An end-to-end network for multiple scattering media imaging

Haoxing Yang[a], Ziyang Yuan[b], Hongxia Wang[a*], Lizhi Cheng[a]
[a] Department of Mathematics, National University of Defense Technology, Changsha 410073, Hunan, China; [b] Academy of Military Sciences, Beijing, China

ABSTRACT

In this paper, we propose an end-to-end neural network abbreviated as TCNN to solve the blind phase retrieval problem in multiple scattering imaging. TCNN is a kind of auto-encoder with a transform layer, which acts as a bridge between transforming domains. Compared to double phase retrieval method, TCNN can directly estimate the image from those phaseless measurements through the nonlinear network structure. During training, the parameters of TCNN are updated by the adaptive moment estimation algorithm Adam. Numerical experiments show that TCNN can recover images with comparable quality to that of state-of-the-art methods. Moreover, TCNN hugely reduces the time cost for recovering images once the training procedure is completed.

Keywords: Neural network, phase retrieval, multiple scattering media, end-to-end

1. INTRODUCTION

The light incident on the multiple scattering media suffers from multiple reflections. When observing the object through the multiple scattering media with coherent light in the focal plane, the speckle pattern displayed on the far side of the scatter bears no resemblance to the image of the object. This phenomenon is attributed to the wavefront interfering with itself destructively when passing through the multiple scattering media. It is sophisticated to recover the image from this speckle pattern. First, a large amount of beneficial structural information for reconstructing the image gets lost. For instance, detectors such as CMOS or CCD can only record the intensity of speckle pattern. It is well-known that recovering an object only from its intensity measurements called phase retrieval is an ill-posed inverse problem. Second, to estimate the transmission matrix of this imaging system is not easy. Suffering from various kinds of interferences, the transmission function of light cannot be modeled accurately. But this problem plays a major role in diversities of applications such as astronomy, crystallography, multiple scattering imaging, etc. With intensive requirements on the high resolution of recovered images, series of techniques are constantly sprung up to enrich this filed, such as TOF (Time of Flight) method[1], multi-slice light-propagation method[2], strong memory effect method[3], holographic interferometry method[4], temporally modulated phase method[5] and double phase retrieval method[6].

Compared to other methods mentioned above, the double phase retrieval can relieve perturbations caused by the depth and complexities of the scatterer. On the other hand, once the transmission matrix has been estimated, this method is able to recover the image by means of capturing only one single speckle pattern. Mathematically, double phase retrieval for 1-D signals can be described as

$$\begin{aligned} &\text{Find} \quad && \mathbf{A}, \mathbf{x} \\ &\text{s.t.} \quad && |\mathbf{A}^{*}\mathbf{x}|^{2} = \mathbf{b}, \end{aligned} \tag{1}$$

where $\mathbf{A} \in \mathbb{C}^{n \times m}$ is the transmission matrix, $\mathbf{x} \in \mathbb{C}^{n}$ is the signal to recovered, **b** is the amplitude-only measurement, * denotes the conjugate and transpose, and $|\cdot|$ denotes the element-wise absolute. The double phase retrieval method contains two main steps: First, estimating **A** from a series of measurements, then recovering x based on **A**. As its name suggested, the core of double phase retrieval is to solve several phase retrieval problems.

Note that the double phase retrieval method may consume a large proportion of computational resources to solve a series of phase retrieval problems for evaluating the transmission matrix **A** from plenty of measurements **b** and signal of

[*] wanghongxia@nudt.edu.cn

Third International Conference on Computer Science and Communication Technology (ICCSCT 2022)
edited by Yingfa Lu, Changbo Cheng, Proc. of SPIE Vol. 12506, 125063V
© 2022 SPIE · 0277-786X · doi: 10.1117/12.2662207

interest **x**. Furthermore, it also seems a little redundant to segment the problem into two steps. As a result, time saved, explicit and efficient methods need to be designed.

Deep learning has received widespread attention in recent years. Diversities of layer structures and optimization methods were come up to accelerate the development of deep neural network. By mining the information from data through networks, inverse problems such as MRI[7], holography[8], super-resolution[9] and phase retrieval[10] can be resolved easier. Specifically, for the phase retrieval problem, a neural network was designed to diminish the effect of twin images[11], and a neural work was proposed to increase the resolution of an image in lensless coherent diffraction images[10]. The proximal gradient descent method was implemented to deal with phase retrieval problem[12], where the RED priority[13] is incorporated in the loss function. A neural network is used as the denoiser in RED priority. Compared to these works mentioned above, the property of transmission media in multiple scattering media imaging is worse, at the same time these speckle patterns bear no resemblance to the images.

In this paper, we proposed a novel neural network method, which is called Transformation Convolution Neural Network (TCNN), to tackle the multiple scattering imaging problem. A special structure called the transforming layer is built, which can help TCNN bridge the relationship between the transforming and object domain. TCNN combines two steps in double phase retrieval together, and learns the relationship between those speckle patterns and images of objects directly. Universality theory is also established to demonstrate the relationship can be approximated by a network with high accuracy. Nonlinear loss function with penalty is built and ADAM[14] is applied to update the parameters of TCNN by training data measured. When the training procedure is done, the image of an object can be recovered by making the speckle pattern as the input of TCNN. Tests on the multiple scattering media imaging data[15] fully show that TCNN has a competitive performance in comparison with double phase retrieval methods. Moreover, the time cost to recover the image by TCNN is much less. Furthermore, if new training data is available, TCNN can be constantly refined based on these datasets. But double phase retrieval method has to evaluate the transmission matrix again from the whole data.

The reminders of this paper are organized as below. Section 2 provides the details of the experiment setup and TCNN. Section 3 presents the numerical results of TCNN are given. Section 4 is the conclusion.

2. NEURAL NETWORK METHOD TO SOLVE DOUBLE PHASE RETRIEVAL

2.1 The feasibility of implementing neural network

As is introduced in Section 1, the computational costs of double phase retrieval method is quite expensive for real-time imaging. We build a neural network with nonlinear structures to approximate operator $g:\mathbf{b}\rightarrow\mathbf{x}$ so that x can be estimated directly from b.

In the following we discuss the existence of the operator g. Based on the inverse function theorem, the operator $f:\mathbf{x}\rightarrow\mathbf{b}$ must be injective or g cannot be existed. But in phase retrieval problem, f is not a single valued operator usually. When $\mathbf{x}\in\mathbb{R}^n$, $f(\mathbf{x})=f(-\mathbf{x})$, while $\mathbf{x}\in\mathbb{C}^n$, $f(\mathbf{x})=f(c\mathbf{x})$, $|c|=1$. As a result, x is defined up to a global phase factor. In this paper, we concern with the map $f:\mathbb{R}^n/\{\pm1\}\rightarrow\mathbb{R}^m$ when x is real, and $f:\mathbb{C}^n/\mathbb{T}\rightarrow\mathbb{R}^m$ when x is complex (where $\mathbb{T}$ is the complex unit circle). When training the network, it is necessary to sift the whole dataset so that every $\mathbf{b}^{(i)}, i=1,\ldots,m$ is not equal. Then Theorem 2.2 and Theorem 3.3[16] guarantee the injective of f.

In the multiple-scattering media imaging, media which may be frost glass or painted wall. Under this condition, rows of transmission matrix A in traditional analysis are often assumed to be the Gaussian or sub-Gaussian which generic vectors. Consequently, if *m* is large enough, f is injective, the existence of g can also be guaranteed. Furthermore, g can also be approximated by neural network g_Θ :

Theorem 2.1. If the inverse operator g of $f:\mathbb{R}^n/\{\pm1\}\rightarrow\mathbb{R}^m$ (or $f:\mathbb{C}^n/\mathbb{T}\rightarrow\mathbb{R}^m$) in phase retrieval problem exists, then g can be approximated by neural network g_Θ (where Θ is the parameters of the neural network) with any desired degree of accuracy.

2.2 Neural network training

When the training set is in $\mathbb{R}^n \,/\, \{\pm 1\}$ or $\mathbb{C}^n \,/\, \mathbb{T}$, we can train a neural network to approximate this inverse operator g. In this paper, a neural network called TCNN is built to train g_Θ as an approximation of g. The mathematical model for neural network training is:

$$\min_{\Theta} L(\Theta) = \sum_{i}^{k} \left\| g_\Theta(\mathbf{b}^{(i)}) - \mathbf{x}^{(i)} \right\|^2 \tag{2}$$

where g_Θ is the neural network with hierarchical structures so that nonlinear information can be mined from the dataset. The encoding and decoding diagram of TCNN is shown in Figure 1, and all the blocks are shown in Figure 2.

·Encoding part

In the encoding part, the multiple flow structure is utilized in g_Θ which can learn the features of input in different scales. The input of g_Θ is decimated by the downsample Block which is constructed by one convolution layer with batch normalization, Relu activation function, and one maxpooling layer, which can be modeled as below:

$$\text{Maxpooling}\ \ \text{Relu}(\text{BN}(\mathbf{b}^{(i)} * \boldsymbol{\omega}^{l}))\ ,i = 1,\ldots,k$$

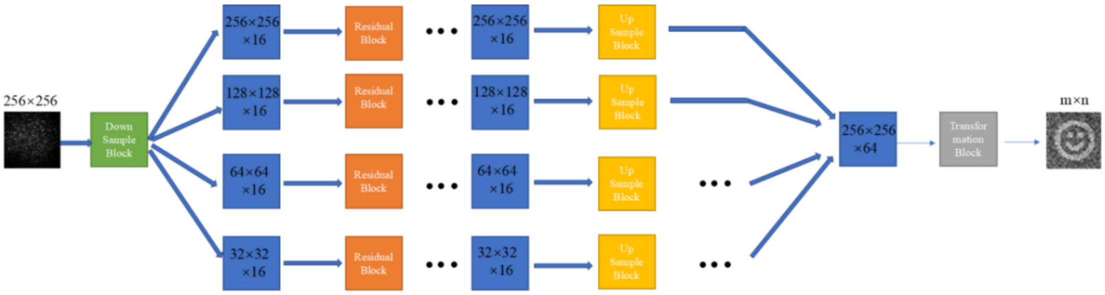


Figure 1. Encoding and decoding part.

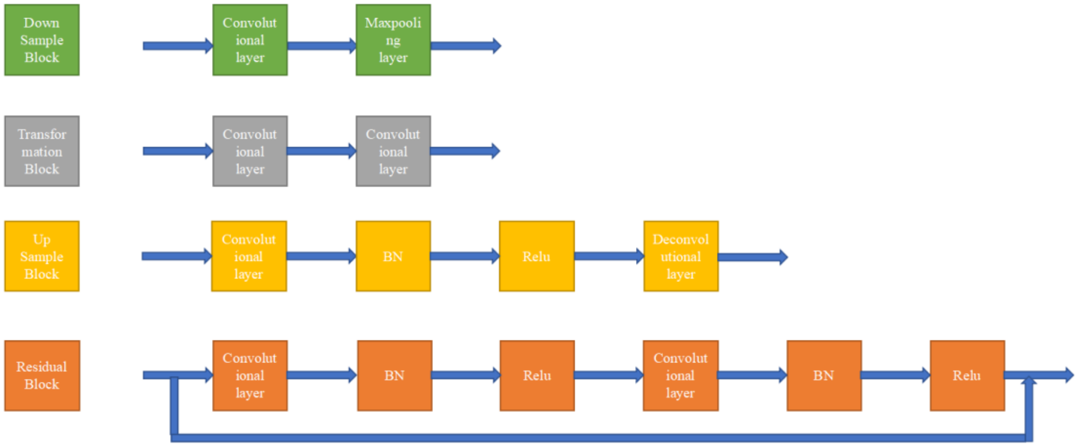


Figure 2. The details of blocks in TCNN.

where $*$ is the convolution, $\boldsymbol{\omega}^l$ is the convolutional kernel, *l* indicates the kernel is in the *l*-th path. The size of those kernels is 1×1×16. The maxpooling layer here decreases the dimension of the input. Specifically, $\mathbf{b}^{(i)}, i = 1,\ldots,k$ are downsampled by ×2, ×4, ×8 respectively. Then, the four different tensors are successively passed through five residue blocks, each block contains two convolution layers with batch normalization and a shortcut from the input to the output. The shortcut can accelerate the convergence of TCNN. After processed by the five residue blocks, the high-order features in different scales can be learned.

·Decoding part

In the decoding part, those features created above will generate four different tensors keeping the same width and length with the $\mathbf{b}^{(i)}, i = 1,\ldots,k$. So there will be 3, 2, 1 upsampling blocks for each tensor respectively. In each upsample block, there will be one convolutional layer with batch normalization and one deconvolutional layer. The deconvolution in this paper adopts the way of upsampling[11] for super-resolution. This method can alleviate the effect of zero padding by traditional deconvolution besides fully utilizing the information in the network.

·Transforming block

After being handled by those upsample blocks, tensors will be fused and passed through the transformation block instituted by one convolutional layer and transformation layer. The transformation layer is actually a full connection layer which acts as a linear transformation between the transform domain and object domain. In the test, to alleviate the influence of over-fit, the units in the transformation layer are randomly neglected.

Because of large quantities of the dataset, mini-batch gradient descent method is implemented to update Θ in (2), which is actually a non-convex optimization, it is hard to guarantee the algorithm converges to the global optimum. But in practice, some efficient gradient based methods outperform when training neural network. In this paper, we utilized ADAM method[14], which can accelerate the decreasing of the loss function besides it can also adaptively calibrate the first moment, second moment and the learning rate. The skeleton of parameter updating is shown in Algorithm 1.

During the training procedure, the learning rate μ decays with a geometrical rate so that the evaluation cannot vibrate violently in the final stage of iteration. In the next section, the results of numerical tests will be given to demonstrate the efficiency of the TCNN.

Algorithm 1: Parameters updating by ADAM.

```
Input:
$\mathbf{x}^{(i)}, i = 1,\ldots,k$ : known images of object;
$\mathbf{b}^{(i)}, i = 1,\ldots,k$ : phaseless measurements of $\mathbf{x}^{(i)}$;
$T$ : the maximum iteration steps;
$\mu$ : the learning rate; $\alpha_1, \alpha_2, \varepsilon$ : parameters.
Output: $\hat{\Theta}$ : the parameters of neural network.
Initialization:
$\Theta_0$ : random initialization of the parameters in the network;
$\mathbf{u}_t$ : initialization of the first order moment;
$\mathbf{v}_t$ : initialization of the second order moment.

For $t = 1 : T$ do
    $\mathbf{g}_t = \nabla_\Theta L(\Theta_{t-1})$
    $\mathbf{u}_t = \alpha_1 \mathbf{u}_{t-1} + (1-\alpha_1)\mathbf{g}_t$
    $\mathbf{v}_t = \alpha_2 \mathbf{v}_{t-1} + (1-\alpha_1)\mathbf{g}_t$
    $\hat{\mathbf{u}}_t = \mathbf{u}_t / (1-\alpha_1^t)$
    $\hat{\mathbf{v}}_t = \mathbf{v}_t / (1-\alpha_2^t)$
    $\Theta_t = \Theta_{t-1} - \mu \hat{\mathbf{u}}_t / (\sqrt{\mathbf{v}_t} + \varepsilon)$
End
```

3. TESTS BY THE EMPIRICAL DATASET

3.1 Datasets of the numerical tests

The experiment setup is made by Rice University[15]. Specifically, a laser beam that is spatially filtered and collimated (λ = 632.8nm) illuminates a spatial light modulator (SLM) from Holoeye. This reflective type display (LC2012), which is equipped with 1024×768 resolution and 36μm square pixels, modulates the phase of the laser beam. Then the lens L (f = 150mm) focus the laser beam on the scattering medium. The experiment uses a microscope objective (Newport, X10, NA: 0.25) to image the SLM calibration pattern to the sensor (Point Grey Grasshopper 2 with pixel size 6.45μm). As the phase only SLM is 8 bit, it can modulate the wavefront by an element of $\{e^0, e^{2\pi j\frac{1}{256}}, \ldots, e^{2\pi j\frac{255}{256}}\}$. In the test, the phase modulation is restricted to $\{0, \pi\}$, which means the value of the x is $\{-1, 1\}$. For the amplitude-only SLM, the source pixel is set to be either completely 0 or completely 1. Then the value of x is $\{0, 1\}$. By constantly modulate the SLM, values of SLM and their speckle patterns can be recorded to estimate transmission matrix.

The dataset can be downloaded from https://rice.app.box.com/v/TransmissionMatrices. Two types of data are utilized in the experiment. For amplitude-only SLM, the size of the image is 16×16. For the phase-only SLM, the size of the image is 40×40. Experiments are implemented by a desktop with GPU NVIDIA 1080 and CUDA 9.0.

3.2 Experimental results

Some hyper parameters of in this test can be found in Table 1. TCNN is built on the framework of tensorflow. The learning rate μ is 10^{-3} at initial and decays after each epoch with the factor 0.85. The total number of the epoch in the training procedure is 100. To preprocess the data in this experiment, we firstly sift the data so that every image and corresponding speckle pattern is unique. Then we train TCNN by two different sizes of data.

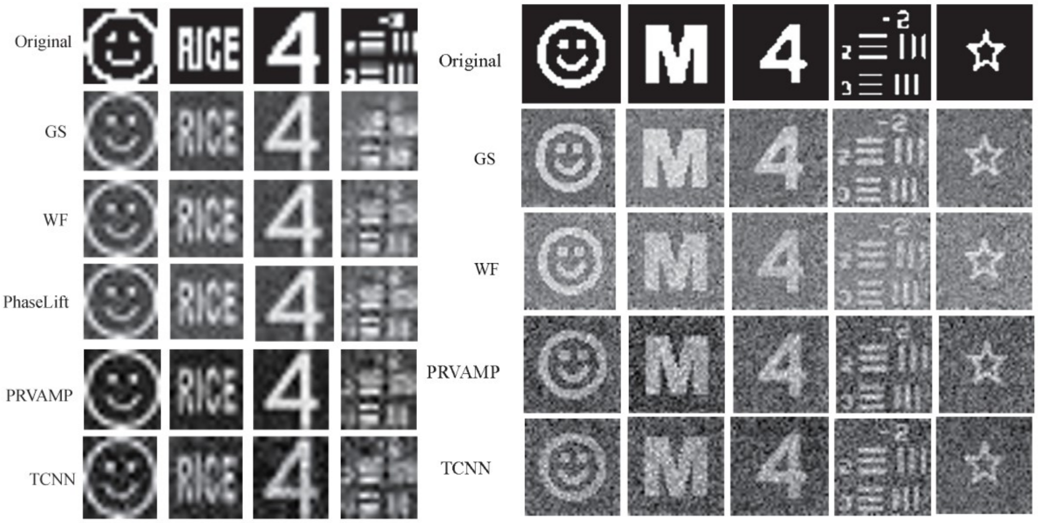
We compare TCNN with GS[17], WF[18], PhaseLift[19], Prvamp[15]. TCNN can directly recover images from those speckle patterns without estimating the transmission matrix. For other methods, we utilize the transmission matrix A estimated by Prvamp to apply phase retrieval algorithm. For fair, those algorithms are all accelerated by the GPU. Several examples of recovered images are shown in Figure 3.

Table 1. Some hyper parameters in TCNN.

	16×16	**40×40**	**64×64**
Training sets	3050	3050	35000
Validation sets	22	20	200
Test sets	5	5	6

From Figure 3, we can observe that pictures recovered by TCNN are competitive with state-of-arts. In fact, it is quite challenging for all these methods recovering high quality images from speckle patterns, since multiple scattering media deteriorates the light besides the phase information get lost by CCD. Moreover, the noise and system error also exist in the practical experiment. But Table 2 shows the time cost by TCNN is far less than other methods. It can save the time 10 to 100 folds. The advantage becomes clearer when 64×64 which demonstrates the possibility of TCNN to realize real time imaging. Moreover, TCNN only performs an end-to-end recovery but other methods must estimate the transformation matrix in advance then getting the evaluation by phase retrieval. Thus, the experiments fully illustrate the power of TCNN in recovering images from the intensity of the speckle pattern.

During the training step, the curves of training error and validation error of amplitude only SLM 16 ×16 are depicted in Figure 4. Considering the large quantities of the training set, training error is the Mean Square Error (MSE) of one of the inputs in training set selected randomly, as a result, there are some oscillations in training error curve. The validation error is the mean MSE of the validation set. From Figure 4, we can find the solution quickly converges to the local optimum after 20 epochs, there is little fluctuation for both errors. Besides the validation error gets close to the training error which fully demonstrates network relieves from over-fit.



(a) (b)

Figure 3. The reconstruction for amplitude-only SLM with image size (a) 16×16 and (b) 40×40.

Table 2. Time cost per image by different methods (“—” means TCNN don’t need iterations).

	16×16		**40×40**		**64×64**	
	Time(s)	Iterations	Time(s)	Iterations	Time(s)	Iterations
WF	4.11	100	19.49	100	289.03	100
GS	3.86	100	18.86	100	43.48	100
Phaselift	239.50	100	----	500	----	1000
PRVAMP	3.41	100	17.40	100	44.52	100
TCNN	0.103	----	0.117	----	0.121	----

The core of TCNN is to utilize two sections g_{Θ_1} and g_{Θ_2} to approximate the inverse function g which can simulate the imaging procedure in multiple scattering media, namely,

$$g(\mathbf{b}) \approx g_{\Theta}(\mathbf{b}) = g_{\Theta_2}\ \ g_{\Theta_1}(\mathbf{b})$$

For g_{Θ_1}, features of intensity of speckle patterns b are learned from different scales which can be seen from Figure 5. We can see that the details from different scales of speckle pattern are learned (the features in Figure 5 are resized to keep the same shape). Then, deconvolution procedures decode those features into a new element $g_{\Theta_1}(\mathbf{b})$ in the transform domain so that it can be approximated to x by transforming layer g_{Θ_2} .

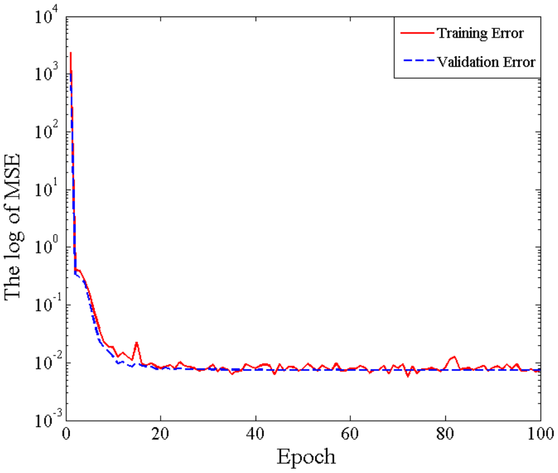


Figure 4. The training and validation errors.

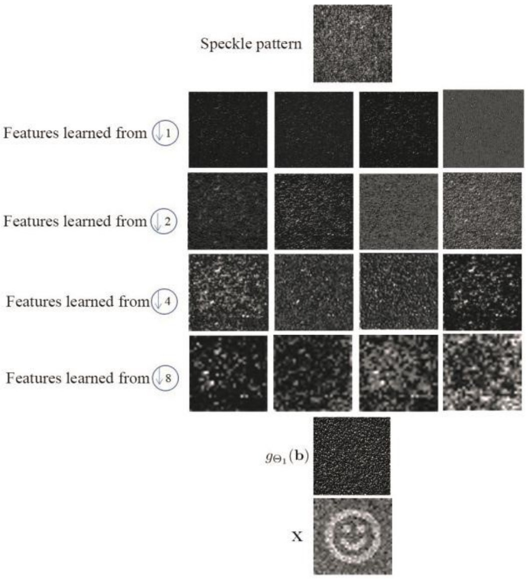


Figure 5. The features learned in TCNN.

4. CONCLUSION

In this paper, we designed a deep neural network TCNN to directly transform the intensity of the speckle pattern via the multiple scattering media into the image of an object. Compared to the traditional double phase retrieval methods, this end-to-end network has no need to model this imaging procedure and calculate the transformation matrix. Instead, it needs a large amount of training data which includes images of objects and their corresponding intensity of speckle patterns to update the parameters of TCNN.

In TCNN, two parts are built deliberately to approximate the inverse function. The nonlinear part encodes the information of the intensity of the speckle pattern besides decoding it into a new vector in the transform domain, then the linear part transforms this vector into the object domain. Compared to the popular methods, TCNN recovers images with competitive quality. Moreover, the time cost by TCNN is much less. Specifically, recovering per image needs no more than 1s.

In the future, the work is to decrease the number of parameters in TCNN. Especially for the parameters in the transformation layer, they occupy large portions of the whole parameters. We consider fuse it implicitly into the convolutional layers where the parameters are comparatively less.

ACKNOWLEDGEMENTS

This work was supported by National Natural Science foundation of China (No.61977065) and 173 Program of China (No.2020-JCJQ-ZD-029).

REFERENCES

[1] Velten, A., Willwacher, T., Gupta, O., Veeraraghavan, A., Bawendi, M. G. and Raskar, R., "Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging," Nat. Commun., 3(1), 745(2012).

[2] Waller, L. and Tian, L., "3D intensity and phase imaging from light field measurements in a led array microscope," Optica, 2(2), 104-111(2015).

[3] Freund, I. I., Rosenbluh, M. and Feng, S., "Memory effects in propagation of optical waves through disordered media," Phys. Rev. Lett., 61(2), 2328-2331(1988).

[4] Popoff, S. M., Lerosey, G., Carminati, R., Fink, M., Boccara, A. C. and Gigan, S., "Measuring the transmission matrix in optics: an approach to the study and control of light propagation in disordered media," Phys. Rev. Lett., 104(10), 100601(2010).

[5] Cui, M., "Parallel wavefront optimization method for focusing light through random scattering media," Opt. Lett., 36(6), 870-872(2011).

[6] Sharma, M. K., Metzler, C. A., Nagesh, S., Baraniuk, R. G., Cossairt, O. and Veeraraghavan, A., "Inverse scattering via transmission matrices: Broadband illumination and fast phase retrieval algorithms," IEEE Trans. Comput. Imaging, 6, 95-108(2020).

[7] Wang, S., Su, Z., Ying, L., Peng, X., Zhu, S., Liang, F., Feng, D. and Liang, D., "Accelerating magnetic resonance imaging via deep learning," Proc. ISBI, 514-517(2016).

[8] Jo, Y., Park S., Jung, J., Yoon, J., Joo, H., Kim, M. H., Kang, S. J., Choi, M. C., Lee, S. Y. and Park, Y., "Holographic deep learning for rapid optical screening of anthrax spores," Sci. Adv., 3(8), e1700606(2017).

[9] Dong, C., Loy, C. C., He, K. and Tang, X., "Image super-resolution using deep convolutional network," IEEE Trans. Pattern Anal. Mach. Intell., 38(2), 295-307(2016).

[10] Sinha, A., Barbastathis, G., Lee, J. and Li, S., "Lensless computational imaging through deep learning," Optica, 4(9), 1117-1125(2017).

[11] Rivenson, Y., Zhang, Y., Gnaydn, H., Teng, D. and Ozcan, A., "Phase recovery and holographic image reconstruction using deep learning in neural networks," Light Sci. Appl., 7(2), 17141(2017).

[12] Metzler, C. A., Schniter, P., Veeraraghavan, A. and Baraniuk, R., "prDeep: Robust phase retrieval with a flexible deep network," Proc. ICML, 3498-3507(2018).

[13] Romano, Y., Elad, M. and Milanfar, P., "The little engine that could: Regularization by denoising (red)," SIAM J. Imaging Sci., 10(4), 1804-1844(2017).

[14] Kingma, D. P. and Ba, J., "Adam: A method for stochastic optimization," Proc. ICLR, (2015).

[15] Metzler, C. A., Sharma, M. K., Nagesh, S., Baraniuk, R. G., Cossairt, O. and Veeraraghavan, A., "Coherent inverse scattering via transmission matrices: Efficient phase retrieval algorithms and a public dataset," Proc. ICCP, 1-16(2017).

[16] Balan, R., Casazza, P. and Dan, E., "On signal reconstruction without phase," Appl. Comput. Harmon. Anal., 20(3), 345-356(2006).

[17] Gerchberg, R. W., "A practical algorithm for the determination of phase from image and diffraction plane pictures," Optik, 35, 237-250(1971).

[18] Candès, E. J., Li, X. and Soltanolkotabi, M., "Phase retrieval via wirtinger flow: Theory and algorithms," IEEE Trans. Inf. Theory, 61(4), 1985-2007(2015).

[19] Candès, E. J., Strohmer, T. and Voroninski, V., "Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming," Commun. Pur. Appl. Math., 66(8), 1241-1274(2013).