

PHOTONICS Research

Incoherent imaging through highly nonstatic and optically thick turbid media based on neural network

SHANSHAN ZHENG,^{1,2} HAO WANG,^{1,2} SHI DONG,^{1,2} FEI WANG,^{1,2} AND GUOHAI SITU^{1,2,3,4,*} 

¹Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Shanghai 201800, China

²Center of Materials Science and Optoelectronics Engineering, University of Chinese Academy of Sciences, Beijing 100049, China

³Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou 310024, China

⁴CAS Center for Excellence in Ultra-intense Laser Science, Shanghai 201800, China

*Corresponding author: ghsitu@siom.ac.cn

Received 1 December 2020; revised 18 January 2021; accepted 18 February 2021; posted 19 February 2021 (Doc. ID 416246); published 21 April 2021

Imaging through nonstatic scattering media is one of the major challenges in optics, and encountered in imaging through dense fog, turbid water, and many other situations. Here, we propose a method to achieve single-shot incoherent imaging through highly nonstatic and optically thick turbid media by using an end-to-end deep neural network. In this study, we use fat emulsion suspensions in a glass tank as a turbid medium and an additional incoherent light to introduce strong interference noise. We calibrate that the optical thickness of the tank of turbid media is as high as 16, and the signal-to-interference ratio is as low as -17 dB. Experimental results show that the proposed learning-based approach can reconstruct the object image with high fidelity in this severe environment. © 2021 Chinese Laser Press

<https://doi.org/10.1364/PRJ.416246>

1. INTRODUCTION

Conventionally, an optical imaging system can be regarded simply as a one-to-one mapping system, as a spherical wavelet emitted from a local point on the object plane is converged by the imaging optic and forms a unique point (subject to diffraction) on the imaging plane. In this way, a clear image of the object can be formed. However, the presence of scattering media between the object and the imaging optic prohibits such a clear image from being formed because a significant part of the light propagating all the way from the object to the optic is scattered many times, producing a noise-like scattered pattern on the image plane. As the optical thickness of the scattering medium is larger than a certain value, the scattered light arriving at the camera becomes more intensive than the unscattered or ballistic one; further, the image is submerged by the noise-like scattered pattern. Thus, an intuitive method for image acquisition under such hazed conditions is to select the ballistic light by using gating techniques in the time [1], space [2], polarization [3], or even coherence [4] domains.

Actually, the propagation and transport of coherent waves in disordered media has been one of the central problems in many different disciplines of science and engineering, ranging from microwaves to electron waves. It has been shown that coherent scattering in a static scattering medium is linear and deterministic and can be described by a transmission matrix [5]. One of

the lines of recent studies then focuses on the ways of reversing the deterministic coherent scattering process. Physically, this can be done by generating a phase conjugated version of the scattered wave [6,7], so that it can trace the way back through the scattering medium in the first place and form a clear image. An alternative way is to measure the transmission matrix [8] or precompensate the phase distortion introduced by the scattering process by the technique of wavefront shaping [5]. When applying these methods in dynamic scattering media, the wavefront measurement and playback must be made within the decorrelation time [9]. Therefore, faster wavefront modulators, parallel processing, and more effective optimization algorithms are highly demanded in these cases [10,11].

Alternatively, one can utilize some intrinsic statistical properties of the scattered light instead of reversing the scattering process. For example, the short-range second-order correlation of a speckle pattern formed by the scattered light, or the memory effect [12,13], has been found to be particularly interesting in this case, as it allows computational reconstruction of an object image from the autocorrelation of the scattered light intensity by using a regular phase-retrieval algorithm [14,15]. However, the memory effect offers a field of view (FoV) that is inversely proportional to the effective thickness of the medium. Thus, this method is usually applicable for imaging through optically thin scattering media [14,15] or around

the corner [16]. We note that efforts have been made to enlarge the FoV by using multiplexing or scanning [17,18]. But these techniques have limited effect in the case when the scattering medium is not static, as the dynamic characteristics of the media and the spectral width of the light in many practical environments will further shrink the correlation length of the scattered light [15,19] and thereby the FoV of the imaging system.

An interesting and practically important question then naturally arises, i.e., if it is possible to reconstruct an image when the coherence of the scattered light is completely lost. At least two factors may lead to the loss of coherence: multiple scattering of the light during its propagation inside the scattering medium [19] and the coherence nature of the light itself.

To answer this question, we attempt to demonstrate that it is still possible to reconstruct the image of an object hidden behind a nonstatic and optically thick diffusive medium and under the illumination of incoherent light. To mimic a realistic environment, we also switch on an ambient light so as to introduce a strong interference noise that is incoherently superimposed onto the scattered light from the object. The method we use to address this problem is deep neural networks.

Deep neural networks have shown great potential in solving many computational optical imaging problems [20], ranging from digital holography reconstruction [21,22], phase imaging [23,24], computational ghost or single-pixel imaging [25,26], to coherent imaging through optically thin [27,28] and even thick scattering media [29,30]. As almost all the deep neural networks for computational optical imaging were trained supervisedly (except, for example, the work reported in Ref. [24]), it is therefore required that each pair of labeled images in the training set should have exclusively unique features [31]. In the application of imaging through scattering media, when the illumination is coherent [27–30], what is recorded by the camera is a high-contrast speckle pattern, the morphology of which is explicitly object-dependent. However, when the objects are illuminated with an incoherent light, as in our case, the coherent effect will be smeared out, yielding milky patterns that are almost visually indistinguishable with respect to one another. Therefore, additional care has to be taken to identify unique features to distinguish each milky pattern.

This article is organized as follows. The problem of incoherent imaging through highly dynamic turbid media is formally described in Section 2. Measurements of the parameters that characterize the turbid medium are presented in Section 3. The experimental demonstration of the proposed method is presented in Section 4.

2. DESCRIPTION OF THE PROBLEM AND IMAGING ENVIRONMENT

Let us describe the problem with the help of Fig. 1. An incoherent light emitted from an LED (M530L3, Thorlabs) with a central wavelength of 530 nm was first collimated by using a Köhler illumination system and then shined on an object [which was displayed on an amplitude-only spatial light modulator (SLM; Pluto 6001, Holoeye)]. The light reflected from the SLM was then guided to a glass tank of fat emulsion suspension, which was prepared by mixing a tank of distilled water

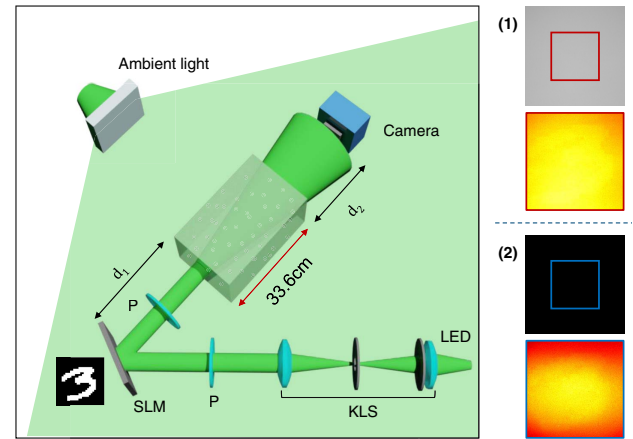


Fig. 1. Incoherent scattering imaging experimental system. (1) and (2) are the captured scattered patterns (the raw data and corresponding partial contrast stretched map) with optical thickness of 8 and 16, respectively. Note that these data are recorded in two sets of experiments: (1) capture data by the camera directly; (2) capture data with two additional apertures placed before the camera. KLS, Köhler lighting system; P, polarizer; ambient light, generated by a high-power LED through a diffuse slate (the distance between the slate and the tank side was around 3.5 cm); camera, working with an imaging lens ($f = 250$ mm, not shown in the figure). $d_1 \approx 41$ cm, $d_2 \approx 15$ cm. The 33.6 cm thick tank is equipped with fat emulsion diluent to simulate a dynamic scattering medium. Note that the scattered patterns shown in (2) look dimmed because a significant part of the large-angle scattered light has been blocked out.

and fat emulsion (Intralipid 20%, Fresenius Kabi). Multiple scattering events occur as the signal-carrying light propagated through the turbid medium and became highly diffusive when it left the tank (geometric length: 33.6 cm) from the back surface. An sCMOS camera (Zyla, Andor Technology Ltd.) with an $f = 250$ mm imaging lens mounted on it was used to take the scattered light pattern at the back surface of the tank. According to the propagation trajectory, the signal scattered light included a small amount of weakly/snake light and a significant part of large-angle scattered light. The snake light contains a small amount of the object information and contributes to the signal on the imaging plane. The camera was adjusted to focus on the SLM plane before adding intralipid into the tank. We also introduced strong interference noise generated by a high-power LED (Solis-525c, Thorlabs) illuminated from one side of the tank. The central wavelength of the ambient LED is 525 nm. This beam of light was also scattered by the intralipid suspensions, and some of it was superimposed with the object scattered light incoherently, which can be treated as interference noise. Although the light was scattered mostly forwardly in the fat emulsion suspensions, the large angle and multiple-scattered ambient light was strong in comparison with the object scattered light. Indeed, the measured signal-to-interference ratio was $\text{SIR} = -17$ dB in our experimental configuration.

The optical thickness of the tank of fat emulsion suspensions was about 16 (refer to Section 3 for the details of calibration). As the absorption of the intralipid is low, we can conclude that it is the scattering that mostly accounts for the attenuation

of the light. Then, the object's scattered light intensity $\mathcal{S}\{I_o\}$ can be expressed as

$$\mathcal{S}\{I_o\} = \mathcal{S}_W\{I_o\} + \mathcal{S}_L\{I_o\}, \quad (1)$$

where $\mathcal{S}_W\{I_o\}$ and $\mathcal{S}_L\{I_o\}$ are the intensity of weak light and large-angle scattered light, respectively, without considering the ambient light.

In our experimental demonstration, we also take the ambient scattered light I_a into account. Thus, the final intensity pattern captured by the camera can be written as

$$I_s = \mathcal{S}_W\{I_o\} + \mathcal{S}_L\{I_o\} + I_a, \quad (2)$$

where the interference noise $I_n = \mathcal{S}_L\{I_o\} + I_a$, and \mathcal{S}_W can be approximately regarded as the signal component of the recorded image I_s . Thus, the signal-to-interference ratio $\text{SIR} = 10 \log_{10} \mathcal{S}_W\{I_o\}/I_n = -17 \text{ dB}$.

Thus, the purpose is to reconstruct the object image I_o from I_s .

3. CHARACTERIZATION OF THE INTRALIPID SUSPENSIONS

Fat emulsion is similar to milk in many ways and mainly consists of soybean oil, water, glycerin, and eggphospholipid. Due to emulsification in the presence of phospholipids, the oil is suspended in water as small droplets with a monolayer lipid membrane. As a large amount of fat droplets are randomly distributed in the emulsion suspensions, a beam of light that propagates through it is distorted owing to multiple scattering, making it difficult to predict the propagation trajectory for each light ray.

Owing to the advantages such as low absorption, easily adjustable scattering coefficient by dilution, standardized particle size distribution, nontoxicity, and low cost [32], intralipid has been frequently used as an optical phantom that mimics a turbid medium, in particular, for biomedical applications. For example, it has been widely used in the calibration of clinical application systems [33,34] and as a diffusive reference standard for optical tissue modeling [35,36]. Many studies have been carried out to measure the scattering properties of intralipid suspensions [32,37]. However, when it is used as a scattering medium in imaging, the decorrelation time and optical thickness are quite important parameters related to the scattering, which are rarely measured or quantified. Thus, we need to calibrate these two parameters in this section.

A. Optical Thickness

The optical thickness (OT) is a parameter deduced from the Lambert–Beer law:

$$I = I_0 e^{-\text{OT}} = I_0 \cdot e^{-\mu \cdot L}, \quad (3)$$

where I_0 and I are the ballistic light intensities measured in front of and behind the tank of intralipid suspensions, μ is the attenuation coefficient, c the concentration of intralipid dilution, and L the path length through the suspensions. The attenuation coefficient μ is the sum of scattering coefficient μ_s and the absorption coefficient μ_a , i.e., $\mu = \mu_s + \mu_a$.

Since we used the 530 nm LED (M530L3, Thorlabs) in our demonstrations, the optical thickness of the fat emulsion should be measured at or at least closed to this wavelength.

We thus used a semiconductor pumped green laser with a wavelength of 532 nm as the light source in the calibration process. An optical power meter (PM200, Thorlabs) with the probe diameter of 9.5 mm was used to measure I_0 and I . Then, the optical thickness can be calculated according to Eq. (3). Note that the contributions to the absorption of fat emulsion are mainly from water and soybean oil, while the influence of glycerin and egg lipid can be ignored due to the low concentration. For the light at around 530 nm, the absorption coefficient μ_a is about 10^{-4} mm^{-1} according to a previous study [32], so that its influence can be ignored; therefore, $\mu \approx \mu_s$.

In the measurement, we first filled in the glass tank with a volume of 11.46 L purified water (33.6 cm × 19.5 cm × 17.5 cm) and then gradually dropped 0–4 mL fat emulsion (intralipid 20%, Fresenius Kabi) into the water to make the intralipid suspensions with different density. Each time, about 0.4 mL intralipid was dropped into the tank of water. Then, we stirred the mixed fluid with a stick, let it stand for about 5 min, and then used a power meter placed on the beam path to measure the power of I_0 and I . A large black cardboard with a 3 mm diameter aperture was used to select the ballistic light when measuring I . Then, the optical thickness can be calculated according to Eq. (3), and the result is plotted in Fig. 2(a). One can see that the experimental data linearly fit the theory very well. From the slope of the fitted line segment, we can obtain the scattering coefficient of the intralipid 20%, $\mu_s = 132 \text{ mm}^{-1}$, which is consistent with previous studies [32]. Deviation of the measurement data from theoretical prediction is observed when the density is high. This is because of the unnegligible contribution of some amount of scattered light received by the power probe. Nonetheless, we can conclude that the optical thickness of the fat emulsion suspensions was about 16 when 4 mL intralipid was applied. This is the turbid medium that we want to look through.

The averaged diameter of the droplet is around $0.3 \mu\text{m}$ [38]; at a scale comparable with the wavelength of the light, Mie scattering occurs when light passes through the fat emulsion suspension. During our experiment, we observed the Tyndall effect from the side of the tank, as shown later in this paper, which was taken by a mobile phone camera at the experiment site. Every droplet in the fat emulsion can be seen as a secondary source that re-emits a wavelet, the superposition of which forms a speckle pattern. However, we observed an interesting

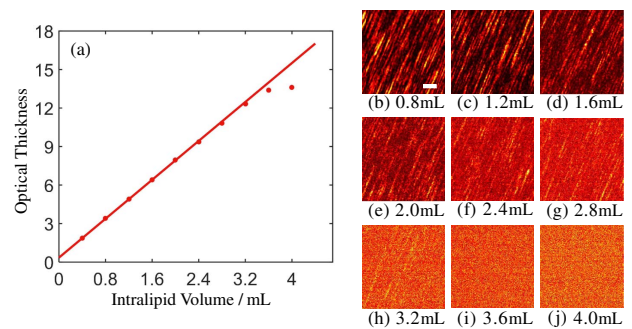


Fig. 2. Optical thickness of intralipid suspensions with respect to its density. (b)–(j) Speckle patterns corresponding to different densities. Scale bar: 200 μm .

phenomenon, i.e., the appearance of the speckle pattern changes with respect to the density of the intralipid suspensions. When the density is low, the speckle grain has a shape like a thin stripe [Fig. 2(b)–2(e)]. The reasons for this can be addressed as follows. First, the density of fat emulsion is low so that only a few scattering events occur to the light that propagates through it. That is why the contrast of the speckle in this case is very high compared with Figs. 2(i) and 2(j). Second, the droplets are moving downward owing to the gravity of the Earth during the course of data acquisition. Finally, the camera plane may be tilted with respect to the incoming light [39]. As the intralipid density increases, many more scattering events occur as the light propagates through the tank. As a result, the speckle pattern is formed by the interference of light waves that experienced various times of scattering. The speckle grain becomes more isotropic, but the contrast is apparently reduced (to 0.08).

B. Decorrelation Time

As noted, a large number of fat droplets are randomly distributed in emulsion suspension. Owing to gravity and Brownian motion in the surrounding liquid [40], these droplet scatterers are not static but moving randomly, resulting in fluctuations of light propagation trajectory in the turbid suspensions. For analyzing the influence of particle motion on the light scattering in this case, one usually attempts to measure the temporal correlation by using diffusing wave spectroscopy (DWS) [40–42]. In this paper, we propose an alternative method to characterize the temporal correlation property of the intralipid diluents with different concentrations in the 33.6 cm thick glass tank.

As illustrated in Fig. 3, when the scatterers move randomly during time interval $\Delta\tau$, the light propagation path, including multiple scattering events, will undergo unpredictable changes. It is obvious that the path fluctuation reflects the motion of the scatterer. Intuitively, we need to evaluate the temporal correlation between electric fields $E(t_0 + \Delta\tau)$ and $E^*(t_0)$ at point r (with a source placed at r_0), which are closely related with the paired trajectories, as represented in Fig. 3. However, since traditional image sensors only record intensity, it is in principle not possible to directly detect the electric fields. According to the Siegert relation [43], the intensity correlation function can be associated with the electric field correlation function as

$$\frac{\langle I(t_0)I(t_0 + \Delta\tau) \rangle}{\langle I(t_0) \rangle \langle I(t_0 + \Delta\tau) \rangle} - 1 = \beta \left| \frac{\langle E(t_0 + \Delta\tau)E^*(t_0) \rangle}{\langle E(t_0 + \Delta\tau) \rangle \langle E^*(t_0) \rangle} \right|^2, \quad (4)$$

where β is an experimental factor between 0 and 1. Obviously, the temporal correlation property of the nonstatic medium can

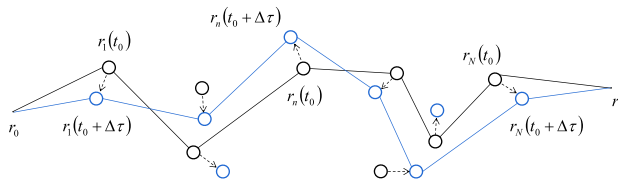


Fig. 3. Multiple scattering trajectories in dynamic media. In this illustration, scatterers move from the black circle to the blue circle during time interval $\Delta\tau$, and r_i ($i = 1, \dots, n, \dots, N$) represents the location where scattering event occurs.

be characterized by using intensity-only measurements and calculate the intensity correlation between the speckle patterns taken at time t_0 and at a later time $t_0 + \Delta\tau$.

The experimental system we designed to measure the temporal correlation of the speckle patterns is shown in Fig. 4. A semiconductor pumped laser emitted at 532 nm (MLL-S-532-A, Changchun New Industries Optoelectronics Tech. Co., Ltd.) was used as the incident light to illuminate the intralipid suspensions. The scattered light was divided into two arms by a beam splitter (BS) and collected, respectively, by two cameras (GS3-U3-23S6M-C, Point Grey). The dual cameras were controlled by an external trigger to achieve precise adjustment of the time lag of the two camera shutters. As shown in Fig. 4(a), the external control board Arduino generated two rectangular signals with a time interval $\Delta\tau$ to trigger the two cameras, each of which acquired a speckle pattern, $I(t_0)$ and $I(t_0 + \Delta\tau)$. This allows us to calculate the correlation coefficient between these two speckle patterns:

$$C(\Delta\tau) = \frac{\overline{\delta I(t_0) \cdot \delta I(t_0 + \Delta\tau)}}{\overline{\delta I(t_0)} \cdot \overline{\delta I(t_0 + \Delta\tau)}}, \quad (5)$$

where $\delta I = I - \bar{I}$, and \bar{I} is the mean of I . This function is basically identical with the temporal intensity correlation on the left-hand side of Eq. (4). By controlling the time delay $\Delta\tau$ via the Arduino controller, it is therefore possible to calculate the curve of $C(\Delta\tau)$ with respect to different $\Delta\tau$, manifesting the decorrelation behavior of the suspensions. In our experiment, the Arduino controller is set to trigger the two cameras with a minimum time lag of 100 μ s. In this way, we can measure the decorrelation behaviors of the intralipid suspensions with different concentrations. The experimental results are plotted in Fig. 5.

In order to determine the decorrelation time, we need to fit the data of $C(\Delta\tau)$ to a theoretical model. Note that there are various contributions to the correlation function that are determined by the different sets of multiple scattering trajectories, which are classified in terms of the crossing number of two diffusive paths [44]. Each contribution has discrepant time-dependence owing to different pairings of two trajectories obtained at time t_0 and $t_0 + \Delta\tau$, respectively. One dominant contribution, $C^{(1)}$, the one without crossings, decreases exponentially with $\Delta\tau$. Further, it specifies the memory effect [12,13]. Another considerable contribution, $C^{(2)}$, has one crossing, which decreases algebraically [45]. Generally speaking, for one more path crossing, the corresponding

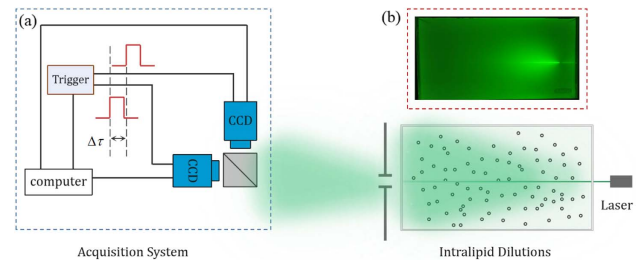


Fig. 4. Experiment setup. (a) Dual camera acquisition system. (b) Experimental site map of intralipid dilution: 11.47 L purified water (33.6 cm \times 19.5 cm \times 17.5 cm) and 2 mL intralipid 20%.

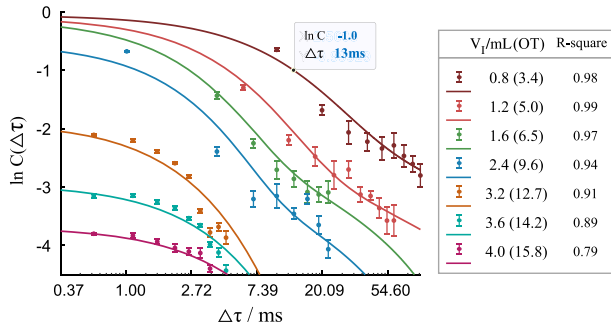


Fig. 5. Decorrelation curves for different concentrations of intralipid dilutions. The data points and the error bars represent the mean value and the standard error of the correlation coefficient calculated from 10 image pairs. The solid lines in different colors are the fitting results, and the corresponding intralipid volume V_I and optical thickness (OT) are shown in the legend. Here, the coefficient of determination (R -square) is used to describe the goodness of fit. Note that the horizontal axis is logarithmic scale.

contribution is reduced by a factor g , which is a large dimensionless number with a value at the order of 10^2 for a visible light in liquid suspensions [44,45]. As the contributions of other correlation terms with multiple crossings can be neglected, we only consider the effect of $C^{(1)}$ and $C^{(2)}$ here. Specifically, these two terms can be expressed as [19]

$$C^{(1)}(\Delta\tau) = \left[\frac{L/L_s}{\sinh(L/L_s)} \right]^2, \quad (6)$$

$$C^{(2)}(\Delta\tau) = \frac{1}{g} \frac{1}{\sinh^2(L/L_s)} \left[\frac{\sinh(2L/L_s)}{2L/L_s} - 1 \right], \quad (7)$$

where L is the geometric thickness of sample, and

$$L_s = \sqrt{D\tau_e} \cdot f(\Delta\tau), \quad (8)$$

where D is the diffusion coefficient that describes wave diffusion and is related to the elastic mean free path l_e , τ_e is the elastic collision time, and

$$f(\Delta\tau) = \left[\frac{e^{-\Delta\tau/(2\tau_b)}}{1 - e^{-\Delta\tau/(2\tau_b)}} \right]^{1/2}, \quad (9)$$

where τ_b describes the Brownian motion of the scatterers. One should note that τ_b is the characteristic time for a scatterer to move a distance in the order of wavelength λ .

We fit the experimental data with the weighted sum of $C^{(1)}$ and $C^{(2)}$ defined in Eqs. (6) and (7), respectively, with the fitting parameters a , b , m , and n as

$$C(\Delta\tau) = a \left\{ \frac{m/f(n \cdot \Delta\tau)}{\sinh[m/f(n \cdot \Delta\tau)]} \right\}^2 + b \frac{1}{\{\sinh[m/f(n \cdot \Delta\tau)]\}^2} \left\{ \frac{\sinh[2m/f(n \cdot \Delta\tau)]}{2m/f(n \cdot \Delta\tau)} - 1 \right\}. \quad (10)$$

Apparently, the parameters a and b represent the related contributions of $C^{(1)}$ and $C^{(2)}$ to C . And m is a parameter related to $L/\sqrt{D\tau_e}$, while n is associated with τ_b . The fitting results are plotted as solid lines in Fig. 5. From these fitted curves, one can

easily determine the decorrelation time τ_d , which is defined as the time where the correlation coefficient drops to $1/e$ of its maximum. We found that τ_d for the suspensions with 0.8, 1.2, 1.6, and 2.4 mL intralipid 20% are 13, 4.8, 2.6, and 1.2 ms, respectively. Unfortunately, owing to the light being scattered too many times when we applied 3.6 mL intralipid or more in the tank, the decorrelation time drops dramatically, and the cameras are not fast enough to capture this process. We can conclude that the decorrelation time in this cases must be less than 100 μ s, which is the shutter speed limit of our cameras. Therefore, we assume that the decorrelation time of the turbid suspensions is around several to several tens of microseconds, corresponding to an optical depth of around 16.

4. EXPERIMENTAL DEMONSTRATIONS

A. Experimental Data Acquisition

For the demonstration of the proposed method, we developed an experimental system, schematically shown in Fig. 1, to acquire all the data. In the first experiment, the optical thickness of the fat emulsion suspensions was 8 (2 mL intralipid 20%). In the second set of experiments, the optical thickness of the suspensions was about 16 (4 mL intralipid 20%), and decorrelation time was at the order of microseconds.

Here, we propose an end-to-end deep neural net to reconstruct object images from the acquired data. For convenience, we used 10,050 images altogether from the MNIST data set [46] as the training and test data. We first resized the original MNIST images to 512×512 pixels and displayed them sequentially onto the central pixels of the 1024×1080 pixel SLM. An incoherent LED light was used to illuminate the system. Reflected from the SLM, the incoherent light was then projected onto the front surface of the tank by using a delayed optic. As the light transmitted from the front surface all the way through the turbid suspensions to the back surface of the tank, it encountered many scattering events, as discussed in Section 3. The scattered patterns at the back surface of the tank were then captured by the sCMOS camera with an imaging optic. Note that there was also strong interference noise produced by an ambient light, with SIR = -17 dB, as described by Eq. (2). Only the central 512×512 out of the 1024×1024 pixels of the raw image taken by the camera were used in the reconstruction process. Our previous study [29] shows that this does not affect the quality of the reconstructed image. Owing to the limited frame rate of the SLM and the sCMOS camera, it took about 3.5 h to capture all 10,050 scattered patterns. Then, we partitioned them exclusively into two groups, one with 9900 patterns as the training set after pairing up with their corresponding ground true images (those displayed on the SLM), and the other as the test set. Typical visible features of the scattered patterns are shown in Fig. 1.

B. Proposed Learning-Based Method for Image Reconstruction

The objective, then, is to reconstruct the object I_o from the corresponding recorded scattered pattern I_s . For the supervised training of a neural network, we need to construct a training set by pairing up every MNIST image in the training set together with the corresponding scattering pattern I_s ,

i.e., $S_{\text{training}} = \{(I_s^{(1)}, I_o^{(1)}), (I_s^{(2)}, I_o^{(2)}), \dots, (I_s^{(N)}, I_o^{(N)})\}$. Then, we can convert the problem to learning the probability distribution $P(I_o|I_s)$ [47] or the mapping relation $R_{\text{learn}}: I_s \rightarrow I_o$ from the training set. Once it is learned, an object I_o in the test set should be predicted from the corresponding scattered pattern I_s [29]. The objective function then can be defined as

$$\mathcal{R}_{\text{learn}} = \arg \min_{\mathcal{R}_\theta, \theta \in \Theta} \sum_{n=1}^N \mathcal{L}(I_o^{(n)}, \mathcal{R}_\theta\{I_s^{(n)}\}) + g(\theta), \quad (11)$$

where the set Θ contains handcraft parameters that specify the network structure and weighting parameters that are automatically adjusted during the training process [48], $\mathcal{L}(\cdot)$ is the loss function that evaluates the error between label I_o^n and predicted output $\mathcal{R}_\theta\{I_s^n\}$, and $g(\theta)$ is a regularizer defined on the parameters with the aim of avoiding overfitting [20].

We adopted the convolutional neural network (CNN) [46] to solve the problem described by Eq. (11). Refer to Appendix A for more details about the structure of the neural net we designed.

To train the neural net, we must specify the loss function in Eq. (11). In this work, we adopted the mean square error (MSE) as the loss function to evaluate the predicted network outputs and ground truth images:

$$\text{MSE} = \min \frac{1}{WHN_1} \sum_{n=1}^{N_1} \sum_{(u,v)}^{W,H} [I_p^{(n)}(u,v) - I_o^{(n)}(u,v)]^2, \quad (12)$$

where $I_p^{(n)}(u,v)$ is defined as the reconstructed image from the n th scattered pattern $I_s^{(n)}(u,v)$, $I_o^{(n)}(u,v)$ is the corresponding ground-truth resized MNIST image, W and H are the width and height of the reconstructed image, and N_1 is the batch size, which was set to 5 in this work. We used the stochastic gradient descent method [49] to evaluate the MSE and Adam [50] to adjust the weights. The training step number was set as 30,000; in practice, however, we were able to use less training steps.

C. Experimental Results

The experimental results are shown in Figs. 6(b) and 6(c). In comparison with the corresponding ground truth images shown in Fig. 6(a), the MNIST images are reconstructed successfully. We use the root of mean square error (RMSE) and structural similarity index (SSIM) [51] to evaluate reconstructed images quantitatively. These functions are defined as

$$\text{RMSE} = \left\{ \frac{1}{WH} \sum_{(u,v)}^{W,H} [I_p(u,v) - I_s(u,v)]^2 \right\}^{\frac{1}{2}}, \quad (13)$$

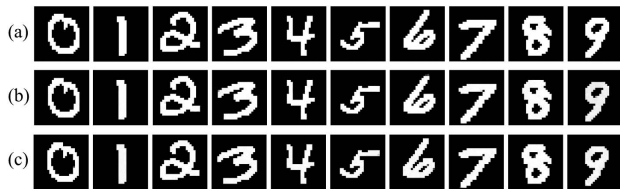


Fig. 6. Experimental results. (a) Ground truths, and the reconstructed images in the case that the optical thickness equals (b) 8 and (c) 16, respectively.

$$\text{SSIM} = \frac{(2\mu_{I_p} \mu_{I_s} + c_1)(2\sigma_{I_p I_s} + c_2)}{(\mu_{I_p}^2 + \mu_{I_s}^2 + c_1)(\sigma_{I_p}^2 + \sigma_{I_s}^2 + c_2)}, \quad (14)$$

where μ_I and σ_I^2 are the mean and variance of the image I , $\sigma_{I_p I_s}$ is the covariance between I_p and I_s , and c_1 and c_2 are regularization parameters. Basically, the RMSE measures the averaged difference between the reconstructed and corresponding ground truth images, whereas the SSIM evaluates the structure similarity between the two images. We calculate the RMSE and SSIM of all the 150 test images. The averaged values are 5.25 and 0.99, respectively.

As mentioned above, we also performed a second set of experiments, in which the optical thickness of the turbid suspension is 16. In this case, we must block out a significant part of the large-angle scattered light by placing two apertures between the back surface of the tank and the camera. The captured scattered patterns then look dimmed, as shown in panel (2), Fig. 1. The corresponding reconstructed images from the proposed neural net are plotted in Fig. 6(c). With respect to the ground truths, the RMSE was about 4.66 and the SSIM was around 0.98, at the same level as those shown in Fig. 6(b).

D. Robustness Analysis

In this subsection, we concisely analyze the robustness of the proposed technique. The optical thickness of the turbid suspension is 16 in all the analysis here.

1. Robustness against the Movement of Object/Camera

The first situation we considered is how the change of the object/camera position in the test with respect to that in the training affects the reconstructed image without retraining the neural net. The experimental results are shown in Figs. 7 and 8. In Fig. 7, we consider the separated effect of the change of position of the object and the camera, both along the diagonal and horizontal directions. One can see that, in all the cases, the images are reconstructed with the SSIM value better than 0.8. The proposed method seems to be more robust to the change of the object position owing to the scattering of light from the object. The RMSE values do not reflect the robustness

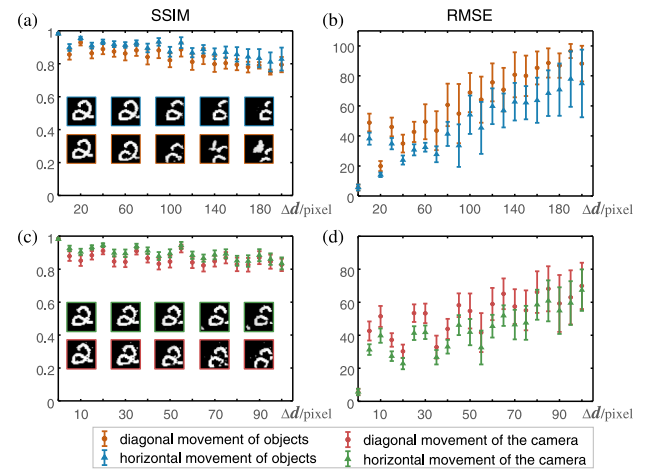


Fig. 7. Robustness against the position change of the object/camera. Δd is the displacement of the object/camera (in pixel). The data points and the error bars represent the mean values and the standard deviations of the SSIM/RMSE of 10 reconstructed images (digits '0–9').

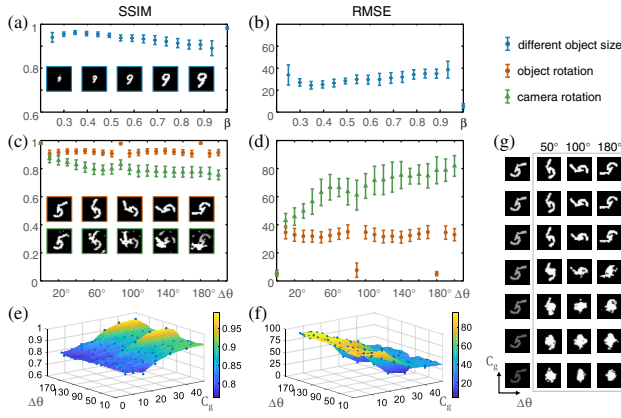


Fig. 8. Robustness against the scaling and rotation of the object/camera. β is the scaling factor of the image size, $\Delta\theta$ the rotation angle, and C_g the image contrast gradient. The data points and the error bars in (a)–(d) represent the mean values and the standard deviations of the SSIM/RMSE of 10 reconstructed images (digits ‘0–9’). (e) and (f) SSIM/RMSE of digit ‘5’ with respect to $\Delta\theta$ and C_g . (g) Visualized reconstructed digits.

in these cases because of the shifting of the reconstructed image, as shown in the insets of Figs. 7(a) and 7(c). In addition, artefacts arise when the object is shifted too far away, as can be seen in the insets in Fig. 7(a).

The robustness to the scaling and rotation of the object/camera is shown in Fig. 8. In both cases, the image can be reconstructed well. Again, it is more robust to the change in the object (both scaling and rotation) owing to the reason mentioned above.

2. Generalization to Nondigit Objects

We also used the trained neural net to reconstruct nondigit objects from the corresponding scattered patterns. The results are shown in Fig. 9. Two types of objects were used: virtual objects, which were displayed on the SLM [Fig. 9(a)], and a physical USAF target [Fig. 9(b)]. The experimental results



Fig. 9. Reconstruction of nondigit objects with the neural network trained by using digits. (a) First and third rows are the ground truths; second and fourth are the corresponding reconstructed images. (b) Reconstructed USAF target and the highlight of some of its portions.

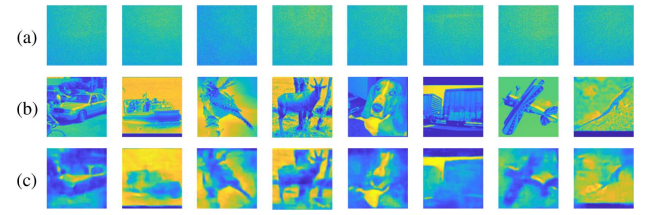


Fig. 10. Experimental results with natural scene object. (a) Scattered patterns. (b) Corresponding ground truth. (c) Reconstructed results.

suggest that the proposed technique can be used to reconstruct objects that can be far from those that were used to train the neural net.

3. Reconstruction of Nature Images

However, the neural network has to be retrained if the object is more complex such as nature scenes that are in gray scale and have complex structures. However, it is still possible to reconstruct the image with a good fidelity. For demonstration, we tried the proposed method on a more complex set of objects. We took the STL-10 data set [52] in our experiment. Specifically, we used 19,000 images to retrain the network model and other 1000 images for testing. Some of the typical results are shown in Fig. 10. The averaged RMSE and SSIM values are 36.14 and 0.63, respectively.

5. CONCLUSION

In conclusion, we have demonstrated an end-to-end deep-learning-based incoherent imaging method through optically thick and dynamic scattering media under a condition of strong interference noise ($SIR = -17$ dB). The proposed method allows direct reconstruction of an object image from the captured scattered pattern with high quality (with an averaged RMSE around 5 and averaged SSIM about 0.86). We have quantitatively measured the optical thickness and decorrelation time of the turbid suspension [which was made by mixing fat emulsion (intralipid 20%) with purified water], with respect to the intralipid concentrations, and calibrated the optical thickness of the tank of intralipid suspensions in our experiments, which is about 16 in the severest case. Although the results were obtained in the lab environment, it is expected that the proposed method can be used in a wider scope of other scattering environments.

APPENDIX A: CNN STRUCTURE

The proposed neural network structure is based on the one we published previously in Ref. [22] and shown Fig. 11. As can be seen, the neural network has a typical U-net structure [53] but with independent branches, which are designed to learn the training set features at different scales.

In our experiments, the central 512×512 pixels of each captured scattered pattern I_s have to be preprocessed before input into the neural network. These images were first processed by three sets of convolution blocks and a max-pooling layer, resizing the image size to 64×64 pixels. Then, the network was divided into four independent paths, each of which equips with a max-pooling layer to implement one, two, four, and

eight times downsampling of the incoming images, respectively, as indicated by the downward arrows in Fig. 11(a). For each data flow path, after the pooling layer, images were put to four identical residual blocks and an appropriate number of upsampling blocks, which was designed to make image size as 64×64 pixels. Next, these independent paths were concatenated into an intact image, which contains 192 feature maps. It was then followed by a series of convolution blocks and three upsampling blocks to resize the output image to 512×512 pixels. Finally, a convolution block was used to reduce the output channels to 1, so that the output results were gray-level images. The final output of our neural network was the reconstructed images I_o when the network was well-trained.

Now, we present a detailed description of the proposed neural network. There are mainly three types of functional blocks in the neural network: the convolution block, the residual block, and the upsampling block, as shown in Fig. 11(b). Compared with the neural network in Ref. [22], a batch normalization layer [54] is added into the convolution and residual blocks, so as to accelerate learning speed and be less sensitive to initialization. Furthermore, it is also expected to act as a regularizer, as described in Eq. (11). In our implementation, the convolution block contains a convolutional layer, a batch normalization layer, and an activation function. The convolution layer is the core layer to build a convolutional neural network and achieves feature extraction at different scales. The activation function used here is the rectified linear units. As shown in Fig. 11(a), a max-pooling layer is periodically inserted between

successive convolution layers, and its function is to gradually reduce the spatial dimension of the data and the number of internal weights required for image reconstruction. Here, max pool is with 2×2 filters and stride 2, so the image size is reduced by a factor of 2. The residual block includes two consecutive groups consisting of a convolutional layer, a batch normalization layer, and an activation function. The shortcut connections between the input and the output enable us to optimize the neural network and improve accuracy by adding considerable depth. In addition, the upsampling block is composed of a transposed convolution layer and an activation function, which is to achieve image enlargement and convolution decoding. The stride of the transposed convolution layer is set as 2 to double the image size by zero-padding in between two adjacent pixels. Moreover, the detailed input channel numbers and output channel numbers of hidden layers in the neural network are represented as 1-16, 16-32, etc., as shown under the block symbols in Fig. 11(a). The size of convolutional kernels is presented as (5, 5) or (3, 3) below the channel numbers.

Funding. Chinesisch-Deutsche Zentrum für Wissenschaftsförderung (GZ1391); National Natural Science Foundation of China (61991452, 62061136005); Chinese Academy of Sciences Key Project (QYZDB-SSW-JSC002).

Acknowledgment. We thank the anonymous reviewers for their constructive comments.

Disclosures. The authors declare no conflicts of interest related to this paper.

REFERENCES

1. L. Wang, P. P. Ho, C. Liu, G. Zhang, and R. R. Alfano, "Ballistic 2-D imaging through scattering walls using an ultrafast optical Kerr gate," *Science* **253**, 769–771 (1991).
2. Q. Z. Wang, X. Liang, L. Wang, P. P. Ho, and R. R. Alfano, "Fourier spatial filter acts as a temporal gate for light propagating through a turbid medium," *Opt. Lett.* **20**, 1498–1500 (1995).
3. S. G. Demos and R. R. Alfano, "Optical polarization imaging," *Appl. Opt.* **36**, 150–155 (1997).
4. E. N. Leith, C. Chen, H. Chen, Y. Chen, J. Lopez, P.-C. Sun, and D. Dilworth, "Imaging through scattering media using spatial incoherence techniques," *Opt. Lett.* **16**, 1820–1822 (1991).
5. I. M. Vellekoop, "Feedback-based wavefront shaping," *Opt. Express* **23**, 12189–12206 (2015).
6. Z. Yaqoob, D. Psaltis, M. S. Feld, and C. Yang, "Optical phase conjugation for turbidity suppression in biological samples," *Nat. Photonics* **2**, 110–115 (2008).
7. M. Cui and C. Yang, "Implementation of a digital optical phase conjugation system and its application to study the robustness of turbidity suppression by phase conjugation," *Opt. Express* **18**, 3444–3455 (2010).
8. S. M. Popoff, G. Lerosey, R. Carminati, M. Fink, A. C. Boccara, and S. Gigan, "Measuring the transmission matrix in optics: an approach to the study and control of light propagation in disordered media," *Phys. Rev. Lett.* **104**, 100601 (2010).
9. M. M. Qureshi, J. Brake, H.-J. Jeon, H. Ruan, Y. Liu, A. M. Safi, T. J. Eom, C. Yang, and E. Chung, "In vivo study of optical speckle decorrelation time across depths in the mouse brain," *Biomed. Opt. Express* **8**, 4855–4864 (2017).
10. D. Wang, E. H. Zhou, J. Brake, H. Ruan, M. Jang, and C. Yang, "Focusing through dynamic tissue with millisecond digital optical phase conjugation," *Optica* **2**, 728–735 (2015).

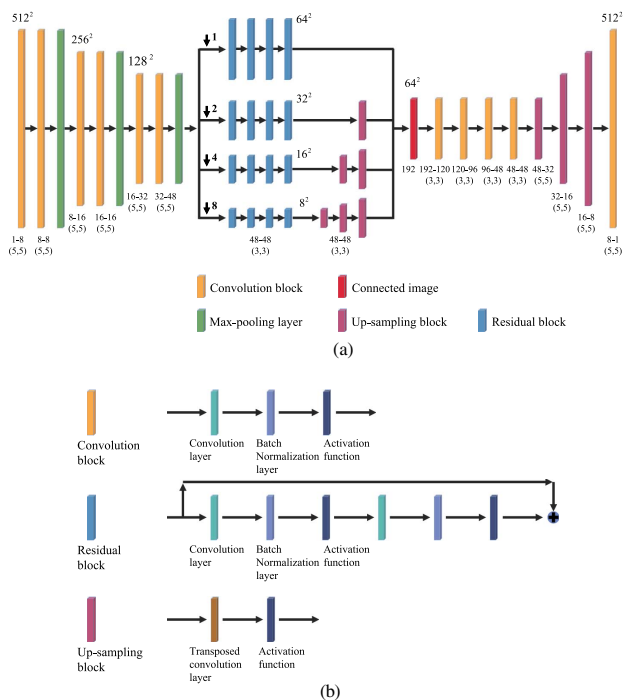


Fig. 11. Proposed neural network architecture. (a) Digits in the format $m-n$ below each layer denote the number of input channels m , and the number of output channels n . (5, 5) and (3, 3) denote the size of the convolution kernel in pixel counts. (b) Detailed information of neural network structure.

11. L. Yan, M. Cheng, S. Yuecheng, S. Junhui, and L. V. Wang, "Focusing light inside dynamic scattering media with millisecond digital optical phase conjugation," *Optica* **4**, 280–288 (2017).
12. S. Feng, C. Kane, P. A. Lee, and A. D. Stone, "Correlations and fluctuations of coherent wave transmission through disordered media," *Phys. Rev. Lett.* **61**, 834–837 (1988).
13. I. Freund, M. Rosenbluh, and S. Feng, "Memory effects in propagation of optical waves through disordered media," *Phys. Rev. Lett.* **61**, 2328–2331 (1988).
14. J. Bertolotti, E. G. van Putten, C. Blum, A. Lagendijk, W. L. Vos, and A. P. Mosk, "Non-invasive imaging through opaque scattering layers," *Nature* **491**, 232–234 (2012).
15. O. Katz, P. Heidmann, M. Fink, and S. Gigan, "Non-invasive single-shot imaging through scattering layers and around corners via speckle correlations," *Nat. Photonics* **8**, 784–790 (2014).
16. C. A. Metzler, F. Heide, P. Rangarajan, M. M. Balaji, A. Viswanath, A. Veeraraghavan, and R. G. Barani, "Deep-inverse correlography: towards real-time high-resolution non-line-of-sight imaging," *Optica* **7**, 63–71 (2020).
17. T. Dongliang, S. S. Kumar, T. Vinh, and D. Cuong, "Single-shot large field of view imaging with scattering media by spatial demultiplexing," *Appl. Opt.* **57**, 7533–7538 (2018).
18. G. Li, W. Yang, H. Wang, and G. Situ, "Image transmission through scattering media using ptychographic iterative engine," *Appl. Sci.* **9**, 849 (2019).
19. E. Akkermans and G. Montambaux, *Mesoscopic Physics of Electrons and Photons* (Cambridge University, 2007).
20. G. Barbastathis, A. Ozcan, and G. Situ, "On the use of deep learning for computational imaging," *Optica* **6**, 921–943 (2019).
21. Y. Rivenson, Y. Zhang, H. Günaydin, D. Teng, and A. Ozcan, "Phase recovery and holographic image reconstruction using deep learning in neural networks," *Light Sci. Appl.* **7**, 17141 (2018).
22. H. Wang, M. Lyu, and G. Situ, "eHoloNet: a learning-based end-to-end approach for in-line digital holographic reconstruction," *Opt. Express* **26**, 22603–22614 (2018).
23. A. Sinha, J. Lee, S. Li, and G. Barbastathis, "Lensless computational imaging through deep learning," *Optica* **4**, 1117–1125 (2017).
24. F. Wang, Y. Bian, H. Wang, M. Lyu, G. Pedrini, W. Osten, G. Barbastathis, and G. Situ, "Phase imaging with an untrained neural network," *Light Sci. Appl.* **9**, 77 (2020).
25. M. Lyu, W. Wang, H. Wang, H. Wang, G. Li, N. Chen, and G. Situ, "Deep-learning-based ghost imaging," *Sci. Rep.* **7**, 17865 (2017).
26. F. Wang, H. Wang, H. Wang, G. Li, and G. Situ, "Learning from simulation: an end-to-end deep-learning approach for computational ghost imaging," *Opt. Express* **27**, 25560–25572 (2019).
27. S. Li, M. Deng, J. Lee, A. Sinha, and G. Barbastathis, "Imaging through glass diffusers using densely connected convolutional networks," *Optica* **5**, 803–813 (2018).
28. Y. Li, Y. Xue, and L. Tian, "Deep speckle correlation: a deep learning approach toward scalable imaging through scattering media," *Optica* **5**, 1181–1190 (2018).
29. 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).
30. Y. Sun, J. Shi, L. Sun, J. Fan, and G. Zeng, "Image reconstruction through dynamic scattering media based on deep learning," *Opt. Express* **27**, 16032–16046 (2019).
31. J. Gua, Z. Wangb, J. Kuenb, L. Mab, A. Shahroudyb, B. Shuaib, T. Liub, X. Wangb, and G. Wang, "Recent advances in convolutional neural networks," *Pattern Recogn.* **77**, 354–377 (2018).
32. R. Michels, F. Foschum, and A. Kienle, "Optical properties of fat emulsions," *Opt. Express* **16**, 5907–5925 (2008).
33. F. Martelli and G. Zaccanti, "Calibration of scattering and absorption properties of a liquid diffusive medium at NIR wavelengths. CW method," *Opt. Express* **15**, 486–500 (2007).
34. J. T. Allardice, A. M. Abulafi, D. G. Webb, and N. S. Williams, "Standardization of intralipid for light scattering in clinical photodynamic therapy," *Laser Med. Sci.* **7**, 461–465 (1992).
35. P. D. Ninni, F. Martelli, and G. Zaccanti, "Intralipid: towards a diffusive reference standard for optical tissue phantoms," *Phys. Med. Biol.* **56**, N21–N28 (2011).
36. B. W. Pogue and M. S. Patterson, "Review of tissue simulating phantoms for optical spectroscopy, imaging and dosimetry," *J. Biomed. Opt.* **11**, 041102 (2006).
37. M. Raju and S. N. Unni, "Concentration-dependent correlated scattering properties of intralipid 20% dilutions," *Appl. Opt.* **56**, 1157–1166 (2017).
38. M. L. Dong, K. G. Goyal, B. W. Worth, S. S. Makkar, W. Calhoun, L. M. Bali, and S. Bali, "Accurate *in situ* measurement of complex refractive index and particle size in intralipid emulsions," *J. Biomed. Opt.* **18**, 087003 (2013).
39. H. T. Yura, S. G. Hanson, R. S. Hansen, and B. Rose, "Three-dimensional speckle dynamics in paraxial optical systems," *J. Opt. Soc. Am. A* **16**, 1402–1412 (1999).
40. G. Maret and P. E. Wolf, "Multiple light scattering from disordered media: the effect of Brownian motion of scatterers," *Z. Phys. B* **65**, 409–413 (1987).
41. M. J. Stephen, "Temporal fluctuations in wave propagation in random media," *Phys. Rev. B* **37**, 1–5 (1988).
42. D. J. Pine, D. A. Weitz, P. M. Chaikin, and E. Herbolzheimer, "Diffusing wave spectroscopy," *Phys. Rev. Lett.* **60**, 1134–1137 (1988).
43. P.-A. Lemieux and D. J. Durian, "Investigating non-Gaussian scattering processes by using *n*th-order intensity correlation functions," *J. Opt. Soc. Am. A* **16**, 1651–1664 (1999).
44. F. Scheffold and G. Maret, "Universal conductance fluctuations of light," *Phys. Rev. Lett.* **81**, 5800–5803 (1998).
45. F. Scheffold, W. Hartl, G. Maret, and E. Matijevic, "Observation of long-range correlations in temporal intensity fluctuations of light," *Phys. Rev. B* **56**, 10942–10952 (1997).
46. Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE* **86**, 2278–2324 (1998).
47. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT, 2016).
48. D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature* **323**, 533–536 (1986).
49. T. S. Ferguson, "An inconsistent maximum likelihood estimate," *J. Am. Stat. Assoc.* **77**, 831–834 (1982).
50. K. Diederik and J. Ba, "Adam: a method for stochastic optimization," arXiv:1412.6980 (2014).
51. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.* **13**, 600–612 (2004).
52. A. Coates, A. Y. Ng, and H. Lee, "An analysis of single-layer networks in unsupervised feature learning," *J. Mach. Learn. Res.* **15**, 215–223 (2011).
53. O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)* (Springer, 2015), pp. 234–241.
54. S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," arXiv:1502.03167 (2015).