



Exploiting holographically encoded variance to transmit labelled images through a multimode optical fiber

LIAM COLLARD,^{1,2,4,†} MOHAMMADRAHIM KAZEMZADEH,^{1,5,†}
LINDA PISCOPO,^{1,3} MASSIMO DE VITTORIO,^{1,2,3,6,‡}
AND FERRUCCIO PISANELLO^{1,2,7,‡} 

¹*Istituto Italiano di Tecnologia, Center for Biomolecular Nanotechnologies, Arnesano, LE 73010, Italy*

²*RAISE Ecosystem, Genova, Italy*

³*Dipartimento di Ingegneria Dell'Innovazione, Università del Salento, Lecce 73100, Italy*

⁴*liam.collard@iit.it*

⁵*Mohammadrahim.Kazemzadeh@iit.it*

⁶*massimo.devittorio@iit.it*

⁷*ferruccio.pisanello@iit.it*

[†]These authors contributed equally and are co-first authors

[‡]These authors jointly supervised and are co-last authors of this work

Abstract: Artificial intelligence has emerged as promising tool to decode an image transmitted through a multimode fiber (MMF) by applying deep learning techniques. By transmitting thousands of images through the MMF, deep neural networks (DNNs) are able to decipher the seemingly random output speckle patterns and unveil the intrinsic input-output relationship. High fidelity reconstruction is obtained for datasets with a large degree of homogeneity, which underutilizes the capacity of the combined MMF-DNN system. Here, we show that holographic modulation can encode an additional layer of variance on the output speckle pattern, improving the overall transmissive capabilities of the system. Operatively, we have implemented this by adding a holographic label to the original dataset and injecting the resulting phase image into the fiber facet through a Fourier transform lens. The resulting speckle pattern dataset can be clustered primarily by holographic label, and can be reconstructed without loss of fidelity. As an application, we describe how color images may be segmented into RGB components and each color component may then be labelled by distinct hologram. A ResUNet architecture was then used to decode each class of speckle patterns and reconstruct the color image without the need for temporal synchronization between sender and receiver.

© 2024 Optica Publishing Group under the terms of the [Optica Open Access Publishing Agreement](#)

1. Introduction

The success of wavefront shaping (WS) techniques [1] has enabled exploitation of modal diversity in multimode optical fibers (MMFs) to finely control light transmission by operating phase-only modulation at the fiber input, overcoming the inherent optical turbidity of the waveguide. WS methods to control light transmission through MMFs are based on the recording of both intensity and phase of the speckle patterns at the fiber output and on the use of phase conjugation or transmission matrix method to create the desired amplitude distribution at the fiber output [2–11]. This has led to set of novel applications, including low invasiveness neural endoscopes with sub-cellular spatial resolution [12–17], far-field imaging [18], holographic optical tweezers [19], and remote control of plasmonic structures [20]. These techniques are complemented by the faster WS-free methods based on reflectance imaging [21,22] and compressive sensing [23], which are however limited in terms of signal to noise ratio.

Alternatively, recent works have shown the transmissive properties of MMFs may be evaluated by artificial intelligence techniques without the need for any phase measurement at the MMF's

output [24,25]. By directly projecting the screen of a phase-only spatial light modulator (SLM) onto the fiber facet via a 4f system, deep learning has been established as a technique to reconstruct the SLM pattern and thus “see” the SLM screen through the fiber, and the ability to “learn” the transmission matrix of a multimode fiber is now well understood to be an efficient method to reconstruct images transmitted through a MMF. For such a problem, the image may either be coupled by a 4f system directly projecting the image onto the fiber facet [24,25], or indirectly by focussing the image through Fourier transform lens onto the fiber facet [26]. Typically, tens of thousands of images are then coupled through the fiber to train the network and several thousand are used to validate it. Since these pioneering works, significant advancements have been made in improving the training fidelity. Recently, an attention layer has been shown to reduce the requirement of a large dataset down to hundreds of images [27] or, alternatively, a simpler network architecture (hidden layer dense neural network) may reduce the training time to several minutes [28]. Utilising training data sets where the fiber is physically or thermally perturbed, image reconstruction has been demonstrated with strong resilience to the bending [29–32], and temperature changes [33]. Wang et al investigated the role of light source line width and stability on image reconstruction [34]. Going full circle, by applying an “actor” and “trainer” network model, the machine learning technique can also be used to project desired light patterns through the fiber [35]. Despite these impressive advancements, little attention has been paid to the possibility of increasing the transmissive capabilities by an “all optical” method.

Here, we show how a Fourier lens-based coupling can be further exploited to increase the variance of the output speckle patterns and can “holographically label the dataset”. When a dataset is modulated by multiple holograms, we find that the data primarily clusters based on the holographic carrier, which in turn act as a “label” for the transmitted data. As an example, we demonstrate how a color image can be segmented and each color component then projected and transmitted with a superimposed holographic label. The developed deep learning method is then able cluster the data based on different labels and also to reconstruct the image without loss of fidelity. Previous works have largely focussed on the transmission of grayscale images with the notable exception of Caramazza et al [26], where time-division is used to transmit color information requiring temporal synchronization between the sender (SLM) and the receiver (CCD). Our technique of holographic labelling allows the data to be classified into distinct clusters based on only the speckle pattern and requires no temporal synchronization between sender and receiver. This supports the more general conclusion that introducing additional variance in the output speckles can help in better exploiting the wealth of information that can be transmitted through MMFs.

2. Methods

2.1. Optical techniques

Figure 1(A) shows the principle of encoding image data into a spatially varying hologram on the input facet of the fiber. Our results leverage on the hypothesis that changing the holographic label changes the basis of modes excited with lower mutual correlation than changing only the image data. This can be observed by studying the stacks of speckle patterns shown in Fig. 1(A), obtained by labelling the image data (ϕ_{data}) with (ϕ_{holo}) to obtain the phase mask:

$$\phi_{mask} = \arg(\exp(i(\phi_{data} + \phi_{holo}))) \quad (1)$$

We selected ϕ_{holo} to be a blazed grating: by varying its pitch and rotation the image data is moved across the fiber core. This generates much higher variance in the output speckle patterns than modulating only the image, which we attribute to a higher level of orthogonality in the plane of the facet over the Fourier plane (screen of SLM) as a basis for the modes of the fiber. ϕ_{data} can be mathematically reconstructed from ϕ_{mask} without loss of information and by taking

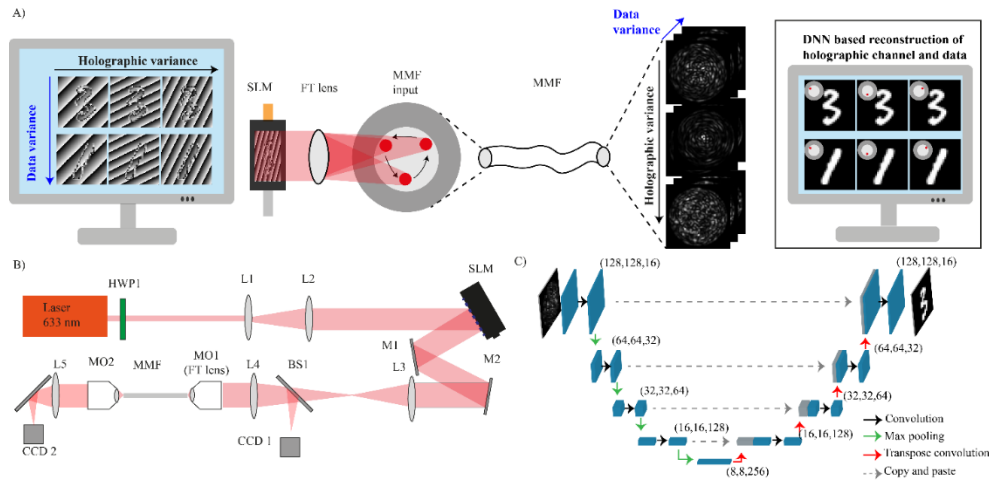


Fig. 1. A) Principle of encoding image data in a hologram. By the addition of a blazed grating, the data is shifted around the fiber core which encodes a higher level of variance in the output speckles than the data itself. B) Optical setup used to transmit data through the MMF, M—mirror, L—lens, MO—objective, MMF—multimode fiber, SLM—spatial light modulator, BS—beam splitter, and CCD—charged coupling device. C) Structure of the implemented ResUNet grayscale image reconstruction convolutional neural networks.

the Fourier transform of ϕ_{mask} , and it can also be verified that the image data is primarily encoded in the first order (as this is where the majority of amplitude lies). In the following, we use a blazed grating as ϕ_{holo} , however any other phase pattern that can increase the variance can be employed to implement the technique.

In the experimental implementations described in next paragraphs, the phase image ϕ_{mask} was injected into a MMF (50 μm core, 0.22 NA, Thorlabs (FG050LGA), approximately 5 cm long) through the optical setup illustrated in Fig. 1(B). The phase image was scaled between 0 (black) and π (white) to ensure maximum contrast between white and black regions of the image [33]. A 633 nm laser beam was expanded by a telescope and had its polarization rotated to match the screen of a reflective, phase-only, SLM. The reflected light passed through a second telescope which de-magnifies the beam to match the back aperture of MO1, coupling light into the MMF. The screen of the SLM is the Fourier plane of the focal point of MO1 and therefore pixels in the image correspond to different insertion angles. Therefore, higher order modes carry information from the edge of the image, while lower order modes represent the center of the image. The zero diffraction from the SLM was spatially filtered at the focal point of the telescope (L3/L4) by a removable razor blade, while the alignment of the hologram and MMF was monitored on a charge coupled device (CCD1). The transmission through the fiber was collected by MO2, and the MMF's output facet imaged on CCD2 to monitor the output speckle patterns. The SLM was given 150 ms to refresh when changing the pattern and the exposure time of the CCD was in total 60 ms.

2.2. Neural network

Speckle patterns generated at the output of the MMF are fed into a deep, fully convolutional neural network based on the ResUNet architecture [36], having the role of unveiling the phase pattern ϕ_{mask} based only on the speckle data. ResUNet is an evolution of the U-Net architecture [37], enhancing its depth by incorporating a residual backbone [38]. This modification is crucial as it addresses issues such as accuracy degradation and vanishing gradients, allowing for more

effective and stable training [39]. A comprehensive visualization of the network's architecture is presented in Fig. 1(C), including the number of layers and the shape of the output tensor at each stage, providing a clear reference for readers to grasp the network's structural intricacies. In terms of activation functions, all neurons in this network utilize the rectified linear unit (ReLU). This choice not only simplifies the network's computations but also aligns with the desired behaviour, making it a suitable choice for this specific application. Moreover, ReLU is preferred for its ability to achieve faster training speeds compared to other activation functions, contributing to the overall efficiency of the learning process [40]. The convolutional operator used in this ResUNet architecture consists of 6 consecutive convolutional layers each followed by batch normalization and ReLU activation. The input and output of this operator are added together to implement the short skip needed in ResUNet architecture.

In the following, we first identify and describe the increased variance in the speckle patterns dataset. Then we verify that the proposed holographic modulation provides state-of-the-art reconstruction quality and, in the final part of the manuscript, we provide the evidence that the additional variance can be employed as holographic label for the transmitted data. For this latter part we employed the approach to generate red, green and blue (RGB) labels and transmit segmented color images. We employed a network architecture comprising three parallel ResUNet components. Each of these components generates an output, which corresponds to transmitted color. The final 3 dimensional tensor within each network (disregarding the fourth dimension for data batches) is then concatenated along the third dimension, resulting in another 3D tensor with a three-fold increase in depth for each channel. This augmented tensor subsequently undergoes further processing through a convolutional network based on the ResNet architecture, aligning its output with a fully realized RGB 3D tensor. In all cases, 50 epochs of the network were tested (except for the RGB images for which 100 were developed). The epoch with minimal validation loss was then used for image reconstruction. It took approximately 1 minute to train each epoch of the network for grayscale image reconstruction and around 3 minutes per epoch for RGB image reconstruction. Therefore, 50 epochs of grayscale reconstruction can be trained in under 1 hour and 100 epochs of RGB reconstruction took around 5 hours to complete. Once the network was trained an image can be reconstructed in under 1 ms. We have chosen to use training datasets consisting of tens of thousands of images to ensure optimal training, but recent techniques based on attention mechanisms allow machine learning TM characterization using much smaller datasets [27].

3. Results

3.1. Encoding variance through holographic coupling

To verify that the amount of variance associated to the additional ϕ_{holo} can be greater than the one associated to ϕ_{data} , 2000 handwritten digits (28 by 28 pixels) in the MNIST dataset were coupled through 35 holographic labels ϕ_{holo} (corresponding to 35 positions on the input facet). The 35 input positions fully overlaid the core of the fiber as shown in Supplement 1. The number of modes carried by the fiber is approximately 1495 at 633 nm and thus capable of carrying the 28 by 28 MNIST digits (784 pixels). The resultant speckle patterns were then analysed using principal component analysis (PCA). Figure 2(A) shows a scatter plot of the two highest variance principal components (PC1 and PC2), colored by holographic label number (1 to 35, left) and Fig. 2(B) shows the data colored by MNIST digits (right). From this visualization it is clear that the two component PCA allows to obtain clusters related to the 35 holographic labels, while there is only a very slight sub-clustering corresponding to digit classification. Although some overlap is observed between the clusters in the two-components graph (PC1 and PC2), including higher order components of the PCA could separate this. This can be confirmed by considering the cumulative variance explained by the PCA, shown in Fig. 2(C). An inflection point at 70% of the cumulative variance is observed after the 34th component which corresponds to number of

holographic labels-1. Decreasing the number of labels moves the inflection point to left of the, confirming that the highest amount of variance is to be assigned to variance generated in the speckle pattern by ϕ_{holo} rather than by ϕ_{data} .

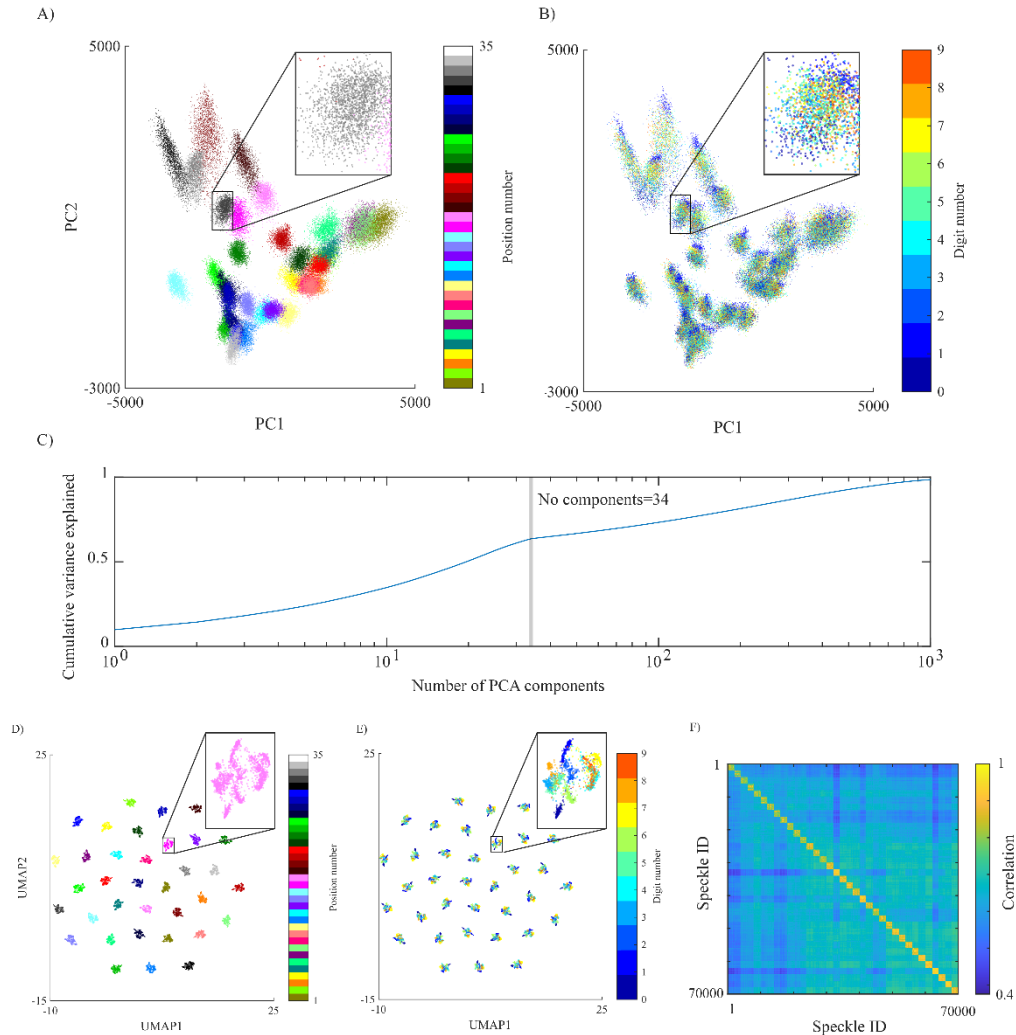


Fig. 2. Multivariate analysis of the speckle patterns corresponding to handwritten digits coupled through 35 unique holograms on the input facet. A) Scatter plot of first two PCA components for all 70000 speckle patterns colored by hologram number. The 35 input positions are evenly distributed over the entire core and shown in the supplement. B) Scatter plot of first two PCA components for all 70000 speckle patterns colored by digit number. C) Plot showing cumulative variance explained by each PC component, the inflection point at No of holograms-1 is marked. D) UMAP projection of all 70000 speckle patterns colored by hologram number E) UMAP projection of all 70000 speckle patterns colored by digit. F) Correlation matrix between the 70000 speckle patterns.

To gain a better representation of the speckle patterns dataset, a method that expresses a much wider amount of variance for each component is therefore required. We therefore applied uniform manifold approximation and projection for dimension reduction (UMAP), exploiting non-linear expression of the dataset and well suited to represent fiber speckle patterns [31,41]. The UMAP

analysis separates the data into 35 main clusters corresponding to each hologram as shown in Fig. 2(D). Compared to the PCA, the label-based clustering is extremely strong. Strikingly, the two component UMAP analysis is also capable of representing variance in the “digit” dataset (Fig. 2(E)) despite being an “unsupervised” technique. Indeed, when colored by handwritten digit, sub-clusters are observed in each of the 35 main clusters, with strong topological symmetry. Within each cluster, digits “0”, “1”, “3” and “6” appear to form distinct sub-clusters whereas “4”, “7” and “9” and “2”, “5” and “8” are overlaid. The clustering of holographic labels can also be observed by considering the correlation matrix between the speckle patterns shown in Fig. 2(F). Speckle ID corresponds with label’s position (that is, the first 2000 IDs correspond to first holographic label, ID 2001 to 4000 to the second holographic label, and so on until 68001 to 70000 corresponding to the 35th ϕ_{holo}). Strong correlation is observed between speckle patterns excited through the same holographic label. To verify that this result, and the principle of holographically encoded variance does not depend on fiber length or bending state, we repeated the above experiment with a 1-metre-long fiber in a coiled confirmation. The result of the PCA/UMAP is shown in [Supplement 1](#) and a similar clustering was observed.

3.2. Reconstruction of holographically coupled data

Whilst the multivariate analysis above allows recovery of input holographic label without supervision, to reconstruct the data within a single cluster a supervised CNN is required. To verify that this is possible with no loss of reconstruction fidelity, we have compared the reconstruction of handwritten digits (grayscale) transmitted through the fiber by setting ϕ_{holo} = zero or a blazed grating. In each case, 40000 digits were transmitted to train the ResUNet outlined in the experimental section, 5000 were used to validate and 5000 were used to test. The network was trained on the handwritten digit (ϕ_{data}) regardless of ϕ_{holo} and the ResUNet network depicted in Fig. 1(C) was used. Figure 3(A) shows reconstruction with ϕ_{holo} = zero, with the first row displaying a subset of the original handwritten input digit from the MNIST dataset, while the second shows ϕ_{mask} (which in this case is equivalent to ϕ_{data}). The third row shows the output speckle patterns when each ϕ_{mask} is set on the SLM and transmitted through the MMF. The ResUNet reconstruction of ϕ_{data} based on the output speckle patterns is shown in the final row. Figure 3(B) instead shows reconstruction with ϕ_{holo} = blazed grating, the rows are organized as in panel A. In both cases, the reconstructions are extremely similar to the original data and each digit is easily identifiable. We quantified this by measuring the structural similarity index (SSIM) between the reconstructed image and the original. For the test data, the average SSIM between the input image and CNN reconstructed was 0.95 ± 0.03 for ϕ_{holo} = Zero and 0.95 ± 0.03 when ϕ_{holo} = blazed grating, which is typical for handwritten digit reconstruction [32]. This is illustrated in the bar charts in Fig. 3(C) and histogram in Fig. 3(D) (only the data used to test the network is included in the statistical analysis and examples shown in Fig. 3). The training loss and validation loss are shown in Fig. 3(E), with the training loss being overlapped and therefore it can be determined that in both cases handwritten digits may be transmitted with high SSIM and equivalent training times. The model is built from the epoch when the validation loss is lowest. In terms of the overall training performance, it can be observed that whilst the validation loss converges, the training loss continued to decline with additional epochs. This suggests a small degree of overfitting of the model to the training data at higher epochs, however, as the validation loss did not rise this shows that they would also provide a viable reconstruction. Ultimately, the reconstruction of ϕ_{data} is not impacted on whether an additional blazed grating is added on.

Moving towards more complex datasets, we applied the same test to the CIFAR (grayscale) dataset. Typically, it is more challenging to transmit natural scenes than handwritten digits as the data has higher variance and, as such, the SSIM of the reconstruction is generally lower. As in the previous experiment, 40000 images were transmitted and used as a training set, while 5000 were used for validation and 5000 used to test. Reconstructions with ϕ_{holo} = zero or a

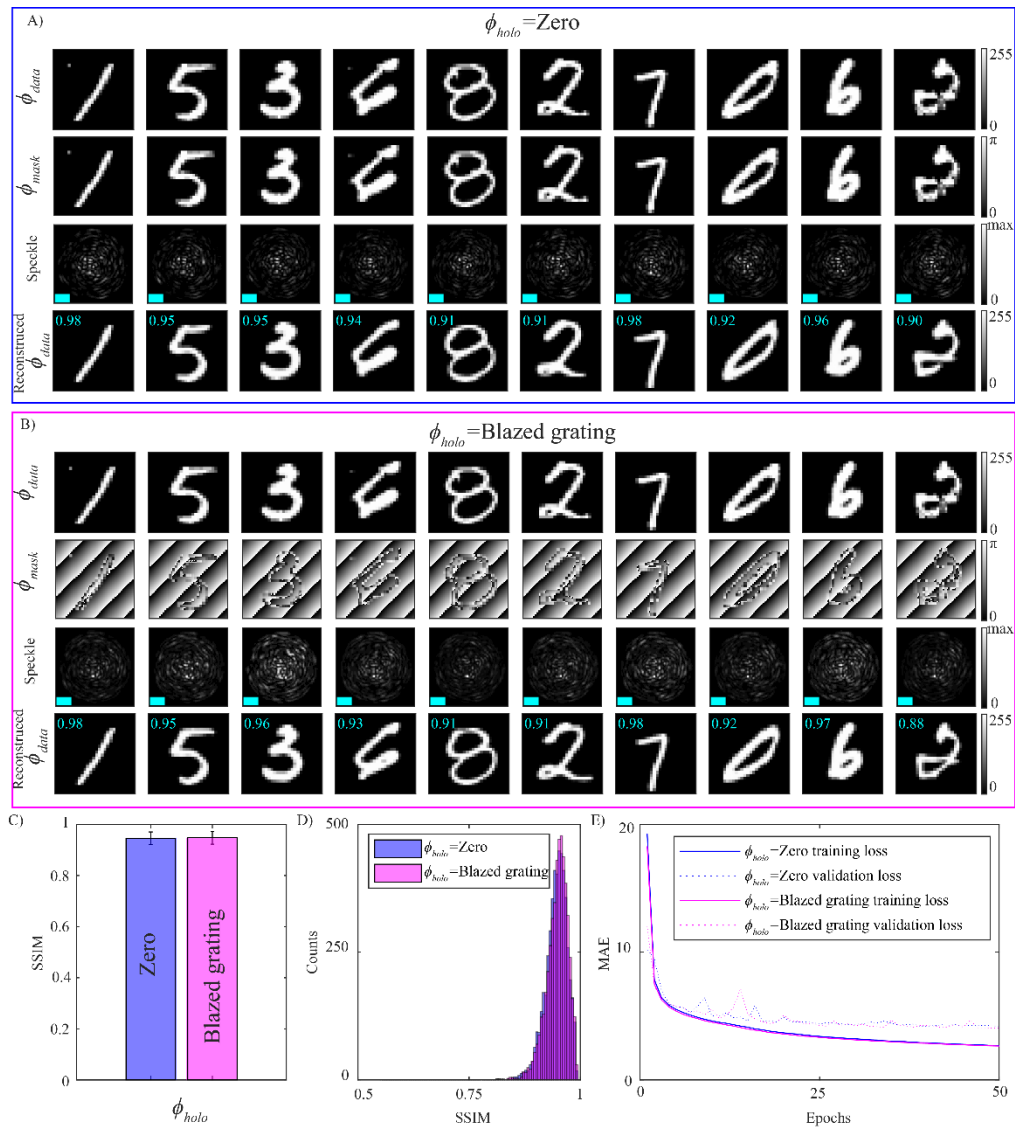


Fig. 3. Reconstruction of grayscale handwritten digits through a MMF, coupled with either $\phi_{holo} = \text{zero}$ or a blazed grating. A) The top row are example handwritten digits to be transmitted through the fiber (ϕ_{data}), the second row shows ϕ_{mask} which is shown on the SLM and in this case is equal to ϕ_{data} the third row shows the resultant output speckle pattern (scale bar is 10 μm) and the final row shows the CNN reconstruction of ϕ_{data} with the SSIM inset. B) The top row are example handwritten digits to be transmitted through the fiber (ϕ_{data}), the second row shows ϕ_{mask} the resultant output speckle pattern (scale bar is 10 μm) and the final row shows the CNN reconstruction of ϕ_{data} with the SSIM inset C) Bar charts showing the average SSIM for the test data set for each coupling D) Histograms of the SSIM for the test data set for each coupling. E) Learning curve for the development of the network for each coupling.

blazed grating are shown in Fig. 4(A), (B). As before, the top row shows ϕ_{data} the second shows ϕ_{mask} the third shows the output speckle patterns and the final shows the reconstruction of ϕ_{data} based on the output speckle. For the test data the SSIM between the input image and reconstructed was 0.67 ± 0.11 for $\phi_{mask} = 0$ coupled data and 0.67 ± 0.11 when ϕ_{mask} is equal to a blazed grating (Fig. 4(C), (D)). Such values are typical for the reconstruction of CIFAR data or natural scenes through an optical fiber using ResUNet type architecture [32]. Again, the result was analysed in terms of the training loss over the network construction and it was determined that for either natural scenes or handwritten digits, the quality of the reconstruction is unaffected by the coupling of the data in the Fourier plane of the SLM. Overall, the above results demonstrate that holographic labelling does not impact the reconstruction fidelity of the neural network. In the following we aim at exploring potential applications of the technique.

3.3. Holographic modulation enables RGB image transmission

We now consider how the holographic label may be used to encode additional information in the data transmission. We associated three different ϕ_{holo} to red, green and blue (RGB) channels of a color image. Firstly, the image was segmented into the single RGB components (ϕ_R, ϕ_G, ϕ_B) and each was added to their a different holographic grating $\phi_{holo,R}, \phi_{holo,G}$ and $\phi_{holo,B}$, selected so that the position of the projection were evenly displaced along an arc of the fiber core. That is, if $\phi_{color} = \phi_R + \phi_G + \phi_B$, then the color image may be coupled through our MMF by sequentially showing $\arg(\exp(i(\phi_R + \phi_{holo,R})))$, $\arg(\exp(i(\phi_G + \phi_{holo,G})))$ and $\arg(\exp(i(\phi_B + \phi_{holo,B})))$ on the screen of the SLM and recording the output speckle on CCD2. This process was repeated for 25000 colored CIFAR images (75000 distinct phase masks). Of which, 22500 colored images were used to build the deep neural network and 2500 to validate. As described in the methods, we chose to train a ResUNet type architecture illustrated in Fig. 5(A) to reconstruct a RGB image. The loss function is defined on the concatenated RGB image as well as each component as follows:

$$L = \frac{\lambda_1}{N} \left(\sum |\phi_R - \widehat{\phi}_R| + \sum |\phi_G - \widehat{\phi}_G| + \sum |\phi_B - \widehat{\phi}_B| \right) + \frac{\lambda_2}{3N} \sum |\phi_{RGB} - \widehat{\phi}_{RGB}|$$

Where $\widehat{\phi}_R, \widehat{\phi}_G, \widehat{\phi}_B$ and $\widehat{\phi}_{RGB}$ refer to the reconstructions of ϕ_R, ϕ_G, ϕ_B and ϕ_{RGB} and $\lambda_1 = 1$ and $\lambda_2 = 0.1$ and N is the number of pixels in the image.

The resultant reconstruction is shown in Fig. 5(B): the top box shows the original images, the second box displays the speckle patterns from each of the 3 holographic channels and the third shows the reconstructed image. The reconstructed color images are extremely similar to the originals and this appears to be independent of color/intensity information, as quantified by measuring the SSIM between original and reconstructed images. CIFAR images are classified as images of airplanes, cars, birds, cats, deer, dogs, frogs, horses, ships, and trucks and the resultant SSIM of the reconstruction for each class is shown in Fig. 5(C) and the histogram across all classes in Fig. 5(D). From the plot, we determine that the reconstruction quality is independent of image class. Across all classes in the validation data, the average SSIM was 0.77 ± 0.09 confirming that our technique can reconstruct color images with high fidelity. The training curves are shown in Fig. 5(E), demonstrating that our network was optimized for image reconstruction.

To test that the data can be clustered by holographic label as in Fig. 2, PCA was applied to the dataset. The holographic variance of the speckle patterns is clearly visible in the scatter plot of PC2 and PC3 shown in Fig. 5(F), confirming that the data is perfectly clustered by holographic label. We remark that PC1 represented an intensity variance between the green and red/blue channels, therefore, the highest three principal components account for variation in holographic label only, as confirmed by the inflection point of the cumulative variance explained data in the inset of Fig. 5(F).

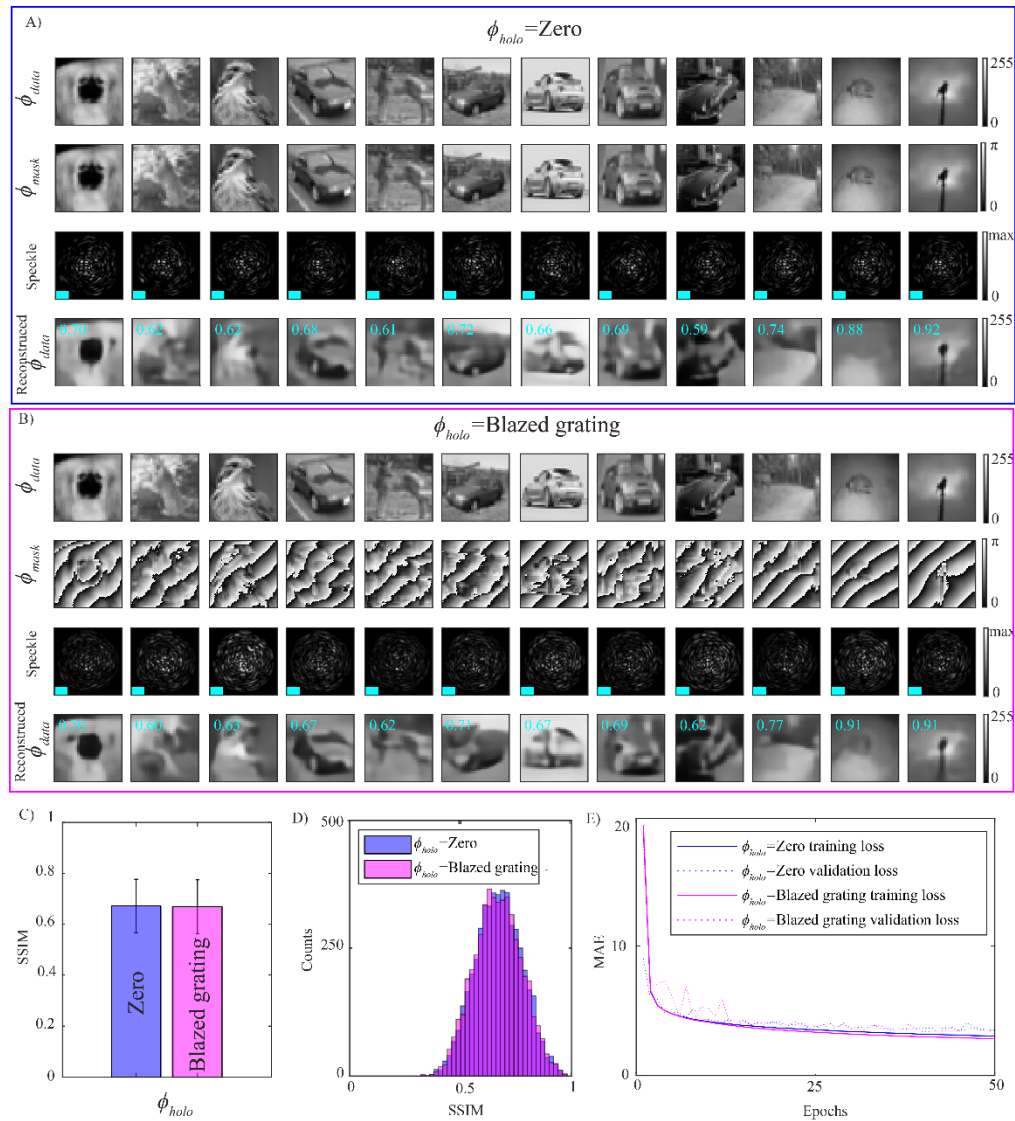


Fig. 4. Reconstruction of grayscale natural scenes (CIFAR) coupled with either $\phi_{holo} = \text{zero}$ or a blazed grating A) The top row are example handwritten digits shown on the SLM and transmitted through the fiber (ϕ_{data}), the second row shows ϕ_{mask} which in this case is equal to ϕ_{data} the third row shows the resultant output speckle pattern (scale bar is 10 μm) and the final row shows reconstruction of ϕ_{data} with the SSIM inset. B) The top row are example handwritten digits shown on the SLM and transmitted through the fiber (ϕ_{data}), the second row shows ϕ_{mask} the third row shows the resultant output speckle pattern (scale bar is 10 μm) and the final row shows reconstruction of ϕ_{data} with the SSIM inset C) Bar charts showing the average SSIM for the test data set for each coupling D) Histograms of the SSIM for the test data set for each coupling. E) Learning curve for the development of the network for each coupling.

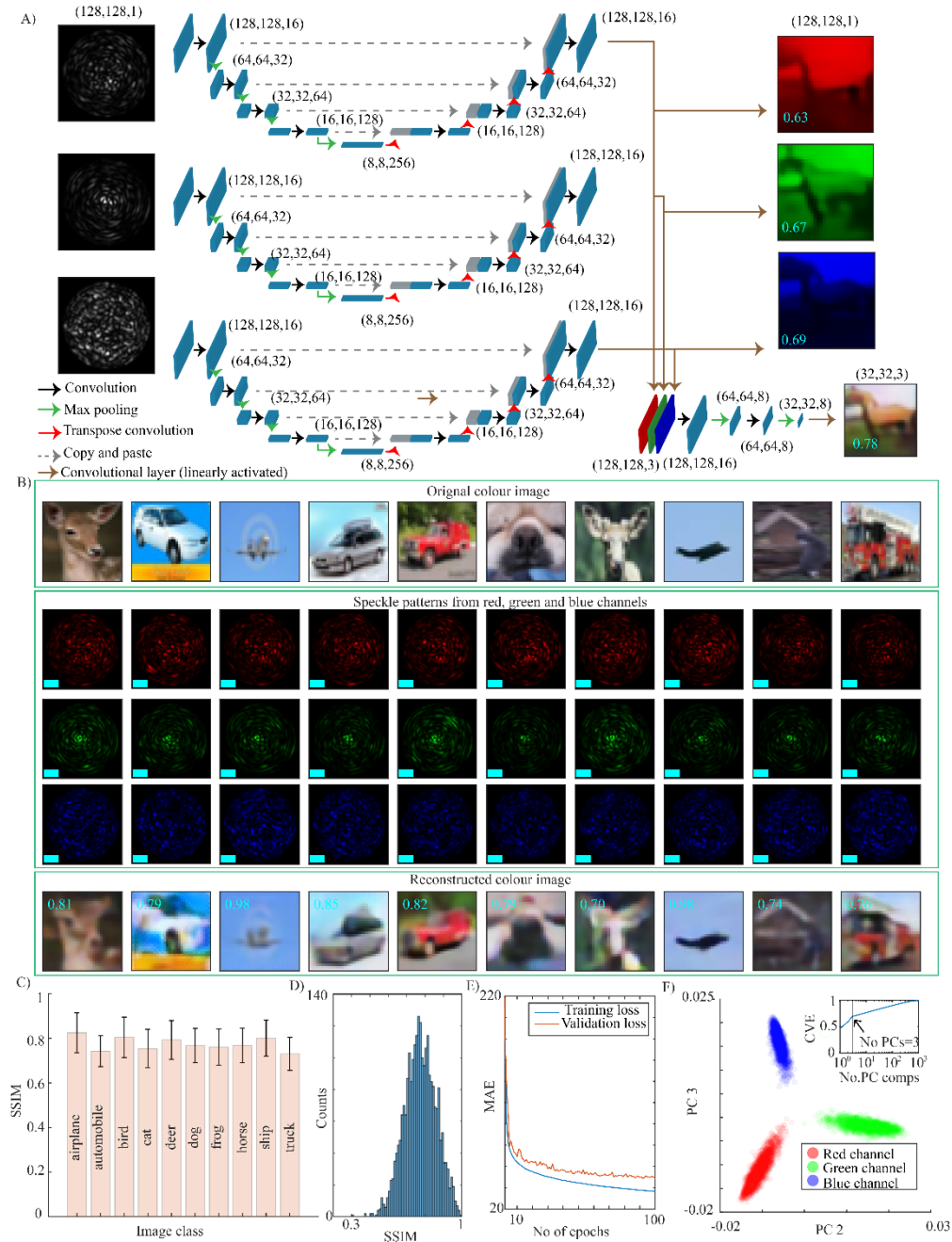


Fig. 5. A) Structure of the implemented ResUNet type color image reconstruction convolutional neural networks. B) Reconstruction of colored CIFAR images through multimode fiber original image (top box), corresponding three speckle patterns (middle box) and reconstructed image (bottom box) with SSIM value inset. C) Bar chart showing average SSIM for each image class (test data) D) Histogram of SSIM (test data). E) Curves showing the training and validation loss over 100 epochs, the epoch achieving the minimal validation loss was used for the reconstruction shown here. F) Scatter plot of second and third PCA component for all 75000 speckle patterns, the data is perfectly clustered by holographic channel, the inset shows the cumulative variance explained by each principal component with the inflection at no of components =3 marked.

Finally, we aim to assess if holographic labelling assists in overall performance of reconstruction of RGB images. To test this, we transmitted the same CIFAR dataset using the conventional technique (with only a temporal label). Example reconstructions, speckle patterns and training curves are shown in [Supplement 1](#). The average SSIM was 0.78 ± 0.09 and the training curves follow that in [Fig. 5](#). Therefore, the label represents an additional information transmitted through the fiber without affecting the overall reconstruction capability of the system. This is evident by the strong clustering of holographic label observed in [Fig. 5\(F\)](#) that cannot be achieved for the temporally labelled data ([Supplement 1](#)).

4. Discussion and conclusion

In this work we show that the optical superimposition of additional variance in phase image transmission through MMF can enrich the wealth of information encoded in the speckle patterns at the fiber output. For this, we propose a modulation technique based on the introduction of additive holograms that generate artificial variance that supersedes the variance of the unmodulated speckle dataset. We demonstrate that this may be done with no loss of reconstruction fidelity by the CNN UNet employed to extract the original image from the speckles pattern. The holograms act as a label for the transmitted data, with multivariate analysis confirming each holographic labels clusters independently. This has the added advantage over a solely temporal label as the label may be reconstructed by considering only the speckle pattern and there is no requirement for temporal synchronization between sender and receiver. This concept is detailed in [Supplement 1](#). Supplementary Figure S4 describes how holographic labelling can allow dynamic reconfiguration of the bandwidth associated to multiple receivers, without any requirements for the receivers and the senders to agree on a new temporal order of the transmitted packages. [Supplement 1](#) instead depicts how the transmission of holographically-labeled streams of RGB data requires no synchronization pulse to associate the received image to being red, green or blue, therefore increasing the actual transmission rate.

As an example application, we show how an RGB image may be transmitted through three holographically-labelled channels. The RGB image can be reconstructed without the need for a temporal synchronization between the detector and the SLM (as in [\[26\]](#)) after the initial training step. We have used a blazed grating to induce holographic variance, however potentially any kind of holographic carrier may be used, provided it can generate greater variance in the output speckle patterns than the one contained in the dataset itself. We note that the average SSIM for the reconstructed color images is significantly higher than that of the grayscale images. In [Fig. 5](#) the average SSIM for colored images was 0.77 ± 0.09 whereas was 0.65 ± 0.11 , 0.65 ± 0.11 and 0.65 ± 0.11 for each R, G and B component. Tentatively, it could be hypothesized that the increase in SSIM is due to a data fusion effect across all channels. Between each RGB component, there exists strong correlation and as the loss function is a linear combination of each component and the concatenated image the network is able to learn this aspect of the structure within the CIFAR dataset. This is also the case for data that is only labelled temporally ([Supplement 1](#)).

Overall, a limitation of our technique and other SLM/MMF optical systems [\[42\]](#) is the speed of transmission modulation (over data transmission through a SMF/MMF) which is determined by the the refresh rate of the SLM, being on the ms scale). This results in a relatively low bandwidth when compared with other optical communication systems. Nevertheless, substituting the SLM for a micro LED arrays which can achieve refresh rates of 100 MHz would present an opportunity to increase frame rate [\[43\]](#).

Finally, we remark that machine learning based techniques are a straightforward technique to achieve image reconstruction through dynamically bent [\[31\]](#) and thermally perturbed [\[33\]](#) multimode fibers, going forward we see our technique as complementary to these methods. Potentially, RGB images could be transmitted through dynamically perturbed fibers without

requirement for temporal synchronicity. As well as this the technique may reveal more broad insights into holographic control of modal propagation in multimode optical fibers.

Funding. Robotics and AI for Socio-economic Empowerment (ECS00000035); National Institutes of Health (1UF1NS108177-01, U01NS094190); Horizon 2020 Framework Programme (101016787, 677683, 692943, 828972, 966674).

Acknowledgments. L.C., M.D.V. and F.P. acknowledge funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement No. 828972. L.C., M.D.V. and F.P. acknowledge funding from the Project "RAISE (Robotics and AI for Socio-economic Empowerment)" code ECS00000035 funded by European Union – NextGenerationEU PNRR MUR - M4C2 – Investimento 1.5 - Avviso "Ecosistemi dell'Innovazione" CUP J33C22001220001. M.D.V. and F.P. acknowledge funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement No 101016787. M.D.V. and F.P. acknowledge funding from the U.S. National Institutes of Health (Grant No. 1UF1NS108177-01). M.D.V. and F.P. acknowledge funding from European Research Council under the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement No. 966674. M.D.V. acknowledges funding from the European Research Council under the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement No. 692943. M.D.V. acknowledges funding from the U.S. National Institutes of Health (Grant No. U01NS094190). F.P. acknowledges funding from the European Research Council under the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement No. 677683.

Disclosures. M.D.V. and F. Pisanello are founders and hold private equity in Optogenix, a company that develops, produces and sells technologies to deliver light into the brain. MDV: Optogenix srl (I). FP: Optogenix srl (I).

Data availability. Data underlying the results presented in this paper are available at Ref. [44].

Supplemental document. See [Supplement 1](#) for supporting content.

References

1. I. M. Vellekoop and A. P. Mosk, "Focusing coherent light through opaque strongly scattering media," *Opt. Lett.* **32**(16), 2309–2311 (2007).
2. R. Di Leonardo and S. Bianchi, "Hologram transmission through multi-mode optical fibers," *Opt. Express* **19**(1), 247–254 (2011).
3. T. Čižmár and K. Dholakia, "Shaping the light transmission through a multimode optical fibre: complex transformation analysis and applications in biophotonics," *Opt. Express* **19**(20), 18871–18884 (2011).
4. I. N. Papadopoulos, S. Farahi, C. Moser, *et al.*, "Focusing and scanning light through a multimode optical fiber using digital phase conjugation," *Opt. Express* **20**(10), 10583–10590 (2012).
5. L. Collard, F. Pisano, M. Pisanello, *et al.*, "Wavefront engineering for controlled structuring of far-field intensity and phase patterns from multimodal optical fibers," *APL Photonics* **6**(5), 51301 (2021).
6. S. H. Li, C. Saunders, D. J. Lum, *et al.*, "Compressively sampling the optical transmission matrix of a multimode fibre," *Light. Appl.* **10**, 88 (2021).
7. A. D. Gomes, S. Turtaev, Y. Du, *et al.*, "Near perfect focusing through multimode fibres," *Opt. Express* **30**(7), 10645–10663 (2022).
8. E. E. Morales-Delgado, S. Farahi, I. N. Papadopoulos, *et al.*, "Delivery of focused short pulses through a multimode fiber," *Opt. Express* **23**(7), 9109–9120 (2015).
9. T. Čižmár and K. Dholakia, "Exploiting multimode waveguides for pure fibre-based imaging," *Nat. Commun.* **3**(1), 1027 (2012).
10. S. Li, S. A. R. Horsley, T. Tyc, *et al.*, "Memory effect assisted imaging through multimode optical fibres," *Nat. Commun.* **12**(1), 3751 (2021).
11. L. Collard, L. Piscopo, F. Pisano, *et al.*, "Optimizing the internal phase reference to shape the output of a multimode optical fiber," *PLoS One* **18**(9), e0290300 (2023).
12. S. A. Vasquez-Lopez, R. Turcotte, V. Koren, *et al.*, "Subcellular spatial resolution achieved for deep-brain imaging in vivo using a minimally invasive multimode fiber," *Light: Sci. Appl.* **7**(1), 110 (2018).
13. S. Turtaev, I. T. Leite, T. Altwegg-Boussac, *et al.*, "High-fidelity multimode fibre-based endoscopy for deep brain in vivo imaging," *Light: Sci. Appl.* **7**(1), 92 (2018).
14. I. Gusachenko, M. Chen, and K. Dholakia, "Raman imaging through a single multimode fibre," *Opt. Express* **25**(12), 13782–13798 (2017).
15. J. Trägårdh, T. Pikálek, M. Šerý, *et al.*, "Label-free CARS microscopy through a multimode fiber endoscope," *Opt. Express* **27**(21), 30055–30066 (2019).
16. A. Cifuentes, T. Pikálek, P. Ondráčková, *et al.*, "Polarization-resolved second-harmonic generation imaging through a multimode fiber," *Optica* **8**(8), 1065–1074 (2021).
17. L. V. Amitonova and J. F. de Boer, "Endo-microscopy beyond the Abbe and Nyquist limits," *Light: Sci. Appl.* **9**(1), 81 (2020).
18. S. Daan, D. B. Phillips, S. Peter, *et al.*, "Time-of-flight 3D imaging through multimode optical fibers," *Science*. **374**(6573), 1395–1399 (2021).

19. I. T. Leite, S. Turtaev, X. Jiang, *et al.*, “Three-dimensional holographic optical manipulation through a high-numerical-aperture soft-glass multimode fibre,” *Nat. Photonics* **12**(1), 33–39 (2018).
20. L. Collard, F. Pisano, D. Zheng, *et al.*, “Holographic manipulation of nanostructured fiber optics enables spatially-resolved, reconfigurable optical control of plasmonic local field enhancement and SERS,” *Small* **18**(23), 2200975 (2022).
21. Y. Choi, C. Yoon, M. Kim, *et al.*, “Scanner-free and wide-field endoscopic imaging by using a single multimode optical fiber,” *Phys. Rev. Lett.* **109**(20), 203901 (2012).
22. S.-Y. Lee, V. J. Parot, B. E. Bouma, *et al.*, “Confocal 3D reflectance imaging through multimode fiber without wavefront shaping,” *Optica* **9**(1), 112–120 (2022).
23. L. V. Amitonova and J. F. de Boer, “Compressive imaging through a multimode fiber,” *Opt. Lett.* **43**(21), 5427–5430 (2018).
24. N. Borhani, E. Kakkava, C. Moser, *et al.*, “Learning to see through multimode fibers,” *Optica* **5**(8), 960–966 (2018).
25. B. Rahmani, D. Loterie, G. Konstantinou, *et al.*, “Multimode optical fiber transmission with a deep learning network,” *Light: Sci. Appl.* **7**(1), 69 (2018).
26. P. Caramazza, O. Moran, R. Murray-Smith, *et al.*, “Transmission of natural scene images through a multimode fibre,” *Nat. Commun.* **10**(1), 2029 (2019).
27. B. Song, C. Jin, J. Wu, *et al.*, “Deep learning image transmission through a multimode fiber based on a small training dataset,” *Opt. Express* **30**(4), 5657–5672 (2022).
28. C. Zhu, E. A. Chan, Y. Wang, *et al.*, “Image reconstruction through a multimode fiber with a simple neural network architecture,” *Sci. Rep.* **11**(1), 896 (2021).
29. P. Fan, T. Zhao, and L. Su, “Deep learning the high variability and randomness inside multimode fibers,” *Opt. Express* **27**(15), 20241–20258 (2019).
30. J. Zhao, X. Ji, M. Zhang, *et al.*, “High-fidelity imaging through multimode fibers via deep learning,” *J. Phys. Photonics* **3**(1), 015003 (2021).
31. S. Resisi, S. M. Popoff, and Y. Bromberg, “Image transmission through a dynamically perturbed multimode fiber by deep learning,” *Laser Photon. Rev.* **15**(10), 2000553 (2021).
32. R. Xu, L. Zhang, Z. Chen, *et al.*, “High accuracy transmission and recognition of complex images through multimode fibers using deep learning,” *Laser Photon. Rev.* **17**(1), 2200339 (2023).
33. N. Bagley, T. Kremp, E. S. Lamb, *et al.*, “Transfer learning and generalization of a neural-network-based multimode fiber position and imaging sensor under thermal perturbations,” *Opt. Fiber Technol.* **70**, 102855 (2022).
34. L. Wang, T. Qi, Z. Liu, *et al.*, “Complex pattern transmission through multimode fiber under diverse light sources,” *APL Photonics* **7**(10), 106104 (2022).
35. B. Rahmani, D. Loterie, E. Kakkava, *et al.*, “Actor neural networks for the robust control of partially measured nonlinear systems showcased for image propagation through diffuse media,” *Nat. Mach. Intell.* **2**(7), 403–410 (2020).
36. F. I. Diakogiannis, F. Waldner, P. Caccetta, *et al.*, “ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data,” *ISPRS J. Photogramm. Remote Sens.* **162**, 94–114 (2020).
37. O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18* (Springer, 2015), pp. 234–241.
38. K. He, X. Zhang, S. Ren, *et al.*, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016), pp. 770–778.
39. Y. Bengio, P. Simard, and P. Frasconi, “Learning long-term dependencies with gradient descent is difficult,” *IEEE Trans. Neural Netw.* **5**(2), 157–166 (1994).
40. K. He and J. Sun, “Convolutional neural networks at constrained time cost,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2015), pp. 5353–5360.
41. Y. Li, S. Cheng, Y. Xue, *et al.*, “Displacement-agnostic coherent imaging through scatter with an interpretable deep neural network,” *Opt. Express* **29**(2), 2244–2257 (2021).
42. Z. Li, W. Zhou, Z. Zhou, *et al.*, “Self-supervised dynamic learning for long-term high-fidelity image transmission through unstabilized diffusive media,” *Nat. Commun.* **15**(1), 1498 (2024).
43. W. Zhao, H. Chen, Y. Yuan, *et al.*, “Ultra-high-speed color imaging with single-pixel detectors at low light level,” *Phys. Rev. Applied* **12**(3), 034049 (2019).
44. L. Collard, M. Kazemzadeh, L. Piscopo, *et al.*, “Optical fiber speckles holographic labelling,” <https://www.kaggle.com/dsv/7904588>, 10.34740/kaggle/dsv/7904588 (2024).