Learning-based lensless imaging through optically thick scattering media

Meng Lyu,[a,b,†] Hao Wang,[a,b,†] Guowei Li,[a,b] Shanshan Zheng,[a,b] and Guohai Situ[a,b,*]

[a]Chinese Academy of Sciences, Shanghai Institute of Optics and Fine Mechanics, Shanghai, China
[b]University of Chinese Academy of Sciences, Center for Materials Science and Optoelectronics Engineering, Beijing, China

Abstract. The problem of imaging through thick scattering media is encountered in many disciplines of science, ranging from mesoscopic physics to astronomy. Photons become diffusive after propagating through a scattering medium with an optical thickness of over 10 times the scattering mean free path. As a result, no image but only noise-like patterns can be directly formed. We propose a hybrid neural network for computational imaging through such thick scattering media, demonstrating the reconstruction of image information from various targets hidden behind a white polystyrene slab of 3 mm in thickness or 13.4 times the scattering mean free path. We also demonstrate that the target image can be retrieved with acceptable quality from a very small fraction of its scattered pattern, suggesting that the speckle pattern produced in this way is highly redundant. This leads to a profound question of how the information of the target being encoded into the speckle is to be addressed in future studies.

Keywords: imaging through scattering media; deep learning; neural network; computational imaging.

Received Nov. 26, 2018; accepted for publication May 16, 2019; published online Jun. 18, 2019.

1 Introduction

Conventional optical imaging techniques work under the condition that spherical wavelets emitting from a target to the entrance pupil cannot be severely distorted.[1] Otherwise, they cannot converge after passing through the lens system and form the image. However, in applications ranging from *in situ* biomedical inspection in life sciences[2] to free space optical communication in cloudy and hazy environments,[3] this condition is not satisfied when scattering media exist between the target and the optical system, distorting the propagation direction of each wavelet in a random manner. When the distortion is weak, one can use adaptive optical techniques to measure it and compensate.[4] However, when the distortion is strong, no image can be formed as the photons are scattered many times randomly inside the medium. The photon-scattering process is explicitly deterministic (described by the so-called scattering matrix) when the scattering medium is static, meaning that it is reversible.[5] Taking advantage of this property, many efforts have been made to exploit imaging through scattering media using either optical or computational techniques. In the optical domain, Goodman et al. have demonstrated that one can measure the scattered wavefront distortion by using holography.[6] Holography has been extended to produce a phase-conjugated beam of the scattered one[7] so that the distortion can be removed when the beam propagates back through the scattering medium once again, and to separate the ballistic photons by taking the advantage of coherence gating.[8] Furthermore, it has been demonstrated that holography has the potential to image a three-dimensional object behind a thin diffuser.[9] Nevertheless, holography only records the complex wavefront of the beam that leaves the scattering medium, without any information about the scattering process occurring inside. Alternatively, wavefront shaping has been proposed to actively control the propagation of the beam through the scattering medium.[10–12] This can be phenomenologically interpreted as the control of the photons so that a certain amount of them can make their way through the scattering medium in the eigenchannels. In this way, wavefront shaping can be used to measure the transmission matrix (TM) of scattering media,[13] allowing the transmission of images through opaque media.[14] However, the state-of-the-art spatial light modulators (SLMs) can neither shape the complex-valued wavefront precisely nor measure the TM at the wavelength scale

*Address all correspondence to Guohai Situ, E-mail: ghsitu@siom.ac.cn
†Meng Lyu and Hao Wang contributed equally to this work

of a scattering medium of large volume due to the pixel size and count, which imposes a limitation to the powerful wavefront-shaping technique. Memory effect,[15] on the other hand, promises noninvasive[16] and single-shot imaging through a scattering layer,[17,18] and has been widely explored recently.[19] However, the memory effect range of the light field that passes through the scattering slab drops very quickly to the wavelength scale as the slab becomes thicker,[20] imposing a strict limitation on the field of view of the scattering imaging system.[17,21]

Here, we exploit the technique of deep learning and develop a special hybrid neural network (HNN) architecture for imaging through optically thick scattering media. It is well known that machine learning techniques including deep learning have been widely employed to solve the problems in recognition and classification.[22,23] It was not until recently that researchers started to use them in the field of computational imaging.[24–32] Horisaki et al.[24] proposed to use the support vector regression architecture to reconstruct the scattering images with a set of human face data for training. Lyu et al.[25] were the first to use deep learning for imaging through scattering media. Li et al.[30] proposed to use a convolutional neural network (CNN) to achieve coherent imaging through thin ground glass, and Li et al.[31] proposed to use a residual neural network that is highly scalable to both scattering medium perturbations and measurement requirements. However, the works reported in Refs. 29 and 30 dealt with thin scattering layers like ground glass, through which the relationship between the acquired speckle and the target is easy to model owing to the memory effect. This is not the case when the scattering media become optically thick. In a previous study,[25] we have demonstrated how to solve this problem with deep neural network (DNN) architecture. But there were three big issues when we applied it in imaging through a thick scattering wall. First, in a DNN, every neuron in one hidden layer is usually connected to every other neuron in the immediate downstream layer, resulting in a parametric space that is time-consuming to optimize. Second, even though the training can converge eventually, overfitting is very likely to occur.[25,33] Third, the trained network is hard to generalize. In this paper, we demonstrate that the above drawbacks can be overcome using an HNN model. In the conceptual demonstration, we develop an HNN that can retrieve the coherent images of various targets behind a 3-mm-thick white polystyrene slab from the corresponding speckle pattern acquired with a digital camera. Furthermore, our experimental results suggest that the speckle pattern contains redundant information, as the target image can be reconstructed from as little as 0.1% information content possessed by the acquired speckle pattern. This suggests great potential for the compression and storage of the information-carrying speckle patterns and their transmission through wired or wireless communication channels.

2 Method

2.1 Experimental Setup

The optical setup used in our experiments is schematically illustrated in Fig. 1(a). The laser we used was a linearly polarized optically pumped semiconductor laser (Verdi G2 SLM, Coherent Inc.) irradiating at 532 nm. A collimated and expanded laser beam was shone on an amplitude-only liquid crystal SLM after passing through the linear polarizer P1 and Iris 1, which were used to control the polarization state and width of the beam. The amplitude-only SLM was a Holoeye Pluto-Vis with a maximum pixel count of 1920×1080 and a pixel size of $8.0~\mu m \times 8.0~\mu m$. The target images were displayed on the SLM to modulate the incident beam. The targets we used were images of handwritten digits from the MNIST database[34] with a size of 28×28 pixels. We first resized these images 18 times to 504×504 and zero-padded them to 1920×1080 before displaying them on the SLM. As a result, the physical size of the handwritten digit was about $2~mm \times 2~mm$ (calculated by counting the number of pixels used to represent the digits). The image-carrying beam reflected from the SLM was then shone onto a 3-mm-thick white polystyrene slab (EDU-VS1/M, Thorlabs Inc.), after a distance $d_1 = 290~mm$ of propagation in free space. At the other side of the slab, a camera (Andor Zyla 4.2 PLUS sCMOS with a 4.2-megapixel sensor) was placed at a distance $d_2 = 35~mm$ to collect the speckle pattern generated by light scattering inside the slab. To obtain amplitude modulation, we placed a linear polarizer P2 whose polarization direction was perpendicular with respect to P1 in the reflected beam. An iris (Iris 2) was also used, to select the first-order reflection. In the experiments, we activated only the central 512×512 pixels of the camera for speckle data acquisition.

2.2 Challenges of the Problem

To visually display the scattering power of the 3-mm-thick polystyrene slab, we first project an image directly on the front surface of the slab, and see how it looks at the back surface. The results are shown in Figs. 1(b) and 1(c), respectively. As can be seen, the image is very clear at the front surface but becomes highly diffusive at the other side of the slab, in contrast to the case of passing through thin ground glass.[16,17] This is because the light has been scattered many times on the way through the slab. Indeed, the measured optical depth of the scattering slab is about 13.40 (see Appendix A), so the light is in the diffusive regime.[2]

Nevertheless, the speckle pattern I captured by the camera is related to the target O through the TM.[35] Our purpose is then to retrieve the target image O from the speckle pattern I. Note that, for weak scattering layers such as a ground glass, their relation is reduced to the convolution $U(R) = \int O(r)h(r;R)d^2r$, where R measures the position on the surface of camera sensor and r measures the position on the surface of the SLM, within the isoplanatic angle. The captured speckle pattern $I(R) = U(R)U(R)^*$. The memory effect[16] allows us to reconstruct the target image by phase retrieval because the autocorrelation of I is approximately equal to the autocorrelation of $|O|^2$. However, when the scattering medium is sufficiently thick, as in our study, the memory effect is destroyed and the isoplanatic angle tends to zero.[36] This means that the spherical wavelets emitted from any two points of a target, although separated by a very small distance, will effectively experience two distinct scattering processes, resulting in two elemental speckle patterns that are totally uncorrelated. One can imagine that the final speckle pattern corresponding to the whole target is produced by the superposition of many of these uncorrelated elemental patterns. As a consequence, one will not expect that an image of the target can be reconstructed using the memory-effect-based techniques. In the case of imaging through static scattering media, one may choose to use TM-based techniques,[13,14] because the TM is deterministic. However, it is unfeasible in
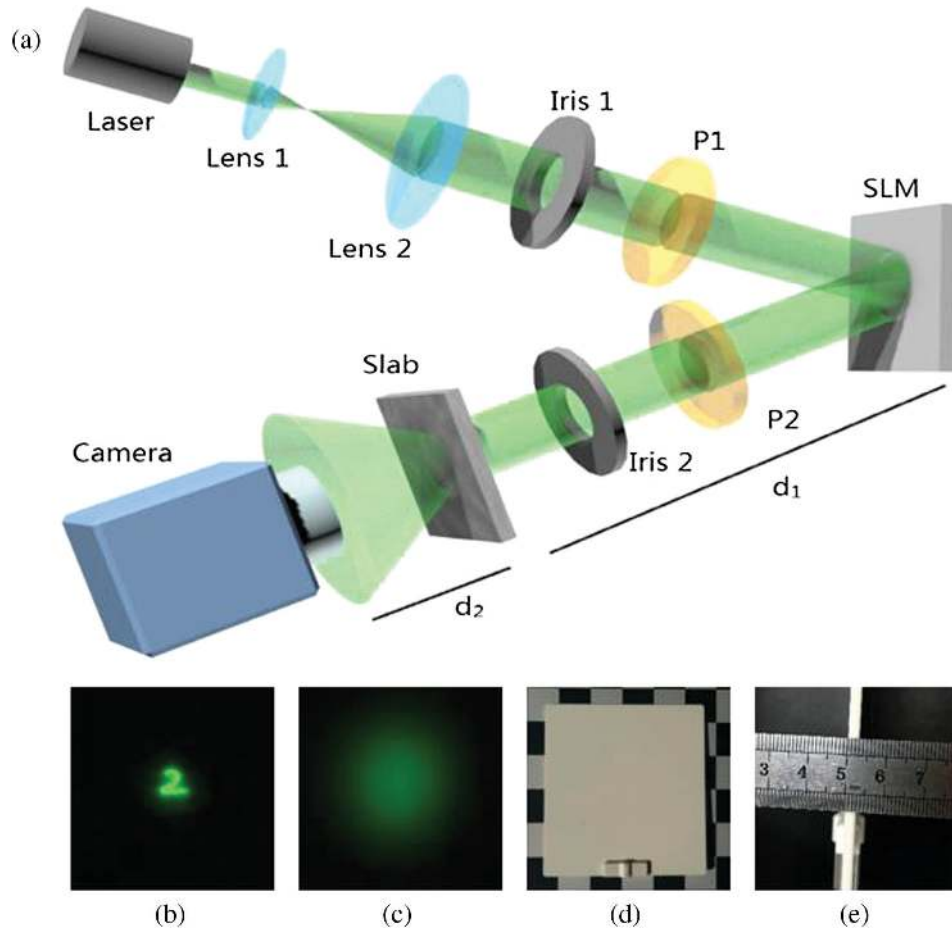
Fig. 1 (a) Experimental setup for imaging through scattering media, SLM represents an amplitude-only SLM, P1 and P2 are linear polarizers and the slab is a 3-mm-thick white polystyrene. The images captured at the (b) front and (c) back surfaces of the scattering medium. (d) The side view and (e) the top view of the polystyrene.

our case here, because the TM should be measured in the wavelength scale and across a large area, whereas current SLM technologies are not ready for that yet.

2.3 HNN Model

The deep-learning-based method that we propose here to image through optically thick media is a data-driven solution. This means that we need a large set of input–output measurements to train the neural network model so as to learn how the speckle patterns are related to the targets.

Even though the universal approximation theorem[37,38] states that a two-layer network can uniformly approximate any continuous function to arbitrary accuracy, such a two-layer network will require a huge amount of neurons, making it difficult to train. Therefore, a neuron network with more layers and fewer neurons in each layer is more favorable. Note that a typical CNN architecture is translation equivariance.[22] But the system we are encountering now becomes spatially variant due to the presence of the thick scattering medium. As a result, attempting to build a spatially invariant neural network to mimic a spatially variant imaging system will take a lot of effort. On the contrary, a fully connected network such as the DNN can be spatially variant. But the training is usually time-consuming and the resulting

network architecture is hard to generalize as aforementioned.[25,33] It typically takes many hours or even days to train depending on the size of the training set and the geometry of the network. For example, the fully connected DNN model that we developed for imaging through scattering[25] took us 18 h to train with a Tesla K20c GPU. Thus, an HNN composed of a few layers of DNN and a few layers of CNN is more desirable.

The HNN architecture we report here has five two-dimensional (2-D) convolutional layers, two reshaping layers, one pooling layer, three densely connected layers, and five dropout layers, as schematically shown in Fig. 2. The three densely connected layers were sandwiched between the five convolutional layers, two upstream and three downstream. For demonstration, we arbitrarily took a block of 64×64 pixels out of each 512×512-pixels speckle pattern acquired by the camera as the input to the network. The two upstream layers were used to extract 32 feature maps of the input speckle, resulting in a $32 \times 32 \times 32$-pixel feature cube. Then it was reshaped to a $1 \times 32,768$ vector and served as the input to the three sequential densely connected layers, which has 1024, 784, and 784 neurons, respectively. Then another reshaping layer r2 was used to shape the 1×784 vector back to a 28×28-pixel image for the sake that we could use the raw handwritten digits from the MNIST database directly to calculate the loss function. Three 2-D
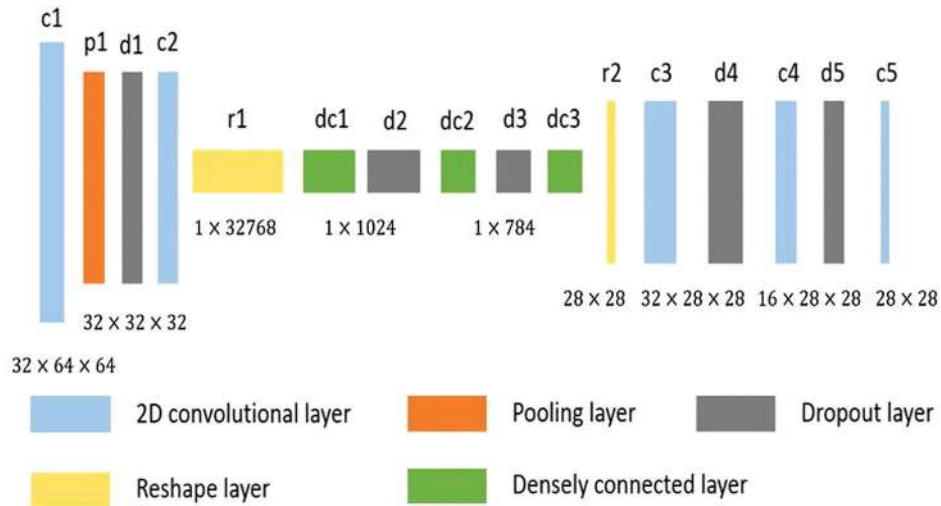
Fig. 2 The diagram of the proposed HNN model.

convolutional layers with the kernel of 3×3, 2×2, and 1×1 pixels, respectively, were used to calculate the feature maps in three different abstraction levels. The output of the network was a 28×28-pixel image. In the proposed HNN, the activation function was the rectified linear units,[39] which allowed fast and effective training of the network compared with the sigmoid.[38] We used the mean square error (MSE) as the loss function, and Adam[40] as the strategy to update the weights in the training process. To reduce overfitting, dropout layers[41] were used throughout the HNN architecture.

To train the HNN, we used images of 3990 handwritten digits downloaded from the MNIST handwritten digit database[34] as the training sets. In the experiment, we sequentially displayed them on the amplitude-only SLM shown in Fig. 1 and captured the corresponding speckle patterns. As aforementioned, we used only the central 512×512 pixels of the sCMOS camera for data acquisition. More specifically, the training set was created by pairing up the 3990 28×28-pixel handwritten digits with the corresponding 64×64 speckle patterns, which were randomly

taken out of the 512×512 speckle patterns acquired by the camera. The set of 3990 pairs' input–output data was fed into the HNN model to optimize the connection weights and biases of all the neurons. The program was implemented on the Keras framework with Python 3.5, and sped up by a GPU (NVIDIA Quadro P6000). The training was converged after only 1 epoch of about 194 s. The time it took to reconstruct an image from its speckle was 0.78 s in our experiment.

3 Results

To test the performance, we experimentally obtained 10 speckle patterns corresponding to 10 other digits from the MNIST database but not in the training set and sent them to the trained neural network. These 10 test speckle patterns are shown in Fig. 3(a). The images reconstructed by the proposed HNN are shown in Fig. 3(b). In comparison with the ground-truth images shown in Fig. 3(c), all the visible features, and particularly the edges of the targets, have been retrieved successfully. Unlike
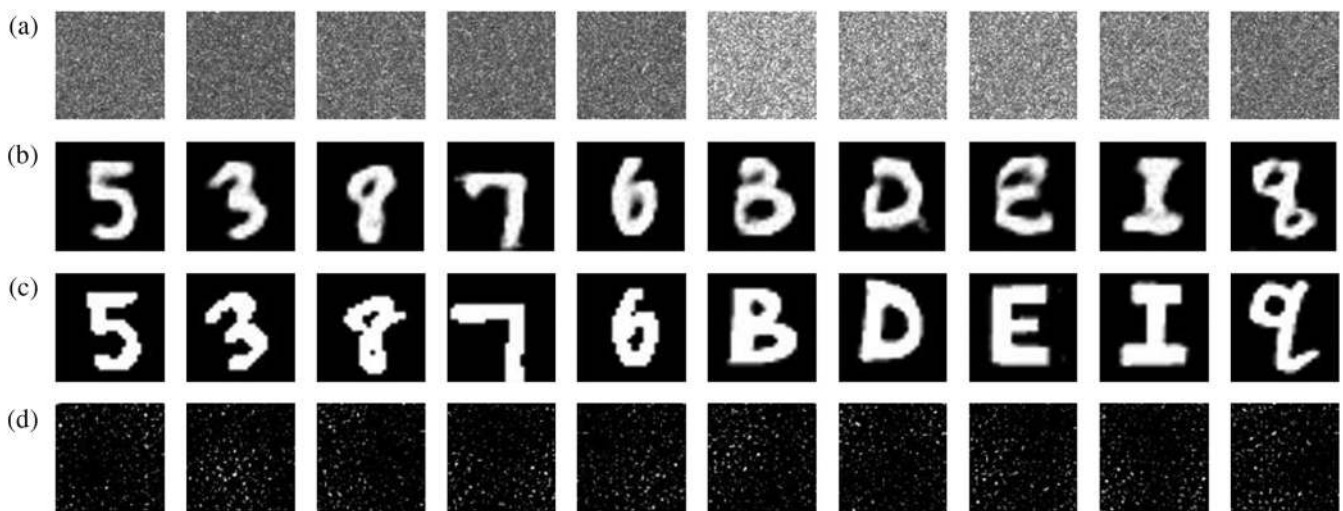

Fig. 3 The reconstructed results. (a) The speckle patterns (64×64 pixels) cropped from the raw acquired scattered pattern (512×512 pixels), (b) the reconstructed images by using the proposed HNN, (c) the ground-truth images, and (d) the reconstructed images by using memory effect.

the images reconstructed using the phase conjugation,[7] wavefront shaping,[10,11] or transmission matrix,[13] the background of the images retrieved using the HNN is very clean. We also plot the images reconstructed using the memory-effect-based method in Fig. 3(d) for comparison. One can see that nothing but noise is reconstructed in this case. This result is expected, and consistent with the theoretical prediction,[15] as the range of memory effect is about 0.01 deg measured by using the standard technique,[20,35] whereas the angular extent of the digits displayed on the SLM is about 0.4 deg.

Usually, a neural network produces a better result if the feature of a test image is closer to that of the training images.[27] Surprisingly, in the case of imaging through thick scattering media, we found that the HNN trained by the digital images alone can retrieve the images of handwritten English letters, which were manually written by one of the authors. The experimentally reconstructed results are shown as the five letter columns [right-hand side of Fig. 3(b)].

4 Analysis and Discussion

4.1 Comparative Analysis of the Performance

Now let us compare the proposed HNN with the CNN and DNN architectures in terms of imaging performance. To do that, we first need to make a CNN and a DNN. For the sake of fairness in comparison, all the networks should have the same number of layers, even though it is not possible to make them the same width. Thus, we changed the three densely connected layers in the proposed HNN to convolutional layers and made it a CNN and changed all the convolutional layers to the densely connected layers to establish a DNN. The 10 reconstructed test images are shown in Fig. 4. One can see from the results that the DNN can reconstruct the handwritten digits with an acceptable

quality (except the digit "5"), but it failed in the case of English letters. This means that the trained DNN has limited generalization. The CNN performed even worse, as only the digit "6" can be reconstructed.

4.2 Information Content of the Speckle

It is instructive to understand where the information of the target is stored in the speckle patterns and how well it can be compressed. To quantify our analysis, we first define the information content (IC) according to Shannon's information theory:[42]

$$IC = n_p \log_2(N_s), \tag{1}$$

where n_p is the total number of image elements (pixels), and N_s is the signal's distinguishable intensity level, which is represented by $g_{max} - g_{min}$, where g_{max} and g_{min} are the maximum gray level and minimum gray level of the captured images. For most of the speckle patterns captured by the camera $n_p = 512 \times 512$; $g_{max} = 2^{16} - 1 = 65{,}535$; and $g_{min} = 16{,}729$ instead of 0, because they were not regularized with respect to the background.

4.2.1 Image reconstruction with randomly selected parts of the speckle pattern

It is well known that holograms contain the whole wavefront of the objects.[1] Even a small fraction can reconstruct the object wavefront, although noise arises. In the case of coherent imaging through a thick scattering medium, the speckle acquired by the camera is formed by the interference of light taking many possible paths inside the medium. The motivation here is to study if the speckle formed by the diffusive light holds the same property. In fact, since we only use part of raw scattering images, it is suggested that our neural network can reconstruct the object from partial information. We arbitrarily select three
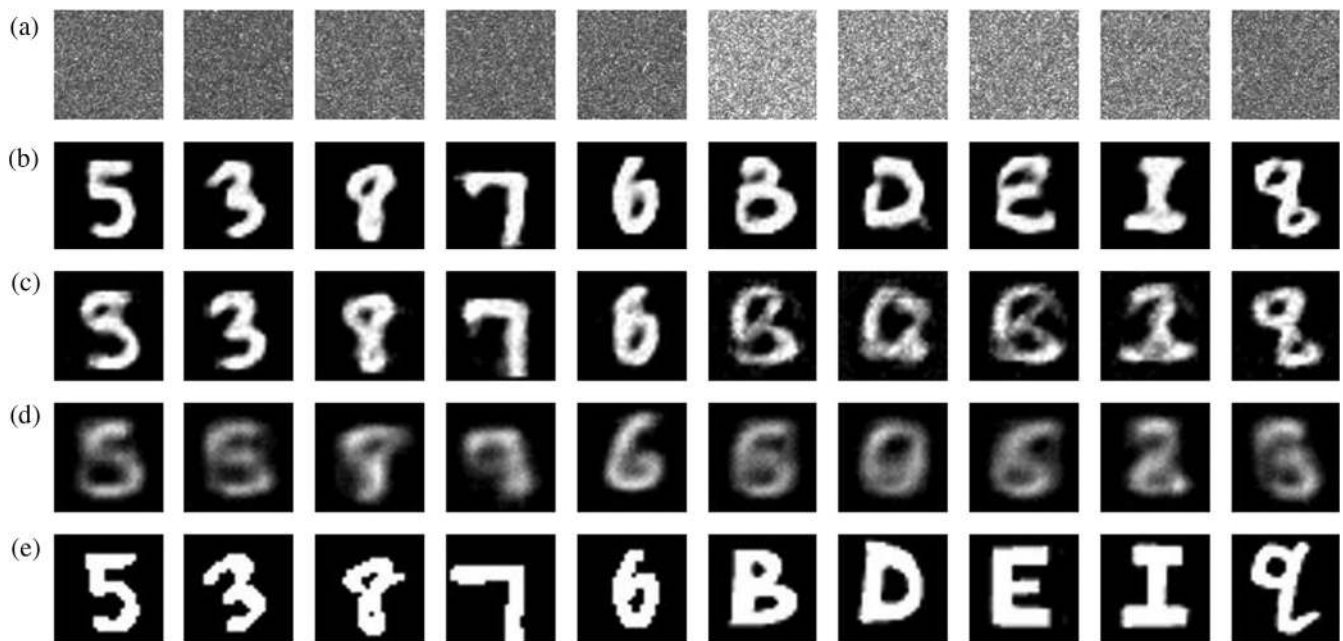

Fig. 4 Comparison of reconstruction performance. (a) The speckle patterns cropped from the raw scattered patterns. The images reconstructed by (b) the HNN, (c) a DNN, and (d) a CNN. (e) The ground-truth images.
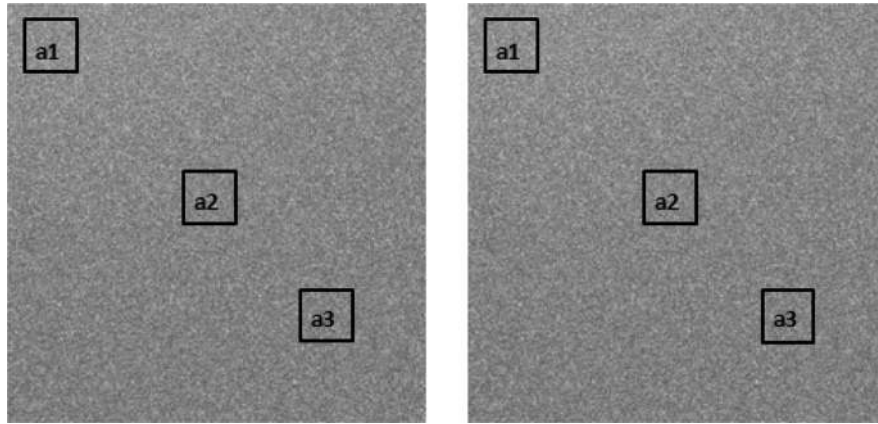
Fig. 5 The positions of the three 64×64-pixel blocks randomly selected from the acquired scattered patterns of 512×512 pixels.

small blocks (marked by a1, a2, and a3) of 64×64 pixels out of each captured raw speckle patterns of 512×512 to train the HNN network, as shown in Fig. 5. The results are shown in Fig. 6. It is clearly seen that both the digits and the English letters are retrieved successfully, although there are some distortions in the results of English letters.

Surprisingly, we could even randomly pick 4096 pixels, not necessarily in a continuous block, out of each raw speckle pattern, and arbitrarily arrange them to form a new 64×64 speckle-like pattern. Then, we paired it up with the associated handwritten digit, although they are not directly physically related to each other. Following the same process (picking the pixels at the same positions and rearranging them in the same way), we obtained 3990 pair data of this type and used them to train the HNN. Then, we generated some test speckle patterns in the same way and used them to test the HNN. The results, as shown in Fig. 7, suggest that the information about the target does spread across the whole speckle pattern, similarly to the hologram. And it does not matter which part of the recorded speckle pattern is used.

A significant concern is to determine how small the fraction of the speckle can be to faithfully reconstruct the target image. To explore this possibility, we took out blocks of six different sizes (8×8, 16×16, 32×32, 64×64, 128×128, and 256×256) from each speckle pattern and paired them up with the corresponding targets to train the network. Of course, the network width should be changed accordingly, since the size of the input speckle that the network will process is different. When the networks were well trained, we used them to process the test speckle patterns of the corresponding size; the reconstructed images are shown in Fig. 8. The results suggest that as the size of speckle reduces, the quality of the reconstructed image degrades. Specifically, the reconstructed image is still recognizable if the input speckle is 32×32 pixels. But the reconstructed image is severely distorted when we use only 256 or fewer pixels of the recorded speckle.

The results suggest that the speckle pattern produced in this way is indeed highly redundant. The information of the target is actually encoded everywhere across the speckle pattern.
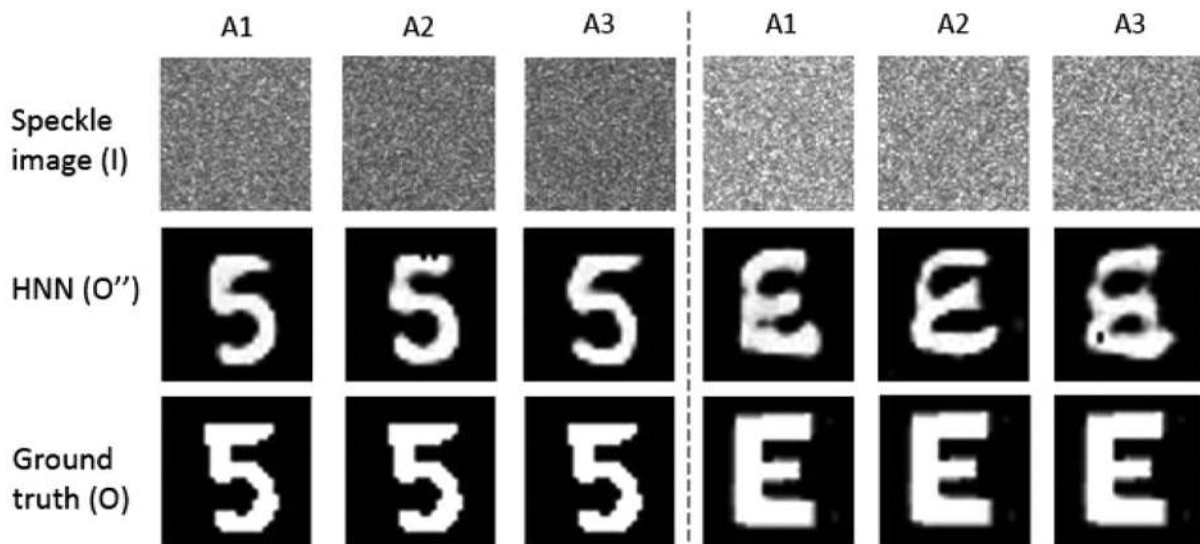

Fig. 6 The images reconstructed by using the subspeckle pattern located at A1, A2, and A3.
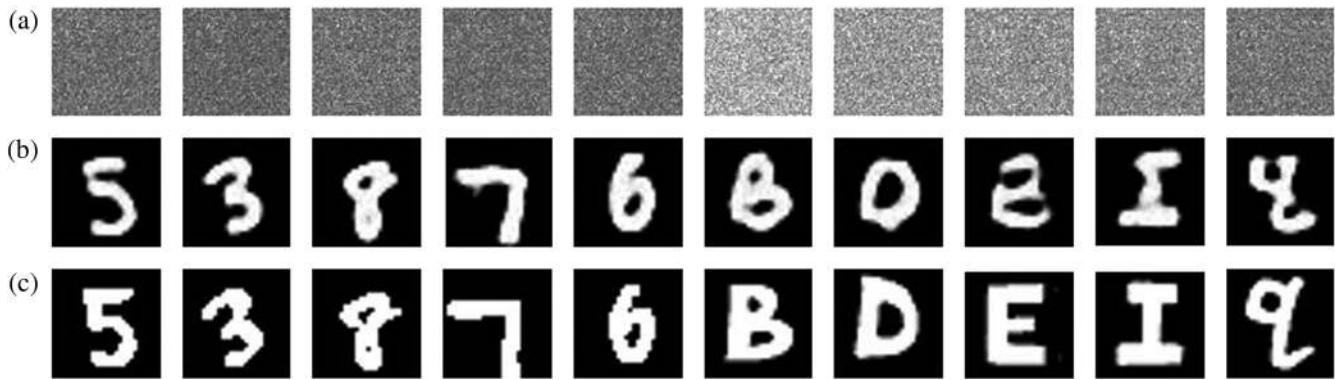
Fig. 7 The images of handwritten digits and English letters reconstructed from the randomly selected pixels. (a) The subspeckle patterns formed by the randomly selected 3096 pixels from the 512 × 512 raw scattered patterns, (b) the reconstructed images, and (c) the ground-truth images.
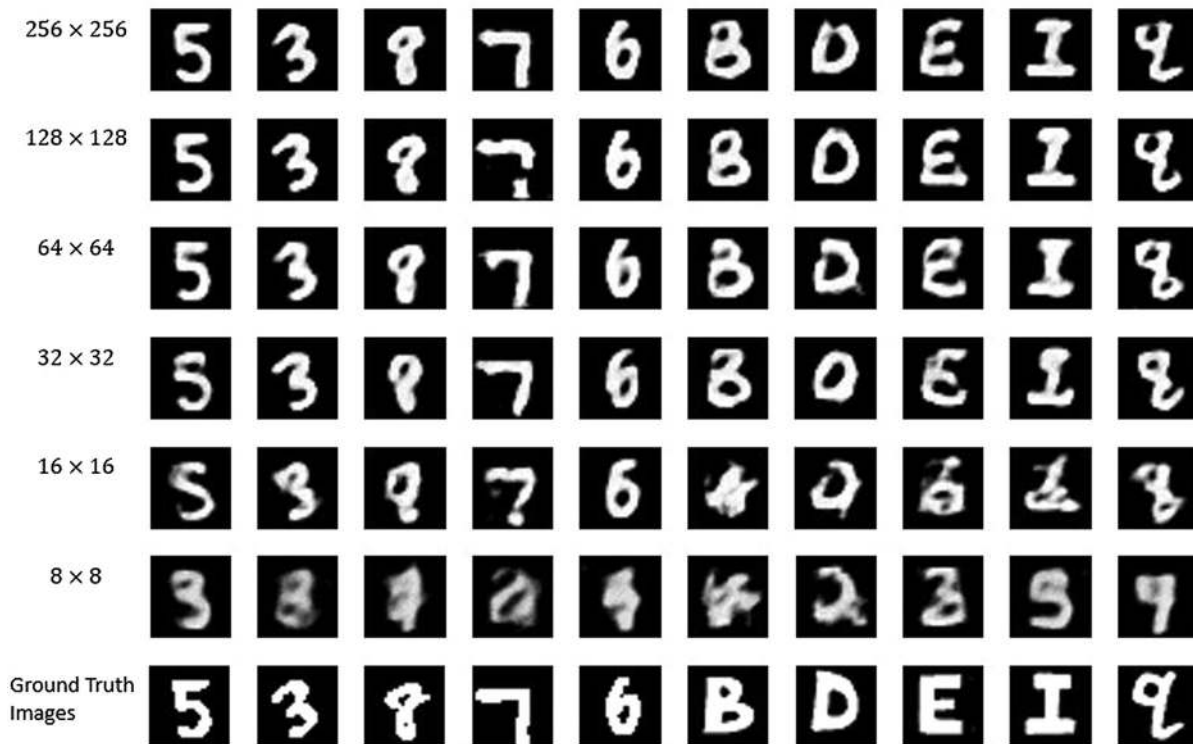

Fig. 8 The images of handwritten digits and English letters reconstructed from the subspeckle patterns of six different sizes.

4.2.2 Image reconstruction with less bit depth

Now we go further to see if it is possible to reconstruct the image by using less bit depth from the 64 × 64-pixel subspeckle pattern. The sCMOS camera we used has a bit depth of 16 bits, quantifying the pixel value of the speckle pattern to the range of 0 to 65,535. We examined the effect in two ways. First, we only used a subrange of the pixel value of the test speckle patterns to reconstruct the target images. For example, we used the pixels with the values in the range of 25,000 to 30,000, setting the value of the pixel to 0 if it is <25,000 and to 65,535 if it is >30,000, and as-is if it is in the range of 25,000 to 30,000. The reconstructed results are shown in Fig. 9(G1). One can see that it is distorted in both the cases of handwritten digits

and English letters. Similarly, when we used the pixels in the range of 35,000 to 40,000 and 40,000 to 45,000 as the input, we have the results that are shown in Figs. 9(G2) and 9(G3), respectively. One can see that the reconstructed digits are recognizable, whereas the English letters are distorted, suggesting the limited ability of generalization in this case. The difference in quality between the images reconstructed using the pixels in different ranges is mainly due to the probability density function of the speckle intensity.[43] Since the pixel values are mainly in the range of 25,000 to 45,000, we are unable to reconstruct the target images by using the pixels whose values are below 25,000 or above 45,000 alone. Second, we examined an extreme case of binarizing the test speckle patterns by setting all the pixel values
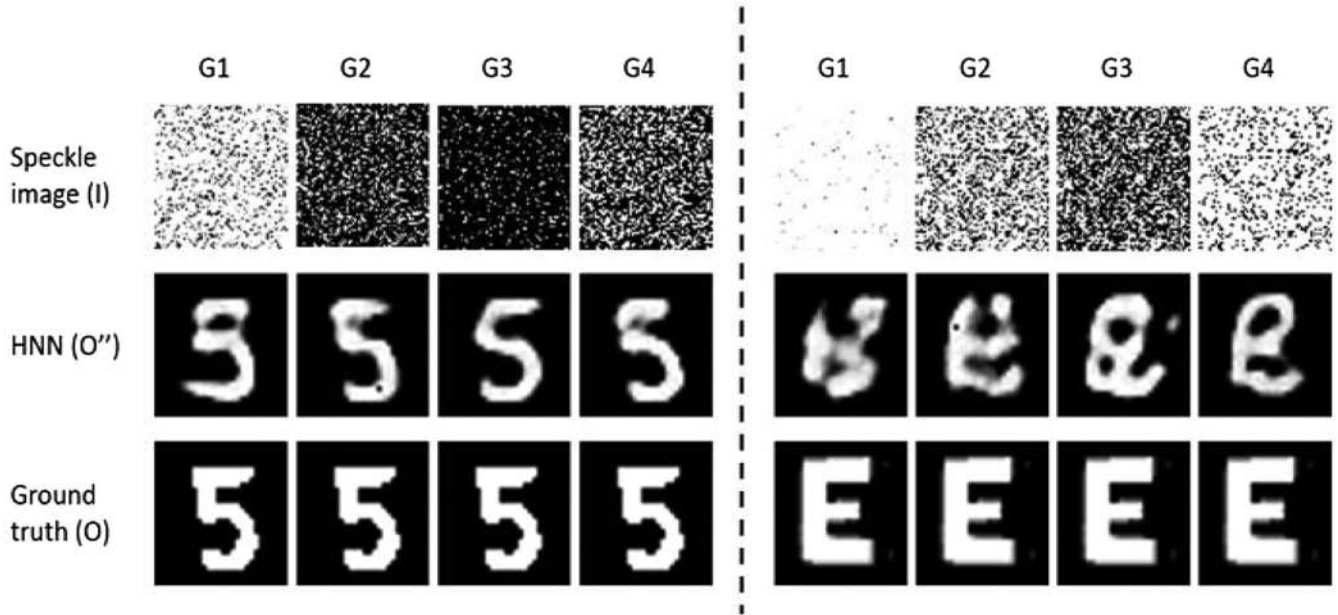
Fig. 9 The result of the digits and English letters with different gray intervals. The images in the first row are the speckle images with different gray-level intervals, the second row shows the predicted objects by the HNN model, images in the third row are the ground-truth images. (G1) Images with gray value interval (25,000 to 30,000), (G2) images with gray value interval (35,000 to 40,000), (G3) images with gray value interval (40,000 to 45,000), and (G4) images with gray value threshold of 35,000.

<35,000 to 0, and otherwise to 65,535. Experimental results show that the target images can be reconstructed even in this case, as shown in the Fig. 9(G4).

4.2.3 Compression ratio of the speckle

The results presented in Secs. 4.2.1 and 4.2.2 suggest that the speckle patterns generated in this way are highly redundant and therefore compressible in both the spatial and the bit depth domains. To measure the redundancy of the speckle pattern, we can calculate the information content according to Eq. (1). For the raw speckle pattern directly captured by the camera, the information capacity is $IC = (512 \times 512)\log_2(65,535 - 16,729) \approx 4.1 \times 10^6$ bits. And for the speckle patterns of minimal size and bit depth that contain sufficient information to reconstruct the target image with acceptable quality, the information content is reduced to $IC = (64 \times 64)\log_2(2) = 4096$ bits. Thus the compression ratio is given as

$$C = \frac{4096}{4.1 \times 10^6} \approx 0.1\%. \quad (2)$$

This means that only 0.1% of the information contained by the captured speckle pattern is sufficient to reconstruct the target image with the proposed HNN.

From the information theory point of view, this result is reasonable because the information content of the handwritten digits is <4096. But no other method except properly designed neural networks can be used to reconstruct the target image from such a small fraction of the recorded speckle, to the best of our knowledge.

5 Conclusions

We have proposed a deep learning method for imaging through thick scattering media. Specifically, we experimentally demonstrated the faithful reconstruction of a image of the target hidden behind a 3-mm-thick polystyrene slab with an optical depth of 13.4. The proposed HNN architecture is composed of fully connected layers and convolutional layers so that it has the advantages of both the DNN and the CNN, and thus gives better performance. As a learning-based method, it requires a large set of labeled data for training. But it just needs to be trained once. The collection of the training data set is also required for the techniques of wavefront shaping and TM measurement. But the requirement of the training data in the proposed method is not as strict as in the other two, which should be a set of spatial modes[15] in wavefront shaping, and a set of plane waves with different spatial frequencies[36] in the TM method. In addition, it does not need to measure the complex amplitude but only the input–output intensities so that the experimental system can be more simple and concise. Furthermore, it does not need to use any microscopic objective to demagnify the images displayed on the SLM to fit the size of the eigenchannel.[11,15] This means that the trained HNN may not reveal any physical details about the transmission matrix of the scattering slab. Nevertheless, it opens up a new possibility for exploring optical imaging through more general complex systems. Finally, we have also demonstrated that the speckle patterns acquired by the camera are highly redundant. Experimentally, we demonstrated that the use of only 0.1% of the information content of the captured speckle pattern is sufficient to reconstruct the target image, meaning that the information-carrying speckle is highly compressible.
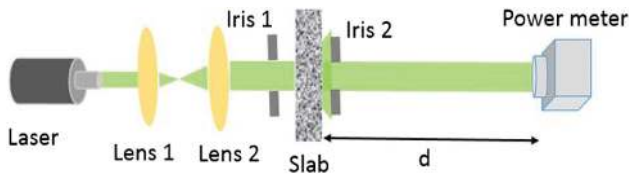
Fig. 10 The setup of the system to measure the optical depth.

6 Appendix A: Measuring Optical Depth

The optical depth d_s of scattering sample is defined as[44]

$$d_s = -\ln\left(\frac{I_b}{I_i}\right), \tag{3}$$

where I_b is the intensity of the ballistic light and I_i is the intensity of the incident light. We used the experimental system shown in Fig. 10 to measure the optical depth. Two variable irises with the diameter $D = 8$ mm were used to adjust the size of the beam. The distance d between the slab and the power meter was 600 mm. First, we measured the incident power I_i to be 8.599 mW. This was done without the slab. Then, we measured the scattered light power $I_s = 17.32$ nW by placing the slab in the light path. Finally, we placed a metal plate behind Iris 2, in order to block the ballistic light, and measured the background noise $I_n = 4.05$ nW. Then, the ballistic light $I_b = I_s - I_n = 13.27$ nW. In our experiments, the optical depth of the white polystyrene slab is 13.40, according to Eq. (3).

Acknowledgments

References

1. J. W. Goodman, *Introduction to Fourier Optics*, Roberts & Co., Greenwood Village, Colorado (2005).
2. V. Ntziachristos, "Going deeper than microscopy: the optical imaging frontier in biology," *Nat. Methods* **7**, 603–614 (2010).
3. A. Ishimaru, *Wave Propagation and Scattering in Random Media*, Academic Press, New York (1978).
4. J. Beckers, "Adaptive optics for astronomy: principles, performance, and applications," *Annu. Rev. Astron. Astrophys.* **31**, 13–62 (1993).
5. S. Rotter and S. Gigan, "Light fields in complex media: mesoscopic scattering meets wave control," *Rev. Mod. Phys.* **89**, 015005 (2017).
6. J. W. Goodman et al., "Wavefront reconstruction imaging through random media," *Appl. Phys. Lett.* **8**, 311–313 (1966).
7. Z. Yaqoob et al., "Optical phase conjugation for turbidity suppression in biological samples," *Nat. Photonics* **2**, 110–115 (2008).
8. Y. Zhang et al., "Application of short-coherence lensless Fourier-transform digital holography in imaging through diffusive medium," *Opt. Commun.* **286**, 56–59 (2013).
9. A. K. Singh et al., "Exploiting scattering media for exploring 3D objects," *Light Sci. Appl.* **6**, e16219 (2017).
10. I. M. Vellekoop and A. Mosk, "Focusing coherent light through opaque strongly scattering media," *Opt. Lett.* **32**, 2309–2311 (2007).
11. A. P. Mosk et al., "Controlling waves in space and time for imaging and focusing in complex media," *Nat. Photonics* **6**, 283–292 (2012).
12. O. Katz, E. Small, and Y. Silberberg, "Looking around corners and through thin turbid layers in real time with scattered incoherent light," *Nat. Photonics* **6**, 549–553 (2012).
13. S. M. Popoff et al., "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).
14. S. Popoff et al., "Image transmission through an opaque material," *Nat. Commun.* **1**, 81 (2010).
15. M. R. I. Freund and S. Feng, "Memory effects in propagation of optical waves through disordered media," *Phys. Rev. Lett.* **61**, 2328–2331 (1998).
16. J. Bertolotti et al., "Non-invasive imaging through opaque scattering layers," *Nature* **491**, 232–234 (2012).
17. O. Katz et al., "Non-invasive single-shot imaging through scattering layers and around corners via speckle correlations," *Nat. Photonics* **8**, 784–790 (2014).
18. G. Li et al., "Cyphertext-only attack on the double random-phase encryption: experimental demonstration," *Opt. Express* **25**, 8690–8697 (2017).
19. W. Yang, G. Li, and G. Situ, "Imaging through scattering media with the auxiliary of a known reference object," *Sci. Rep.* **8**, 9614 (2018).
20. S. Schott et al., "Characterization of the angular memory effect of scattered light in biological tissues," *Opt. Express* **23**, 13505–13516 (2015).
21. G. Li et al., "Image transmission through scattering media using ptychographic iterative engine," *Appl. Sci.* **9**, 849 (2019).
22. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, Cambridge (2016).
23. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature* **521**, 436–444 (2015).
24. R. Horisaki, R. Takagi, and J. Tanida, "Learning-based imaging through scattering media," *Opt. Express* **24**, 13738–13743 (2016).
25. M. Lyu et al., "Exploit imaging through opaque wall via deep learning," arXiv:1708.07881 (2017).
26. Y. Rivenson et al., "Phase recovery and holographic image reconstruction using deep learning in neural networks," *Light Sci. Appl.* **7**, 17141 (2018).
27. A. Sinha et al., "Lensless computational imaging through deep learning," *Optica* **4**, 1117–1125 (2017).
28. K. H. Jin and M. T. McCann, "Deep convolutional neural network for inverse problems in imaging," *IEEE Trans. Image Process.* **26**, 4509–4522 (2017).
29. P. Caramazza et al., "Neural network identification of people hidden from view with a single-pixel, single-photon detector," *Sci. Rep.* **8**, 11945 (2018).
30. S. Li et al., "Imaging through glass diffusers using densely connected convolutional networks," *Optica* **5**, 803–813 (2018).
31. 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).
32. G. Barbastathis, A. Ozcan, and G. Situ, "On the use of deep learning for computational imaging," *Optica* accepted (2019).
33. M. Lyu et al., "Deep-learning-based ghost imaging," *Sci. Rep.* **7**, 17865 (2017).
34. L. Deng, "The MNIST database of handwritten digit images for machine learning research [best of the web]," *IEEE Signal Process. Mag.* **29**, 141–142 (2012).
35. I. Freund, "Looking through walls and around corners," *Phys. A* **168**, 49–65 (1990).
36. I. Freund, "Image reconstruction through multiple scattering media," *Opt. Commun.* **86**, 216–227 (1991).

37. K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Netw.* **2**, 359–366 (1989).

38. G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Math. Control Signals Syst.* **2**, 303–314 (1989).

39. V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. of the 27th Int. Conf. on Machine Learning*, pp. 807–814 (2010).

40. K. Diederik and J. Ba, "Adam: a method for stochastic optimization," arXiv:1412:6980 (2014).

41. N. Srivastava et al., "Dropout: a simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.* **15**, 1929–1958 (2014).

42. C. M. Michail et al., "Figure of image quality and information capacity in digital mammography," *BioMed Res. Int.* **2014**, 297–308 (2015).

43. J. W. Goodman, *Statistical Optics*, 2nd Ed., Wiley and Sons, Hoboken, New Jersey (2015).

44. R. Michels, F. Foschum, and A. Kienle, "Optical properties of fat emulsions," *Opt. Express* **16**, 5907–5925 (2008).

Meng Lyu received his BS degree at Sichuan University and PhD degree in optical engineering at Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences. His research interests include computational imaging, digital holography, and deep learning.

Hao Wang received his BS degree at University of Science and Technology of China. He is a third-year master's student at Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences. His research interests include computational imaging, digital holography, and deep learning.

Guowei Li received his BS degree at Sichuan University and he is a fifth-year PhD student in optical engineering at Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences. His research interests include computational imaging and scattering imaging.

Shanshan Zheng received her BS degree at Hunan University and she is a third-year PhD student in optical engineering at Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences. Her research interests include computational imaging, light fields, and scattering imaging.

Guohai Situ is a professor at the Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences. His research interests span a wide field of computational optical imaging, ranging from developing novel techniques, including deep learning, to wavefront engineering for optical imaging and image processing. He serves on the editorial board of *Applied Optics*, *Applied Physics B*, *Chinese Journal of Lasers*, and *Scientific Reports*.