



Learning-based nonparametric autofocusing for digital holography

ZHENBO REN,¹ ZHIMIN XU,² AND EDMUND Y. LAM^{1,*}

¹Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam, Hong Kong China

²SharpSight Limited, Hong Kong, China

*Corresponding author: elam@eee.hku.hk

Received 23 January 2018; revised 28 February 2018; accepted 28 February 2018 (Doc. ID 320511); published 22 March 2018

In digital holography, it is crucial to extract the object distance from a hologram in order to reconstruct its amplitude and phase. This is known as autofocusing, which is conventionally solved by first reconstructing a stack of images, and then the sharpness of each reconstructed image is computed using a focus metric such as entropy or variance. The distance corresponding to the sharpest image is considered the focal position. This approach, while effective, is computationally demanding and time-consuming. To cope with this problem, we turn to machine learning, where we cast the autofocusing as a regression problem, with the focal distance being a continuous response corresponding to each hologram. Therefore, distance estimation is converted to hologram prediction, which we solve by designing a powerful convolutional neural network trained by a set of holograms acquired *a priori*. Experimental results show that this allows fast autofocusing without reconstructing an image stack, even when the physical parameters of the optical setup are unknown. © 2018 Optical Society of America under the terms of the OSA Open Access Publishing Agreement

OCIS codes: (090.1995) Digital holography; (110.1758) Computational imaging; (110.3010) Image reconstruction techniques.

<https://doi.org/10.1364/OPTICA.5.000337>

1. INTRODUCTION

Digital holography (DH) is a powerful imaging technique that can capture the diffracted wavefront of a three-dimensional (3D) object by recording the interference pattern with an electronic sensor [1,2]. With the hologram preserving the entire complex wavefront, one can reconstruct both the amplitude and phase information by back-propagating to a proper distance. As a non-invasive and label-free imaging technique, DH has been applied to biological imaging [3,4], MEMS defect inspection [5], and surface topography [6,7].

A fundamental problem in DH is to obtain the exact location of the object, a process known as autofocusing. In certain applications, such as a continuous monitoring of living specimens, autofocusing allows for tracking of axial movements; in other industrial uses, it provides a means to measure the surface profile [8]. Autofocusing is also critical for robust imaging against unstable environmental conditions [9]. Furthermore, in numerical reconstruction, we need to find the true location of an object in order to reconstruct an in-focus and sharp image. All in all, numerically searching for the object distance from a hologram would benefit various subsequent reconstruction, such as sectioning [10], extended focused imaging [11], and 3D imaging [12].

Specific optical configurations with additional components in the setup to handle autofocusing have been developed for digital holographic microscopy [13,14]. However, more often, autofocusing is handled computationally. Several algorithms, based on magnitude differential [8], variance [15,16], entropy [17], structure

tensor [18], and edge sparsity [19] have been proposed in recent years. These methods, however, all require sequential numerical reconstructions within an estimated distance range. An image-based focus metric is then used to compute the sharpness of each reconstructed image, and the position that corresponds to the sharpest one is considered the focal distance. Often, in practical applications, a refinement with a shorter step is needed after the coarse search in order to improve the accuracy. Such approaches, while effective, are computationally demanding and time-consuming, especially for large holograms and a small step width. It should also be mentioned that Oh *et al.* have proposed a frequency-based method to estimate the focal distance without any numerical reconstruction. Yet, the authors acknowledge that their method cannot be applied to objects with multiple distances [20], which severely limits its use.

In addition to autofocusing for amplitude objects, it is desirable to develop techniques that can work for pure phase objects. However, due to the wrapped phase values' modulo 2π , phase jumps are often misinterpreted as sharp structures, leading to further difficulty in autofocusing using the above approaches. References [9,21] propose four sharpness-based methods for phase contrast DH. However, they require sequential unwrapped phase images to be reconstructed first, which consume even more time than the sequential reconstructions for the amplitude-only objects due to phase unwrapping.

Since the hologram records the entire wavefront information of the object, we aim to extract the distance parameter directly

from it without back-propagation. The method we propose to use is based on deep learning, which has shown to be useful for many problems ranging from computer vision [22,23] to medical image analysis [24]. Deep learning for holographic imaging is a nascent area. Kamilov *et al.* sparked interest in a learning approach to imaging by showing that they can recover the phase from scattered light using a layered structure of a neural network in a tomographic configuration [25]. Sinha *et al.* applied deep neural networks to solve inverse problems in computational imaging, demonstrating through a setup that has some resemblance to DH and transport-of-intensity imaging [26]. Rivenson *et al.* constructed a deep neural network to eliminate twin-image and self-interference artifacts for the in-line holography setup. To generate the training data, which consist of the reconstructed images, an autofocusing algorithm based on the axial magnitude differential is used for a coarse scan, and a golden section search algorithm is then applied for refinement [27]. Nguyen *et al.* employ a simplified *U*-net model to generate a binary mask for phase aberration compensation. Afterwards, the phase map is reconstructed using an angular spectrum method, where the location of the sample is still needed either by manual measurement or autofocusing [28]. As for autofocusing, Pitkääho *et al.* propose to use the AlexNet architecture to estimate the focal position in holography. However, the hologram has to be pre-processed first to remove the zero-order and twin terms and subsequently back-propagated to a set of manually selected axial positions. Then, the in-focus depth is found among the reconstructed images using deep learning [29]. More recently, in Ref. [30], autofocusing is treated as a classification problem tackled by deep learning. Although effective, this technique assumes a discrete set of distances, which is not as versatile as the method being proposed in the current work.

In this paper, we harness the convolutional neural network (CNN) to achieve autofocusing in DH. For a sectional object, the focal distance of the individual section is regarded as a response of the hologram. Thus, we transform the problem of estimating the section distance to predicting a hologram, making it equivalent to a regression problem that can be tackled with machine learning effectively. After describing our method in detail, we quantitatively compare it with other learning-based algorithms, such as multilayer perceptron (MLP) [31], support vector machine (SVM) [32], and *k*-nearest neighbor (kNN) [33], as well as other model-based autofocusing algorithms. Experimental results show that the proposed CNN is capable of predicting the distance without reconstructing or knowing any physical parameters about the setup, and has better performance than other competing methods.

2. PRINCIPLE

A. Digital Holography

As shown in Fig. 1, the optical setup we use consists of a Mach-Zehnder interferometer. The beam emitting from the laser source (He-Ne with 632.8 nm) is filtered and collimated using a spatial filter and a lens. Then, the beam size is expanded, and it enters the interferometer by splitting into two paths. One is the reference beam and the other is the object beam, which carries the information of the object. Two half-wave plates are placed in the setup to adjust the intensity ratio of the two beams. At the exit of the interferometer, a hologram is created by the interference between

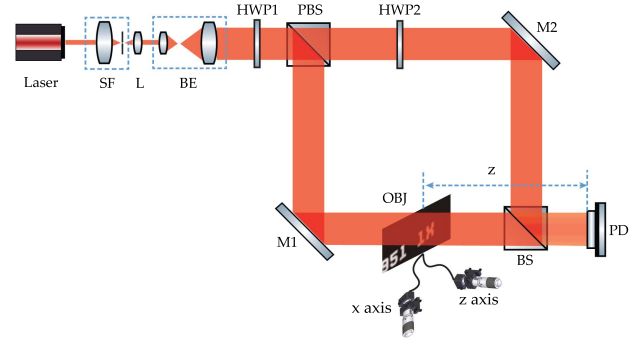


Fig. 1. Schematic diagram of a DH system. SF stands for the spatial filter. L is the collimation lens. BE is the beam expander. HWP1 and HWP2 are the half-wave plates. PBS and BS are the polarization and non-polarization beam splitters, respectively. M1 and M2 are the mirrors. OBJ is the object. PD is the camera. *x* axis and *z* axis denote the two motion controllers along the two directions. *z* is the distance between the object and the camera. The object shown here is a small region of a negative USAF 1951 resolution chart.

the object and the reference waves, and this is recorded with a camera (pixel pitch is 5.2 μm).

In a regular DH setup, after data acquisition, the hologram is back-propagated to an estimated distance to reconstruct both the amplitude and phase distributions. For our experiment, we need to train an algorithm with an extensive set of hologram data. Thus, two linear motion controllers (Newport, CONEX-LTA-HL, typical absolute accuracy is ±1.2 μm) are used to precisely control the movement of the object. The main controller, annotated as “*z* axis,” is used to move the object axially, while the other one is used to move laterally.

B. Deep-Learning-Based Method

CNN has shown to be powerful for a variety of recognition, classification, and segmentation tasks [22,34]. Conventionally, it is viewed as an extension of the multilayer neural network consisting of convolutional layers followed by one or more fully connected layers. Such an architecture has the advantages of being shift, scale, and distortion invariant, making it especially suitable for image processing applications [35]. To the best of our knowledge, this is the first report of using CNN to estimate the focal distance of an object and thereby achieving autofocusing in digital holographic imaging.

Motivated by LeNet, which does not require any knowledge of the viewing geometry [36], we propose the framework as shown in Fig. 2. The CNN architecture consists of several functional layers: convolution layer, pooling layer, fully connected layer, and output layer. For the ℓ th convolutional layer, suppose there are $N^{(\ell)}$ feature maps with a uniform size of $k \times k$, which can be denoted as $h_j^{(\ell)}$, for $j = 1, 2, \dots, N^{(\ell)}$. The convolutional layer can then be represented as

$$h_j^{(\ell)} = \psi \left(\sum_{i=1}^{N^{(\ell-1)}} h_i^{(\ell-1)} * w_{ij}^{(\ell)} + b_j^{(\ell)} \right), \quad (1)$$

where w_{ij} and b_j are the weight and bias that need to be learned through training, and $\psi(\cdot)$ denotes an activation function. In this work, the rectified linear unit (ReLU), $\psi(x) = \max(0, x)$,

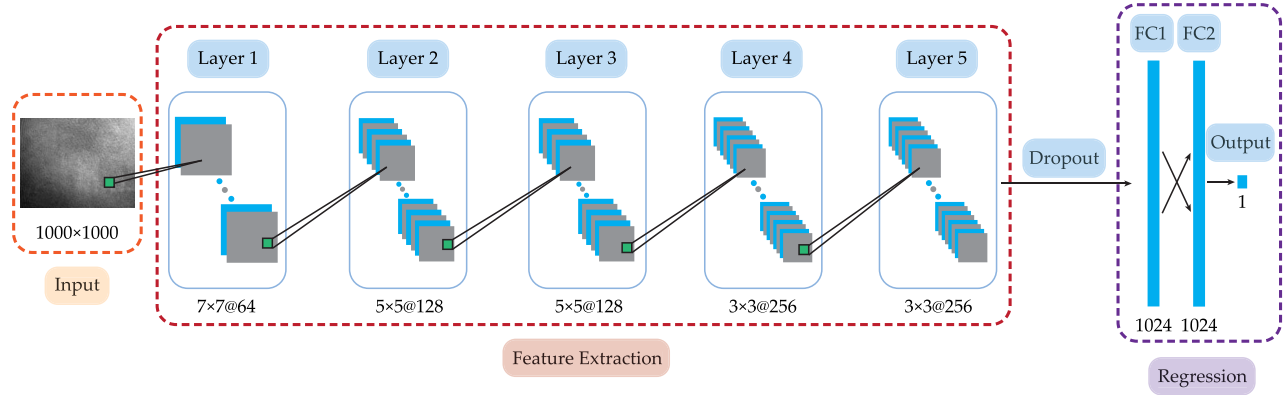


Fig. 2. Framework of the proposed CNN for autofocusing. In each “Layer,” a convolutional layer, a ReLU layer, a batch normalization layer, and a max-pooling layer are included. “FC1” and “FC2” represent fully connected layers, and “Dropout” means dropout layer. In “Input,” the input hologram size, which is cropped from the raw 1280×1024 image, is shown beneath. In “Feature Extraction,” the kernel size and depth are given at the bottom of each layer. In “Regression,” the input hologram is predicted with a response, denoting the estimated focal distance.

is used [35]. After each convolutional layer and application of the ReLU, we compute the batch normalization, using both mean and variance to normalize each batch to improve the performance of the proposed network [35]. The network then has a pooling layer, which downsamples the feature map, before the next convolutional layer. This operation can significantly reduce the spatial dimension of the representation and the number of parameters, and consequently the total amount of computation. It also helps to prevent the network from over-fitting. There are generally two pooling methods: max pooling and average pooling [35]. We make use of the former to obtain the maximum value as the representation among a small region. The combination of convolution and pooling layers is the basic building block in our network, which is repeated a total of five times, before leading to the final feature extraction module. These steps are grouped as the feature extraction module in Fig. 2.

The main purpose of this module is to identify the underlying characterization of the hologram data. Before the final classification, a dropout layer is added to prevent the network from over-fitting. At each training iteration, the individual nodes are either kept with a probability of p or discarded from the network with a probability of $1 - p$, so that consequently we have a reduced network [37]. Finally, the extracted feature representation is fed into the fully connected layer for regression analysis, and the output layer gives a predicted response, which is the axial distance estimate of the input hologram.

To train a deep CNN, we need a substantial collection of holograms and their true distances as the response assigned to each hologram. We only train with holograms resulting from several discrete distances, but through regression the resulting neural network can output a continuum of distances for the test cases. To compute the loss between the predicted quantity and the true value, we measure the square of the L_2 norm of their difference. Suppose \hat{y} is the estimated output and y is the corresponding true value, the square of the L_2 norm is the loss function \mathcal{L} given by

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^N \|y_i - \hat{y}_i\|_2^2, \quad (2)$$

where N is the total number of terms.

3. RESULTS

A. Evaluation Metrics and Training Details

We make use of three quantitative evaluation metrics, namely, mean absolute error (MAE), explained variance regression score function (EV), and R^2 (coefficient of determination) regression score function, for the assessment of the regression performance. These functions are defined as [38]

$$\text{MAE}(y, \hat{y}) = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|, \quad (3)$$

$$\text{EV}(y, \hat{y}) = 1 - \frac{\text{Var}\{y - \hat{y}\}}{\text{Var}\{y\}}, \quad (4)$$

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}, \quad (5)$$

where $\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i$, and $\text{Var}\{\cdot\}$ denotes the biased variance. The first function is an average measure of the absolute difference between two variables, and its best possible score is 0. The last two functions provide measurement of how well future samples are likely to be predicted by the model. The best possible scores of EV and R^2 are 1.0, and smaller values are worse; while for MAE, the range is unbounded.

We now evaluate the model based on the typical train-validation-test approach. The hologram data are randomly split into three subsets with a ratio of 75:15:10 for training, validation, and testing, respectively. The network is trained using the Adam optimizer [35], which is a form of gradient descent, where a parameter known as the learning rate needs to be set beforehand. Here, we set it empirically to 0.001, and allow it to decay exponentially with a rate of 0.9 as the training progresses. The dropout probability is set to be 0.75 in training, while in validation and testing it becomes 1, i.e., no dropout. This allows 25% of the nodes to be randomly chosen and intentionally disabled in training to reduce over-fitting, while keeping all nodes alive can effectively check how well the network has learned.

To tackle the issues of limited computer memory and stagnation in local minima during optimization, every time only a small batch of 8, called a mini-batch, of the entire training data is fed

into the network for training instead of the entire set of holograms. All the weights are initialized using truncated normalization, with a standard deviation of 1, and the biases are initialized with a constant value of 0.1. In each mini-batch training, one iteration of the optimization is performed, and the parameters of the network are updated. The whole training stage is stopped after 20 epochs training. We implement the CNN using TensorFlow [39], and all the experiments are performed on a Ubuntu 16.04.2 environment with CPU Intel Core i7@2.67 GHz and an Nvidia GTX 760.

B. Amplitude Object

The optical setup is shown in Fig. 1, and we select various samples, including different local areas of a negative USAF 1951 resolution chart and biological specimens, as the amplitude objects. In Fig. 3, we present several used areas of the test target and several biological specimens including a *testis slice*, a *ligneous dicotyledonous stems*, an *earthworm crosscut*, etc., for example. To control the movement of the object accurately during recording, we use two linear actuators for the optical axis and the lateral direction. At distances 250, 252, 254, 256, 258, 260, 263, 266, 269, and 272 mm, we collect 500 holograms with different lateral positions of various objects. Therefore, in total, we have 5000 holograms with 10 possible responses. In Fig. 4, we provide an example showing 16 of the holograms.

To train the network, a mini-batch of 8 holograms is fed into the network for every iteration. Then, for every 625 iterations (which is 1 epoch), the network is evaluated using the validation dataset. After finishing the training with a predetermined number of epochs, the network is assessed with the test subset. In Fig. 5, the validation loss is presented. As can be seen from the plot, the CNN converges gradually as the loss decreases along the training. This agrees with our expectation that the network is continuously updating its parameters and learning representative features of the holograms.

The quantitative comparisons using the three evaluation metrics with kNN, SVM, and MLP are given in Table 1. For kNN, we set k , which defines the number of nearest neighbors we wish to take vote from in regression, to range from 1 to 10 and select the value that gives the best performance, which is 5. For SVM, the kernel of the radial basis function is used, and the two parameters,

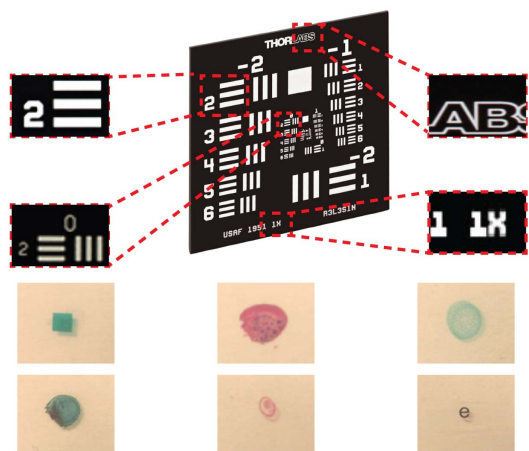


Fig. 3. USAF test target and its local area, as well as several biological specimens used in the experiment.

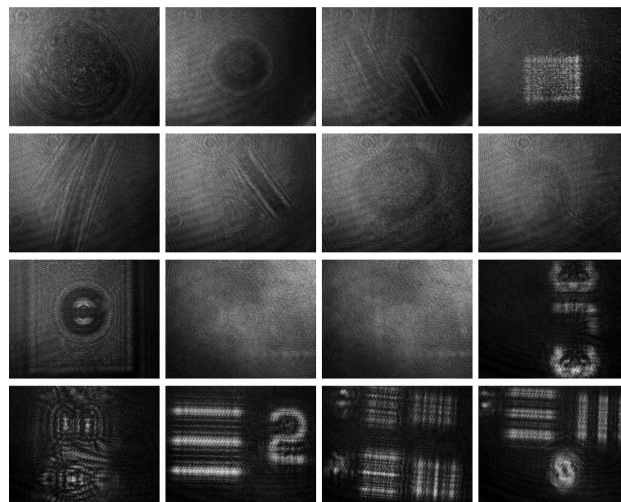


Fig. 4. Sixteen of the experimentally collected testing holograms recording various amplitude objects.

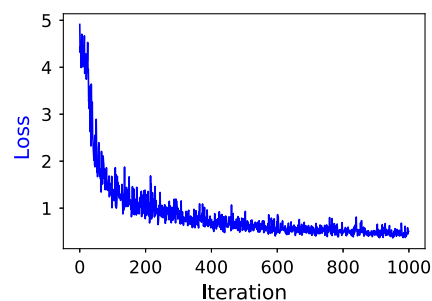


Fig. 5. Validation loss decreases along the training process. Only 1000 iterations are shown here.

γ (the standard deviation of the Gaussian Kernel) and C (penalty parameter of the error term) as given in Ref. [32], are set as 1 and 2 according to multiple trials. For MLP, we construct a five-layer neural network using the same weight initialization method and loss function as CNN to perform regression analysis. The results,

Table 1. Comparison of the Regression Performance on the Validation and Test Datasets among kNN, SVM, MLP, and CNN for the Amplitude Object

Measure	Methods	Amplitude Dataset	
		Validation	Test
MAE	kNN	1.63	1.54
	SVM	1.04	0.99
	MLP	0.98	1.12
	CNN	0.05	0.06
EV	kNN	0.3990	0.3922
	SVM	0.4528	0.4705
	MLP	0.5163	0.5016
	CNN	0.9843	0.9836
R^2	kNN	0.3041	0.2701
	SVM	0.6447	0.5922
	MLP	0.5727	0.5846
	CNN	0.9901	0.9907

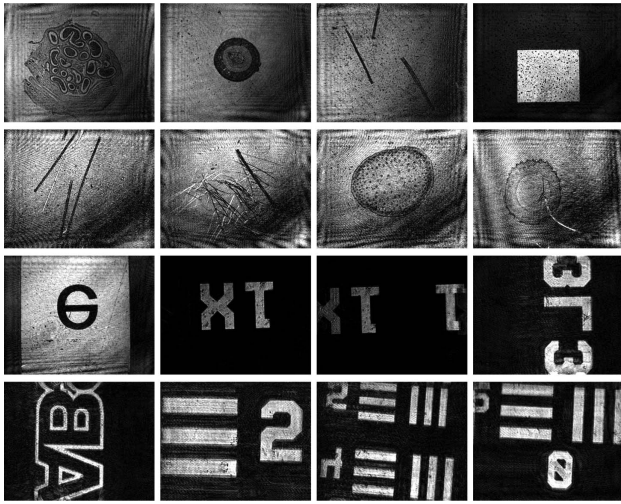


Fig. 6. Back-propagated images of the testing holograms in Fig. 4 using the predicted distances with CNN.

given in the table, indicate that these three comparison methods have very similar performance for autofocusing with generally large MAE and low EV and R^2 scores. In contrast, our CNN-based autofocusing delivers a much better performance in terms of MAE, with substantially smaller error. The EV score and R^2 score also corroborate with this finding, where their values are much closer to one. In addition, based on the predicted focal distance, in Fig. 6 we reconstruct the 16 images from the holograms presented in Fig. 4. The reconstructed images are sharp, further supporting the claim that CNN is able to extract the correct focal distance from the holograms directly.

C. Phase Object

Apart from autofocusing an amplitude object, we also collect experimental hologram data of a phase-only object. We use a customized groove with tiny structures on an optical wafer, where a magnified image using a microscope with a 4x objective is shown in Fig. 7. In total, we collect 2000 holograms at 270, 272, 274, 276, and 278 mm. Figure 8 shows several holograms (without the magnification as shown in Fig. 7) randomly chosen from the dataset. The network is trained in a similar fashion as before. The validation loss curve is shown in Fig. 9.

In Table 2, the quantitative comparisons with kNN, SVM, and MLP are presented. The parameters of the three regressors remain unchanged. We can observe that our CNN-based autofocusing again has the best performance among the four methods,

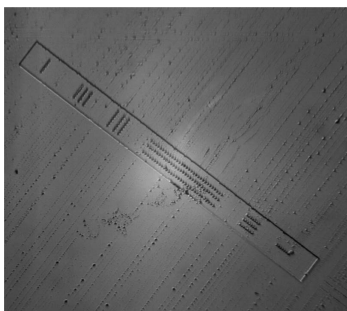


Fig. 7. Customized groove used as the phase object.

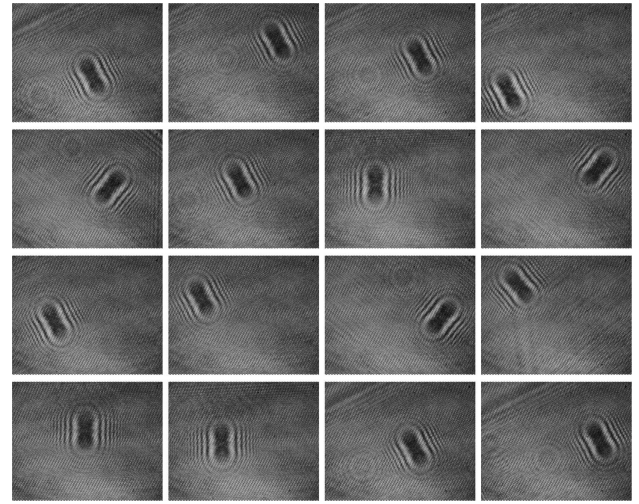


Fig. 8. Sixteen experimentally collected testing holograms of a phase object shown in Fig. 7.

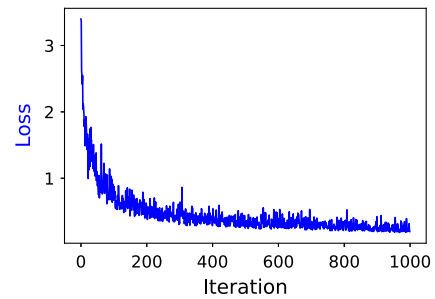


Fig. 9. Validation loss decreases and accuracy increases as the network is trained. Only 1000 iterations are shown here.

whether in terms of MAE, EV, or R^2 . Nevertheless, it can be noted that all methods generally have a better performance compared to Table 1 involving the amplitude object. This is because there is only one phase object used during recording in this case, while various amplitude objects are used for the collection of

Table 2. Comparison of the Regression Performance on the Validation and Test Datasets among kNN, SVM, MLP, and CNN for the Phase Object

Measure	Methods	Phase Dataset	
		Validation	Test
MAE	kNN	1.17	1.12
	SVM	1.01	1.06
	MLP	0.87	0.79
	CNN	0.03	0.02
EV	kNN	0.4471	0.4945
	SVM	0.6559	0.6940
	MLP	0.6633	0.6435
	CNN	0.9849	0.9838
R^2	kNN	0.4247	0.4417
	SVM	0.6287	0.6333
	MLP	0.6123	0.6388
	CNN	0.9911	0.9907

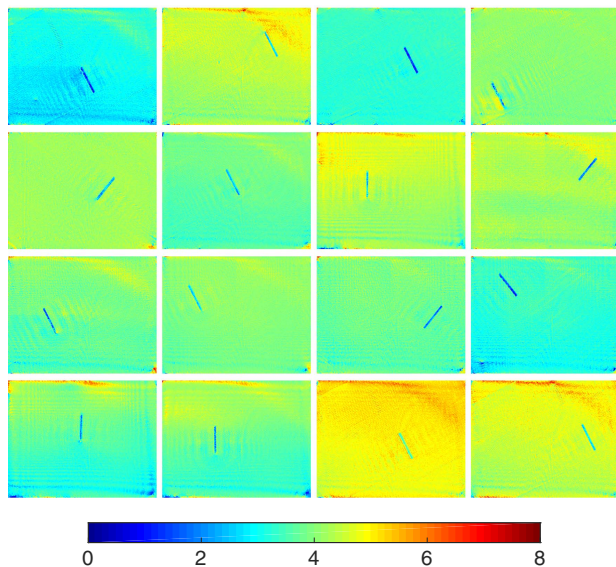


Fig. 10. Reconstructed and unwrapped phase images of the testing holograms in Fig. 8 using the predicted focal distances and double exposure method. The unit of the color bar is radian.

holograms in the previous experiment. This may result in reducing the difficulty of training a model.

Similarly, based on the predicted focal distances, in Fig. 10 we show the reconstructed and unwrapped phase images using convolution and double exposure [40] of the corresponding holograms in Fig. 8. The thickness of the groove on the optical wafer (fused silica, refractive index is 1.4585) is around 140 nm. With a He–Ne laser as the illumination source (wavelength is 632.8 nm), its phase difference with the wafer surface is about 2 radian. Thus, the phase of the groove is smaller than that of the surface, which is in agreement with what we expect of the true height. Nevertheless, there are still some artifacts in the unwrapped phase images. Since we are using double exposure to compensate for the phase aberration due to experimental convenience, the reference holograms are not captured instantly after their corresponding holograms. The noise introduced by the laser, the camera, and the ambiance can then differ with time. The phase images therefore are not identical; yet, by comparing the phase values of the groove and the surface in one phase image, the relative difference shows the correct estimate of the distance.

4. DISCUSSIONS

Here we further explore the capability of the trained network under various situations.

A. Different Exposure Times

Exposure time affects the contrast of the interference pattern. The holograms in the training set are recorded with a fixed exposure time of 12 ms, and we then record 20 additional holograms exposed for 0.5, 2, 6, 18, and 80 ms. These holograms are directly fed into the trained network for testing. Results show that, except for the two extreme cases with severe underexposure and overexposure where the images are very close to pure black and white images, the network can make correct estimates with a MAE score and R^2 score of around 0.04 and 0.97, respectively. In Fig. 11 we give two examples that are captured with an exposure time of 5 ms

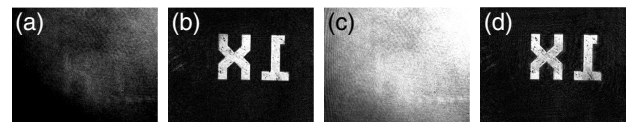


Fig. 11. Hologram and reconstructed image, respectively, with an exposure time of (a), (b) 5 ms and (c), (d) 18 ms.

and 18 ms, as well as their reconstructions with the individual predicted distances. This illustrates that even the interference pattern is rather dim or bright, the proposed method is robust enough to give an accurate prediction.

B. Different Axial Distances

Since we are training a regression network for autofocusing, it is natural to consider how well it can perform if an object is located at distances different from those in the training set. To test this, we collect 60 holograms where the objects are at integer distances between 259 and 271 mm inclusive, except for the specific distances used in the training. We feed these holograms into the trained network and compare the estimated output with the true target. Results show that our regression model can correctly produce the mapping function between the hologram and its corresponding response, and can give an accurate estimated distance. The MAE score and R^2 score are 0.06 and 0.97, respectively. In Fig. 12, we show several reconstructed images from this experiment, which support the generalization capability of the proposed autofocusing CNN.

When the object is located outside the training range, predicting its position directly using the trained network may lead to a significant error. However, this situation is rare in applications where the object is normally located in a fixed range. An example is in-line holography, in which the object has to be located between the point source and the detector. Many commercial DH products also have a designated object localization. Even for other situations, the problem can be mitigated with a rather straightforward solution, where we simply extend the range of collected distances. By doing so, it is possible to train a network that has a better and wider capability to deal with various holograms with various distances.

C. Different Incident Angles

In off-axis digital holography, prior to Fresnel back-propagation, the 0 and -1 spectra are normally filtered out, and only the $+1$ spectrum remains in the frequency domain. For different configurations, we then need to set the center of the $+1$ spectrum manually. In real experiments, the incident angle between the two beams varies and may not be the same as the angle used for training. Moreover, even for each acquisition, the angle may also change due to the adjustment of the fringe contrast. Since we compute the autofocusing using the raw holograms, it is critical to test the performance of the network under this situation of

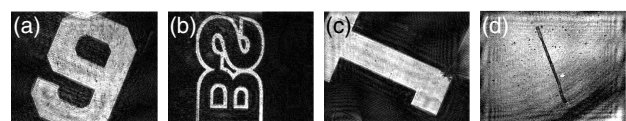


Fig. 12. Reconstructed images with the holograms recorded at different distances.

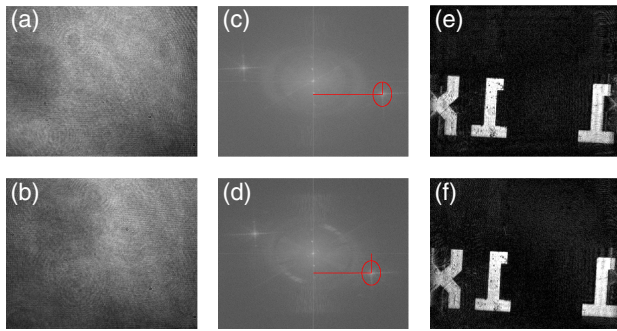


Fig. 13. (a), (b) Holograms, (c), (d) frequency spectra, and (e), (f) reconstructed images under different angles.

different incident angles. We slightly change the incident angle (by no more than 2–3 deg) 5 times and record 10 holograms for each angle. In Fig. 13, two holograms captured under different angles and their corresponding frequency spectra are shown. We can see that the +1 spectra of the two holograms are different, as annotated with the red markers.

These holograms are tested by the same trained network, which achieves an MAE score and R^2 score of 0.02 and 0.98, respectively. With the estimated distances computed from the neural network, the two holograms are then back-propagated, and the reconstructions are given in Figs. 13(e) and 13(f). This result illustrates that the network is capable of autofocusing with little regard for a variation in the incident angles. In other words, in an optical setup, even if the mirrors have a slight rotation, the proposed method can still perform well.

D. Comparison with Conventional Autofocusing Algorithms

We consider the advantages and disadvantages of a learning-based approach to DH autofocusing versus traditional image-sharpness-based techniques. The major strength of the former is certainly speed; no sequential reconstruction is needed at all, which significantly saves time in autofocusing. In addition, in conventional methods, we normally need the parameters of the optical setup such as the wavelength, pixel pitch of the camera, incident angle of the two beams, and sampling rate for a good numerical reconstruction. Unfortunately, some of them may not be known *a priori*, and this limits the applicability of many approaches that require computing the sharpness of the reconstructed images.

On the other hand, learning-based approaches require a sizable database consisting of the hologram data and the true distance as the label for training. However, once the network is trained, it has a very short prediction time compared to conventional autofocusing methods. In Table 3, quantitative comparisons in terms of absolute autofocusing error [19] and computational speed of the proposed CNN with several selected conventional methods, including integrated amplitude modulus (AMP), self-entropy (SEN), variance (VAR), gradient (GRA), summed Laplacian (LAP), Tenenbaum gradient (TEN), Gini of the gradient (GoG), and Tamura of the gradient (ToG), are presented. The detailed definition of each method can be found in Ref. [19]. The holograms in Fig. 4 are used to test the conventional metrics, and the average results are given in Table 3. The hologram has the same size as the input of the network, and the number of sampling steps for a stack of reconstructed images is 50. For a fair

Table 3. Comparison of Absolute Error and Computation Time among CNN and Conventional Methods on the Amplitude Samples

Method	Error (mm)	Time (s)
CNN	0.01	0.189
Integrated amplitude modulus (AMP)	1.6	5.324
Self-entropy (SEN)	1.6	5.035
Variance (VAR)	2.0	4.961
Gradient (GRA)	0.8	6.479
Summed Laplacian (LAP)	9.6	7.853
Tenenbaum gradient (TEN)	6.4	7.528
Gini of the gradient (GoG)	0.4	8.187
Tamura of the gradient (ToG)	0.4	7.231

comparison, conventional metrics are also run on the same GPU. The fact that CNN has the least error and shortest computation time demonstrates the superior performance of the proposed method over others.

5. CONCLUSIONS

In this paper, a deep-learning-based autofocusing method is proposed. Holograms of various amplitude and phase objects are collected to verify its effectiveness. Compared to conventional autofocusing algorithms and other machine learning methods, this approach outperforms without any numerical reconstruction in digital holography.

Funding. Research Grants Council, University Grants Committee (RGC, UGC) (N_HKU714/13, 17203217).

Acknowledgment. The authors thank Nan Meng at the University of Hong Kong for fruitful discussions and Dr. Ping Su at the Graduate School at Shenzhen, Tsinghua University for providing some samples.

REFERENCES

1. J. W. Goodman, *Introduction to Fourier Optics*, 4th ed. (W.H. Freeman, 2017).
2. U. Schnars, C. Falldorf, J. Watson, and W. Jüptner, *Digital Holography and Wavefront Sensing: Principles, Techniques and Applications* (Springer, 2015).
3. A. Doblas, E. Sánchez-Ortiga, M. Martínez-Corral, G. Saavedra, and J. Garcia-Sucerquia, "Accurate single-shot quantitative phase imaging of biological specimens with telecentric digital holographic microscopy," *J. Biomed. Opt.* **19**, 046022 (2014).
4. P. Marquet, C. Depeursinge, and P. J. Magistretti, "Review of quantitative phase-digital holographic microscopy: promising novel imaging technique to resolve neuronal network activity and identify cellular biomarkers of psychiatric disorders," *Neurophotonics* **1**, 020901 (2014).
5. Y. Pourvais, P. Asgari, P. Abdollahi, R. Khamedi, and A.-R. Moradi, "Microstructural surface characterization of stainless and plain carbon steel using digital holographic microscopy," *J. Opt. Soc. Am. B* **34**, B36–B41 (2017).
6. E. Cucho, P. Marquet, and C. Depeursinge, "Simultaneous amplitude-contrast and quantitative phase-contrast microscopy by numerical reconstruction of Fresnel off-axis holograms," *Appl. Opt.* **38**, 6994–7001 (1999).
7. D. J. Brady, K. Choi, D. L. Marks, R. Horisaki, and S. Lim, "Compressive holography," *Opt. Express* **17**, 13040–13049 (2009).
8. M. Lyu, C. Yuan, D. Li, and G. Situ, "Fast autofocusing in digital holography using the magnitude differential," *Appl. Opt.* **56**, F152–F157 (2017).

9. P. Langehanenberg, G. von Bally, and B. Kemper, "Autofocusing in digital holographic microscopy," *3D Res.* **2**, 1–11 (2011).
10. X. Zhang, E. Y. Lam, and T.-C. Poon, "Reconstruction of sectional images in holography using inverse imaging," *Opt. Express* **16**, 17215–17226 (2008).
11. Z. Ren, N. Chen, and E. Y. Lam, "Extended focused imaging and depth map reconstruction in optical scanning holography," *Appl. Opt.* **55**, 1040–1047 (2016).
12. A. C. Chan, K. K. Tsia, and E. Y. Lam, "Subsampled scanning holographic imaging (SuSHI) for fast, non-adaptive recording of three-dimensional objects," *Optica* **3**, 911–917 (2016).
13. P. Gao, B. Yao, J. Min, R. Guo, B. Ma, J. Zheng, M. Lei, S. Yan, D. Dan, and T. Ye, "Autofocusing of digital holographic microscopy based on off-axis illuminations," *Opt. Lett.* **37**, 3630–3632 (2012).
14. J. Zheng, P. Gao, and X. Shao, "Opposite-view digital holographic microscopy with autofocusing capability," *Sci. Rep.* **7**, 4255 (2017).
15. M. Subbarao and J. K. Tyan, "Selecting the optimal focus measure for autofocusing and depth-from-focus," *IEEE Trans. Pattern Anal. Mach. Intell.* **20**, 864–870 (1998).
16. H. A. İlhan, M. Doğar, and M. Özcan, "Digital holographic microscopy and focusing methods based on image sharpness," *J. Microsc.* **255**, 138–149 (2014).
17. Z. Ren, N. Chen, A. Chan, and E. Y. Lam, "Autofocusing of optical scanning holography based on entropy minimization," in *Digital Holography and Three-Dimensional Imaging* (Optical Society of America, 2015), paper DT4A-4.
18. Z. Ren, N. Chen, and E. Y. Lam, "Automatic focusing for multisectional objects in digital holography using the structure tensor," *Opt. Lett.* **42**, 1720–1723 (2017).
19. Y. Zhang, H. Wang, Y. Wu, M. Tamamitsu, and A. Ozcan, "Edge sparsity criterion for robust holographic autofocusing," *Opt. Lett.* **42**, 3824–3827 (2017).
20. S. Oh, C.-Y. Hwang, I. K. Jeong, S.-K. Lee, and J.-H. Park, "Fast focus estimation using frequency analysis in digital holography," *Opt. Express* **22**, 28926–28933 (2014).
21. P. Langehanenberg, B. Kemper, D. Dirksen, and G. von Bally, "Autofocusing in digital holographic phase contrast microscopy on pure phase objects for live cell imaging," *Appl. Opt.* **47**, D176–D182 (2008).
22. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems* (2012), pp. 1097–1105.
23. Y. LeCun, Y. Bengio, and G. E. Hinton, "Deep learning," *Nature* **521**, 436–444 (2015).
24. D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Ann. Rev. Biomed. Eng.* **19**, 221–248 (2017).
25. U. S. Kamilov, I. N. Papadopoulos, M. H. Shoreh, A. Goy, C. Vonesch, M. Unser, and D. Psaltis, "Learning approach to optical tomography," *Optica* **2**, 517–522 (2015).
26. A. Sinha, J. Lee, S. Li, and G. Barbastathis, "Lensless computational imaging through deep learning," *Optica* **4**, 1117–1125 (2017).
27. Y. Rivenson, Y. Zhang, H. Gunaydin, D. Teng, and A. Ozcan, "Phase recovery and holographic image reconstruction using deep learning in neural networks," *Light Sci. Appl.* **7**, 17141 (2018).
28. T. Nguyen, V. Bui, V. Lam, C. B. Raub, L.-C. Chang, and G. Nehmetallah, "Automatic phase aberration compensation for digital holographic microscopy based on deep learning background detection," *Opt. Express* **25**, 15043–15057 (2017).
29. T. Pitkäaho, A. Manninen, and T. J. Naughton, "Performance of autofocus capability of deep convolutional neural networks in digital holographic microscopy," in *Digital Holography and Three-Dimensional Imaging* (Optical Society of America, 2017), paper W2A-5.
30. Z. Ren, Z. Xu, and E. Y. Lam, "Autofocusing in digital holography using deep learning," *Proc. SPIE* **10499**, 104991V (2018).
31. M. W. Gardner and S. Dorling, "Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences," *Atmos. Environ.* **32**, 2627–2636 (1998).
32. J. Huang, J. Lu, and C. X. Ling, "Comparing naive Bayes, decision trees, and SVM with AUC and accuracy," in *Proceedings of International Conference on Data Mining* (IEEE, 2003), pp. 553–556.
33. K. Q. Weinberger, J. Blitzer, and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," in *Advances in Neural Information Processing Systems* (2006), pp. 1473–1480.
34. M. Jaderberg, A. Vedaldi, and A. Zisserman, "Deep features for text spotting," in *European Conference on Computer Vision* (Springer, 2014), pp. 512–528.
35. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT, 2016).
36. Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE* **86**, 2278–2324 (1998).
37. N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.* **15**, 1929–1958 (2014).
38. S. Menard, "Coefficients of determination for multiple logistic regression analysis," *Am. Stat.* **54**, 17–24 (2000).
39. M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. J. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Józefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. G. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. A. Tucker, V. Vanhoucke, V. Vasudevan, F. B. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "Tensorflow: large-scale machine learning on heterogeneous distributed systems," arXiv:1603.04467 (2016).
40. T. Colomb, J. Kühn, F. Charriere, C. Depeursinge, P. Marquet, and N. Aspert, "Total aberrations compensation in digital holographic microscopy with a reference conjugated hologram," *Opt. Express* **14**, 4300–4306 (2006).