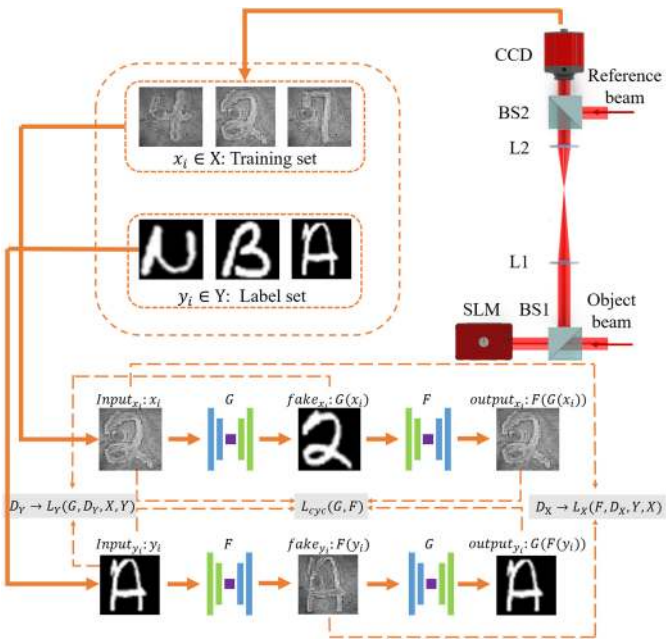
IEEE Photonics Journal
An IEEE Photonics Society Publication

Open Access

Digital Holographic Reconstruction Based on Deep Learning Framework With Unpaired Data

Volume 12, Number 2, April 2020

Da Yin
Zhongzheng Gu
Yanran Zhang
Fengyan Gu
Shouping Nie
Jun Ma
Caojin Yuan



DOI: 10.1109/JPHOT.2019.2961137

Digital Holographic Reconstruction Based on Deep Learning Framework With Unpaired Data

Da Yin,[1] **Zhongzheng Gu**,[1] **Yanran Zhang**,[1] **Fengyan Gu**,[1] **Shouping Nie**,[1] **Jun Ma**,[2] **and Caojin Yuan** [1,3]

[1]Key Laboratory for Opto-Electronic Technology of Jiangsu Province, Nanjing Normal University, Nanjing 210023, China
[2]School of Electronic Engineering and Optoelectronic Techniques, Nanjing University of Science and Technology, Nanjing 210094, China
[3]Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China

DOI:10.1109/JPHOT.2019.2961137

Manuscript received November 12, 2019; revised December 5, 2019; accepted December 17, 2019. Date of publication December 20, 2019; date of current version March 26, 2020. This work was supported in part by the National Science Foundation of China (NSFC) under Grants 61775097 and 61975081; in part by the National key research and development program under Grant 2017YFB0503505; and in part by the Open Foundation of Key Lab of Virtual Geographic Environment (Nanjing Normal University), Ministry of Education under Grant 2017VGE02. Corresponding authors: Jun Ma; Caojin Yuan (e-mail: majun@njust.edu.cn; yuancj@njnu.edu.cn).

Abstract: Convolutional neural network (CNN) has great potentials in holographic reconstruction. Although excellent results can be achieved by using this technique, the number of training and label data must be the same and strict paired relationship is required. Here, we present a new end-to-end learning-based framework to reconstruct noise-free images in absence of any paired training data and prior knowledge of object real distribution. The algorithm uses the cycle consistency loss and generative adversarial network to implement unpaired training method. It is demonstrated by the experiments that high accuracy reconstruction images can be obtained by using unpaired training and label data. Moreover, the unpaired feature of the algorithm makes the system robust to displacement aberration and defocusing effect.

Index Terms: Holography, deep learning, phase recovery, aberration.

1. Introduction

Digital holography (DH) is a non-contact, high-resolution and highly sensitive technique for biomedical imaging [1]–[3], inspection [4], and metrology [5]. DH often uses electronic camera to record the interference pattern and calculate both amplitude and phase information of the reconstructed object wavefront [6], [7]. However, DH suffers from challenges in reconstruction process. The phase distribution is vulnerable to the wrapped phase and system aberration. To solve these problems, a large amount of unwrapping algorithms [8]–[11] and the compensation methods for distorted phase have been studied.

In recent years, deep learning with CNN becomes a powerful tool to solve inverse problem in various optical imaging fields, including scattered image recovery [12], [13], wavefront sensing [14], [15], super-resolution [16], [17], fluorescence microscopy [18], [19], noise reduction [20]–[22] and

phase recovery [23]. Since the reconstruction of DH can be also regarded as an inverse problem, the deep learning has been introduced into DH. Many problems of DH that were difficult to solve in the past now can be solved by CNN, such as classification [24], [25], aberration compensation [26], adaptive spatial filtering [27] and focus prediction [28]. There are many algorithms based on CNN for the reconstruction of DH [29], [30], which use the CNN to eliminate the twin image term and get the accurate focusing distance. In order to achieve one-step reconstruction of the hologram, the end-to-end framework is introduced into the algorithm [31], [32]. After learning the relationship between holograms and corresponding complex distribution, the wavefront information can be directly recovered from a single-shot hologram. In this process, the twin image, the zero-order image and the aberration components are simultaneously eliminated. Most of these algorithms are based on UNet model [33], which is a classical paired training method, and has been widely used in holographic field [34], [35]. However, the hologram and the corresponding object real distribution need to be strictly paired in the training process. The prior knowledge of the object distribution determines the accuracy of reconstruction result. In fact, the object distribution is difficult to obtain in the experiment. At the same time, the displacement caused by the instability system during the recording process is likely to reduce the learning accuracy of pairing requirement algorithm. The accuracy of reconstruction ought to be high theoretically but it will not, since the above reasons.

In this paper, we propose a new end-to-end deep learning framework to solve the problems in one-step approach, which employs unpaired and less images for training. The special feature of our proposed network is that the real distribution of the object is not required. The holograms used in training process and even their amounts can be irrelevant to their labels, which is called unpaired. The strict relationship between hologram and label set in traditional end-to-end framework is replaced by three loss functions in the proposed framework. It is meaningful for reconstructing dynamic phase-contrast objects, like biological cells, since the paired object phase distribution and its hologram need to be strictly aligned. With our algorithm, we can learn the reconstruction mechanism through similar static cells and unpaired holograms. The experimental results prove that the proposed network can reconstruct the hologram well even when the number of label set can be a half of the training image. This greatly reduces the difficulty of getting the data set. When displacement occurs, the proposed algorithm is of higher accuracy than traditional algorithm. Moreover, it has strong robustness to the aberration of the imaging system and the defocus astigmatism.

2. Principle of the Technique

2.1 Architecture of CNN

Our deep convolutional neural network is built on a cycle-generative adversarial network (cycle-GAN) architectural design [36], which has been used in medical image segmentation [37] and coherent noise reduction [22]. Our CNN is a cycle network, which contains two mirror symmetric generative adversarial networks (GAN). We define X as a training set consisting of holograms and Y as label set consisting of object real distributions. However, the data in Y does not match any holograms in X, where the relationship is shown in Fig. 1(a). We can use a number of letters as the label set and the holograms of Arabic numerals as the training set. Therefore, the data in X and Y do not overlap at all.

Partial recording system is sketched in Fig. 1(b), which is a modified Mach-Zehnder interferometer. A He-Ne laser (with the wavelength $=$ 632.8 nm) is used as a light source, which is divided into reference and object beams. The real object information is loaded by the phase-only spatial light modulator (SLM). A 4f system consisting of lens L1 (f $=$ 250 mm) and lens L2 (f $=$ 75 mm) is used to zoom the object. The charge-coupled device (CCD) is fixed on a translation stage to control the diffraction distance. The hologram that is formed by the interference of reference beam and object beam is captured by the CCD. The recorded holograms are used as the training set, but its corresponding label doesn't require to be known with our method.

As shown in Fig. 1(c), the holograms (the training set) and the uncorrelated label data simultaneously input into two respective networks. Two generator functions, $G: X \rightarrow Y$ and $F: Y \rightarrow X$ in

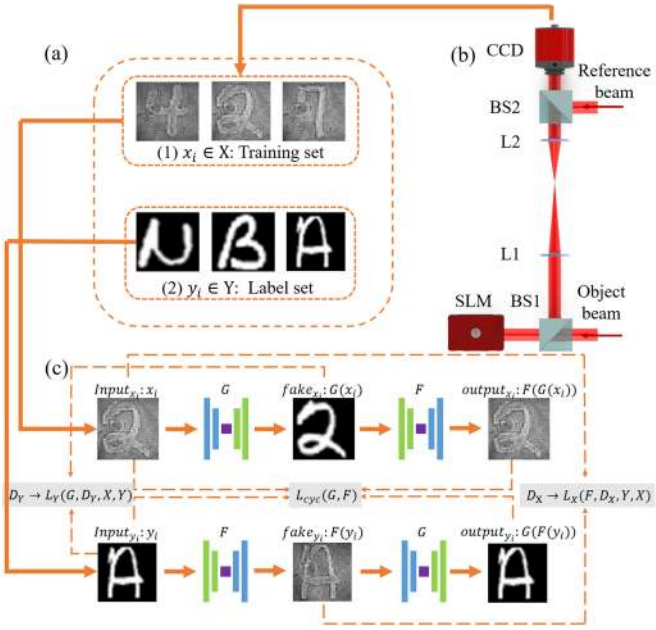


Fig. 1. Simplified view of network architecture. (a) Training and label set prepared for the network. $X_i \in X$: holograms and $Y_i \in Y$: labels. (b) Partial recording system. (c) Overview of training process. Generator function: G and F. D_Y: images generated by G indistinguishable from real images. D_X: images generated by F indistinguishable from Holograms. $L_{cyc}(D, F)$: cycle consistency loss. $L_Y(G, D_Y, X, Y)$: Adversarial loss between D_Y and G. $L_X(F, D_X, Y, X)$: Adversarial loss between D_X and F.

each network and two associated discriminators (D_Y and D_X) are used to generate the adversarial data. The role of D_Y is to measure the difference between the real phase distribution and the fake image generated by G, and vice versa for D_X. This constitutes a complete cycle-GAN framework.

Each generator consists of 9 residual blocks [38] to learn the relationship between domain X and domain Y. We use an encoder-decoder network to design the generator architecture including down-sampling encoder and corresponding up-sampling decoder, which is shown in Fig. 2(a). In order to increase the universality of the network, we designed the input as three channels, which can input color images, and for grayscale images, we convert the input training data into a three-channel tensor to facilitate the design of the network. The input hologram is first sent to a 7×7 convolutional block with 64 filters and stride 1, which yields 64 feature maps. In order to accelerate model convergence and increase training speed, each feature map has been processed by an instance normalization block [39] and rectified linear unit (ReLU) block [40]. The next two convolutional blocks are designed by 3×3 convolution kernels and stride 2. Hence, the numbers of features increase from 64 to 256. A set of 9 identical residual blocks are sent after the end of down-sampling layer. The residual blocks consist of two 3×3 convolutional layers with the same number of filters on both layer and a ReLU block is set after the shortcut connection with two convolution layers. The remaining part is an up-sampling decoder, which aims to sample the obtained features to generate a single image. There are two corresponding 3×3 deconvolutional layers after the residual blocks with 128 and 64 filters, respectively. In order to ensure the output image with the same dimension as the input image, the last 7×7 deconvolutional layers with 3 filters and stride 1. This architecture can generate either fake hologram or fake label, which depends on the composition of input and output data and its respective training.

We design the discriminator as the architecture of PatchGAN [41], which is shown in Fig. 2(b). The 4×4 convolutional block with 64 filters and stride 2 is followed by non-linear leaky rectifier unit (LeakyReLU) [42]. The next part is the combination of a convolutional block, an instancenorm block and a LeakyReLU block which is repeated 3 times. The PatchGAN splits the input into

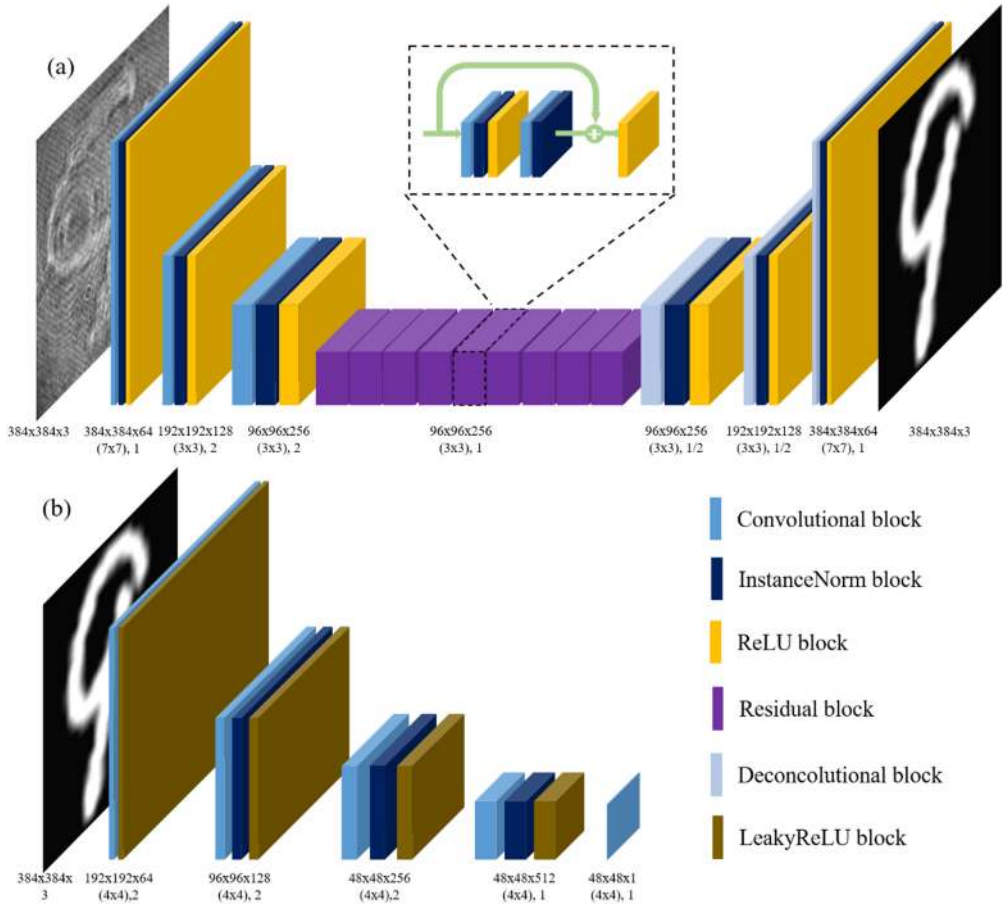


Fig. 2. The architecture of CNN. (a) Architecture of generator. The part of dotted line is represented as the architecture of residual block, which consists of the shortcut connection with two same convolution layers. (b) Architecture of discriminator.

overlapping patches, which run a regular discriminator over each patch, and average the result. By designing PatchGAN to discriminate the local area of the image, the ability to model high-frequency components is improved, so that a more detailed image than the original GAN discriminator can be generated.

2.2 Loss Function

The purpose of discriminator network D_Y is to make the outputs generated by G difficult to be distinguished. For the generator G and discriminator D_Y, we define the loss function as follows [43]:

$$\min_{G}\max_{D_Y} L_Y(G, D_Y, X, Y) = E_{y\to P_{data(y)}}[(D_Y(y)-1)^2] + E_{x\to P_{data(x)}}[(1-D_Y(G(x)))^2] \quad (1)$$

where $E[\cdot]$ represents the expectation of random variable. Generator G is used to generate an output that cannot be identified, while the D_Y is used to estimate the probability that a sample comes from the real data. Therefore, the two form an adversarial relationship. The model G and D_Y are simultaneously trained: fixed discriminator D_Y, adjusting the parameters of generator G to minimize the expectation of $(1-D_Y(G(x)))^2$; fixed generator G, adjusting the parameters of discriminator D_Y to maximize the expectation of $(D_Y(y)-1)^2+(1-D_Y(G(x)))^2$. This optimization process can be summarized as a "minimax two-player game" problem. Similarly, for generator F and D_Y, the loss

function can be expressed as follows:

$$\min_{F}\max_{D_X} L_X(F, D_X, Y, X) = E_{x\to P_{data(x)}}[(D_X(x)-1)^2] + E_{y\to P_{data(y)}}[(1-D_X(F(y)))^2] \quad (2)$$

Next, the cycle consistency loss is introduced to further optimize the model. As shown in Fig. 1, for each hologram x_i from domain X, the output $F(G(x_i))$ by generator G and F should be the same as itself. Similarly, for each label y_i from domain Y, the output $G(F(y_i))$ by generators F and G should be same. They are forward cycle consistency and backward cycle consistency, respectively. Hence, we express the cycle consistency loss as:

$$L_{cyc}(G, F) = E_{x\to P_{data(x)}}[||F(G(x)) - x||_1] + E_{y\to P_{data(y)}}[||G(F(y)) - y||_1] \quad (3)$$

where the symbol $//\,//_1$ represents the L1 norm. Finally, the full loss function including three terms can be expressed as

$$L(G, F, D_Y, D_X) = L_Y(G, D_Y, X, Y) + L_X(F, D_X, Y, X) + \lambda L_{cyc}(G, F) \quad (4)$$

where λ is the cycle consistency loss coefficient, which controls the relative importance between $L_{cyc}(G,\ F)$ and $L_Y(G,\ D_Y,\ X,\ Y) + L_X(F,\ D_x,\ Y,\ X)$. If the value of λ is large, the generated image can better preserve the contour of the training data. If the value of λ is small, the generated image can better approximate the label set distribution. In other words, a lower λ should be used for data with simple textures, and a higher λ should be used for data with complex textures. According to this feature, we usually set λ to 10 by default [44]. We solve the following "minimax two-player game" problem:

$$G^*, F^* = \min_{G,F}\max_{D_Y,D_X} L(G, F, D_Y, D_X) \quad (5)$$

The weights of our network are optimized by Adaptive Moment Estimation (Adam) based optimization and the learning rate is set as 0.0002. Our proposed network is implemented using TensorFlow and all data processing is done in ubuntu 18.04.1 environment with Tesla T4 (14115 MB memory).

3. Results and Discussion

3.1 Data Preparation

For experimental data, we employ processed images loaded onto the phase-only SLM (1080 $\times$ 1920 pixel, pixel size: 6.4 $\times$ 6.4 um) interfered with the reference beam and the holograms are captured by CCD (960 $\times$ 1280 pixel, pixel size: 3.75 $\times$ 3.75 um). We set the phase-only SLM to change the loaded image every second and configure the CCD to synchronize with it. Then we cut the central 384 $\times$ 384 pixel of the captured holograms, forming the holographic images input to our proposed network. We select 100 images as the test sets.

We use the EMNIST dataset [45] and Faces-LFW [46] as the experimental data for our CNN, which are character dataset and face dataset, respectively. EMNIST dataset is a set of handwritten character digits derived from the NIST Special Database 19, which contains handwritten character of numbers, uppercase and lowercase letters. In order to test more complicated situations, we employ the open source facial images from the Faces-LFW dataset. The details of the face images are more abundant, and the grayscale changes are rich.

3.2 Experimental Results Analysis

Fig. 3(a) and Fig. 3(b) represent the results from EMNIST and Faces-LFW, respectively. The holograms from EMNIST and Faces-LFW are captured by CCD at 1mm from the focal plane of 4f system, which are shown in Fig. 3(I). The reconstructed phase images as shown in Fig. 3(II) are obtained through traditional aberration compensation algorithm based on PCA [47]. The phase distribution free of aberration is retrieved from the largest principal component in decomposed orthogonal components. It can be obviously seen from the Fig. 3(II) that the recovered phase

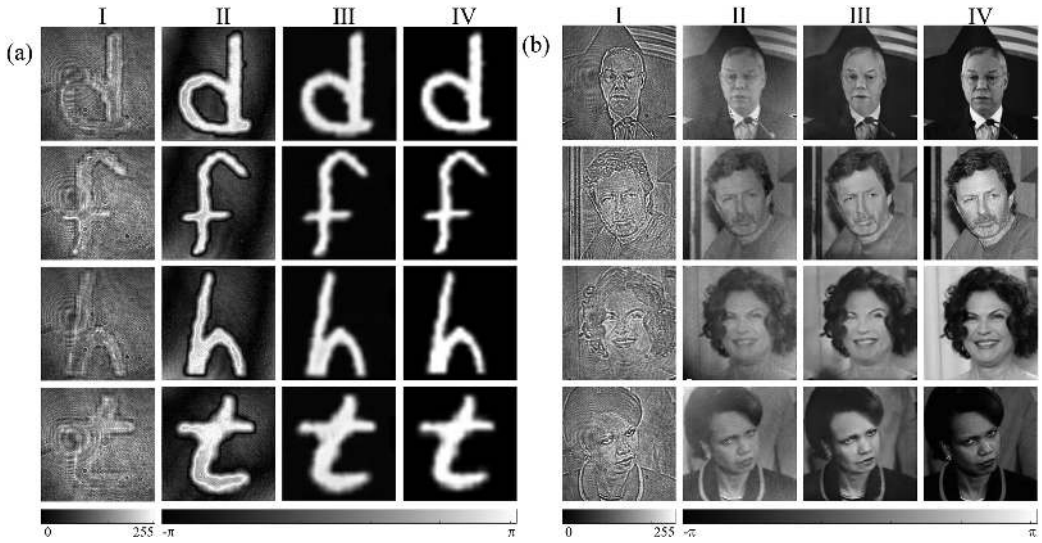


Fig. 3. Reconstruction results using proposed convolutional neural network. (I) Untrained test holograms. (II), (III) Reconstruction images through traditional method and our proposed algorithm. (IV) Ground truths of the proposed test set. (a)-(b): The test images are drawn from: (a) EMNIST [45] and (b) Faces-LFW [46], respectively.

images are still disturbed by aberration and background noise. We can get the images as shown in Fig. 3(III) through the network we have proposed and the images as shown in Fig. 3(IV) are the ground truths in Fig. 3(I). For EMNIST dataset, we select 2400 images of handwritten number as training set, 2000 images of handwritten uppercase letters as label set and 100 images of handwritten lowercase letters as test set. Therefore, there is no overlap between the training set, the label set and the test set, which reflects the unpaired features of the proposed network. We perform a similar operation on the Faces-LFW dataset.

In order to verify the accuracy of reconstruction results, we use the complex wavelet structural similarity (CW-SSIM) index [48] to quantify the degree of similarity between the reconstruction results and the ground truths. CW-SSIM compares images in the complex wavelet transform domain, regardless of the spatial position of the image. CW-SSIM as well as Peak Signal to Noise Ratio (PSNR) and structural similarity index (SSIM) are always adopted as the accuracy metric. Since the positions and sizes of the reconstructed images are hard to match the label images in the spatial domain, CW-SSIM is more suitable to calibrate the accuracy. It is found that CW-SSIM is not only suitable with the human's judgment on image quality than PSNR, but also CW-SSIM can still run well than SSIM in the case of image displacement, scaling and other non-structural distortion. In the complex wavelet transform domain, we define that $C_x = \{C_{x,i} \mid i = 1, \ldots, N\}$ and $C_y = \{C_{y,i} \mid i = 1, \ldots, N\}$ as two sets of coefficients extracted at the same spatial location in the same wavelet subbands between the reconstruction results and the ground truths. Therefore, the CW-SSIM can be expressed as:

$$CW-SSIM = \frac{2|\sum_{i=1}^{N} C_{x,i} C_{y,i}^{*}| + K}{\sum_{i=1}^{N} |C_{x,i}|^2 \sum_{i=1}^{N} |C_{y,i}|^2 + K} \tag{6}$$

where the symbol $*$ indicates the complex conjugate. K is a small positive constant, which is used to improve the robustness of CW-SSIM in the case of low local signal-to-noise ratio of images.

The value of CW-SSIM lies between 0 and 1, the larger the value of CW-SSIM the smaller the difference between the reconstruction results and the ground truths, that is, the better the image quality. When the two images are identical, CW-SSIM is 1. We calculate the CW-SSIM averaging over the all test images as the metric values. The two CW-SSIM of reconstruction results from

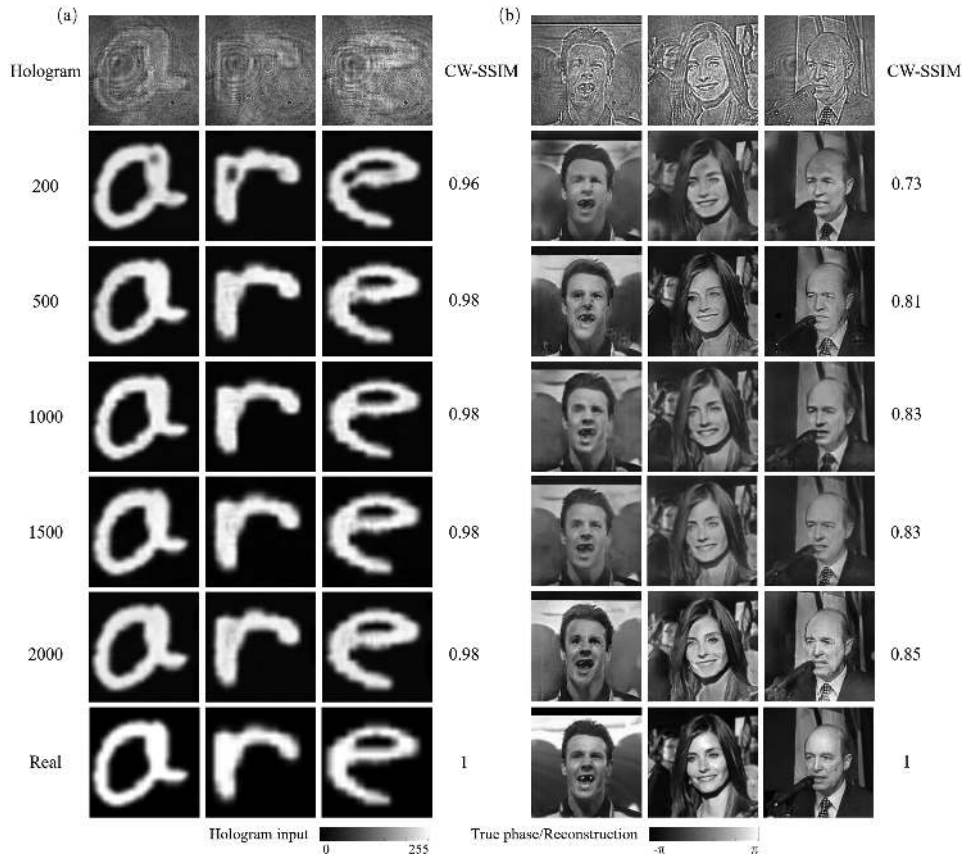


Fig. 4. (a), (b) The reconstructed results with different number of labels from EMNIST [45] and Faces-LFW [46].

EMNIST and LFW are 0.98 and 0.85. However, the CW-SSIM for the reconstruction results of EMNIST and Faces-LFW obtained by the conventional DH algorithm with PCA are 0.71 and 0.78, respectively, which are lower than the CW-SSIM index obtained by using CNN. The reason is that image factors such as aberration and background noise are difficult to be completely eliminated by the conventional DH reconstruction algorithm. However, our proposed algorithm can eliminate these adverse effects, thus the imaging quality is greatly improved. In the experiment, we try to use the binary EMNIST dataset for training and the gray Faces-LFW dataset for test. It can be found that the test reconstructed image is casted to the binary one, although the outlines of object can be roughly reconstructed. Conversely, the gray training data will bring background noise to the binary test data. Therefore, the similar kind of training data and test data are preferred to guarantee the high accuracy.

3.3 Different Number of Labels During Training

In the past experiments using CNN to recover holograms, the same number of one to one paired label set as the training set are needed. Since the real images are difficult to obtain, our goal is to use our CNN to obtain excellent holographic recovery images with fewer labels. This will greatly reduce the difficulty of obtaining the real distribution of objects. The 2400 holograms of handwritten numbers are captured by CCD which are used as the training set and smaller number of labels with handwritten uppercase letters are used as the labels set. We train the network with 200, 500, 1000, 1500 and 2000 labels respectively, which is shown in Fig. 4. The results in Fig. 4(a) and (b) are from EMNIST and Faces-LFW datasets, respectively.

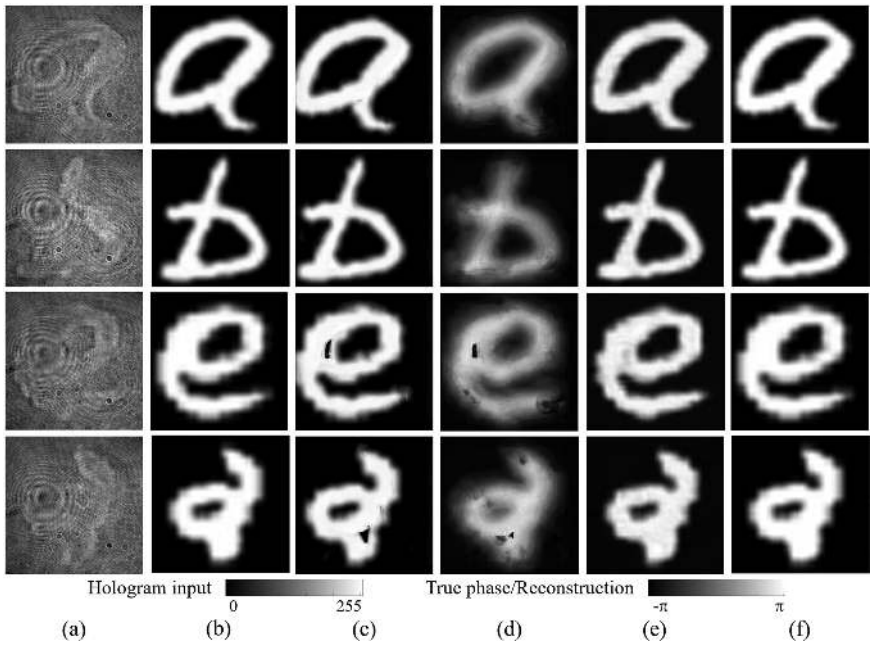


Fig. 5. (a) Raw holograms. (b), (c), (d) are the reconstructed results without any shifting and rotation of the label, with shifting 20 pixels of the label, with random shifting 0–25 pixels and random rotation 0–25 degrees, respectively. (e) The experimental results from our proposed network. (f) The ground truths of (a).

The above experimental results show that the holograms can be reconstructed with fewer labels by using our proposed neural network. We use 100 test images for reconstruction and calculate their CW-SSIM index. The averaged CW-SSIM index of the reconstruction results under different number of labels is shown on the right side of Fig. 4(a) and (b). As can be seen from the Fig. 4, as the number of labels is decreasing, the average CW-SSIM index of the reconstruction results does not decrease much. Therefore, with different numbers of labels, both simple datasets and complex datasets can use less real distribution of objects to train the network, which greatly reduces the difficulty of data acquisition. However, when the number of trained labels is less than 200, the CW-SSIM index of reconstruction is greatly reduced. This is because when the number of labels is reduced to a certain extent, the neural network learns too few features, so that holograms can't be reconstructed very well.

3.4 Comparison Between Two Training Methods

In order to compare with the paired training method, we reconstructed the holograms using the classic UNet model and proposed hologram, respectively. The test holograms and true phase distribution are shown in Fig. 5(a) and (f), respectively. We simulated three cases: strict pair between the label and the hologram, the label and the hologram with 20 pixels dislocation, the label and the hologram with random shifting within [0, 25] pixels and the random rotation within [0, 25] degrees, which are respectively shown in Fig. 5(b), (c), and (d). It can be seen from the Fig. 5 that as the displacement becomes more and more complicated, details of the reconstructed image become blurred. The black spots represent the lost information. We calculate the CW-SSIM of Fig. 5(b), (c), (d) respectively, which are 0.99, 0.95 and 0.65, respectively. However, the reconstruction results obtained using our method as shown in Fig. 5(e) still maintain good reconstruction qualities, and their CW-SSIM values remain at 0.98. This shows that the network proposed by us is robust to displacement and rotation problems.

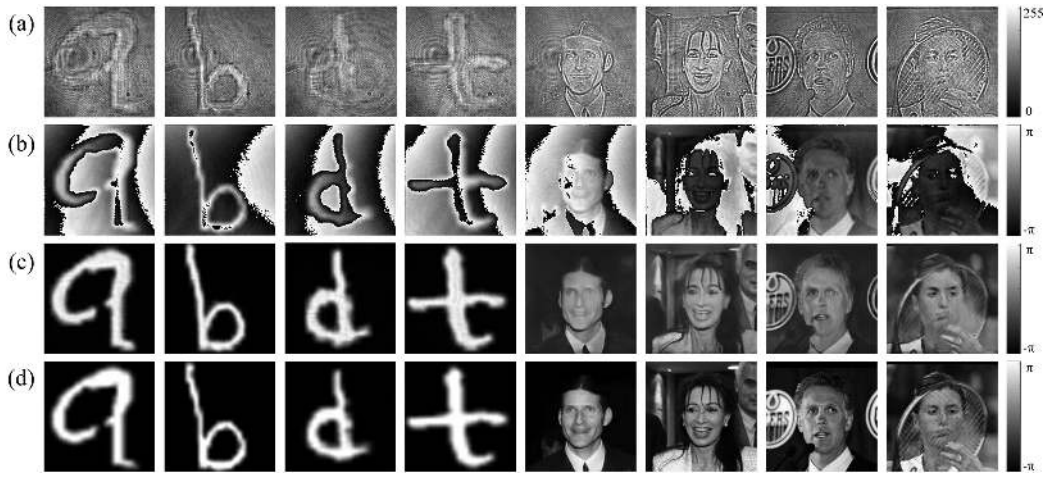


Fig. 6. (a) The raw holograms of reconstructed images. (b) Reconstructed holograms with aberrations randomly selected. (c) Reconstructed results from (a) with our proposed method. (d) The ground-truth image of (a). The ground truths come from EMNIST [45] and Faces-LFW [46].

3.5 Robustness of Aberration

In traditional DH reconstruction methods, it is often necessary to compensate for aberrations using physical [49] or numerical [50] methods. Physical methods require accurate alignment of optical elements, which is impractical. In numerical methods, it is necessary to manually select the region without object or to require the object to be very thin, and it is often impossible to fully compensate for the aberration. It is known that our method takes a lot of time to collect holograms. In this process, the optical system inevitably changes more or less. It can be said that the training data are actually obtained under different circumstances. Therefore, the aberration in each hologram is different from each other. Fig. 6(b) shows several reconstructed holograms with aberration randomly selected. To show the phase aberration compensation ability of our algorithm, PCA algorithms are not used to deal with the aberrations in Fig. 6(b). They show several reconstructed holograms with aberration randomly selected from data. It is apparent from the Fig. 6(b) that the aberration of the imaging system is variable and is difficult to completely eliminate by physical or numerical methods. In our method, we can completely compensate the aberration, which is shown in Fig. 6(c). Fig. 6(a) shows the raw holograms and the Fig. 6(d) is the ground truths of Fig. 6(a). The experimental results show that our method is not too strict for the stability of optical system, which can completely eliminate the aberration of the optical system under different conditions.

3.6 Different Axial Locations

To further study the ability of the aberration compensation, we analyze the ability to deal with defocus distortion. The distance between the test object and CCD should be exactly the same with that used in the training process. But in practice, it is difficult to satisfy the requirement. This process can be seen as defocusing. The defocused image reconstructed from the hologram is an additional quadratic phase attached to the optical system. Then the phase deviation of the defocus hologram can be written as:

$$\begin{aligned}\exp[jkW(x,y)] &= \exp\left\{-j\frac{k}{2z}(x^2+y^2)-\left[-j\frac{k}{2(z+\Delta z)}(x^2+y^2)\right]\right\}\\ &= \exp\left[j\frac{k\varepsilon}{2}(x^2+y^2)\right] \end{aligned} \tag{7}$$

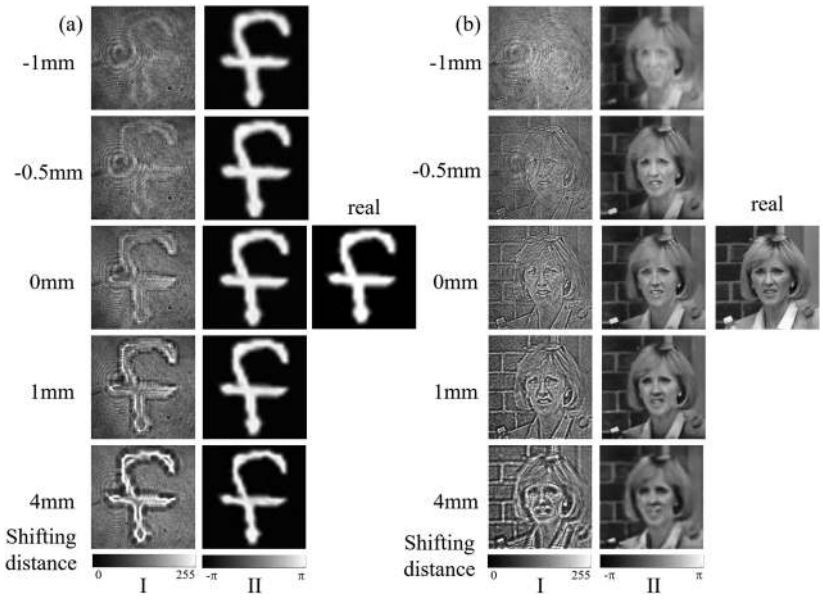


Fig. 7. The effects of recovered images in different axial locations. (I) Holograms. (II) Results. (a), (b) Data from EMNIST [45] and Faces-LFW [46].

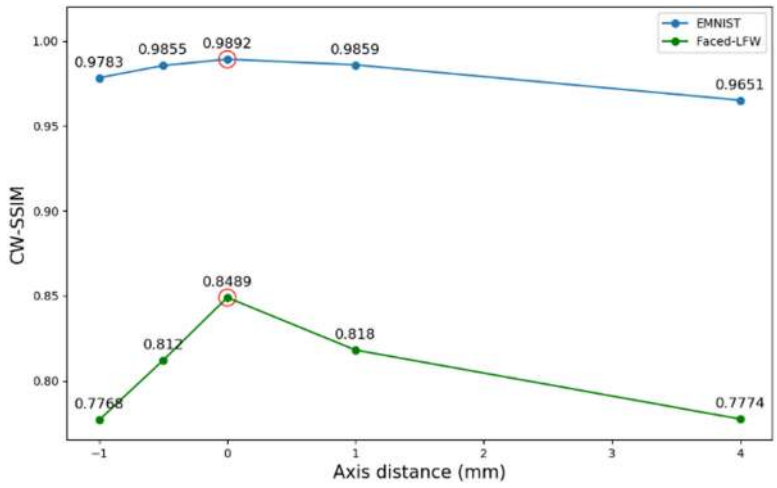


Fig. 8. CW-SSIM curve with different axis locations. The blue line is the CW-SSIM curve of EMNIST. The green line is the CW-SSIM curve of Faces-LFW. All the CW-SSIM has been marked on it.

where Δz represents the distance from original position, ε represents the degree of defocus, and $\varepsilon = \frac{1}{z+\Delta z} - \frac{1}{z} \approx \frac{\Delta z}{z^2}$. To study it, we shift the CCD away from the original position to capture test holograms with different axial locations. The captured holograms for test are shown in the (I) column of Fig. 7(a) and Fig. 7(b), where the focus to defocus process can be observed. The corresponding shifting distance is denoted on the left side of column (I). The reconstructed results are shown in Fig. 7(II). We place the ground truth images in the right of the column (II). As shown in Fig. 7, our method learns the features of the hologram, which can eliminate the defocusing aberration, so it can restore the hologram at different distances to a certain extent. However, with the increase of defocusing distance, the quality of reconstructed results decreases.

In order to compare the reconstruction results of holograms at different axial distances, the CW-SSIM curves of two datasets are plotted in Fig. 8. The data circled by the red circle in the Fig. 8 is the CW-SSIM index obtained at the position of the trained holograms. This experiment demonstrates that although the network is only trained in one particular location, the holograms recorded in a nearby location can still be recovered well through the CNN. However, when the recording position of the hologram is too far from the position of the hologram during training, the CW-SSIM index will drop significantly. This phenomenon indicates that CNN has learned the features of holograms during the experiment, so the proposed method is robust to the slight movement of holograms.

4. Conclusion

In a summary, we have proposed an end-to-end deep learning framework for holographic reconstruction. Our proposed algorithm has strong robustness with the aberration of imaging system and the defocus distortion, and is more capable of compensating for aberration than traditional methods. The wavefront information of the object can be directly extracted from a single hologram in the absence of paired training examples. It reduces the difficulty of data acquisition. Moreover, our method shows great advantages, when real label data are lack in practice. For example, in experiments such as living cell, the label data which are regarded as phase-contrast images and their corresponding holograms cannot be obtained simultaneously. Traditional methods based on CNN are not up to the tasks mentioned above. However, using our approach, we can train the known and similar label set instead the real one to make CNN work.

References

[1] F. Charriere *et al.*, "Cell refractive index tomography by digital holographic microscopy," *Opt. Lett.*, vol. 31, no. 2, pp. 178–180, 2006.
[2] B. Kemper and G. von Bally, "Digital holographic microscopy for live cell applications and technical inspection," *Appl. Opt.*, vol. 47, no. 4, pp. A52–A61, 2007.
[3] B. Kemper *et al.*, "Investigation of living pancreas tumor cells by digital holographic microscopy," *J. Biomed. Opt.*, vol. 11, no. 3, 2006, Art. no. 34005.
[4] B. Kemper, S. Stürwald, C. Remmersmann, P. Langehanenberg, and G. von Bally, "Characterisation of light emitting diodes (LEDs) for application in digital holographic microscopy for inspection of micro and nanostructured surfaces," *Opt. Laser Eng.*, vol. 46, no. 7, pp. 499–507, 2008.
[5] J. Dong, C. Jiang, and S. Jia, "Digital holographic metrology based on multi-angle interferometry," *Opt. Lett.*, vol. 41, no. 18, pp. 4301–4304, 2016.
[6] U. Schnars, J. ptner, and P. O. Werner, "Digital recording and numerical reconstruction of holograms," *Meas. Sci. Technol.*, vol. 13, no. 9, pp. R85–R101, 2002.
[7] L. Xu, X. Peng, Z. Guo, J. Miao, and A. Asundi, "Imaging analysis of digital holography," *Opt. Exp.*, vol. 13, no. 7, pp. 2444–2452, 2005.
[8] S. Lai, B. King, and M. A. Neifeld, "Wave front reconstruction by means of phase-shifting digital in-line holography," *Opt. Commun.*, vol. 173, no. 1-6, pp. 155–160, 2000.
[9] B. Gutmann and H. Weber, "Phase unwrapping with the branch-cut method: Role of phase-field direction," *Appl. Opt.*, vol. 39, no. 26, pp. 4802–4816, 2000.
[10] J. R. Buckland, J. M. Huntley, and S. R. Turner, "Unwrapping noisy phase maps by use of a minimum-cost-matching algorithm," *Appl. Opt.*, vol. 34, no. 23, pp. 5100–5108, 1995.
[11] D. C. Ghiglia and L. A. Romero, "Robust two-dimensional weighted and unweighted phase unwrapping that uses fast transforms and iterative methods," *J. Opt. Soc. Amer. A*, vol. 11, no. 1, pp. 107–117, 1994.
[12] Y. Li, Y. Xue, and L. Tian, "Deep speckle correlation: A deep learning approach toward scalable imaging through scattering media," *Optica*, vol. 5, no. 10, pp. 1181–1190, 2018.
[13] M. Yang *et al.*, "Deep hybrid scattering image learning," *J. Phys. D, Appl. Phys.*, vol. 52, no. 11, 2019, Art. no. 115105.
[14] Y. Nishizaki *et al.*, "Deep learning wavefront sensing," *Opt. Exp.*, vol. 27, no. 1, pp. 240–251, 2019.
[15] S. W. Paine and J. R. Fienup, "Machine learning for improved image-based wavefront sensing," *Opt. Lett.*, vol. 43, no. 6, pp. 1235–1238, 2018.
[16] T. Nguyen, Y. Xue, Y. Li, L. Tian, and G. Nehmetallah, "Deep learning approach for Fourier ptychography microscopy," *Opt. Exp.*, vol. 26, no. 20, pp. 26470–26484, 2018.
[17] Y. Rivenson *et al.*, "Deep learning microscopy," *Optica*, vol. 4, no. 11, pp. 1437–1443, 2017.
[18] W. Ouyang, A. Aristov, M. Lelek, X. Hao, and C. Zimmer, "Deep learning massively accelerates super-resolution localization microscopy," *Nature Biotechnol.*, vol. 36, no. 5, pp. 460–468, 2018.
[19] H. Wang *et al.*, "Deep learning enables cross-modality super-resolution in fluorescence microscopy," *Nature Methods*, vol. 16, no. 1, pp. 103–110, 2019.

[20] W. Jeon, W. Jeong, K. Son, and H. Yang, "Speckle noise reduction for digital holographic images using multi-scale convolutional neural networks," *Opt. Lett.*, vol. 43, no. 17, pp. 4240–4243, 2018.
[21] K. Yan *et al.*, "Fringe pattern denoising based on deep learning," *Opt. Commun.*, vol. 437, pp. 148–152, 2019.
[22] G. Choi *et al.*, "Cycle-consistent deep learning approach to coherent noise reduction in optical diffraction tomography," *Opt. Exp.*, vol. 27, no. 4, pp. 4927–4943, 2019.
[23] A. Sinha, J. Lee, S. Li, and G. Barbastathis, "Lensless computational imaging through deep learning," *Optica*, vol. 4, no. 9, pp. 1117–1125, 2017.
[24] C.-J. Cheng, K.-C. C. Chien, and Y.-C. Lin, "Digital hologram for data augmentation in learning-based pattern classification," *Opt. Lett.*, vol. 43, no. 21, pp. 5419–5422, 2018.
[25] Y. Jo *et al.*, "Holographic deep learning for rapid optical screening of anthrax spores," *Sci. Adv.*, vol. 3, no. 8, 2017, Art. no. e1700606.
[26] T. Nguyen *et al.*, "Automatic phase aberration compensation for digital holographic microscopy based on deep learning background detection," *Opt. Exp.*, vol. 25, no. 13, pp. 15043–15057, 2017.
[27] W. Xiao *et al.*, "Adaptive frequency filtering based on convolutional neural networks in off-axis digital holographic microscopy," *Biomed. Opt. Exp.*, vol. 10, no. 4, pp. 1613–1626, 2019.
[28] T. Pitkaaho, A. Manninen, and T. J. Naughton, "Focus prediction in digital holographic microscopy using deep convolutional neural networks," *Appl. Opt.*, vol. 58, no. 5, pp. A202–A208, 2019.
[29] Y. Wu *et al.*, "Extended depth-of-field in holographic imaging using deep-learning-based autofocusing and phase recovery," *Optica*, vol. 5, no. 6, pp. 704–710, 2018.
[30] 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.*, vol. 7, 2018, Art. no. 17141.
[31] Z. Ren, Z. Xu, and E. Y. Lam, "End-to-end deep learning framework for digital holographic reconstruction," *Adv. Photon.*, vol. 1, no. 1, 2019, Art. no. 016004.
[32] H. Wang, M. Lyu, and G. Situ, "eHoloNet: A learning-based end-to-end approach for in-line digital holographic reconstruction," *Opt. Exp.*, vol. 26, no. 18, pp. 22603–22614, 2018.
[33] O. Ronneberger, P. Fischer, and T. Brox., "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Medical Image Comput. Comput.-Assisted Intervention.*, 2015, pp. 234–241.
[34] G. Zhang *et al.*, "Fast phase retrieval in off-axis digital holographic microscopy through deep learning," *Opt. Exp.*, vol. 26, no. 15, pp. 19388–19405, 2018.
[35] K. Wang, J. Dou, Q. Kemao, J. Di, and J. Zhao, "Y-Net: A one-to-two deep learning framework for digital holographic reconstruction," *Opt. Lett.*, vol. 44, no. 19, pp. 4765–4768, 2019.
[36] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2223–2232.
[37] J. M. Wolterink *et al.*, "Deep MR to CT synthesis using unpaired data," in *proc. Int. Workshop Simul. Synth. Med. Imag.*, 2017, pp. 14–23.
[38] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, Cham, Switzerland, 2016, pp. 694–711.
[39] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," 2016, *arXiv:1607.08022*.
[40] X. Glorot *et al.*, "Deep sparse rectifier neural networks," in *Proc. 14th Int. Conf. Artif. Intell. Statist.*, 2011, pp. 315–323.
[41] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1125–1134.
[42] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," *in Proc. Int. Conf. Mach. Learn.*, 2013, vol 3, no. 30, pp. 1–6.
[43] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.
[44] Z. Yi, H. Zhang, P. Tan, and M. Gong, "DualGAN: Unsupervised dual learning for image-to-image translation," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2849–2857.
[45] G. Cohen, S. Afshar, J. Tapson, and A. V. Schaik, "EMNIST: An extension of MNIST to handwritten letters," 2017, *arXiv:1702.05373v2*.
[46] G. B. Huang *et al.*, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Univ. Massachusetts, Amherst, MA, USA, Tech. Rep. 07, 2007.
[47] C. Zuo, Q. Chen, W. Qu, and A. Asundi, "Phase aberration compensation in digital holographic microscopy based on principal component analysis," *Opt. Lett.*, vol. 38, no. 10, pp. 1724–1726, 2013.
[48] M. P. Sampat, Z. Wang, S. Gupta, A. C. Bovik, and M. K. Markey, "Complex wavelet structural similarity: A new image similarity index," *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2385–2401, Nov. 2009.
[49] Q. Weijuan, Y. Yingjie, C. O. Choo, and A. Asundi, "Digital holographic microscopy with physical phase compensation," *Opt. Lett.*, vol. 34, no. 8, pp. 1276–1278, 2009.
[50] L. Miccio *et al.*, "Direct full compensation of the aberrations in quantitative phase microscopy of thin objects by a single digital hologram," *Appl. Phys. Lett.*, vol. 90, no. 4, 2007, Art. no. 41104.