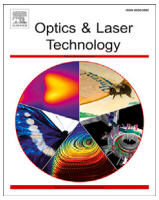
Optics & Laser Technology 190 (2025) 113246

Contents lists available at ScienceDirect

Optics and Laser Technology

journal homepage: www.elsevier.com/locate/jolt



Full length article

Defocus and speckle noise suppression in optical scanning holography

Ruiwei Xu [a], Haiyan Ou [a, b,*], Edmund Y. Lam [c]

[a] Shenzhen Institute of Advanced Study, University of Electronic Science and Technology of China, Shenzhen, 518110, Guangdong, China
[b] School of Physics, University of Electronic Science and Technology of China, Chengdu, 611731, Sichuan, China
[c] Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam, Hong Kong, China

HIGHLIGHTS

- Suppression of both defocus and speckle noise generated in OSH.
- Propose an effective training strategy for small datasets with reconstruction noise in OSH.
- The approach is rigorously validated through simulations and experimental data, demonstrating robustness and efficiency for industrial applications.

ARTICLE INFO

Keywords:
Multi-type noise
Digital holography
Small dataset
Deep learning
Image denoising

ABSTRACT

Optical scanning holography (OSH) is a versatile and digital technique for recording and reconstructing holograms, with a wide range of potential applications in three-dimensional (3D) imaging, display, data storage, and non-destructive testing. However, undesirable noise is inevitable in OSH for sectional image reconstruction. In this manuscript, we implemented a classic Convolutional Neural Network (CNN), specifically U-net, which can be leveraged for denoising with only a small dataset corrupted by both defocus and speckle noise. Furthermore, we show that this is possible by applying various training strategies with different hyperparameters. The proposed method has been validated using both simulation and experimental data, which demonstrate that the approach is both effective and efficient.

1. Introduction

Optical scanning holography (OSH) is a precise digital holography (DH) technique that records information of three-dimensional (3D) objects through a single two-dimensional (2D) pixel-by-pixel scan. The reconstruction process in OSH leverages digital signal processing to recreate the original 3D wavefront and enable various applications, such as microscopy [1], 3D pattern recognition [2], remote sensing [3], and tumor segmentation from magnetic resonance images (MRI), *etc* [4].

To retrieve the information contained within a digital hologram of 3D objects in OSH, two key challenges must be addressed: (1) depth of focus determination, and (2) in-focus pattern reconstruction. Numerous studies have been conducted to optimize the solutions to these two problems.

To retrieve the accurate depth of focus for reconstruction, various methods have been proposed to automatically locate the object depth. These include the connected domain approach [5], time-reversal multiple signal classification [6], structure tensor analysis [7], and numerous other techniques. Moreover, alternative methods like dual-wavelength laser [8] and configurable pupil [9] have also been proposed for further improvement of the depth resolution during the reconstruction process.

While for the in-focus pattern reconstruction, the challenge lies in the inevitable noise from the other sections, also known as defocus noise. Conventional methods employ techniques like iterative shrinkage-threshold algorithms [10], inverse imaging [11], and modified image fusion [12] to reduce or minimize the impact of this noise. Besides these image processing methods, people have also explored modifying the optical system itself to suppress the defocus noise. An interesting work was published by Zhou et al. in ref. [13], in which the author demonstrated how to transfer the defocus noise into speckle-like patterns by random phase pupil. This noise can be further suppressed by either average method or connected domain, improving the overall image quality and signal-to-noise ratio [14].

The choice of reconstruction and denoising methods in OSH is critical for achieving high-quality images. While conventional methods are valuable, they often face limitations. They either ignore certain parts

* Corresponding author at: Shenzhen Institute of Advanced Study, University of Electronic Science and Technology of China, Shenzhen, 518110, Guangdong, China
Email address: ouhaiyan@uestc.edu.cn (H. Ou).

https://doi.org/10.1016/j.optlastec.2025.113246
Received 3 July 2024; Received in revised form 15 November 2024; Accepted 21 May 2025
Available online 31 May 2025

of the original signals during the iterative process and signal transfer [10,11], or indiscriminately restore all the signal from the hologram without considering the unique characteristics of different regions [13]. This results in the inferior quality of the OSH image reconstructions and denoising recoveries, as the information from the hologram is not properly utilized. Additionally, the runtime of these algorithms can pose significant challenges, especially in complex imaging scenarios.

Recent years have witnessed a surge in the application of deep learning techniques across various industries, thanks to the rapid advancements in artificial intelligence. Deep learning, a subset of machine learning, has proven to be highly effective in solving complex problems and has been successfully applied in various domains [15–17]. It has also been increasingly applied to the field of digital holography, showing promising results in addressing various challenges mentioned above. Several proposed works have demonstrated the effectiveness of deep learning strategies in DH: Ren et al. presented a CNN based auto-focusing method with regression prediction [18], in Ref. [19] Rivenson and colleagues used deep learning for phase recovery and image reconstruction in a single process; while Kim established a direct hologram classification neural network model for cell classification [20]. These deep learning methods have shown promise in improving accuracy and efficiency over traditional algorithms. In addition, deep learning has shown great potential in image reconstruction enhancement: Ren et al. proposed a deep learning method for fringe pattern super-resolution [21]; Liu et al. demonstrate the performance of U-net architecture under a Generative Adversarial Networks(GANs) structure in super-resolving both pixel-limited and diffraction-limited in-line holograms of lung tissue sections and smear samples [22]; Fang and colleagues also applied a GAN, which is structured with a U-net descriminator and a DenseNet generator, in DH interferometry speckle denoising. Their method shows better performance than other denoising algorithms for processing experimental strain data from DH [23]; Tang et al. then propsed a CNN method for continuous phase denoising in DH microscopy [24], Zhuang and colleagues proposed an OSH reconstruction method that directly retrieves the reconstruction image from the hologram [25], and a lot more deep learning based studies have been conducted on super-resolution and noise suppression tasks [26–28].

In this paper, we aimed to reduce OSH-reconstruction noise using a fine-tuned U-shaped convolutional neural network (U-net) on a multi-type noise small dataset. The U-net architecture was chosen for its ability to learn both local and global features from the input data, and the fine-tuning process allowed the model to adapt to the specific task at hand. Both simulation and experimental results have demonstrated the effectiveness and generalization performance of the proposed method.

This paper is structured as follows: In Section 2, the principle of OSH and the network architecture of U-net are illustrated. Simulation results are presented and discussed in Section 3. In Section 4, the effectiveness and generalization performance of the proposed method are demonstrated through the utilisation of experimental data. A brief summary and prospects for future work are provided in Section 5.

2. Principle

2.1. Optical scanning holography

The OSH system depicted in Fig. 1 consists of several components, which work together to create a hologram. The coherent light generated by the He-Ne laser will be split into two parts by beam splitter (BS1). An acousto-optic frequency shifter (AOFS) is used to shift the frequency of the upper beam from ω_0 to $\omega_0+\Omega$. The two beams with distinct frequencies will then be directed to the X-Y scanner by separate sets of optical elements (mirrors, pupils and lenses). The X-Y scanner works in a raster pattern, allowing the system to capture the 3D information of the object. After that, a photo detector (PD) is used to convert light information collected by lens L3 into an electric signal. It will then be transformed into a hologram for further signal processing, which includesband-pass filtering, demodulation processes, et al. The generated hologram can be expressed as [29]

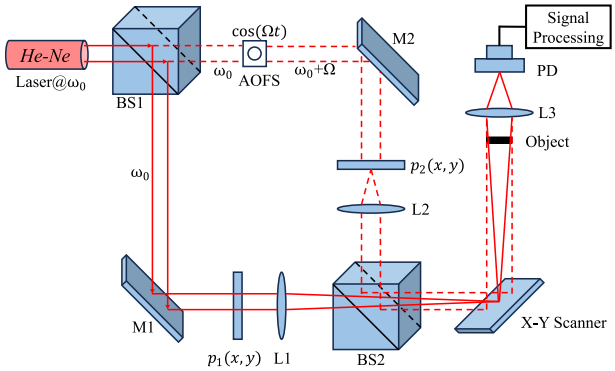


Fig. 1. Typical OSH system setup. BS1&BS2, beam splitter; M1&M2, mirror; AOFS, acousto-optic frequency shifter; $p_1(x,y)$& $p_2(x,y)$, pupils; L1, L2&L3, lens; PD, photo detector.

$$holo(x,y)=\sum_{i=1}^{N}\mathcal{F}^{-1}\{\mathcal{F}[I(x,y;z_i)]\cdot OTF(k_x,k_y;z_i)\} \tag{1}$$

where $holo(x,y)$ is the generated hologram, $I(x,y;z_i)$ is the complex amplitude of the object and it has been discretized into N sections along the z-axis. x and y are space coordinates of the object and z_i is the axial coordinate of the i-th section. $\mathcal{F}$ and $\mathcal{F}^{-1}$ indicate Fourier transformation and inverse Fourier transformation, respectively. k_x and k_y are frequency domain coordinates, $\mathrm{OTF}(k_x,k_y;z_i)$ is the optical transfer function (OTF) of the system given by:

$$\begin{aligned}\mathrm{OTF}(k_x,k_y;z_i)=exp\left[-j\frac{z_i}{2k_0}(k_x^2+k_y^2)\right]\times\iint p_1^*(x',y')\\ \times p_2\left(x'+\frac{f}{k_0}k_x,y'+\frac{f}{k_0}k_y\right)\\ \times exp\left[j\frac{z_i}{f}(x'k_x+y'k_y)\right]dx'dy'\end{aligned} \tag{2}$$

where $k_0=2\pi/\lambda$ is the wavenumber (or propagation constant), and λ represents the wavelength of the He-Ne laser. $*$ represents conjugate operation on complex matrix, and f is the focal length of the lens L3. The variables x' and y' denote the intermediate variables used within the convolution of p_1 and p_2. It's notable in Eq. (2) that the OTF changes according to the choise of pupils, as is the spatial impulse response of the system given below:

$$h(x,y;z_i)=\mathcal{F}^{-1}\{OTF(k_x,k_y;z_i)\} \tag{3}$$

To retrieve the information at a certain distance (*e.g.*, $z=z_d$), the spatial impulse response for reconstruction $h_r(x,y;z_d)$ is required. The reconstructed sectional image can therefore be expressed as:

$$\begin{aligned}I_{out}(x,y)&=holo(x,y)\otimes h_r(x,y;z_d)\\&=\sum_{i=1}^{N}\mathcal{F}^{-1}\{\mathcal{F}[I(x,y;z_i)]\cdot OTF(k_x,k_y;z_i)\\&\quad\cdot OTF_r(k_x,k_y;z_d)\}\end{aligned} \tag{4}$$

where $I_{out}(x,y)$ is the reconstructed sectional image, and $\otimes$ denotes the convolution operation. $OTF_r(k_x,k_y;z_d)$ denotes the expression of $h_r(x,y;z_d)$ in spatial frequency domain, derived from Eq. (3)

As previously stated in Eq. (2), the overall performance of an OSH system can be tailored to meet specific application needs by meticulously selecting and designing the pupils. Here, we mainly consider two strategies for pupil selection:

a) In the normal OSH system, the pupils are chosen as $p_1(x,y)=1$, $p_2(x,y)=\delta(x,y)$. (Set p_1 as a collimating mirror, and p_2 as a pinhole board

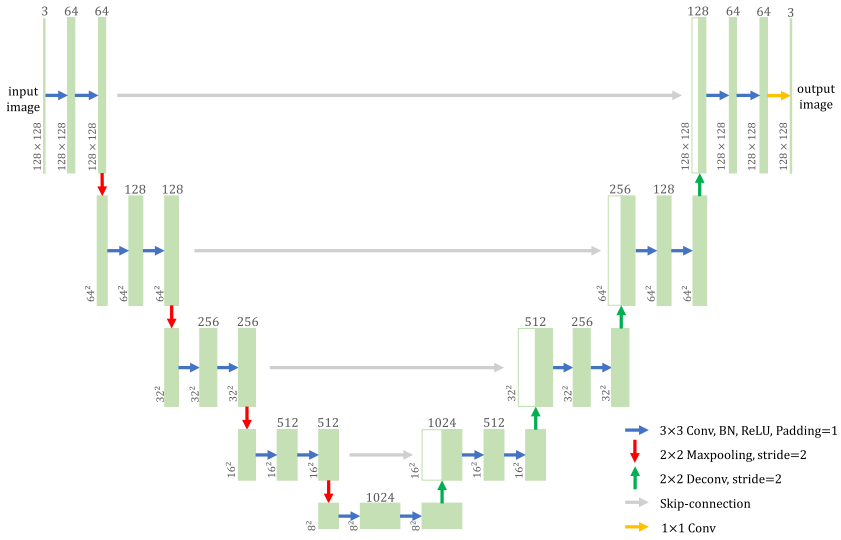


Fig. 2. The network architecture of U-net. The arrows indicate the layers of the sampling process within the network, and the green blocks represent the feature maps generated by each sampling process.

for example.) Then the reconstructed sectional image in Eq. (4) can be expressed as [29]:

$$I_{out}(x,y) = \sum_{i=1}^{N} \mathcal{F}^{-1}\left\{\mathcal{F}[I(x,y;z_i)] \times exp\left[j\frac{z_d - z_i}{2k_0}\left(k_x^2 + k_y^2\right)\right]\right\}$$
$$= I(x,y;z_d) + \sum_{i\neq d}^{} \mathcal{F}^{-1}\left\{\mathcal{F}[I(x,y;z_i)] \times exp\left[j\frac{z_d - z_i}{2k_0}(k_x^2 + k_y^2)\right]\right\} \quad (5)$$

where $I_{out}(x,y)$ is composed of two terms. The first term is the desired sectional image $I(x,y;z_d)$, while the second term denotes defocus noise generated from other sections. The pupils' functions are simplified according to their values of 1 or $\delta(x,y)$.

b) In the case of an OSH system with random-phase pupil, the specific design can be set as follows: $p_1(x,y) = exp[j2\pi \cdot s(x,y)]$, where $s(x,y)$ denotes an independent random function uniformly distributed in the interval $[0,1]$ and $p_2(x,y) = 1$. To achieve this mechanism, p_1 would be set as a random phase mask with a spatial light modulator, and p_2 as a collimating mirror. The resulting interference pattern created by this setup can be recorded to form a hologram. During the numerical reconstruction phase, a spatial impulse response for reconstruction $h_r(x,y;z_d)$ is also required. In this case, it is constructed by two numerical pupils, with $p_{1r}(x,y) = 1$ and another random phase mask $p_{2r}(x,y)$ to meet the condition $p_1^*(x,y) \cdot p_{2r}(x,y) = 1$. In this way, Eq. (4) can be rewritten as follows [13]:

$$I_{out}(x,y) = \sum_{i=1}^{N} \mathcal{F}^{-1}\left\{\mathcal{F}[I(x,y;z_i)] \times exp\left[j\frac{z_d - z_i}{2k_0}(k_x^2 + k_y^2)\right]\right.$$
$$\left. \times P_1^*\left(-\frac{z_i k_x}{f}, -\frac{z_i k_y}{f}\right) \cdot P_{2r}\left(\frac{z_d k_x}{f}, \frac{z_d k_y}{f}\right)\right\}$$
$$= I(x,y;z_d) + \sum_{i\neq d}^{} \mathcal{F}^{-1}\left\{\mathcal{F}[I(x,y;z_i)] \times exp\left[j\frac{z_d - z_i}{2k_0}(k_x^2 + k_y^2)\right]\right.$$
$$\left. \times P_1^*\left(-\frac{z_i k_x}{f}, -\frac{z_i k_y}{f}\right) \cdot P_{2r}\left(\frac{z_d k_x}{f}, \frac{z_d k_y}{f}\right)\right\} \quad (6)$$

where P_1 and P_{2r} represent the Fourier transformation of p_1 and p_{2r}, respectively. Both P_2 and P_{1r} are omitted due to their values being equal to 1. By comparing both the second items of Eqs. (5) and (6), it can be inferred from Eq. (6) that the random phase pupil has transformed the out-of-focus noise into forms that differ from those of the traditional dofocus noise, as shown in Eq. (5).

2.2. U-net

U-net is a popular CNN architecture that was originally designed for image segmentation tasks [30]. It is based on the fully convolutional network (FCN) architecture, which was first proposed by Long et al. in 2016 [31].

The network is characterised by an encoder-decoder structure, and the name 'U-net' is derived from its symmetric architecture, which resembles the letter 'U'. The network architecture of U-net is composed of two paths, as illustrated in Fig. 2. The encoding path comprises layers of 'double convolution and down-sampling', while the decoding path contains layers of 'up-sampling and double convolution'. The term 'Double convolution' refers to the consecutive application of two 3×3 convolutional kernels, accompanied by a batch normalization (BN) operation and a reflected linear unit (ReLU, defined as $ReLU(x) = \max(0,x)$) activation [32]. The padding of the convolutional kernels is set to a value of 1, to ensure that the dimensions of the input and output are identical In the double convolution layers of U-net, the output of the l-th layer, denoted as x^l, is obtained from the output of the previous layer x^{l-1}. This process can be expressed as follows:

$$x_j^l = activation\left(\sum_{i=1}^{N_{l-1}} x_i^{l-1} \otimes w_{ij}^l + b_j^l\right) \quad (7)$$

where N_l indicates the number of feature maps at the l-th layer, and $j = 1,2,\ldots,N_l$ is the index accordingly. The weight map w_{ij}^l and bias map b_j^l are updated during the model training process.

Subsequently, down-sampling is performed using a 2×2 max pooling operation, which reduces the size of the feature maps while increasing the number of channels to facilitate further weight updates. Conversely, in the upsampling path, this process is reversed by applying a 2×2 deconvolutional kernel that halves the number of feature channels. In the case of feature maps with the same number of feature channels in the down-sampling and up-sampling paths, U-net employs skip-connections between corresponding layers in the downsampling and upsampling paths to preserve important features throughout the training process. Finally, a 1×1 convolutional kernel is employed to generate the output image.

U-net has been applied to tasks beyond the biomedical image segmentation domain, such as image denoising and restoration etc. [33,34]. It has consistently demonstrated exceptional performance, highlighting its ability to learn robust features and generalize well even with limited data [30,35,36]. Previously, we employed the U-net algorithm to reduce

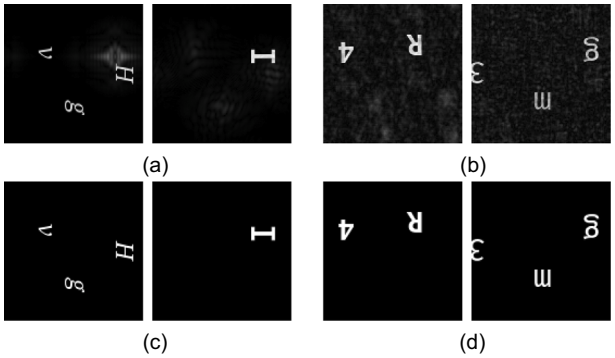


Fig. 3. Reconstruction images with (a) defocus noise and (b) speckle noise in the input dataset, (c) and (d) are the original noise-free images.

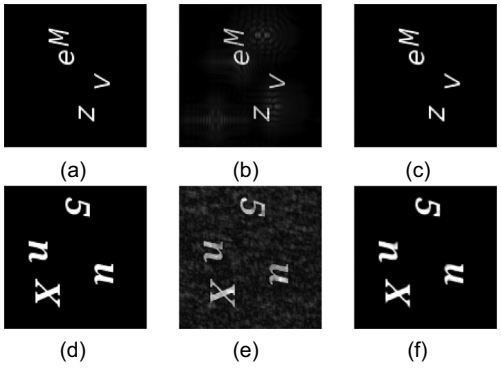


Fig. 4. Model performance tested with simulated noise images: (a) and (d) the original noise-free images; (b) image with defocus noise; (e) image with speckle noise. (c) and (f) are the recovered output images of the fine-tuned U-net.

speckle noise based on a standard OSH dataset comprising 10,422 images [37]. In this manuscript, we will further investigate the efficacy of the U-net algorithm in processing multi-type noise on a small dataset.

3. Simulation and discussion

In this section, the proposed method was demonstrated via simulation. The laser wavelength was centered at 632.8 nm, and the focal length of the lens L1 and L2 was $f = 75$ mm. Objects with two sections were used to generate the dataset. The coordinates of the two sections in z-axis (i.e. z_1 and z_2) are randomly distributed in the interval $[10, 300]$ mm. The size of each section was 130 mm $*$ 130 mm, and was sampled to 128×128 pixels.

Fig. 3(a) and (b) show some of the sample images with defocus noise as well as speckle-like noise. These noisy images are generated from the other section based on Eqs. (5) and (6), respectively. The original, noise-free images are shown in Fig. 3(c) and (d). These images constitute the dataset and are employed as the input images for the U-net model.

The training process of the deep learning models was conducted on a computer equipped with an NVIDIA RTX 4090 GPU which is allocated 24GB of graphics memory, an Intel Xeon Gold 6430 CPU, and 120GB of cache memory. This configuration was employed to accelerate the training process. The process is implemented using the PyTorch framework in particular.

The hyper parameters of the fine-tuned model are set as follows: original learning rate $lr_0 = 8e-3$, which is cosine annealed to $lr_{end} = 1e-7$. The weight decay is set to 0.09 if required, and the batch size is 16. Each model was trained for 60 epochs. The optimizer utilized is AdamW. In conjunction with this, the Charbonnier loss function can be expressed as follows:

$$Loss_{Charbonnier}(X,Y) = \sum_{i=1}^{M} \sqrt{(y_i - x_i)^2 + \epsilon^2},\ \epsilon > 0 \tag{8}$$

where M is the total number of pixels in each image. The regularization term, ϵ, is introduced to prevent the loss from becoming too small, which could result in computational errors due to the limitations of computer data storage. In order to construct the input datasets, each image containing noise was paired with a noise-free image, which served as the learning target. 1000 sets of image pairs were generated via simulation for training purposes, with an additional 300 pairs selected for evaluation. These input datasets comprise images with two different types of noise: defocus noise and speckle noise. The ratio of image pairs with them was maintained at an equal ratio in both sets, i.e. 50 % : 50 %. Following the training of the model, another 1000 simulated noise-free images are used to generate two test sets based on Eqs. (5) and (6)

Fig. 4 illustrates the denoising performance of our fine-tuned U-net model on two example images in the test sets, one with defocus noise (Fig. 4(b)) and the other with speckle noise (Fig. 4(e)). Fig. 4(a) and (d) show the original, noise-free images, while the recovered output of the trained model is displayed in Fig. 4(c) and (f), respectively. It is noteworthy that the model performs well on both the two test images with distinct noise features.

The evaluation criteria for the recovered images in this work are the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM), as defined by Eqs. (9) and (10) respectively.

$$\mathrm{PSNR}(X,Y) = 10\log\frac{(2^b-1)^2}{\mathrm{MSE}(X,Y)} \quad \text{for gray images, b} = 8 \tag{9}$$

$$\mathrm{SSIM}(X,Y) = \frac{(2u_xu_y + c_1)(2\sigma_{xy} + c_2)}{(u_x^2 + u_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \in [0,1] \tag{10}$$

where $\mathrm{MSE}(X,Y)$ is the mean square error of the 2-D matrices X and Y. The expected value of X and Y are denoted by u_x and u_y, respectively. The standard deviations of X and Y are represented by σ_x and σ_y, while the covariance of X and Y is denoted by σ_{xy}. The constants c_1 and c_2 are used to avoid zero-division errors. It is important to note that these two criteria for a dataset are calculated into the average value of all images within the dataset.

3.1. Dataset size analysis

In this subsection, we will investigate the model's performance and stability on small datasets with the proposed fine-tuned U-net model. The purpose of this simulation is to see if the output of the model remains stable when the size of the training set is changed by a certain percentage, in this case 10 % up and down at sizes of 1000, 2000 and 3000.

Fig. 5 shows the outcomes of the PSNR and SSIM metrics in relation to the training set sizes, which range from 900 to 3300. These metrics generally exhibit a positive correlation with the increase in the size of the training set. And the SSIM value has already reached a stable range of more than 0.997 in both test sets. Typically, the quality of the recovery images will be commensurate with that of the original ones when the PSNR value reaches 40 dB or above. It can be seen from Fig. 5 that the fine-tuned U-net model, trained with a mere 1000 training set, is effective in both the defocus-noise and speckle-noise test sets: the PSNR and SSIM values for the test set with defocus noise are 42.904 dB and 0.9982 Fig. 5(a); while the PSNR and the SSIM values for the speckle noise test set is 40.531 dB and 0.9975 Fig. 5(b).

Consequently, the model trained with U-net can achieve good performance with only 1000 dataset size, as mentioned above, when the proposed fine-tuning strategies are employed. A detailed analysis of the fine-tuning of optimizers and loss functions will be presented subsequently in Section 3.4.

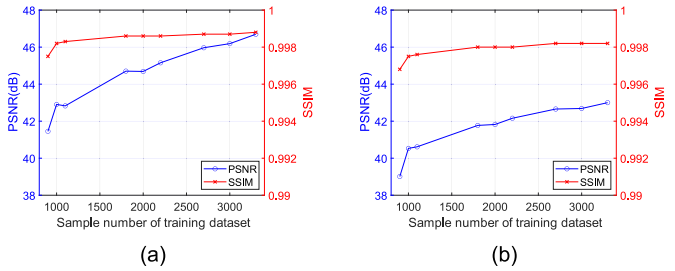


Fig. 5. Model performance tested with (a) defocus-noise test set and (b) speckle-noise test set; the *x*-axis represents the size of the training set, while the left and right *y*-axes indicate the criteria PSNR and SSIM, respectively.

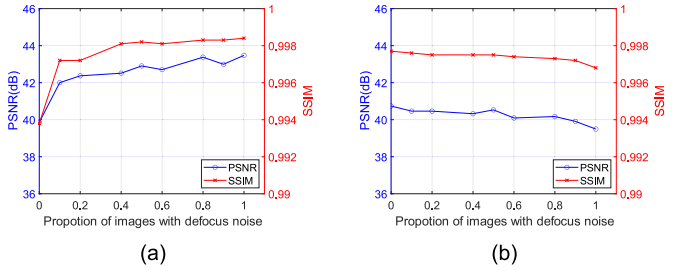


Fig. 6. Model performance tested with (a) defocus-noise test set and (b) speckle-noise test set; the *x*-axis represents the proportion of images with defocus-noise in the training dataset; the left and right *y*-axes indicate the criteria PSNR and SSIM, respectively.

3.2. Multi-type noise dataset analysis

This subsection examines the impact of the data ratio of images with defocus noise and speckle noise in the training sets. This is to investigate the stability of the fine-tuned model when varying the composition of the training data.

Fig. 6 illustrates the influence of the data ratio of two different types of noise images in the 1000 training set on the performance of the model. One can observe from this figure that as the proportion of images with defocus noise increases, the performance of the model improves on the defocus noise test set, and slightly decreases on the speckle noise test set. This is to be expected, given that the model is trained with a greater number of images containing defocus noise.

However, the variation in performance is not particularly significant when increasing the proportion of images with defocus noise in multi-type noise datasets. As shown in Fig. 6(a), the PSNR varies from 42.002 dB to 43.376 dB, while the SSIM remains constantly above 0.997 across these datasets. In contrast, when the model is trained exclusively on a full speckle noise dataset (i.e., no images exhibit defocus noise), both PSNR and SSIM values degrade to 39.895 dB and 0.9938, respectively. In Fig. 6(b), the model's performance on the speckle noise test set shows minimal changes when the proportion of images with defocus noise is below 50 %. However, as the proportion increases from 50 % to 100 %, PSNR and SSIM decrease to 39.494 dB and 0.9968, respectively. These results indicate the robustness of the proposed model, which is trained on datasets with a moderate proportion of both types of noise.

3.3. Information entropy analysis

The objective of this subsection is to investigate the outcome that the model trained with a small dataset can still achieve acceptable recovery results for images with different complexities. To this end, an information entropy (IE) analysis was conducted on various datasets [38].

First-order information entropy (denoted as IE_1) measures the unpredictability of individual pixel values, while second-order information entropy (IE_2) provides a deeper understanding of the image's texture and structure by considering the relationships between adjacent pixels. Analyzing both can be very useful for image quality assessment.

Table 1
Average value of first order and second order information entropy (IE_1 & IE_2) with different datasets.

Dataset type	IE_1	IE_2	IE_1 bias	IE_2 bias
Original_Letter	0.2420	0.4873	–	–
Defocus_Letter	2.1105	3.0233	1.8685	2.536
Speckle_Letter	5.4671	9.5064	5.2251	9.0191
Original_DSP	7.3205	11.6923	–	–
Defocus_DSP	7.3780	11.8695	0.0575	0.1772
Speckle_DSP	7.2099	11.8189	0.0206	0.1266

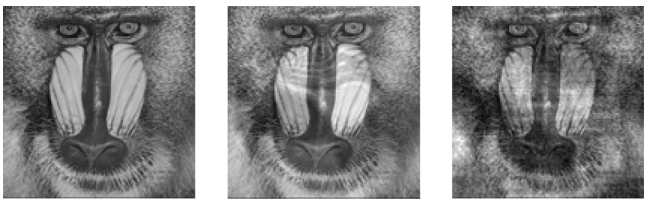


Fig. 7. Examples of standard 'DSP' images: (a) original image, reconstruction images with (b) defocus noise and (c) speckle noise.

In the simulation datasets, the average value of information entropy is calculated based on 10 images from each image type. The Average value of first order and second order information entropy (IE_1 & IE_2) with different datasets are listed in Table 1, where 'Letter' denotes the simulated images with Alphabet letters, while 'DSP' indicates standard digital signal processing images. 'Original' represents the original, noise-free images, whereas 'defocus' and 'speckle' indicate the two aforementioned types of OSH noise, respectively. Bias is defined as the difference of absolute value between noise images and clean images. (*e.g.*, the IE_1 bias of Defocus_Letter dataset equals to $abs(2.1105 - 0.2480) = 1.8685$).

The simulated images with Alphabet letters as well as the corresponding reconstruction noisy ones are previously illustrated in Fig. 3. One of the original 'DSP' images and the reconstruction images with different types of noise, are presented in Fig. 7.

As illustrated in Table 1, the IE of original images differs significantly from that of noise images for simulated images with Alphabet letters. In contrast, the IE remains relatively constant for DSP images with different noise.

The results of the corresponding denoised output of the models trained on different datasets are presented in Table 2. Each training set contains 1000 noisy images. These models are then tested on a test set with different noise data. One can observe that the models trained on 'Letter' datasets both reach PSNR and SSIM values of more than 40 dB and 0.997, respectively. On the other hand, the models trained on 'DSP' datasets performed terribly, with PSNR below 15 dB and SSIM around 0.5. In order to enable the models to achieve satisfactory recovery results, large-scale datasets are required for cases with closer complexity between clean images and noise images. As previously demonstrated in our research [37], approximately 10,920 training samples were required to achieve a PSNR and SSIM score that was barely acceptable.

This outcome aligns with the functionality and characteristics of the U-net architecture. It is challenging to achieve satisfactory denoising results if there is a closed IE bias between the original images and the training datasets.

In conclusion, for noisy datasets with large bias of IE, it is possible to achieve good results with a smaller number of samples. The bias index of IE thus provides a reference point (*e.g.*, the IE bias is greater than 1.8

Table 2
Denoising results of the models trained on different datasets.

Dataset type	Reconstrcution result	
	PSNR	SSIM
Defocus_Letter	43.518	0.9994
Speckle_Letter	40.779	0.9987
Defocus_DSP	14.510	0.6123
Speckle_DSP	13.330	0.4974

in our case) for the learning of small datasets, which can be applied to fields where the collection of information is challenging, such as medical bio-image processing.

3.4. Optimizer and loss function analysis

In this subsection, we will analyze the performance of the proposed fine-tuned U-net model with different optimizers and loss functions.

Optimizer, as a crucial component of the deep learning training pipeline, employs specific optimization algorithms, such as Stochastic Gradient Descent (SGD), Adaptive moment estimation (Adam), and others, to determine the direction and magnitude of parameter updates during training [39]. SGD is a classic optimization strategy that can maintain a reliable optimization direction by introducing the concept of momentum. Adam is a strategy that adapts the learning rates for each parameter based on the first and second moments of the gradients, while AdamW (Adam with weight decay) is an extension of the Adam optimization strategy that incorporates weight decay directly into the parameter update rule. Weight decay is a regularization technique used to prevent overfitting by penalizing large weights, effectively encouraging simpler models.

The loss function, as another fundamental component in machine learning, defines the training objective for a learning model. There is a wide variety of loss functions that can be used in the learning process, including Mean Squared Error (MSE) loss, L1 loss, Smooth L1 loss, Charbonnier loss, *etc*. The MSE and the L1 loss are loss functions that essentially calculate the norms between the denoised output and the original image [40]. The Smooth L1 loss is a combination of MSE and L1 at different ranges of the arguments. The Charbonnier loss can be defined as the L1 loss with a regularization term. This loss function can converge to the optimal model more rapidly during the training process [41]. The selection of an appropriate loss function for a given task is of crucial importance for the model's performance and generalization capability.

A comparative study of model performance with different combinations of optimizers and loss functions was conducted, and the results are listed in Table 3. In the simulation, a dataset with size of only 1000 and an equal ratio of two types of noise images are employed for training. And the two formerly generated test sets are used to analyse different models, as listed in Table 3. 'Test Set 1' refers to the test dataset with defocus noise, while 'Test Set 2' is the test dataset with speckle noise. Charbon. stands for the Charbonnier loss function. We selected three optimizers (SGD, Adam, and AdamW) and four loss functions (MSE, L1, Smooth L1 and Charbonnier) to identify the optimal combination. It is notable that the combination of the AdamW optimizer and the Charbonnier loss function shows the best performance in terms of the PSNR value of the test set 1 and both two criteria of the test set 2. Furthermore, in the context of SSIM in the test set 1, the value reaches 0.9982.

The results indicated that the combination of AdamW and Charbonnier Loss is the most effective with the proposed fine-tuning hyper parameters.

3.5. Comparison of other learning-based network architectures

To further explore the potential for enhanced performance on small datasets with close IEs, as discussed in Section 3.3, this subsection presents an analysis conducted on datasets featuring different IEs. We employed two additional distinct learning-based network architectures: DnCNN [42] and Uformer [34].

Table 3
A comparison of the performance of models trained with different optimizers and loss functions. The bold numbers represent the best performance under all listed conditions.

Optimizer	Loss func.	Test set1		Test set2	
		PSNR	SSIM	PSNR	SSIM
SGD	$Charbon.$	24.194	0.8946	24.377	0.7721
$Adam$	$Charbon.$	41.071	0.9955	30.253	0.9971
$AdamW$	$Charbon.$	**42.904**	0.9982	**40.531**	**0.9975**
$AdamW$	$L1$	42.518	**0.9983**	40.326	**0.9975**
$AdamW$	$Smooth\ L1$	41.864	0.9962	39.798	0.9946
$AdamW$	$MSE(L2)$	41.712	0.9899	39.529	0.9897

Table 4
A comparison of the performance of models trained with different learning-based network architectures. The bold numbers represent the best performance among all listed conditions.

Test sets	Model					
(Dataset size: 1000)	U-net		DnCNN		Uformer	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Defocus_Letter	43.518	0.9994	**44.945**	**0.9996**	37.576	0.9718
Speckle_Letter	40.779	0.9987	**41.087**	**0.9989**	24.173	0.9625
Defocus_DSP	**14.510**	**0.6123**	14.464	0.5165	13.395	0.3965
Speckle_DSP	**13.330**	**0.4974**	12.387	0.4310	11.597	0.2072

DnCNN is a CNN model introduced by K. Zhang and colleagues in 2017. Like U-Net, it is effective for noise suppression tasks. The architecture employs residual learning and batch normalization techniques, which reduce the input data complexity. This allows for the use of larger datasets even with limited computational resources. Additionally, DnCNN features a deeper architecture, accommodating more layers than most standard CNNs.

Uformer was first presented in 2022 by Wang et al. and is designed for image restoration tasks, including those involving motion blur and defocus blur. This model combines U-net and Transformer architecture [43], which has been widely used in large language models (LLMs). The name "Uformer" reflects its U-shaped Transformer design. Within the network architecture of Uformer, a novel locally-enhanced window (LeWin) Transformer block and a learnable multi-scale restoration modulator are utilized. These components introduce a self-attention mechanism and adjust features across multiple layers of the decoder. However, Uformer typically requires hundreds of millions of images as input data due to its Transformer-based construction. Consequently, it can be reasonably assumed that the performance of Uformer on a small dataset would not meet expectations, a point that will be further illustrated in our subsequent analysis.

A comparative study of model performance with various learning-based networks was conducted, with results presented in Table 4. Each training set contains 1000 noisy images. Under the same training parameters, DnCNN shows the most optimal performance in the simulated datasets comprising Alphabet letters, that are with distinct IEs. Models trained with U-net outperform the other two networks on the Standard DSP datasets, which are with close IEs, while they perform slightly lower than DnCNN in the datasets with Alphabet letters. As anticipated, Uformer performs the worst on such small dataset task among the three networks.

This outcome indicates that, for small datasets with close IEs, existing deep learning methods may struggle to achieve high-quality results in noise suppression tasks. However, it may be feasible to utilize techniques such as transfer learning, leveraging a large model trained on billions of images, to obtain acceptable results in future applications [44].

4. Experimental data analysis

The proposed method has also been experimentally verified. In the OSH experimental setup, a He-Ne laser with a wavelength centered at $\lambda = 632.8$ nm is used to generate coherent light. The diameter of the collimated beam is $D = 25$ mm, and the focal length of the lens is $f = 500$ mm. A 3D object is employed with two slides located at $z_1 = 87$ cm and $z_2 = 107$ cm, respectively. In this experimental setup, we focused exclusively on images that exhibited defocus noise. Therefore, the choice of the pupils was set as follows: $p_1(x, y) = 1$, $p_2(x, y) = \delta(x, y)$ in the hologram recording process, and $p_{1r}(x, y) = \delta(x, y)$, $p_{2r}(x, y) = 1$ in the reconstruction process.

The hologram was sampled to a size of 500×500 pixels and resized to 128×128 pixels to fit the model training input. Fig. 8(a) illustrates the structure of the 3D object. The real part and the imaginary part of the recorded hologram are shown in Fig. 8(b) and (c), respectively.

The original sectional images of the object with two slides are shown in Fig. 9(a). And the recovery sectional images with traditional reconstruction method as well as the inverse-image method are shown in Fig. 9(b) and (c), respectively. Fig. 9(d) presents the output of our fine-tuned U-net model, which was trained with only the 1000 dataset and an equal ratio of the two types of noise. It is evident that the U-net model has the most optimal performance in our experimental data. The performance comparison of different reconstruction methods on experimental data is listed in Table 5, where 'S'&'H' represent the two sections of the object, respectively. The results of the numerical analysis also demonstrated the best performance of the U-net model.

The poor reconstruction quality observed with the conventional OSH method can be attributed to its reliance on the conjugate function of the optical transfer function (OTF) for image recovery. This approach inherently leads to the introduction of defocus noise, as contributions from unfocused sections are reflected in the reconstructed images. Consequently, the presence of this noise significantly degrades the overall image quality, making it challenging to accurately analyze the details of the object being imaged.

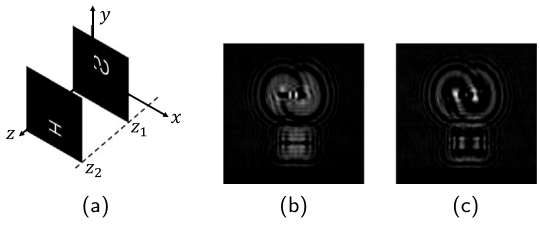


Fig. 8. (a) Object with two slides; (b) the real part and (c) the imaginary part of the recorded hologram.

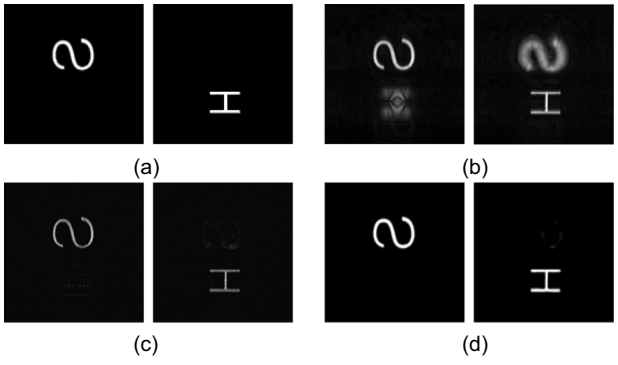


Fig. 9. (a) The original sectional images of the 3D object; (b) the sectional images with defocus noise recovered from the experiment with the classic OSH method; (c) the inverse-image method and (d) the U-net trained with our proposed method.

Table 5
Performance comparison of different reconstruction methods on experimental data. The bold numbers represent the best performance under all listed conditions.

Size of training data	Section 'S'		Section 'H'	
	PSNR	SSIM	PSNR	SSIM
Classic OSH method	22.714	0.3370	17.395	0.2337
Inverse-image method	23.449	0.2311	23.343	0.2030
U-net method(ideal)	**36.170**	**0.9971**	27.903	0.9834
U-net method (with Gaussian envelope)	33.254	0.9937	**31.215**	**0.9922**

Table 6
Performance comparison of different denoising network architectures on experimental data. The bold numbers represent the best performance among all listed conditions.

Network architecture	Section 'S'		Section 'H'	
	PSNR	SSIM	PSNR	SSIM
U-net	**33.254**	**0.9937**	**31.215**	**0.9922**
DnCNN	27.556	0.9839	22.597	0.9619
Uformer	17.718	0.4944	19.409	0.4885

Similarly, the inverse-image method demonstrates limitations, primarily due to its neglect of the imaginary component of defocus noise. It relies on a Conjugate Gradient (CG) process to approximate the inverse imaging solution. As a result, both the conventional OSH method and the inverse-image method experience reconstruction quality issues stemming from retained or unaccounted defocus noise. These challenges highlight the inherent limitations of each approach.

Additionally, it is noteworthy that the recovery pattern of the section with letter 'S' exhibits higher denoising criteria than the one with letter 'H'. This is due to the fact that the section with letter 'H' is located at a greater depth along the z-axis, which results in a higher level of Gaussian envelope attenuation. In the simulation process, we have simulated an ideal OSH condition, simplifying some detailed parameters such as the numerical aperture (NA), Gaussian beam properties, and others that will be differ from the specific experimental setup. Consequently, the inferior recovery quality of section 'H' relative to section 'S' is to be expected.

As a supplementary analysis, we generated an additional simulated dataset based on the actual experimental configuration and parameters. The U-net model trained on this dataset, which incorporated Gaussian envelope attenuation, demonstrated the expected performance in the reconstructions of both sections. This improvement is illustrated in the last row of Table 5. One can observe from this result that the model trained with dataset, involving the Gaussian envelope attenuation, performs much better in section 'H' than in the ideal situation. Additionally, the performance in section 'S' remains at a favorable denoising result, with a PSNR of 33.254 dB and SSIM of 0.9937.

Moreover, a comparison of the performance of different networks on experimental data is presented in Table 6. All models were trained on a dataset incorporating more realistic parameters, including Gaussian envelope attenuation. The U-net model shows the best performance among the three. This outcome highlights the universality of U-net in handling data captured in complex experimental scenarios.

Fig. 10 shows the performance of the U-net trained with different dataset sizes, tested on experimental images. The training sets in Fig. 10(a) and (b) are of sizes $2k$ and $3k$, respectively. And the corresponding numerical results are listed in Table 7. It's notable in Fig. 10 that the denoising results of section 'S' indicate a good performance on all listed sizes of training sets. This can be demonstrated in Table 7 that the criterion of sizes from 1k to 3k all exhibited a PSNR above 35 dB and a SSIM over 0.997 in section 'S'. It is as expected according to the previous simulation result in Fig. 5(a). While in the section 'H', the

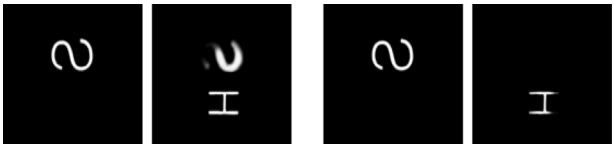


Fig. 10. The performance of U-net tested on experimental images with different dataset sizes: (a) 2k; (b) 3k.

Table 7
The performance of U-net trained with different training set sizes on experimental data. The bold numbers represent the best performance under all listed conditions.

Size of training data	Section 'S'		Section 'H'	
	PSNR	SSIM	PSNR	SSIM
1k	36.170	0.9971	**27.903**	**0.9834**
2k	37.434	**0.9976**	17.346	0.9248
3k	**37.775**	**0.9976**	25.681	0.9766

Table 8
The performance of U-net trained with different proportions of noisy images on experimental data. The bold numbers represent the best performance under all listed conditions.

Proportion of images-with defocus noise	Section 'S'		Section 'H'	
	PSNR	SSIM	PSNR	SSIM
10 %	**37.352**	**0.9971**	16.266	0.8911
20 %	37.178	**0.9971**	16.047	0.8896
40 %	36.939	**0.9971**	27.637	0.9756
50 %	36.170	**0.9971**	**27.903**	**0.9834**
80 %	35.248	0.9951	25.986	0.9766
90 %	34.341	0.9956	26.321	0.9766

Fig. 11. The performance of U-net trained on datasets with proportions of (a) 40 %; and (b) 90 % images with defocus noise, respectively

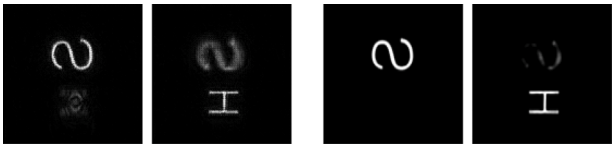


Fig. 12. The performance of U-net trained with different combinations of optimizers and loss functions tested on experimental data: the model trained with (a) SGD optimizer and Charbonnier loss function; and (b) AdamW optimizer and L1 loss function.

denoising results of the models trained on the training sets with sizes of $2k$ and $3k$ exhibited underfitting and overfitting. The model trained with our proposed method on only 1000 size of training set shows the best performance. This discrepancy between the simulation outcome and the experimental data on Section 'H' can be attributed to the simplification of the physical parameters in the OSH simulation.

Table 9
The performance of U-net trained with different optimizers and loss functions on experimental data. The bold numbers represent the best performance under all listed conditions.

Optimizer	Loss func.	Section 'S'		Section 'H'	
		PSNR	SSIM	PSNR	SSIM
SGD	$Charbon.$	23.187	0.8984	22.300	0.8389
$Adam$	$Charbon.$	35.729	0.9966	15.988	0.8838
$AdamW$	$Charbon.$	36.170	**0.9971**	**27.903**	**0.9834**
$AdamW$	$L1$	36.923	**0.9971**	27.745	**0.9834**
$AdamW$	$Smooth\ L1$	37.422	0.9951	27.686	0.9805
$AdamW$	$MSE(L2)$	**37.809**	0.9956	27.714	0.9824

The numerical results of U-net tested on experimental data with different proportions of noisy images in the training datasets are listed in Table 8. It can be observed that almost all the models, trained with a dataset containing images with defocus noise, perform well on the recovery of section 'S'. However, the recovery of section 'H' exhibited underfitting when the proportion of images with defocus noise was at a lower level, but overfitting at a higher level. This is illustrated in Fig. 11(a) and (b), respectively. In consideration of the aforementioned simplification of the physical parameters, it is reasonable to conclude that a certain amount of images with defocus noise are necessary for U-net to achieve a good performance.

Furthermore, the performance of different combinations of optimizer and loss function is also tested with the experimental images. Fig. 12 illustrates the performance of two models trained with combinations of (a) SGD optimizer with Charbonnier loss function and (b) AdamW optimizer with L1 loss function. The training set comprises an only 1000 size dataset with an equal ratio of the two noise types. As listed in Table 9, the PSNR values of most of the combinations are all above 35 dB in Section 'S'. With regard to the other criterions, the combination of the AdamW optimizer and the Charbonnier loss function demonstrates the most optimal performance on the defocus experimental data. This outcome aligns with the simulation result in Section 3.4.

5. Conclusion

The removal of undesired noise in sectional image reconstruction in DH is a longstanding challenge that has been extensively studied. This work demonstrated the capability of U-net in handling multi-type OSH noise images with small training datasets. Information entropy analysis as well as simulation results indicate that the proposed method performs well on images with a certain level of IE bias in a small-size dataset. This outcome serves as a reference point for deep learning tasks in domains where data acquisition is challenging, such as medical bio-image processing and optical microscopy. Moreover, the experimental analysis demonstrated the generalization ability of the U-net model trained with our method.

It is believed that the proposed method can achieve better performance if trained with actual experimental datasets, which is feasible due to its effectiveness with small training sizes.

Additionally, it is worth noting that the test samples used in our experiments may not fully represent the complexity of real-world applications, such as medical or microscopic imaging. Future studies will benefit from collaborations with biomedical imaging experts to incorporate more biologically relevant specimens.

Moreover, the application of the proposed method to phase-only holograms remains an intriguing area that requires further exploration.

CRediT authorship contribution statement

Ruiwei Xu: Writing – review & editing, Writing – original draft, Methodology, Conceptualization. **Haiyan Ou:** Writing – review & editing, Conceptualization. **Edmund Y. Lam:** Writing - reviewing & editing.

Data and code availability

The data and code supporting the findings of this study are available at https://github.com/CrackCzi/Defocus-and-speckle-noise-suppression-in-optical-scanning-holography

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used *DeepL* in order to polish the article to get a higher readability. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

I have shared the link to my data at the end of the article.

References

[1] J. Swoger, M. Martınez-Corral, J. Huisken, E.H. Stelzer, Optical scanning holography as a technique for high-resolution three-dimensional biological microscopy, J. Opt. Soc. Am. A Opt. Image Sci. Vis. 19 (2002) 1910–1918, https://doi.org/10.1364/JOSAA.19.001910

[2] Y. Zhang, T.-C. Poon, P.W. Tsang, R. Wang, L. Wang, Review on feature extraction for 3-D incoherent image processing using optical scanning holography, IEEE Trans. Ind. Inf. 15 (2019) 6146–6154, https://doi.org/10.1109/TII.2019.2938806

[3] B.W. Schilling, G.C. Templeton, Three-dimensional remote sensing by optical scanning holography, Appl. Opt. 40 (2001) 5474–5481, https://doi.org/10.1364/AO.40.005474

[4] A. Essadike, E. Ouabida, A. Bouzid, Optical scanning holography for tumor extraction from brain magnetic resonance images, Opt. Laser Technol. 127 (2020) 106158, https://doi.org/10.1016/j.optlastec.2020.106158

[5] H. Ou, Y. Wu, E.Y. Lam, B.-Z. Wang, New autofocus and reconstruction method based on a connected domain, Optics Lett. 43 (2018) 2201–2203, https://doi.org/10.1364/OL.43.002201

[6] H. Ou, Y. Wu, E.Y. Lam, B.-Z. Wang, Axial localization using time reversal multiple signal classification in optical scanning holography, Opt. Express 26 (2018) 3756–3771, https://doi.org/10.1364/OE.26.003756

[7] Z. Ren, N. Chen, E.Y. Lam, Automatic focusing for multisectional objects in digital holography using the structure tensor, Opt. Lett. 42 (2017) 1720–1723, https://doi.org/10.1364/OL.42.001720

[8] J. Ke, T.-C. Poon, E.Y. Lam, Depth resolution enhancement in optical scanning holography with a dual-wavelength laser source, Appl. Opt. 50 (2011) H285–H296, https://doi.org/10.1364/AO.50.00H285

[9] H. Ou, T.-C. Poon, K.K. Wong, E.Y. Lam, Enhanced depth resolution in optical scanning holography using a configurable pupil, Photonics Res. 2 (2014) 64–70, https://doi.org/10.1364/PRJ.2.000064

[10] F. Zhao, X. Qu, X. Zhang, T.-C. Poon, T. Kim, Y.S. Kim, J. Liang, Solving inverse problems for optical scanning holography using an adaptively iterative shrinkage-thresholding algorithm, Opt. Express 20 (2012) 5942–5954, https://doi.org/10.1364/OE.20.005942

[11] X. Zhang, E.Y. Lam, T.-C. Poon, Reconstruction of sectional images in holography using inverse imaging, Opt. Express 16 (2008) 17215–17226, https://doi.org/10.1364/OE.16.017215

[12] L. Jin-Xi, Z. Ding-Fu, Y. Sheng, Z. Yuan-Yuan, Z. Luo-Zhi, H. Dong-Ming, Z. Xin, Modified image fusion technique to remove defocus noise in optical scanning holography, Opt. Commun. 407 (2018) 234–238, https://doi.org/10.1016/j.optcom.2017.08.057

[13] Z. Xin, K. Dobson, Y. Shinoda, T.-C. Poon, Sectional image reconstruction in optical scanning holography using a random-phase pupil, Opt. Lett. 35 (2010) 2934–2936, https://doi.org/10.1364/OL.35.002934

[14] H. Ou, H. Pan, E.Y. Lam, B.-Z. Wang, Defocus noise suppression with combined frame difference and connected component methods in optical scanning holography, Opt. Lett. 40 (2015) 4146–4149, https://doi.org/10.1364/OL.40.004146

[15] D.W. Otter, J.R. Medina, J.K. Kalita, A survey of the usages of deep learning for natural language processing, IEEE Trans. Neural Netw. Learn. Syst. 32 (2020) 604–624, https://doi.org/10.1109/TNNLS.2020.2979670

[16] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, M.S. Lew, Deep learning for visual understanding: a review, Neurocomputing 187 (2016) 27–48, https://doi.org/10.1016/j.neucom.2015.09.116

[17] H. Purwins, B. Li, T. Virtanen, J. Schlüter, S.-Y. Chang, T. Sainath, Deep learning for audio signal processing, IEEE J Sel Top Signal Process 13 (2019) 206–219, https://doi.org/10.1109/JSTSP.2019.2908700

[18] Z. Ren, Z. Xu, E.Y. Lam, Learning-based nonparametric autofocusing for digital holography, Optica 5 (2018) 337–344, https://doi.org/10.1364/OPTICA.5.000337

[19] Y. Rivenson, Y. Zhang, H. Günaydın, D. Teng, A. Ozcan, Phase recovery and holographic image reconstruction using deep learning in neural networks, Light Sci. Appl. 7 (2018) 17141–17141. https://doi.org/10.1038/lsa.2017.141

[20] S.-J. Kim, C. Wang, B. Zhao, H. Im, J. Min, H.J. Choi, J. Tadros, et al., Deep transfer learning-based hologram classification for molecular diagnostics, Sci. Rep. 8 (2018) 17003, https://doi.org/10.1038/s41598-018-35274-x

[21] Z. Ren, H.K.-H. So, E.Y. Lam, Fringe pattern improvement and super-resolution using deep learning in digital holography, IEEE Trans. Ind. Inf. 15 (2019) 6179–6186, https://doi.org/10.1109/TII.2019.2913853

[22] T. Liu, Z. Wei, Y. Rivenson, K. de Haan, Y. Zhang, Y. Wu, A. Ozcan, Deep learning-based color holographic microscopy, J. Biophoton. 12 (2019) e201900107, https://doi.org/10.1002/jbio.201900107

[23] Q. Fang, H. Xia, Q. Song, M. Zhang, R. Guo, et al., Speckle denoising based on deep learning via a conditional generative adversarial network in digital holographic interferometry, Opt. Express 30 (2022) 20666–20683, https://doi.org/10.1364/OE.459213

[24] J. Tang, B. Chen, L. Yan, L. Huang, Continuous phase denoising via deep learning based on perlin noise similarity in digital holographic microscopy, IEEE Trans. Ind. Inf. 20 (6) (2024) 8707–8716, https://doi.org/10.1109/TII.2024.3375375

[25] X. Zhuang, A. Yan, P.W.M. Tsang, T.-C. Poon, Deep-learning based reconstruction in optical scanning holography, Opt. Lasers Eng. 158 (2022) 107161, https://doi.org/10.1016/j.optlaseng.2022.107161

[26] Z. Luo, A. Yurt, R. Stahl, A. Lambrechts, V. Reumers, D. Braeken, L. Lagae, Pixel super-resolution for lens-free holographic microscopy using deep learning neural networks, Opt. Express 27 (2019) 13581–13595, https://doi.org/10.1364/OE.27.013581

[27] S. Park, Y. Kim, I. Moon, Automated phase reconstruction and super-resolution with deep learning in digital holography, Opt. Laser Technol. 176 (2024) 111030, https://doi.org/j.optlastec.2024.111030

[28] S. Lee, S.-W. Nam, J. Lee, Y. Jeong, B. Lee, HoloSR: deep learning-based super-resolution for real-time high-resolution computer-generated holograms, Opt. Express 32 (2024) 11107–11122, https://doi.org/10.1364/OE.516564

[29] T.-C. Poon, Optical Scanning Holography with MATLAB, Springer, 2007.

[30] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, in: Medical image computing and computer-assisted intervention 2015 international conference, Springer, 2015, pp. 234–241, https://doi.org/10.1007/978-3-319-24574-4_28

[31] E. Shelhamer, J. Long, T. Darrell, Fully convolutional networks for semantic segmentation, arXiv E-Prints 1605.06211, 2016, https://doi.org/10.48550/arXiv.1605.06211

[32] N. Siddique, S. Paheding, C.P. Elkin, V. Devabhaktuni, U-net and its variants for medical image segmentation: a review of theory and applications, IEEE. Access 9 (2021) 82031–82057, https://doi.org/10.1109/ACCESS.2021.3086020

[33] J. Gurrola-Ramos, O. Dalmau, T.E. Alarcón, A residual dense U-net neural network for image denoising, IEEE Access 9 (2021) 31742–31754, https://doi.org/10.1109/ACCESS.2021.3061062

[34] Z. Wang, X. Cun, J. Bao, W. Zhou, J. Liu, H. Li, Uformer: a general u-shaped transformer for image restoration, arXiv preprint arXiv:2106.03106, 2021, https://doi.org/10.48550/arXiv.2106.03106

[35] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, et al., Transunet: transformers make strong encoders for medical image segmentation, arXiv preprint arXiv:2102.04306 (2021) https://doi.org/10.48550/arXiv.2102.04306

[36] X.-X. Yin, L. Sun, Y. Fu, R. Lu, Y. Zhang, [Retracted] u-net-based medical image segmentation, J. Healthc. Eng. 2022 (2022) 4189781, https://doi.org/10.1155/2022/4189781

[37] H. Ou, Y. Wu, K. Zhu, E.Y. Lam, B.-Z. Wang, Suppressing defocus noise with u-net in optical scanning holography, Chin. Optics Lett. 21 (8) (2023) 080501, https://doi.org/10.3788/COL202321.080501

[38] R.M. Gray, Entropy and Information Theory, Springer Science & Business Media, 2011.

[39] D. Soydaner, A comparison of optimization algorithms for deep learning, Intern. J. Pattern. Recognit. Artif. Intell. 34 (13) (2020) 2052013, https://doi.org/10.1142/S0218001420520138

[40] H. Zhao, O. Gallo, I. Frosio, J. Kautz, Loss functions for image restoration with neural networks, IEEE. Trans. Comput. Imaging. 3 (1) (2016) 47–57, https://doi.org/10.1109/TCI.2016.2644865

[41] B. Gajera, S.R. Kapil, D. Ziaei, J. Mangalagiri, E. Siegel, D. Chapman, Ct-scan denoising using a charbonnier loss generative adversarial network, IEEE. Access 9 (2021) 84093–84109, https://doi.org/10.1109/ACCESS.2021.3087424

[42] K. Zhang, W. Zuo, Y. Chen, D. Meng, L. Zhang, Beyond a gaussian denoiser: residual learning of deep cnn for image denoising, IEEE Trans. Image Process 26 (7) (2017) 3142–3155, https://doi.org/10.1109/TIP.2017.2662206

[43] A. Vaswani, Attention is all you need, Adv. Neural Inf. Process. Syst. (2017) https://doi.org/10.48550/arXiv.1706.03762

[44] H.E. Kim, A. Cosa-Linan, N. Santhanam, M. Jannesari, M.E. Maros, T. Ganslandt, Transfer learning for medical image classification: a literature review, BMC Med. Imaging 22 (1) (2022) 69, https://doi.org/10.1186/s12880-022-00793-7