Title:
Displacement-agnostic coherent imaging through scatter with an interpretable deep neural network

Authors:
Yuzhe Li, Shiyi Cheng, Yujia Xue, Lei Tian*

Affiliation:
Department of Electrical and Computer Engineering, Boston University, Boston, MA 02215, USA
* Corresponding Author: leitian@bu.edu
Abstract:

Coherent imaging through scatter is a challenging topic in computational imaging. Both model-based and data-driven approaches have been explored to solve the inverse scattering problem. In our previous work, we have shown that a deep learning approach for coherent imaging through scatter can make high-quality predictions through unseen diffusers. Here, we propose a new deep neural network (DNN) model that is agnostic to a broader class of perturbations including scatter change, displacements, and system defocus up to 10X depth of field. In addition, we develop a new analysis framework for interpreting the mechanism of our DNN model and visualizing its generalizability based on an unsupervised dimension reduction technique. We show that the DNN can unmix the diffuser/displacement information and distill the object-specific information to achieve generalization under different scattering conditions. Our work paves the way to a highly scalable deep learning approach to different scattering conditions and a new framework for network interpretation.
**Introduction:**

Imaging through scatter remains one of the most challenging topics in computational imaging. The difficulty stems from the scattering process scrambling the object’s spatial frequency information and generating a complex system matrix. As a result, computational retrieval of the object requires solving an ill-posed inverse problem based on the speckle measurements and a careful characterization of the random media. Despite these challenges, many useful computational imaging techniques have been demonstrated for various applications, such as wavefront shaping, deep tissue imaging, and dynamic biological imaging.

In general, the *coherent* scattering process can be characterized by a linear transmission matrix (TM). This coherent TM establishes a one-to-one relation between the input and the output wavefronts. However, since the scattering is in general linearly shift-variant (LSV), the complete characterization of the TM is often time-consuming due to the large size of the matrix. Computational imaging techniques based on inverting the TM are susceptible to calibration errors, which may come from medium change and other perturbations to the system. One useful simplification utilizing the memory effect approximates the system to be linearly shift-invariant (LSI). Under this approximation, an invariant speckle intensity pattern only translates when rotating the incident beam over a small angle. This implies that under the *incoherent* imaging condition, the output speckle intensity is the convolution between the object’s intensity distribution and the medium’s speckle intensity point spread function (PSF). Based on this principle, 2D imaging through scatter can be achieved by a single-shot by either utilizing the pre-calibrated PSF or solving a phase retrieval problem based on the autocorrelation of the output intensity.

Recently, deep learning (DL) has proven to be a powerful technique for solving highly ill-posed computational imaging problems. In particular, deep neural network (DNN) models have been proposed to replace the standard TM that relates the output speckle patterns and the input objects, and shown superior performance over traditional methods. Most importantly, the DNN models have shown to be resilient against various perturbations and instabilities of the scattering media. For example, DNN models for *coherent* imaging through multimode fibers can make robust predictions under temperature and mechanical instabilities. In our previous work, we have shown that a DNN model for *coherent* imaging through scatter trained on a few thin diffusers can make high-quality predictions through unseen diffusers, indicating the model’s robustness against medium perturbations. Specifically, during this study, we changed the diffuser placed in-between the object and the imaging optics while keeping the diffuser’s location and the imaging system fixed. In general, many other factors can perturb the scattering medium and thus affect the imaging results.
In this work, we further consider the effect of axial displacements of the scatter itself and the imaging optics. We demonstrate a more robust DNN model for coherent imaging through scatter that is agnostic to a broader class of perturbations. Generally speaking, axial displacements of both the scatter and the imaging optics reduce the correlation of the speckle intensity measurements. Recently, the 3D memory effect has been used to expand the imaging range and achieve extended depth-of-field (DOF) in incoherent imaging through scatter. Coherent imaging under defocus is further complicated by diffraction effects. In Wu et al.’s work, 10X DOF improvement is achieved by incorporating defocus measurements in the training process for in-line holographic imaging in free space. In general, to train a robust DNN model against a broad class of perturbations, a diverse training dataset is needed to provide sufficient statistical information of the underlying process. As a result, we design our DNN training by incorporating the measurements taken from changing the scatter, axially displacing the scatter, and defocusing the imaging optics. Specifically, our training data includes speckle intensity patterns captured from four different diffusers at each training position; the training positions includes one diffuser position at 5X DOF displacement and two camera positions at ±5X DOF defocus (Fig. 1(a)). Most importantly, we show that the trained DNN can make high-quality predictions beyond the training range which is across 10X DOF through previously unseen diffuser.

To achieve robustness to displacement and better generalizability, we propose a DNN model using a hybrid network structure to better model the shift-variant property of the imaging problem. Our network is built on the encoder-decoder structure in our previous work. To improve the network’s expressivity for modeling the SV properties, we add two fully-connected layers in the bottleneck of the network, as denoted by the transformation module in Fig. 1(b). This module takes input from the encoded features and transforms the 2D information to a 1D latent code, which is then fed into the decoder to reconstruct the 2D object. The operations on the latent code in the transformation module enlarges the effective receptive field of the DNN model, which in turn accounts for the shift variance of the system.

To interpret the working principle of our DNN model and better understand its generalization capability, we further develop a new analysis framework based on an unsupervised dimensionality reduction technique, UMAP. In particular, our analysis provides several new insights into both the information contained in the raw speckle data, and the training and prediction processes. First, we show that directly decomposing the raw speckle intensity data onto a nonlinear manifold using the unsupervised technique already reveals scatter and displacement-specific information without the need for any prior information. Second, by further analyzing the latent code in our DNN model, we show that the model reveals object-
specific information and disentangles scatter/displacement information using the encoder path. The generalization of our DNN model is analyzed in two steps. First, we set up the learned latent manifold using the learned latent codes extracted by feeding only the training data to the trained network. Next, the predicted latent codes from the unseen speckle patterns under different scattering conditions are extracted and projected onto the learned manifold. We show that the predicted latent codes match well with the learned latent codes, which indicates that indeed the DNN model can generalize well to the unseen scattering cases.

In summary, we demonstrated a deep learning method for coherent imaging through scatter that is agnostic to scatter change, displacements, and system defocus. We demonstrated robust predictions across 10X DOF by a new data acquisition procedure and a novel DNN model. We further developed a new analysis framework for revealing the information contained in the speckle data, interpreting the mechanism of the network, and visualizing the generalizability of the DNN model.

Results

We develop and demonstrate a robust DNN model for coherent imaging through scatter that is agnostic to both change of scatters and axial displacements of scatter and imaging optics, as summarized in Fig. 1(a).

Experiment setup

The imaging setup is shown in Fig. 2(a). A spatial light modulator (SLM) (Holofeye NIR-011, pixel size 8 um) was coherently illuminated by a collimated beam from a HeNe laser (632 nm, Thorlabs HNL210L). We used the SLM as a programmable amplitude-only object by placing two orthogonally oriented polarizers before and after. It was relayed onto the camera (Thorlabs Quantalux, pixel size 5.04 µm) by a 4F system. Two lenses with focal lengths 200 mm (L1) and 125 mm (L2) were used to provide a 0.625 magnification. A 14 mm iris was placed at the pupil plane of the 4F system to control the speckle size. A diffuser was placed in between the SLM and L1. We placed five diffusers (Thorlabs N-BK7 Ground Glass Diffuser, 220 Grit DG10-220) on a filter wheel (Thorlabs FW1A) in order to capture data through different diffusers. Both the camera and filter wheel were attached to linear motion stages that can be moved axially. The initial positions for the camera $Z_{C10}$ and diffuser $Z_{D0}$ were set at the back focal plane of L2 and 100 mm in front of L1, respectively. By controlling the motion stages, we took speckle patterns from multiple combinations of diffuser/camera displacements and with five different diffusers. The intervals between the neighboring displacement positions for the diffuser and the camera are 1X DOF, which are set by the corresponding speckle sizes.
System characterization

The system was characterized by measuring the 3D speckle size. We first captured a speckle intensity stack by moving the camera from 2 mm before $Z_{C10}$ to 2 mm after $Z_{C10}$ with step size 0.02 mm while fixing the diffuser at its initial position $Z_{D0}$. We then measured the 3D speckle size by calculating the 3D autocorrelation of the speckle intensity stack, as shown in Fig. 2(b). The experimentally measured lateral and axial speckle sizes at the sensor plane are 10.08 µm and 0.41 mm, respectively, which match with the theory (see Methods). The axial speckle size at scatter plane is enlarged to 1.04 mm due to the system magnification.

Data acquisition

When taking the displacement measurements, we set the interval between two neighboring positions of the diffuser and the camera to be 1 mm and 0.5 mm, respectively. The total displacement range of the diffuser and the camera covers 10X DOF and 20X DOF, respectively. First, the diffuser was moved from the initial position $Z_{D0}$ towards L1 while taking data at 10 different positions ($Z_{D1}$ to $Z_{D10}$) with the camera being fixed at $Z_{C10}$, as shown in Fig. 2(c). Next, the camera was moved across 20 positions from $Z_{C0}$ to $Z_{C20}$ with the diffuser being fixed at $Z_{D0}$. Among the 20 camera positions, 10 positions ($Z_{C0}$ to $Z_{C9}$) were moved towards L2, the other 10 were moved away from L2.

We studied the statistical distribution of the measured speckle intensity data. As shown in Fig. 2(c), the estimated probability density functions (PDF) of the speckle patterns captured from different objects through different diffusers and/or at different displacement positions can all be fitted to the same speckle intensity distribution function (see Methods). This highlights that all the scattering-specific and object-specific information that we will investigate next are encoded in the higher order statistics, which is hard to extract using basic statistical fitting techniques.

Our training data consists of 4200 image pairs, each of which consisted of the input object and the measured speckle pattern. The input objects contained 400 MNIST handwritten digits, 350 of which were used as the training objects. The speckle patterns for training were taken through four different diffusers, one diffuser position, and two camera positions. Specifically, the training diffuser position was 5X DOF from its initial position at $Z_{D5}$; the two training camera positions were ±5X DOF away from its initial positions at $Z_{C5}$, $Z_{C15}$, as shown in Fig. 3(a). The testing data consisted of 50000 speckle patterns taken under two different imaging conditions. In the first testing condition, we tested our network using speckle patterns through four seen diffusers (i.e. used for taking the training data) at 9 unseen diffuser positions and 18
unseen camera positions, and using both seen and unseen objects. In the second testing condition, we tested our network using speckle patterns from one unseen diffuser (i.e. never used during network training) at all diffuser and camera positions and using both seen and unseen objects.

**Network implementation**
We built a DNN shown in Fig. 1(b) to learn a statistical model relating the speckle patterns and the unscattered objects. The overall structure of the proposed DNN follows the encoder-decoder “Unet” architecture\(^3\) with the modifications of replacing the convolutional layers with the dense blocks\(^4\) and the additional fully connected layers at the “bottleneck” to perform latent code transformation. The input to the CNN is a preprocessed 128×128 speckle intensity. The input then goes through the “encoder path”, which yields a stack of 4×4 latent code. The latent code includes case-specific information that encodes the displacement and diffuser parameter. Next, the latent code is flattened to a 1D vector, which is input to two fully connected layers and then reshaped to 2D. Together, this composes the latent code transformation module. This module enables transforming the case-specific information to meaningful object-specific features. These operations on the latent code also enlarges the effective receptive field of the DNN model, which facilitates modeling our shift-variant system. The decoder reverses the process that recombines the information into feature maps with gradually increased lateral details. The output is a binary object. Additional details of our DNN model and the benefit of the transformation module are provided in Methods and SI Figs. S1 and S2.

**Experimental Results**
We demonstrated the robustness of our network against scatter and the camera displacement on two types of experiments. All the experimental results were obtained using the same single network trained with the four diffusers at three different training positions.

**Results on axial displacements through seen diffusers.** We first tested our network using the speckle patterns from the same four trained diffusers at different unseen positions, as summarized in Fig. 3. The testing objects consisted of both seen digits used for training and unseen digits. The testing displacement positions for both the diffuser and the camera were up to 10X DOF, as shown in Fig. 3(a). The speckle patterns appear notably different when the diffuser or camera is displaced over 1X DOF in the axial direction. Our DNN demonstrated the ability to make high-quality predictions at previously unseen positions across 10X DOF. Representative examples of the speckle and prediction pairs for both seen and unseen objects are shown in Fig. 3(b). We first show the results on diffuser displacements with the camera placed at its initial position \(Z_{C10}\). On the left panel of Fig. 3(b), the speckle patterns and the prediction
results are listed as the diffuser displacement ranging from 1X, 3X, 7X and 10X DOF, respectively. Next, we present the testing results of camera displacements with the seen diffusers placed at $Z_{D0}$ on the right panel of Fig. 3(b). We show the speckle patterns and the network predictions when the camera was displaced by -9X, -7X, 2X, 10X DOF away from $Z_{C10}$.

Next, we quantify the prediction performance using the Jaccard Index (JI), as summarized in Fig. 3(c). Our experimental results show consistent predictions on both seen and unseen objects through all four diffusers. Accordingly, the statistics shown in the plot in Fig. 3(c) were calculated by accumulating all the objects (seen and unseen) and all four diffusers. Each cross marker indicates the mean JI over all the predictions at each position. Each error bar indicates the standard deviation of prediction results at each position. We observed that the network performs the best at the trained positions (on unseen objects). When the displacement increases, the JI gradually decreases. The variations of the predictions (quantified by the standard deviation) also increases with the displacement distance.

Overall, our DNN showed the ability to consistently make high-quality predictions for both camera and diffuser displacements. As compared to image classification, our network performs pixel-wise prediction that is considerably more difficult since the network needs to effectively learn a pixel-level input-output relation\textsuperscript{23}. For both seen and unseen objects belonging to the same class, our network can make high-quality pixel-level predictions. The degradation of the predictions as a function of displacement is gradual, as visualized in Fig. 3(b). This shows the robustness of our DNN model under these physical perturbations.

Results on imaging through unseen diffusers across different displacements. In the second experiment, we further tested our network using the speckle patterns obtained with the unseen diffuser and across a range of displacement positions, as shown in Fig. 4(a). We tested our network on both seen and unseen objects from the MNIST digit dataset. As summarized in Fig. 4(b), the left panel shows that the prediction results with the diffuser displaced at 1X, 3X, 7X DOF, respectively. The right panel shows the prediction results with the camera displaced at -9X, 0X, 7X DOF, respectively. By using an unseen diffuser, the problem becomes more challenging as the network needs to overcome both scatter and position perturbations. As a result, the performance degrades as compared to the seen diffuser case. Still, the main structure of the objects are accurately recovered across a range of tested displacement positions.

Analysis
Next, we investigated the correlation across different speckle patterns imaged under different imaging conditions and further developed a framework to interpret the mechanism of how our DNN model generalizes over different scattering conditions. To do so, we used the state-of-the-art unsupervised dimensionality reduction technique, UMAP\textsuperscript{32}. UMAP models the entire dataset into a single nonlinear
manifold by learning the underlying topological structure contained in the high-dimensional data. In its simplest form, UMAP considers each data (e.g. an image) as a single vector in the learned manifold and models the entire dataset as a 2D (nonlinear) representation. We apply this technique to analyze both the raw speckle patterns and the DNN model’s latent codes, and propose a procedure to visualize the training and prediction process.

**Raw speckle patterns contain scatter and displacement-specific information.** We analyzed the input data and the corresponding measured speckle patterns to discover the *intrinsic* correlations, as shown in Fig. 5(a). First, the UMAP learned manifold of the input object dataset is visualized as a 2D map. For better visualization, We randomly selected 4000 images from the same MNIST dataset as our input data. As clearly shown in Fig. 5(a)(i), this dataset naturally (i.e. without any supervision / labels for UMAP) clusters into 10 groups corresponding to the underlying 10 digit classes, each of which is marked by a different color. Here, each point (i.e. a vector) on this 2D map represents an input object image. This visualization shows that the raw input object intrinsically contains object-specific information in its image structure as expected. Next, we visualized the UMAP learned manifold from 9600 speckle patterns taken under 24 different scattering conditions, including 4 diffusers, each with 3 diffuser positions and 3 camera positions. Importantly, we observe that the learned manifold for the speckle patterns are clustered into 24 distinct groups according to the underlying scattering condition, while the object-specific information has been scrambled by the scattering, as shown in Fig. 5(a)(ii). This result shows that the speckle patterns captured under the same scattering condition contain intrinsic correlations; the speckle patterns become more decorrelated as the scattering condition changes. This observation matches well with our previous study based on the classical Pearson correlation analysis. Here, by using a more advanced dimensionality reduction technique to analyze the raw speckle patterns, we show that the raw speckle patterns contain scatter/displacement-specific information that can be revealed *without the need for any supervised learning procedure*.

**Interpreting the mechanism of the DNN model’s generalization.** Next, we develop a novel procedure to interpret the working principle of our DNN model and its generalization capability to different scattering conditions (i.e. different diffusers and/or displacement positions). Our main idea is to analyze the training and prediction processes and quantify the underlying information content by UMAP. To do so, we take the following two-step process. First, we set up the *learned latent space* by the trained DNN model, which will be used as the fixed “coordinate system” to quantify the information content in both the training and the prediction. Specifically, we fed each training data into the trained network and extracted the corresponding latent code. We then used all the latent code from the entire training dataset and set up the learned latent
space using UMAP. Second, we projected the latent code extracted from the testing data under different conditions to the learned latent space (coordinate system) and visualized the discrepancy between the learned and predicted latent codes in order to assess the DNN model’s generalization capability. In the following, we first discuss our analysis results on unseen objects under the same training scattering conditions and show that our DNN model can reveal object-specific information and disentangles scatter/displacement-specific information using the DNN model’s encoder path. Next, we discuss the testing results underlying different scattering conditions and demonstrate our DNN’s generalization capability to different scattering cases.

The DNN model reveals object-specific information. We first visualized the speckle patterns used for the training that include 12 different scattering conditions (i.e. four diffusers at one diffuser position and two camera positions) by UMAP in the 2D map in Fig. 5(b)(i), which is termed the training input manifold. As expected based on our previous analysis, 12 distinct clusters are formed matching the underlying scattering condition. Next, we fed all the training speckle patterns to our trained DNN and extracted the learned latent codes, which are then used to compute the learned latent manifold by UMAP. In Fig. 5(b)(v), the learned latent space is visualized as a 2D map. Importantly, it contains 10 clusters based on the corresponding digit label instead of the scattering conditions. This shows that the trained network learns to distill object-specific features and “unmix” the scattering effects.

Next, we projected the testing speckle patterns captured from unseen objects and under the same 12 scattering conditions onto the existing training input space under the same (nonlinear) UMAP transformation. As shown in Fig. 5(b)(ii), the projection aligns well with the existing training input manifold and the corresponding cluster, which further indicates that speckles captured with a given scattering condition are correlated regardless of the input objects. Next, we fed all the testing speckle patterns to the trained DNN and extracted the predicted latent codes. Finally, we projected the predicted latent codes onto the previously learned latent manifold, as shown in Fig. 5(b)(vi). The predicted latent code clusters align well with that from the training data. This result indicates that our DNN can reveal object-specific information under the same scattering conditions from unseen speckle patterns.

Interpreting the DNN model’s generalizability to different scattering conditions. We discuss the analysis results from different scattering conditions in two cases. In the first case, we analyzed the testing data from different displacements through the same four training diffusers, including four unseen diffuser positions and four unseen camera positions. After projecting the input speckle patterns onto the previously established training input manifold, the clusters no longer align with the existing input manifold, as shown in Fig. 5(b)(iii). In the second case, we analyzed speckle patterns captured from the unseen diffuser with 4
diffuser positions and 4 camera positions. As shown in Fig. 5(b)(iv), the projected clusters significantly differ from the training input manifold. In both cases, it shows again that speckles captured from different scattering conditions are decorrelated. Although the manifold learned by UMAP is nonlinear, Fig. 5(b)(iii) and Fig. 5(b)(iv) can be intuitively interpreted as follow: given the “coordinate system” set up by the training speckles, the testing speckles can no longer be represented by any single cluster (“axis”). This is because speckle patterns from different cases exhibit very distinct features so that their corresponding 2D representations are far apart. Specifically, combining with our previous analysis, the scattering condition dictates the unique features in the speckle patterns, and will differentiate them from other cases. This elucidates on the challenge for deep learning to generalize over different scattering conditions.

The next step is to project the predicted latent codes extracted from the testing data onto the learned latent manifold under the same UMAP transformation. As shown in Fig. 5(b)(vii) and Fig. 5(b)(viii), for both cases under different scattering conditions, the predicted latent code clusters align well with that of the training data. The good alignment between the predicted and learned latent manifolds illustrates our DNN model’s ability to generalize in terms of diffuser displacements, camera displacement, and diffuser change. Artifacts including mixing across different clusters are also observed as compared to the original learned latent manifold. These artifacts become more obvious for the unseen diffuser case (Fig. 5(b)(viii)). Fundamentally, this is because the learned encoder is trained based on the training data distribution, which may not sufficiently capture the testing data distribution.

**Conclusion and Discussion**

In this paper, we presented a new deep learning framework for coherent imaging through scatter and pushing the robustness against scatter displacement and imaging optics defocus. We developed a new analysis framework for revealing the information contained in the speckle dataset, interpreting the mechanism of our DNN, and visualizing the generalizability of the DNN model.

We demonstrated that our DNN model is agnostic to scatter changes, scatter displacement and camera defocus for coherent imaging through scatter. By improving the data acquisition strategy and improving the network structure, the generalization capability may be further improved, which will be explored in our future work. These promising results show that deep learning can robustly solve challenging inverse problems of coherent imaging through complex media under various perturbations.

Our analysis framework shows that the speckle patterns intrinsically carry scatter/displacement information. After training, the encoder can unmix the scattering and distill the object-specific information. Our analysis framework allows us to visualize the DNN’s generalizability under different imaging
conditions. This provides a powerful tool to explore the underlying correlations within the data and to interpret the learning mechanism of the DNN model. The caveat of using a nonlinear dimensionality reduction process, like UMAP, is that the traditional distance measures for linear spaces, such as the Euclidean norm, can no longer be used\textsuperscript{32,33}. This poses challenges to quantify the discrepancy between the learned and the predicted distributions, which will be investigated in our future work.

Here, to simplify the problem we have focused on establishing an understanding of the mechanism of network generalization against scatter and displacement changes while using a simple dataset containing only 10 classes of handwritten digits. This approach has the benefits of facilitating direct visualization of the learned latent space into a small number of visually distinguishable clusters using the dimensionality reduction technique. However, the simplicity of the object dataset inevitably limits the network’s generalizability for predicting more complex structures, which requires increased diversity of the training objects, as shown in our previous work\textsuperscript{25}. In general, it has been recently shown that the network’s generalizability against object variations can be improved by increasing the information capability, i.e. entropy, of the training dataset\textsuperscript{19}. In our work, the information content in the input data is visualized by the dimensionality reduction technique. Importantly, our results show that for systems involving complex transformations, such as scattering, directly measuring the information content in the raw input data may not provide sufficient information for generalization. Instead, we develop a latent code analysis framework for understanding these challenging computational imaging problems. We envision a comprehensive analysis framework developed by analyzing the information content of both the network input and the latent space may provide additional insights into improving and interpreting generalization against both system perturbations and object variations, which will be considered in our future work.

Method
Theoretical 3D speckle size. The speckle size in 3D was calculated by computing the 3D autocorrelation of a speckle intensity stack\textsuperscript{26}. The theoretical lateral speckle size in free space propagation geometry is defined by the full width at half maximum (FWHM) of the normalized autocorrelation function along the lateral direction and is $1.0\lambda \frac{z}{D}$, where $\lambda$ is the wavelength of the coherent source, $z$ is the distance between the scattering surface and observation region, $D$ is the diameter of the scattering spot. The axial speckle size is defined by the FWHM of the normalized correlation function along the axial direction and is $7.1\lambda (\frac{z}{D})^2$, which also defines the system’s DOF\textsuperscript{26}. According to this, the theoretical DOF of our system at the object side is 0.92 mm and at imaging side is 0.38mm. The theoretical lateral speckle size at the camera side is 5.56 µm. Our experimentally calculated speckle intensity autocorrelation is shown in Fig. 2(b), which
shows that the lateral speckle size is 10.08 µm. Due to under-sampling by the camera, the experimentally measured lateral speckle size is larger than the theoretical value.

**Speckle statistical distribution.** We investigated the statistical distribution of speckle intensity patterns of our dataset. For fully developed speckles, the intensity follows the negative exponential distribution. For \( N \) incoherently summed independent speckles, the PDF is a Gamma density function:

\[
P(I) = \frac{N^N I^{N-1}}{\Gamma(N)I^N} \exp \left( -N \frac{I}{I} \right)
\]

In practice, the probability density function (PDF) was estimated from the normalized intensity histogram of experimentally measured speckle patterns. In Fig. 2(c), we show several representative cases that cover all the cases we studied, including the speckle patterns from different objects, different scatters, different camera positions, and different diffuser positions. Applying the theoretical PDF in MATLAB to match the experimental measurement, \( N \) was estimated to be 1.8, which is consistent with our under-sampled imaging condition. All speckle intensity patterns approximately follow the same distribution.

**Data preprocessing.** The SLM input and the camera measurements are collected in pairs for generating the dataset. The central 512×512 SLM pixels were used as the object; the corresponding central 512×512 camera pixels were used as the speckle intensity input for our DNN. The objects displayed on the SLM were 8-bit grayscale images from the MNIST handwritten digit. Due to the computation and memory limitations, all input and output images were down-sampled from 512×512 pixels to 128×128 pixels by taking the average within each 4×4 neighboring pixels. For both training and testing, intensity outliers were removed by histogram clipping. The speckle images were then normalized between 0 and 1 by dividing each image by its maximum. Although we display grayscale images on the SLM, our DNN was designed to make binary predictions.

**Dimensionality Reduction and Data Visualization.** UMAP was used to model the data on a high-dimensional manifold with a graph structure to reduce the dimension for visualization. Once a large dataset without labels is fed into UMAP, the algorithm adaptively outputs an unsupervised transformation mapping between the high-dimensional dataset and a low-dimensional representation. Consider the entire speckle dataset as a \( N \times 128 \times 128 \) matrix, where \( N \) denotes the number of images. For perform dimensionality reduction analysis, we first preprocessed the data by reshaping the matrix into a \( N \times 128^2 \) matrix. Next, we compute the training input manifold as a 2D representation using UMAP. To analyze the testing data, we use the transform method in UMAP to directly project preprocessed 2D testing input onto the training input.
manifold. For latent code analysis, the learned manifold was first computed from the latent code by feeding the entire training dataset into the trained network and extracting the code from the bottleneck layer. Next, we used the transform method in UMAP to project the testing data’s latent code to the previously established learned manifold.

**Neural network implementation.** Fig. S1 shows the details of our neural network implementation. The input to the CNN is a preprocessed 128×128 speckle intensity. Next, the input goes through the “encoder path”, which consists of 4 dense blocks connected by a max pooling layer for down-sampling. Each dense block consists of 2 layers, in which each layer performs batch-normalization (BN), the rectified linear unit (ReLU) nonlinear activation, and convolution (conv) with 3 filters. The intermediate output from the encoder has small lateral dimensions (4×4), but encodes rich information along the “depth”. A transformation module is then concatenated to the encoder, consisting of a flatten layer that outputs a 1-D latent vector, 2 fully connected layers with ReLU activation and a constant reshaping layer that transforms back to a 2D feature map. Next, the low-resolution feature maps go through the “decoder path”, which consists of 4 additional dense blocks connected by up-sampling followed by convolutional layers. The information across different spatial scales are tunneled through the encoder-decoder paths by skip connections to preserve high-frequency information. After the decoder path, an additional convolutional layer with sigmoid followed by the last layer produces the network output. The last layer is designed to solve a pixel-wise binary prediction problem. The CNN makes decisions on if an object is present. The CNN training was performed on BU Shared Computing Cluster with one GPU (NVIDIA Tesla P100) using Keras/Tensorflow. Each CNN was trained with 200 epochs by the ADAM optimizer for up to 3 hours. The learning rate of $10^{-4}$ was used. Once the CNN was trained, each prediction was made in 0.0156 s.

**Funding**
National Science Foundation (1813848).
Fig. 1. Overview of our deep learning-based approach to achieve generalization in coherent imaging through scatter. (a) Our coherent imaging model and our data acquisition approach to obtain a diverse dataset. The dataset including scatter changes, sensor and scatter displacement over 10X DOF. Speckle patterns from training diffusers at training positions are used to train the DNN. Others are used as testing data. (b) We implemented a DNN structure including a transformation module to achieve generalizability.
Fig. 2. **Experimental setup and speckle characterization.** (a) The experiment setup for coherent imaging through scatter. The SLM is used as the amplitude-only object. Both the diffuser and the camera are placed on motion stages to control the axial displacement. (b) The 3D speckle size is characterized by calculating the speckle intensity stack’s autocorrelation. (c) The raw speckle intensity distributions $P(I/I)$ approximately follow the same probability density distribution regardless of the input object, diffuser, camera position $Z_{Ci}$, and diffuser position $Z_{Di}$. 
Fig. 3. Data acquisition and results on coherent imaging through seen diffusers at unseen displacement positions. (a) The summary of the training and testing dataset. The training data are captured through 4 different training diffusers at three training positions. All the rest are testing positions. (b) Representative testing results at different diffuser displacement positions (Left panel) and camera displacement positions (Right panel) for both seen objects (Row 1 and Row 2) and unseen objects (Row 3 and Row 4). (c) Quantitative evaluation of the DNN performance. Each cross marker represents the mean JI of the predictions calculated on all four diffusers at each displacement position. Each error bar indicates the standard deviation of the prediction results at each position. The training displacement positions are marked by the grey box.
Fig. 4. Results on coherent imaging through seen diffusers at unseen displacement positions. (a) The summary of the training and testing dataset. Training data are the same as the data in the first experiment. Testing data is from the unseen diffuser $D_s$ at different positions. (b) Testing results through unseen diffuser $D_s$ at different diffuser displacement positions (Left panel) and camera displacement positions (Right panel) for both seen objects (Row 1 and Row 2) and unseen objects (Row 3 and Row 4).
Fig. 5. Data and latent space analysis based on dimensionality reduction. (a) UMAP based visualization of the input data and speckle measurements. (i) The manifold of the input data shows 10 clusters matching the underlying 10 digits. (ii) The manifold of the speckle patterns form clusters based on the underlying scattering conditions. (b) Network analysis during training and making predictions. (i) The training input manifold computed from the speckles captured under 12 scattering conditions correspondingly forms 12 distinct clusters. (ii-iv) The testing data under different imaging conditions are projected onto the same training input manifold. (v) The learned latent manifold is computed from the leaned latent code of the training data. (vi-viii) The corresponding predicted latent code for testing data under different imaging conditions are projected onto the learned latent manifold.
References

1. Kim M, Choi W, Choi Y, Yoon C, Choi W. Transmission matrix of a scattering medium and its applications in biophotonics. *Opt Express* 2015; 23: 12648–12668.

2. Popoff S, Lerosey G, Fink M, Boccara AC, Gigan S. Image transmission through an opaque material. *Nat Commun* 2010; 1: 1–5.

3. Vellekoop IM, Mosk AP. Focusing coherent light through opaque strongly scattering media. *Opt Lett* 2007; 32: 2309–2311.

4. Ohayon S, Caravaca-Aguirre A, Piestun R, DiCarlo JJ. Minimally invasive multimode optical fiber microendoscope for deep brain fluorescence imaging. *Biomed Opt Express* 2018; 9: 1492–1509.

5. Dunn AK, Bolay H, Moskowitz MA, Boas DA. Dynamic Imaging of Cerebral Blood Flow Using Laser Speckle. *J Cereb Blood Flow Metab* 2001; 21: 195–201.

6. Popoff SM, Lerosey G, Carminati R, Fink M, Boccara AC, Gigan S. Measuring the Transmission Matrix in Optics: An Approach to the Study and Control of Light Propagation in Disordered Media. *Phys Rev Lett* 2010; 104: 100601.

7. Hillman TR, Yamauchi T, Choi W, Dasari RR, Feld MS, Park Y *et al.* Digital optical phase conjugation for delivering two-dimensional images through turbid media. *Sci Rep* 2013; 3: 1–5.

8. Jang M, Ruan H, Vellekoop IM, Judkewitz B, Chung E, Yang C. Relation between speckle decorrelation and optical phase conjugation (OPC)-based turbidity suppression through dynamic scattering media: a study on in vivo mouse skin. *Biomed Opt Express* 2015; 6: 72–85.

9. Liu Y, Lai P, Ma C, Xu X, Grabar AA, Wang LV. Optical focusing deep inside dynamic scattering media with near-infrared time-reversed ultrasonically encoded (TRUE) light. *Nat Commun* 2015; 6: 1–9.

10. Qureshi MM, Brake J, Jeon H-J, Ruan H, Liu Y, Safi AM *et al.* In vivo study of optical speckle decorrelation time across depths in the mouse brain. *Biomed Opt Express* 2017; 8: 4855–4864.

11. Feng S, Kane C, Lee PA, Stone AD. Correlations and Fluctuations of Coherent Wave Transmission through Disordered Media. *Phys Rev Lett* 1988; 61: 834–837.

12. Freund I, Rosenbluh M, Feng S. Memory Effects in Propagation of Optical Waves through Disordered Media. *Phys Rev Lett* 1988; 61: 2328–2331.

13. Freund I. Looking through walls and around corners. *Phys Stat Mech Its Appl* 1990; 168: 49–65.

14. Schott S, Bertolotti J, Léger J-F, Bourdieu L, Gigan S. Characterization of the angular memory effect of scattered light in biological tissues. *Opt Express* 2015; 23: 13505–13516.
15 Bertolotti J, van Putten EG, Blum C, Lagendijk A, Vos WL, Mosk AP. Non-invasive imaging through opaque scattering layers. *Nature* 2012; **491**: 232–234.

16 Edrei E, Scarcelli G. Memory-effect based deconvolution microscopy for super-resolution imaging through scattering media. *Sci Rep* 2016; **6**: 1–8.

17 Katz O, Heidmann P, Fink M, Gigan S. Non-invasive single-shot imaging through scattering layers and around corners via speckle correlations. *Nat Photonics* 2014; **8**: 784–790.

18 Barbabastis G, Ozcan A, Situ G. On the use of deep learning for computational imaging. *Optica* 2019; **6**: 921–943.

19 Li S, Deng M, Lee J, Sinha A, Barbabastis G. Imaging through glass diffusers using densely connected convolutional networks. *Optica* 2018; **5**: 803–813.

20 Lyu M, Wang H, Li G, Zheng S, Situ G. Learning-based lensless imaging through optically thick scattering media. *Adv Photonics* 2019; **1**: 036002.

21 Rahmani B, Loterie D, Konstantinou G, Psaltis D, Moser C. Multimode optical fiber transmission with a deep learning network. *Light Sci Appl* 2018; **7**: 1–11.

22 Turpin A, Vishniakou I, Seelig J d. Light scattering control in transmission and reflection with neural networks. *Opt Express* 2018; **26**: 30911–30929.

23 Borhani N, Kakkava E, Moser C, Psaltis D. Learning to see through multimode fibers. *Optica* 2018; **5**: 960–966.

24 Fan P, Zhao T, Su L. Deep learning the high variability and randomness inside multimode fibers. *Opt Express* 2019; **27**: 20241–20258.

25 Li Y, Xue Y, Tian L. Deep speckle correlation: a deep learning approach toward scalable imaging through scattering media. *Optica* 2018; **5**: 1181–1190.

26 Goodman JW. *Speckle Phenomena in Optics: Theory and Applications*. Roberts and Company Publishers, 2007.

27 Xie X, Zhuang H, He H, Xu X, Liang H, Liu Y *et al.* Extended depth-resolved imaging through a thin scattering medium with PSF manipulation. *Sci Rep* 2018; **8**: 1–8.

28 Liao M, Lu D, Pedrini G, Osten W, Situ G, He W *et al.* Extending the depth-of-field of imaging systems with a scattering diffuser. *Sci Rep* 2019; **9**: 1–7.

29 Horisaki R, Horisaki R, Okamoto Y, Tanida J. Single-shot noninvasive three-dimensional imaging through scattering media. *Opt Lett* 2019; **44**: 4032–4035.

30 LABEYRIE A. Attainment of diffraction limited resolution in large telescopes by Fourier analysing speckle patterns in star images. *Attain Diffr Ltd Resolut Large Telesc Fourier Anal Speckle Patterns Star Images* 1991; **28**: 427–429.

31 Wu Y, Rivenson Y, Zhang Y, Wei Z, Günaydin H, Lin X *et al.* Extended depth-of-field in holographic imaging using deep-learning-based autofocusing and phase recovery. *Optica* 2018; **5**: 704–710.
32 McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *ArXiv180203426 Cs Stat* 2018. http://arxiv.org/abs/1802.03426 (accessed 14 Apr 2020).

33 Becht E, McInnes L, Healy J, Dutertre C-A, Kwok IWH, Ng LG *et al*. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol* 2019; 37: 38–44.

34 MNIST handwritten digit database, Yann LeCun, Corinna Cortes and Chris Burges. http://yann.lecun.com/exdb/mnist/ (accessed 11 May 2020).
Fig. S1. Network structure.
Fig. S2. Performance comparison on different downsize methods and different network models. We use the same training dataset and downsized it using pixel-binning and subsampling respectively. Pixel-binning is averaging among a 4×4 pixel block. And subsampling is to take a pixel among a 4×4 pixel block. We fed two types of preprocessed data to trained two networks. The CNN we trained is without the transformation module.