Supplementary information: Unwrapping Non-Locality in the Image Transmission Through Turbid Media

Mohammadrahim kazemzadeh[1,2], Liam John Collard[1,2,3], Filippo Pisano[1,4], Linda Piscopo[1,5], Massimo De Vittorio[1,5,6], Ferruccio Pisanello[1,6]

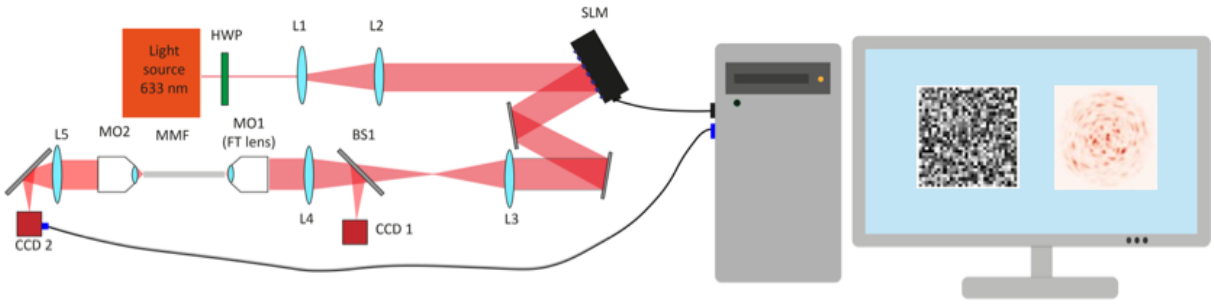
1 stituto Italiano di Tecnologia Center for Biomolecular Nanotechnologies
2 These two authors contributed equally
3 RAISE Ecosystem, Genova, Italy
4 Department of Physics and Astronomy, University of Padua, Via F. Marzolo, 8, 35131 Padua (Italy)
5 Dipartimento di Ingegneria dell'Innovazione, Università del Salento
6 These two authors are co-last authors

Email: mohammadrahim.kazemzadeh@iit.it, liam.collard@iit.it , Ferruccio.Pisanello@iit.it, Massimo.DeVittorio@iit.it

Optical setup

The detailed configuration of the optical setup utilized in this study is illustrated in Supplementary Figure 1. This system comprises three telescopic systems, a half-wave plate (HWP), a spatial light modulator (SLM), two reflecting flat mirrors, a beam splitter, a multimode optical fiber (MMF), and two microscope objectives (MO), each paired with a CCD camera. Lenses denoted as L1 and L2 (in this figure) are employed to couple the laser light into the SLM screen. The reflected light from the SLM is then directed into the first microscope objective (MO1), effectively conjugating the SLM screen onto the proximal side of the MMF.

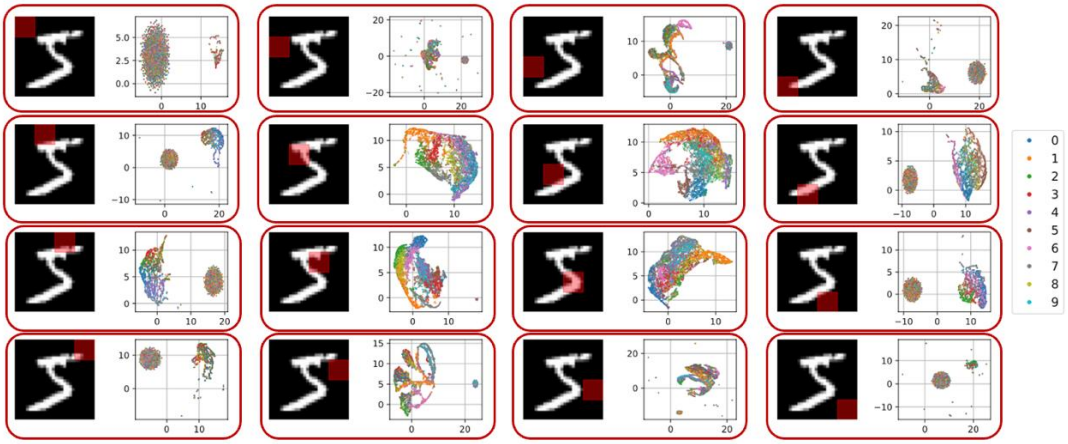
Monitoring the proximal side is achieved using CCD1 and a telescope setup (L4 and MO1). The generated spackle, as well as the distal side of the fiber, is recorded using CCD2, facilitated by a telescopic arrangement (MO2 and L5). The SLM pattern is commanded, and the spackle pattern is recorded using a computer in this setup.



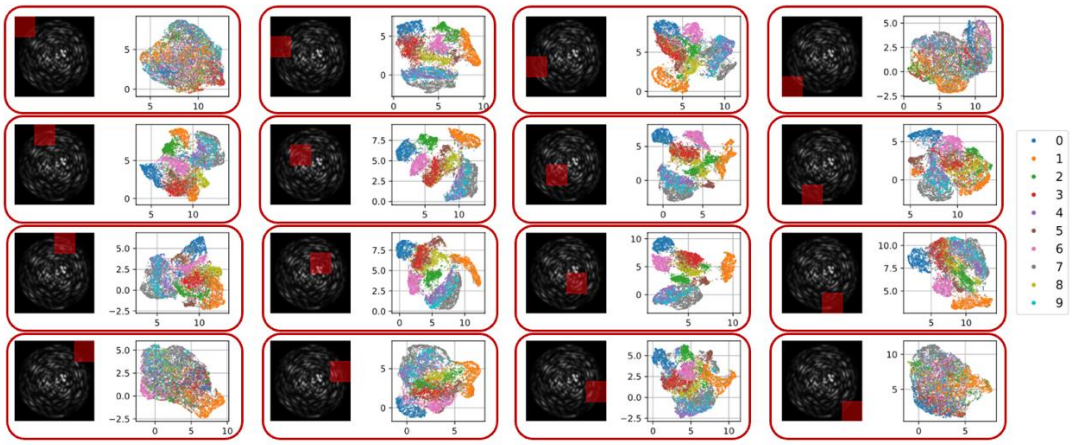
Supplementary Figure 1: detailed schematic of the optical setup used in this study.

Additional showcasing of non-locality

All the multivariate and deep learning regarding the study of MMF has been done through the pair data saved on this computer. In addition to the UMAP presented for the cropped section of MNIST and its corresponding spackle data in the manuscript, we conducted a similar analysis over additional cropped sections of both datasets. These results are showcased in supplementary Figures 2 and 3 for the actual digit images and their corresponding spackle, respectively.
It is evident that in the digit images, cropping leads to a complete collapse of digit clustering in the 2D UMAP projection. However, this phenomenon does not occur in the UMAP projection of the cropped spackle data. As explained in the manuscript, this effect is due to the introduced non-locality that stems from the physics of the multimode fiber.

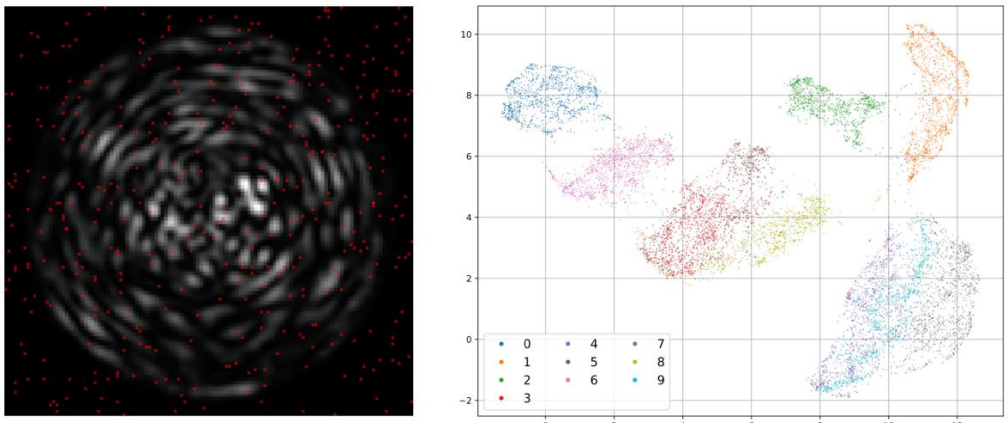


Supplementary Figure 2: showcases the cropped section of the MNIST image datasets, outlined by a red square within the digit image. The UMAP projection of this cropped section is depicted, where each individual dot represents an image of handwritten digits, color-coded according to the image label.

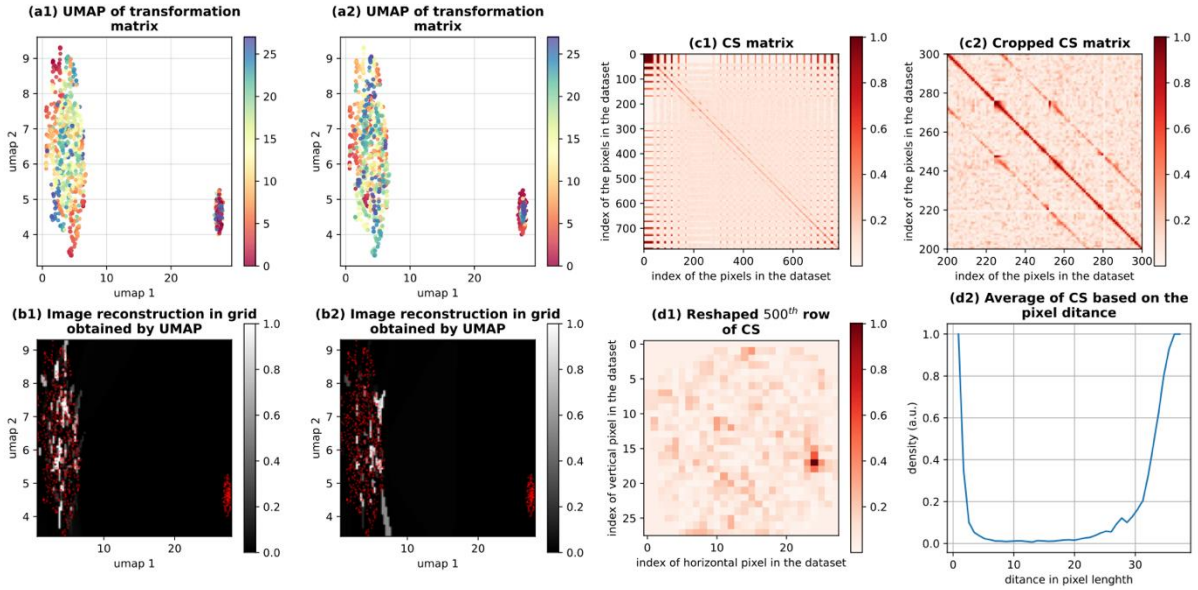


Supplementary Figure 3: the cropped section of the spackle generated with the MNIST dataset using our optical setup, delineated by a red square within the spackle image. The UMAP projection of this cropped section is showcased, with each individual dot representing a spackle and color-coded according to the image label.

To delve further into this matter, we applied this non-linear projection to a few randomly selected points from the spackle, as illustrated in supplementary Figure 4. The UMAP applied to these randomly selected points from the spackle image, marked by the red dots on the left side of the figure, reveals distinct clusters of digits. This observation supports the application of linear regression in reconstructing images through randomly selected points from the spackle datasets. As these points, despite their far spatial location, carries the topology of the digits in the MNIST datasets and possibly provides enough data for the image reconstruction throught them.



Supplementary Figure 4: left, shows the location of the sampled pixels in the spackle datasets. Right, the UMAP projection of the data from sampled points on the right. Each projected data is color coded based on their corresponding labels in the MNIST dataset.

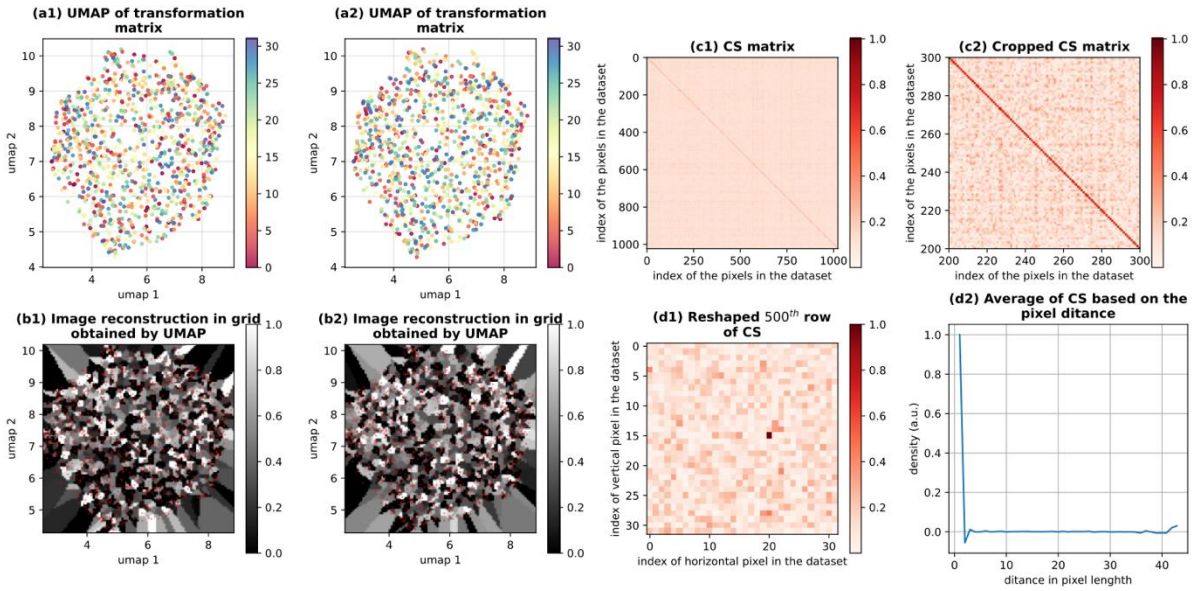


Supplementary Figure 5: (a1)-(a2) UMAP projections of the Transformation Matrices in Figure 2 (main manuscript) for the MNIST datasets. The projected points are color-coded based on the horizontal (a1) and vertical (a2) positions of the pixels on the SLM screen. (b1) and (b2) Representation of the reconstructed image using linear regression based on the grid obtained in (a1) and (a2). (c1) and (c2) Cosine similarity matrices comparing the transformation matrix of the MNIST dataset with a closer look at its details, respectively. (d1) Reshaped version of the 500th row of the cosine similarity matrix in (c1). (d2) The average cosine similarity based on the distances between pixels.

Analysis of linear transformation matrices of MNIST and randomly generated images datasets

In addition to the multivariate analysis conducted on the linear transformation matrix derived from the CIFAR dataset (main manuscript), similar analyses were performed on the transformation matrices obtained from the MNIST and randomly generated datasets. The outcomes of these analyses are presented in Supplementary Figures 5 and 6 for the MNIST and randomly generated datasets, respectively.

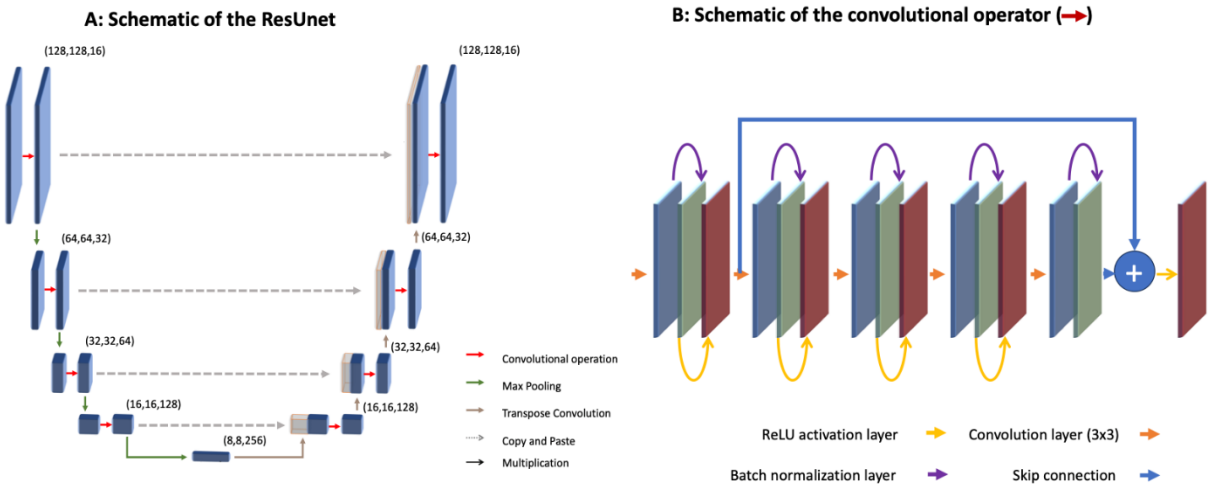
In Supplementary Figure 5, it is evident that the UMAP projection of the transformation matrix fails to capture the two-dimensional manifold of the spatial light modulator (SLM), in contrast to the CIFAR dataset. This is particularly pronounced when using the projection points for data visualization in (b1) and (b2), where the digits do not appear as clearly as in the CIFAR case. This discrepancy arises from the unused pixels on the SLM screen (corners of the digit images), resulting in trivial solutions of the linear regression matrix (zero lines in the obtained transmission matrix) and sharp pixel value changes compared to CIFAR.



Supplementary Figure 6: (a1)-(a2) UMAP projections of the Transformation Matrices in Figure 2 (main manuscript) for the randomly generated datasets. The projected points are color-coded based on the horizontal (a1) and vertical (a2) positions of the pixels on the SLM screen. (b1) and (b2) Representation of the reconstructed image using linear regression based on the grid obtained in (a1) and (a2). (c1) and (c2) Cosine similarity matrices comparing the transformation matrix of the randomly generated dataset with a closer look at its details, respectively. (d1) Reshaped version of the 500th row of the cosine similarity matrix in (c1). (d2) The average cosine similarity based on the distances between pixels.

Another significant observation in this figure is the presence of large areas with perfect similarity away from the main diagonal elements, corresponding to the zero lines in the MNIST transformation matrix. The CS matrix in this case exhibits fewer similarity features compared to that of the CIFAR, notably lacking the second parallel line to the main diagonal (shown in Supplementary Figure 5 (c2)), a clear contrast to the CIFAR dataset. This difference may be attributed to the more nonlinear behavior of the MNIST dataset, as discussed in the manuscript.

Supplementary Figure 6 illustrates a parallel analysis for randomly generated datasets. While this dataset doesn't encounter the blind spots observed in the transformation matrix of the MNIST dataset, its heightened nonlinearity prevents a successful reconstruction of the spatial light modulator (SLM) manifold through the acquired transformation matrix. Additionally, in this instance, the cosine similarity (CS) matrix does not manifest any similarity features.



Supplementary Figure 7: a detailed version of the ResUNet architecture. A, schematic of the employed network. B, the detailed layers used in the convolutional operator.

Detailed ResUNet architecture

The employed ResUNet in this study is depicted in Supplementary Figure 7. On the left side of this figure, we observe the schematic of the UNet, while the inner structure of the convolutional operator (illustrated by a red arrow) is displayed on the right side. This convolutional operator serves as a nested ResNet backbone, forming the ResUNet. It comprises 5 convolutional layers, each succeeded by batch normalization and a rectified linear (ReLU) activation layer. Additionally, the output of the first activation layer is skipped and added to the output of the last batch normalization layer to create the skip connection in the ResNet architecture. As mentioned earlier, this approach addresses issues such as accuracy degradation and gradient vanishing in a deep neural network.

The number of filters utilized in each of these convolutional layers is illustrated in panel A of this figure. For instance, the size of the tensor on the upper layers of the ResUNet is denoted as (128, 128, 16). This signifies that the feature map size in this layer is 128x128, with a channel size of 16. This channel essentially indicates the number of filters employed in the convolutional layer at that stage of the ResUNet. It's important to note that as we progress deeper into the network, the feature map size diminishes due to the presence of max-pooling layers, while we simultaneously employ more filters to capture additional features and prevent information loss.

Supplementary Figure 8: More examples of the image reconstruction through the proposed method. The celebrity faces dataset used for this analysis was obtained from the source [1].

1 Speckle 2 Reconstruction with GAM that trained on random dataset

3 Result after post training 4 Ground truth

Supplementary Figure 9: More examples of the image reconstruction through the proposed method. The celebrity faces dataset used for this analysis was obtained from the source [1].

References

[1] Liu, Z., Luo, P., Wang, X., & Tang, X. (2015). Deep learning face attributes in the wild. In *Proceedings of the IEEE international conference on computer vision* (pp. 3730-3738).