a coupled intensity modulation with maximum ratio ~ 2.7 and a small fill factor of 58%. These will also influence the performance of the DNN. Last, there will be inevitably mismatch between the training arm and testing arm, e.g. different noise levels, different aberrations, misalignment,..., which will make the two forward models become different. Further improvement is required to reduce the influence of all the above mentioned factors to achieve a good phase retrieval quality.

6. SUMMARY

In this paper, we apply deep learning to quantitative phase microscopy. Our DNN, which is trained using a transmissive SLM, is able to qualitatively reconstruct the profile of a phase target. But quantitatively, large error still exists. Possible reasons for this unsatisfactory performance are discussed. Further improvement is still required to achieve a better performance.

ACKNOWLEDGMENTS

We thank Alvaro Fernandez Galiana for his help in calibrating the SLM. This research was funded by the Singapore National Research Foun- dation through the SMART program (Singapore-MIT Alliance for Research and Technology) and by the Intelligence Advanced Research Projects Activity (iARPA) through the RAVEN Program. Justin Lee acknowledges funding from the U.S. Department of Energy Computa- tional Science Graduate Fellowship (CSGF) (DE-FG02-97ER25308).

REFERENCES

[1] Zernike, F., "Phase contrast, a new method for the microscopic observation of transparent objects part ii," *Physica* **9**(10), 974IN1981IN3983–980982986 (1942).

[2] Kou, S. S., Waller, L., Barbastathis, G., and Sheppard, C. J., "Transport-of-intensity approach to differential interference contrast (ti-dic) microscopy for quantitative phase imaging," *Optics letters* **35**(3), 447–449 (2010).

[3] Carré, P., "Installation et utilisation du comparateur photoélectrique et interférentiel du bureau international des poids et mesures," *Metrologia* **2**(1), 13 (1966).

[4] Creath, K., "Phase-shifting speckle interferometry," *Applied Optics* **24**(18), 3053–3058 (1985).

[5] Teague, M. R., "Deterministic phase retrieval: a greens function solution," *JOSA* **73**(11), 1434–1441 (1983).

[6] Goodman, J. W. and Lawrence, R., "Digital image formation from electronically detected holograms," *Applied physics letters* **11**(3), 77–79 (1967).

[7] Sinha, A., Lee, J., Li, S., and Barbastathis, G., "Lensless computational imaging through deep learning," *Optica* **4**, 1117–1125 (Sep 2017).

[8] Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., and Wang, G., "Recent advances in convolutional neural networks," *arXiv preprint arXiv:1512.07108* (2015).

[9] Krizhevsky, A., Sutskever, I., and Hinton, G. E., "Imagenet classification with deep convolutional neural networks," in [*Advances in neural information processing systems*], 1097–1105 (2012).

[10] Simonyan, K. and Zisserman, A., "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556* (2014).

[11] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A., "Going deeper with convolutions," in [*Proceedings of the IEEE conference on computer vision and pattern recognition*], 1–9 (2015).

[12] Girshick, R., Donahue, J., Darrell, T., and Malik, J., "Rich feature hierarchies for accurate object detection and semantic segmentation," in [*Proceedings of the IEEE conference on computer vision and pattern recognition*], 580–587 (2014).

[13] Ren, S., He, K., Girshick, R., and Sun, J., "Faster r-cnn: Towards real-time object detection with region proposal networks," in [*Advances in neural information processing systems*], 91–99 (2015).

[14] Li, H., Li, Y., Porikli, F., et al., "Deeptrack: Learning discriminative feature representations by convolutional neural networks for visual tracking.," in [*BMVC*], **1**(2), 3 (2014).

[15] Tompson, J. J., Jain, A., LeCun, Y., and Bregler, C., "Joint training of a convolutional network and a graphical model for human pose estimation," in [*Advances in neural information processing systems*], 1799–1807 (2014).

[16] Lyu, M., Wang, H., Li, G., and Situ, G., "Exploit imaging through opaque wall via deep learning," *arXiv preprint arXiv:1708.07881* (2017).

[17] Li, S., Deng, M., Lee, J., Sinha, A., and Barbastathis, G., "Imaging through glass diffusers using densely connected convolutional networks," *arXiv preprint arXiv:1711.06810* (2017).

[18] Rivenson, Y., Zhang, Y., Gunaydin, H., Teng, D., and Ozcan, A., "Phase recovery and holographic image reconstruction using deep learning in neural networks," *arXiv preprint arXiv:1705.04286* (2017).

[19] Jin, K. H., McCann, M. T., Froustey, E., and Unser, M., "Deep convolutional neural network for inverse problems in imaging," *IEEE Transactions on Image Processing* **26**(9), 4509–4522 (2017).

[20] He, K., Zhang, X., Ren, S., and Sun, J., "Deep residual learning for image recognition," in [*Proceedings of the IEEE conference on computer vision and pattern recognition*], 770–778 (2016).

[21] Yu, F. and Koltun, V., "Multi-scale context aggregation by dilated convolutions," *arXiv preprint arXiv:1511.07122* (2015).
[22] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al., "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision* **115**(3), 211–252 (2015).
[23] Huang, G. B., Ramesh, M., Berg, T., and Learned-Miller, E., "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," tech. rep., Technical Report 07-49, University of Massachusetts, Amherst (2007).
[24] LeCun, Y., Cortes, C., and Burges, C. J., "Mnist handwritten digit database," *AT&T Labs [Online]. Available: http://yann. lecun. com/exdb/mnist* **2** (2010).
[25] Krizhevsky, A. and Hinton, G., "Learning multiple layers of features from tiny images," (2009).
[26] "At&t database of faces," tech. rep., AT&T Laboratories Cambridge.