Deep Learning Based Coherence Holography (DCH) Reconstruction of 3D Objects

Quang Trieu and George Nehmetallah
Departments of Electrical Engineering and Computer Science,, The Catholic University of America, 620 Michigan Avenue N.E., Washington, D.C. 20064
trieu@cua.edu,nehmetallah@cua.edu

Abstract

We propose a novel reconstruction method for coherence holography using deep neural networks. cGAN and Unet models were developed to reconstruct 3D complex objects from recorded interferograms. Our proposed methods dubbed deep coherence holography (DCH) predict the non-diffracted fields or the sub-objects included in the 3D object from the captured interferograms yielding better reconstructed objects than the traditional analytical imaging methods in terms of accuracy, resolution, and time. The DCH needs one image per sub-object as opposed to N images for the traditional sin-fit algorithm and hence the total reconstruction time is reduced by $N\times$. Furthermore, with noisy interferograms the DCH amplitude mean square reconstruction error (MSE) is $5\times10^4\times$ and $10^4\times$ and phase MSE is 100× and 3000× better than Fourier fringe and sin-fit algorithms, respectively. The amplitude peak signal to noise ratio (PSNR) is 3× and 2× and phase PSNR is 5× and 3× better than Fourier fringe and sin-fit algorithms, respectively. The reconstruction resolution is same as sin-fit but 2× better than the Fourier fringe analysis technique.

1. Introduction

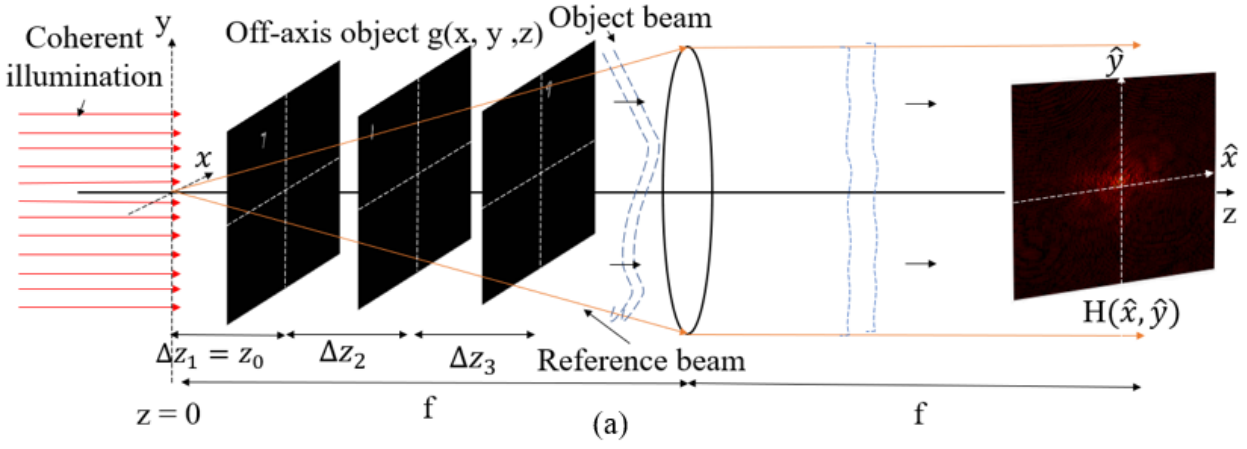
Incoherently illuminating a hologram generates an optical field having its phase depending on an instantaneous random phase instead of generating a desired object field. Fortunately, the mutual intensity between a pair of points or the coherence function obtained from this optical field can provide a time independent field which gets rid of the instantaneous random phase resulting from incoherent illumination. Due to this advantage, the mutual intensity has been applied in coherence holography besides being applied in other applications such as metrology [1, 2]. Coherence holography (CH) is a novel unconventional holographic method that is capable of reconstructing an object as the 3D distribution of a spatial coherence function (CF) [3, 4]. Detecting this spatial CF requires a suitable interferometer and an appropriate analysis technique to quantify the contrast and the phase of the interference fringe pattern generated by the interferometer. One of these interferometers can be a common-path imaging Sagnac radial shearing interferometer (SI) which is a robust way to correlate optical fields to detect the 3D spatial CF representing the object recorded in the coherence hologram. One way to quantify the CF is to introduce a tilt between the wave fronts of the interfering beams. This tilt generates a carrier spatial frequency in the Sagnac interferogram and thus we can apply the Fourier Transform (FT) method of fringe analysis to reconstruct the object field at the axial plane of interest as a complex CF using only a single interferogram [3, 4]. However, the resolution of the reconstructed image is limited by the finite frequency pass-band of the filtered carrier fringe pattern. In addition, in practice, the object is space-limited so its spectrum is unlimited in frequency domain, and thus applying Fourier fringe analysis can only obtain a part of the object's spectrum containing the information of the object. This results in an inaccurate reconstruction. To overcome these limitations, phase-shift CH was proposed, in which the 4-parameter sine wave fit algorithm is applied to a set of Sagnac interferograms generated by a set of phase-shifted computer-generated FT holograms to

reconstruct the object field at the axial plane of interest, rather than the spatial CF [5, 6]. This method however is slower than FT method of fringe analysis since it requires more interferograms to reconstruct the object. Furthermore, both methods are subject to noise such as dark noise and shot noise that are inevitable when capturing images by a CCD camera. Fundamentally, it is almost impossible to reconstruct an object under ideal conditions. Therefore, their recovery performance is influenced by external factors, some of which are inevitable.

Recently, with the significant improvement of computational power, deep learning (DL) has shown great potential in computational imaging. With powerful data mining and mapping abilities, the data-informed DL methods is able to extract the key features and build a reliable model. For that reason, the DL approach has been successfully applied in holographic display [7,8], super-resolution microscopic imaging [9-12], and imaging through scattering media [13-17], aberration compensation [32], computational multi-wavelength (MW) phase synthesis [33], and computational optical tomography [34]. Not only does DL solve complex imaging problems, but DL can also significantly improve the core performance and reduce the computational time [7, 8]. Hence, in this work we developed a DL neural network based method, dubbed deep coherence holography (DCH) to reconstruct the 3D object from captured interferograms, yielding a better result than the above methods in accuracy and time.

2. Principle of coherence holography

The principle of coherence holography (CH) has been described in detail in [18]. Briefly, due to the formal similarity between the diffraction integral and Van Cittert-Zernike theorem [19,20], when the light field from the hologram illuminated by a spatially incoherent light source is directed into the shearing interferometer (SI), the coherence function (CF) or the mutual intensity between a pair of points separated by a length $\delta\boldsymbol{r} = \boldsymbol{r_2} - \boldsymbol{r_1}$ on the image plane is proportional to the optical field reconstructed at a point $\boldsymbol{r} = m\delta\boldsymbol{r}$ by conventional holography, where m is the scale parameter depending on the setup of the SI. This is demonstrated in Figure 1(b). In the first step, the simulated computer-generated hologram (CGH) is generated based on the concept of Fourier transform coherence holography as shown in Figure 1(a).



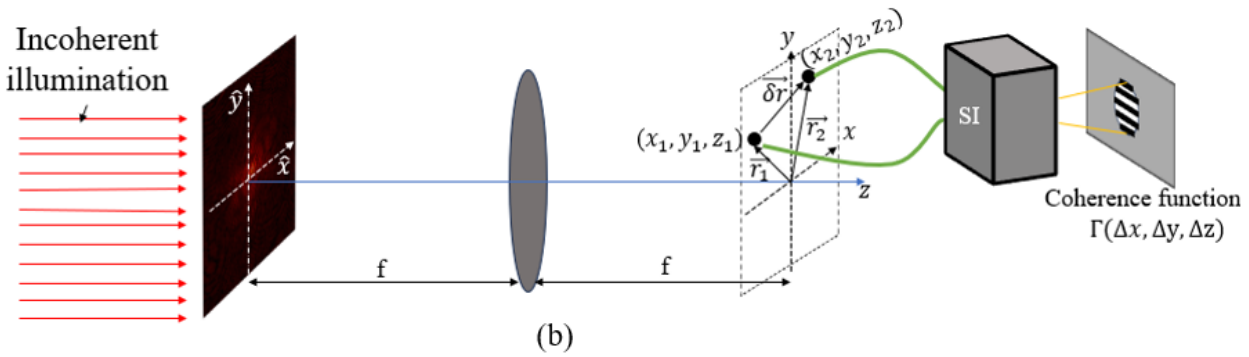


Figure 1. (a) Generation of a Fourier transform coherence hologram; (b) Generation of coherence function.

The complex amplitude of an off-axis 3D object $g(x,y,z) = |g(x,y,z)|\exp[i\phi(x,y,z)]$ located at a distance z_0 away from the front focal plane of the Fourier transform lens at the hologram plane is given by

$$\begin{aligned} G(\hat{x},\hat{y}) &= |G(\hat{x},\hat{y})|\exp[i\Phi(\hat{x},\hat{y})] \\ &\propto \int \left\{ \iint g(x,y,z)\exp\left[-i\frac{2\pi}{\lambda f}(x\hat{x}+y\hat{y})\right]dxdy\right\}\exp[-ik_z(\hat{x},\hat{y})z]\,dz, \\ &\propto \int G_z(\hat{x},\hat{y},z)\exp[-ik_z(\hat{x},\hat{y})z]\,dz, \end{aligned} \tag{1}$$

where $G_z(\hat{x},\hat{y},z) = F\{g(x,y,z)\}|_{k_x=\frac{2\pi}{\lambda f}\hat{x},k_y=\frac{2\pi}{\lambda f}\hat{y}}$, $F\{\cdot\}|_{k_x=\frac{2\pi}{\lambda f}\hat{x},k_y=\frac{2\pi}{\lambda f}\hat{y}}$, denotes the 2D Fourier transform operation and will be denoted as $F\{\cdot\}$ for convenience in this paper. λ is the wavelength of light, and f is the focal length of the lens. The term $G_z(\hat{x},\hat{y},z)$-represents the angular spectrum of the object field distribution across the plane z = z with the spatial frequencies represented by the coordinates $\hat{x}$ and $\hat{y}$. The $\exp[-ik_z(\hat{x},\hat{y})z]$ term accounts for defocusing, and propagating the angular spectrum of the field by a distance z with $k_z(\hat{x},\hat{y}) = \frac{2\pi}{\lambda}\sqrt{1-\left(\frac{\hat{x}}{f}\right)^2-\left(\frac{\hat{y}}{f}\right)^2}$.

To create the final CGH, the term $|G(\hat{x},\hat{y})|^2$ was removed from the interference fringe intensity to avoid unwanted autocorrelation image, and the intensity (rather than amplitude) transmittance of the hologram was made proportional to the interference fringe pattern [6]. The equation of the CGH is then given by

$$H(\hat{x},\hat{y}) \propto |G(\hat{x},\hat{y})| + 0.5[G(\hat{x},\hat{y}) + G^*(\hat{x},\hat{y})] = |G(\hat{x},\hat{y})|\{1+\cos[\Phi(x,y,z)]\}. \tag{2}$$

The term $|G(\hat{x},\hat{y})|$ is added to make $H(\hat{x},\hat{y})$ positive as shown in the equation above.

In the second step, the simulated CGH is illuminated with spatially incoherent light which is modeled as an optical field with unit amplitude and instantaneous random phase $\Phi_r(\hat{x},\hat{y})$ in the hologram plane. The experiment setup for recording SI interferograms that are used to reconstruct the object is displayed in Figure 2. The experiment setup has been described in detail in [6].

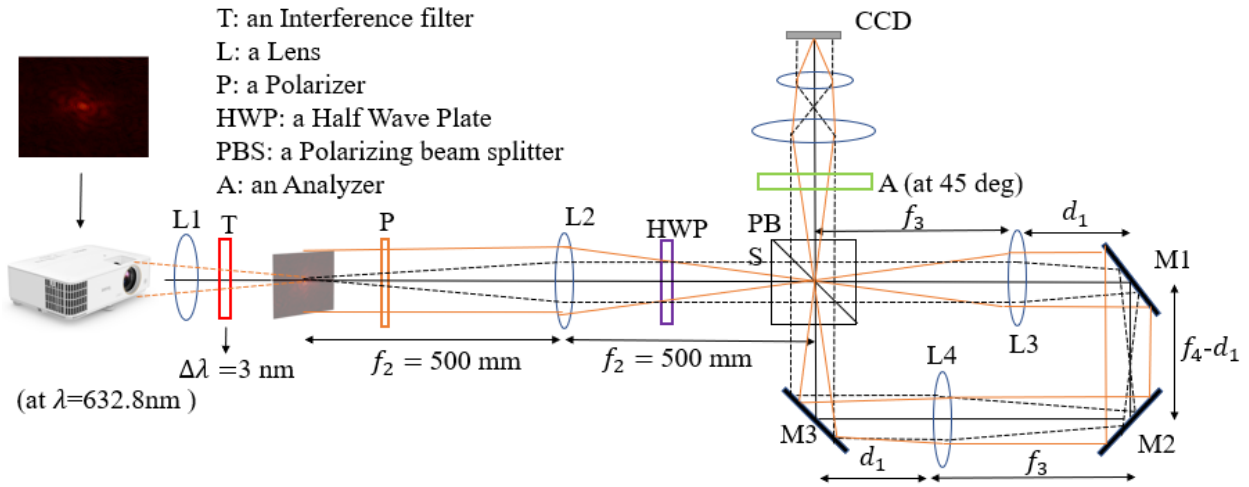


Figure 2. Experiment setup for recording the interferogram used for reconstruction [6]. The CGH is numerically generated. An interference filter T with a bandwidth of $\Delta\lambda$= 3nm at $\lambda = 632.8$ nm is used to mitigate chromatic aberrations on the recorded interferogram due to the optical elements. The values of the focal lengths f_3, f_4 dictates α and the lateral and axial magnifications for the reconstructed image are (m_x, m_y) and m_z, respectively.

The instantaneous field right at the rear focal plane of the FT lens L2 is

$$u(x,y,z) = \iint \sqrt{H(\hat{x},\hat{y})}\exp\left[i\,\Phi_r(\hat{x},\hat{y})\right]\,\exp[ik_z(\hat{x},\hat{y})z]\exp[i(2\pi/\lambda f)(x\hat{x}+y\hat{y})]\,d\hat{x}d\hat{y}, \tag{3}$$

where the square root transforms the intensity transmittance of the hologram to amplitude transmittance. The instantaneous time parameter is implicitly included in $\Phi_r(\hat{x},\hat{y})$ and $u(x,y,z)$. This field by itself does not reconstruct the object wave since the phase has been scrambled. Fortunately, a time independent field can be obtained from $u(x,y,z)$ by finding the mutual intensity between a pair of points or the coherence function, $\Gamma(\Delta x,\Delta y,\Delta z)$, which eliminates the effect of the instantaneous random phase $\Phi_r(\hat{x},\hat{y})$ resulting from the incoherent illumination. Therefore, $u(x,y,z)$ is directed into the SI to correlate the field and find the mutual intensity or coherence function. This correlating process using the SI is described in detail in [6]. The field intensity at the output of the SI is given by

$$I(\Delta x,\Delta y,\Delta z) = \langle|u(x_1,y_1,z_1)+u(x_2,y_2,z_2)|^2\rangle = 2\Gamma(0,0,0) + 2\mathrm{Re}[\Gamma(\Delta x,\Delta y,\Delta z)], \tag{4}$$

where < > denotes ensemble average, $\Delta x = x_2 - x_1, \Delta y = y_2 - y_1, \Delta z = z_2 - z_1$ are the difference of the coordinates and

$$\begin{aligned}\Gamma(\Delta x,\Delta y,\Delta z) &= \langle u^*(x_1,y_1,z_1)u(x_2,y_2,z_2)\rangle \\ &= \iint\iint \sqrt{H(\hat{x}_1,\hat{y}_1)}\sqrt{H(\hat{x}_2,\hat{y}_2)}\,\langle\exp[-i\Phi_r(\hat{x}_1,\hat{y}_1)]\exp[i\Phi_r(\hat{x}_2,\hat{y}_2)]\rangle \\ &\times \exp[-ik_z(\hat{x},\hat{y})z_1]\exp[-ik_z(\hat{x},\hat{y})z_1] \\ &\times \exp\left[-i\tfrac{2\pi}{\lambda f}(x_1\hat{x}_1+y_1\hat{y}_1)\right]\exp\left[i\tfrac{2\pi}{\lambda f}(x_2\hat{x}_2+y_2\hat{y}_2)\right]d\hat{x}_1d\hat{y}_1d\hat{x}_2d\hat{y}_2 \\ &= \iint H(\hat{x}_1,\hat{y}_1)\exp[ik_z(\hat{x}_1,\hat{y}_1)\Delta z]\exp\left[i\tfrac{2\pi}{\lambda f}(\hat{x}_1\Delta x+\hat{y}_1\Delta y)\right]d\hat{x}_1d\hat{y}_1. \end{aligned} \tag{5}$$

$$\Gamma(0,0,0) = \langle u^*(x_1,y_1,z_1)u(x_1,y_1,z_1)\rangle = \langle u^*(x_2,y_2,z_2)u(x_2,y_2,z_2)\rangle = \iint H(\hat{x},\hat{y})\,d\hat{x}d\hat{y}. \tag{6}$$

Since the hologram is illuminated by spatially incoherent quasi-monochromatic light with high temporal coherence, $\langle \exp[-i\Phi_r(\hat{x}_1,\hat{y}_1)]\exp[i\Phi_r(\hat{x}_2,\hat{y}_2)]\rangle = \delta(\hat{x}_1-\hat{x}_2,\hat{y}_1-\hat{y}_2)$ is used to derive Eq. (5) and Eq. (6). Under the assumption that the random field $u(x,y,z)$ is stationary both in time and space, the coherence function only depends on the difference of the coordinates.
By substituting Eq. (2) into Eq. (5), we have,

$$\Gamma(\Delta x,\Delta y,\Delta z) \propto \tilde{g}(\Delta x,\Delta y,\Delta z) + \frac{1}{2}[\breve{g}(\Delta x,\Delta y,\Delta z) + \breve{g}^*(-\Delta x,-\Delta y,-\Delta z)], \tag{7}$$

where

$$\tilde{g}(\Delta x,\Delta y,\Delta z) \propto \iint |G(\hat{x},\hat{y})|\exp[ik_z(\hat{x},\hat{y})\Delta z]\exp\left[i\frac{2\pi}{\lambda f}(\hat{x}\Delta x+\hat{y}\Delta y)\right]d\hat{x}d\hat{y}. \tag{8a}$$

$$\begin{aligned}\breve{g}(\Delta x,\Delta y,\Delta z) &\propto \iint G(\hat{x},\hat{y})\exp[ik_z(\hat{x},\hat{y})\Delta z]\exp\left[i\frac{2\pi}{\lambda f}(\hat{x}\Delta x+\hat{y}\Delta y)\right]d\hat{x}d\hat{y},\\ &\propto \iint F^{-1}\{\int F\{g(\Delta x,\Delta y,\Delta z_1)\}\exp[-ik_z(\hat{x},\hat{y})(\Delta z_1-\Delta z)]\mathrm{d}\Delta z_1\}.\end{aligned} \tag{8b}$$

$$\begin{aligned}\breve{g}^*(-\Delta x,-\Delta y,-\Delta z) &\propto \iint G^*(\hat{x},\hat{y})\exp[ik_z(\hat{x},\hat{y})\Delta z]\exp\left[i\frac{2\pi}{\lambda f}(\hat{x}\Delta x+\hat{y}\Delta y)\right]d\hat{x}d\hat{y},\\ &\propto \iint F^{-1}\{\int F\{g^*(-\Delta x,-\Delta y,-\Delta z_1)\}\exp[ik_z(\hat{x},\hat{y})(\Delta z_1+\Delta z)]\mathrm{d}\Delta z_1\}.\end{aligned} \tag{8c}$$

where $F^{-1}\{\cdot\}$ represents the inverse 2D Fourier Transform operation. The $\breve{g}(\Delta x,\Delta y,\Delta z)$ contains the non-diffracted object field at the Δz plane of interest for $\Delta z_1 = \Delta z$, and other diffracted fields resulting from the angular spectrum propagation of the object field from the other Δz_1 planes ($\Delta z_1 \neq \Delta z$) to the Δz plane of interest. Similarly, the $\breve{g}^*(-\Delta x,-\Delta y,-\Delta z)$ includes the non-diffracted conjugate object field at the $-\Delta z$ plane of interest for $\Delta z_1 = -\Delta z$, and other diffracted fields resulting from the angular spectrum propagation of conjugate object field from the other $-\Delta z_1$ planes ($\Delta z_1 \neq -\Delta z$) to the Δz plane of interest.
By substituting Eq. (7) into Eq. (4), the field intensity at the output of the SI is given by

$$\begin{aligned}I(\Delta x,\Delta y,\Delta z) \propto Re\{&2\tilde{g}(0,0,0) + 2\tilde{g}(\Delta x,\Delta y,\Delta z) + \breve{g}(0,0,0)\\ &+\breve{g}(\Delta x,\Delta y,\Delta z) + \breve{g}^*(0,0,0) + \breve{g}^*(-\Delta x,-\Delta y,-\Delta z)\}\end{aligned} \tag{9}$$

It is clear to observe from Eq. (9) that the recorded interferogram does contain the desired information of the object field $g(\Delta x,\Delta y,\Delta z)$ at the Δz of interest, beside the appearance of the other terms. The first term in Eq. (9) generates a uniform background. The third and the fifth ones also generate uniform backgrounds, but these terms can be removed by placing the object far off from the optical axis, which is the reason why the off-axis object is chosen to generate the hologram. By being able to physically eliminate the third and the fifth terms, we enhance the contrast of the image, which improves the ability of the DL model to extract the information of the desired object from the interferogram. Furthermore, if the information of the object is not known before reconstructing by using traditional imaging methods or DL approach, it is impossible to distinguish the non-diffracted field resulting from the object field from the one causing by conjugate object field. Fortunately, this can be solved by placing the 3D off-axis object at a distance z_0 away from the front focal plane ($z = 0$).

From the properties of the SI which is described in details in [6], $\Delta x = -(\alpha-\alpha^{-1})\tilde{x}$, $\Delta y = -(\alpha-\alpha^{-1})\tilde{y}$, and $\Delta z = (\alpha^2-\alpha^{-2})\tilde{z}$, where $\tilde{x}, \tilde{y}, \tilde{z}$ are the coordinates of the output plane of the interferometer, and $\alpha = f_3/f_4$. This means that reconstructed image is magnified in lateral and axial directions by factors $-(\alpha-\alpha^{-1})^{-1}$ and $(\alpha^2-\alpha^{-2})^{-1}$, respectively. We should note that the lateral magnification must be chosen such that the reconstructed image size fits the CCD aperture.

3. Analytical reconstruction algorithms

To obtain the information of the object from the recorded interferograms, two main imaging methods that either physically introduce a tilt in the SI or adjust the recorded CCD have been proposed [3,6].

The first technique is called Fourier Fringe analysis [3]. A small tilt is introduced into one of the mirrors in the Sagnac interferometer to generate an interferogram with a spatial carrier frequency f_c. The recorded CCD image under the introduction of a spatial carrier frequency is formulated below

$$I(\Delta x, \Delta y, \Delta z) = 2\Gamma(0,0,0) + 2|\Gamma(\Delta x, \Delta y, \Delta z)| \cos[\varphi(\Delta x, \Delta y, \Delta z) + 2\pi f_c(\Delta x + \Delta y)] \quad (10)$$

where $\Gamma(\Delta x, \Delta y, \Delta z) = |\Gamma(\Delta x, \Delta y, \Delta z)| \exp\big(i\, \varphi(\Delta x, \Delta y, \Delta z)\big)$. Then, the coherence function $\Gamma(\Delta x, \Delta y, \Delta z)$ is computed from the recorded interferogram by the Fourier transform method of fringe pattern analysis. The steps of this method are summarized in Figure 3.

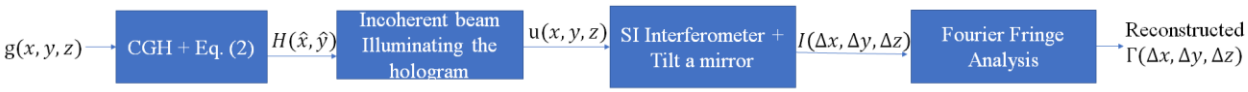


Figure 3. The steps of the Fourier Fringe analysis approach.

The second technique is based on the Sin-fit algorithm [6]. Instead of numerically generating only one hologram, a set of phase-shift computer generated holograms are created by introducing known phase shifts to the object spectrum

$$G(\hat{x}, \hat{y}; m) = G(\hat{x}, \hat{y}) \exp\,(i2\pi m/N), \quad (11)$$

where $m = 0, 1, 2, \ldots, N$ and N must be larger than three to generate over-determined equations from the N recorded interferograms $I(\Delta x, \Delta y, \Delta z; m)$. These over-determined equations are solved by the four-parameter sine wave fit algorithm to reconstruct $\breve{g}(\Delta x, \Delta y, \Delta z)$ and $\breve{g}^*(-\Delta x, -\Delta y, -\Delta z)$. The steps of this approach are summarized in Figure 4.

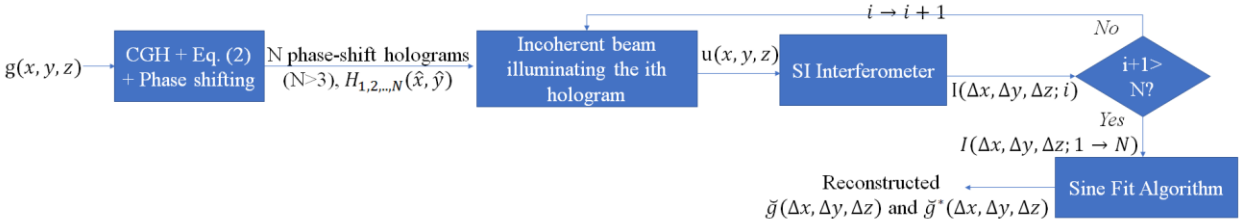


Figure 4. The steps of the Sin-fit algorithm.

In this work, we propose a deep learning based coherence holography (DCH) approach. It can be observed from the CCD images or even from Eq. (9) that the part which is not diffracted in each recorded interferogram corresponds to the sub-object at the z plane where it is located. The task of the DL model is to predict the in focused sub-objects from the captured images, which is the image synthesis task. The Conditional generative adversarial network **(**cGAN) learns a conditional generative model, where output image is conditioned on an input image, which makes cGANs suitable for image-to-image translation tasks especially for image segmentation. In addition, based on the conditional information, cGANs can generate images with high quality. Therefore, to acquire the reconstructed off-axis 3D object from the recorded interferograms, we adopted the *pix2pix* cGAN architecture [21]. The steps of reconstructing the magnified object $g(\Delta x, \Delta y, \Delta z)$ using DCH are summarized in Figure 5.

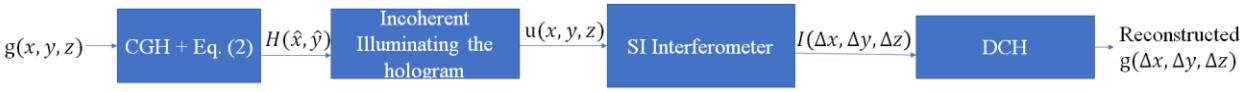
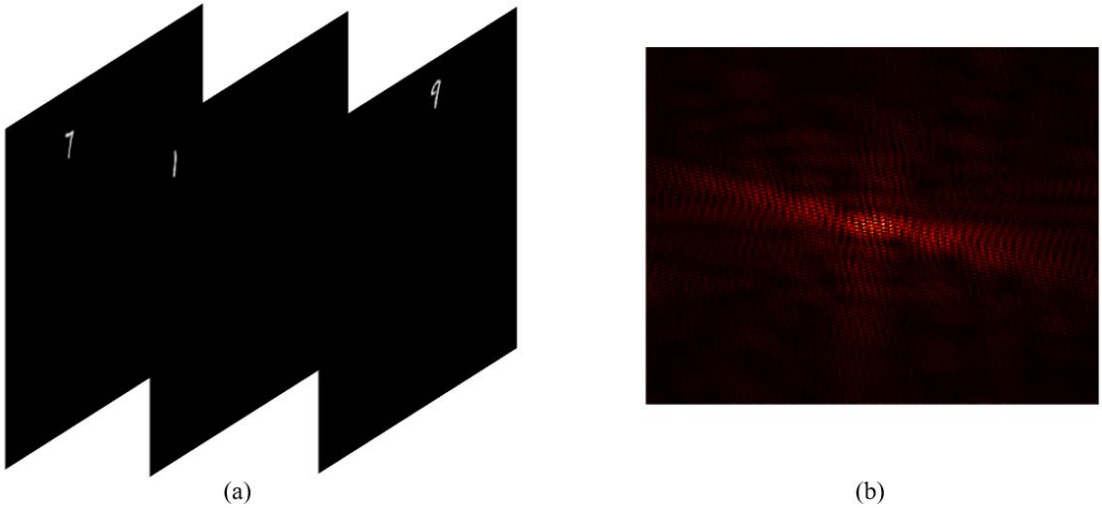


Figure 5. The steps of the proposed DCH approach.

4. Creating the dataset

The Mixed National Institute of Standards and Technology (MNIST) dataset [23] is used to generate off-axis 3D objects. Each of the objects consists of three sub-objects which are three random numbers from the MNIST dataset kept at different depth locations. The amplitude values of each number correspond to its 8-bit values in the MNIST dataset**.** Then, its phase values are proportional to its amplitude values, which means that 255 amplitude value corresponds to π phase value so that it is unnecessary to apply phase unwrapping when the phase is reconstructed. These numbers are positioned far away from the optical axis mentioned in Section 2, and in such a way that there is no occlusion when the off-axis 3D objects are created. The size of the sub-objects at each z plane was chosen to be 256 × 256 which is also the size of the recorded interferograms to enable training on commonly available GPUs. The pixel size of the object was set to be 0.5 μm, which leads to the physical size of the sub-objects is 1.28 ×1.28 mm. With this physical size, the distance between the front focal plane and the object is 1 mm which is large enough for the conjugate sub-objects to be highly diffracted. In addition, the physical size also affects the performance of the DCH, which is described in detail in Section 9**.** To achieve the best performance of the proposed network, the minimum distance between the z planes is set to be 0.1 mm, the smaller the distance between the z planes the harder for the DCH to predict the sub-objects from the recorded interferograms. For that reason, most of the off-axis 3D objects have the distance between the z planes close to this minimum distance, which is achieved by choosing the values on the right side of the Gaussian distribution with the mean equal to 0.1 and the standard deviation equal to 0.1. From these configurations, 4000 off-axis 3D objects were generated. The Fourier transform of these off-axis objects are used to generate the coherence holograms $H(\hat{x}, \hat{y})$ as described in Eq. (2). Figure 6 shows one of the off-axis 3D objects and its coherence hologram.



(a) (b)

Figure 6. (a) The off-axis 3D object; (b) The corresponding Fourier Transform coherence hologram.

The lateral and axial magnifications for the reconstructed image were chosen to be -5.23 and 2.6, respectively. Then, Eq. (4) or Eq. (9) is used to generate the simulated interferograms at different $\tilde{z}$ positions at the output of the SI. There are 4000 off-axis 3D objects created from the MNIST dataset, which leads to

12000 recorded interferograms at different positions. The input to the network is the recorded interferograms, and the outputs are the corresponding amplitude and phase of the sub-objects at the positions of interest. After obtaining this dataset, all the data is normalized between -1 and 1 to be suitable for the values of the output of the Tanh activation function which only goes from -1 to 1. Moreover, by normalizing the dataset, the values of the amplitude and phase of the object are transformed to be on a similar scale, which improves the performance and training stability of the model. The dataset is divided into 80% for training, 10 % for validating, and 10% for testing. Figure 7 shows an example of the recorded interferogram or the input and the amplitude and phase of corresponding sub-object or the ground truth.

Figure 7. (a) The recorded interferogram; (b) The amplitude and (c) The phase of the "two" sub-object.

5. Noise Simulation mimicking experimental data

To simulate the captured image of a camera, noise is added to the recorded interferogram. Noise in a camera image is the aggregate spatial and temporal variation in the measured signal, assuming constant, uniform illumination. There are several components of noise which are photon shot noise, read noise, and dark noise. Photon shot noise is the statistical noise associated with the randomness of emission of photons at the pixel. Due to this random emission, photon measurement obeys Poisson distribution. This noise is dependent on the signal level measured and is independent of sensor temperature. Read noise is the uncertainty in measuring the detected signal in a camera, determined by speed of readout and quality of electronic design. It is independent of signal level but is dependent on the temperature of the sensor. However, read noise is only significant in low light imaging. Dark noise is caused by the appearance of the dark current that flows even though no photons are incident on camera. It is thermal noise that builds up during the duration of an exposure caused by electrons spontaneously generated within the silicon chip. Like the read noise, dark noise is also independent of signal level but is dependent on the temperature of the sensor. Therefore, when a cooled camera is chosen to capture the images, this significantly reduces these two types of noise. The photon shot noise becomes the dominant one in terms of these three types of noise. Hence, photon shot noise is added to the recorded interferogram.

Not all the photons arriving at the pixels generate a signal level, but only a portion is converted to a signal level. This portion is the number of photons generating electrons or photoelectrons. The relation between the number of photons, p, hitting the pixels and the generated electrons, $e-$, is called quantum efficiency (Q.E). From the datasheet, the absolute Q.E for the red wavelength of the camera used in [6] is approximately 0.35

$$e^{-} = Q.E \times p. \tag{12}$$

As mentioned, the arriving photons follow Poisson distribution, the photoelectrons also follow the same distribution based on the equation above. The formular to compute the necessary number of generated electrons stored in each pixel from a recorded interferogram $I(\Delta x, \Delta y, \Delta z)$ is

$$e^{-}(\Delta x, \Delta y, \Delta z) = \frac{I(\Delta x,\Delta y,\Delta z)\times F_w}{R} \tag{13}$$

where R is the possible gray scale level values derived from the bit depth of the camera, and F_w is full well capacity defining the number of photoelectrons an individual pixel can hold before saturating which means the signal level reaches the maximum gray scale level. The larger the bit depth is, the larger the value of the full well capacity is since the increasing of the maximum gray scale level allows more photoelectrons that can be stored in a pixel. After the necessary number of photoelectrons at each pixel is computed, this number is considered as a mean value to generate a random number of photoelectrons, e^{-}_{noise}, following the Poisson distribution in order to simulate the shot noise. Then, the noisy recorded interferogram is computed by using the equation below which is based on Eq. (13)

$$I_{noise}(\Delta x, \Delta y, \Delta z) = \frac{e^{-}_{noise}(\Delta x,\Delta y,\Delta z)\times R}{F_w} = \frac{Pois(e^{-}(\Delta x,\Delta y,\Delta z))\times R}{F_w} \tag{14}$$

where $Pois(\cdot)$ represents the generation of a random number following the Poisson distribution. Based on the chosen camera in [6], the bit depth is 14, which yields 16383 gray scale levels. The full well capacity is not specified in the datasheet, so it is chosen in such a way that the Signal to Noise Ratio (SNR) is about 18 db, its value is set to be 500000 e-/gray. In addition, a small noise following a normal distribution accounting for the noise in experiment is added to the noisy image $I_{noise}(\Delta x, \Delta y, \Delta z)$. The function used for computing the SNR is shown below

$$SNR = 10\log_{10}\left[\frac{\mu_{I(\Delta x,\Delta y,\Delta z)}}{\sigma_{noise}}\right] \tag{15}$$

where $\mu_{I_{noise}}$ represents the expected value of the ideal recorded interferogram, σ_{noise} is the variance of the noise added to the ideal recorded interferogram.

6. Deep Coherence Holography Network Architectures

6.1 cGAN Network Architecture

cGANs consist of two adversarial models; a generator is defined by a function *G* that takes an observed image *x* and a random vector *z* as input and uses $\theta^{(G)}$ as parameters to capture the data distribution. A cGan is trained to produce output image *y* that cannot be distinguished from "real" images by an adversarially trained discriminator model which is a function *D* that takes *x* and *y* as input and uses $\theta^{(D)}$ as parameters. The discriminator is trained to identify "fake" images produced by the generator and estimates the likelihood that a sample belongs to the training data. cGANs learn a mapping from an observed image x and a random noise vector z to an output image y, e.g. G(x, z) → y [21]. The composite objective of a conditional GAN can be expressed as [21,22]

$$\mathcal{L}_{cGAN}(G, D) = \mathbb{E}_{x,y}[\log D(x, y)] + \mathbb{E}_{x,z}\big[\log\big(1 - D(x, G(x, z))\big)\big], \tag{16}$$

where $\mathbb{E}$ is the expectation operator. The generator attempts to minimize the objective against an adversarial discriminator that tries to maximize it, i.e.

$$G^{*} = arg\min_{G}\max_{D}\mathcal{L}_{cGAN}(G, D). \tag{17}$$

6.2 U-NET Generator Network

The generator network implementation is the well-known U-NET architecture based on the fully convolutional network [22], shown in Figure 8. It consists of a contracting path or encoder and an expansive path or decoder. The contracting path consists of eight 4×4 2D convolutional layers with stride equal to 2 to downsample the observed image. Each layer is followed by a leaky rectified linear unit (LReLU) activation function with slope 0.2 to alleviate the vanishing gradient problem and improve learning efficiency [24], except the last layer of the encoder which is followed by a ReLU activation function. The number of feature channels (i.e. the numbers of independent 2D spatial filters in the network node) of the 2D convolution layers are [64, 128, 256, 512, 512, 512, 512, 512]. The decoder includes seven 4×4 2D transpose convolutional layers with stride 2 to upsample the important features that are encoded by the contracting path. The first three layers of the expansive path are enforced with 50% dropout rate to avoid overfitting. The numbers of the feature channels of the decoder's layers are [512, 512, 512, 512, 256, 128, 64]. After the last decoder layer, a 2D convolution is applied to map the number of output channels and a *Tanh* activation function is applied to acquire the reconstructed image. Except for the first layer of the contracting path, Batch- Normalization is applied to the rest of the layers of both encoder and decoder to standardize the data, stabilize and improve network training [25]. The operations performed at each layer interface are detailed in the legend of Figure 8. While classical CNN solutions require information to flow through the entire encoder/decoder structure, the U-NET implementation provides skip connections between mirror symmetric layers in the encoder/decoder. Skip connections are a way to protect the information from loss during transport in neural networks and provide access to low level information which is crucial for image representation [22,26].

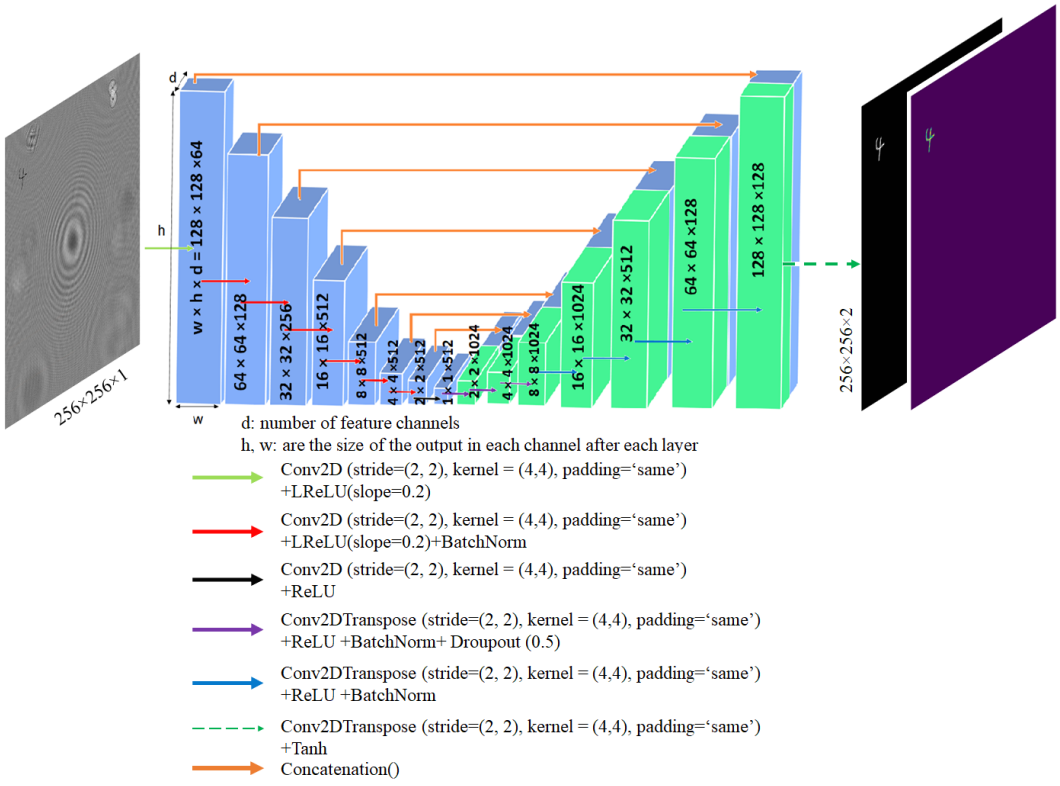


Figure 8. U-Net generator network consisting of convolutional encoder layers and de-convolutional decoder layers; skip connections between mirrored layers, depicted as orange arrows, provide access to additional low level information.

6.3 PatchGAN Discriminator Network

To further enforce high spatial frequency representation *pix2pix* adopts the well-known PathGAN discriminator framework [22], which only penalizes structure at the scale of local image patches. The discriminator is run for a series of convolution patches across the image, averaging all the responses to provide an aggregate likelihood that each $M \times M$ patch is a real image produced by the generator. By assuming independence between the pixel separations which are large relative to the patch size, the network will learn high order image texture [22]. The discriminator network, displayed in Figure 9, mimics the first four convolution layers in the contracting path of the generator network to extract the important features. Then, the image batches are predicted as real, or fake based on these extracted features by applying other two 2D convolution layers with stride equal one. The activation functions following these layers are a LReLU with slope 0.2 and a Sigmoid, respectively. The last layer is applied to map to a 1-dinmensional output. With this configuration, the discriminator network is applied independently to 256 image patches.

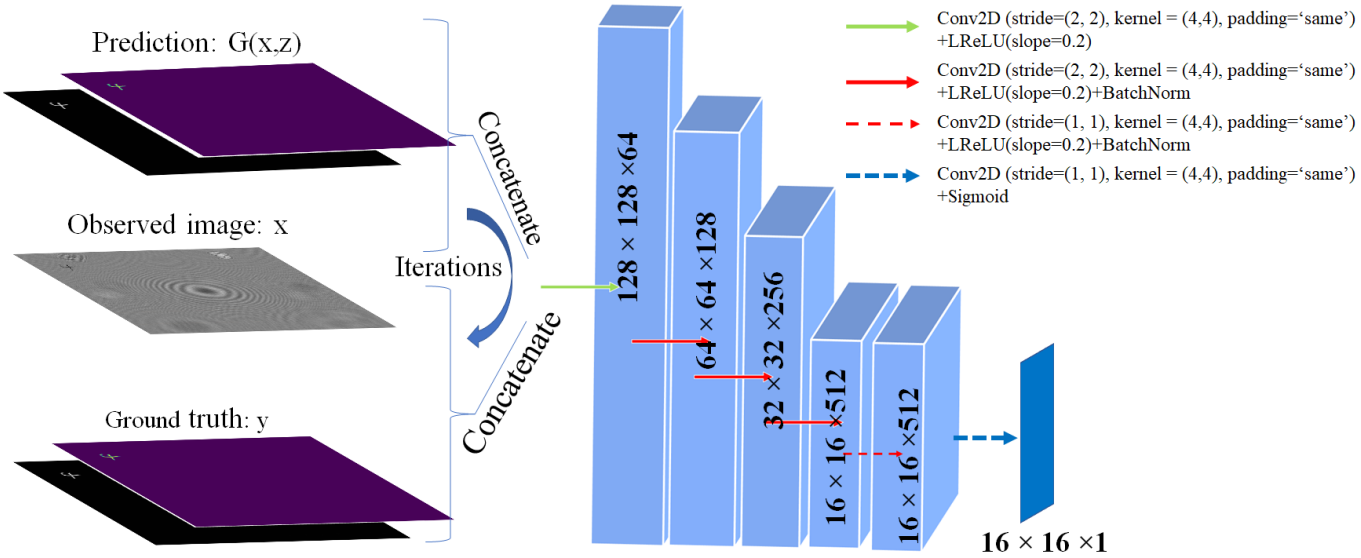


Figure 9. PatchGAN discriminator network is run for a series of convolution patches across the image and only penalizes structure at the scale of local image patches.

6.4 Loss Function

6.4.1 Discriminator loss $L^{(D)}$

Described mathematically, the discriminator seeks to maximize the average of the log probability of the real image and the log of the inverted probability of the fake image. If it is implemented directly, the stochastic ascent rather than the stochastic descent will be used to model the $\theta^{(D)}$, which is not more commonly implemented in practice. Instead of doing that, the discriminator wishes to minimize the loss function $L^{(D)}$ and must do so while controlling only $\theta^{(D)}$. Based on Eq. (11), the loss function used for the discriminator is [27]:

$$L^{(D)} = -\frac{1}{2}\mathbb{E}_{x,y}[\log D(x,y)] - \frac{1}{2}\mathbb{E}_{x,z}\left[\log\left(1 - D\left(x, G(x,z)\right)\right)\right]. \tag{18}$$

This is just the standard cross-entropy loss function that is minimized when training a standard binary classifier with a sigmoid output, where $D(\cdot)$ is the probability of the input concatenation being real. This loss function consists of two parts because it is trained on two minibatches of data; one coming from the

observed image and the "real" image in the dataset, where the label is 1 for all the examples, and one coming from the observed image and the output of the generator which is called "fake" image, where label is 0 for all examples. The purpose of minimizing the first part is to classify the concatenation of the real image and the observed image as real which means $D(x, y)$ =1, while minimizing the second part is to classify the concatenation of the fake image and the observed image as fake corresponding to $D(x, G(x, z)) = 0$.

It is clear to observe from the target images, the background class is much larger than the other classes, which means that the pixels containing the information of the object account for a small percentage of the total pixels in the image. Therefore, it makes the base loss low, which limits the ability of optimizing $\theta^{(D)}$ to successfully classify the real and fake images. To solve the data imbalance problem, the Dice loss function is added to $L^{(D)}$. Let D_{patch} denote the set of 256 image patches which are the output of the PatchGAN discriminator and each of the image patch $d_{i=1,2,\dots,256} \in D_{patch}$ is associated with a binary label $b_i = [b_{i0}, b_{i1}]$ corresponding to the fake or real class that d_i belongs to; the predicted probabilities of these two classes are $p_i = [p_{i0}, p_{i1}]$, where $b_{i0}, b_{i1} \in \{0, 1\}$, $b_{i=1,2,\dots,256} \in$ b, $p_{i0}, p_{i1} \in [0, 1]$ and $p_{i0} + p_{i1} = 1$. The dice loss and binary cross entropy loss functions of this PatchGAN discriminator are:

$$L_{Dice}(D_{patch}, b) = 1 - \frac{2\sum_{i=1}^{256} p_{1i} b_i + \gamma}{\sum_{i=1}^{256} p_{1i}^2 + \sum_{i=1}^{256} b_i^2 + \gamma}\ , \tag{19}$$

$$L_{BNE}(D_{patch}, b) = \frac{1}{256}\sum_{i=1}^{256} -(b_i \log(p_{i1}) + (1 - b_i)\log(1 - p_{i1})). \tag{20}$$

The γ factor that is set to be 10^{-6} is added to both the numerator and denominator of the Dice loss to prevent the undefined case when $p_{1i} = b_i = 0$. Comparing to the traditional dice loss function, the denominator is changed to the square form for faster convergence instead of non-square form [28]. Let L_{DBNE} be the summation of the Dice loss and the binary cross entropy

$$L_{DBNE}(D_{patch}, b) = L_{Dice}(D_{patch}, b) + L_{BNE}(D_{patch}, b). \tag{21}$$

The discriminator's loss function becomes

$$L^{(D)} = \left[L_{DBNE}(D_{patch}(x, y), b_{real}) + L_{DBNE}(D_{patch}(x, G(x, z)), b_{fake})\right], \tag{22}$$

where $b_{real} = [1, 1, \dots, 1]$ and $b_{fake} = [0, 0, \dots, 0]$.

6.4.2 Generator loss $L^{(G)}$

The generator wishes to minimize the loss function $L^{(G)}$ and must do so while controlling only $\theta^{(G)}$. Based on Eq. (16), Eq. (17), Eq. (22), the simplest version of the loss function of the generator is [27]

$$L^{(G)} = -L^{(D)}. \tag{23}$$

This loss function is only useful for theoretical analysis, but it does not perform especially well in practice. From the above equation, it is clear to observe that the discriminator minimizes a combination of cross-entropy and dice loss function, while the generator maximizes the same combination loss function. When the discriminator successfully rejects generator samples with high confidence, the generator's gradient vanishes, which prevent the generator to continue optimizing $\theta^{(G)}$ to generate better fake images that should not be able to be distinguished by the discriminator. To overcome this problem, the cross-entropy and dice loss minimizations are used to model $\theta^{(G)}$. Instead of flipping the sign on the discriminator's loss to get the loss function for the generator, we flip the target used to construct the cross-entropy and dice loss functions. The loss function of the generator is

$$L^{(G)} = L_{DBNE}(D(x, G(x,z)), b_{real}). \tag{24}$$

The purpose of minimizing the loss function above is to optimize $\theta^{(G)}$ in order for the generator to produce the "fake" image $G(x,z)$ concatenated with the observed image that is classified as "real" by the well-trained discriminator, which corresponds to $D\big(x, G(x,z)\big) = 1$. Previous improvement approaches [22,29] for cGANs have found it beneficial to mix the loss of the generator with other loss, such as $L2$ distance. Therefore, the negative Pearson correlation coefficient (NPCC) [30] loss and mean square error (MSE) are added to the generator's loss function. Adding these two loss functions improves the capability of reconstructing objects with different complexity and sparsity. The loss functions can be formulated as

$$L_{NPCC}(G(x,z), y) = \sum_{k=1}^{d} \frac{\sum_{i=1}^{w}\sum_{j=1}^{h}\left[G(x,z)_{ijk} - \widehat{G_k}\right]\left[y_{ijk} - \widehat{y_k}\right]}{\sqrt{\sum_{i=1}^{w}\sum_{j=1}^{h}\left[G(x,z)_{ijk} - \widehat{G_k}\right]^2}\sqrt{\sum_{i=1}^{w}\sum_{j=1}^{h}\left[y_{ijk} - \widehat{y_k}\right]^2}}, \tag{25}$$

$$L_{MSE}(G(x,z), y) = \frac{1}{whd}\sum_{k=1}^{d}\sum_{i=1}^{w}\sum_{j=1}^{h}\left|G(x,z)_{ijk} - y_{ijk}\right|^2, \tag{26}$$

where $\widehat{G_k}, \widehat{y_k}$ are the mean value of the predicted output and ground truth corresponding to channel k, respectively, d, w, h are the feature channels, the width and height of the predicted output of the generator, respectively.

The loss function of the generator is the weighted sum of the summation of Dice loss and Binary cross entropy, the NPCC and the MSE

$$L^{(G)} = \lambda_1 L_{DBNE}(D(x, G(x,z)), b_{real}) + \lambda_2 L_{NPCC}(G(x,z), y) + \lambda_3 L_{MSE}(G(x,z), y). \tag{27}$$

The values of $\lambda_1, \lambda_2, \lambda_3$ were chosen to be 1, 200, 350, respectively.

7. Results

7.1 Implementation

To optimize the network, the standard approach from [22] was followed: We alternate between one gradient descent step on the discriminator, then one step on the generator to minimize their loss functions defined above. In addition, the loss function of the discriminator is divided by 2 while optimizing the discriminator, which slows down the rate at which the discriminator learns relative to the Generator. The adaptive moment estimation (ADAM) optimizer is used for stochastic descent optimization with a learning rate of 0.0002 and momentum parameters $\beta_1 = 0.5$, $\beta_2 = 0.999$ [31]. The Batch size was set to 5.

All results were generated using a Window 10 based system with Intel(R) Core(TM) I9-9940X 14-core (28 threads) processor operating at 3.30 GHz base frequency with 131 GB RAM. To accelerate computations, a NVIDIA GeForce RTX 2080 TI with 11 GB of memory and 4352 CUDA cores were used. The algorithm was developed using Python 3.9.16 and CUDA Toolkit 11.3. The cGAN architecture for DCH was implemented using the TensorFlow (v2.6.0) framework, matrix operations were implemented using single-precision (32-bit) floating point operations evaluated using available CUDA cores.

7.2 DCH reconstruction results

To predict a reconstructed 3D off-axis object from the recorded interferograms at different axial positions, the captured images are input to the trained generator of the DCH network. As an example, a 3D off-axis object includes 3 sub-objects and the distances between the z planes and the origin of the input plane are 1 mm, 1.2 mm, and 1.4 mm, respectively. With the setup of the Sagnac SI, the recorded interferograms are

recorded at 2.6 mm, 3.12 mm, and 3.64 mm away from the origin of the output plane. Figure 10 shows these captured images and the corresponding results of the fringe contrast and phase of the sub-objects.

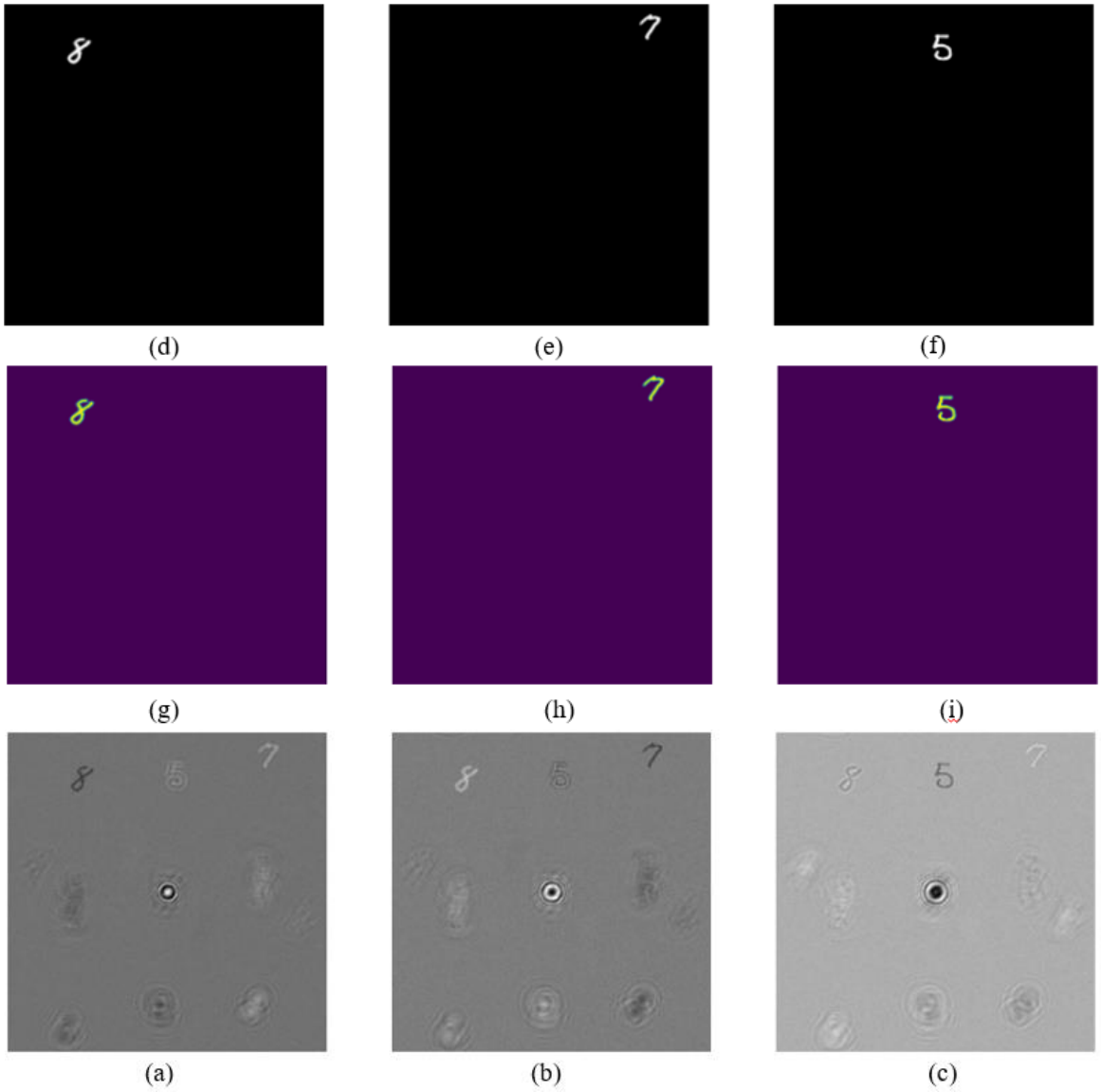


Figure 10. (a, b, c) Recorded interferograms at depths $\tilde{z} = 2.6\ mm, \tilde{z} = 3.12\ mm, \tilde{z} = 3.64\ mm$ corresponding to "eight", "seven", "five" sub-objects, respectively. (d, e, f) The reconstructed amplitude corresponds to each sub-object. (g, h, i) The corresponding true phase of these sub-objects.

The sub-objects "eight", "seven", "five" were reconstructed from the first, second, third capture images, respectively. They correspond to the focused part in each of the recorded interferogram. For clearer presentation, the data of the 135th column and 20th to 65th rows of the "five" sub-object is plotted and shown in Figure 11. Since the zero background is dominant in the sub-objects, the structural similarity index measure (SSIM) value between the reconstructed results and ground truth is still high even if the sub-object is wrongly predicted. Therefore, the Pearson Correlation coefficient (PCC) was used to evaluate the similarity between the reconstructed results and the ground truth instead of the (SSIM) since it takes the

dominance of the dark background into account, which prevents from getting a high value of similarity value when the sub-object is reconstructed incorrectly. In addition, the peak signal to noise ratios (PSNR) of the reconstructed amplitude and phase are also computed. These values are shown in Table 1 below.

Table 1. The MSEs, PCCs, PSNRs of the reconstructed sub-objects

	"Eight" sub-object	**"Seven" sub-object**	**"Five" sub-object**
Amplitude MSE	1.3905	0.5611	1.5977
Phase MSE	2.1116e-04	8.5176e-05	2.4291e-04
Amplitude PCC	0.9949	0.9966	0.9937
Phase PCC	0.9948	0.9967	0.9936
Amplitude PSNR	46.6990	50.6398	46.0957
Phase PSNR	46.6970	50.6397	46.0884

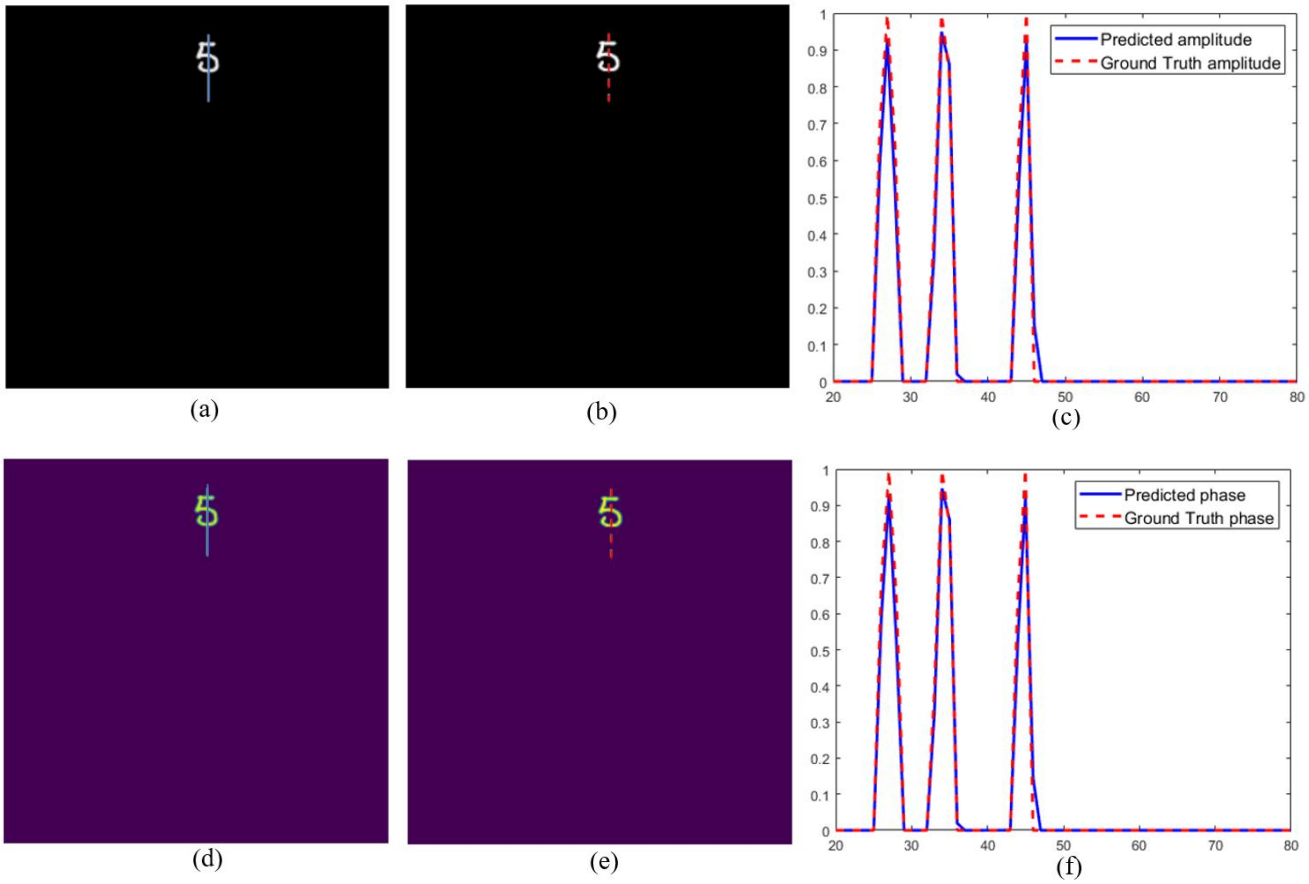


Figure 11. (a, b) The predicted and the ground truth amplitudes, the blue and red lines are at the 135th column and from 20th to 65th rows of the predicted and ground truth amplitudes, respectively;(c) The section curves of the 135th column and 20th to 65th rows. The zero values correspond to the dark background. (d, e) The predicted and the ground truth phases, the blue and red lines are at the 135th column and from 20th to 65th rows of the predicted and ground truth phases, respectively;(f) The section curves of the 135^{h} column and 20th to 65th rows. The zero values correspond to the zero phase values.

In addition, an object consisting of two sub-objects and an object consisting of one sub-object were reconstructed by using the same trained DCH network above to show the generalization of this network. The captured images and results of these reconstructions are displayed in Figures 12 and 13.

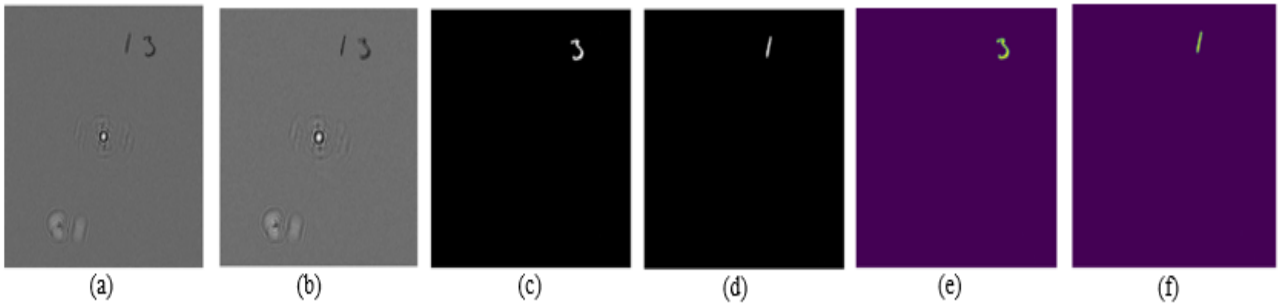


Figure 12. The results of an object consisting of two sub-objects. (a, b) The captured interferograms corresponds to “three”, “one” sub-objects. (c, d) The reconstructed amplitudes correspond to each sub-object. (e, f). The reconstructed phases of these sub-objects.

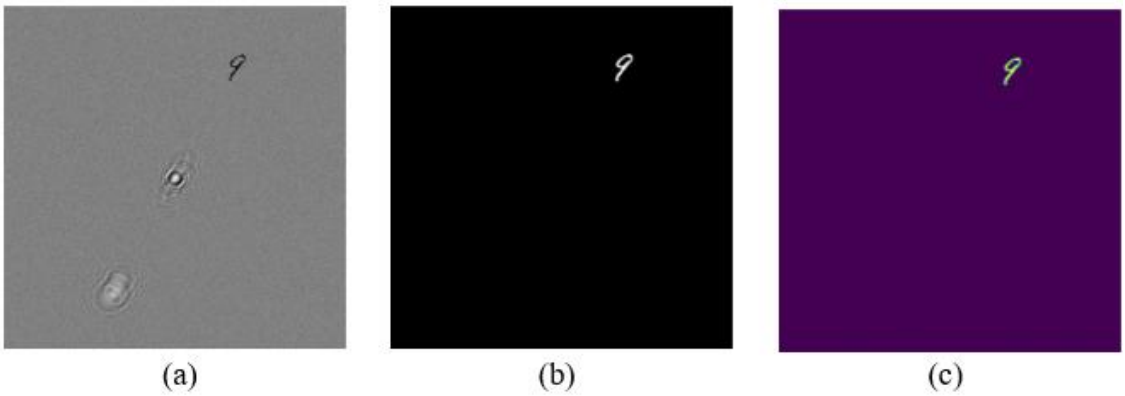


Figure 13. The results of a 2D object. (a) The captured CCD image. (b, c) The reconstructed amplitude and phase of the 2D object.

8. Comparison between DCH and analytical techniques

In this section, the reconstruction results of three methods are compared in terms of accuracy, reconstruction time, and resolution.

8.1 Reconstructing time

The reconstructing time consists of the total exposure time of the CCD to record the necessary interferograms and the execution time of the reconstructing algorithm applied in each method. Typically, the exposure time to record an interferogram is set to about 1 second in order to destroy spatial coherence by averaging out a large number of superposed fields created by all the possible random states of the incoherent illumination [4]. According to Figures 3, 4, and 5, the numbers of recorded CCDs needed to reconstruct a sub-object of Fourier Fringe analysis, Sin-fit algorithm, DCH are 1, N (N>3), and 1, respectively. After the required number of interferograms is recorded, the corresponding approach is applied to this set of interferograms. The execution time is the amount of time that it takes for the algorithm to yield the output. While the execution time to reconstruct the 3D off-axis object of the Sin-fit algorithm and Fourier Fringe analysis is proportional to the number of the sub-objects, the DCH time will slightly increase with the increasement of the number of sub-objects since it was trained with a large batch size. The reconstructing time of each method is shown in Table 2.

Table 2. Reconstruction time comparison. N represents the number of phase-shifted CGH in the case of sin-fit algorithm [6].

Methods	Fourier Fringe analysis	Sin-fit algorithm	DCH
Number of sub-objects (n)	5	5	5
Number of CCD images	1 image/sub-object	N=4 (N>3) images/sub-object	1 image/sub-object
Exposure time	1 s/image	1 s/image	1 s/image
Average executed time for a sub-object	18 ms	20 ms	85 ms
Average total execution time	$18 \times n = 18 \times 5 = 90\ ms$	$20 \times n = 20 \times 5$=100 ms	90 ms
Total reconstruction time	5.090 s	20.100 s	5.090

8.2 Resolution of the reconstructed objects

Let us first consider the resolution of the Fourier Fringe analysis technique. Rewriting Eq (10), we have

$$\begin{aligned} I(\Delta x, \Delta y, \Delta z) &= 2\Gamma(0,0,0) + |\Gamma(\Delta x, \Delta y, \Delta z)|\{\exp(j[\varphi(\Delta x, \Delta y, \Delta z) + 2\pi f_c(\Delta x + \Delta y)]) \\ &+ \exp(-\mathrm{j}[\varphi(\Delta x, \Delta y, \Delta z) + 2\pi f_c(\Delta x + \Delta y)])\} \\ &= 2\Gamma(0,0,0) + \Gamma(\Delta x, \Delta y, \Delta z)\exp[j2\pi f_c(\Delta x + \Delta y)] \\ &+ \Gamma^*(\Delta x, \Delta y, \Delta z)\exp[-j2\pi f_c(\Delta x + \Delta y)]. \end{aligned} \quad (28)$$

The 2D Fourier Transform of the recorded interferogram under the introduction of the carrier frequency is

$$\begin{aligned} F\{\mathrm{I}(\Delta x, \Delta y, \Delta z)\} &= C\delta(\Delta k_x, \Delta k_y, \Delta z) + F\{\Gamma(\Delta x, \Delta y, \Delta z)\} * \delta(\Delta k_x + f_c, \Delta k_y + f_C, \Delta z) \\ &+ F\{\Gamma^*(\Delta x, \Delta y, \Delta z)\} * \delta(\Delta k_x - f_C, \Delta k_y - f_c, \Delta z) \end{aligned} \quad (29)$$

where $\Delta k_x, \Delta k_y$ are the spatial frequency of the $\Delta x, \Delta y$ respectively, $C\delta(\Delta k_x, \Delta k_y, \Delta z) = F\{2\Gamma(0,0,0)\}$ with C is a constant. According to Eq. 29, $C\delta(\Delta k_x, \Delta k_y, \Delta z)$ is located at the center of the spatial frequency domain, the angular spectra of the coherence function and its conjugate are located at the $f_c, -f_c$, respectively. Figure 14 demonstrates the positions of these three terms in the spatial frequency domain. Let $\Delta k_{xmax}, \Delta k_{ymax}$ are the maximum spatial frequencies which are determined by the pixel size of the CCD, and $\Delta k_{W_x}, \Delta k_{W_y}$ are half of the bandwidth of the spectrum of the coherence function which determine the pixel size of the reconstructed field. It is clear to observe that $\Delta k_{W_x} < 0.5\Delta k_{xmax}, \Delta k_{W_y} < 0.5\Delta k_{ymax}$, which leads to the resolution of the reconstructed CF of the Fourier Fringe analysis approach is less than 0.5 times the resolution of the recorded interferogram.

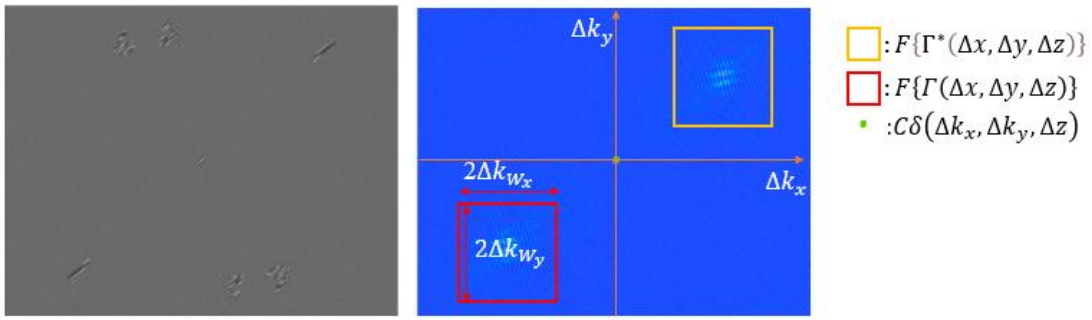


Figure 14. (a) The captured interferogram with the introduction of carrier frequency. (b) Fourier spectrum of this interference image.

Based on the configurations of the DCH network and the algorithm of the Sin-fit approach, the resolution of the reconstructed off-axis object using these methods is the same as the resolution of the recorded interferogram. The resolutions of the reconstructed signal of each approach are displayed in Table 3.

8.3 Accuracy of the reconstructed objects

The Fourier fringe analysis approach reconstructs the coherence function $\Gamma(\Delta x, \Delta y, \Delta z)$. As an example, the result of the reconstructed CF at the z plane of interest is shown in Figure 15. Moreover, the recorded interferogram is space-limited so its spectrum is unlimited in spatial frequency domain, applying Fourier filter can only get a part of the spectrum of the coherence function, which leads to an inaccuracy between the reconstructed and desired spatial coherence function.

While the DCH only reconstructs the magnified sub-object at the z plane of interest, the Sin-fit (phase-shift hologram) approach reconstructs the magnified sub-object at the z plane of interest and the diffracted fields resulting from the propagation of angular spectrum of other sub-objects and their conjugates from the other z planes to the z plane of interest. The reconstruction results of Sin-fit and DCH methods are displayed in Figures 16 and 17, respectively. Basically, these methods are superior to the Fourier fringe analysis approach in terms of accuracy. Furthermore, since the resolution of the reconstructed result of the Fourier Fringe analysis is smaller than the other methods, when accuracy is computed, before being applying inverse Fourier Transform, the Fourier filtered signal is padded for the reconstructed signal to have the same size as other approaches' results. Three methods are comparable to each other when the mean square errors (MSE) and Peak signal to noise ratios (PSNR) are computed only on the pixels where the sub-object is located. The MSEs and the PSNRs of the amplitude and phase of each approach are shown in Table 3.

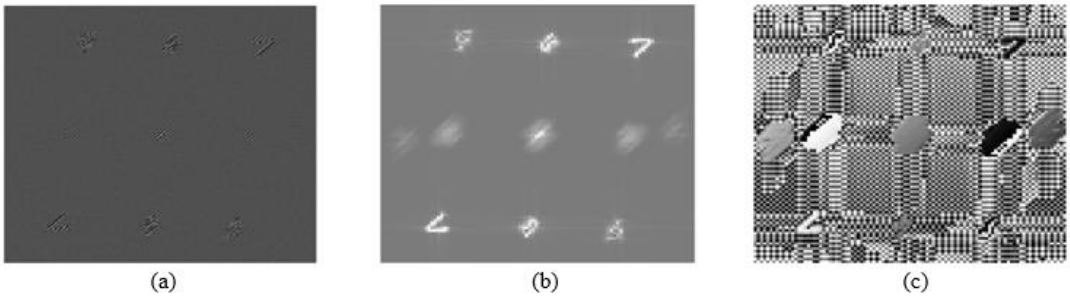
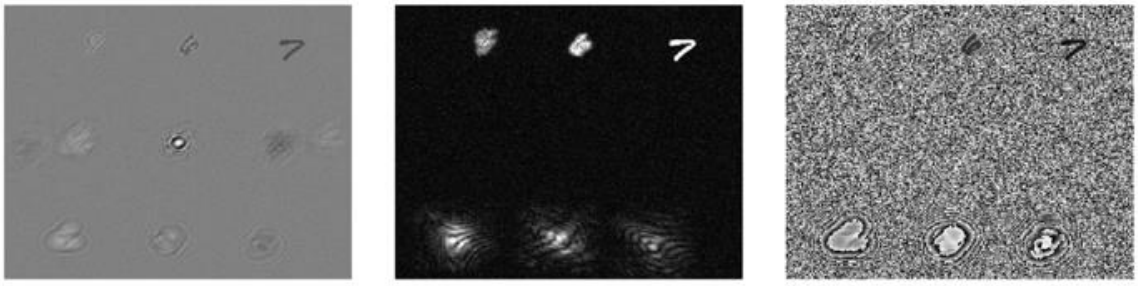


Figure 15. (a) Recorded Interferogram with fringes at location corresponding to a "seven" sub-object; (b) The reconstructed amplitude of the CF; (c) The reconstructed phase of the CF.



(a) (b) (c)

Figure 16. (a) One of the phase-shift recorded interferograms at location corresponding to a "seven" sub-object; Sin-fit: (b) reconstructed amplitude of the field; (c) reconstructed phase of the field.

(a) (b) (c)

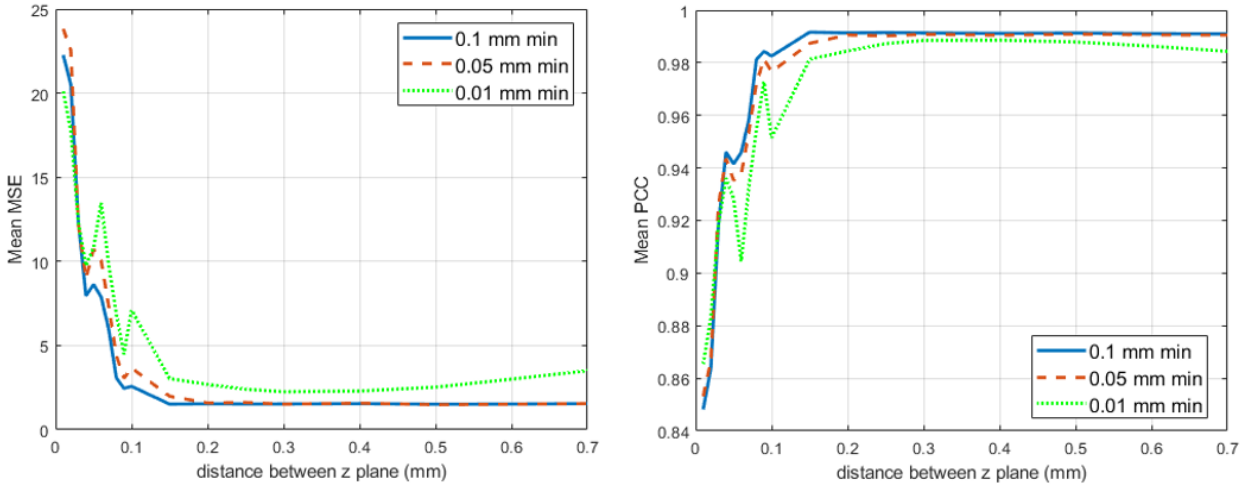
Figure 17. (a) The recorded interferogram; (b, c) DCH reconstructed amplitude and phase of the "seven" sub-objects

Table 3. Accuracy and Resolution of the reconstructed "seven" sub-object of each approach.

Methods	Fourier Fringe analysis	Sin-fit algorithm	Deep Coherence Holography
Amplitude MSE	2860.8	220.6338	1.0031
Phase MSE	1.2010	0.3197	1.5511e-04
Amplitude PSNR	13.5660	24.6941	48.1173
Phase PSNR	9.1477	14.8961	48.0.366
$\frac{\text{Reconstruction Resolution}}{\text{Object resolution}}$	Less than 0.5	1	1

9. Discussion

As mentioned in section 4, the distance between the z planes affects the performance of our proposed method. To demonstrate this effect, the DCH network was trained by different datasets having various minimum distances between the z planes or the depth resolutions which are 0.01 mm, 0.05 mm, and 0.1 mm under the same number of epochs. Then, the performances of these trained networks are compared to each other by looking at the average mean square error and Pearson correlation coefficient of the reconstructed amplitude of sets of 1500 sub-objects from 500 off-axis 3D objects having distance between the z planes at a certain value ranging from 0.01 mm to 0.7 mm. Figure 18 shows the mean MSE and the mean PCC at different values of the distance between the z planes.



(a) (b)

Figure 18. (a, b) The mean MSEs, PCCs of three networks trained by datasets having 0.01 mm, 0.05 mm, 0.1 mm minimum distances between the z planes.

According to the plots above, the network trained with the dataset having 0.1 mm depth resolution shows the best performance. In addition, when the distance between the z plane is less than 0.1 mm, the performances of the three trained networks start to decrease significantly even with the ones that was trained with the dataset consisting of these values of distance between the z planes. It is well-known that the angular spacing of the features in the diffraction pattern is proportional to the wavelength and inversely proportional to the dimensions of the object causing the diffraction. In other words, the smaller the diffracting object or the larger the angular spacing is, the more the diffraction pattern spreads. Since the DCH predicts the sub-object which is not diffracted from the recorded interferogram at the z plane of interest, if the distance between the z planes is not large enough for the diffraction pattern of sub-objects from the other z planes to spread enough, DCH will be confused to predict these diffracted fields as the part of the sub-object. In addition, for the same object's size, the larger the wavelength used to record the hologram and reconstruct the object, the higher the depth resolution.

In summary, due to the mechanism of the DCH, the distance between the z planes or the depth resolution depends on the physical size of the object. The smaller the object is, the smaller the depth resolution is, if the DCH is used to reconstruct the object. Therefore, with the physical size mentioned in Section 4, there are cases in which the fields from other z planes do not spread enough, which confuses the DCH which gives wrong results. The corresponding distances between the z planes for these cases are less than 0.1 mm. Figure 19 shows an example of this situation and the reconstruction results of the Sin-fit approach.

The advantage of this proposed method is that there is no need to retrain the DCH network when the physical size of the object changes. The only thing that needs to be considered is that the distance between the z planes should not be smaller than the optimal depth resolution depending on the size of the object in order for the DCH network to successfully predict the non-diffracted field from the captured image.

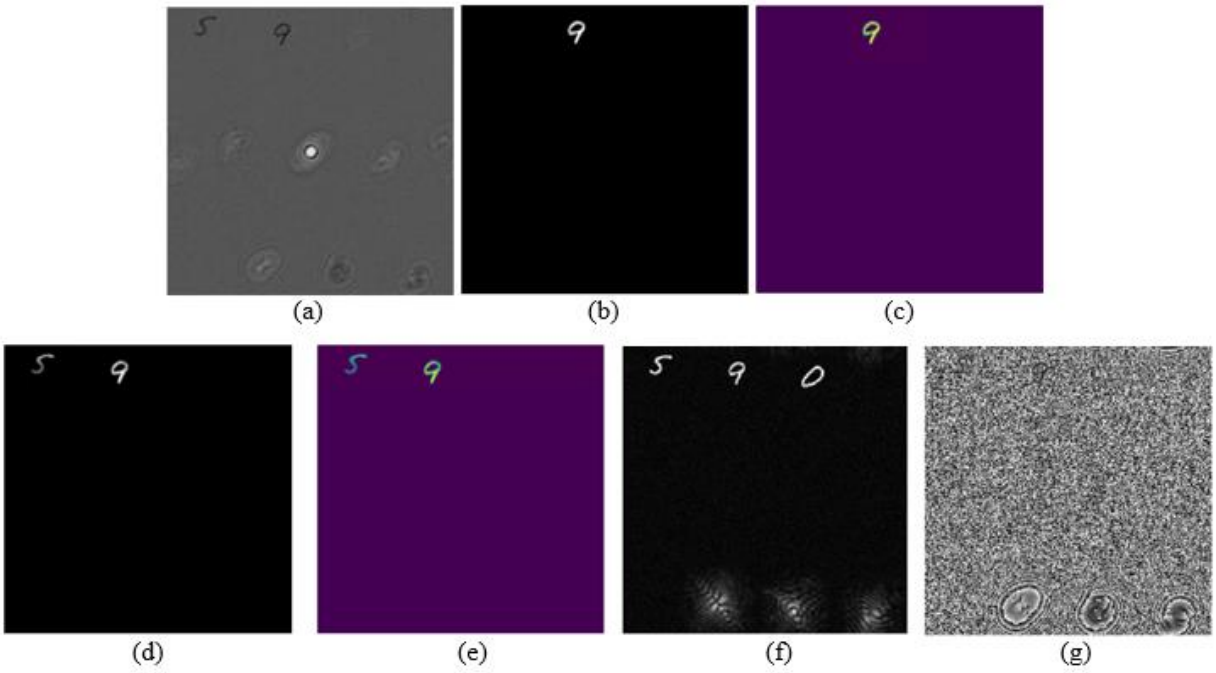


Figure 19. (a) The recorded interference image at the location corresponding to a "nine" sub-object, the distance between the z planes of the object is 0.01 mm; (b, c) The ground truth of the amplitude and phase of the "nine" sub-object; (d, e) The reconstructed amplitude and phase of the sub-object using DCH. Since the distance between z planes is not large enough, the diffracted field of a "five" sub-object does not spread enough, which results in this diffracted field becomes a part of the reconstructed results beside the "nine" sub-object; (f, g) The reconstructed amplitude and phase of the Sin-fit approach. Since this approach is subject to noise, the effect of the noise can be clearly observed from the reconstructed amplitude and especially from the reconstructed phase of the "nine" sub-object. In addition, the Sin-fit method predicts the non-diffracted field or the "nine" sub-object and other diffracted fields resulting from the angular spectrum propagation of the object and conjugate object fields. Thus, using this method is not able to only reconstruct the non-diffracted field.

10. Conclusion and Future work

Under the same reconstruction conditions, our proposed method DCH predicts the non-diffracted fields from the captured interferograms yielding a more accurate reconstructed object than the traditional analytical methods. The depth resolution of the DCH depends on the physical size of the object, the performance of the DCH reduces when the distances between the z planes are less than the depth resolution. Furthermore, even though the DCH can be applied to objects having various number of sub-objects, the same set of network parameters cannot cope with all types of samples since a DL method is based on data training. This is a common challenge for all DL imaging approaches based on data training at present. Among the existing methods to mitigate this challenge are: (a) training the network with a large and various types of data sets, (b) using the transfer learning technology to transfer the network parameters for different kinds of samples, or (c) use unsupervised learning techniques.

Funding.

Acknowledgments. Portions of this work were presented at the Optica Imaging Congress 2023, Digital Holography and 3D Imaging.

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper and Python code to perform DCH are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

[1] C. Falldorf, "Taking the next step: The advantage of spatial covariance in optical metrology," in Imaging and Applied Optics 2016, OSA Technical Digest (online) (Optica Publishing Group, 2016), paper DW3E.1.

[2] Falldorf, Claas, Mostafa Agour, and Ralf B. Bergmann. "Digital holography and quantitative phase contrast imaging using computational shear interferometry." Optical Engineering 54, no. 2 (2015): 024110-024110.

[3] M. Takeda, H. Ina, and S. Kobayashi, “Fourier-transform method of fringe-pattern analysis for computer-based topography and interferometry,” J. Opt. Soc. Am. **72**(1), 156–160 (1982).

[4] D. N. Naik, T. Ezawa, Y. Miyamoto, and M. Takeda, “Real-time coherence holography,” Opt. Express **18**(13), 13782–13787 (2010)

[5] P. Handel, “Properties of the IEEE-STD-1057 four-parameter sine wave fit algorithm,” IEEE Trans. Instrum. Meas. **49**(6), 1189–1193 (2000)

[6] D. N. Naik, T. Ezawa, R. K. Singh, Y. Miyamoto, and M. Takeda, “Coherence holography by achromatic 3-D field correlation of generic thermal light with an imaging Sagnac shearing interferometer”, Opt. Express **20**(18), 19658- 19669 (2012)

[7] Shi, L. et al. “Towards real-time photorealistic 3D holography with deep neural networks”. Nature 591, 234–239 (2021).

[8] Shi, L., Li, B. & Matusik, W. End-to-end learning of 3D phase-only holograms for holographic display. Light Sci Appl 11, 247 (2022).

[9] W. Ouyang, A. Aristov, M. Lelek, X. Hao, and C. Zimmer, “Deep learning massively accelerates super-resolution localization microscopy,” Nat. Biotechnol. 36, 460–468 (2018).

[10] Nehme, L. E. Weiss, T. Michaeli, and Y. Shechtman, “Deep-storm: super-resolution single-molecule microscopy by deep learning,” Optica 5, 458–464 (2018).

[11] H. Wang, Y. Rivenson, Y. Jin, Z. Wei, R. Gao, H. Günaydın, L. A. Bentolila, C. Kural, and A. Ozcan, “Deep learning enables crossmodality super-resolution in fluorescence microscopy,” Nat. Methods 16, 103–110 (2019).

[12] C. Ling, C. Zhang, M. Wang, F. Meng, L. Du, and X. Yuan, “Fast structured illumination microscopy via deep learning,” Photon. Res. 8, 1350–1359 (2020).

[13] Shuo Zhu, Enlai Guo, Jie Gu, Lianfa Bai, and Jing Han, "Imaging through unknown scattering media based on physics-informed learning," Photon. Res. 9, B210-B219 (2021).

[14] Shuai Li, Mo Deng, Justin Lee, Ayan Sinha, and George Barbastathis, "Imaging through glass diffusers using densely connected convolutional networks," Optica 5, 803-813 (2018).

[15] M. Lyu, H. Wang, G. Li, S. Zheng, and G. Situ, “Learning-based lensless imaging through optically thick scattering media,” Adv. Photon. 1, 036002 (2019).

[16] E. Guo, S. Zhu, Y. Sun, L. Bai, C. Zuo, and J. Han, “Learning-based method to reconstruct complex targets through scattering medium beyond the memory effect,” Opt. Express 28, 2433–2446 (2020).

[17] E. Guo, Y. Sun, S. Zhu, D. Zheng, C. Zuo, L. Bai, and J. Han, “Single-shot color object reconstruction through scattering medium based on neural network,” Opt. Lasers Eng. 136, 106310 (2020).

[18] M. Takeda, W. Wang, Z. Duan, and Y. Miyamoto, “Coherence holography,” Opt. Express 13(23), 9629–9635 (2005).

[19] M. Born, and E. Wolf, “Principles of Optics”, 4th ed. (Pergamon, London, 1970), Chap. 10.

[20] J. W. Goodman, Statistical Optics, 1st ed. (Wiley, New York, 1985), Chap. 5.

[21] M. Mirza and S. Osindero. “Conditional generative adversarial nets,” *arXiv,* 1411.1784 (2014).

[22] P. Isola, J. Zhu, T. Zhou, and A. Efros, “Image-to-Image Translation with conditional adversarial networks,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 5967-5976 (2017).

[23] Y. LeCun, C. Cortes, and C. J. C. Burges, “The MNIST database of handwritten digits,” http://yann.lecun.com/exdb/mnist/.

[24] D. Clevert, T. Unterthiner, and S. Hochreiter, “Fast and accurate deep network learning by exponential linear units (ELUs)”, *arXiv,* 1511.07289 (2016).

[25] S. Ioffe and C. Szegedy, “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Coveriate Shift,” *arXiv*, 1502.03167 (2015).

[26] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” *arXiv,* 1505.04597 (2015).

[27] I. Goodfellow, “NIPS 2016 Tutorial: Generative Adversarial Networks”, arXiv, 1701.00160 (2016)

[28] F. Milletari, N. Navab and S. -A. Ahmadi, "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation," *2016 Fourth International Conference on 3D Vision (3DV)*, Stanford, CA, USA, 2016, pp. 565-571, doi: 10.1109/3DV.2016.79.

[29] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, “Context encoders: Feature learning by inpainting,” In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (2016), pp. 2536–2544

[30] A. M. Neto, A. C. Victorino, I. Fantoni, D. E. Zampieri, J. V. Ferreira, and D. A. Lima, “Image processing using Pearson’s correlation coefficient: applications on autonomous robotics,” in 13th International Conference on Autonomous Robot Systems (Robotica) (IEEE, 2013), pp. 1–6.

[31] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” arXiv preprint arXiv:1412.6980 (2014).

[32] 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. Exp., 25(13) 15043-15057 (2017).

[33] Brad Bazow, Thuc Phan, Christopher B. Raub, and George Nehmetallah, “Computational Multi-wavelength (MW) Phase Synthesis Using Convolutional Neural Networks (CNNs),” [Invited], Applied Optics, 61(5), B132-B146 (2022).

[34] T. Nguyen, V. Bui, and G. Nehmetallah, “Computational Optical Tomography Using 3D Deep Convolutional Neural Networks (DCNNs),” Optical Engineering 57(4), 043111 (April 2018).