ARTICLE

Actor neural networks for the robust control of partially measured non-linear systems showcased for image propagation through diffuse media

Babak Rahmani[†1], Damien Loterie[2], Eirini Kakkava[3], Navid Borhani[4], Uğur Teğin[5], Demetri Psaltis[6] and Christophe Moser[7]

Abstract

The output of physical systems such as the scrambled pattern formed by shining the spot of a laser pointer through the fog is often easily accessible by direct measurements. The selection of the input of such a system to obtain a desired output on the other hand is difficult because it is an ill-posed problem i.e. there are multiple inputs yielding the same output. Information transmission through scattering media is an example of this problem. Machine learning approaches for imaging have been very successfully implemented in photonics to recover the original input phase and amplitude objects of the scattering system from the distorted intensity diffraction pattern outputs. However, controlling the output of such a system, without having examples of inputs that can produce outputs in the class of the output objects the user wants to produce, is a challenging problem. In this paper, we propose an online learning approach for projection of arbitrary shapes through a multimode fiber (MMF) when a sample of intensity-only measurements is taken at the output. This projection system is nonlinear because the intensity, not the amplitude is detected. We show image projection fidelity as high as ~90 %, which is on par with the gold standard methods which characterize the system fully by phase and amplitude measurements. The generality as well as simplicity of the proposed approach could potentially provide a new way of target-oriented control in real-world applications when only partial measurements are available.

INTRODUCTION

Tailoring a physical system to behave in the way that leads to a desired outcome entails careful characterization of the system. The situation becomes worse when the physical system is difficult to model. Even when a model is available, finding the appropriate input that would experimentally result in the desired output is usually not straightforward. Light propagation in scattering media, or through a network of on-chip waveguides are examples of such systems[1-18]. In these situations, the system needs to be characterized by closely monitoring the system's response to a series of inputs and then inversely solving for the system's *response*

[1] Ecole Polytechnique Fédérale de Lausanne, Laboratory of Applied Photonics Devices, CH-1015 Lausanne, Switzerland.

[2] Ecole Polytechnique Fédérale de Lausanne, Laboratory of Optics, CH-1015 Lausanne, Switzerland

†Correspondence: Babak Rahmani, E-mail: babak.rahmani@epfl.ch

* The paper has been accepted for publication Nature Machine Intelligence. DOI: https://doi.org/10.1038/s42256-020-0199-9

function H. Often when outputs have a linear relationship with their corresponding inputs and more importantly are fully measured, it is possible to establish a model to relate the inputs to the outputs. Even when a forward response function of a system is found, calculating its inverse H^{-1} could still be challenging because of the large size of H. Fully measuring the outputs of a system requires nontrivial sensory apparatus. For example, the electromagnetic fields satisfying Maxwell equations at the output of a multimode waveguide are complex-valued fields having both amplitude and phase information. Measuring both requires phase extraction using interferometry-based schemes[3]. Being sensitive to environmental perturbation, the phase information needs to be carefully traced and corrected for while characterizing the system. The conventional methods cannot provide an optimal solution for a desired output of the system if no phase information is provided.

Neural networks perform well in solving ill-posed inverse problems[19, 20] for various applications such as biology[21], design of photonic devices[22] or novel drugs[23]. Here, we propose a neural-network based approach to solve the highly ill-posed problem of predicting a scattering medium system's *forward and backward* response functions. To this end, we only need a signal modulator to shape the control signal entering the system as well as an intensity detection apparatus (e.g. camera) to measure the output signal. We show that it is possible to recover the sought-after functions even using the most basic sensory devices that only detect a portion of the outputs' information. In photonics, one can readily envision using the proposed method for light propagation control in spatial and\or spectral domains. In particular, we showcase the success of the approach for the scenario of imaging through MMFs: an extreme case for which thousands of modes of the fiber are controlled in order to transmit tens of thousands of pixels though the fiber to project a user defined image. To the best of our knowledge, this is the first demonstration of controlling a nonlinear system with a fidelity on a par with the case for which all parameters of the system are measured (amplitude and phase).

RESULTS

Image transmission through Multimode fibers. The energy splitting among different modes of the multimode fiber (MMF) gives rise to a distorted field at the output and creating speckle patterns[1]. To undo modal scrambling inside the MMF, several approaches ranging from digital iterative algorithms[2-6] to analog\digital phase conjugation[7-13] as well as interferometry methods have been proposed to compensate the modal dispersion in MMFs[14-18]. The former requires calibrating the system for each image to project, which is too computationally intensive and thus too slow for practical applications while the latter entails measuring the complex output field (both phase and amplitude), which is cumbersome to implement. Here, we seek an approach that can be implemented with a simple optical setup, i.e. an input modulator together with an amplitude-only output detection (neglecting phase information), and yet has the generality of approaches that measure both phase and amplitude information. The schematic of such a system is depicted in Fig. 1. Neural networks have previously been shown to reconstruct the undistorted input fields of scattering media including MMFs[24-28] from the amplitude-only scrambled speckle patterns at its output for fibers. Here we intend to accomplish the reverse: to learn the correct inputs that will generate a desired output of the MMF. This is challenging because first and foremost, no straightforward way (without

using the transmission matrix) exists for acquiring a training set with the desired output patterns to train the neural network. The training set for this task consists of examples of the desired shapes (for example recognizable drawings such as smiley face, alphabet letters, etc.) on the distal end of the fiber (camera side) and examples of the corresponding inputs that generate those patterns when sent through the fiber. Unfortunately, this is exactly the problem we intend to solve. If we had the ability to generate those examples, the problem would already be solved, making it a chicken and egg dilemma. In this regard, an initial attempt towards projecting patterns through MMFs using neural networks has been made that is limited only to focusing spots[29] after the fiber.

Learning algorithm. We use a combination of neural networks (referred together as *the projector network*) to generate input control signals that create the desired target output on the detector. Specifically, our projector network, schematically depicted in Fig. 2a, is made of two sub-networks, the *Actor* and the *Model*. The Model sub-network tries to learn the forward propagation path of the light through the system (input to output) and the Actor learns the inverse path. In other words, the Actor generates proper inputs constrained by the physical propagation rules of the system embedded in the network as the Model. The two networks are trained synergistically so that the Model network forces the Actor to generate control patterns that, upon sending through the fiber, produce images on the detector belonging to the desired set. Training the Model to emulate the forward path of the light propagation in fact allows *in situ* back propagation of the error between the desired target images and what appears on the detector in the real experiment through the virtual fiber (i.e. the Model network). When the back propagated error reaches the Actor, the learnable parameters of the Actor sub-network is adjusted so as to the error is reduced. The error is the smallest when the Actor is the inverse of the Model network. Therefore, this training method, which has also been proposed for use in control systems nearly three decades ago[30], is effectively trying to find the inverse path of light propagation or analogously the time reversal operator. Some recent works in adaptive optics have benefited from synergistic combinations of networks to obtain the system's optimal input[31]. These systems usually enjoy a *A-priori* well-established relationship between their inputs and outputs insofar as the extra networks can be used to correct the environmental variations perturbing the system.

The architecture of the projector network, on the other hand, is *reminiscent* of the so-called Actor-critic neural networks used in Reinforcement Learning[32]. The similarity is rooted from the fact that the training of the two networks is carried out in interaction with each other so that the performances of both networks gets better over time. *A-priori* no direct label (ground truth control patterns) for target output images exists. Therefore, the Actor uses the available control modules and sends random signals through the Model. The Model network evaluates the patterns produced by the Actor and sends feedback to the Actor to teach it to produce better control patterns. In the meantime, the Model network also improves its own performance. The training procedure is explained more formally in the Materials and Methods section.

Image projection. We first train our network with grayscale images of handwritten Latin alphabet characters from EMNIST[33] as targets (see data preparation subsection in Materials and Methods for more information). Fig. 3 depicts examples of these images and the physical outputs on the camera projected using the proposed learning algorithm. For the sake of comparison, target images are also projected with the transmission matrix approach, an example of methods that require full complex field measurement and control of phase information. In each image, the inset indicates the projection fidelity according to the Pearson coefficient

defined in equation 4 in the Methods section. Note that the fidelity is calculated over the entire image area (both zero and non-zero pixels).

Without any fine tuning, the network trained with Latin characters is used directly to project a different category of images. Examples of projected images are shown in Fig. 4, Extended Data Figure 1 and Supplementary Video 1. These results demonstrate the generalization ability of the projector neural network and show that it can extend its ill-posed inverse problem ability to images never seen by the network in the training step.

Table I summarizes the projection fidelity of the neural network approach and that of the transmission matrix approach for various types of images.

The performance of the neural network in inferring the required SLM modulations is correlated with the complexity of the target images that the network is trained with. The Latin alphabet images, used for training the network in the first experiment, are of sparse nature: having a constant zero background and a grayscale feature centered in the middle of the image. Therefore, it is expected that projecting target images with richer contexts will be more challenging.

We explored this, by using our approach to project continuous gray-scale natural-scene-like pictures. Examples of which are shown in Fig. 5. The projected images (red, green, blue, 3-channel RGB and the superposition of all three as one channel) are also depicted. The complexity in the target images makes the training difficult; however, our method is able to provide the appropriate SLM patterns for projecting images of natural scenes with fidelities on par with that of full-measurement schemes.

DISCUSSION

Learning trajectory. The fidelity plot in Extended Data Figure 2a demonstrates the algorithm's convergence speed in finding an appropriate solution for the system. The convergence speed depends on the complexity of the target images, the modulation scheme and the extent by which this modulation can be implemented via the modulator, and finally the rate at which the system changes over time. For example, instead of complex-value solutions, the network can be constrained to find solutions that are amplitude-only. The former type of solutions is better suited for SLMs and the latter is used more conveniently with amplitude-only modulators (such as digital micromirror devices). It can be shown that the number of iterations required to achieve certain fidelity is higher when an amplitude-only solution is favored (5 iterations versus 1 to 2 in the complex-value case). Refer to the Extended Data Figure 3 for comparing the quality of the projected patterns for amplitude-only inputs. Still, for complex-value solutions, an intermediate conversion scheme is required to convert them into phase-only solutions compatible with phase-only SLMs (see methods).

Robustness. The system is also prone to multiple time-dependent processes including mechanical perturbations, instabilities associated with drifts in power and working wavelength of the laser source, among others, which influence the learning trajectory. In a first step to investigate the robustness of the projector algorithm, we use the measured transmission matrix of the system (measured once) to virtually project the control patterns provided by our algorithm. Doing this, we are able to obtain the resulting fiber's outputs without sending them directly through the fiber. The projector network is then trained as before with this new dataset. After training, the network is run to produce the SLM patterns that correspond to a user defined output. The patterns are then loaded on the SLM of the experimental system. The fidelity trajectory of the projected images is shown in Extended Data Figure 2c (solid lines) for three colors (Red, Green, Blue). It can be inferred from the plots that the fidelities of the projected images converge to slightly higher values than those of the experimentally projected images (solid circles). The lower fidelity of the

latter is because of the degradation due to time variation and non-perfect modulation scheme. The transmission matrix used for projecting the SLM patterns could also be re-measured after each round of training as to bring the system's variation with time into play (dashed lines in Extended Data Figure 2c). Hence although the system is changing over time, it is effectively being corrected. Interestingly, the close overlap between trajectory of graphs in Extended Data Figure 2c (dashed lines) and the experimentally projected images (solid circles) shows that the neural network approach is automatically compensating drifts but without the need to continuously measure and invert the matrix as it is required in the transmission matrix approach. Further analysis of the robustness of the proposed approach in the event of a misalignment in the system is shown in the Methods section. It should be noted that the recovery time of the system could be considerably improved by using a digital micromirror devices with a projection rate of a few KHz rather than an SLM with a rate limited to a maximum of 60 Hz.

Perspective. In this work, we showed the success of our approach in controlling the output of a system whose nonlinearity is due to the nonlinearity at the input and output of the system. In fact, the nonlinearity of the MMF system we consider here originates not only from the intensity measurement on the camera side at the output of the fiber, but also to some degree from the nonlinear response of the spatial light modulator (SLM) at the input of the MMF as well. But generally, a system, as a whole, becomes nonlinear due to the nonlinearity at its output and/or input or due to the non-linearity of the propagation in-between (or both). For example in MMFs with high power input fields, the medium itself shows nonlinear response. Using neural networks with nonlinear elements may be a valid approach to mimic the nonlinear response of the fiber. In fact, in some recent works, it has been shown that neural networks can predict\control some relatively simple nonlinear phenomena such as spectral shape and Raman scattering[34-37]. It remains to be seen if the results in these recent works can be improved by the method presented in this manuscript. This will be further explored in our future works. Here, we focused on cases in which the nonlinearity is at the input and output of the system and with the additional constraint of partial measurements which makes the system extremely ill-posed, a scenario that is common in various fields and applications, and could be of high importance. For example, our approach most promisingly finds applications in large scale systems (large number of modes) for which inverting the forward transfer function of the system is practically impossible for a standalone standard Personal Computer. For example, the size of the forward response function of a MMF system increases with the 4^{th} power of the fiber core size. Inverting such a large scale system is challenging and prone to numerical errors. However, using our approach both the forward and backward responses of the system can be directly calculated. The latter point becomes important for making a 3D display, virtual/augmented reality systems, by projecting patterns with controlled amplitude and phase after diffuse reflection from a surface (walls, glasses). Similarly, the proposed control method could be potentially applied to other disciplines such as robotics where, for example, the 3D position of a robotic arm could be controlled via learning the system's control parameters such as pressure in pneumatic actuators, voltage in piezo or thermal actuators all thanks to the straightforward way of learning the forward and backward responses of the system.

MATERIALS AND METHODS

Experimental set-up: The experimental setup for image transmission through the fiber is depicted in Extended Data Figure 4. Three continuous input beams at wavelengths 488, 532, and 633 nm are delivered one at a time to the system via a single mode fiber. The beam entering the system (attenuated to an average power of 4 mW), is collimated by lens L2 (*f*=100 mm) and then directed to the SLM. The beam spatially modulated by the phase-only SLM (HOLOEYE PLUTO) is imaged on the input facet of a multimode fiber using a 4-f system composed of L3 (*f*=250 mm) and OBJ 1 (60x, NA=0.85). After transmission through the graded-index fiber with length L=75 cm, core diameter D=50 μm and a NA of 0.22 (corresponding to ~1050 fiber modes for one polarization), the output field is imaged onto the camera using an identical 4-f configuration. The experimental setup for measuring the transmission matrix of the system requires the extra reference path shown faded in the schematic. The beam in the reference path is superimposed to the main path's beam on the camera via a series of mirrors and a beam splitter. A number of input patterns (basis vectors) modulated with either phase-, amplitude- or both are then sent through the fiber and their corresponding complex output fields are measured. The transmission matrix is then constructed using these input-outputs[17].

Neural network architecture: The projector network, schematically depicted in Extended Data Figure 5, is composed of two sub-networks: the Actor (A) and the Model (M) which together generate the SLM patterns required to obtain the desired images at the output of the fiber. To allow complex valued modulation of the inputs and outputs of the system, so as to be able to closely mimic the complex physical fields entering the fiber and hence taking advantage of higher degree of freedom in shaping the input fields of the system, the Actor and Model themselves are made of two smaller sub-networks (A_{real}, A_{imag}) and (M_{real}, M_{imag}). The training of the pair A_{imag} and M_{imag} is always carried out separately right after the training of the pair A_{real} and M_{real} in an identical manner; therefore, henceforth we only refer to them collectively as the Actor and Model. Each of the sub-networks have the architecture of a fully-connected neural network in which, the input images are fed to the network via the input layer and passed on to the output layer via weights $w_{i,j}$ connecting node *i* to node *j*. All the incoming connections at node n_j in the output-layer are first summed up and bias corrected via bias b_j and then passed on to the nonlinear unit with the Sigmoid nonlinearity function $\sigma(x) = 1/(1+\exp(-x))$. Thereby, the *j*-th output of the fully-connected layer reads as:

$$n_j = \sigma[\sum_i w_{i,j} n_i + b_j] \tag{1}$$

The training of the two networks is carried out in an alternating fashion. Figure 2b schematically illustrates the training procedure comprised of 3 sub-steps A-C:

Training algorithm: **A-** first the training examples composed of the patterns on the SLM, denoted by *X*, and their corresponding amplitude patterns captured on the camera referred to as *Y* are collected.

B- the Model is trained on examples obtained from sub-step A to emulate the physical forward path from the SLM to the camera. Accordingly, the Model is trained by minimizing the mean squared error between $\hat{Y} = M(X)$ and the ground truth labels *Y*, i.e.

$$MSE(\theta) = \frac{1}{N \times L \times L} \sum_{l=1}^{N} \sum_{i=1}^{L} \sum_{j=1}^{L} | \hat{Y}_{i,j}^{l}(\theta_D) - Y_{i,j}^{l} |^2 \tag{2}$$

where θ_D are weights and biases of the Model, *i* and *j* are the indices of the neural network reconstructed image $\hat{Y}^l$ and the label image Y^l belonging to the *l*-th image pairs, where *l* and *N* are the samples' mini-batch index and size, respectively, and *L* is the width and height of the images.

C- next is the training of the Actor. The inputs to the Actor, denoted by Z, are examples of the target images that we wish to see on the camera. The output of the Actor, $A(Z)$, is passed through the Model. The Model's output, $M(A(Z))$, is then compared against the inputs of the Actor (i.e. Z) using the logarithm of the Pearson coefficient, i.e.

$$-\log[(1+\text{Pearson}(\text{M}(\text{A}(\text{Z})),\text{Z}))/2] \quad (3)$$

where the Pearson coefficient of variables *x* and *y* is defined as:

$$\text{Pearson}(x,y)=\frac{\sigma_{x,y}}{\sigma_x\sigma_y} \quad (4)$$

in which $\sigma_{x,y}$ is the covariance between *x* and *y*, σ_x and σ_y are the standard deviations of *x* and *y*, respectively.

Even though the back propagated error from this comparison reaches the Actor via propagation through the Model, but the weights and biases of the Model are kept constant while training the Actor.

Once the training step is complete, the target image is fed to the Actor to produce the required SLM pattern that generates the desired image on the camera. The training process can be repeated for several iterations to improve the fidelity between the target images and the projected images on the camera. To avoid over-fitting of the fiber model and proper modeling of the physical system, we adopt an early stopping strategy by finding the sufficient number of steps for training the Model network. For the first iteration, this number can be found empirically. From the second iteration and thereafter, the number of steps for which the fiber-model is trained can be decreased since an estimate of the proper SLM inputs for the system is already available.

An adaptive moment estimation optimization (ADAM) algorithm with a learning rate of 10^{-4} and mini-batch size of 32 is used to execute the training on an NVIDIA RTX 2080 Titan GPU.

Semi-supervised learning. We have also investigated different variations of the proposed learning algorithm. In particular, in one scenario the network is trained in a *semi-supervised* manner, namely the images that are used to train the Model network are also used to train the Actor network. For this, we modified the loss function of the actor network and added an extra mean-squared term with weight λ to (3) so that the Actor network also learns to predict the input SLM patterns of the output images that were used for training of the Model (the reverse of the Model). We compare the performances of the variations of the proposed approaches with that of the original algorithm in Extended Data Figure 6. Evidently the alternative approaches acquire slightly higher fidelities in shorter amount of time while facing over-fitting and a degradation of performance in the long run. The original proposed approach slightly lags behind but keeps the same performance asymptotically. The latter is due to the fact that the forward model network works as a means of regularizing for the Actor to avoid over-fitting.

Data preparation: The input and ground truth data for training the sub-network Model consist of images of sizes 51×51 and 200×200, respectively. The former, referred to as SLM images are uploaded to the SLM and sent through the fiber and captured on the camera. After taking the square root of camera images (shifting from intensity to amplitude images), a rectangular area of size 200×200 pixels (corresponding to an area of size 19×19 μm^2 on the output facet of the fiber) are cropped from the fiber's imaged facet. These images constitute the ground truth for training the Model. The training of the network for projection of EMNIST images is carried out with 20k examples.

For cases in which the fidelity of the projected images after the first iteration of training is not optimal, another round of training could substantially increase the quality of images. As indicated in the description of the learning algorithm, the

training dataset for the second round is comprised of the predicted solutions found by the network for the set of target images as well as what those solutions produce experimentally. However, when the target images are few in number, the training dataset is accordingly small. In this case, we can simply augment the dataset by adding noise to the found solutions from the first iteration (inputs) and measure the corresponding outputs of the system for the augmented inputs. Doing this allows one to increase the number of training dataset arbitrarily. We followed the same procedure for projecting 4 images in Fig. 5 using a Gaussian noise with a zero mean and a standard deviation of 0.5.

Robustness analysis: We further show the robustness of the proposed approach in two additional scenarios in which the system gets perturbed because of a misalignment in the system. In the first scenario, we assume the system has encountered a minor perturbation: the proximal side of the fiber (the one on the side of the SLM) has been slightly misaligned so that near 20% of its facet area is blocked. We apply this by means of inserting a spatial filter in our system depicted in Extended Data Figure 7a. The filter has value 1 in the yellow parts and 0 elsewhere. Despite this misalignment, we show that the network is able to compensate this loss. The fidelity plot of the projected images before and after the perturbation event is shown in Extended Data Figure 7b. It can be seen that the system converged before the perturbation. Once the system is perturbed, at iteration 11 (marked by a star symbol), the fidelity goes down. Yet, the decrease in the fidelity is compensated and recovered when the neural network finds new pathways after a few more iterations. In a second scenario, we assume the fiber facet has changed direction (the entire facet is tilted in orientation). This translates into a shift in the input bandwidth of the fiber which is schematically shown in Extended Data Figure 7c. It should be noted that this is a severe change in the system because some of the input spatial frequencies are now disturbed\changed. However, this does not decrease the capacity of the system to recover after a few iterations as in the previous case. The fidelity plot of the projected images is depicted in Extended Data Figure 7d. As expected, the system undergoes much more degradation as compared with the previous case but is still able to recover from it.

Modulation scheme: To convert the complex-field control patterns produced by the network to a pattern that can be used by a phase-only SLM, we use a modulation referred to as the checkerboard scheme. In the proposed approach, the amplitude information is encoded in the distribution of neighboring fields having conjugate phases with respect to one another (hence the name checkerboard). In this way, the amplitude of the desired complex field is proportional to the cosine of the phases and therefore can be physically implemented via a phase-only SLM[38].

Amplitude modulation vs. complex modulation: Extended Data Figure 3 depicts examples of projected images on the camera and the convergence speed with amplitude-modulation, i.e. when the network is configured to produce real-value control patterns. Comparing the fidelity, signal to noise ratio as well as the convergence speed, it can be concluded that the complex-field modulation scheme is more efficient than its amplitude-only counterpart. This is not surprising and well known.

Comparison between performances of the two-subnetwork approach and that of a single neural network for amplitude-modulation

In a supervised manner, the projector network is trained to reconstruct input SLM images going through the fiber from the corresponding amplitude-only patterns observed on the camera. For training this network, a dataset of 20000 image pairs consisting of SLM patterns at the input of the fiber and amplitude-only patterns at the output of the fiber are used. For acquiring this dataset, we first experimentally measure the transmission matrix of the system[17] and then use the matrix to obtain the amplitude patterns at the output of the

fiber. It should be noted that using the transmission matrix for virtually projecting the SLM patterns is equivalent to sending the SLM patterns directly through the fiber. We do this in order to have a constant means to relate the inputs of the fiber to the outputs when comparing the two architectures of the neural networks. It should be noted that in the initial step, the SLM images are chosen randomly because, *a-priori*, no information about the correct SLM modulations for producing the desired target images on the camera is available. Therefore, due to this randomness, the output amplitude-only images are basically speckle patterns. By training the network with amplitude-speckle patterns as inputs and SLM images as outputs of the neural network (opposite to the path that light travels through the fiber), the network is then given a set of desired target patterns (20000 images of Latin alphabet characters) directly as inputs to give out a set of 20000 SLM images predicted for the target images. These SLM images are then tested to observe the projected patterns on the camera. An example of the projected images is shown in Extended Data Figure 8. For the sake of comparison, the *projector network* with two sub-networks is also trained with the same dataset and then is used to predict the SLM images of the Latin alphabet dataset. It can be observed that the two sub-network approach gives a better projection performance even from the first iteration. One can continue with the training of this network by replacing the dataset used in the first iteration with another dataset that is comprised of the SLM images predicted for the Latin alphabet set and their corresponding fiber outputs with the hope that starting from SLM images that are not completely random, better projection fidelities could be achieved. This new dataset is then used for the second iteration of training. This procedure can be repeated a few times. The trajectory of the projected images is also plotted in Extended Data Figure 8. Noticeably, the single network is not able to improve its projection performance even after few iterations of the training. This is due to the fact that the architecture of the network does not allow it to compare the actual projected images on the camera with the desired ground truth target images at any point in the course of training while the two sub-network approach does. This is expected because the network with a single sub-network is trained solely for the purpose of predicting the corresponding SLM patterns of the amplitude images observed on the camera and not to correct for the differences between the target images and projected images. This becomes even more important as time elapses and the system changes.

Data availability

The directory to the dataset required to reproduce results appearing in Table1 can be found at https://github.com/Babak70/Projector_network.

Code availability

Our neural network framework is available at https://github.com/Babak70/Projector_network.

DOI: https://zenodo.org/record/3727136

Any correspondence and reasonable request for extra materials should be directed to the corresponding author babak.rahmani@epfl.ch.

ACKNOWLEDGMENTS

The authors would like to acknowledge the anonymous reviewers whom contributed to the manuscript.

C.M. and D.L. acknowledge the financial support of the Gebert Rüf Stiftung via the grant “Flexprint” (GRS-057/18, Pilot Projects track).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

CONTRIBUTIONS

B.R. carried out the experiment and analysis of the data and wrote the manuscript. D.L. wrote the code for the transmission matrix measurements. E.K, N.B and U.T participated in the analysis and in the SLM modulation format. D.P. and C.M proposed, supervised the project and contributed to writing the manuscript.

REFERENCES

1 Spitz, E. & Werts, A. Transmission des images à travers une fibre optique. *Comptes Rendus Hebd. Des. Seances De. L Acad. Des. Sci. Ser. B* **264**, 1015 (1967)

2 Di Leonardo, R. & Bianchi, S. Hologram transmission through multi-mode optical fibers. *Optics express* **19**, 247–254 (2011).

3 Čižmár, T. & Dholakia, K. Shaping the light transmission through a multimode optical fibre: complex transformation analysis and applications in biophotonics. *Optics Express* **19**, 18871–18884 (2011).

4 Čižmár, T. & Dholakia, K. Exploiting multimode waveguides for pure fibre-based imaging. *Nature communications* **3**, 1027 (2012).

5 Bianchi, S. & Di Leonardo, R. A multi-mode fiber probe for holographic micromanipulation and microscopy. *Lab on a Chip* **12**, 635–639 (2012).

6 Andresen, E. R., Bouwmans, G., Monneret, S. & Rigneault, H. Toward endoscopes with no distal optics: video-rate scanning microscopy through a fiber bundle. *Optics letters* **38**, 609–611 (2013).

7 Gover, A., Lee, C. P. & Yariv, A. Direct transmission of pictorial information in multimode optical fibers. *JOSA* **66**, 306–311 (1976).

8 Friesem, A. A., Levy, U. & Silberberg, Y. Parallel transmission of images through single optical fibers. *Proceedings of the IEEE* **71**, 208–221 (1983).

9 Yariv, A., AuYeung, J., Fekete, D. & Pepper, D. M. Image phase compensation and real-time holography by four-wave mixing in optical fibers. *Applied Physics Letters* **32**, 635–637 (1978).

10 Yamaguchi, I. & Zhang, T. Phase-shifting digital holography. *Optics letters* **22**, 1268–1270 (1997).

11 Cuche, E., Bevilacqua, F. & Depeursinge, C. Digital holography for quantitative phase-contrast imaging. *Optics letters* **24**, 291–293 (1999).

12 Papadopoulos, I. N., Farahi, S., Moser, C. & Psaltis, D. Focusing and scanning light through a multimode optical fiber using digital phase conjugation. *Optics express* **20**, 10583–10590 (2012).

13 Papadopoulos, I. N., Farahi, S., Moser, C. & Psaltis, D. High-resolution, lensless endoscope based on digital scanning through a multimode optical fiber. *Biomed. Opt. Express* **4**, 260–270 (2013).

14 Choi, Y. *et al.* Scanner-free and wide-field endoscopic imaging by using a single multimode optical fiber. *Physical review letters* **109**, 203901 (2012).

15 Caravaca-Aguirre, A. M., Niv, E., Conkey, D. B. & Piestun, R. Real-time resilient focusing through a bending multimode fiber. *Optics express* **21**, 12881–12887 (2013).

16 Gu, R. Y., Mahalati, R. N. & Kahn, J. M. Design of flexible multi-mode fiber endoscope. *Optics express* **23**, 26905–26918 (2015).

17 Loterie, D. *et al.* Digital confocal microscopy through a multimode fiber. *Optics express* **23**, 23845–23858 (2015).

18 Popoff, S., Lerosey, G., Fink, M., Boccara, A. C. & Gigan, S. Image transmission through an opaque material. *Nat. Commun.* **1**, 81 (2010).

19 LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).

20 McCann, M. T., Jin, K. H. & Unser, M. Convolutional neural networks for inverse problems in imaging: A review. *IEEE Signal Process Mag.* **34**, 85–95 (2017).

21 Rivenson, Y. et al. Deep learning microscopy. *Optica* **4**, 1437–1443 (2017).

22 Molesky, S. et al. Inverse design in nanophotonics. *Nature Photonics* **12**, 659 (2018).

23 Paruzzo, F. M. *et al.* Chemical shifts in molecular solids by machine learning. *Nature communications* **9**, 4501 (2018).

24 Rahmani, B., Loterie, D., Konstantinou, G., Psaltis, D. & Moser, C. Multimode optical fiber transmission with a deep learning network. *Light: Science & Applications* **7**, 69 (2018).

25 Borhani, N., Kakkava, E., Moser, C. & Psaltis, D. Learning to see through multimode fibers. *Optica* **5**, 960–966 (2018).

26 Caramazza, P., Moran, O., Murray-Smith, R. & Faccio, D. Transmission of natural scene images through a multimode fibre. *Nature communications* **10**, 2029 (2019).

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

28 Li, S., Deng, M., Lee, J., Sinha, A. & Barbastathis, G. Imaging through glass diffusers using densely connected convolutional networks. *Optica* **5**, 803–813 (2018).

29 Turpin, A., Vishniakou, I. & d Seelig, J. Light scattering control in transmission and reflection with neural networks. *Optics express* **26**, 30911–30929 (2018).

30 Psaltis, D., Sideris, A. & Yamamura, A. A. A multilayered neural network controller. *IEEE control systems magazine* **8**, 17–21 (1988).

31 Xu, Z., Yang, P., Hu, K., Xu, B. & Li, H. Deep learning control model for adaptive optics systems. *Applied optics* **58**, 1998–2009 (2019).

32 Sutton, R. S. & Barto, A. G. Reinforcement learning: An introduction. (2011).

33 Cohen G, Afshar S, Tapson J, van Schaik A. EMNIST: An extension of MNIST to handwritten letters. arXiv preprint arXiv:1702.05373, (2017).

34 Närhi, M. *et al.* Machine learning analysis of extreme events in optical fibre modulation instability. *Nature communications* **9**, 1–11 (2018).

35 Xiong, W. *et al.* Deep learning of ultrafast pulses with a multimode fiber. *arXiv preprint arXiv:1911.00649* (2019).

36 Salmela, L., Lapre, C., Dudley, J. M. & Genty, G. Machine learning analysis of rogue solitons in supercontinuum generation. *arXiv preprint arXiv:2003.05871* (2020).

37 Teğin, U. *et al*. Controlling spatiotemporal nonlinearities in multimode fibers with deep neural networks. *APL Photonics* **5**, 030804 (2020).

38 Davis, J. A., Cottrell, D. M., Campos, J., Yzuel, M. J. & Moreno, I. Encoding amplitude information onto phase-only filters. *Applied optics* **38**, 5004–5013 (1999).

Table 1 Neural network and transmission matrix image projection average fidelities (in percent) for various dataset

Dataset (1000 samples)	SUM		Red		Green		Blue		Min		Max		Average		Variance	
	NN	TM	NN	TM	NN	TM	NN	TM	NN	TM	NN	TM	NN	TM	NN	TM
Latin alphabet	95.0	98.1	92.5	96.0	90.7	96.6	91.2	97.2	90.7	95.0	96.0	98.1	92.4	96.9	3.7	0.8
Digits	95.2	98.2	92.4	96.2	90.9	96.7	91.6	97.3	90.9	95.2	96.2	98.2	92.5	97.1	3.5	0.7
Random sketches	86.6	91.2	83.3	89.0	83.6	89.8	82.0	91.2	82.0	86.6	89.0	91.2	83.9	90.3	3.9	1.2

FIGURES

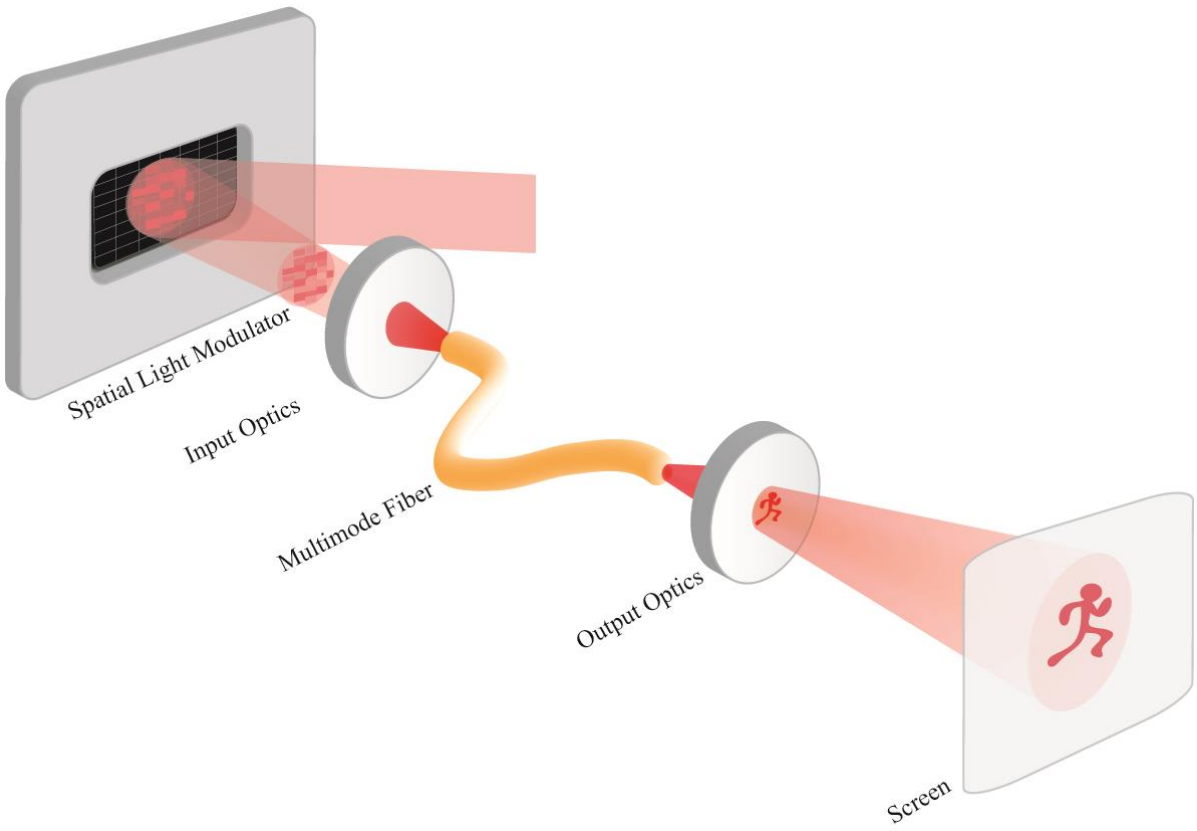


Fig. 1| Fiber projector. The setup consists of an input signal (laser source), a light modulator (SLM), the system (multimode fiber), and a screen. In the training phase, a detector (camera) is used in place of the screen to record the system's responses to the inputs. The neural network is then fed with the pairs of input signals and the system's responses so that it finds the forward and backward response functions of the system. Once trained, the network is given the target output (the picture of the "running man" here) that is to be projected on the screen and in return it gives out the appropriate control signal that would result in the target image (the network finds the modulation required to undo the light's scrambling inside the fiber).|

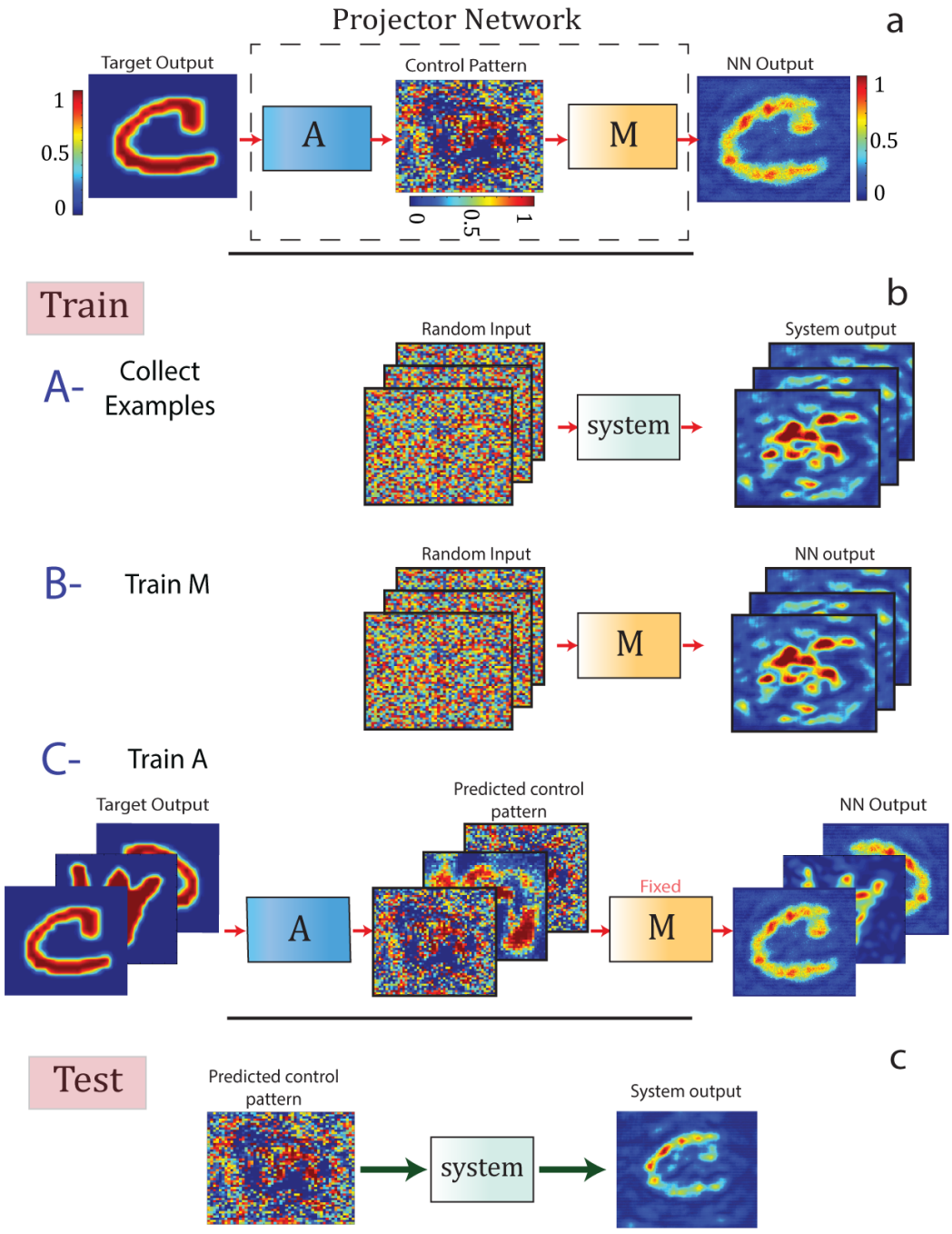


Fig. 2| Neural network's architecture and training procedure. a, the overall schematic of the projector network consists of two sub-networks: the Model (M) and the Actor (A). Once trained, the sub-network Actor accepts a target pattern desired to be projected at the output of the system (here MMF) and accordingly generates a control pattern (here SLM image) corresponding to the target pattern. The role of the sub-network Model is to help the Actor to come up with control patterns that are bound by the physics of light propagation through the fiber. **b,** the training procedure is carried out in three steps: A- a number of input control patterns are sent through the system and the corresponding outputs are captured on the camera. B- the sub-network Model is trained on these images to learn the mapping from the SLM to camera; hence the Model is essentially learning the optical forward path of light starting from its reflection from the SLM, propagation through the MMF and finally impinging on the camera. C- while the sub-network Model being fixed,

the Actor is fed with a target image and is asked to produce an SLM image corresponding to that target image. The Actor-produced SLM image is then passed to the fixed sub-network Model now mimicking the fiber. The error between the output of the Model and the target image is back propagated via the Model to Actor to update its trainable weights and biases. **c,** the test procedure is carried out by feeding the target image to the trained sub-network Actor and acquiring the appropriate SLM image corresponding to that target image and sending it through the system.

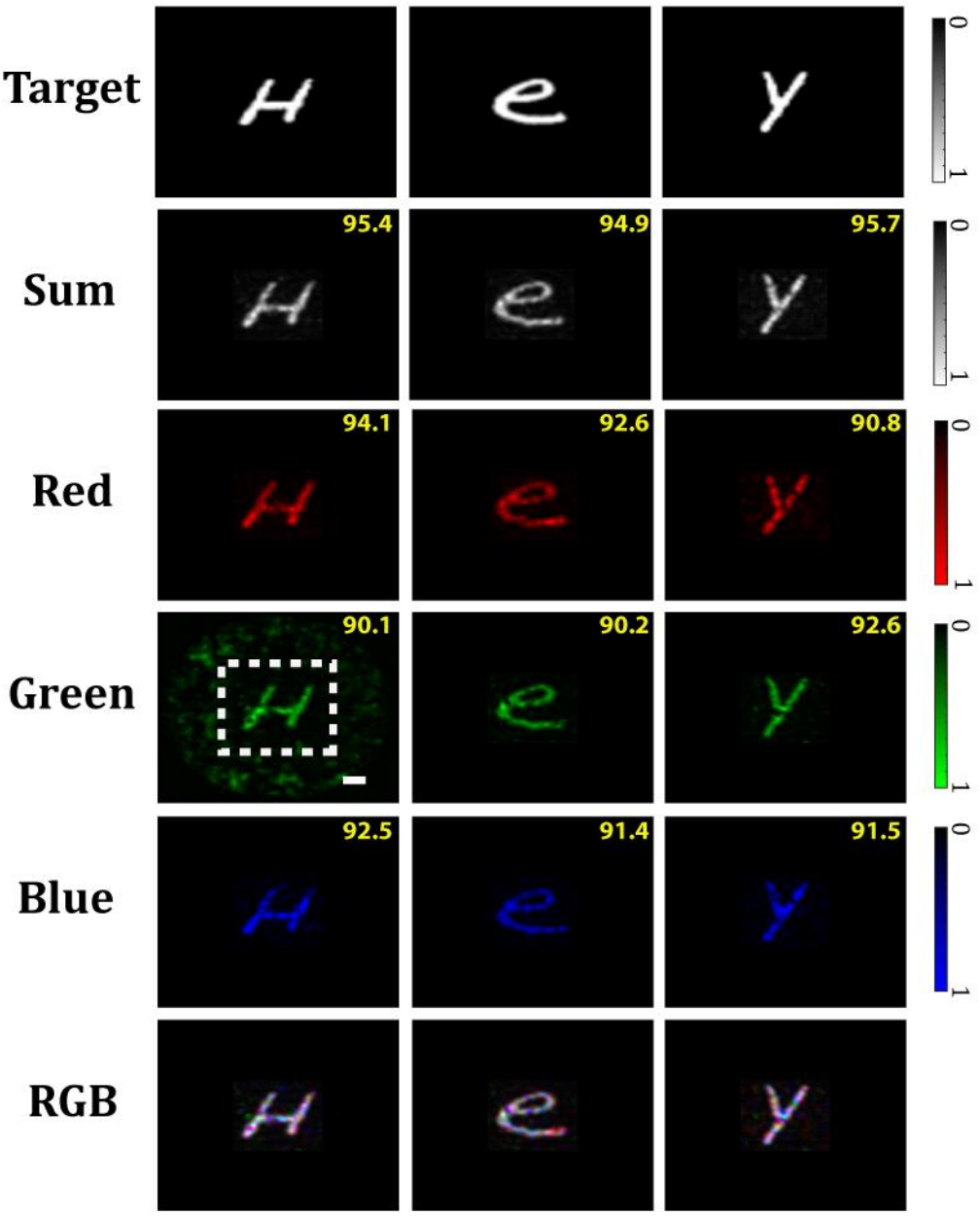


Fig. 3| MMF system arbitrary output control with partial measurements. Examples of images projected onto a camera at the output of a MMF are shown. The projection of images are carried out for three different wavelengths (633 nm, 532 nm, 488 nm) corresponding to red (R), green (G) and blue (B) as well as the superposition of those colors either as a 3 channel RGB image or as a one channel incoherent image produced by summing R, G and B. The neural network is trained with EMNIST dataset as target images. The appropriate SLM patterns generated by the network are sent to the system to obtain the desired targets on a rectangular area of size 200×200 pixels on the camera (corresponding to an area of 19×19 μm^2 on the output facet of the fiber). This area is shown as a dashed box on one of the examples. Scale bar 5 μm. The fidelity of projected images with respect to the corresponding target images is also shown.

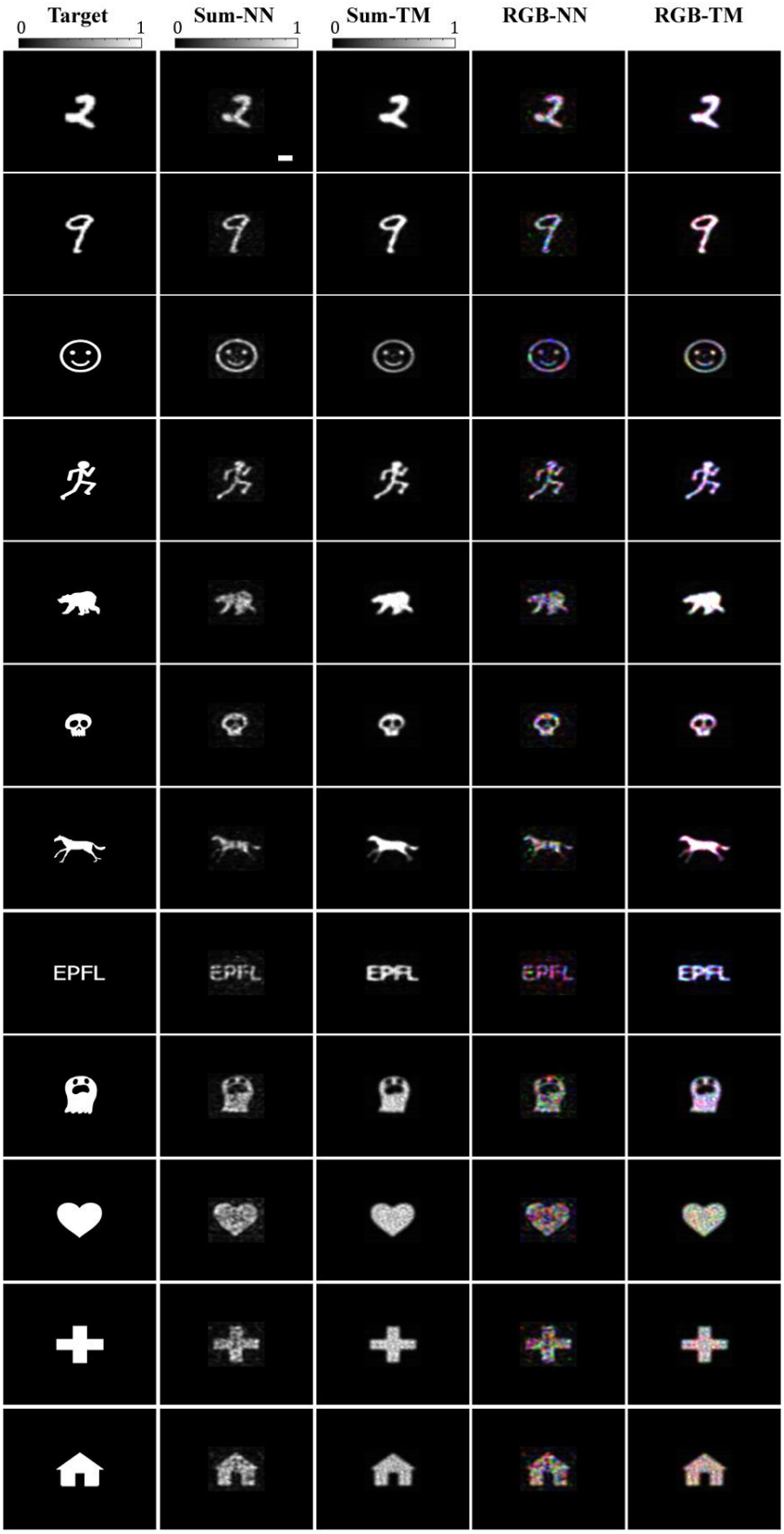


Fig. 4| Neural network generalization ability for controlling the MMF output. Examples of images projected onto a camera at the output of a MMF are shown. The control patterns that produce the output images on the camera (the incoherent summation of red, green and blue wavelengths as well as the 3 channel RGB images) are generated either via

a neural network (NN) trained on the dataset of Latin alphabet characters (different from the category of target images) or via the transmission matrix full measurement approach (TM). The generalization of the network is demonstrated in its ability to provide control patterns for target images that come from a different class as that of the images originally used for training. Scale bar 5 μm.

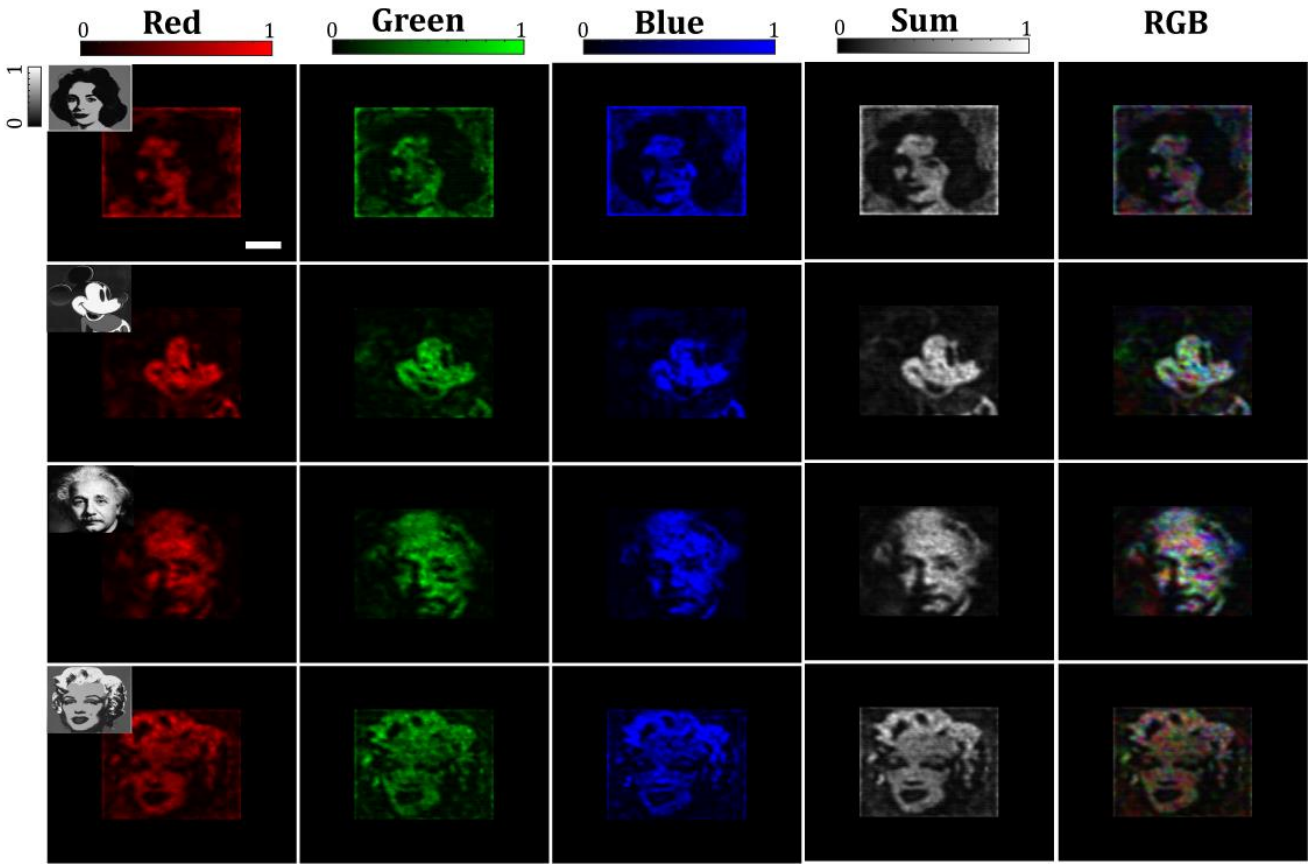
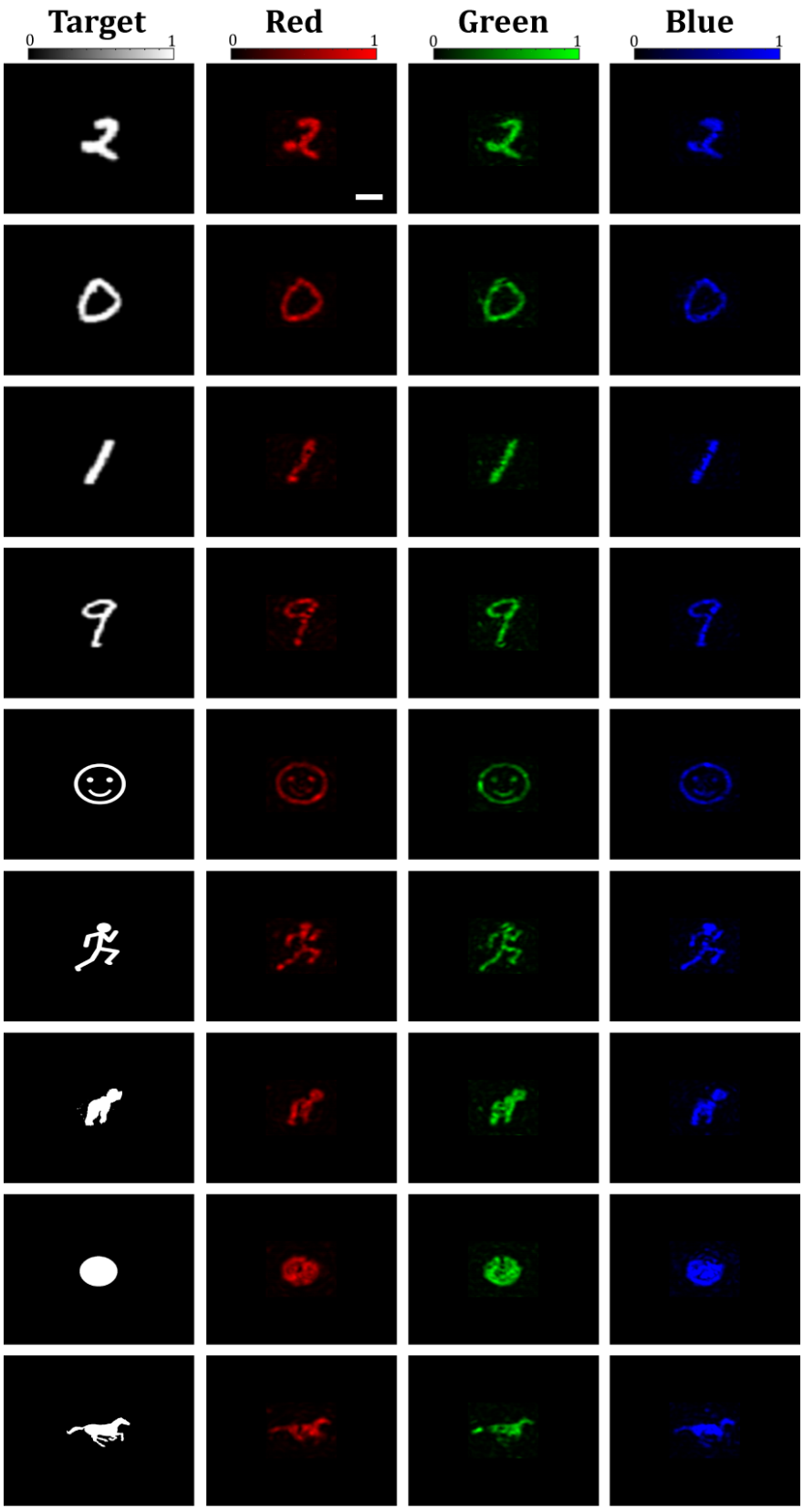
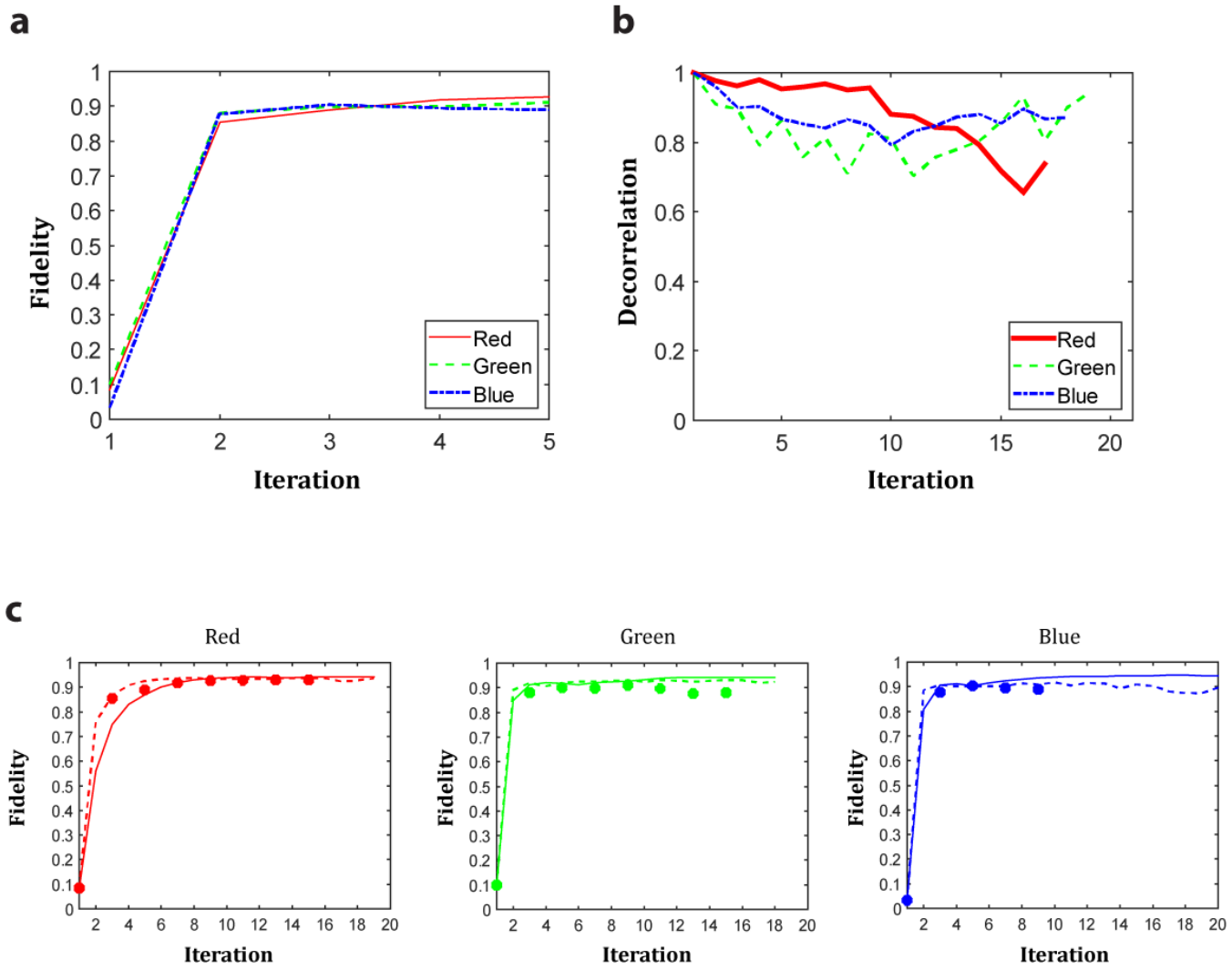


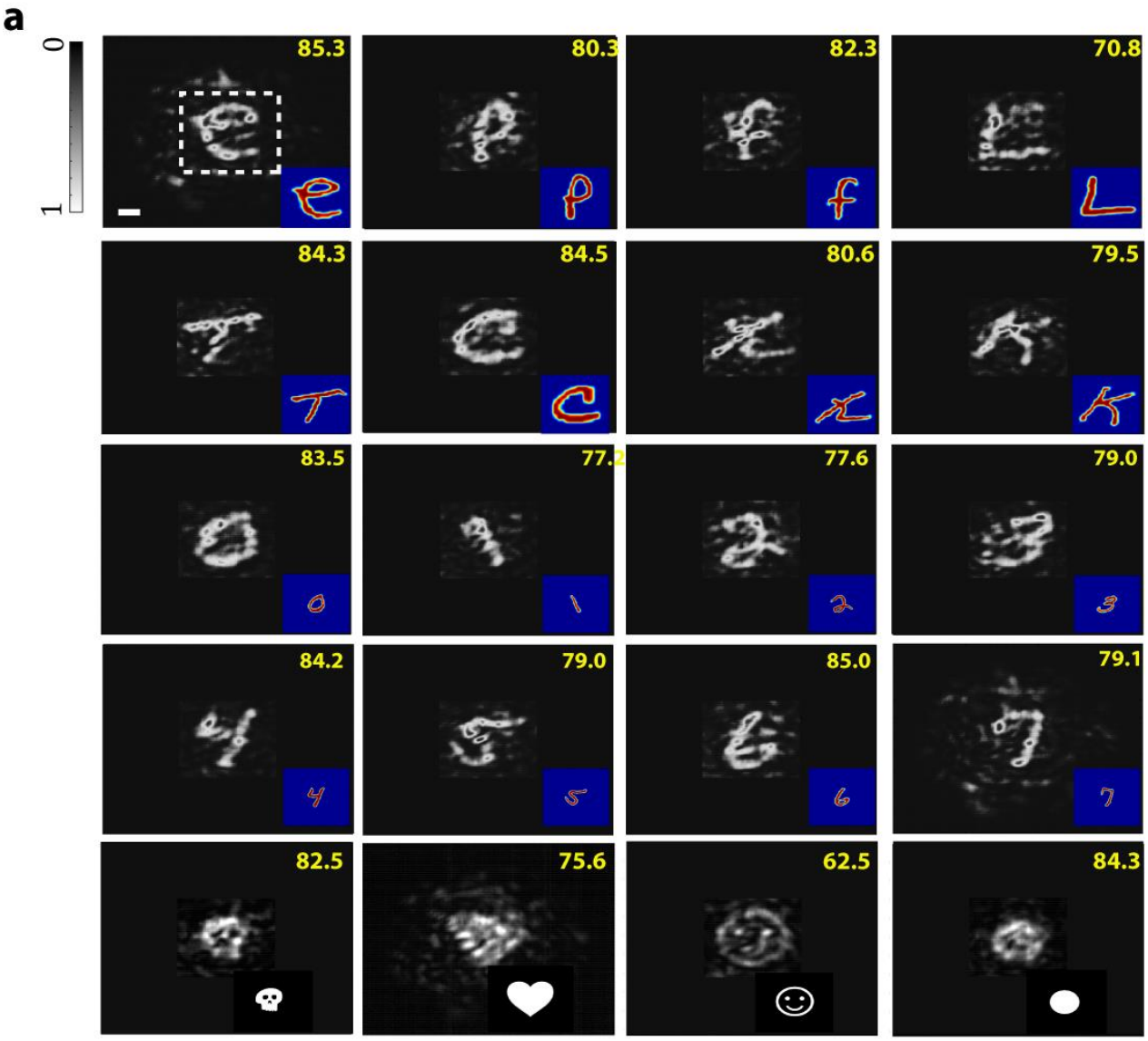
Fig 5| Continuous gray-scale image projection. Examples of natural-scene continuous gray-scale target and experimentally projected images being sent through the MMF and captured on the camera for colors red, green, blue and the 3-channel RGB as well as the superposition of all three colors in 1 channel (sum) are shown. Scale bar 5 μm. Inset images in column 1: **a**, Liz (Elizabeth) Taylor by Andy Warhol 1964. **b**, Mickey Mouse by Andy Warhol (Diamond Dust) 1981 **c**, Marilyn Monroe 31 by Andy Warhol 1967. © The Andy Warhol Foundation for the Visual Arts, Inc. / 2020, ProLitteris, Zurich for WARHOL ANDY's works **d,** Portrait of Albert Einstein 1936 Photo by Bachrach/Getty Images.

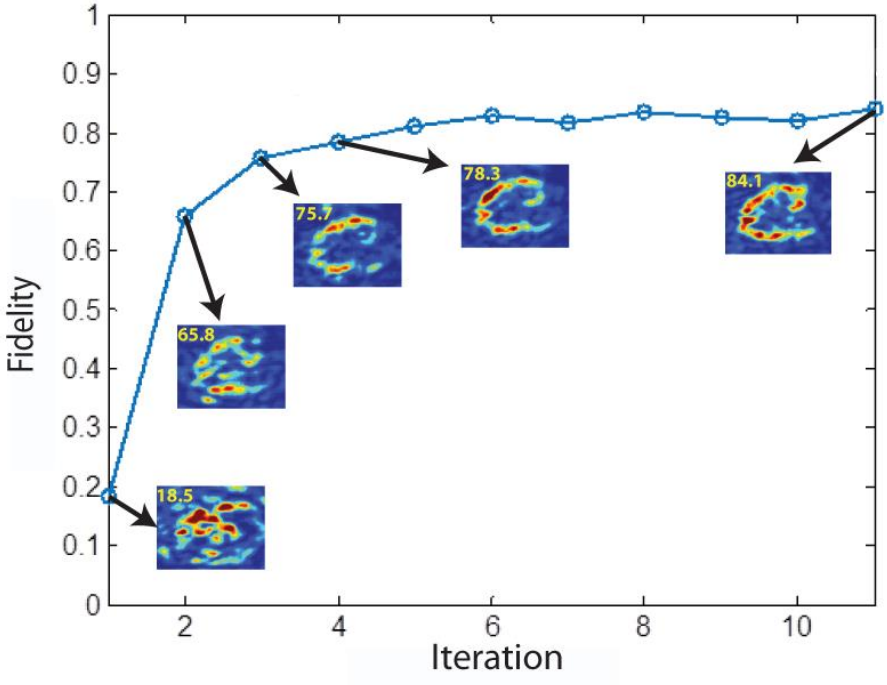


Supplementary Fig 1. Additional examples as in Fig. 3.

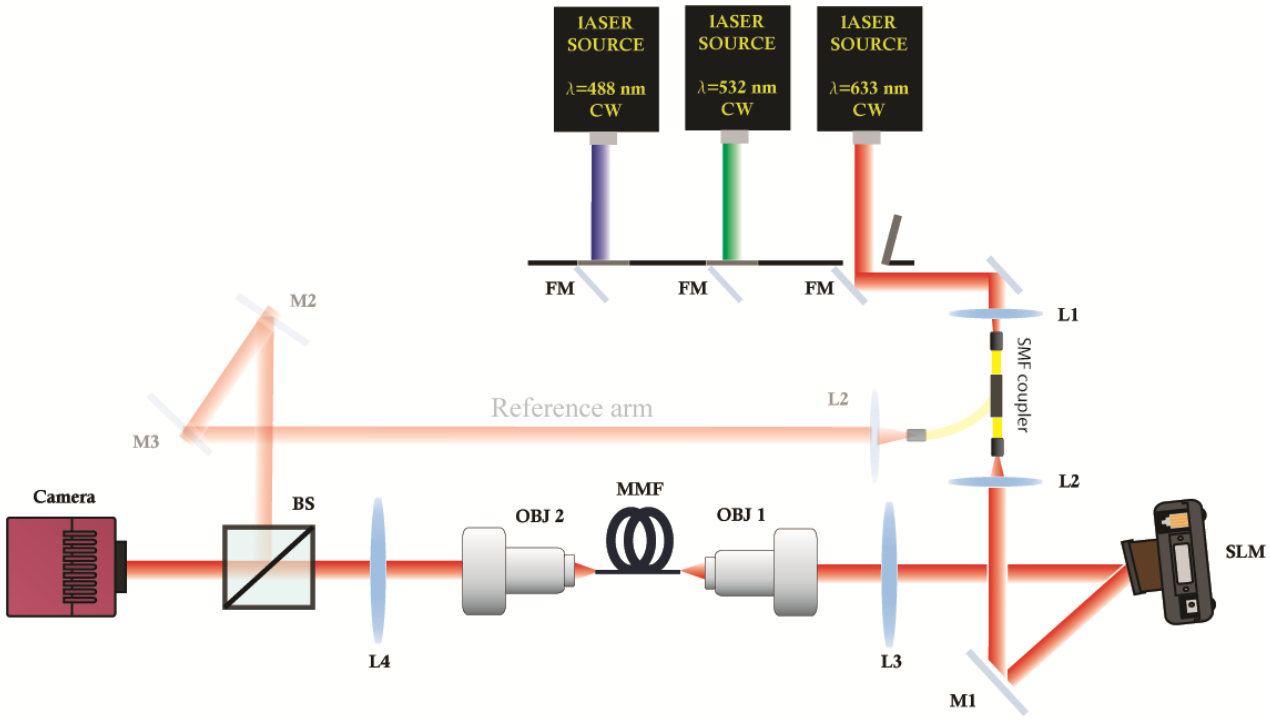


Supplementary Fig 2. a, the fidelity trajectory of experimentally projected images versus the training iteration number is plotted for all three colors. **b**, while training, the instability of the system (estimated as the correlation between instances of the system's response to a constant input signal being sent through the system over and over) is monitored over time (If the system is time-invariant, then the decorrelation plot holds a value of one continually). **c**, degradation in the fidelity of projected images due to the non-perfect modulation scheme as well as the variation of the system with time is shown by using the experimentally measured transmission matrix (TM) to forward the neural network's predicted SLM images for all three colors. The fidelities in part (**a**) are redrawn in part (**c**) for comparison. As observed, the experimentally projected images (solid circles) closely follow the track of time variant TM-based relayed projections (dashed lines) and both eventually fall below the track of time-invariant TM-based relayed projections (solid lines). In the former, what is taken out from the learning algorithm is only the effect of modulation scheme, whereas in the latter, it is the lumped effect of time variation as well as the modulation scheme. The ripples in the trajectory of the graphs in (**c**) (dashed lines) show that the network is continuously trying to correct for the drifts.

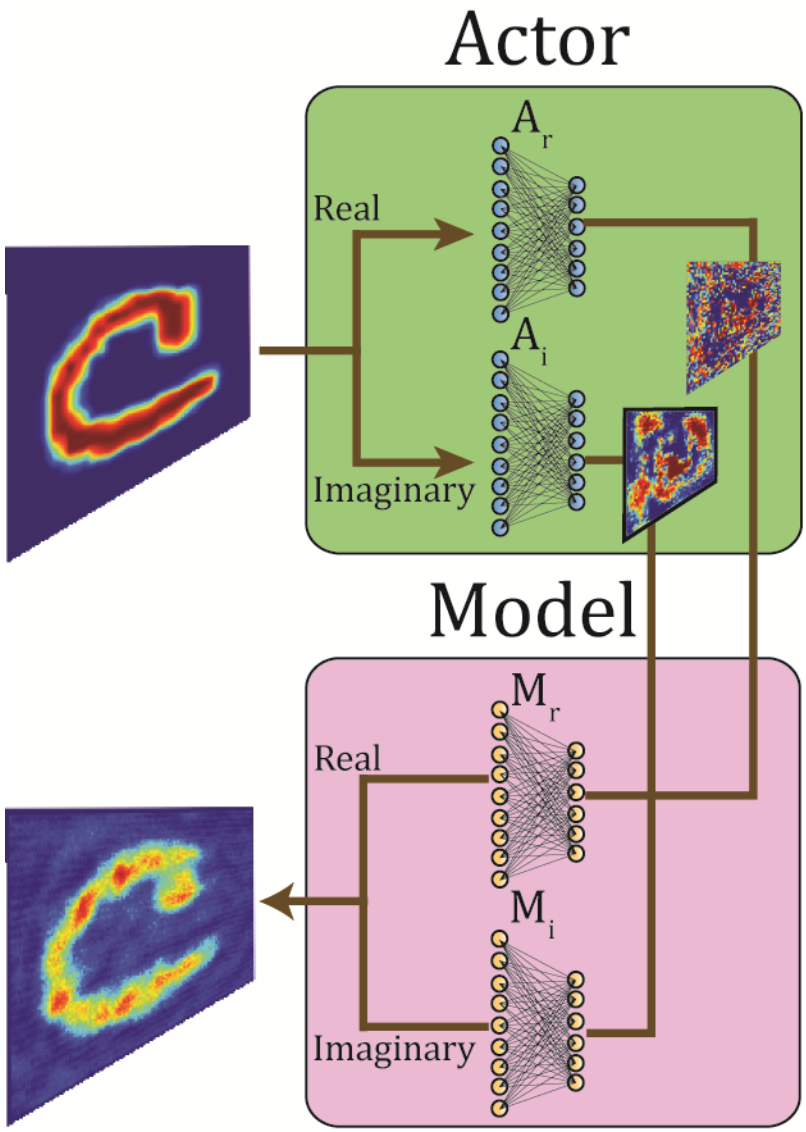

a
0
1
85.3
80.3
82.3
70.8
84.3
84.5
80.6
79.5
83.5
77.6
79.0
84.2
79.0
85.0
79.1
82.5
75.6
62.5
84.3


b
Fidelity
Iteration
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
2
4
6
8
10
18.5
65.8
75.7
78.3
84.1

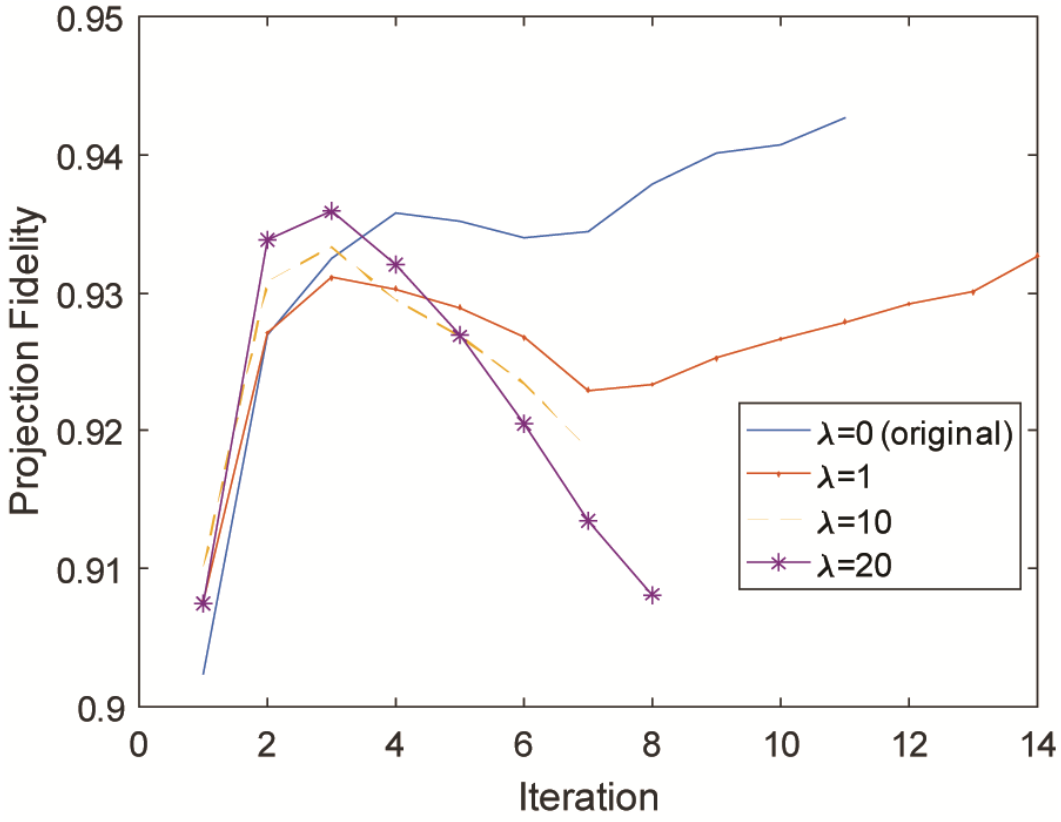
Supplementary Fig 3. a, examples of images projected onto a camera at the output of a MMF (wavelength 780 nm) are shown. The network is forced to generate amplitude-only control patterns. These patterns are then sent to the system and the outputs on the camera are captured. The network is trained with target images of Latin characters but it is also used to predict control patterns for target images from different categories. The visible background of the projected images accounts for the lower signal to noise ratio of the images (also lower fidelities) as compared with that of the complex value control patterns. This is attributed to missing out on controlling the phase of control signals. **b**, plot of the convergence speed for amplitude-only input controls.



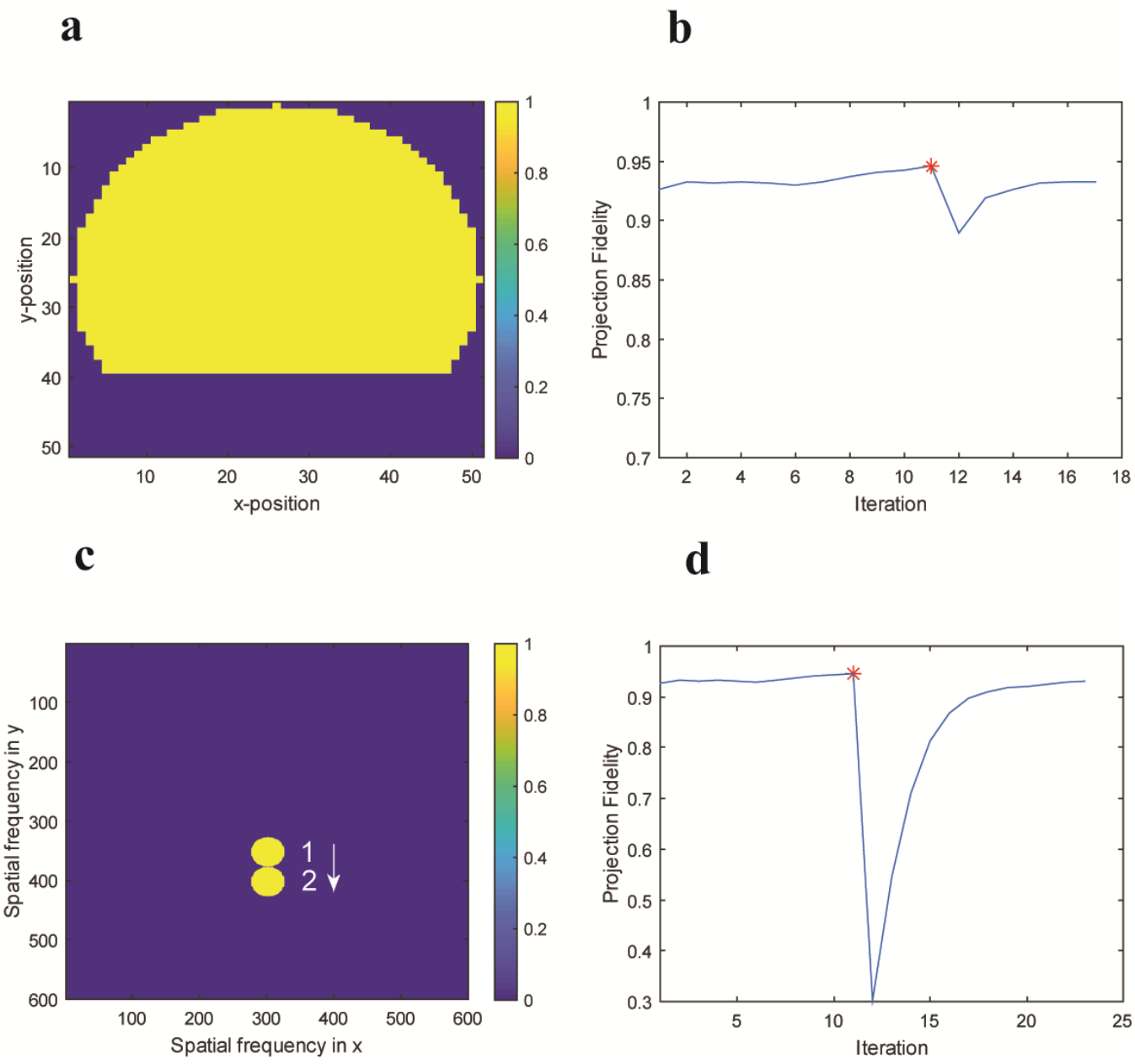
Supplementary Fig 4. Detailed diagram of the optical setup. Control patterns are generated via the SLM, guided through the fiber and captured by the camera. L1: Aspheric lens, L2: f = 100mm lens; L3: f = 250mm lens; L4: f = 250mm lens; OBJ1, OBJ2: 60x microscope objective; SLM: spatial light modulator; M1: mirror; FM: flip mirror; SMF: single mode fiber; MMF: multimode fiber, BS: beam splitter.



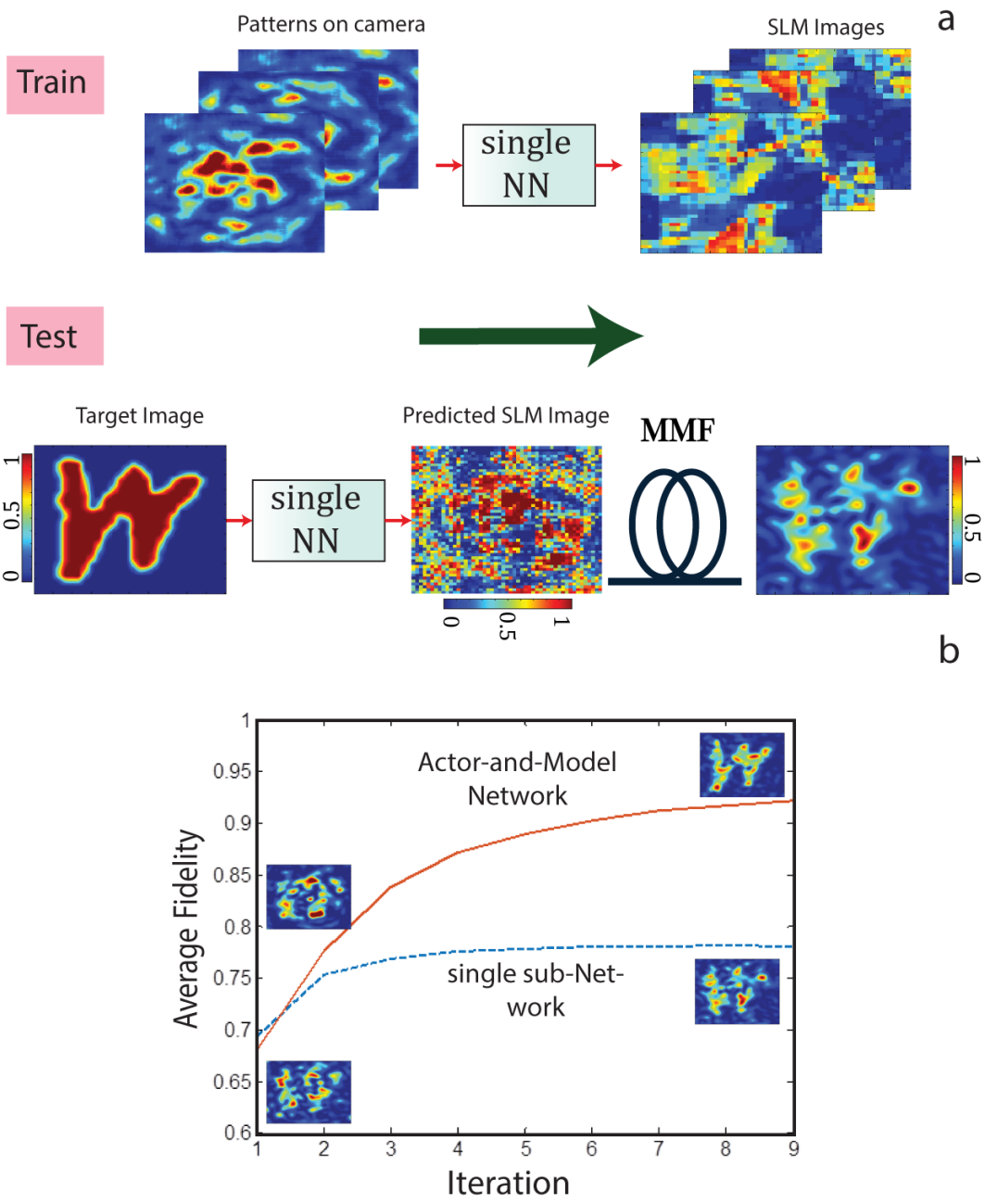
Supplementary Fig 5. The Actor and Model networks are comprised of two sub-networks, (A_{real}, A_{imag}) and (M_{real}, M_{imag}) respectively, to cope with the real and imaginary parts of input-output fields. All sub-networks are fully-connected. Thereby, the input images to the Actor network is first flattened out (from size 200×200 pixels to 40000×1 vectors) and then fed to the sub-networks A_{real} and A_{imag} (input nodes 40000, output nodes 2601). In the training step, the output vectors (size 2601×1) of A_{real} and A_{imag} are passed on to M_{real} and M_{imag} (input nodes 2601, output nodes 40000), respectively. The virtual neural network output image of the target image (size 200×200) is then produced at the output of the Model. Once trained, the output vectors of the Actor network can be directly reshaped to produce the real and imaginary parts of the SLM images (reshaping from size 2601×1 to 51×51 pixels). If these SLM images are uploaded to the SLM and sent through the fiber, they produce projected images on the camera that are similar to the target images.



Supplementary Fig 6. The Actor network is additionally trained with images used for training of the Model network. The loss function for training the Actor is comprised of two terms: the loss term in equation (3) and an extra mean square term with weight λ. The fidelity is obtained for multiple values of λ. The original learning algorithm corresponds to the case in which λ=0.



Supplementary Fig 7. Two scenarios of minor (a, b) and major (c, d) perturbations due to misalignment of the MMF system is studied. **a-** The proximal side of the fiber has been slightly blocked with a spatial filter depicted in (a). **b-** the fidelity plot of the projected images before and after the perturbation event (denoted by a star symbol) is shown. **c-** In the second scenario, the proximal side of the fiber has been slightly tilted (changed in angle). This results in a shift in the reception bandwidth of the fiber facet from spatial frequency configuration 1 to 2. **d-** The fidelity plot of the projected images for the perturbed system is shown.



Supplementary Fig 8. a- Schematic of a neural network (NN) that is made of one single network as an alternative architecture for the Actor-Model network proposed in the main text. The single NN is trained with images obtained from the camera and the corresponding input control patterns (here amplitude-only control patterns). Once trained, the target image is directly passed on to the network and is asked for its corresponding control pattern. The predicted pattern is then sent through the fiber. **b-** The average fidelity trajectory of the projected images of 20000 Latin alphabet characters versus the training iteration number.