Dataset for the paper "Statistical Dependencies Beyond Linear Correlations in Light Scattered by Disordered Media"

Authors: Ilya Starshynov, Alex Turpin, Philip Binner, and Daniele Faccio, School of Physics & Astronomy, University of Glasgow, Glasgow G12 8QQ, United Kingdom

The purpose of the dataset are:

To demonstrate the existence of a nonlinear correlation in randomly scattered fields that connects different realisations of the scattering medium.

To show an example of image reconstruction through a scattering medium in the absence of any known linear correlations in particular, in the absence of the speckle memory effect.

The dataset is split into 3 folders each related to the corresponding figure in the manuscript:

Fig1: Contains the far field speckle patterns collected with the input image (input_pattern.bmp) containing 3 dots placed on the SLM and varying the disorder. The 2 files 1diff.npz and 1diff.npz correspond to one and two diffuser configuration respectively. The python script that generates the average autocorrelation, processing.py, is also attached.

Fig2\MI contain the code used to calculated the mutual information between the input and output patterns. The entry script is combinedMI_calc.m

Fig2\ML contains the numerically calculated (using Fresnel formula) datasets of the MNIST digits propagating through a double layer diffuser. Those data was used for training and testing an ANN predicting the digit from the speckle through an unknown disorder configuration, as shown in Fig2d of the paper.

Fig3: Contains the raw data for the experiment of imaging through 2 and 1 diffuser unseen scattering configurations. This is achieved by training a U-map ANN to reconstruct the images from the speckle patterns while varying the disorder (translating the diffusers). The files One_diffuser.zip and Two_diffusers.zip in the raw_data subfolder contain 96 P0NN.npy files each. Each of those files contains 1000 different 256x256 speckle images, corresponding to the first 1000 MNIST digits being demonstrated on the DMD while illuminating it with a laser and then passing the light through a diffuser/s and collecting the speckle by a camera. Each of the P0NN.npy files corresponds to one of the 96 different realizations of the disorder (different spots on a diffuser/s). The code subfolder contains all the scripts required to implement, train and test the ANN. averageCalc.py calculates the average intensity at the camera (with the speckle averaged out), which is then subtracted from each speckle during training/testing. It is calculated using all the spots, but it's not essential. Only a few (even one) spot could be used. averageCalc.py creates a file average_speckle.npy, which is already calculated and attached for each of the datasets. The MNIST_64.npy file contains first 1000 MNIST digits, in exactly the format used for DMD projection (also binarized before being projected on the DMD exactly in the same way as in the training code). trainingGPU.py is actual training code, in which each P0NN.npy file is read and fed into the network. The testing.py takes the saved model weights and makes a prediction on a specified subset of input examples. Finally, the actual network is built within the unetMod.py file. It is very similar to the code used in [1] with an important difference that the skip connection and up-sampling layer outputs are normalized before merging. Finally the pretrained models are in the trained_models subfolder. They are trained on 95 disorder configurations and are meant to be tested on the 96th.

[1] Li, Yunzhe, Yujia Xue, and Lei Tian. "Deep speckle correlation: a deep learning approach toward scalable imaging through scattering media." Optica 5.10 (2018): 1181-1190.