Supplemental Document

Optics EXPRESS

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

YUNZHE LI, SHIYI CHENG, YUJIA XUE, AND LEI TIAN*

Department of Electrical and Computer Engineering, Boston University, Boston, MA 02215, USA
**leitian@bu.edu*

Supplement DOI: https://doi.org/10.6084/m9.figshare.13489350

Parent Article DOI: https://doi.org/10.1364/OE.411291

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

In this supplement, we provide additional information of the main paper. The details of network implementation are described in section 1. In addition, we analyze the functionalities of two major paths in our network including encoder-decoder path and skip connection path in section 2. The section includes the procedure of constructing the effective plain encoder-decoder network by blocking the skip connection information flow, the procedure of constructing the effective skip-connection-only path by blocking the encoder-decoder flow and their visualization and quantification results.

1. PROPOSED NETWORK AND PERFORMANCE COMPARISONN

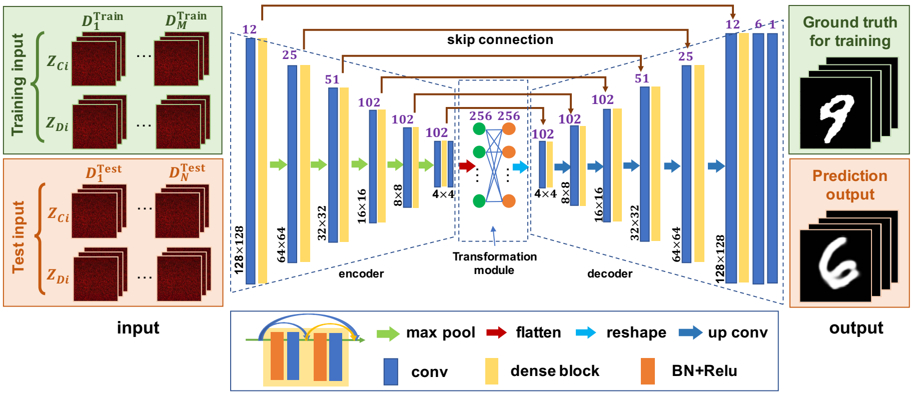

Fig. S1. The proposed DNN. The input is a preprocessed 128×128 speckle intensity. The input then goes through the "encoder path", which yields a stack of 4×4 latent code. Next, the latent code is flattened to a 1D vector, which is input to two fully connected layers and then reshaped to 2D. The decoder reverses the process that recombines the information into feature maps with gradually increased lateral details. Skip connections are used to transfer additional information from the encoder to the decoder without going through the bottleneck. The final output is a binary object prediction.

The proposed DNN structure is shown in Fig. S1. The input to the DNN is a preprocessed 128×128 speckle intensity. Two preprocessing techniques, including pixel binning and subsampling, are compared in Fig. S2. Next, the input goes through the "encoder path", which consists of four dense blocks connected by a max pooling layer for down-sampling. Each dense block consists of two layers, in which each layer performs batch-normalization (BN), the rectified linear unit (ReLU) nonlinear activation, and convolution (conv) with three filters. The intermediate output from the encoder has small lateral dimensions (4×4), but encodes rich information along the "depth". The latent code includes case-specific information that encodes the displacement and diffuser parameter. A transformation module is then concatenated to the encoder, consisting of a flatten layer that outputs a 1-D latent vector, two fully connected layers with ReLU activation and a constant reshaping layer that transforms back to a 2D feature map. 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 the shift-variance effect of the imaging process. The benefit of the transformation module are discussed in Fig. S2. Next, the low-resolution feature maps go through the “decoder path”, which consists of four additional dense blocks connected by up-sampling followed by conv 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 conv 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.

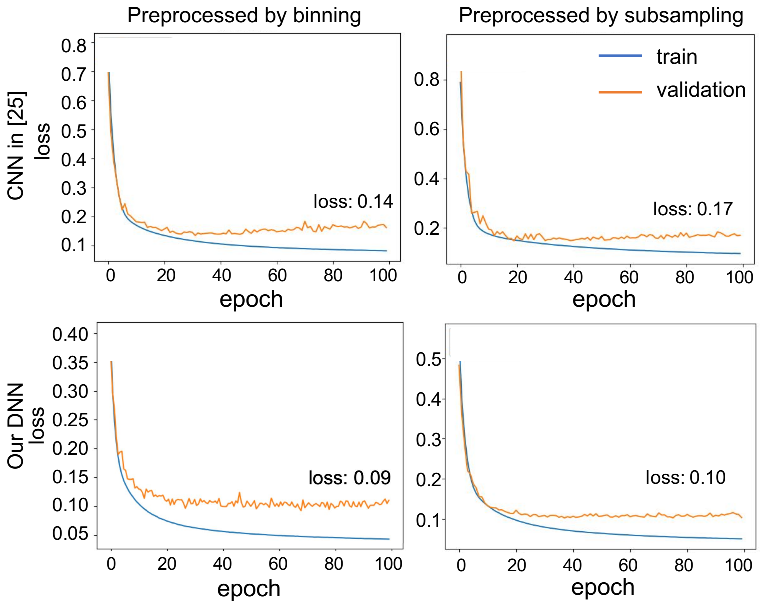

Fig. S2. Performance comparisons on different speckle preprocessing methods and different network structures. Two down-sampling strategies in preprocessing are compared, including pixel-binning and subsampling. The pixel-binning is performed by averaging the intensity values for every 4×4 pixels. The subsampling is performed by taking the intensity value from the upper-left corner pixel for every 4×4 pixels. We fed these two types of preprocessed data to train two DNNs, including the one from [1] and the proposed DNN with the transformation module. The network with the transformation module paired with the pixel-binning preprocessing method provides the lowest validation loss.

2. THE ENCODER-DECODER PATH AND THE SKIP CONNECTIONS PROVIDE DISTINCT FUNCTIONALITIES

We compared three different network structures using the weights directly loaded from our trained network.

In Fig. S3(a)(i), the modified U-net used in this work and a few representative prediction results on seen and unseen diffusers as shown as the benchmark. Next, in Fig. S3(a)(ii) we blocked the skip connections of the trained network to effectively construct a plain encoder-decoder network. To block the information flow of the skip connections or the latent code, we modified the layer weights using the following procedures. Since each skip connection is implemented by a merge layer (concat) that concatenates the feature maps from the encoder layer to the matching decoder layer followed by a 2D conv, we extracted the weights of 2D conv layer and set those for bridging the information from the encoder layer to be zero. By doing so way, each concat layer effectively only passes the decoder feature maps to the next layer while blocking the feature maps from the skip connections. To validate this approach, we also constructed a plain encoder-decoder network without the skip connections, while keeping the rest of hyper-parameters the same as our trained network. We then passed the corresponding weights from the trained network to the plain encoder-decoder network. Next, we performed predictions using both the plain encoder-decoder network and our network with blocked skip connections on the same testing data, and showed identical prediction results. The corresponding prediction results showed that only minor blurring is resulted in the reconstruction. We further quantitatively evaluate the predictions from the plain encoder-decoder network in Fig. S3(b) using the same testing data as Fig.5 in the manuscript. Notably, the performance on the unseen diffuser across all positions remain consistently around 0.6 and is similar to that from Fig.5 in manuscript using the full U-net structure. Slight performance drop was observed on the seen diffuser cases with reduced mean PCC and increased std. Lastly, in Fig. S3(a)(iii) we blocked the information flow from the latent code to the first layer of the decoder. The latent code information can be blocked by setting the weights of 2D conv layer immediately after the bottleneck layer to be zero. By doing so, the first decoder layer always outputs zero feature maps regardless of the latent code. The corresponding prediction results showed severe degradation.

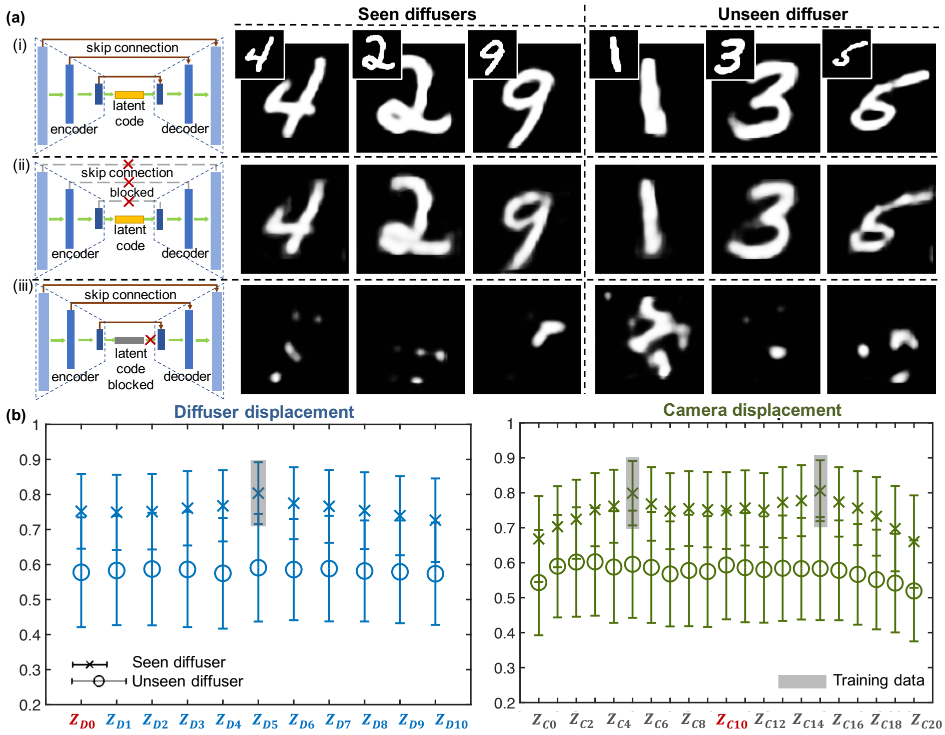

Fig. S3. Analysis of the roles of the encoder-decoder path and the skip connections. (a) Overview of three different network structures under study and their representative prediction results on seen and unseen diffusers. (i) Our DNN model contains both an encoder-decoder path and skip connections. (ii) The encoder-decoder network with the skip connections being blocked. (iii) The skip-connection only network with the latent code layer being blocked. (b) Quantitative evaluation of the performance from the network in (a)(ii) without skip connections. Each cross (seen diffusers) or circle (unseen diffuser) marker represents the mean PCC of the predictions at each position. Each error bar quantifies the standard deviation of the prediction results at each position. The training displacement positions are marked by the grey box.

Empirically, we also found that adding the skip connections helps preventing overfitting. To show this, we optimized the hyper-parameters of the plain encoder-decoder network (in Fig. S3(a)(ii)) and trained it from scratch using random initial weights on the same training dataset. The prediction results showed that the average PCCs for the seen diffuser cases improved slightly to around 0.9, while the PCCs on the unseen diffuser dropped to around 0.5, as summarized in Fig. S4. This study shows that the plain encoder-decoder network tends to overfit to the seen diffuser dataset and does not generalize well to unseen diffusers. This issue was effectively overcome by the skip connections in our network.

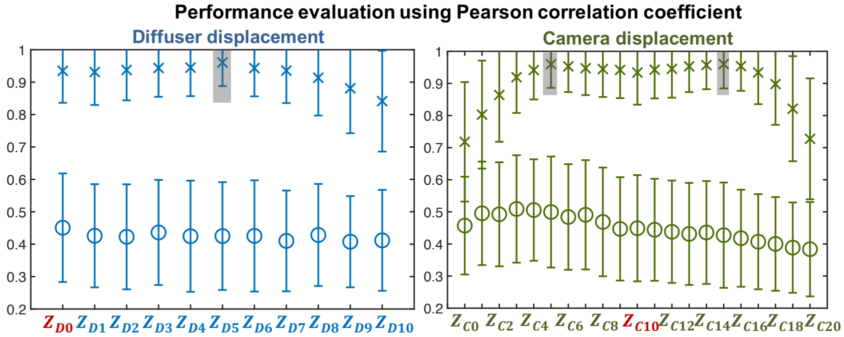

Fig. S4. Quantitative performance evaluation results of directly training a tuned encoder-decoder network. Each cross (seen diffuser) or circle (unseen diffuser) marker represents the mean PCC of the predictions at each displacement position. Each error bar indicates the standard deviation of the prediction results at each position. The training positions are marked by the grey box.

REFERENCES

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