Recurrent neural network reveals transparent objects through scattering media

IKSUNG KANG,[1,*] ⓘ SUBEEN PANG,[2] QIHANG ZHANG,[1] NICHOLAS FANG,[2] AND GEORGE BARBASTATHIS[2,3]

[1]*Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
[2]*Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
[3]*Singapore-MIT Alliance for Research and Technology (SMART) Centre, 1 Create Way, Singapore 138602, Singapore*
*iskang@mit.edu

Abstract: Scattering generally worsens the condition of inverse problems, with the severity depending on the statistics of the refractive index gradient and contrast. Removing scattering artifacts from images has attracted much work in the literature, including recently the use of static neural networks. S. Li et al. [Optica **5**(7), 803 (2018)] trained a convolutional neural network to reveal amplitude objects hidden by a specific diffuser; whereas Y. Li et al. [Optica **5**(10), 1181 (2018)] were able to deal with arbitrary diffusers, as long as certain statistical criteria were met. Here, we propose a novel dynamical machine learning approach for the case of imaging phase objects through arbitrary diffusers. The motivation is to strengthen the correlation among the patterns during the training and to reveal phase objects through scattering media. We utilize the on-axis rotation of a diffuser to impart dynamics and utilize multiple speckle measurements from different angles to form a sequence of images for training. Recurrent neural networks (RNN) embedded with the dynamics filter out useful information and discard the redundancies, thus quantitative phase information in presence of strong scattering. In other words, the RNN effectively averages out the effect of the dynamic random scattering media and learns more about the static pattern. The dynamical approach reveals transparent images behind the scattering media out of speckle correlation among adjacent measurements in a sequence. This method is also applicable to other imaging applications that involve any other spatiotemporal dynamics.
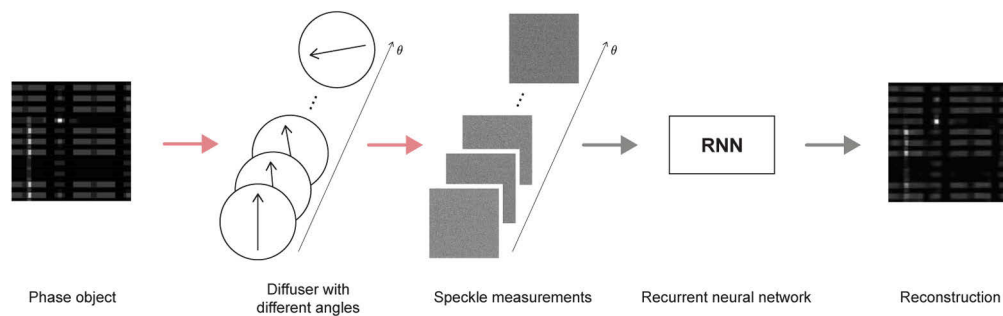
1. Introduction

Imaging through scattering media is considered challenging because the speckle-like raw images are strongly ill-conditioned. Moreover, the scattering process is generally a stochastic and nonlinear operation, meaning that the forward operator is not readily available for use in the inverse algorithm. One standard approach is to characterize the random medium through the transmission matrix (TM) [1–3]. Alternatively, other approaches involve angular scanning of the illumination through the medium and then utilize the memory effect in the speckle correlations [4–9].

Recently, machine learning algorithms have been used for image regression in a variety of inverse problems [10–14] including imaging through scatter. The first effort, to our knowledge, used a support vector machine (SVM) in 2016 [15], but it was subject to strong hallucinations when tested outside its typical training domain. Subsequently, static, convolutional neural networks (CNN) have been used to retrieve *amplitude* objects behind the random medium either using a single [16] or multiple diffusers since 2018 [17]. This approach has attracted further attention recently in even more challenging conditions, such as low photon limit [18], dynamical emulsion scattered media [19] and CNN-SVM cascade classification case [20].

Considerable amount of effort has also been devoted to image *recognition* through scattering media, also starting *circa* 2015 [21,22]. Subsequent works have addressed, for example, motion detection of figurines hidden by a sheet of paper [23], imaging through multi-mode fibers [24], and exploiting diffusers as a form of spread-spectrum [25].

In this paper, we limit our scope to image regression, as opposed to recognition, in the case where the diffuser is placed between the object and the camera, obscuring rather than aiding the imaging process; in other words, we aim to reveal the object hidden behind the scatter. For the first time, to our knowledge, we propose a recurrent neural network (RNN) as a novel dynamical machine learning approach for this problem. The RNN has exhibited good performance on exploiting correlations in spatial and temporal sequences for dynamics applications, e.g. video frame prediction [26–30], shape inpainting [31–33], depth map prediction [34,35], or multi-dimensional segmentation [36–38]. Most of these works make use of the spatiotemporal dynamics, and thus the RNN acts along the temporal and/or spatial axis. The present paper is based on the previous studies on dynamical sequences with the RNN [39–43].

We impart the dynamics to the network with a diffuser that rotates on-axis for several different angles, and thus corresponding speckle measurements sequentially form an *angular sequence* as shown in Fig. 1. Unlike the static neural network in [17] where object-measurement pairs with different realizations of the diffuser are randomly batched during the training process, with the RNN the multiple measurements of the same object remain in the same batch and contribute to training collectively in a sequential order. This way speckle correlations among the multiple measurements can be more strongly learned as priors.

Fig. 1. Collimated beam illuminates a phase object, and the diffracted optical field is strongly scattered by a diffuser, which is rotated on-axis for several different angles. For each angle of the rotation, speckle measurements are recorded by a camera, and they are processed by our proposed recurrent neural network architecture, hence the reconstruction. Red bold arrows indicate the physical propagation, and black bold ones denote the computational pipeline.
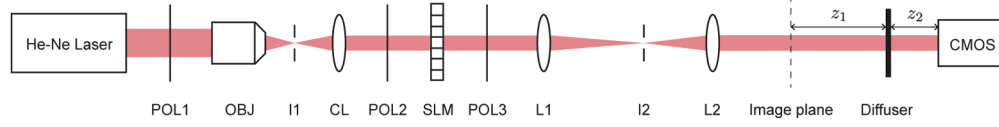
In the following sections, we first provide in-depth details on our optical apparatus for experimental data acquisition and the RNN architecture in Section 2. Next, in Section 3, we share the results on several generalization tasks: (1) seen angles of the rotation with unseen objects (3.1), (2) unseen angles with unseen objects (3.2), (3) cross-domain generalization along starkly different priors (3.3), and (4) the effect of randomizing measurements in a sequence (3.4). Concluding thoughts are in Section 4.

2. Methods

2.1. Experiment

Transparent objects, in this paper, are realized with a transmissive spatial light modulator (SLM; Holoeye LC2012, 36 μm pixel pitch, 1024 $\times$ 768) in phase modulation mode – although it is not

perfectly pure-phase, as there is a coupled amplitude modulation approximately within a range between [0.95, 1.15]. Angles of two linear polarizers in Fig. 2 are carefully chosen to maximize the phase depth of the SLM up to 4.5 rad and to minimize the spurious amplitude modulation.

Fig. 2. Proposed optical apparatus with the visible-wavelength laser ($\lambda = 633$ nm). POL: linear polarizer, OBJ: objective lens, I: iris, CL: collimating lens, SLM: transmissive spatial light modulator, L: lens. $z_1 = 70$ mm, $z_2 = 30$ mm.

A diffuser (Thorlabs DG10-600) is mounted on a motorized precision rotation stage (Thorlabs PRM1Z8) for the on-axis rotation of the diffuser. It is rotated for 20 degrees with a 1°-degree increment, and thus the number of measurements in each sequence is 20. The diffuser is placed at $z_1 = 70$ mm downstream from the image plane, where the image of a phase object on the SLM is delivered with a 3 : 2 telescope ($f_{L_1} = 150$ mm, $f_{L_2} = 100$ mm).

Speckle measurements are recorded with a CMOS camera (Basler A504k, 12 μm pixel pitch, 1280×1024), placed after another $z_2 = 30$ mm defocus from the diffuser. Integration time per frame is set to be 800 μs and the background is subtracted from each measurement.

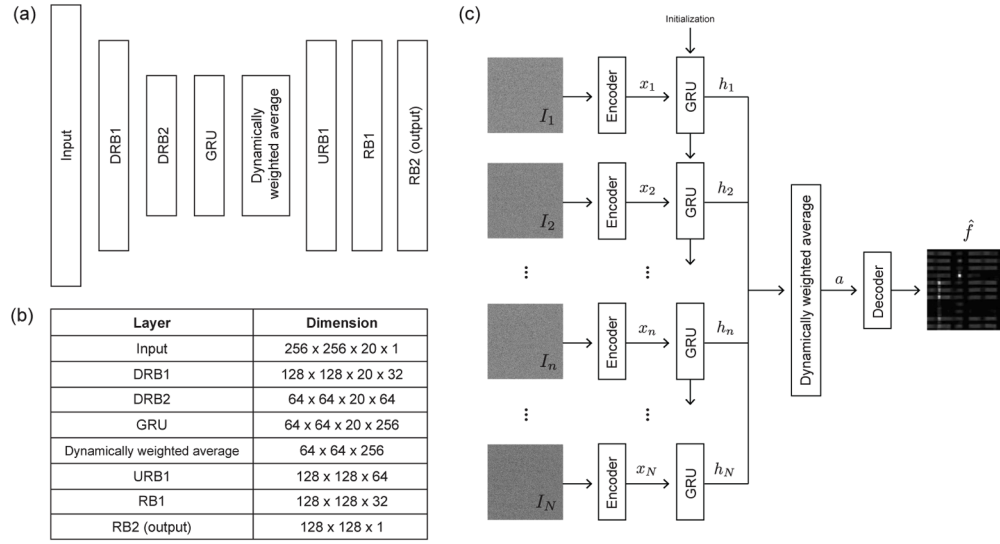## 2.2. Computational architecture

Speckle measurements of the same phase object form an *angular sequence* along the angular axis of the rotation of the diffuser. Each sequence is first encoded by two down-residual blocks (DRB) as shown in Fig. 3(a). The DRB itself contains several convolutional layers with residual pathways, adopted from [39,44]. The reason is that residual learning is known to improve generalization in deep networks [45]. This encoding process before a recurrent block extracts useful features out of raw images and reduces the number of trainable parameters thereby facilitating the training process [46–48].

As recurrent block we chose the Gated Recurrent Unit (GRU) [49], because it has fewer parameters than the older and wider-used alternative Long-Short Term Memory (LSTM) [50] without compromising performance. The GRU consists of two gates, *i.e.* reset and update gates, with fully connected layers W_r, U_r, W_z, U_z, W, and U whereas LSTM has more complex computational path. We apply one modification on the original design GRU that the native tanh activation function is now substituted with ReLU [39,51–53]. The governing equations of the modified GRU are

$$
\begin{aligned}
r_n &= W_r * x_n + U_r * h_{n-1} + b_r \\
z_n &= W_z * x_n + U_z * h_{n-1} + b_z \\
\tilde{h}_n &= \text{ReLU}\left(W * x_n + U * (r_n \circ h_{n-1}) + b_h\right) \\
h_n &= (1 - z_n) \circ \tilde{h}_n + z_n \circ h_{n-1},
\end{aligned}
\tag{1}
$$

where r_n, z_n are tensors of reset and update gates, respectively, and b_r, b_z are biases for each gate ($*$ is the matrix-vector multiplication, and $\circ$ is the Hadamard product). Throughout several recurrences, hidden features (h_n) obtain useful information for reconstructions while the redundancies are selectively discarded.

Although the n-th hidden feature h_n is a nonlinear function of $(n - 1)$ previous hidden features, the previous history of inputs is weighted in favor of the most recent measurements, *i.e.* closer to the n-th in the angular sequence [54]. To aggregate information from the hidden features, we additionally adopt the dynamically weighted average on h_n's, whose weights on input features

Fig. 3. (a) Layered diagram of our proposed recurrent neural network architecture, (b) tensorial dimensions of each layer in a tabular view, and (c) an unrolled view of the network with notations. DRB, URB, and RB stand for down-residual block, up-residual block, and residual block, adopted from [44], whose details can be found in Fig. 8 in Appendix A. In this paper, $n = 1, 2, \ldots, N (= 20)$ as 20 different angles of the on-axis rotation of the diffuser are considered.

are dynamically determined according to their scores. It follows the convention of the additive attention mechanism [54] as below.

$$e_n = \tanh\left(W_e h_n\right),$$

$$\alpha_n = \text{softmax}\left(e_n\right) = \frac{\exp(e_n)}{\sum_{n=1}^{N} \exp(e_n)}, \tag{2}$$

$$n = 1, 2, \ldots, N \,(=20),$$

where e_n and α_n are the score and the normalized weight of n-th hidden feature, and thus the output as a weighted average becomes

$$a = \sum_{n=1}^{N} \alpha_n h_n. \tag{3}$$

However, in practice, the learned weights α_n end up being approximately equal, indicating that there is no preferred diffuser orientations—as it should be. This means that we could have replaced this dynamic weighting mechanism with a simple average; yet we chose to maintain the approach of learning the α_n's for generality and as a sanity check.

Finally, the decoder is composed of an up-residual block (URB) and two residual blocks (RB) that also contain several convolutional layers with residual pathways [44], and receives a to restore its dimension to that of phase objects, thus the reconstruction $\hat{f}$ of a phase object f.

2.3. Training and testing procedures

For training, 2400 phase objects are used with the validation split ratio of 1/6. Each training sequence consists of 20 speckle measurements from 20 different angles of the on-axis rotation of the diffuser ($\theta = 0°, 1°, \ldots, 19°$). We make three separate groups of training datasets using three

starkly different priors, *i.e.* MNIST, IC layout, and ImageNet, and the network is trained with each of the training datasets. Negative Pearson correlation coefficient (NPCC) is used as training loss function [16], and it is defined as

$$\mathcal{L}_{\text{NPCC}}\left(f,\hat{f}\right) \equiv -\frac{\sum_{x,y}\left(f(x,y) - \left\langle f\right\rangle\right)\left(\hat{f}(x,y) - \left\langle \hat{f}\right\rangle\right)}{\sqrt{\sum_{x,y}\left(f(x,y) - \left\langle f\right\rangle\right)^2}\sqrt{\sum_{x,y}\left(\hat{f}(x,y) - \left\langle \hat{f}\right\rangle\right)^2}}. \tag{4}$$

We use the *Adam* optimizer [55] with initial learning rate of 10^{-3}. The learning rate is halved every time the validation loss plateaus for 5 epochs. The minimum learning rate is set to be 10^{-8}.

Testing procedures vary according to each generalization task in Section 3. For the first task, as we aim to see if the network can generalize for unseen phase objects with its trained weights, we test the trained network with 100 angular sequences from 100 non-overlapping phase objects with the same angles of rotation. We assess the results for each prior separately.

For the second task, we now test the trained network with 100 sequences from 100 unseen phase objects and from partly or wholly different angles of the rotation, where we control the percentage of measurements from unseen angles in each test sequence by 0, 25, 50, 75 and 100%. Letting m be an integer, the sequences with $m\%$ of measurements from unseen angles consist of measurements from the angles of $\left\lfloor\frac{m}{5}\right\rfloor^{\circ}, \left(1 + \left\lfloor\frac{m}{5}\right\rfloor\right)^{\circ}, \ldots, \left(19 + \left\lfloor\frac{m}{5}\right\rfloor\right)^{\circ}$.

Next, on top of the second task, we further want to see if the network trained with one prior can generalize with phase objects from different priors. For this task, we test three trained networks with 100 angular sequences from 100 unseen phase objects of MNIST, IC layout, and ImageNet [56] with measurements from 20 unseen angles ($m = 100\%$ or $\theta = 20°, 21°, \ldots, 39°$) of diffuser rotation. Thus, we get 9 cross-domain generalization cases.

Lastly, the final task is an extension of the first task in the sense that the network is trained with sequences of measurements from seen angles in an increasing order, but we test it with sequences of randomized order of measurements. Therefore, we use 100 angular sequences of 100 unseen phase objects with the same angles, *i.e.* $m = 0\%$ or $\theta = 0°, 1°, \ldots, 19°$.

The computer used for both training and testing processes has Intel Xeon Gold 6248 CPU at 2.50 GHz with 27.5 MB cache, 384 GB RAM, and dual NVIDIA Volta V100 GPUs with 32 GB VRAM.
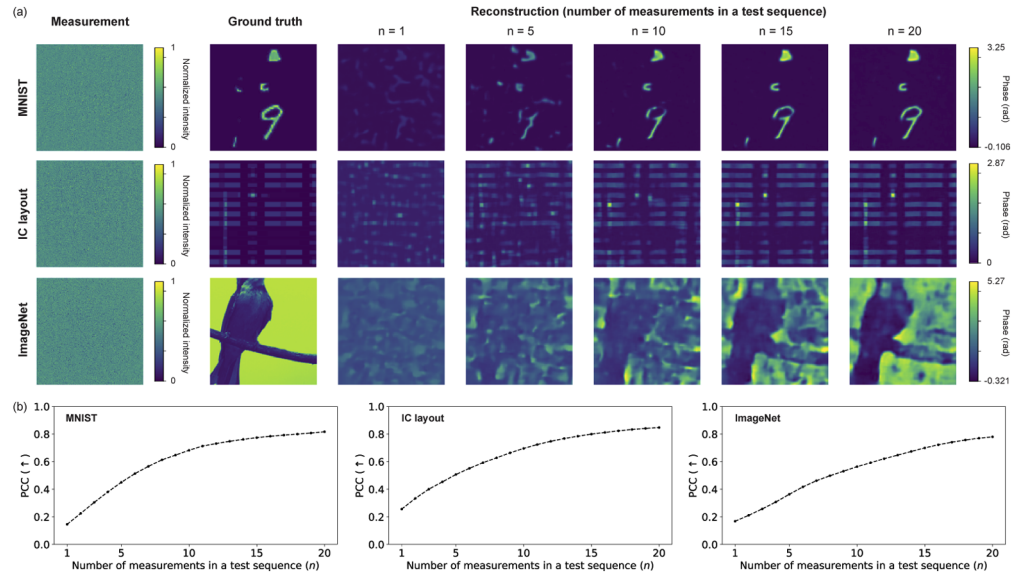
3. Results

3.1. Seen angles with unseen phase objects

The trained recurrent networks are tested with sequences of measurements of unseen phase objects from the same prior. Figure 4(a) shows qualitatively the progression of reconstructions for each prior as the number of measurements in a test sequence increases, showing the visual quality becoming similarly incrementally enhanced. This finding is quantitatively supported by Fig. 4(b) using PCC as metric. Note that the network trained with IC layout database generalizes the best, followed by MNIST and ImageNet.

3.2. Unseen angles with unseen phase objects

The previous generalization task involves test sequences of measurements from the same angles of the rotation as the training sequences (or seen angles). Here, test sequences consist of measurements from the different angles (or unseen angles). As previously mentioned in Section 2.3, the ratio of measurements from unseen angles to seen angles vary from 0% to 100%. According to Fig. 5(a), the visual quality of reconstructions degrades as more measurements from unseen angles are taken into account in test sequences, which is also quantitatively shown

Fig. 4. Seen angles of the rotation with unseen phase objects. (a) Qualitatively shown are the progressions of reconstructions according to the number of measurements in a test sequence (n) for three different priors, *i.e.* MNIST, IC layout, and ImageNet. n ranges from 1 to 20 in our case. (b) Progressions in a quantitative view using Pearson correlation coefficient (PCC) as a metric. Plotted are the means of PCC of 100 objects. (See Visualization 1 for these progressions in a video format.)

in Fig. 5(b). Still, the network is capable of retrieving prominent features even when the measurements are replaced in its entirety as long as the priors are restrictive enough, *i.e.* MNIST and IC layout.
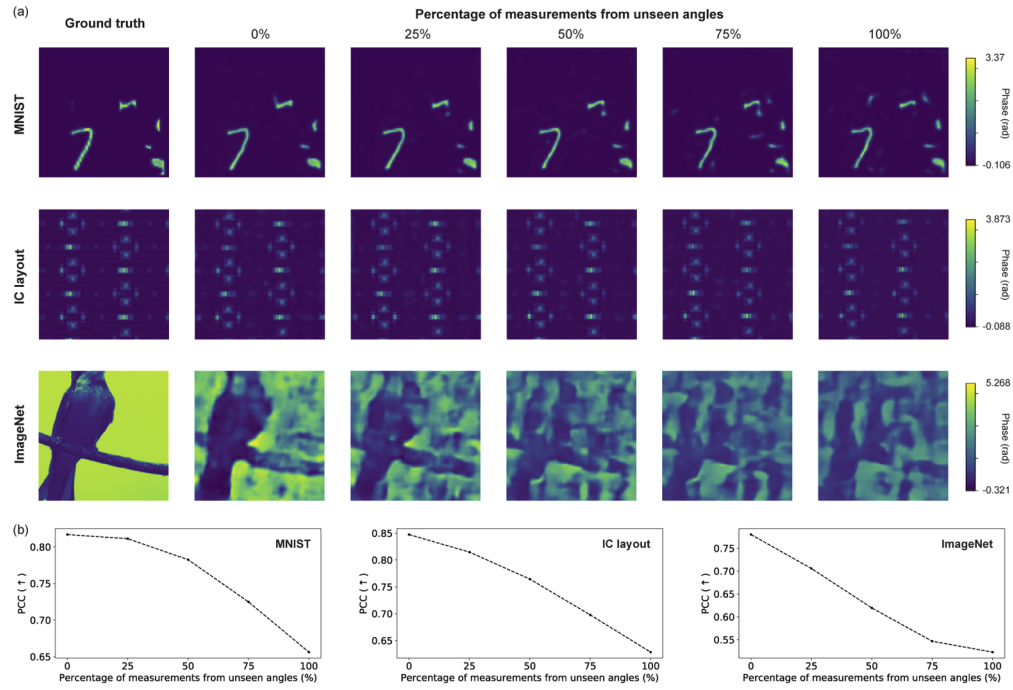
3.3. Cross-domain generalization

In this section we investigate the most restrictive generalization task, where the trained networks are now tested with sequences of phase objects sampled from different prior. This is referred to as cross-domain generalization. Here, measurements are totally from unseen angles ($m = 100\%$ or $\theta = 20°, 21°, \ldots, 39°$). The network trained with ImageNet database offers some level of cross-domain generalizability to other databases as visually in Fig. 6(a). It is possible as the deep neural networks trained with ImageNet are better generalizable [57]. As both IC layout and MNIST databases are more restrictive than ImageNet, images from the network trained with IC layout strongly resemble ("hallucinate") the shape of IC designs, and those from the MNIST-trained network are seen to be very sparse.
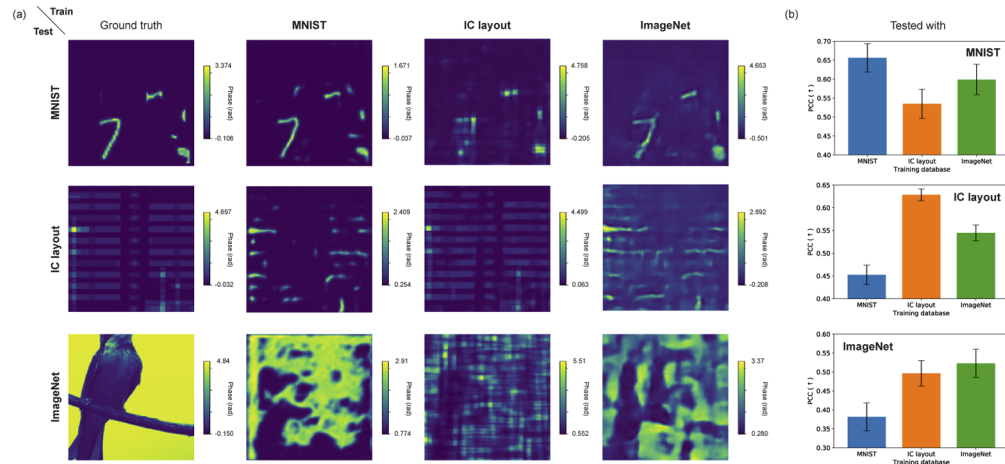
3.4. Effect of randomizing measurements in a sequence

The last section is an extension of Section 3.1. The only difference is how measurements are ordered in test sequences. When training the network, measurements are aligned in an increasing order of rotation angles; whereas for testing, we randomize the order of angles to assess the effect of the randomization. Interestingly, in Fig. 7, the severity of the effect varies with the type of database; namely, the degradation becomes more severe for less restrictive databases.
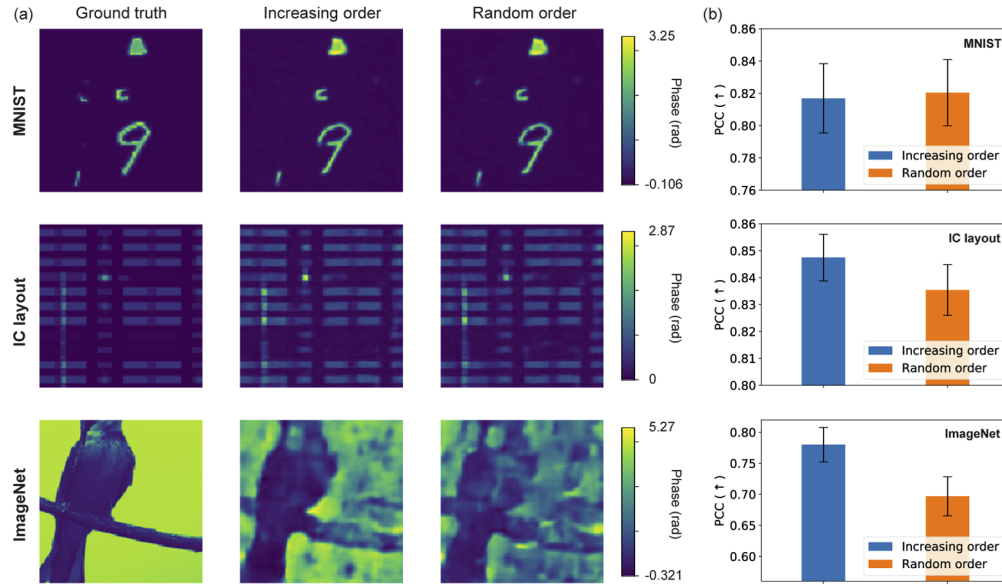
Fig. 5. Unseen angles of the rotation with unseen phase objects. (a) 0, 25, 50, 75, and 100% of measurements in a test sequence are replaced with ones from unseen angles of the rotation. Although the quality of the reconstructions degrade as the percentage of measurements from unseen angles increases, progressions of two examples show fair generalizability even the measurements are replaced with the unseen ones in its entirety. (b) The result is quantitatively shown with PCC as a metric. Plotted is the mean of PCC of 100 objects.

Fig. 6. (a) Cross-domain generalization is qualitatively shown along three different priors. The network trained with ImageNet database generalizes better than others, and (b) this result is quantitatively shown by bar graphs with the means and 95% confidence intervals of PCC of 100 objects.

Fig. 7. The network is trained with image sequences composed of measurements in an increasing order, *i.e.* $n = 1, 2, \ldots, 20$, and tested with the sequences of the measurements in either the same or a randomized order. (a) Visually, the network trained with MNIST is the least affected by the order of the measurements, followed by IC layout and ImageNet, and (b) it is quantitatively shown with bar graphs with the means and 95% confidence intervals of PCC of 100 objects.
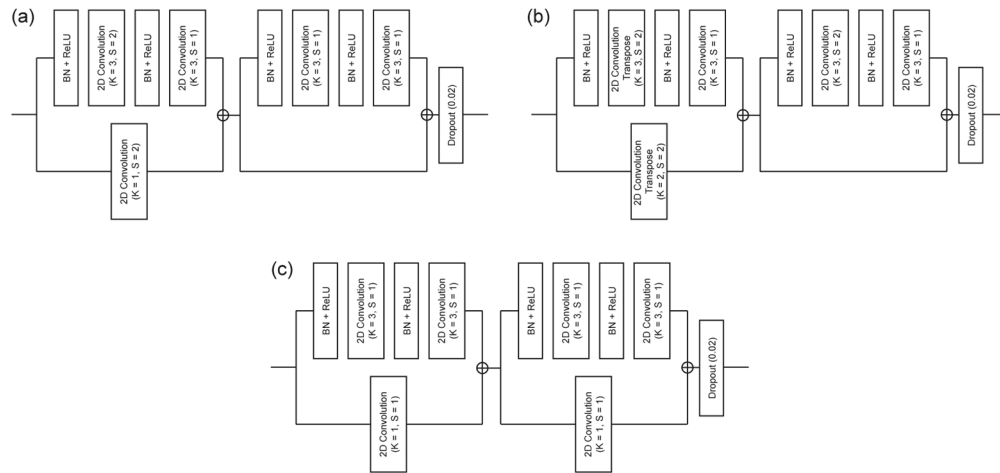
4.   Conclusion

The recurrent neural network that we constructed and trained is capable of retrieving transparent or pure-phase objects behind the random scattering media. Speckle measurements from different angles of the on-axis rotation of the diffuser form a sequence, so speckle correlation among adjacent measurements can be strongly learned by the training process. The RNN effectively inverts the speckle patterns due to the dynamic scattering media and with help from the learned prior; it reveals the correct static patterns, *i.e.* the test pure-phase objects. The trained recurrent neural network is generalizable to unseen phase objects, and unseen angles of diffuser rotation. When training priors are restrictive, the approach is generalizable even across different domains, to some extent. We expect that this approach will offer insights to other imaging applications that involve spatiotemporal dynamics combined with scattering.

Appendix A. Additional details of the architecture

This section provides additional details of the RNN architecture in Fig. 3. DRB, URB and RB consist of several convolutional layers with batch normalization, activation and dropout layers. These blocks are adopted from [44].

Fig. 8. (a) Down-residual block (DRB), (b) Up-residual block (URB) and (c) Residual block (RB) in Fig. 3. K and S denote the size of kernel and strides in convolutional layers.

References

1. I. M. Vellekoop and A. Mosk, "Focusing coherent light through opaque strongly scattering media," Opt. Lett. **32**(16), 2309–2311 (2007).
2. S. Popoff, G. Lerosey, R. Carminati, M. Fink, A. Boccara, and S. Gigan, "Measuring the transmission matrix in optics: an approach to the study and control of light propagation in disordered media," Phys. Rev. Lett. **104**(10), 100601 (2010).
3. M. Kim, W. Choi, Y. Choi, C. Yoon, and W. Choi, "Transmission matrix of a scattering medium and its applications in biophotonics," Opt. Express **23**(10), 12648–12668 (2015).
4. N. Antipa, S. Necula, R. Ng, and L. Waller, "Single-shot diffuser-encoded light field imaging," in *2016 IEEE International Conference on Computational Photography (ICCP)*, (IEEE, 2016), pp. 1–11.
5. N. Stasio, C. Moser, and D. Psaltis, "Calibration-free imaging through a multicore fiber using speckle scanning microscopy," Opt. Lett. **41**(13), 3078–3081 (2016).
6. O. Katz, P. Heidmann, M. Fink, and S. Gigan, "Non-invasive single-shot imaging through scattering layers and around corners via speckle correlations," Nat. Photonics **8**(10), 784–790 (2014).
7. J. Bertolotti, E. G. Van Putten, C. Blum, A. Lagendijk, W. L. Vos, and A. P. Mosk, "Non-invasive imaging through opaque scattering layers," Nature **491**(7423), 232–234 (2012).
8. G. Osnabrugge, R. Horstmeyer, I. N. Papadopoulos, B. Judkewitz, and I. M. Vellekoop, "Generalized optical memory effect," Optica **4**(8), 886–892 (2017).
9. A. Porat, E. R. Andresen, H. Rigneault, D. Oron, S. Gigan, and O. Katz, "Widefield lensless imaging through a fiber bundle via speckle correlations," Opt. Express **24**(15), 16835–16855 (2016).
10. K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *Proceedings of the 27th International Conference on Machine Learning (ICML)*, (2010), pp. 399–406.

11. M. Mardani, H. Monajemi, V. Papyan, S. Vasanawala, D. Donoho, and J. Pauly, "Recurrent generative adversarial networks for proximal learning and automated compressive image recovery," arXiv preprint arXiv:1711.10046 (2017).

12. G. Barbastathis, A. Ozcan, and G. Situ, "On the use of deep learning for computational imaging," Optica **6**(8), 921–943 (2019).

13. A. Song, F. J. Flores, and D. Ba, "Convolutional dictionary learning with grid refinement," IEEE Trans. Signal Process. **68**, 2558–2573 (2020).

14. K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," IEEE Trans. Image Process. **26**(9), 4509–4522 (2017).

15. R. Horisaki, R. Takagi, and J. Tanida, "Learning-based imaging through scattering media," Opt. Express **24**(13), 13738–13743 (2016).

16. S. Li, M. Deng, J. Lee, A. Sinha, and G. Barbastathis, "Imaging through glass diffusers using densely connected convolutional networks," Optica **5**(7), 803–813 (2018).

17. Y. Li, Y. Xue, and L. Tian, "Deep speckle correlation: a deep learning approach toward scalable imaging through scattering media," Optica **5**(10), 1181–1190 (2018).

18. L. Sun, J. Shi, X. Wu, Y. Sun, and G. Zeng, "Photon-limited imaging through scattering medium based on deep learning," Opt. Express **27**(23), 33120–33134 (2019).

19. Y. Sun, J. Shi, L. Sun, J. Fan, and G. Zeng, "Image reconstruction through dynamic scattering media based on deep learning," Opt. Express **27**(11), 16032–16046 (2019).

20. P. Wang and J. Di, "Deep learning-based object classification through multimode fiber via a cnn-architecture specklenet," Appl. Opt. **57**(28), 8258–8263 (2018).

21. T. Ando, R. Horisaki, and J. Tanida, "Speckle-learning-based object recognition through scattering media," Opt. Express **23**(26), 33902–33910 (2015).

22. H. Chen, Y. Gao, and X. Liu, "Speckle reconstruction method based on machine learning," in *Biomedical Imaging and Sensing Conference*, vol. 10711 (International Society for Optics and Photonics, 2018), p. 107111U.

23. G. Satat, M. Tancik, O. Gupta, B. Heshmat, and R. Raskar, "Object classification through scattering media with deep learning on time resolved measurement," Opt. Express **25**(15), 17466–17479 (2017).

24. N. Borhani, E. Kakkava, C. Moser, and D. Psaltis, "Learning to see through multimode fibers," Optica **5**(8), 960–966 (2018).

25. N. Antipa, G. Kuo, R. Heckel, B. Mildenhall, E. Bostan, R. Ng, and L. Waller, "Diffusercam: lensless single-exposure 3d imaging," Optica **5**(1), 1–9 (2018).

26. N. Srivastava, E. Mansimov, and R. Salakhudinov, "Unsupervised learning of video representations using lstms," in *International conference on machine learning*, (2015), pp. 843–852.

27. X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting, in," *Adv. Neural Inf. Process. Syst. (NIPS)*, (2015), pp. 802–810.

28. Y. Wang, L. Jiang, M.-H. Yang, L.-J. Li, M. Long, and L. Fei-Fei, "Eidetic 3D LSTM: A model for video prediction and beyond," in *International Conference on Learning Representations (ICLR)*, (2018).

29. Y. Wang, M. Long, J. Wang, Z. Gao, and S. Y. Philip, "PredRNN: Recurrent neural networks for predictive learning using spatiotemporal LSTMs," in *Adv. Neural Inf. Process. Syst. (NIPS)*, (2017), pp. 879–888.

30. Y. Wang, Z. Gao, M. Long, J. Wang, and P. S. Yu, "PredRNN++: Towards a resolution of the deep-in-time dilemma in spatiotemporal predictive learning," arXiv preprint arXiv:1804.06300 (2018).

31. W. Wang, Q. Huang, S. You, C. Yang, and U. Neumann, "Shape inpainting using 3D generative adversarial network and recurrent convolutional networks," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, (2017), pp. 2298–2306.

32. D. Kim, S. Woo, J.-Y. Lee, and I. S. Kweon, "Deep video inpainting," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, (2019), pp. 5792–5801.

33. C. Wang, H. Huang, X. Han, and J. Wang, "Video inpainting by jointly learning temporal structure and spatial details," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33 (2019), pp. 5232–5239.

34. A. C. S. Kumar, S. M. Bhandarkar, and M. Prasad, "Depthnet: A recurrent neural network architecture for monocular depth prediction," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR) Workshops*, (2018), pp. 283–291.

35. R. Wang, S. M. Pizer, and J.-M. Frahm, "Recurrent neural network for (un-) supervised learning of monocular video visual odometry and depth," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2019), pp. 5555–5564.

36. T. Le, G. Bui, and Y. Duan, "A multi-view recurrent neural network for 3D mesh segmentation," Comput. Graph. **66**, 103–112 (2017).

37. M. F. Stollenga, W. Byeon, M. Liwicki, and J. Schmidhuber, "Parallel multi-dimensional LSTM, with application to fast biomedical volumetric image segmentation," in *Adv. Neural Inf. Process. Syst. (NIPS)*, (2015), pp. 2998–3006.

38. Q. Huang, W. Wang, and U. Neumann, "Recurrent slice networks for 3D segmentation of point clouds," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2018), pp. 2626–2635.

39. I. Kang, A. Goy, and G. Barbastathis, "Limited-angle tomographic reconstruction of dense layered objects by dynamical machine learning," arXiv preprint arXiv:2007.10734 (2020).

40. Y. Yao, Z. Luo, S. Li, T. Shen, T. Fang, and L. Quan, "Recurrent mvsnet for high-resolution multi-view stereo depth inference," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2019), pp. 5525–5534.
41. J. Liu and S. Ji, "A novel recurrent encoder-decoder structure for large-scale multi-view stereo reconstruction from an open aerial dataset," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, (2020), pp. 6050–6059.
42. C. B. Choy, D. Xu, J. Gwak, K. Chen, and S. Savarese, "3D-R2N2: A unified approach for single and multi-view 3D object reconstruction," in *European Conference on Computer Vision (ECCV)*, (Springer, 2016), pp. 628–644.
43. T. Andersen, M. Owner-Petersen, and A. Enmark, "Image-based wavefront sensing for astronomy using neural networks," J. Astron. Telesc. Instrum. Syst. **6**(3), 034002 (2020).
44. A. Sinha, J. Lee, S. Li, and G. Barbastathis, "Lensless computational imaging through deep learning," Optica **4**(9), 1117–1125 (2017).
45. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, (2016), pp. 770–778.
46. J. Gehring, M. Auli, D. Grangier, and Y. N. Dauphin, "A convolutional encoder model for neural machine translation," arXiv preprint arXiv:1611.02344 (2016).
47. T. Hori, S. Watanabe, Y. Zhang, and W. Chan, "Advances in joint CTC-attention based end-to-end speech recognition with a deep CNN encoder and RNN-LM," arXiv preprint arXiv:1706.02737 (2017).
48. R. Zhao, R. Yan, J. Wang, and K. Mao, "Learning to monitor machine health with convolutional bi-directional LSTM networks," Sensors **17**(2), 273 (2017).
49. K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," arXiv preprint arXiv:1406.1078 (2014).
50. S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Comput. **9**(8), 1735–1780 (1997).
51. R. Dey and F. M. Salemt, "Gate-variants of gated recurrent unit (GRU) neural networks," in *2017 IEEE 60th international midwest symposium on circuits and systems (MWSCAS)*, (IEEE, 2017), pp. 1597–1600.
52. V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th International Conference on Machine Learning (ICML)*, (2010).
53. X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proceedings of the fourteenth International Conference on Artificial Intelligence and Statistics*, (2011), pp. 315–323.
54. D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," arXiv preprint arXiv:1409.0473 (2014).
55. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980 (2014).
56. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, (2009), pp. 248–255.
57. M. Deng, S. Li, Z. Zhang, I. Kang, N. X. Fang, and G. Barbastathis, "On the interplay between physical and content priors in deep learning for computational imaging," Opt. Express **28**(16), 24152–24170 (2020).