IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 21, 2021, accepted April 23, 2021, date of publication May 4, 2021, date of current version May 14, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3077560

Mutual Transfer Learning of Reconstructing Images Through a Multimode Fiber or a Scattering Medium

XUETIAN LAI[1], QIONGYAO LI[1], XIAOYAN WU[2,3], GUODONG LIU[2,3], ZIYANG CHEN[1], AND JIXIONG PU[1], (Member, IEEE)

[1]College of Information Science and Engineering, Fujian Provincial Key Laboratory of Light Propagation and Transformation, Huaqiao University, Xiamen 361021, China
[2]Institute of Fluid Physics, China Academy of Engineering Physics, Mianyang 621900, China
[3]Key Laboratory of Science and Technology on High Energy Laser, China Academy of Engineering Physics, Mianyang 621900, China

Corresponding author: Jixiong Pu (jixiong@hqu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 11674111, and in part by the Fujian Province Science Funds for Distinguished Young Scholar under Grant 2018J06017.

ABSTRACT In recent years, convolutional neural network (CNN) has been successfully applied to reconstruct image from the speckle, which is generated as an object passes through a scattering medium or a multimode fiber (MMF). To reconstruct image from the speckle, the CNN must be trained with a large number of object-speckle pairs (training dataset), and the trained CNN is capable of reconstructing image from dataset (test dataset), which is taken in the same condition as the training dataset. However, in some cases, data type and the scattering medium may vary with the situation. In this case, the CNN has to be re-trained using a large number of new data taken from the new scattering media for reconstructing image. In this paper, we develop a CNN called as Mobiledense-net (MDN) to realize the *mutual transfer learning*. Specifically, the MDN is first pre-trained with a large number of object-speckle pairs taken from MMF or scattering slab, then tuned with quite small number of object-speckle pairs from scattering slab or MMF. It is shown that in this case the MDN can reconstruct image from the speckle with quite good quality, in which the speckle is taken from MMF or the scattering slab. We also show that using a more complex dataset for pre-training, the amount of data for pre-training can be largely reduced and reconstruction quality can be further improved. Using *transfer learning*, the reconstruction quality is quite good, being up to 99%. The results in this paper provide a more generalized method for studying the imaging through scattering imaging or MMF by using CNN.

INDEX TERMS Imaging, convolutional neural network (CNN), Mobiledense-net (MDN), transfer learning, scattering medium, multimode fiber (MMF).

I. INTRODUCTION

In recent years, there have been increasing interests in the field of computational imaging with deep learning. It has been demonstrated that deep neural networks can help to deal with problems that cannot be analytically addressed using conventional methods [1]. Thus, the deep learning schemes are applied effectively in areas such as computer vision, unmanned technology, and natural language processing *etc*. [2], [3], [4]. With a deeper understanding and increased processing speed, the deep neural networks are found to be used in the field of computational imaging. The deep neural networks can solve many remaining problems in fields of scattering imaging, computational ghost imaging, Fourier laminar microscopic imaging, phase retrieval and hologram data compression *etc*. [5]–[11]. It is shown that as an object passes through a scattering medium or multimode fiber (MMF), the information of the image is deteriorated. Therefore it is difficult for us to directly obtain the original image. In the past decade, various approaches have been proposed for imaging through scattering media or MMF, such as wavefront shaping, speckle correlation and

The associate editor coordinating the review of this manuscript and approving it for publication was Eduardo Rosa-Molinar.

digital phase conjugation *etc*. [12]–[16]. These approaches are found to be capable of reconstructing images from speckles. However, these approaches require very complicated and precise experimental control and measurement, and are susceptible to external disturbances. Therefore, a slight vibration of the scattering media may largely degrade the performance of reconstructing images. Recently many researchers show that apart from the limitations of traditional methods for scattering imaging, deep learning can provide a new kind of imaging technique with more sustainability and stability [5], [17]–[22]. Specifically, the convolutional neural networks (CNNs) can learn the input-output relationships effectively without prior knowledge of parameters of the optical system and light propagation processes. However, to achieve good imaging quality, it needs a large dataset from a certain scattering medium to train the CNN. When we shift to reconstruct images from speckles taken from other scattering medium or MMF, a lot of new training dataset is needed for training the CNN. It shows that this method is time-consumption, in which we must take a lot of dataset and use this dataset to train the CNN.

Fortunately, CNN is capable of extracting the spatial invariant information from speckle, which can be re-used to extract high-dimension features from new dataset. This kind of deep learning scheme is called as *transfer learning* [25]. In this paper, we design a CNN structure to perform the *transfer learning* between MMF and scattering slabs. The process of scattering can be described as $y = F(x)$, where x is the input of the optical system and y is the measured speckle [5]. The role of the CNN is to reverse this process to $x = F^{-1}(y)$. The advantage of using the CNN is that we can obtain $F^{-1}(\cdot)$ without knowing the process of light propagation. The CNN performs end-to-end mapping to reconstruct images. Once the CNN is trained by speckles form scattering slab or MMF, the parameters of the network are well fitted to the objective feature. For example, the CNN is pre-trained by speckles from MMF to reconstruct handwritten digit images, the parameters of each layer in this trained network are well trained to extract digit structural features (such as arches, angles, horizontal and vertical structure *etc*.) from speckles. Therefore, when the pre-trained network is used to reconstruct images through scattering slabs, we only need to tune the parameters with much few speckles from scattering slabs. It is apparent that this training scheme needs less training data and saves training time.

The CNN used in this work is built based on U-net [26], dense-net [27] and mobilenet [28], and is called as ''Mobiledense-net (MDN)''. The U-Net architecture is first proposed for biomedical image segmentation and then be used in variety of image relevant tasks. It has been shown that, the U-Net is of high efficiency in end-to-end image processing tasks. The dense-net encourages feature reuse and the mobilenet is often used in mobile and embedded vision applications, which is lightweighted. The experimental results indicate that the MDN has strong robustness and few parameters, making the training process effectively. In the experiment, we pre-train the MDN with a lot of object-speckle pairs from MMF, and then use a few speckles taken from scattering slab to tune the pre-trained model. Likewise, we pre-train the MDN with a relatively large number of speckles taken from three different scattering slabs respectively. The pre-trained models are then tuned by a small number of speckles taken from MMF. It is shown that these trained MDNs is capable of reconstructing object images from speckles taken from both scattering slabs and MMF. We define this kind of *transfer learning* as the *mutual transfer learning*. It will be shown that the *mutual transfer learning* through scattering slab or MMF can reconstruct images from speckles with high fidelity.

II. MATERIALS AND METHODS

A. NETWORK STRUCTURE

The network structure is consisted of U-net [26] and dense blocks [27]. Specifically, each dense block contains a depthwise separable convolution from Mobilenet [28]. The architecture diagram is shown in Fig. 1, which follows the symmetrical path of encoder and decoder [29], [30] of the U-net and replaces part of the convolutional layers in typical U-net with dense blocks. The dense block contains several sets, each consisting of a depthwise separable convolution, a batch normalization layer, a rectified linear unit activation, a convolutional layer, and a dropout layer. Among these layers, depthwise separable convolution is proved to possess less parameters compared with a standard convolution, thus it is often used in some lightweight networks, such as MobileNetV1, MobileNetV3 [31]. Here we apply it to each dense block in our network for reducing the number of parameters.

The input of the CNN is a processed 64×64 speckle intensity image. The image is first encoded through 4 dense blocks, each followed by a max pooling layer for down-sampling and the feature maps become a 4×4 low-solution image on the bottom. Then, the feature maps are decoded through 4 up-sampling convolutional layers and dense blocks. The skip connection helps to avoid pixel loss. The corresponding spatial scale feature information in encode and decode path are tunneled through skip connection. Finally, the network output is obtained after two convolutional layers.

The CNN is trained by using our experimental data and the trained CNN can reconstruct images from speckles. In Fig.2, we show the visualization of internal activation maps between the layers (the letter G is taken as example). The feature extracted from each layer is gradually shown to be representative as the layers go deeper, demonstrating the network's ability to extract spatial statistical characteristics from highly complex speckles.

B. EXPERIMENTAL SETUP

Figure 3 is the experimental setup, in which the object patterns are displayed on a phase-only spatial light modulator (SLM, 1920×1080 pixels, Pluto-Vis, Holoeye). The laser

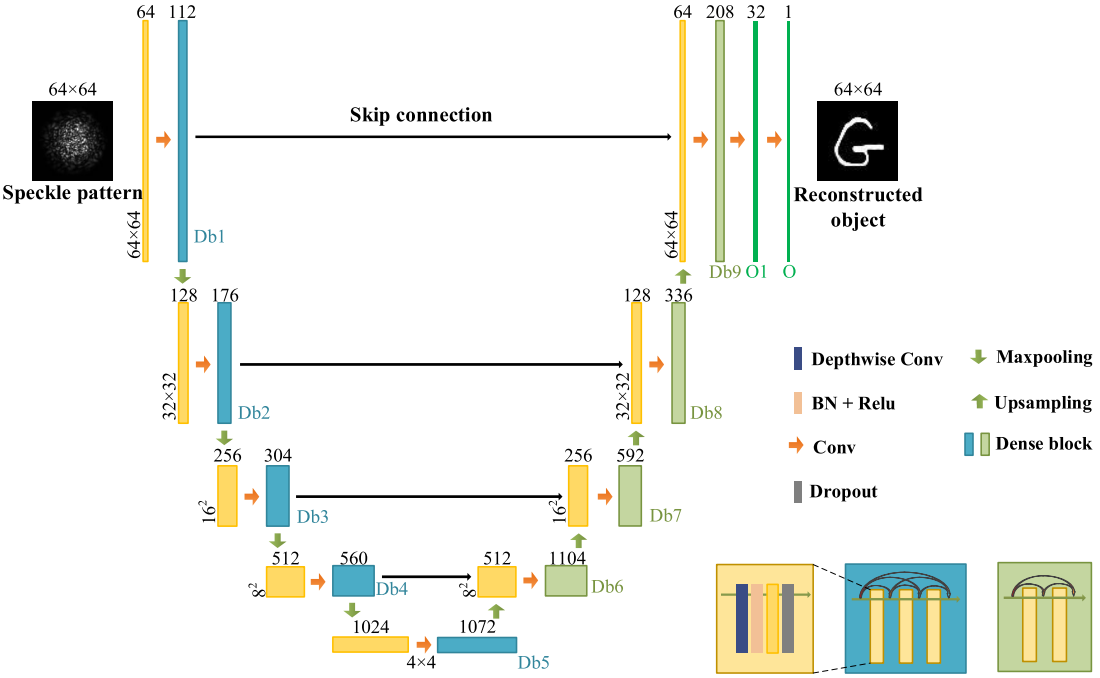


FIGURE 1. The diagram of the proposed MDN architecture, which is consisted of a general U-net and dense blocks behind convolution. Each dense block consists of multiple subblocks that each is made up of depthwise conv+BN+Relu+conv+dropout. The input speckle goes through the encode and decode path to generate a one channel reconstructed image pixel by pixel.

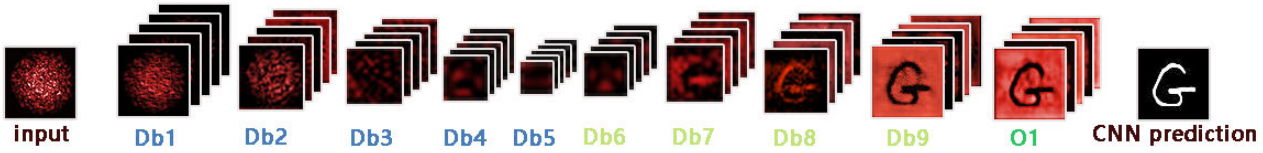


FIGURE 2. Visualization of internal activation maps. The layers shown here are the last layer in dense blocks (Db1-Db9, which are marked in blue color in Fig. 1). The spatial information of CNN input is gradually resembled as it flows through the encode-decode path. The final output is obtained after an additional convolution layer (O1).

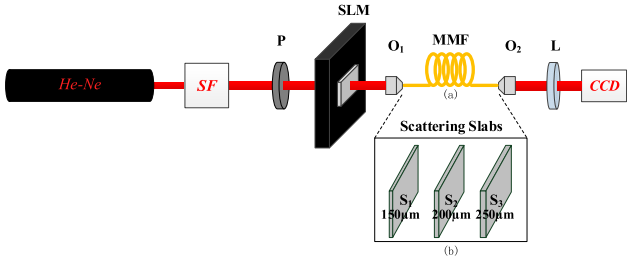


FIGURE 3. Experiment setup for realizing imaging based on deep learning, in which the object images pass through MMF or scattering slabs. (a) The light with object information couples into MMF, and the speckle is collected by O2 and lens (L) (b) the light with object information passes through three scattering slabs, respectively. The output speckles are obtained by CCD. P is a polarizer and SF is a spatial filter.

beam of He-Ne laser passes through a spatial filter (SF) and a horizontal polarizer (P), and illustrates the SLM. As the light beam with object information passes through MMF (62.5μm core diameter, NA of 0.25, Thorlabs), the speckle is recorded by a CCD (Chameleon 3, 1024 × 1280 pixels, Mono, Point Grey). A lens (L) is used to collect scattered light in the front of the CCD. The change of the object will result in slight change of the speckle. We program the SLM to load object patterns one by one and collect a lot of object-speckle pairs for training our network. We replace the MMF by scattering slabs of three different thickness (the thickness of the scattering slabs are 150μm, 200μm and 250μm, respectively), and the speckle is obtained as a light beam with object information passes through a scattering slab. Similarly, the datasets with a lot of object-speckle pairs can be obtained through scattering slabs, for training the network.

Here, handwritten digits from MNIST [32], handwritten letters from NIST [33], CIfar-10 [34] and Fashion-mnist [39] are used as object patterns. Before loaded onto the SLM, all object patterns are resized into 512 × 512 pixels and handwritten digits and letters are binarized. The corresponding speckles with the same size are collected. We collect a lot of object -speckle pairs in experiment, which are as following:

M1 : The object patterns are handwritten digits, and the CCD collects 11,000 speckle images as the objects pass through MMF.

M2 : The object patterns are handwritten letters, and the CCD collects 6000 speckle images as the objects pass through MMF.

M3 : The object patterns are Cifar-10 images, and the CCD collects 2000 speckle images as the objects pass through MMF.

M4 : The object patterns are Fashion-mnist images, and the CCD collects 4000 speckle images as the objects pass through MMF.

S1 : The object patterns are handwritten digits, and the CCD collects 11000 speckle images as the objects pass through scattering slab with thickness of 150μm.

S2 : The object patterns are handwritten digits, and the CCD collects 11000 speckle images as the objects pass through scattering slab with thickness of 200μm.

S3 : The object patterns are handwritten digits, and the CCD collects 11000 speckle images as the objects pass through scattering slab with thickness of 250μm.

S4 : The object patterns are Fashion-mnist images, and the CCD collects 4000 speckle images as the objects pass through scattering slab with thickness of 150μm.

III. DATA PROCESSING

All the speckle images of different datasets are divided randomly into training sets and test sets. The training speckles are downsampled to 64×64 pixels, normalized between 0, 1. Images of digits and letters are thresholded by setting all the nonzeros to 1 and also resized into 64×64 pixels. The pixel value of each image after processing is the classification category of corresponding speckle pixel point, namely, 0, and 1. During training, the model optimizer is Adam [35], and the loss function is binary cross-entropy. We introduce Pearson correlation coefficient (PCC) [36], Peak Signal to Noise Ratio (PSNR), and Structure Similarity (SSIM) [37] *etc*. to evaluate the imaging quality.

PCC is used to calculate the similarity between features and categories to determine whether the extracted features and categories are positively correlated, negatively correlated or not correlated. PCC formula is defined as follows [36]:

$$PCC = \frac{\sum_{i=1}^{H}\sum_{j=1}^{W}\hat{y}(i,j)-\mu_{\hat{y}})(y(i,j)-\mu_y)}{\sqrt{\sum_{i=1}^{H}\sum_{j=1}^{W}(\hat{y}(i,j)-\mu_{\hat{y}})^2}\sqrt{\sum_{i=1}^{H}\sum_{j=1}^{W}(y(i,j)-\mu_y)^2}}, \quad (1)$$

where H, W represent the height and width of the image, $\hat{y}(i,j)$ and $y(i.j)$ represent the pixel value of row i, column j of reconstructed image and the object image respectively. $\mu_{\hat{y}}$ and μ_y refer to the average gray-value of the reconstructed image and the object image.

PSNR is a widely used method for calculating similarity between images, and it is defined as [37]:

$$PSNR = 20\log_{10}\frac{2^n-1}{\sqrt{MSE}} \quad (2)$$

$$MSE = \frac{1}{H*W}\sum_{i=1}^{H}\sum_{j=1}^{W}(\hat{y}(i,j)-y(i,j)), \quad (3)$$

TABLE 1. Average PCC, PSNR, and SSIM of test results for M1, M2, and M1_M2.

Training Data-set	PCC(dB)	SSIM	PSNR
M1	0.79	0.86	14.94
M2	0.74	0.93	18.50
M1_M2	0.72	0.92	17.97

where n is the number of bits per pixel of the image, here is 8, that is, the gray scale of the pixel is 256.

SSIM is an index for measuring the structural similarity between two images, which is more in line with the observation law of human eyes on natural things. SSIM formula is written as:

$$SSIM = \frac{(2\mu_y\mu_{\hat{y}}+c_1)(2\sigma_{y\hat{y}}+c_2)}{(\mu_y^2+\mu_{\hat{y}}^2+c_1)(\sigma_y^2+\sigma_{\hat{y}}^2+c_2)}, \quad (4)$$

where, $\sigma_{\hat{y}y}$ is the covariance between reconstructed image and original object. $\sigma_{\hat{y}}$ and σ_y represent the standard deviation of reconstructed image and object image, respectively. c_1 and c_2 are constants.

IV. EXPERIMENTAL RESULTS

A. IMAGE RECONSTRUCTION THROUGH MMF

First of all, we apply the MDN to reconstruct images from speckles through MMF. The training steps and results are presented as follows:

1) The trained MDN is used to reconstruct digitals from speckles. We train the MDN by using 10k speckles of M1 and test it by 1k speckles from M1. Representative examples of the speckles and reconstruction pairs are shown on the left column of Fig. 4(a).
2) The trained MDN is used to reconstruct letters from speckles. We train the MDN with M2 training set of 5k speckles and test it by using 1k test speckles. The test results are presented on the right column of Fig. 4(a).
3) We apply *transfer learning* to do the reconstruction task. We use the trained MDN (from step 1), which has been trained by 10k speckles of digits, as prior model. Then we tune it by using only 1k speckles of letters and test the final model performance on 1k speckles of letters. Figure 4(b) presents the reconstructed letters from the speckles. It shows that, even we train the network with a few (1k) speckles of letters, the reconstructed images are of high-quality. It is due to that the MDN has the prior knowledge in which it has been trained using a lot of speckles of digits.

These results indicate that the MDN can extract and classify the information from the highly complex speckles through MMF, to reconstruct object images with good quality. Table 1 illustrates the average PCC, SSIM, PSNR of the test results of the above three experiments. During training, the MDN makes pixel-wise classification, learning the per-pixel input-output relation. Accordingly, the results demonstrate superior generalization of the MDN and the model obtained after *transfer learning* retains good performance, just as the

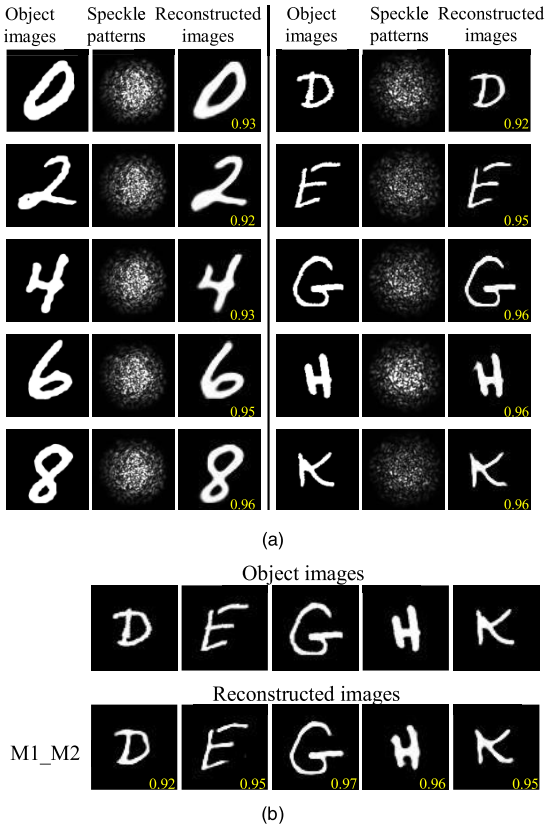


FIGURE 4. MDN test results through MMF. (a) The left column: M1 (digits), the right column: M2 (letters). In each column, from the left to the right: ground truths images, corresponding speckles and test results, (b) the performance of the transfer learning, in which the MDN is trained by 10 K speckles of digitals, and tuned by 1K speckles of letters. The trained MDN can reconstruct letters with quite good quality. The PCC value is given in yellow color (other figures are the same).

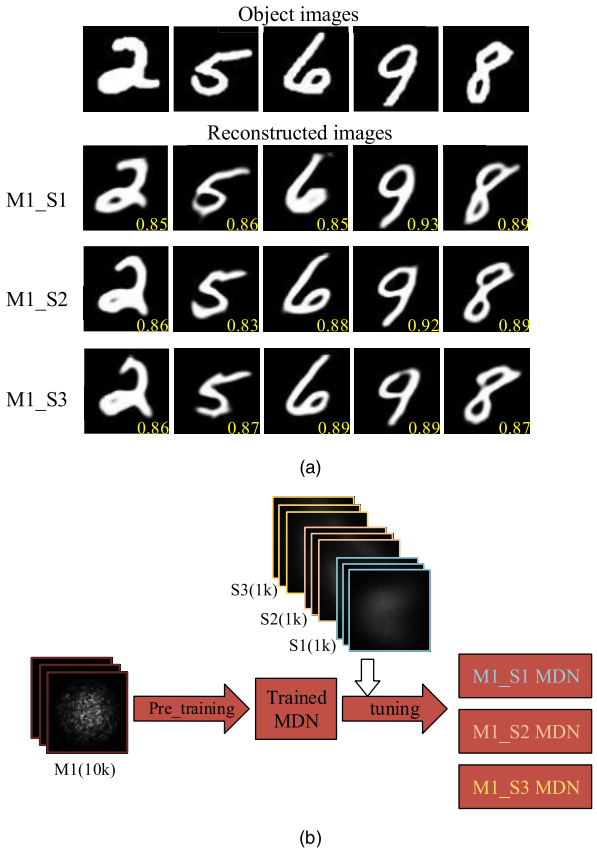


FIGURE 5. The performance of the trained MDN for achieving the transfer learning from MMF to scattering slabs. The MDN has been pre-trained by using each 10k object-speckle pairs from MMF (M1), and each 1k speckles from each scattering slab Si (i = 1,2,3) are used for tuning the pre-trained MDN. (a) The performance of the reconstructed images from speckles from S1, S2 and S3, respectively, (b)the schematic diagram of training process.

MDN does when it is directly trained by a lot of speckles of M2.

B. MUTUAL TRANSFER LEARNING BETWEEN MMF AND SCATTERING SLABS

Now we want to demonstrate that the MDN does have the characteristic of *mutual transfer learning* between MMF and scattering slabs. The experimental results are presented as follows:

1) First of all, we prove the *transfer learning* from MMF to scattering slabs for reconstructing images as the objects pass through three different scattering slabs. In the experiment, we take a large number of speckles (10k) from M1 (MMF dataset) for pre-training the MDN, then take a small number of speckles from S1, S2, and S3 (1k speckles from each scattering slab) for tuning the MDN, respectively. To test the obtained model generalization ability, 10k speckles from each S1, S2, and S3 are used as test set respectively, in which these test data are never used in training. The test results are shown in Fig. 5(a). Since the weights used to initialize the pre-trained model is the optimal one which is obtained after training M1, the speckle features of the tuning data can be well extracted based on the pre-trained model. Therefore, the desired new model can be obtained quickly, within a few minutes. The training process is displayed in Fig. 5(b). The average PCC, PSNR and SSIM of the test results are shown in Table 2. It is shown from Table 2 that the fidelity of reconstructing image is quite high.
2) The second experiment will demonstrate the *transfer learning* from each of three different scattering slabs to MMF for reconstructing images as the objects pass

TABLE 2. The average PCC, PSNR, SSIM values of the transfer learning M1_S1, M1_S2, M1_S3, S1_M1, S2_M1, S3_M1, and Mix_M1, respectively. The average PCC, PSNR, SSIM values are calculated by using 10k test data.

Training Data-set	PCC	SSIM	PSNR
M1_S1	0.79	0.85	15.00
M1_S2	0.79	0.86	15.00
M1_S3	0.79	0.86	15.00
S1_M1	0.65	0.80	13.00
S2_M1	0.70	0.76	13.12
S3_M1	0.70	0.80	13.02
Mix_M1	0.69	0.69	12.90

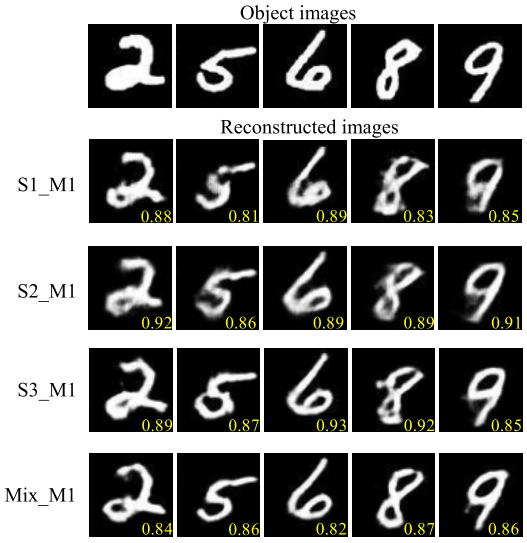


FIGURE 6. The performance of the trained MDN for achieving the transfer learning from scattering slabs to MMF. The MDN has been pre-trained by using each 10k object-speckle pairs from S1, S2, and S3, respectively, or 9k mixture of S1, S2, and S3. 1k speckles from MMF are used for tuning the pre-trained MDN. (a) The performance of the reconstructed images from speckles taken from MMF (M1). The PCC value in yellow color is marked, (b)the schematic diagram of the training process.

through MMF. In the experiment, each 10k speckles from S1, S2, and S3 respectively, and a totally 9k speckles from mixture of S1, S2, S3 (each 3k speckles), are used for pre-training the MDN. 1k speckles from M1 are used for tuning the pre-trained MDNs. The training process is displayed in Fig. 6(b). The trained MDN is then tested by 10k speckles from M1, in which these test data have never been used in training. The test results are shown in Fig. 6(a), and the average PCC, PSNR and SSIM of the reconstructed images are shown in Table 2.

It can be seen from the results that both *transfer learning* tasks do have high imaging fidelity. The results indicate that using a lot of speckles from scattering slab or MMF to train the MDN as a prior model, we can reconstruct images from speckles of other scattering slab or MMF. From Table 2, we find that the imaging quality for *transfer learning* from MMF to scattering slab is slightly better than that for *transfer learning* from scattering slab to MMF. This is due to the difference of intensity distribution between the pre-training dataset and tuning dataset. We will further discuss this problem in Section IV-D.

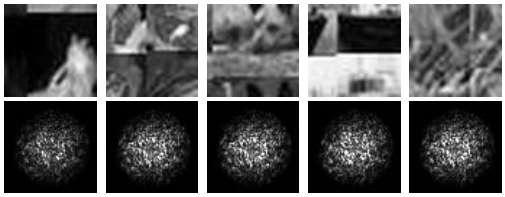


FIGURE 7. Examples of Cifar-10 images and the corresponding speckles for the objects passing through MMF.

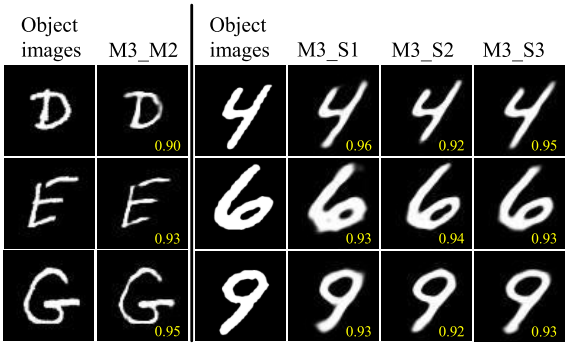


FIGURE 8. Transfer learning results, in which the MDN is pre-trained by using M3(Cifar-10 images). The letters are the test results on transfer learning task within MMF from M3(natural images) to M2 (letters).

TABLE 3. Average PCC, PSNR, SSIM values for transfer learning M3_M2, M3_S1, M3_S2, and M3_S3, respectively.

Training Data-set	PCC	SSIM	PSNR
M3_M2	0.71	0.92	18.01
M3_S1	0.85	0.86	15.95
M3_S2	0.84	0.87	15.70
M3_S3	0.85	0.87	15.91

C. TRANSFER LEARNING WITH OTHER DATASETS

1) IMPROVEMENT OF MUTUAL TRANSFER LEARNING

In order to further reduce the amount of pre-training data, we use M3 (2k) (image-speckle pairs as shown in Fig. 7) for pre-training the MDN. It is shown that these images are more complex than digit, letter and Fashion-mnist images. Using these relatively complex data to pre-train the CNN can help us reduce data requirement [38].

The loss function for pre-training the MDN is *MSE*. After pre-training, 1k speckles from M2, S1, S2, and S3, respectively are used for tuning the model with loss function binary-crossentropy. The test results are shown in Fig. 8, and the performance of the *transfer learning* is evaluated by PCC, PSNR and SSIM, which are presented in Table 3.

Compared with Table2, the results in Table3 indicate that the pre-trained model using M3 possesses stronger generalization ability than that using M1, showing better imaging quality. Accordingly, with the method of *transfer learning*, the MDN is capable of reconstructing objects of different classes. Moreover, the data requirement to pre-train the MDN

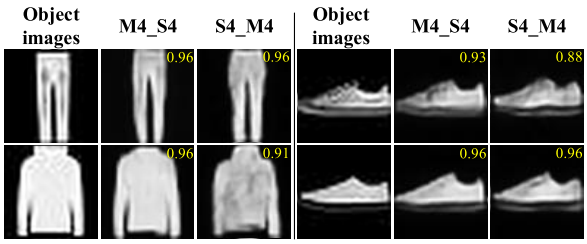


FIGURE 9. Examples of Cifar-10 images and the corresponding speckles for the objects passing through MMF.

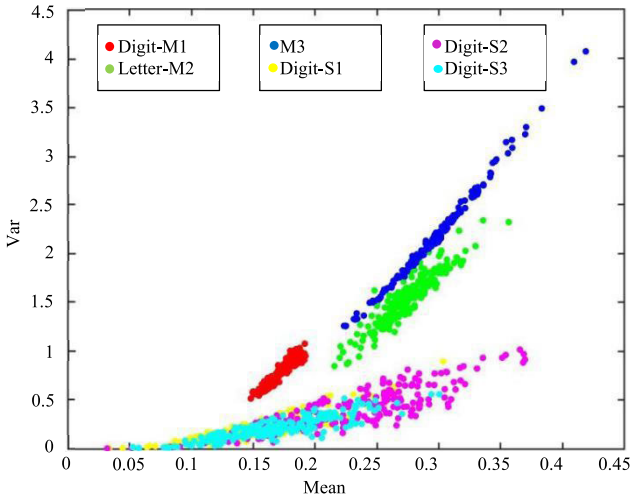


FIGURE 10. Mean-variance diagram of the six datasets.

for reconstructing objects through scattering media or MMF is notably reduced. This method provides great convenience for object reconstruction in cases that the scattering medium turns to another different one or MMF.

2) RECONSTRUCTING GRAYSCALE OBJECT IMAGES

The *mutual transfer learning* can be used to reconstruct grayscale object images between scattering slab and MMF. Now we show the *transfer learning* from MMF to a scattering slab. To do this, we use 4k speckles of M4 for pre-training, and randomly select 1k speckles of S4 for tuning. The rest 3k speckles of S4 are used for testing. For the *transfer learning* from a scattering slab to MMF, we use 4k speckles of S4 for pre-training, and randomly select 1k speckles of M4 for tuning. The test results are shown in Fig 9. It is shown from Fig. 9 that the test results are of high PCC value, indicating that the MDN can reconstruct almost the texture information of the images. Compared to digit and letter images, these gray images are more complex, indicating that the MDN can build more complex mappings of object-speckle pairs for reconstruction.

D. DATA ANALYSIS AND NETWORK DISCUSSION

1) DATA ANALYSIS

Transfer learning is used to enhance the generalization ability of a trained network across different datasets. Therefore, the difference of data distribution between source domain and target domain [25] will affect the network performance. We analyze the influence of the distribution of speckle intensity on the MDN test results. We randomly select 200 speckles

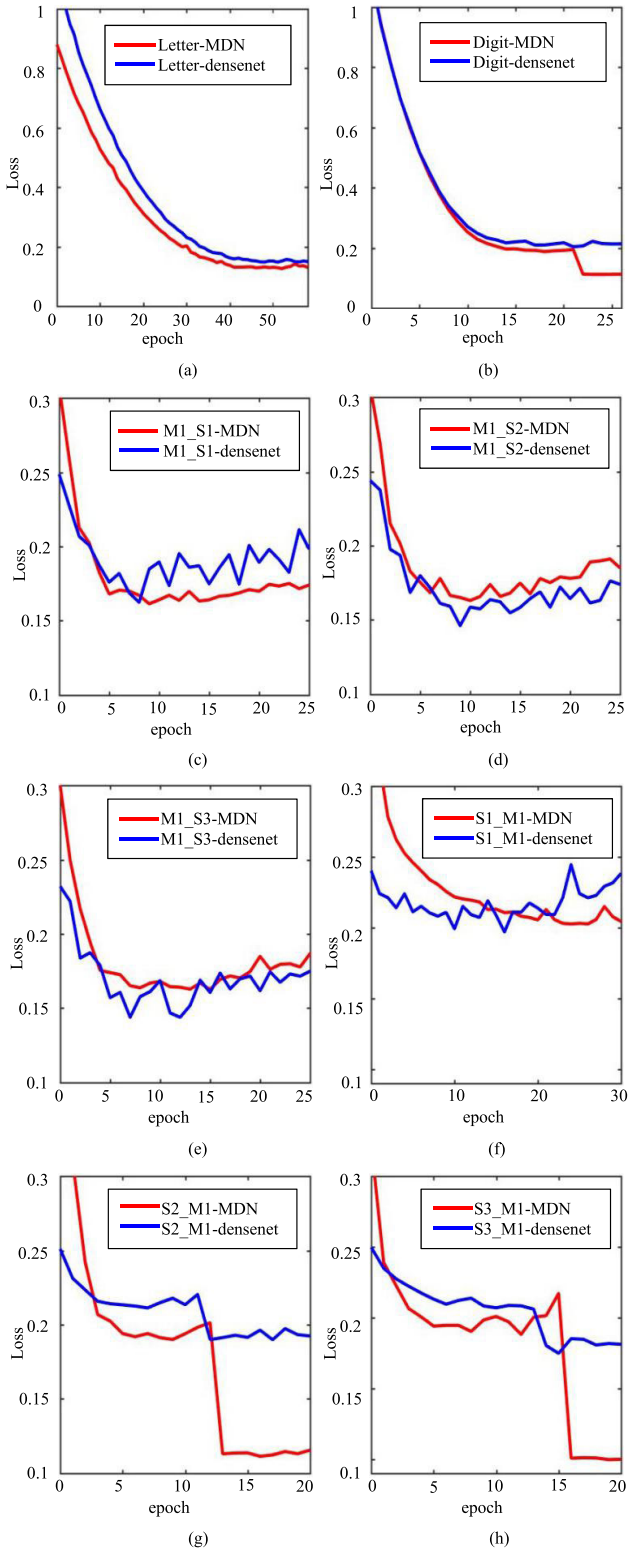


FIGURE 11. Comparison of validation loss curves of MDN and dense-net. (a) trained on M1, (b) trained on M2. (c), (d) and (e) are validation loss curves of transfer learning from MMF to scattering slabs. (f), (g) and (h)are validation loss curves of transfer learning from scattering slabs to MMF.

from each of the six datasets (M1∼M3, S1∼S3), and calculate their average means and variances respectively. Figure 11 presents the diagram of variance as a function of mean. It can be seen that the distributions of speckles from S1, S2 and S3 are relatively close, and their distributions are far different from those of M1∼M3. It seems that the intensity distribution of speckles for pre-training and tuning are significantly different. Moreover, the distribution of M1 is relatively concentrated, so it is more difficult for the MDN to perform pixel-wise classification and to reconstruct image. While the speckle distributions of S1, S2, and S3 are more uniform, making it easier for the model to do classification. In other words, the MDN performs better when it is pre-trained with data that is hard to be classified and then tuned with data that is easy to be classified. So, the average evaluation values of *transfer learning* from MMF to scattering slab in Table 2 is better than that from scattering slab to MMF. Based on this reason, M1 and M3 are both chosen for pre-training.

2) NETWORK DISCUSSION

Our MDN uses depthwise separable convolution and dense blocks to help reduce training parameters and improve network stability. For comparison, we remove the depthwise separable convolution from the dense block, so that the network becomes ''dense-net'', and we use the method of training MDN to train this new network. The validation loss curves during training are recorded and shown in Fig. 11. The curves of Fig. 11(a) and (b) indicate the validation losses of the MDN and dense-net trained by M1 and M2 respectively. The curves from Fig. 11(c) to Fig. 11(h) are the *transfer learning* validation losses of MDN and dense-net, where Fig. 11(c) to Fig. 11(e) indicate *transfer learning* from MMF to scattering slabs, Fig. 11(f) to Fig. 11(h) indicate *transfer learning* from scattering slabs to MMF. It can be seen from the curves that the MDN is more stable than dense-net during training, since the curves of dense-net oscillates severely. These curves show that MDN achieves faster loss decrease and lower loss value than dense-net. We also replace all the dense blocks by standard convolutional layers to make it a typical U-net, which has parameters over 30-million [24]. Compared to the typical U-net, the number of parameters of our network is reduced by nearly 63%, which improves the training efficiency. Therefore a pre-trained MDN can realize *transfer learning* to reconstruct images quickly from different kinds of speckles from MMF or scattering medium.

V. CONCLUSION

In this paper, we designed a CNN structure to realize the *mutual transfer learning* between MMF and scattering slabs. The CNN is built based on U-net, dense-net and mobilenet, and is called as ''Mobiledense-net (MDN)''. The MDN is pre-trained using a large number of object-speckle pairs from MMF, then tuned with quite small number of object-speckle pairs from scattering slab. This trained network can reconstruct images with high fidelity from the speckles taken from MMF or scattering slabs. Conversely we can pre-train the MDN with a large number of object-speckle pairs from the scattering slabs, and tune the network with quite small number of object-speckle pairs from MMF. It has been shown that the trained MDN can recover images from the speckles taken from MMF with high fidelity. It was also shown that training the MDN with a relatively complex dataset can help enhance network's *transfer learning* ability, and greatly reduce the training data. These results in this paper may provide a more generalized method for the study of scattering imaging and improve the quality of imaging across MMF and scattering media using CNN.

REFERENCES

[1] Y. LeCun, Y. Bengio, and G. Hinton, ''Deep learning,'' *Nature*, vol. 521, pp. 436–441, 2015.
[2] S. Hoermann, M. Bach, and K. Dietmayer, ''Dynamic occupancy grid prediction for urban autonomous driving: A deep learning approach with fully automatic labeling,'' in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2018, pp. 2056–2063.
[3] L. Aziz, M. S. B. Haji Salam, U. U. Sheikh, and S. Ayub, ''Exploring deep learning-based architecture, strategies, applications and current trends in generic object detection: A comprehensive review,'' *IEEE Access*, vol. 8, pp. 170461–170495, 2020.
[4] C.-H. Chiang, C.-H. Kuo, C.-C. Lin, and H.-T. Chiang, ''3D point cloud classification for autonomous driving via dense-residual fusion network,'' *IEEE Access*, vol. 8, pp. 163775–163783, 2020.
[5] R. Horisaki, R. Takagi, and J. Tanida, ''Learning-based imaging through scattering media,'' *Opt. Exp.*, vol. 24, no. 13, pp. 13738–13743, 2016.
[6] Y. He, G. Wang, G. Dong, S. Zhu, H. Chen, A. Zhang, and Z. Xu, ''Ghost imaging based on deep learning,'' *Sci. Rep.*, vol. 8, no. 1, pp. 1–7, Dec. 2018.
[7] T. Nguyen, Y. Xue, Y. Li, L. Tian, and G. Nehmetallah, ''Deep learning approach for Fourier ptychographic microscopy,'' *Opt. Exp.*, vol. 26, pp. 26470–26484, 2018.
[8] J. R. Fienup, ''Phase retrieval algorithms: A comparison,'' *Appl. Opt.*, vol. 21, no. 15, pp. 2758–2769, Aug. 1982.
[9] S. Jiao, Z. Jin, C. Chang, C. Zhou, W. Zou, and X. Li, ''Compression of phase-only holograms with JPEG standard and deep learning,'' *Appl. Sci.*, vol. 8, no. 8, p. 1258, Jul. 2018.
[10] T. Shimobaba, D. Blinder, M. Makowski, P. Schelkens, Y. Yamamoto, I. Hoshi, T. Nishitsuji, Y. Endo, T. Kakue, and T. Ito, ''Dynamic-range compression scheme for digital hologram using a deep neural network,'' *Opt. Lett.*, vol. 44, no. 12, pp. 3038–3041, 2019.
[11] R. V. Vinu, Z. Chen, R. K. Singh, and J. Pu, ''Ghost diffraction holographic microscopy,'' *Optica*, vol. 7, no. 12, pp. 1697–1704, 2020.
[12] H. He, Y. Guan, and J. Zhou, ''Image restoration through thin turbid layers by correlation with a known image,'' *Opt. Exp.*, vol. 21, no. 10, pp. 12539–12545, 2013.
[13] 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*, vol. 491, no. 7423, pp. 232–234, Nov. 2012.
[14] Z. Yaqoob, D. Psaltis, M. S. Feld, and C. Yang, ''Optical phase conjugation for turbidity suppression in biological samples,'' *Nature Photon.*, vol. 2, no. 2, pp. 110–115, Jan. 2008.
[15] I. Freund, ''Looking through walls and around corners,'' *Phys. A, Stat. Mech. Appl.*, vol. 168, no. 1, pp. 49–65, Sep. 1990.
[16] I. Freund, M. Rosenbluh, and S. Feng, ''Memory effects in propagation of optical waves through disordered media,'' *Phys. Rev. Lett.*, vol. 61, no. 20, pp. 2328–2331, Nov. 1988.
[17] Y. Li, Y. Xue, and L. Tian, ''Deep speckle correlation: A deep learning approach toward scalable imaging through scattering media,'' *Optica*, vol. 5, no. 10, pp. 1181–1190, 2018.
[18] N. Borhani, E. Kakkava, C. Moser, and D. Psaltis, ''Learning to see through multimode fibers,'' *Optica*, vol. 5, pp. 960–966, Aug. 2018.
[19] A. Sinha, J. Lee, S. Li, and G. Barbastathis, ''Lensless computational imaging through deep learning,'' *Optica*, vol. 4, no. 9, pp. 1117–1125, 2017.
[20] K. Wang, D. J. Dou, Q. Kemao, J. Di, and J. Zhao, ''Y-Net: A one-to-two deep learning framework for digital holographic reconstruction,'' *Opt. Lett.*, vol. 44, pp. 4765–4768, Oct. 2019.

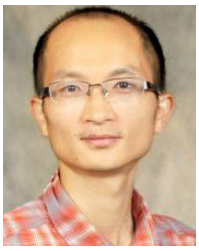
[21] S. Li, M. Deng, J. Lee, A. Sinha, and G. Barbastathis, "Imaging through glass diffusers using densely connected convolutional networks," *Optica*, vol. 5, no. 7, pp. 803–813, 2018.
[22] J. Zhao, X. Ji, M. Zhang, X. Wang, Z. Chen, Y. Zhang, and J. Pu, "High-fidelity imaging through multimode fibers via deep learning," *J. Phys., Photon.*, vol. 3, no. 1, Jan. 2021, Art. no. 015003.
[23] L. Wu, J. Zhao, M. Zhang, Y. Zhang, X. Wang, Z. Chen, and J. Pu, "Deep learning: High-quality imaging through multicore fiber," *Current Opt. Photon.*, vol. 4, pp. 286–292, 2020.
[24] Q. Li, J. Zhao, Y. Zhang, X. Lai, Z. Chen, and J. Pu, "Imaging reconstruction through strongly scattering media by using convolutional neural networks," *Opt. Commun.*, vol. 477, Dec. 2020, Art. no. 126341.
[25] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
[26] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segatimenton," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, 2015, pp. 234–241.
[27] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.
[28] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*. [Online]. Available: http://arxiv.org/abs/1704.04861
[29] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
[30] X.-J. Mao, C. Shen, and Y.-B. Yang, "Image restoration using convolutional auto-encoders with symmetric skip connections," 2016, *arXiv:1606.08921*. [Online]. Available: http://arxiv.org/abs/1606.08921
[31] A. Howard, M. Sandler, B. Chen, W. Wang, L.-C. Chen, M. Tan, G. Chu, V. Vasudevan, Y. Zhu, R. Pang, H. Adam, and Q. Le, "Searching for MobileNetV3," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 1314–1324, doi: 10.1109/ICCV.2019.00140.
[32] L. Deng, "The MNIST database of handwritten digit images for machine learning research [Best of the Web]," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 141–142, Nov. 2012. [Online]. Available: http://yann.lecun.com/exdb/mnist/
[33] P. J. Grother and K. K. Hanaoka, "NIST special database 19 handprinted forms and characters database," *World Wide Web-Internet Web Inf. Syst.*, pp. 1–30, 2016, doi: 10.18434/T4H01C.
[34] A. Krizhevsky, V. Nair, and G. Hinton. *CIFAR-10 (Canadian Institute for Advanced Reasearch)*. Accessed: Jan. 2020. [Online]. Available: http://www.cs.toronto.edu/~kriz/cifar.html
[35] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent.*, San Diego, CA, USA, 2015, pp. 1–15.
[36] A. G. Asuero, A. Sayago, and A. González, "The correlation coefficient: An overview," *Crit. Rev. Anal. Chem.*, vol. 36, no. 1, pp. 41–59, 2006.
[37] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? A new look at signal fidelity measures," *IEEE Signal Process. Mag.*, vol. 26, no. 1, pp. 98–117, Jan. 2009.
[38] Y. Yang, L. Deng, P. Jiao, Y. Chua, J. Pei, C. Ma, and G. Li, "Transfer learning in general lensless imaging through scattering media," 2019, *arXiv:1912.12419*. [Online]. Available: http://arxiv.org/abs/1912.12419
[39] H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms," 2017, *arXiv:1708.07747*. [Online]. Available: http://arxiv.org/abs/1708.07747

XUETIAN LAI was born in Fujian, China, in 1996. She is currently pursuing the master's degree in electronics and communication engineering with Huaqiao University, Fujian. Her research interest includes imaging through the scattering medium using deep learning.

QIONGYAO LI was born in Fujian, China, in 1994. She is currently pursuing the master's degree with Huaqiao University, Fujian. Her research interests include multiple scattering, transmission characteristics of multimode fiber, and quantum information.

XIAOYAN WU received the Ph.D. degree from Jilin University, Changchun, China, in 2016. He is currently an Associate Research Fellow with the Institute of Fluid Physics, China Academy of Engineering Physics, Mianyang, China. He is also a member of the Key Laboratory of Science and Technology on High Energy Laser, China Academy of Engineering Physics. His current research interests include the intelligent optics and electro-optical countermeasure technology.

GUODONG LIU received the Ph.D. degree from the China Academy of Engineering Physics, Mianyang, China, in 2015. He is currently an Associate Research Fellow with the Institute of Fluid Physics, China Academy of Engineering Physics. He is also a member of the Key Laboratory of Science and Technology on High Energy Laser, China Academy of Engineering Physics. His current research interests include the intelligent optics and laser-matter interaction.



ZIYANG CHEN received the Ph.D. degree in physics from Zhejiang University, Hangzhou, China, in 2014. He is currently an Associate Professor with the College of Information Science and Engineering, Huaqiao University, Fujian, China. He is also a member of the Fujian Key Laboratory of Light Propagation and Transformation. His current research interests include propagation and modulation of laser beams.

JIXIONG PU (Member, IEEE) received the Ph.D. degree from the University of Tsukuba, in 2002. From 1998 to 1999, he worked as a Visiting Scholar with the University of Tsukuba, Japan. From 2001 to 2002, he was a JSPS Fellow with the University of Tsukuba. He is currently a Professor of optics engineering with Huaqiao University, Xiamen, China. He is the author of over 200 articles. His research interests include modulating of laser beams, and the propagation and focusing of modulated laser beams, and nonlinear optics.