

# Imaging through glass diffusers using densely connected convolutional networks: supplementary material

SHUAI LI<sup>1,\*</sup>, MO DENG<sup>2</sup>, JUSTIN LEE<sup>3</sup>, AYAN SINHA<sup>1,5</sup>, AND GEORGE BARBASTATHIS<sup>1,4</sup>

<sup>1</sup>Department of Mechanical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139

<sup>2</sup>Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139

<sup>3</sup>Institute for Medical Engineering Science, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139

<sup>4</sup>Singapore-MIT Alliance for Research and Technology (SMART) Centre, One Create Way, Singapore 117543, Singapore

<sup>5</sup>Current address: Magic Leap Inc, 1376 Bordeaux Drive, Sunnyvale, CA 94089

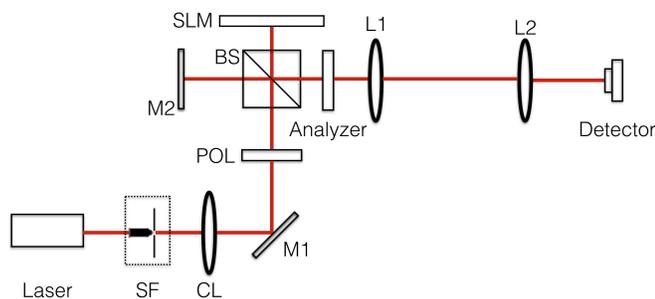
\*Corresponding author: shuaili@mit.edu

Published 6 July 2018

This document provides supplementary information to "Imaging through glass diffusers using densely connected convolutional networks," <https://doi.org/10.1364/OPTICA.5.000803>.

## 1. CALIBRATION OF THE SLM

In order to calibrate the modulation performance of the SLM, we built a Michelson interferometer shown in Fig. S1. Compared with our experimental setup shown in Fig. 1 in the main manuscript, we remove the glass diffuser in this calibration setup and place a reflecting mirror M2 to the left of the beam splitter, creating the reference beam.



**Fig. S1.** The optical setup for calibrating the phase and intensity modulation of SLM. SF: spatial filter; CL: collimating lens; M1, M2: mirror; L1, L2: lens; POL: linear polarizer; BS: beam splitter; SLM: spatial light modulator.

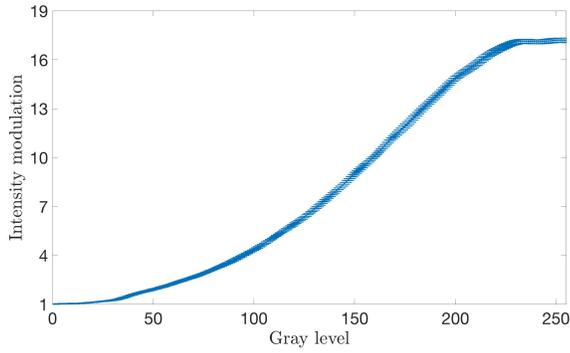
While calibrating the intensity modulation, we use a photon diode sensor (Newport, 818-SL) as the detector. The mirror M2

is blocked so that the light being detected comes from the SLM only. All the pixels of the SLM are driven by the same value  $V$  (uniform) and the corresponding intensity values measured by the photon diode are recorded. By changing  $V$  from 0 to 255 and repeating the measurement, we obtain the intensity modulation curve as shown in Fig. S2, which is normalized to the intensity values at  $V = 0$ . The intensity modulation of the SLM follows a monotonic relationship with respect to the assigned pixel value and a maximum intensity modulation ratio of  $\sim 17$  can be achieved.

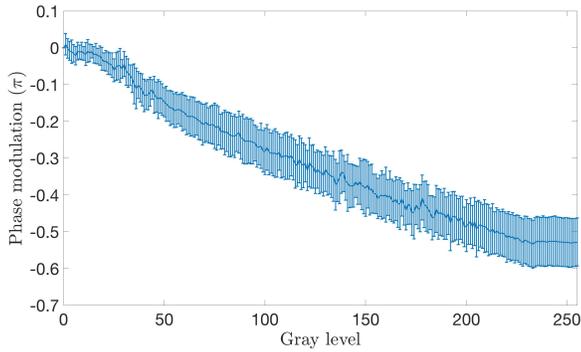
While calibrating the phase modulation, we use the CMOS camera as the detector and the interferometer setup is used. All the pixels of the SLM are driven by the same value  $V$  (uniform) and the corresponding interference patterns are captured by the CMOS camera. Let  $\phi(V)$  denote the relative phase between the light reflected by the uniform SLM when its pixels are all driven with gray scale value  $V$ , relative to M2; and  $I_0$  denote the intensity reflected by M2. In the uniform illumination case, given  $M(V)$ , which is the calibrated intensity modulation, the intensity recorded by the CMOS should be of the form

$$\frac{I(V)}{I_0} = 1 + M(V) + 2\sqrt{M(V)} \cos \phi(V) \quad (\text{S1})$$

We arbitrarily assigned  $\phi(0) = 0$  radians. The phase modulation curve is shown in Fig. S3. The phase depth is  $\sim 0.6\pi$ .



**Fig. S2.** Experimentally calibrated intensity modulation curve with error bounds in the grayscale range of [0,255] for the SLM.



**Fig. S3.** Experimentally calibrated phase modulation curve with error bounds in the grayscale range of [0,255] for the SLM.

## 2. ANALYSIS OF THE INFLUENCE OF PHASE MODULATION

In this paper, we want the SLM to perform as a pure intensity object. However, as we showed in the last section, due to the optical anisotropy of the liquid crystal molecules, the SLM will always perform a correlated phase modulation and the phase depth is  $\sim 0.6\pi$  for our experimental arrangement. In order to analyze the influence of this phase modulation in the formation of the speckle patterns, we carry out the following simulations.

For a randomly given image uploaded to the SLM, we simulate the corresponding speckle patterns on the CMOS camera using equation (2) and (3) in the main manuscript for two cases:

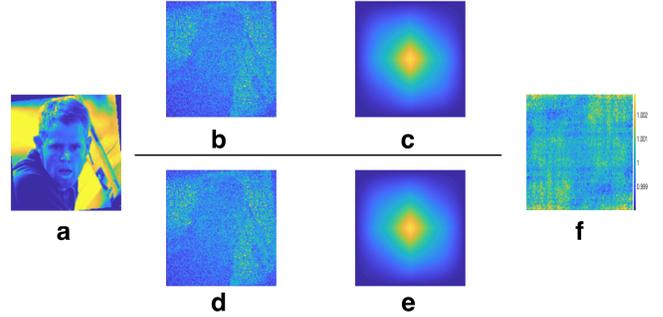
(a): Assuming the SLM to perform both intensity and phase modulation, which is the actual modulation in practice, i.e.  $g(x, y) = \sqrt{M[V(x, y)]} \exp\{iP[V(x, y)]\}$ . Here,  $V(x, y)$  is the pixel size of the uploaded image,  $M(V)$  and  $P(V)$  are the intensity modulation and phase modulation curves as shown in Fig. S2 and Fig. S3, respectively.

(b): Assuming the SLM to perform intensity modulation only, i.e.  $g(x, y) = \sqrt{M[V(x, y)]}$ .

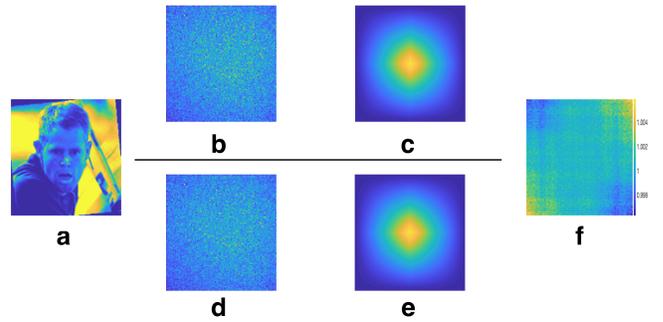
In our simulation, we set those parameters as:  $\mu = 16\mu\text{m}$ ,  $\sigma_0 = 5\mu\text{m}$ ,  $\sigma = 4\mu\text{m}$  for the 600-grit diffuser and  $\mu = 63\mu\text{m}$ ,  $\sigma_0 = 14\mu\text{m}$ ,  $\sigma = 15.75\mu\text{m}$  for the 220-grit diffuser. Other simulation parameters are set to be the same as the actual experiment:  $z_d = 15\text{mm}$ ,  $R = 12.7\text{mm}$  and  $\lambda = 632.8\text{nm}$ .

For the speckle patterns obtained in case (a) and (b), we also

compute their respective autocorrelation functions and take the element-wise ratios between them. This process is repeated for both diffusers and the corresponding results are shown in Fig. S4 and S5.

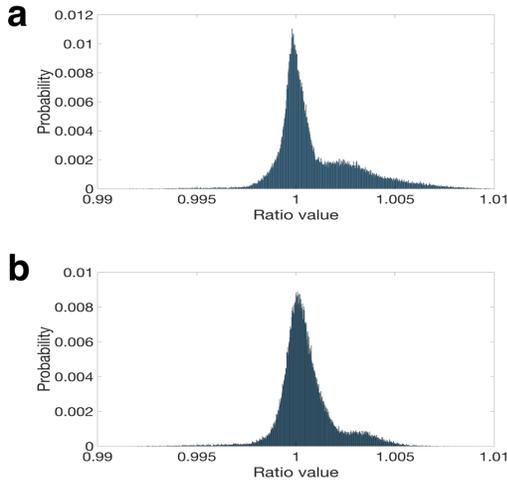


**Fig. S4.** Analysis of the influence of phase modulation in the formation of speckle patterns for 600-grit diffuser. (a) Input image; (b) Simulated speckle pattern for the complex object; (c) Autocorrelation of the speckle in (b); (d) Simulated speckle pattern for the pure-intensity object; (e) Autocorrelation of the speckle in (d); (f) Element-wise ratios between the autocorrelations in (c) and (e).



**Fig. S5.** Analysis of the influence of phase modulation in the formation of speckle patterns for 220-grit diffuser. (a) Input image [1]; (b) Simulated speckle pattern for the complex object; (c) Autocorrelation of the speckle in (b); (d) Simulated speckle pattern for the pure-intensity object; (e) Autocorrelation of the speckle in (d); (f) Element-wise ratios between the autocorrelations in (c) and (e).

We find that for both diffusers, the speckle patterns obtained in case (a) and (b) are qualitatively similar. Most importantly, the autocorrelation ratio values are all  $\sim 1$  with a small standard deviation, which indicates that the correlated phase modulation will not change the statistics of the resulted speckle patterns. To further demonstrate this point, we repeat this simulation process for 10 times for both diffusers and plot the probability histograms for the resulting autocorrelation ratios. The results are shown in Fig. S6. As expected, the autocorrelation ratio values are all  $\sim 1$ . We also compute the mean values  $\mu_r$  and standard deviations  $\sigma_r$  for the two histograms:  $\mu_r = 1.0009$ ,  $\sigma_r = 0.0020$  for the 600-grit diffuser and  $\mu_r = 1.0005$ ,  $\sigma_r = 0.0013$  for the 220-grit diffuser. Therefore, we conclude that the influence of the correlated phase modulation can be neglected and we can reasonably approximate the SLM as a pure intensity object.



**Fig. S6.** Quantitative analysis of the influence of phase modulation in the formation of speckle patterns. (a) 600-grit diffuser; (b) 220-grit diffuser.

### 3. ADDITIONAL DETAILS OF THE IDIFFNET

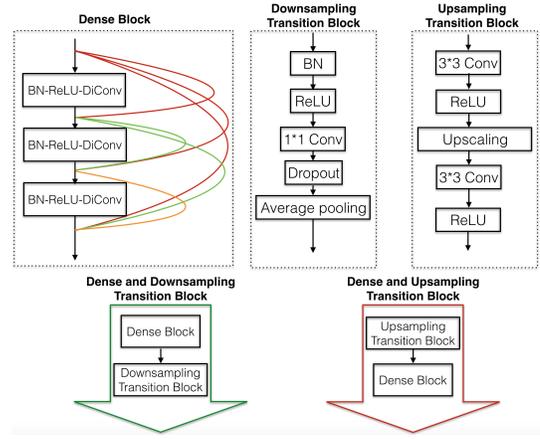
As described in the main manuscript, we use densely connected convolutional networks (DenseNets) to construct our IDiffNet. Fig. S7 shows the detailed architectures of the different blocks used in our IDiffNet. Each dense block consists of three composite convolutional layers and each layer connects to every other layer within the same block in a feed-forward fashion. The growth rate  $k$  is set to be 12 and the initial number of filters is set to be 16. Each composite convolutional layer is comprised of three consecutive operations: batch normalization (BN), rectified linear unit (ReLU) and dilated convolution (DiConv) with filter size  $5 \times 5$  and dilation rate 2. We use dilated convolutions so as to increase the receptive field of the convolution filters. The downsampling transition block consists of an average pooling operation with stride (2,2). As a result, the dimension of the input to this block is reduced by a factor of 2 at the output. The upsampling transition block increases the dimension of the input by a factor of 2. This is achieved by the subpixel upsampling operation [2]. The dense and downsampling transition block is built by placing a downsampling transition block after a dense block, while the dense and upsampling transition block is built by placing a dense block after an upsampling transition block.

We used  $L_2$  regularization with weight decay of  $1E-4$  in all convolutional filters initialized with random numbers from a Gaussian distribution. The same regularization was used in batch normalization as well. A small dropout rate of 0.05 was set to prevent overfitting. Because of GPU memory constraints, our IDiffNet was trained with a mini-batch size 8 using ADAM optimizer in Tensorflow. We started the training with a learning rate of 0.001 and dropped it by a factor of 2 after every 5 epochs. Additionally, we clipped the gradients at value 1 to stabilize the training. The neural network was trained for 20 epochs with the training samples being shuffled at every epoch.

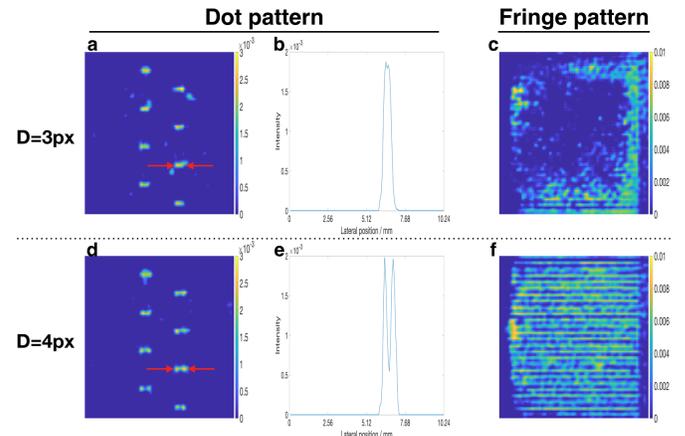
### 4. RESOLUTION TEST FOR IDIFFNET TRAINED USING NPCC AS LOSS FUNCTION

Here, we show the resolution test results of the IDiffNets trained using NPCC as loss function. The diffuser used is 600-grit.

As shown in Fig. S8, the IDiffNet trained on MNIST database

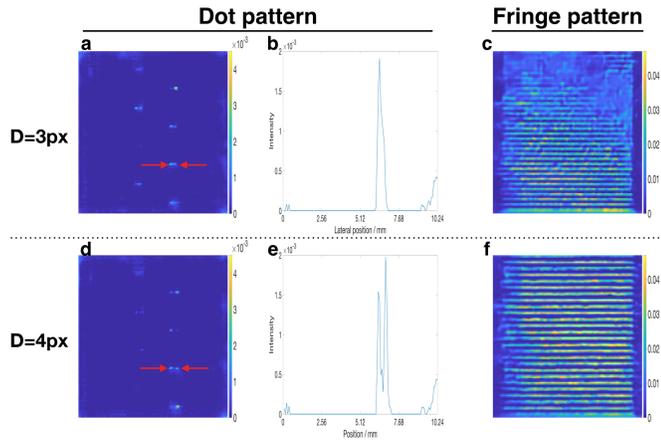


**Fig. S7.** Detailed architectures of the different blocks in our IDiffNet



**Fig. S8.** Experimental resolution test result for IDiffNet trained on MNIST using NPCC as loss function. The diffuser used is 600-grit. (a) Reconstructed dot pattern when  $D = 3$  super-pixels. (b) 1D cross-section plot along the line indicated by red arrows in (a). (c) Reconstructed fringe pattern when  $D = 3$  super-pixels. (d) Reconstructed dot pattern when  $D = 4$  super-pixels. (e) 1D cross-section plot along the line indicated by red arrows in (d). (f) Reconstructed fringe pattern when  $D = 4$  super-pixels.

is able to resolve two dots with spacing  $D = 4$  super-pixels, but fails to distinguish two dots with spacing  $D = 3$  super-pixels. Same spatial resolution is demonstrated using fringe patterns as well, where nearby fringes with spacing  $D = 4$  super-pixels are resolved while fringes with spacing  $D = 3$  super-pixels are unable to be distinguished. In addition, we find that the reconstruction qualities of dot patterns are better than those of the fringe patterns. This result is as expected since the MNIST training database imposes a strong sparsity prior, making the IDiffNet perform better on sparse test samples (dot patterns) than other less sparse test samples (fringe patterns). Therefore, dot patterns are more appropriate to be used to test the resolution of IDiffNet trained on MNIST. For the IDiffNet trained on ImageNet, its spatial resolution is the same as the MNIST training case, which is demonstrated in Fig. S9. However, the reconstruction qualities of fringe patterns are better than those of the dot patterns since the ImageNet training database contains



**Fig. S9.** Experimental resolution test result for IDiffNet trained on ImageNet using NPCC as loss function. The diffuser used is 600-grit. (a) Reconstructed dot pattern when  $D = 3$  super-pixels. (b) 1D cross-section plot along the line indicated by red arrows in (a). (c) Reconstructed fringe pattern when  $D = 3$  super-pixels. (d) Reconstructed dot pattern when  $D = 4$  super-pixels. (e) 1D cross-section plot along the line indicated by red arrows in (d). (f) Reconstructed fringe pattern when  $D = 4$  super-pixels.

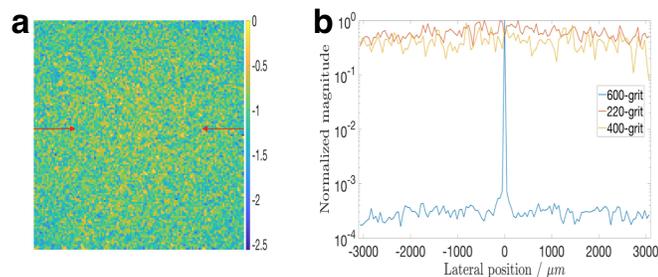
more general images and no longer conform to the sparsity prior. Hence, fringe patterns should be used to test the resolution of IDiffNet trained on ImageNet. All the above observations are the same as those in the MAE training case. Therefore, the choice of loss function does not affect the spatial resolution of the trained IDiffNet in the 600-grit diffuser case.

## 5. RESULTS OF 400-GRIT DIFFUSER

In this section, we summarize the results from the 400-grit diffuser (OptoSigma, DFB1-30C02-400).

### A. Simulated PSF

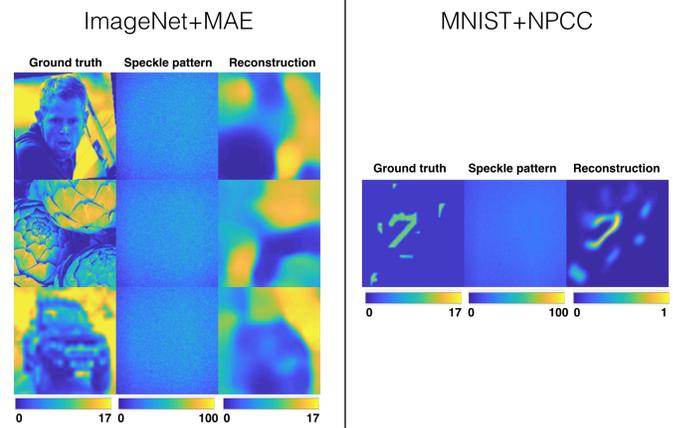
We first simulated the point spread function (PSF) of our imaging system using Eqs.(2) and (3) in the main manuscript. The results are shown in Fig.S10. For comparison, we also plot the cross-sections of PSFs in the 600-grit and 220-grit case. Similar to the 220-grit diffuser case, the PSF for the 400-grit diffuser spreads more widely than that of the 600-grit one. This indicates that the 400-grit diffuser also scatters the light strongly. As a result, we expect that the raw image that we obtain through the 400-grit diffuser should be a pure speckle with no discernible features .



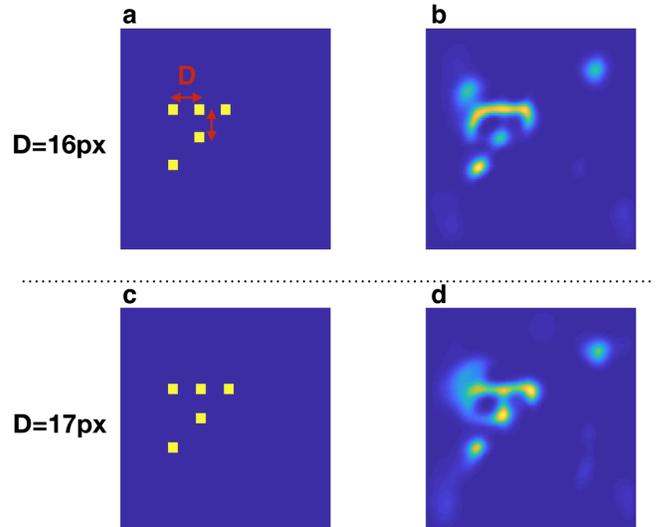
**Fig. S10.** (a) PSF for the 400-grit diffuser. (b) Cross-section of the PSFs along the lines indicated by the red arrows.

### B. Reconstruction

With the 400-grit diffuser inserted, we experimentally collected the training data and test data. Then, we trained two IDiffNets separately: (i) Trained on ImageNet using MAE as loss function; (ii) Trained on MNIST using NPCC as loss function. The reconstruction results on the test data are shown in Fig. S11.



**Fig. S11.** Qualitative analysis of IDiffNets in 400-grit diffuser case



**Fig. S12.** Experimental resolution test result for IDiffNet trained on MNIST using NPCC as loss function. The diffuser used is 400-grit. (a) Resolution test pattern when  $D = 16$  super-pixels. (b) Reconstructed test pattern when  $D = 16$  super-pixels. (c) Resolution test pattern when  $D = 17$  super-pixels. (d) Reconstructed test pattern when  $D = 17$  super-pixels.

As expected, due to the strong scattering of the 400-grit diffuser, the raw images that we obtained were all pure speckles, which is the same as the 220-grit case. The reconstruction performance is also the same as the 220-grit case, where the IDiffNet trained on ImageNet is only able to recover the salient features of the object, while the IDiffNet trained on MNIST using NPCC loss function is able to achieve high quality reconstruction for those sparse test objects.

### C. Resolution test

We also tested the spatial resolution of the IDiffNet in the 400-grit diffuser case. The results are shown in Fig. S12. We can find that the trained IDiffNet is able to resolve nearby dots with spacing  $D = 17$  super-pixels, but fails to distinguish two dots with spacing  $D = 16$  super-pixels. This result is again the same as the 220-grit diffuser case.

Above all, since the scattering strength of the 400-grit diffuser is pretty close to that of the 220-grit diffuser (as shown in Fig. S10(b)), the reconstruction performance of the trained IDiffNets in these two cases are also very similar.

### REFERENCES

1. G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Tech. rep., Technical Report 07-49, University of Massachusetts, Amherst (2007).
2. W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in "Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition," (2016), pp. 1874–1883.