CHINESE OPTICS LETTERS

Vol. 21, No. 8 | August 2023

Suppressing defocus noise with U-net in optical scanning holography

Haiyan Ou (欧海燕)[1,2*], Yong Wu (吴 勇)[1], Kun Zhu (朱 坤)[3], Edmund Y. Lam (林彦民)[4], and Bing-Zhong Wang (王秉中)[1]

[1] School of Physics, University of Electronic Science and Technology of China, Chengdu 610054, China

[2] Shenzhen Institute for Advanced Study, University of Electronic Science and Technology of China, Shenzhen 518000, China

[3] Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hong Kong, China

[4] Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong, China

*Corresponding author: ouhaiyan@uestc.edu.cn

Received February 21, 2023 | Accepted April 23, 2023 | Posted Online August 7, 2023

Optical scanning holography (OSH) records both the amplitude and phase information of a 3D object by a 2D scan. To reconstruct a 3D volumetric image from an OSH hologram is difficult, as it suffers from the defocus noise from the other sections. The use of a random phase pupil can convert defocus noise into speckle-like noise, which may require further processing in sectional image reconstruction. In this paper, we propose a U-shaped neural network to reduce this speckle haze. Simulation results show that the proposed method works effectively and efficiently both in simple and complex graphics.

Keywords: digital holography; image reconstruction; defocus noise; neural network.

DOI: 10.3788/COL202321.080501

1. Introduction

As a versatile digital holography technique, optical scanning holography (OSH) has been widely used in microscopy[1], remote sensing[2], image encryption, etc[3,4]. In OSH, the object is scanned by the heterodyne beams, which are launched from the same light source with frequency difference generated by the frequency shifter. Unlike traditional digital holography, OSH can record 3D objects into 2D holograms by single-pixel 2D scanning. The hologram can preserve the amplitude as well as the phase information of the object. One can obtain this object's information using hologram reconstruction.

In the image reconstruction, there are two important issues. The first one is auto-focusing, i.e., finding the accurate reconstruction distance, and the other is choosing an applicable reconstruction method with increasing depth resolution as well as less defocus noise. A great deal of research have proposed to retrieve the reconstruction distance automatically, such as the extended focused imaging[5], structure tensor[6], and connected domain[7]. Researchers also used the time-reversal (TR) technique to get the depth information by calculating the pseudo-spectrum of the TR matrix generated from the hologram[8]. To further improve the depth resolution, methods based on double measurements have also been proposed, including the use of a dual-wavelength laser[9], double-location detection[10], and the reconfigurable pupil[11].

The out-of-focus haze, also known as the defocus noise, is undesired residual signals from other sections. Many methods have been presented to conduct image reconstruction and to suppress the defocus noise, such as inverse imaging[12,13], Wiener filtering[14], and 3D imaging[15]. For example, Zhou *et al.* used a random phase pupil to transfer the defocus noise into speckle-like patterns[16]. This noise can be further suppressed by average[16], connected component[17], and image fusion[18].

In recent years, deep learning has undergone rapid development and has found wide applications in some areas, such as language processing, image processing, biomedical and machine visions, as well as digital holography[19–22]. Ren *et al.* presented a convolution neural network (CNN) based on the regression method to achieve fast auto-focusing[23]. The CNN has been trained by a set of holograms *a priori*. Pitkäaho *et al.* showed that CNNs can also predict the in-focus depth by learning from half a million hologram amplitude images in advance[24]. Compared with traditional methods, their work showed better precision and efficiency. Rivenson *et al.* used deep learning to rapidly perform phase recovery and image reconstruction simultaneously. The calculation was based on only one hologram, and could reconstruct both the phase and amplitude images of the objects[25]. Nguyen *et al.* presented a phase aberration compensation method based on deep learning CNN[26]. It could perform automatic background region detection for most aberrations. Deep learning has also proved an effective tool in molecular diagnostics[27], as Kim *et al.* trained the neural networks to classify the holograms without reconstruction. The captured

holograms of the cells were used as raw holograms to train the neural networks, which were able to classify individual cells afterwards.

In this paper, we present for the first time and to the best of our knowledge, a reconstruction method based on a U-shaped convolutional neural network (U-net) to remove the speckle-like defocus noise in a OSH system. The U-net approach is adopted to learn the mapping between various holograms and the corresponding sectional images. Unlike other CNN methods, which require large training data sets, U-net can work with very few training images and yields more precise results. The proposed method can eliminate the speckle-like noise generated by the random phase pupil. Simulation results show that the algorithm works well with both simple and complex graphics. It also outperforms the traditional reconstruction methods in terms of better sectional image quality and significantly faster processing speed.

This paper is organized as follows. In Section 2, we first introduce the OSH system and the principle of random phase pupil system. The theory of U-net deep learning is also explained in this section. Simulation results are presented and discussed in Section 3 to demonstrate the visibility of the proposed method. The conclusion remarks are given in Section 4.

2. Principle

2.1. Optical scanning holography

The holographic system is illustrated in Fig. 1. A He–Ne laser is set at the starting position of the system to launch a cluster of planar waves. The waves will be split into two beams by beam splitter BS1. An acousto-optic frequency shifter (AOFS) is used to shift one of the frequencies from ω_0 to $\omega_0 + \Omega$. The two beams will pass though different mirrors, pupils, and lenses, respectively, and then converge at BS2. After that, the combined beam will be used to scan the object via the X–Y scanner. In the meantime, lens L3 will collect the light from the object. The generated electrical signal from the photo detector (PD) will be transformed into a hologram via a band-pass filter and further demodulation process.

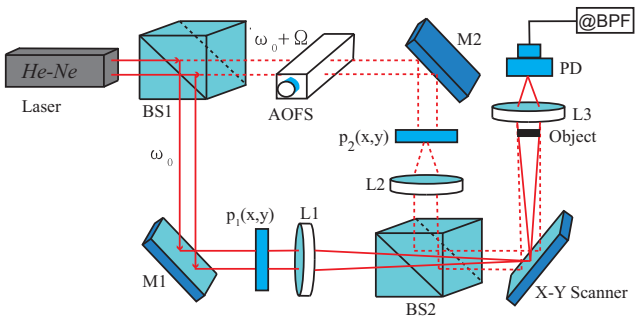

Fig. 1. OSH system setup[28]. BS1 and BS2, beam splitter; M1 and M2, mirror; AOFS, acousto-optic frequency shifter; $p_1(x, y)$ and $p_2(x, y)$, pupils; L1, L2, and L3, lens; PD, photo detector; BPF, band-pass filter.

In a random phase pupil holographic system, the two pupils are set as $p_1(x, y) = \exp[j2\pi r(x, y)]$ and $p_2(x, y) = 1$, where $r(x, y)$ is a random function, and its value is chosen from a uniform distribution between (0, 1). The random phase pupil is widely used in OSH to disperse the defocus noise from other sections into speckle-like patterns[16]. If we discretize the object into N sections along the z-axis, then the hologram can be expressed as

$$g(x,y) = \sum_{i=1}^{N} \mathcal{F}^{-1}\left\{\mathcal{F}[\mathcal{O}(x,y;z_i)] \times \exp\left[-j\frac{z_i}{2k_0}(k_x^2 + k_y^2)\right] \times P_1^*\left(-\frac{z_i k_x}{f}, -\frac{z_i k_y}{f}\right)\right\}, \tag{1}$$

where $\mathcal{O}(x, y; z_i)$ is the complex amplitude function of the object, x and y are space coordinates of the object, and z_i indicates the axial coordinates of the ith section. k_x and k_y are frequency domain coordinates, and $k_0 = \frac{2\pi}{\lambda}$ is the wavenumber, where λ represents the wavelength of the laser. $\mathcal{F}$ and $\mathcal{F}^{-1}$ indicate the Fourier transformation and inverse Fourier transformation, respectively. P_1 is the frequency domain expression of $p_1(x, y)$, * represents the conjugate operation, and f is the focus distance of the lens.

To recover the jth section, we can set the decoding pupils as $p_{1d}(x, y) = 1$ and $p_1^*(-x, -y)p_{2d}(x, y) = 1$. The reconstruction distance is set as z_j. In this case, the recovered image becomes

$$\begin{aligned} I_{\text{out}}(x,y;z_j) &= \mathcal{F}^{-1}\left\{\mathcal{F}\{g(x,y)\} \times \exp\left[j\frac{z_j}{2k_0}(k_x^2 + k_y^2)\right] \times P_{2d}\left(\frac{z_j k_x}{f}, \frac{z_j k_y}{f}\right)\right\} \\ &= \mathcal{O}(x,y;z_j) + N(x,y{:}z_j), \end{aligned} \tag{2}$$

where $I_{\text{out}}(x, y; z_j)$ is the reconstructed sectional image at z_j, which contains useful signal $\mathcal{O}(x, y; z_j)$ and the speckle-like noise signal $N(x, y; z_j)$. The speckle noise can be derived by substituting Eq. (1) into Eq. (2) and can be expressed as[16]

$$N(x,y{:}z_j) = \sum_{i\neq j}^{N} \mathcal{F}^{-1}\left\{\mathcal{F}\left\{\mathcal{O}(x,j;z_i) \times \exp\left[j\frac{z_j - z_i}{2k_0}(k_x^2 + k_y^2)\right]\right\} \times P_1^*\left(-\frac{z_i k_x}{f}, -\frac{z_i k_y}{f}\right) \times P_{2d}\left(\frac{z_j k_x}{f}, \frac{z_j k_y}{f}\right)\right\}. \tag{3}$$

For a better sectioning effect, this noise should be further eliminated. One can suppress the speckle haze by averaging multiple section images or using the connected component methods[16,17]. These methods succeed in suppressing overriding noise. However, they all require multiple frames to solve the problem. This would greatly reduce the efficiency. Here, we present a special CNN-based method, U-net, to suppress the speckle-like noise. In addition to having no requirement of prior information, the U-net method also features a simpler,

shorter operation time and is more robust than the conventional methods mentioned above.

Unlike other CNN methods, which require large training data sets, U-net can work with very few training images and yield more precise results[29,30]. This is the main advantage, especially in the situation where it is difficult to retrieve large data sets, such as biological applications.

2.2. U-shaped convolutional neural network

The neural network is organized by a contracting path as well as an expansive path. As the appearance is very similar to the letter 'U', this neural network has been named U-net. U-net was first presented to segment the biomedical image and has proved to be a very effective end-to-end image processing tool[30,31].

Figure 2 illustrates the architecture of the U-net, which contains two paths: the contracting one (convolution + downsampling) and the expansive one (deconvolution + upsampling). The contracting path has many layers. Each layer consists of two 3×3 convolutions, followed by a rectified linear unit (ReLU). For downsampling, a 2×2 max pooling operation with stride 2 is used. While for the expansive path, the pattern of deconvolution with upsampling is repeated. A 2×2 convolution is used for upsampling, and two 3×3 convolutions followed by a ReLU activation function are included in the deconvolution layers. In the final layer, the required features are retained while those structures related to speckle noise are discarded.

It is worth noting that the number of feature channels are copied and cropped after each downsampling process, as denoted by the white boxes in Fig. 2. These contracted high resolution features are then merged and combined with the upsampled output to generate a more precise output.

For the $\mathcal{L}$-th convolutional layer, we assume there are $N_{\mathcal{L}}$ feature maps with a size $k\times k$, and it can be expressed as $h_j^{\mathcal{L}}$, $j=1,2,\ldots,N_{\mathcal{L}}$. The next convolutional layer $h_j^{\mathcal{L}+1}$ can be expressed as

$$h_j^{\mathcal{L}+1}=\mathrm{ReLU}\left(\sum_{i=1}^{N_{\mathcal{L}}}h_i^{\mathcal{L}}\otimes w_{ij}^{\mathcal{L}+1}+b_j^{\mathcal{L}+1}\right),\tag{4}$$

where $\otimes$ is the convolution operation, and w_{ij} and b_j are the weight and bias that need to be learned via training, respectively. ReLU is an activation function, which can be denoted as $\mathrm{ReLU}(x)=\max(0,x)$. In a U-net model, when the inputs are transmitted between neurons, the weights are applied to the inputs and passed into an activation function ReLU along with the bias. The weights are essentially reflecting how important an input is, while the value of bias controls when the activation function is activated. In the process of downsampling, we need to pad the feature maps with zero first. This process can make the next convolutional layer feature maps have the same size after convolution. In the next step, we utilize the max pooling to choose the maximum value as the representation among a small region. In this way, the size of the feature maps will be smaller. The process of upsampling is similar with the downsampling, except that the upsampling needs to extend the size of the feature maps with zero. It is worth noting that lots of useful information will be dropped out in the process of downsampling. To preserve enough significant information, U-net provides an approach to merge the symmetric parts of convolution and deconvolution.

To suppress the defocus noise, a sufficient training set should be collected. The training set contains the encoded hologram with speckle-like noise and the labeled image without noise. In the training process, the reconstructed images with noise are set as input images, which would propagate forward to obtain the predicted images. The loss function L is defined as a function between the predicted image and the labeled standard image without noise by mean square error (MSE),

$$L=\frac{1}{M}\sum_{i=1}^{M}|y_i-\hat{y}_i|^2,\tag{5}$$

where M is the total count of pixels, y_i is the labeled standard image, and $\hat{y}_i$ represents the predicted image.

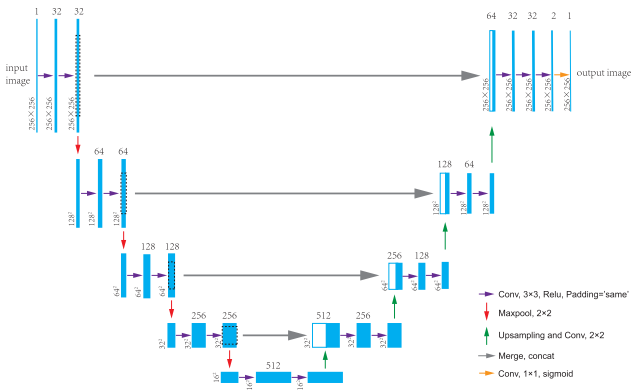

Fig. 2. The architecture of U-net. 'Conv, 3×3' represents a 3×3 convolution kernel with the ReLU activation function. 'Padding=same' means that the matrix dimensions of the input and output in the convolution layer are the same. 'Maxpool 2×2' represents the function to choose the maximum value from a 2×2 matrix. 'Upsampling and conv, 2×2' stands for upsampling using a 2×2 convolution kernel. Each blue box represents a multi-channel feature map, while the white ones represent the copied feature maps.

3. Results and Discussion

The proposed method was demonstrated via simulation. The optical process was simulated with Matlab, while the reconstruction results based on the proposed U-net method were generated with Pytorch. The GPU used in the simulation was NVIDIA GTX 1080Ti with 16 GB memory. A He–Ne laser centered at 632.8 nm was used as the source. The focal length of the lens L1 and L2 was $f=75$ mm. An object with two sections located at $z_1=30$ mm and $z_2=30.3$ mm was used in the simulation. The size of each section was 2 mm × 2 mm, which was sampled to 256×256 pixels.

3.1. Simple graphics

The U-net method was first verified with simple graphics, such as the English alphabet. In the training process, it is important to generate enough data sets. In the simulation, 386 original images were used, with each one passing through 27 different random phase pupils in the OSH system as shown in Fig. 1. In this way, 10,422 data sets were produced for training. Some of the sample images with speckle-like noise are shown in Fig. 3, which are used as the input images of the U-net model. The speckle-like noises are generated from the other section based on Eqs. (1) and (2). The sectional images are in the database as mentioned above. One can observe from this figure that different speckle-like noise was added according to Eq. (2). The corresponding noise-free images, also denoted as the standard images or reference images of the U-net, are shown in Fig. 4.

To accelerate the convergence, we chosed the method for stochastic optimization with a learning rate equal to 0.0001[32]. The parameters less than 0.5 were dropped out to prevent over-fitting[33]. The relationship between the training loss and the iteration times is shown as the blue curve in Fig. 5, while the orange one represents the relationship between the validation loss and the iteration times. It can be seen from this figure that the loss of the training data as well as the validation data both decrease with the iteration times.

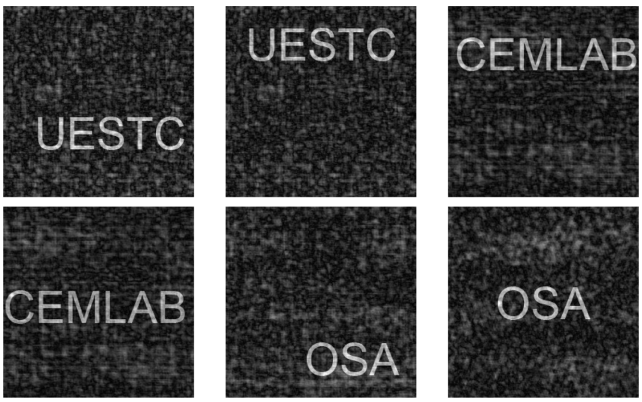

Fig. 3. Input images for the U-net model.

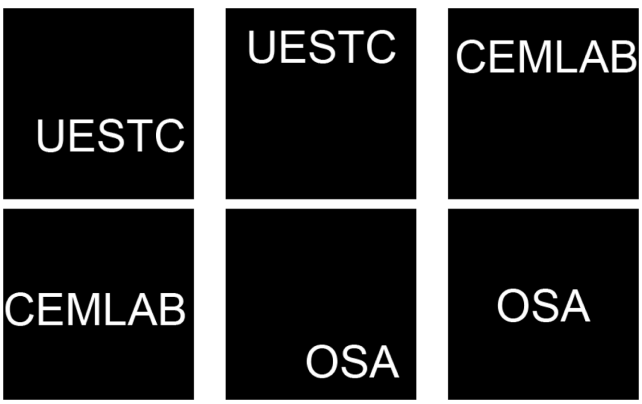

Fig. 4. Standard images of the U-net model.

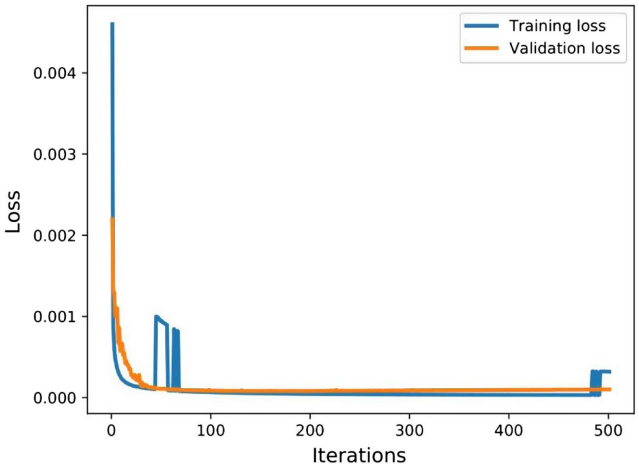

Fig. 5. The relationship between the loss function and iteration times.

The first simulation results proved that the U-net architecture really did a good job of learning the characteristics of speckle-like noise. Here, we demonstrate the reconstruction results under different random phase pupils in Fig. 6. Three different test images were used to verify the proposed method. The image with the letters 'ABC' was in the training data sets, while the ones with letter '*XYZ*' and a Chinese character were not included. One can observe from this figure that the speckle-like noise has been eliminated successfully.

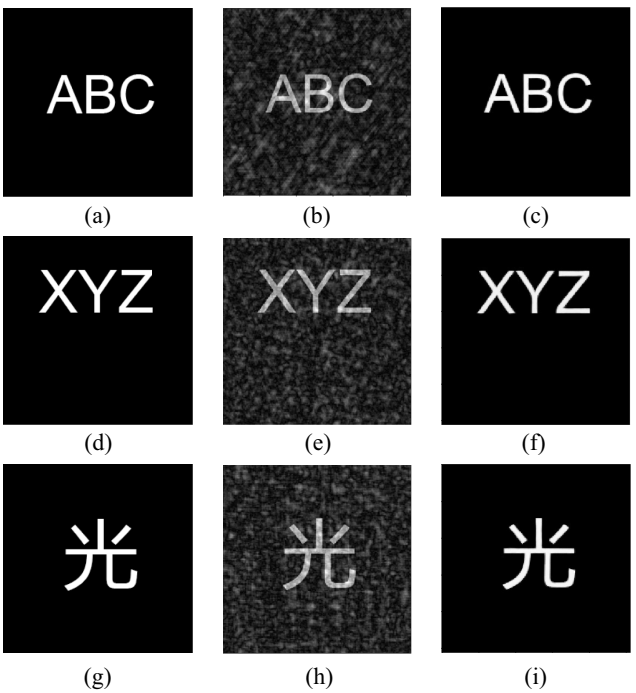

Fig. 6. The reconstruction results with U-net. (a), (d), and (g) are the original images. (b), (e), and (h) are input images with speckle-like noise generated by different random phase pupils. (c), (f), and (i) are the corresponding reconstructed output images.

To evaluate the reconstruction effect with the U-net, we compare the results with two important factors: the peak signal-to-noise ratio (PSNR) and the structural similarity (SSIM). The PSNR can be defined as[34]

$$\mathrm{MSE} = \frac{1}{M_1 N_1} \sum_{i=1}^{M1} \sum_{j=1}^{N_1} [X(i,j) - Y(i,j)]^2, \tag{6}$$

$$\mathrm{PSNR} = 10 \lg\left[\frac{(2^n - 1)^2}{\mathrm{MSE}}\right], \tag{7}$$

where $X(i,j)$ is the output image, while $Y(i,j)$ is the corresponding reference image. M_1 and N_1 are the number of columns and rows, respectively, and MSE stands for mean square error.

The SSIM is used to quantify the visibility of differences between the output image and the corresponding reference image. The quality assessment is based on the degradation of structural information and can be expressed as[35]

$$\mathrm{SSIM} = l(X,Y) \times c(X,Y) \times s(X,Y), \tag{8}$$

where

$$l(X,Y) = \frac{2\mu_X\mu_Y + C_1}{\mu_X^2 + \mu_Y^2 + C_1},$$

$$c(X,Y) = \frac{2\sigma_X\mu_Y + C_2}{\sigma_X^2 + \sigma_Y^2 + C_2},$$

$$s(X,Y) = \frac{\sigma_{XY} + C_3}{\sigma_X\sigma_Y + C_3},$$

where X and Y represent the output image and the reference image, respectively. μ_X and μ_Y are the means of X and Y, σ_X and σ_Y are the variances of X and Y, σ_{XY} is the covariance of X and Y, respectively, and C_1, C_2, and C_3 are constants. In general, $C_1 = (K_1 \times L)^2$, $C_2 = (K_2 \times L)^2$, and $C_3 = C_2/2$. Here, we assume that $K_1 = 0.01$, $K_2 = 0.03$, and $L = 255$.

The quantified assessment results are shown in Table 1. One can observe from the row 'Input VS Original' that, the PSNR for all cases is all around 15 dB, and the SSIM is quite small at around 0.06. This means that the speckle-like noise has greatly degraded the signal-to-noise ratio, and the similarity between the noisy images and the original ones has also been reduced to a rather low degree. While for the scenario of 'Output VS Original', the PSNR for all cases is increased above 32 dB, and the SSIM is also raised close to 1. This indicates that speckle haze has been eliminated successfully, and the output images of the U-net have small distortion and high similarity with the original ones.

It is worth noting that the computation time for all cases are around 33.3 ms, and the computer is configured as NVIDIA GTX 1080Ti with 16 GB memory.

We have also measured the sectioning results of the Chinese character in Fig. 6(h) with different noise ratios. The noise is generated from different sections with the traditional method[16], as is shown in Figs. 7(a)–7(c). The sectional images generated by the U-net method are shown in Figs. 7(d)–7(f), respectively. Table 2 presents the corresponding quantified results. It can be seen from Fig. 6 and Table 2 that for sectional images with different noise ratios, the values of the PSNR and the SSIM all increase significantly (with the PSNR around 32 dB and the SSIM close to 1). This result shows that the U-net method can successfully remove the defocus noise with different noise ratios.

Table 1. The Quantified Performance of the U-net Using Different Test Images[a].

	Test sample	MSE	PSNR (dB)	SSIM (∈[0,1])
Input	'ABC'	1909.14	15.32	0.5782
VS	'XYZ'	1942.50	15.25	0.6171
Original	'光'	2138.30	14.80	0.5584
Output	'ABC'	15.8	36.15	0.9535
VS	'XYZ'	36.4	32.52	0.9326
Original	'光'	32.7	32.98	0.9383

[a]'Input' is the input image of the U-net, in which the speckle-like noise is added. 'Original' stands for the original reference image. 'Output' means the denoised output image of the U-net.

Table 2. The Quantified Performance of the U-net with Different Noise Ratios.

	Test sample in Fig. 7	MSE	PSNR (dB)	SSIM (∈[0,1])
Input	(a)	2256.34	13.35	0.3732
VS	(b)	2387.33	12.33	0.3245
Original	(c)	2489.95	10.58	0.2788
Output	(d)	33.1	32.56	0.9305
VS	(e)	32.1	32.47	0.9289
Original	(f)	33.7	32.18	0.9245

3.2. Complex graphics

In this subsection, the situation of complex graphics is tested. Some samples of the complex graphics are shown in Fig. 8, in which some standard digital image processing images are included, such as Barbara, Cameraman, and Peppers. These images were used in the OSH system to generate holograms with complex graphics in order to test the noise-suppressing ability of the proposed method.

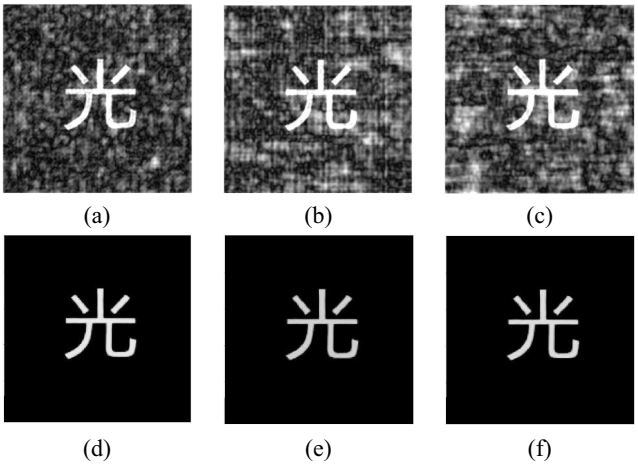

Fig. 7. Sectioning results with different noise ratios based on the U-net method.

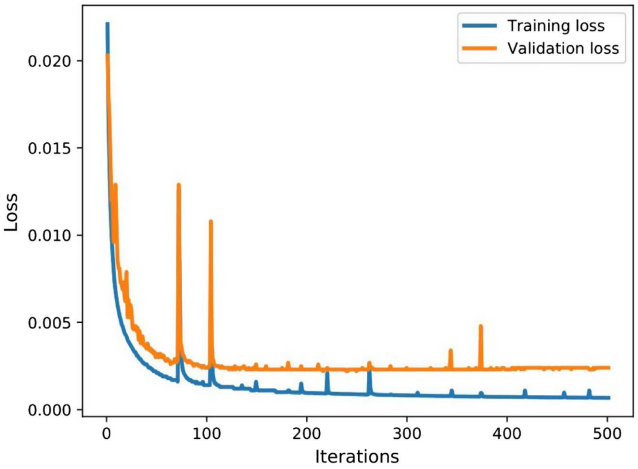

Fig. 9. The relationship between the loss function and iteration times in the complex graphics.

Fig. 8. The original images for generating the training data sets.

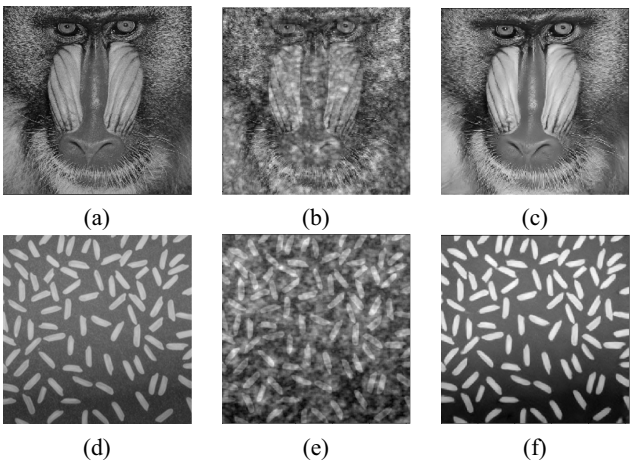

Fig. 10. The test results of the complex graphics with U-net. (a),(d) are the original images of 'Monkey' and 'Rice'. (b),(e) are the input images with speckle-like noise generated by different random phase pupils. (c),(f) are the corresponding output images of the U-net.

To generate the training data sets, 364 original images were used, with each one passing through 30 different random phase pupils. In this way, 10,920 data sets were produced for the U-net model. It is important to mention that the speckle-like noises are generated from the other section based on Eqs. (1) and (2). The sectional images are in the database as mentioned above.

The loss function for both the training data and the validation data were calculated in the training process. The results are shown in Fig. 9. As can be seen from this figure, the loss function decreases with the iteration times. One can expect that the defocus noise can be suppressed after 400 iterations. The generalization gap between training loss and validation loss is measured to be around 0.002 in this case.

The reconstruction results using the U-net with complex graphics are shown in Fig. 10. Two complex graphics named 'Monkey' and 'Rice' were used to test the U-net method. It can be seen from this figure that most of the speckle haze has been eliminated.

The quantified evaluation results are listed in Table 3. By comparing the data in the row 'Input VS Original' with that in the row 'Output VS Original', one can observe that the values of the PSNR and the SSIM have both increased after the U-net processing. The increased PSNR represents the improvement of

Table 3. The Quantified Performance of the U-net Using Complex Test Images.

	Test sample	MSE	PSNR (dB)	SSIM (∈[0,1])
Input VS	'Monkey'	996.3	18.15	0.5613
Original	'Rice'	1218.2	18.15	0.5613
Output VS	'Monkey'	278.5	23.68	0.8127
Original	'Rice'	57.6	30.53	0.8378

the signal-to-noise ratio, which means that the noise has been decreased. While the variation of the SSIM suggests higher similarity between the output images of the U-net and the original ones.

It can also be concluded from Tables 1 and 3 that as the input image becomes more complex, the improvement of the PSNR would degrade. This indicates that it is harder for U-net to distinguish the features between the image and the speckle noise when the complexity of the image increases.

3.3. Reconstruction of 3D objects

To verify the feasibility of eliminating the defocus noise in OSH, we have also evaluated the reconstruction results of two different 3D objects in this subsection. The performances between the conventional reconstruction method and the proposed one are also analyzed.

The first 3D object used in the simulation contains two slices, as is shown in Fig. 11(a). Each slice has a size of 1 mm × 1 mm and are sampled to 256 × 256 pixels. The focal length of the lens in the optical system is set as 75 mm. The locations of the two slices are $z_1 = 9$ mm and $z_2 = 10$ mm, respectively. The wavelength of the laser is 632.8 nm. Figure 11(b) shows the recorded hologram.

Figures 12(a) and 12(b) show the retrieved sectional images using the conventional algorithm[16], with reconstruction distance at $z_1 = 9$ mm and $z_2 = 10$ mm, respectively. One can see from this figure that at the front plane ($z_1 = 9$ mm), the slice "UESTC" is in focus, and the image from the other section has become defocused noise spreading around the figure. This speckle-like noise obviously degrades the image quality, which results in a poor 3D reconstruction effect. While the corresponding results of the proposed U-net based method are shown in Figs. 12(c) and 12(d), in which each section has been recovered clearly with degraded defocus noise.

We have also tested the U-net method with a simulated hologram of a rocket. The semi-transparent 3D rocket is shown in Fig. 13. It has been divided into six uniformly separated sections along the z-axis. The ith section is located z_i away from the focal plane, with $z_1 = 31$ mm and $z_6 = 36$ mm. Each slice has a size of 1 mm × 1 mm, and are sampled to 256 × 256 pixels, as can be seen in Figs. 14(a)–14(f).

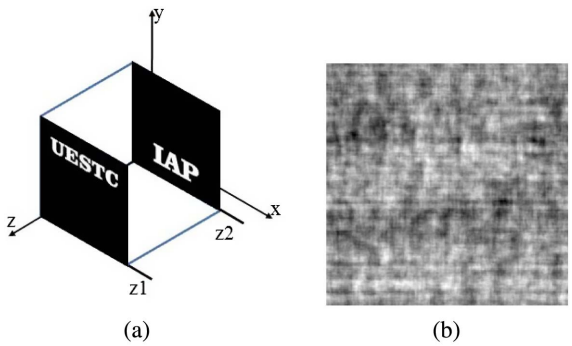

Fig. 11. (a) Object with two slices, and (b) the recorded hologram.

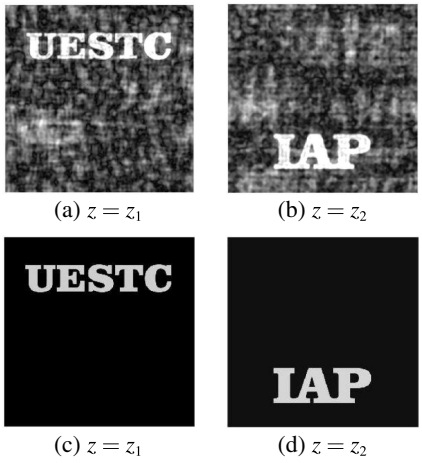

Fig. 12. (a), (b) Sectional results of the conventional method. (c), (d) Sectional results of the proposed U-net method, with z_1 = 9 mm and z_2 = 10 mm.

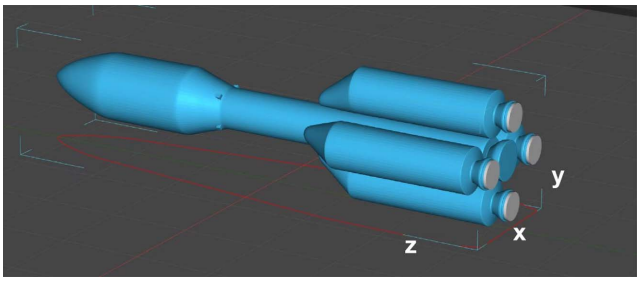

Fig. 13. The 3D rocket.

The reconstructed images with the traditional method are shown in Figs. 14(g)–14(l), while the ones with the U-net-based method are shown in Figs. 14(m)–14(r). The corresponding reconstruction distances are set as $z_1 = 31$ mm, $z_2 = 32$ mm, ..., and $z_6 = 36$ mm. It can be seen from these images that the proposed method outperforms the traditional method in suppressing the defocus noise.

The quantified assessment results of each section are also analyzed. The results are shown in Figs. 15 and 16, in which number of section denotes the ith sectional image. One can observe from Fig. 15 that with the traditional method, the PSNR drops from 16 dB at Section-1 to 11 dB at Section-6. This is caused by the defocus noise from other sections. While with the U-net method, the output PSNR have been upgraded to around 35 dB, which indicates the improvement of the image quality. The measured SSIM results of each method are shown in Fig. 16. One can find that the value of the SSIM are all above 0.9 with the proposed U-net method. While for the case with traditional method, the SSIM decreases from 0.36 at Section-1 to 0.28 at Section-6. This also implies better sectioning results with the U-net method.

In conclusion, the U-net based method can be adapted to 3D objects with multiple sections. It outperforms the traditional

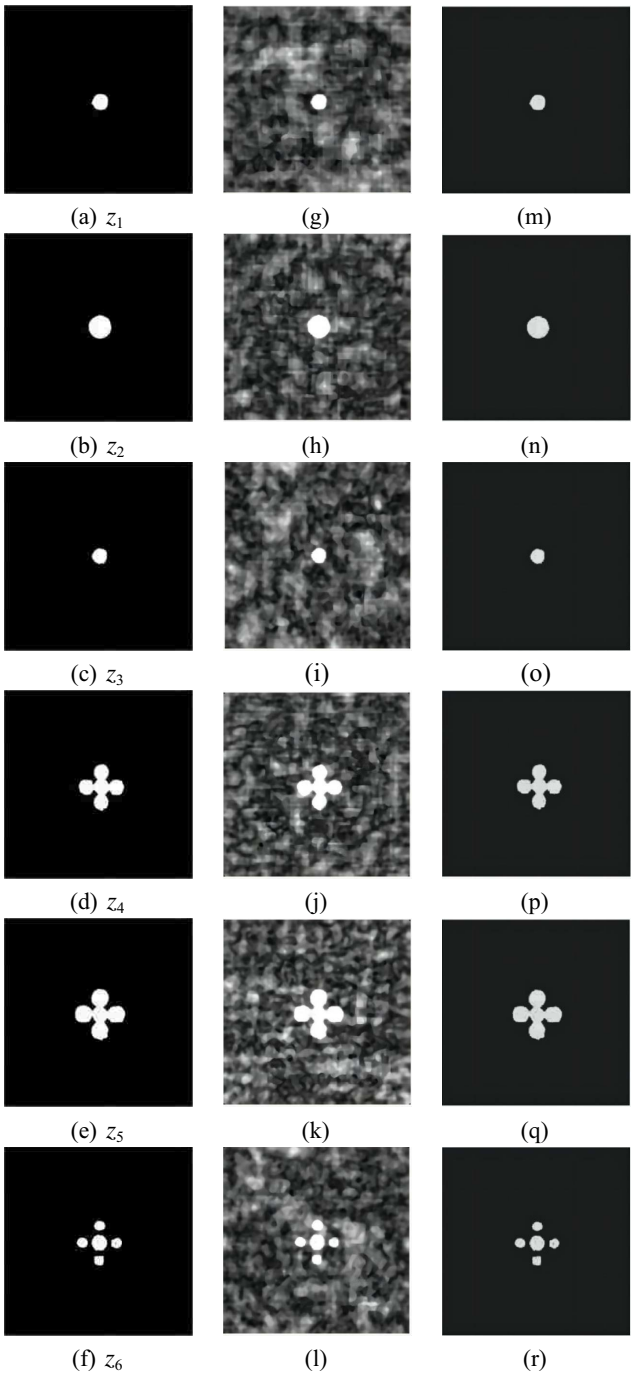

Fig. 14. (a)–(f) Sectional images of the 3D rocket. (g)–(l) Reconstructed images with the traditional method. (m)–(r) Reconstructed images with the proposed method.

method in removing defocus noise as well as recovering multiple sections. One can also deduce from the tables and figures in section 3 that the improvement of the PSNR ranges from around 5 dB to 20 dB, which indicates better sectioning results over the conditional method. However, the scenario of the object with complex sectional images still needs further investigation. This can be done either by increasing the training data sets or adjusting the training mode in the deep learning algorithm. These are all the future tasks to be done.

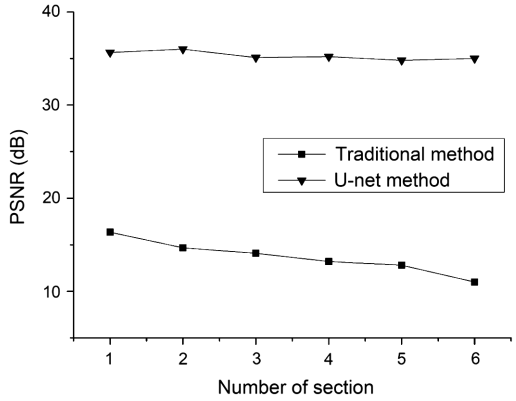

Fig. 15. PSNR of the sectioning results.

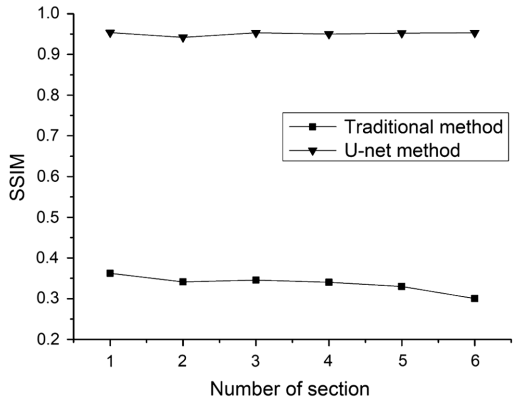

Fig. 16. SSIM of the sectioning results.

4. Conclusion

Speckle-like noise is generated by the random phase pupil in an OSH system, which is hard to reduce. This work verifies the feasibility of the U-net neural network for this task, which provides a new way for us to realize fast and effective defocus noise suppressing in OSH. Simulation results show that the proposed method works well both in simple and complex graphics. We believe that the proposed method can also be applied to other digital holography systems, especially for biomedical applications where it is hard to get enough training data sets.

References

1. J. Swoger, M. Martınez-Corral, J. Huisken, and E. H. K. Stelzer, "Optical scanning holography as a technique for high-resolution three-dimensional biological microscopy," J. Opt. Soc. Am. A **19**, 1910 (2002).
2. B. W. Schilling and G. C. Templeton, "Three-dimensional remote sensing by optical scanning holography," Appl. Opt. **40**, 5474 (2001).

3. T.-C. Poon, T. Kim, and K. Doh, "Optical scanning cryptography for secure wireless transmission," Appl. Opt. **42**, 6496 (2003).
4. H. Di, K. Zheng, X. Zhang, E. Y. Lam, T. Kim, Y. S. Kim, T.-C. Poon, and C. Zhou, "Multiple-image encryption by compressive holography," Appl. Opt. **51**, 1000 (2012).
5. Z. Ren, N. Chen, and E. Y. Lam, "Extended focused imaging and depth map reconstruction in optical scanning holography," Appl. Opt. **55**, 1040 (2016).
6. Z. Ren, N. Chen, and E. Y. Lam, "Automatic focusing for multisectional objects in digital holography using the structure tensor," Opt. Lett. **42**, 1720 (2017).
7. H. Ou, Y. Wu, E. Y. Lam, and B.-Z. Wang, "New autofocus and reconstruction method based on a connected domain," Opt. Lett. **43**, 2201 (2018).
8. H. Ou, Y. Wu, E. Y. Lam, and B.-Z. Wang, "Axial localization using time reversal multiple signal classification in optical scanning holography," Opt. Express **26**, 3756 (2018).
9. J. Ke, T.-C. Poon, and E. Y. Lam, "Depth resolution enhancement in optical scanning holography with a dual-wavelength laser source," Appl. Opt. **50**, H285 (2011).
10. H. Ou, T.-C. Poon, K. K. Y. Wong, and E. Y. Lam, "Depth resolution enhancement in double-detection optical scanning holography," Appl. Opt. **52**, 3079 (2013).
11. H. Ou, T.-C. Poon, K. K. Y. Wong, and E. Y. Lam, "Enhanced depth resolution in optical scanning holography using a configurable pupil," Photonics Res. **2**, 64 (2014).
12. E. Y. Lam, X. Zhang, H. Vo, T.-C. Poon, and G. Indebetouw, "Three-dimensional microscopy and sectional image reconstruction using optical scanning holography," Appl. Opt. **48**, H113 (2009).
13. X. Zhang, E. Y. Lam, and T.-C. Poon, "Reconstruction of sectional images in holography using inverse imaging," Opt. Express **16**, 17215 (2008).
14. T. Kim, "Optical sectioning by optical scanning holography and a Wiener filter," Appl. Opt. **45**, 872 (2006).
15. A. C. S. 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 (2016).
16. Z. Xin, K. Dobson, Y. Shinoda, and T.-C. Poon, "Sectional image reconstruction in optical scanning holography using a random-phase pupil," Opt. Lett. **35**, 2934 (2010).
17. H. Ou, H. Pan, E. Y. Lam, and B.-Z. Wang, "Defocus noise suppression with combined frame difference and connected component methods in optical scanning holography," Opt. lett. **40**, 4146 (2015).
18. L. Jin-xi, Z. Ding-fu, Y. Sheng, Z. Yuan-yuan, Z. Luo-zhi, H. Dong-ming, and Z. Xin, "Modified image fusion technique to remove defocus noise in optical scanning holography," Opt. Commun. **407**, 234 (2018).
19. T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent trends in deep learning based natural language processing," IEEE Comput. Intell. Mag. **13**, 55 (2018).
20. A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep learning for computer vision: a brief review," Comput. Intell. Neurosci. **2**, 13 (2018).
21. S. Ghosal, D. Blystone, A. K. Singh, B. Ganapathysubramanian, A. Singh, and S. Sarkar, "An explainable deep machine vision framework for plant stress phenotyping," Proc. Natl. Acad. Sci. U.S.A. **115**, 4613 (2018).
22. T. Zeng, Y. Zhu, and E. Y. Lam, "Deep learning for digital holography: a review," Opt. Express **29**, 40572 (2021).
23. Z. Ren, Z. Xu, and E. Y. Lam, "Learning-based nonparametric autofocusing for digital holography," Optica **5**, 337 (2018).
24. T. Pitkäaho, A. Manninen, and T. J. Naughton, "Focus prediction in digital holographic microscopy using deep convolutional neural networks," Appl. Opt. **58**, A202 (2019).
25. Y. Rivenson, Y. Zhang, H. Günaydin, D. Teng, and A. Ozcan, "Phase recovery and holographic image reconstruction using deep learning in neural networks," Light Sci. Appl. **7**, 17141 (2017).
26. 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 (2017).
27. S.-J. Kim, C. Wang, B. Zhao, H. Im, J. Min, H. J. Choi, J. Tadros, N. R. Choi, C. M. Castro, R. Weissleder, H. Lee, and K. Lee, "Deep transfer learning-based hologram classification for molecular diagnostics," Sci. Rep. **8**, 17003 (2018).
28. T.-C. Poon, *Optical Scanning Holography with MATLAB* (Springer, 2017).
29. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv:1409.1556 (2015).
30. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," arXiv:1505.04597 (2015).
31. G. Zhang, T. Guan, Z. Shen, X. Wang, T. Hu, D. Wang, Y. He, and N. Xie, "Fast phase retrieval in off-axis digital holographic microscopy through deep learning," Opt. Express **26**, 19388 (2018).
32. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv:1412.6980 (2014).
33. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," J. Mach. Learn. Res. **15**, 1929 (2014).
34. Y. Fisher, *Fractal Image Compression* (Springer, 1995).
35. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Trans. Image Process. **13**, 600 (2004).