Light scattering control with neural networks in transmission and reflection

Alex Turpin, Ivan Vishniakou, and Johannes D. Seelig
Center of Advanced European Studies and Research, 53175 Bonn, Germany
(Dated: May 28, 2018)

Scattering often limits the controlled delivery of light in applications such as biomedical imaging, optogenetics, optical trapping, and fiber-optic communication or imaging. Such scattering can be controlled by appropriately shaping the light wavefront entering the material. Here, we demonstrate a machine-learning approach for light control. Using pairs of binary intensity patterns and intensity measurements we train neural networks (NNs) to provide the wavefront corrections necessary to shape the beam after the scatterer. Additionally, we demonstrate that NNs can be used to find a functional relationship between transmitted and reflected speckle patterns. As a proof of the validity of this relationship, we demonstrate focusing and scanning of light in transmission through opaque media using reflected light. Our approach demonstrates the versatility of NNs for light shaping and for efficiently and flexibly correcting for scattering. In particular, the feasibility of transmission control based on reflected light opens up new opportunities for applications in optical imaging, sensing, and light delivery.

When light propagates through a linear non-homogeneous and non-isotropic material its wavefront becomes distorted due to aberrations and scattering, resulting in an apparently random interference pattern of granular speckles [1, 2]. Such scattering conditions hamper the controlled delivery of light and the engineering of the PSF, which is a basic requirement for many applications [3, 4]. To counteract this effect, methods based on shaping the light wavefront entering the scattering material have been developed. Wavefront shaping is typically achieved by using spatial light modulators (SLMs) [7, 8], which, with their millions of degrees of freedom (pixels), allow focusing through diffusers [9, 11], multimode fibers [8, 12, 13], and biological tissue [14, 17]. Different techniques have been developed to determine the appropriate wavefront corrections to be displayed on the SLM. The first demonstration of scattering control took advantage of iterative wavefront optimization [9, 15, 19], which approaches the targeted light distribution, typically a single or multiple focal spots, by updating the wavefront depending on the result after each optimization step [1, 19, 20, 21]. These feedback-based algorithms calculate the wavefront correction separately for each focal position or shape and can be optimized to achieve very fast focusing times [11, 22]. A second approach, also typically used to control a single focus, is digital optical phase conjugation which uses interferometry to measure the scattered light field and reverses it with an SLM [16, 23, 26]. This technique has the advantage of achieving update rates approaching the millisecond range needed for imaging in dynamic biological tissue [27], while however requiring that the scattering signal originates from a coherent source suitable for reversal. A third group of methods aims for describing and controlling the scattering process simultaneously across an entire field of view which was first achieved with the help of a transmission matrix [10, 28, 29]. For obtaining the transmission matrix, one needs to measure light phase, which, similar to digital optical phase conjugation, requires technically more demanding interferometric approaches. To simplify such experiments, computational methods for estimating incompletely measured information have been implemented which for example can infer the light phase from intensity measurements [30, 31].

Another set of computational techniques that, thanks to the development of programming frameworks together with the computational power of GPUs, is increasingly being applied in imaging and microscopy relies on machine learning (ML) [72, 33] and in particular on NNs [36–39]. The usefulness of these techniques [40–49] has been demonstrated in the context of light scattering for image analysis [37, 44, 45, 47], where the goal lies in the classification of an object across a scattering layer, or image reconstruction based on a predefined data set [48]. In astronomy, NNs have been applied for the correction of weak scattering encountered when imaging through the atmosphere, for example for the control of multi mirror telescopes [49]. For light control, genetic algorithms, a class of iterative optimization algorithms, have been used for optimizing focusing across scattering materials [40, 41, 43]. Single- and multi-focus single-shot control (after training) over a 5 × 5 pixel area has been achieved using support vector regression [40], but the reported small field of view, low signal to noise ratio, and long training times (97 min) are limiting for high-resolution PSF engineering.

While the methods outlined above allow focusing through scatterers in transmission, an additional set of challenges arises when the focal plane lies hidden behind or inside the scatterer, remote from direct optical access. For applications such as sensing, imaging or communication this is the more relevant configuration [50]. For example in biological imaging, fluorescent signals or guidestars can be used to monitor excitation intensity inside or behind a scatterer [2, 13, 19, 21]. Particularly in the presence of strong scattering, however, these signals are often dim and generally have a limited pho-
ton budget [51]. Alternatively, back-scattered excitation light can provide feedback about the beam [52, 54]. These signals need to be additionally filtered to remove out-of-focus light and using various combinations of temporal, frequency, or spatial gating [52, 58], one aims for extracting photons that are scattered little and therefore retain image information. Since such weakly scattered photons disappear exponentially with depth, they in turn limit imaging depth.

However, even under strongly scattering conditions reflected (or backscattered) photons carry information about transmitted light [50, 59, 60]. Mutual information between speckle patterns generated in these two opposite scattering directions indicates that reflected light might potentially be used to control transmitted light [50, 59, 60]. This would require that a functionally explicit relationship between these two scattering signals could be found and that the available information would be sufficient for controlling one signal through the other. So far, reflected light has been used to maximize the energy sent into a sample [61, 62], but without control over the resulting light distribution inside the sample, or required an embedded highly scattering target to achieve a localized light distribution [63]. Other schemes to take advantage of backscattered light have been suggested in theoretical work [64, 65], but these concepts have so far not been implemented experimentally.

Here, we discuss how neural networks can be used to image through materials with different scattering characteristics such as glass diffusers, multi-mode fibers, or paper. First, we show that single-layer NNs (SLNNs) and multi-layer convolutional NNs (CNNs) can be trained to control the light distribution behind scattering materials with high accuracy. Second, we show that NNs can be used to find functional relationships between transmitted and reflected light, i.e., they can predict transmitted speckle patterns from reflected speckle patterns with sufficient accuracy for light control through opaque materials. Taking advantage of this relationship we then show that NNs can be used for focusing in transmission using reflected light.

RESULTS

NN approach for scattering control

We here first outline the underlying approach of using NNs for light control through a scatterer, which is also sketched in Fig. 1(a). In an initial step, we generate a dataset consisting of pairs of binary illumination patterns displayed on the SLM and corresponding speckle patterns recorded with a CCD camera after transmission through the scatterer (64 × 64 macropixels for illumination patterns and 96 × 96 pixels for the CCD camera, see also Supplementary Fig. 1). These pairs of illumination and speckle patterns (typically on the order of 10000, but see below for training with a reduced number of patterns) are used to train the NNs as detailed in Methods, with the goal of inferring the relationship between the resulting scattered light distributions and the illumination patterns. We then feed the desired distribution into the trained NNs to predict the corresponding illumination pattern. This pattern is finally displayed on the SLM and the resulting light pattern is recorded with the camera. Each pattern, C (k), can be considered as a combination of plane waves with different wave vectors k. This distribution of plane waves is modified by the scattering material through a function F [C (k)] in a deterministic way and results in the speckle pattern, S, i.e., S = F[C (k)]. Through training, the NN learns an approximation of the function F needed to generate any light distribution after the scatterer.

The experimental setup is schematically shown in Fig. 1(b) and explained in detail in Methods. A laser beam is sent to the SLM, a high-speed digital micromirror device (DMD), which displays binary patterns of high and null intensity values (0s and 1s) with a frame rate of up to 22.7 kHz. We have tested our system, both with pseudo-random checkerboard-like patterns and with patterns obtained from Hadamard matrices. These patterns are imaged onto the back aperture of a microscope objective that focuses the light beam through the scatterer. A second identical microscope objective and a pair of lenses are used to collect and image the speckle patterns onto the CCD camera. A second detection path, including a non-polarizing beam splitter and an additional CCD camera are added for experiments with reflected light (see below and Methods for details).

SLNNs for light control through glass diffusers

In Fig. 2(a) we demonstrate the ability of SLNNs to generate diffraction-limited Gaussian foci through a glass diffuser (as used for example in [17]) at different positions in the image plane. Top images schematically illustrate the NN architecture and training process (see Methods for details); below that, the first rows show the intensity distribution captured with the CCD camera, while the second and third rows display horizontal and vertical cross sections through the center of the focus. Insets and red-dashed lines show the position and shape of the target distribution that is fed into the neural network and for which the network then calculates the appropriate SLM mask (the target distribution is displayed normalized to the experimentally recorded intensity).

The images show an excellent agreement between the desired and recorded patterns with a signal-to-noise ratio > 10 and an enhancement η = 32 ± 5 (see Methods, and Supplementary Fig. 1 for scanning across the entire field of view). The time to achieve light control depends on the number of recorded frames and the training time. For the typical datasets of 10000 frames (with a resolution of 64 × 64 macropixels on the DMD and 96 × 96 pixels on the CCD, recorded at 1000 Hz) training on a
single GPU required 34 s, and could be reduced down to 18 s while keeping an enhancement $\eta > 10$ (see Supplementary Fig. 5).

CNNs for light control through glass diffusers

While SLNNs are easy to implement and train, the underlying linearity limits their performance for many tasks [66]. A plethora of other network architectures have therefore been developed with the goal to improve over the performance of SLNNs. The most straightforward generalization of SLNNs combines multiple NN layers with connections between all neurons, resulting in a densely connected network. While such densely connected networks are not limited by linearity, the increased number of parameters also makes them more challenging to train, particularly for large data sets such as stacks of high resolution images. Network architectures were therefore developed that take into account the structure of the underlying data and convolutional neural networks (CNNs) have emerged as one of the most successful solutions for image processing [66]. The typical architecture of a CNN consists of multiple convolutional layers that extract features across an entire field of view, interspersed with pooling layers that down-sample the image, and fully connected layers. While a large number of different networks are applied for different tasks, with a few to a few hundred convolutional layers [32, 33, 66], we here found that a three-layer CNN (see Fig. 2b) and Methods for details) could be used for scattering control through a glass diffuser. To circumvent the difficulties of training nonlinear networks we pretrained the network with an autoencoder [66], a network that compresses and then uncompresses the data into a close approximation of the input. The part of the network that was used for compression then served as the initial CNN for scattering control (see Methods for details). In Fig. 2b we demonstrate the ability of CNNs to generate diffraction-limited Gaussian foci through a glass diffuser. The images again show an excellent agreement between the target pattern that was fed into the CNN (red dashed lines and inset) and the recorded patterns with a signal-to-noise ratio $> 10$ and an enhancement $\eta = 3.6 \pm 0.9$ (measured over an ensemble of 25 different focus positions, see Methods and Supplementary Fig. 2 for scanning across the entire field of view). CNNs in this case reduce the number of network parameters by 80% compared to SLNNs.

SLNNs for point spread function engineering

Since the SLNN is a linear network, we reasoned that after training it should be able to take advantage of the linearity of light scattering in non-absorbing media to generate arbitrary light distributions. To demonstrate the validity of our approach for controlling the light intensity distribution after the scatter we generated in Fig. 3 a variety of non-trivial shapes using SLNNs. Again, there is an excellent correspondence between the target distribution that enters that network (insets) and the recorded patterns. We note that thanks to the high frame rate of the DMD (22.7 kHz), alternatively, any shape can be generated with high fidelity by painting it spot by spot, e.g. similar to approaches for trapping ultra cold atoms [67] or optogenetics [68], as shown in Supplementary Fig. 2 and in Supplementary Movies 1–6.

SLNNs for light control through optical fibers

Our system is suited well to correct for scattering in materials with slow dynamics (on the order of a few tens of seconds, see Supplementary Fig. 5), such as optical fibers [12, 69, 70]. In particular, multimode optical fibers are ideal for applications in imaging and optogenetics, but modal dispersion and cross-talk distribute light into an apparently random speckle pattern. We therefore tested the performance of SLNNs for controlled light delivery through multimode fibers. In Fig. 4 single focus is scanned ($\eta = 10 \pm 3$) with different paths across the field of view of the fiber (including a circle, a square, and a $5 \times 5$ array of points), demonstrating that SLNNs are able to precisely control light through optical fibers.

NNs find functional relationships between transmitted and reflected speckle patterns

While many of the methods for focusing light through strongly scattering media rely on measuring transmitted light (as in the experiments described so far), many applications could benefit from using reflected light. Towards that goal we tested whether neural networks can take advantage of mutual information between transmission and reflection images [50, 59, 60] for light control. In the following we show that with the help of a functional relationship between reflected and transmitted speckle patterns one can control transmitted light using reflected light.

For this experiment we simultaneously recorded transmitted and reflected light by adding a non-polarizing beam splitter, a pair of imaging lenses and a CCD camera to the setup, as shown in Fig. 1. To achieve good signal to noise ratio of transmitted as well as reflected speckle patterns, we used paper as scattering material, which was more strongly scattering than the glass diffuser [56] and led to an increased amount of backscattered light. Movies of simultaneously recorded transmitted and reflected speckle patterns (with size of $128 \times 128$ pixels) were then generated by illuminating the sample with a series of checkerboard projections ($64 \times 64$). Once the speckle patterns were recorded, we trained a SLNN (SLNN1 in Fig. 6a) to find the relationship between transmitted and reflected light. To quantify the performance of this network we used the Pearson correlation co-
efficient as a similarity measure \[71]\) between transmission speckle patterns predicted by the network and measured transmission speckle patterns. Fig. \[5(a)\] shows the histogram of these correlation coefficients and the correlation coefficient between transmitted and reflected speckle patterns for comparison. Figs. \[5(b)\] and \(c\) show, respectively, an example of measured and predicted speckle patterns with median correlation \(\rho = 0.50\), while and Figs. \[5(d)\] and \(e\) show an example of measured reflected speckle pattern when a corresponding focus is generated in transmission.

**Focusing through opaque media in transmission using reflected light**

To take advantage of this mutual relationship between transmitted and reflected light for light control, we trained a second independent network (SLNN2 in Fig. \(3(a)\)) to infer the relation between reflected speckles and illumination patterns, similar to the training of the SLNN in the transmission configuration in the previous sections of the article (that is, with the reflected speckles as input of the SLNN and the illumination patterns as output, see Methods for details). Combining these SLNNs, as shown in Fig. \(3(a)\) (see Methods for details), allowed us to form transmission foci by only taking advantage of reflected light, based on SLNN1 relating reflected to transmitted speckle patterns. In Figs. \(3(b)-d\) we show, respectively, that we can scan a circle, a square, and a grid, demonstrating full control of transmitted modes using reflected modes over the entire field of view. This additionally demonstrates that the predicted speckle patterns (Figs. \(5\)) are sufficiently accurate for high-resolution light control. The measured enhancement in this case was \(\eta = 12 \pm 4\) (measured over an ensemble of 25 different focus positions). Note also that even though paper is more strongly scattering \([59]\) than the glass diffuser, this does not hinder the SLNN from light control. Focusing through paper with a SLNN as reported in the previous sections is shown in Supplementary Fig. 3.

**DISCUSSION**

In summary, we showed that NNs can be used to efficiently shape light through a variety of media with different scattering characteristics (Figs. \(4\)) to \(4\)). Once the NNs are trained, we achieve real-time, single-shot light control through the scattering material with high fdelity, in a fashion similar to transmission matrix approaches \([10, 29, 72]\). Specifically, we demonstrated the ability of SLNNs to focus and scan light through glass diffusers, multimode fibers, and paper, and to generate arbitrary light distributions through glass diffusers. We further showed that nonlinear networks, specifically a three layer CNN, can focus light through a glass diffuser.

In a second set of experiments, we demonstrated that with the help of two networks, one establishing an explicit functional relationship between light that is transmitted through a scatterer and light that is reflected, and one relating reflected light to illumination patterns, we can control transmission using reflection at diffraction limited resolution. SLNNs therefore prove to be well suited to take advantage of a recently described mutual information between transmitted and backscattered light for light control \([54, 59, 60]\).

To compare the performance of the NN method for focusing in transmission with other schemes, we quantified the enhancement as in \([73]\) (see Methods) and obtained values in the lower range of what was reported for intensity-only modulation for the SLNN \(30, 41, 73\) and lower values for the CNNs (but still with a sufficient SNR and enhancement for imaging applications), with the caveat that a direct quantitative comparison needs to take into account the specific combination of scatterer, optical setup (we used lower N.A. objectives than many of the reports with higher enhancement), and the number of controlled modes. The maximum number of controlled modes in our experiments was 4096 (64 \times 64); while the SLM has 768 \times 1024 pixels, using more degrees of freedom was not warranted given our optical configuration. The number of controllable modes depends on the memory of the GPU of 12 GB which needs to accommodate the NN model and a single batch of training data, in our case typically 150 frames, but the memory could be expanded by using multiple GPUs.

To increase the enhancement and light control one could in addition modulate phase (see p. 42 of \([1]\) for the effect on enhancement and signal-to-noise ratio) and the NN approach could be extended to any combination of stimulus-response pairs, including phase or polarization on either the detection or projection side, or both. An advantage of only using binary intensity modulation and intensity measurements \([20]\), is that it simplifies the setup compared to approaches that also rely on phase information. Additionally, although we used monochromatic light, our approach could also be used with pulsed light \([74]\).

For applications, the time it takes to achieve light control is critical. For the transmission matrix approach as well as the NN approach this time can be broken down into two parts, the time for acquiring the data and the time to compute the wavefront correction. The acquisition time is ultimately limited by the number of required frames. Typical numbers for the transmission matrix approach in recent reports range from 4000 \([29]\) to 12000 \([70]\), which is similar to the number of frames used in our experiments, which was typically 10000 or less, see Fig. 4(b). The time-limiting factor in our experiments was training of the NNs. For the largest data sets the time required for training was less than 35 seconds for the SLNNs and less than 50 seconds for fine-tuning the CNNs. To accelerate the process we tested training with a reduced amount of data (Supplementary
Fig. 4) which sped up training at the cost of lower enhancement. The shortest training time on a single GPU with the SLNN that lead to a focus with significant enhancement ($\eta > 10$) was obtained with 5000 frames in 18 seconds. For comparison, the time required for calculating the transmission matrix varies for different techniques, from a simple Hermitian conjugation operation, to computationally more demanding approaches which require 15 seconds matrix multiplication on a GPU [70]. While some methods that optimize a single mode can be very fast (for example 33.8 ms in [11]), this still results in a comparable correction time for a full field of view (of about 5 minutes for 96 × 96 pixels in this example).

Further optimization of the NN approach could be achieved by optimizing the network architectures. Here, we compared two basic networks, SLNNs and three-layer CNNs. SLNNs take advantage of the linearity of scattering (as does the transmission matrix approach) and therefore can generalize from speckle patterns to arbitrary light distributions. Multi-layer NNs in contrast need to be specifically designed and trained to generate a desired type of light distribution. At the same time, that multi-layer NNs are independent of assumptions about the underlying physical model (such as linearity of scattering) and can efficiently reduce the dimensionality of the images through convolutional layers as well as lower the number of parameters required for training (by 80% compared to the SLNNs in our case), will likely prove advantageous for applications.

For many applications of light control through scattering media, such as imaging, sensing or communication, it will be necessary to develop methods that can work with reflected light [50]. For example in biological microscopy, fluorescence signals can serve as feedback for scattering correction [219], but they require labeling of the sample and are often dim, particularly before wavefront correction. Other schemes for light control in tissue resort to the assistance of acoustics waves [575] but do not achieve diffraction limited optical resolution [19]. The most broadly applicable implementation for wavefront correction takes advantage of backscattered light as for example in optical coherence tomography or related approaches [52-55]. However, ultimately the availability of weakly scattered photons is limiting the imaging depth of these methods and ways to take advantage of strongly scattered light are therefore needed. Strategies for light control using strongly scattered, reflected light have indeed been developed [61,63] for maximizing the energy delivered into the material [61,62] or an embedded strongly scattering target [63] without, however, exerting full independent control over the transmitted modes.

We here took advantage of mutual information between transmitted and reflected speckle patterns [50,51] 60 and used NNs to show that it is indeed possible to control transmitted light with reflected light with sufficient accuracy for high-resolution focusing and scanning (Fig. 5). We achieved this by establishing an explicit functional relation (with NNs) between transmitted and reflected speckle patterns (Fig. 5) with sufficient accuracy for high-resolution transmission control. That such a relationship can be established (with a linear network) could not necessarily be expected based on the mutual information relationship in [50].

The limitation of the current approach for applications is that it first requires characterizing the transmission and reflection properties of the scatterer for the specific field of view, which still requires unobstructed access to the focal plane behind the scatterer. How could this limitation be overcome? One of the distinctions of neural networks is their ability to generalize. One potential avenue would therefore be to train appropriate NN models on sufficiently broad training sets and to adapt these models to the specific sample or field of view, e.g. using backscattered light. For example, CNNs are the building blocks for many of the more advanced network techniques that analyze novel visual scenes based on previously learned data sets [32,33,66], and such methods might also be harnessed for light scattering.

Independent of this, the simplicity, effectiveness and flexibility of the method presented here makes it suitable for many different applications that can take advantage of scattering control in transmission or now also in reflection (for example super-resolution imaging [76]), as well as for the further analysis of the relationship between transmission and reflection in scattering materials.

**METHODS**

**Experimental setup**

The laser beam ($\lambda = 640$ nm, power up to $P = 100$ mW; iBeamSmart, Toptica) is expanded with a telescope ($f_1 = 15$ mm, $f_2 = 150$ mm) and sent to the SLM. The SLM is a high-speed digital micromirror device (DMD, 768 × 1024 pixels, pixel size = 13.7 $\mu$m$^2$; model V-7000 from Vialux) allowing binary amplitude modulation at a maximum frame rate of 22.7 kHz and is used to display binary pseudo-random checkerboard or Hadamard patterns with typically 64 × 64 macropixels extending over the central 768 × 768 pixels of the DMD (12 × 12 micromirrors per macropixel). Two additional lenses ($f_3 = 200$ mm, $f_4 = 50$ mm) combined with a pinhole are used after the DMD to filter the maximum-intensity diffraction order mode and to demagnify and image the DMD onto the back aperture of the microscope objective (10X, 0.25 NA, Olympus). The objective focuses the light beam through the scatterer (a glass diffuser, Thorlabs DG20-120, a step-index multimode fiber optic patch cable, Thorlabs M38L02; and a piece of white paper) and a second identical microscope objective is used to collect the scattered light. Finally, a pair of lenses ($f_5 = 100$ mm, $f_6 = 75$ mm) in 2f configuration images the back aperture of the second microscope objective onto the CCD camera (acA640-750um, Basler), with
a frame rate of 500 fps at full resolution of 480 × 640 pixels (pixel size 4.8 μm²). Both microscope objectives and the scattering material are mounted on XYZ stages (omitted in Fig. [1] for aligning the system and moving the sample to different positions, as well as for displacing the image plane axially. In our experiments typically 10000 checkerboard patterns are uploaded to the internal memory of the DMD. Then, the projection of a pattern on the DMD triggers the frame capture of the CCD camera. The maximum frame rate of the DMD is 22.7 kHz and the maximum frame rate of the CCD camera is about 1000 fps at a resolution of 96 × 96 pixels, which allow us to record the whole sequence in about 10 s. We also note that our approach is valid for larger fields of view than those shown in the main figures (20 × 20 μm², see Supplementary Information and Supplementary Fig. 1). For experiments with reflected light, a non-polarizing beam splitter redirects the backscattered speckles towards a pair of lenses (f₇ = 50 mm, f₈ = 25 mm) in 2f configuration that image the back aperture of the first microscope objective onto a CCD camera identical to and synchronized with the one used to capture the transmitted speckles.

**Enhancement factor metric**

The quality of the generated foci is analyzed with an automated procedure that generates spots at different positions (up to 25) placed in a grid throughout the whole field of view and measures the enhancement, defined as:

\[
\eta \equiv \frac{I_{\text{focus}}}{\langle I_{\text{speckle}} \rangle},
\]

where \(I_{\text{focus}}\) is the intensity at the generated foci and \(\langle I_{\text{speckle}} \rangle\) is the mean value of the background speckle.

**Computer specifications**

Our computer has a Linux-Ubuntu operating system, an Intel Xeon CPU E5-1620 v4 @ 3.50 GHz, 32Gb of DDR5 RAM memory, and a Nvidia Titan XP GPU possessing 3840 CUDA cores running at 1.60 GHz and with 12Gb of GDDR5X memory running at over 11 Gbps.

**Neural network design and performance**

We use the Keras library [77] with the TensorFlow [78] back-end for GPU-accelerated neural network training. The networks are trained to map grayscale speckle images to the corresponding binary illumination patterns with a subset of the total dataset of image pairs (8000 pairs in our case) and tested on the previously not introduced data (the remaining 2000 pairs). Once the network is trained, we input the desired PSF and the output binary map is uploaded to the DMD for light control through the diffuser or fiber.

The SLNN we used is a single-layer perceptron, which is a network consisting of one fully connected layer followed by a non-linear activation function bounding the output to the 0-1 range. In principle, it can be represented as a matrix dot product, with bias addition and a sigmoid function applied element-wise to the resulting vector. We found that with the activation function applied per each individual element the model is prone to over-fitting and does not make good generalizations. As a solution, we replaced the nonlinear activation function with a binarization function with a threshold common for the whole predicted pattern (mean value of the prediction) which results in a more robust model with better focus enhancement and faster training. The training time depends on the number of images used (8000 in our case), the batch size (number of images taken for each iteration of the training algorithm, 150 in our case), and the number of epochs (up to 20 for the results presented here, see Supplementary Figs. 3 and 4 for SLNN training performance). With these parameters the single-layer perceptron requires less than 35 seconds for training, while the predicted patterns take about 1 s to be calculated. However, we have verified that lower training times with acceptable enhancement can be obtained by reducing the number of image pairs used in the training. For an analysis on how the training times and enhancement of the focus depend with the number of images pairs used in the training, see Supplementary Fig. 5.

For exploring functional relations between transmitted and reflected speckle patterns we concatenated two SLNNs similar to the one described above. The first SLNN (SLNN1) uses the transmitted speckle as input and connects it to the reflected speckle pattern (output) through a single fully-connected layer (without binarization). The second SLNN (SLNN2) connects the reflected speckle (input) to the illumination patterns (output) also through a fully connected layer including the mean threshold binarization. Both sets of speckle patterns can be generated with independent checkerboard illumination patterns. For training of these two SLNNs we used the same number of images, batch size, and number of epochs as for the SLNN discussed above, with similar performance. The desired illumination in transmission is finally fed into SLNN1, which predicts a speckle pattern as output, which in turn serves as input for SLNN2. The output of SLNN2 is a binary illumination pattern that is sent to the DMD in order to experimentally obtain the desired illumination.

In perceptron-like models a single fully-connected layer contains a large number of parameters (the product of input and output vector dimensions) which makes these models more demanding to train as the resolution of the illumination and speckle images and the memory demand increase. CNNs can efficiently reduce the number of trainable parameters and we used a model with three convolutional layers with \(48\ (9 \times 9),\ 24\ (5 \times 5)\) and \(12\ (3 \times 3)\) filters respectively, each succeeded by rectified linear unit (ReLU) activation and \((2 \times 2)\) max pooling.
operation, followed by a fully connected layer with 0.25 dropout rate. This configuration achieves a performance similar to the SLNN in controlling a single focal spot while having 20% of the SLNN’s number of parameters.

As any deeper network, the CNN requires longer training time and a more extensive dataset. A workaround is offered by the fine tuning technique: the convolutional layers are pretrained separately on a dataset of 40000 speckle images in an autoencoder. An autoencoder is a network trained to map its input to itself, however it contains a bottleneck - a lower dimensional middle layer (latent space) where a compressed representation of the data is learned. Our autoencoder has three convolutional layers as needed for the proposed CNN model and a symmetrical deconvolutional decoder. The training time largely varies with dataset size and speckle image resolution, and it is best to provide as much data as possible. Good training results were achieved after 20 minutes of training with 40000 samples sized 256 × 256. The illumination-predicting CNN is then constructed from the pretrained encoder and an untrained dense output layer. During the training through backpropagation, predominantly the output layer is adjusted, while the convolutional layers are already initialized with adequate values to compress the speckle images and extract features. This reduces the CNN training time to under 50 seconds while using the same 8000 image-pair dataset, comparable to the SLNN.

**ACKNOWLEDGEMENTS**

We thank Andres Flores for Python support and Bernd Scheiding for electronics support. We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research. This work was supported by the Max Planck Society and the Center of Advanced European Studies and Research (caesar).

**AUTHOR CONTRIBUTIONS**

A.T. and I.V. performed the experiments and analyzed the data. I.V. designed the neural network analysis. A.T., I.V., and J.D.S. designed the project and wrote the manuscript.

**COMPETING INTERESTS**

The authors declare no competing financial interests.

---

[1] Rotter, S. & Gigan, S. Light fields in complex media: Mesoscopic scattering meets wave control. *Reviews of Modern Physics* **89**, 015005 (2017).

[2] Ji, N. Adaptive optical fluorescence microscopy. *Nature Methods* **14**, 374–380 (2017).

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

[4] Čižmár, T., Mazilu, M. & Dholakia, K. In situ wavefront correction and its application to micromanipulation. *Nature Photonics* **3**, 388–394 (2010).

[5] H. W. Ruan et al. Deep tissue optical focusing and optogenetic modulation with time-reversed ultrasonically encoded light. *Science Advances* **3**, eaao5520 (2017).

[6] Choi, Y. et al. Scanner-free and wide-field endoscopic imaging by using a single multimode optical fiber. *Physical Review Letters* **109**, 203901 (2012).

[7] Forbes, A., Dudley, A. & McLaren, M. Creation and detection of optical modes with spatial light modulators. *Adv. Opt. Photon* **8**, 200–227 (2016).

[8] Turtsev, S. et al. Comparison of nematic liquid-crystal and dmd based spatial light modulation in complex photonics. *Optics Express* **25**, 29874–29884 (2017).

[9] Vellekoop, I. M. & Mosk, A. P. Focusing coherent light through opaque strongly scattering media. *Optics Letters* **32**, 2309–2311 (2007).

[10] Popoff, S. M. et al. Measuring the transmission matrix in optics: An approach to the study and control of light propagation in disordered media. *Physical Review Letters* **104**, 100601 (2010).

[11] Conkey, D. B., Caravaca-Aguirre, A. M. & Piestun, R. High-speed scattering medium characterization with application to focusing light through turbid media. *Optics Express* **20**, 1733–1740 (2012).

[12] Leonardo, R. D. & Bianchi, S. Hologram transmission through multi-mode optical fibers. *Optics Express* **19**, 247–254 (2011).

[13] Kim, D. et al. Toward a miniature endomicroscope: pixelation-free and diffraction-limited imaging through a fiber bundle. *Optics Express* **39**, 1291–1294 (2014).

[14] Ruan, H. et al. Deep tissue optical focusing and optogenetic modulation with time-reversed ultrasonically encoded light. *Science Advances* **3**, eaao5520 (2017).

[15] Yoon, J. et al. Optogenetic control of cell signaling pathway through scattering skull using wavefront shaping. *Scientific Reports* **3**, 13289 (2015).

[16] Wang, D. et al. Focusing through dynamic tissue with millisecond digital optical phase conjugation. *Optica* **2**, 728–735 (2015).

[17] Liu, Y., Lai, P., Ma, C., Xu, X. & andLihong V. Wang, A. A. G. Optical focusing deep inside dynamic scattering media with near-infrared time-reversed ultrasonically encoded (true) light. *Nature Communications* **6**, 5904 (2015).

[18] Vellekoop, I. M. Feedback-based wavefront shaping. *Optics Express* **23**, 12189–12206 (2015).

[19] Horstmeyer, R., Ruan, H. & Yang, C. Guidestar-assisted wavefront-shaping methods for focusing light into biological tissue. *Nature Photonics* **9**, 563–571 (2015).
[20] Mosk, A. P., Lagendijk, A., Lerosey, G. & Fink, M. Controlling waves in space and time for imaging and focusing in complex media. *Nature Photonics* **6**, 283–292 (2012).

[21] Tanga, J., Germain, R. N., & Cui, M. Superpenetration optical microscopy by iterative multiphoton adaptive compensation technique. *Proceedings of the National Academy of Sciences* **129**, 8434–8439 (2012).

[22] Blochet, B., Bourdieu, L. & Gigan, S. Focusing light through dynamical samples using fast continuous wavefront optimization. *Optics Letters* **42**, 4994–4997 (2017).

[23] Cui, M. & Yang, C. Implementation of a digital optical phase conjugation system and its application to study the robustness of turbidity suppression by phase conjugation. *Optics Express* **18**, 3444–3455 (2010).

[24] Hsieh, C.-L., Pu, Y., Grange, R., Laporte, G. & Psaltis, D. Imaging through turbid layers by scanning the phase conjugated second harmonic radiation from a nanoparticle. *Optics Express* **18**, 20723–20731 (2010).

[25] Wang, Y. M., Judkewitz, B., DiMarzio, C. A. & Yang, C. Deep-tissue focal fluorescence imaging with digitally time-reversed ultrasound-encoded light. *Nature Communications* **3**, 928 (2012).

[26] Hillman, T. R. et al. Digital optical phase conjugation for delivering two-dimensional images through turbid media. *Scientific Reports* **3**, 1909 (2013).

[27] Liu, Y., Ma, C., Shen, Y., Shi, J. & Wang, L. V. Focusing light inside dynamic scattering media with millisecond digital optical phase conjugation. *Optica* **4**, 280–288 (2017).

[28] Popoff, S., Lerosey, G., Fink, M., Boccara, A. C. & Gigan, S. Image transmission through an opaque material. *Nature Communications* **1**, 81 (2010).

[29] Boniface, A., Mounaix, M., Blochet, B., Piestun, R. & Gigan, S. Transmission-matrix-based point-spread-function engineering through a complex medium. *Optica* **4**, 54–59 (2017).

[30] Drémeau, A. et al. Reference-less measurement of the transmission matrix of a highly scattering material using a dmd and phase retrieval techniques. *Optics Express* **23**, 11898–11911 (2015).

[31] Metzler, C. A. et al. Coherent inverse scattering via transmission matrices: Efficient phase retrieval algorithms and a public dataset. In *2017 IEEE International Conference on Computational Photography (ICCP)*, 1–16 (2017).

[32] Jordan, M. I. & Mitchell, T. M. Machine learning: Trends, perspectives, and prospects. *Science* **349**, 255–260 (2015).

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

[34] Waller, L. & Tian, L. Machine learning for 3d microscopy. *Nature* **523**, 416–417 (2015).

[35] Deans, C., Griffin, L. D., Marmugi, L., & Renzoni, F. Machine learning based localization and classification with atomic magnetometers. *Physical Review Letters* **120**, 033204 (2018).

[36] Kamilov, U. S. et al. Learning approach to optical tomography. *Optica* **2**, 517–522 (2015).

[37] Sinha, A., Li, J. L. S. & Barbastathis, G. Lensless computational imaging through deep learning. *Optica* **4**, 1117–1125 (2017).

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

[39] Nehme, E., Weiss, L. E., Michaeli, T., & Shechtman, Y. Deep-storm: Super resolution single molecule microscopy by deep learning. *arXiv preprint arXiv:1801.09631v1* (2018).

[40] Conkey, D. B., Brown, A. N., Caravaca-Aguirre, A. M. & Piestun, R. Genetic algorithm optimization for focusing through turbid media in noisy environments. *Journal of Optics* **20**, 4840–4849 (2012).

[41] Zhang, X. & Kner, P. Binary wavefront optimization using a genetic algorithm. *Journal of Optics* **17**, 125704 (2014).

[42] Tehrani, K. F., Xu, J., Zhang, Y., Shen, P. & Kner, P. Adaptive optics stochastic optical reconstruction microscopy (ao-storm) using a genetic algorithm. *Optics Express* **23**, 13677–13692 (2015).

[43] Zhang, B. et al. Focusing light through strongly scattering media using genetic algorithm with sbr discriminant. *Journal of Optics* **20**, 025601 (2017).

[44] Ando, T., Horisaki, R. & Tanida, J. Speckle-learning-based object recognition through scattering media. *Optics Express* **23**, 33902–33910 (2015).

[45] Horisaki, R., Takagi, R. & Tanida, J. Learning-based imaging through scattering media. *Optics Express* **24**, 13738–13743 (2016).

[46] Horisaki, R., Takagi, R. & Tanida, J. Learning-based focusing through scattering media. *Applied Optics* **56**, 4358–4362 (2017).

[47] Li, S., Deng, M., Lee, J., Sinha, A. & Barbastathis, G. Imaging through glass diffusers using densely connected convolutional networks. *arXiv preprint arXiv:1711.06810v1* (2017).

[48] Lyu, M., Wang, H., Li, G. & Situ, G. Exploit imaging through opaque wall via deep learning. *arXiv preprint arXiv:1711.06810v1* (2017).

[49] J. R. P. Angel, Wizinowich, P., Lloyd-Hart, M. & Sandler, D. Adaptive optics for array telescopes using neural-network techniques. *Nature* **348**, 221–224 (1990).

[50] N. Fayard, Goetschy, A., Pierrat, R. & Carminati, R. Mutual information between reflected and transmitted speckle images. *Physical Review Letters* **120**, 073901 (2018).

[51] Zhe Liu, Lavis, L. D. & Betzig, E. Imaging live-cell dynamics and structure at the single-molecule level. *Molecular Cell* **58**, 644–659 (2015).

[52] M. Feierabend, Ruckel, M. & Denk, W. Coherence-gated wave-front sensing in strongly scattering samples. *Optics Letters* **29**, 2255–2257 (2004).

[53] R. Fiolka, Si, K. & Cui, M. Parallel wavefront measurements in ultrasound pulse guided digital phase conjugation. *Optics Express* **20**, 24827–24834 (2012).

[54] Drexler, W. & Fujimoto, J. G. Optical Coherence Tomography (Springer International Publishing, 2015).

[55] S. Kang et al. Imaging deep within a scattering medium using collective accumulation of single-scattered waves. *Nature Photonics* **9**, 253–258 (2015).

[56] A. Badon et al. Smart optical coherence tomography for ultra-deep imaging through highly scattering media. *Science Advances* **2**, e1600370 (2016).

[57] S. Kang et al. High-resolution adaptive optical imaging within thick scattering media using closed-loop accumulation of single scattering. *Nature Communications* **8**, 2157 (2017).

[58] M. Kadobianskyi, Papadopoulos, I. N., Chaigne, T., Horstmeyer, R. & Judkewitz, B. Scattering correlations
of time-gated light. *Optica* 5, 389–394 (2018).

[59] N. Fayard, Caze, A., Pierrat, R. & Carminati, R. Intensity correlations between reflected and transmitted speckle patterns. *Physical Review A* 92, 033827 (2015).

[60] I. Starshynov et al. Non-gaussian correlations between reflected and transmitted intensity patterns emerging from opaque disordered media. *Physical Review X* 8, 021041 (2018).

[61] Youngwoon Choi et al. Measurement of the time-resolved reflection matrix for enhancing light energy delivery into a scattering medium. *Physical Review Letters* 111, 243901 (2013).

[62] Hyeonseung Yu, Park, J. H. & Park, Y. Measuring large optical reflection matrices of turbid media. *Optics Communications* 352, 33–38 (2015).

[63] Seungwon Jeong et al. Focusing of light energy inside a scattering medium by controlling the time-gated multiple light scattering. *Nature Photonics* 12, 277–283 (2018).

[64] Curtis Jin, Nadakuditi, R. R., Michielsen, E. & Rand, S. C. Iterative, backscatter-analysis algorithms for increasing transmission and focusing light through highly scattering random media. *J. Opt. Soc. Am. A* 30, 1592–1602 (2013).

[65] Curtis Jin, Nadakuditi, R. R., Michielsen, E. & Rand, S. C. Backscatter analysis based algorithms for increasing transmission through highly scattering random media using phase-only-modulated wavefronts. *J. Opt. Soc. Am. A* 31, 1788–1800 (2014).

[66] Jurgen Schmidhuber. Deep learning in neural networks: An overview. *Neural Networks* 61, 85–117 (2014).

[67] Henderson, K., Ryu, C., MacCormick, C. & Boshier, M. G. Experimental demonstration of painting arbitrary and dynamic potentials for bose–einstein condensates. *New Journal of Physics* 11, 043030 (2009).

[68] John Peter Rickgauer & Tank, D. W. Two-photon excitation of channelrhodopsin-2 at saturation. *PNAS* 106, 15025–15030 (2009).

[69] Caravaca-Aguirre, A. M. & Piestun, R. Single multimode fiber endoscope. *Optics Express* 25, 1656–1665 (2017).

[70] Ohayon, S., Caravaca-Aguirre, A. M., Piestun, R. & DiCarlo, J. J. Deep brain fluorescence imaging with minimally invasive ultra-thin optical fibers. *arXiv preprint arXiv:1703.07633* (2017).

[71] Goshtasby, A. A. *Image Registration* (Springer, 2012).

[72] Damien Loterie, D. P. C. M., Salma Farahi. Complex pattern projection through a multimode fiber (2015).

[73] Akbulut, D., Huisman, T. J., van Putten, E. G., Vos, W. L. & Mosk, A. P. Focusing light through random photonic media by binary amplitude modulation. *Optics Express* 19, 4017–4029 (2011).

[74] Sun, B. *et al.* Four-dimensional light shaping: manipulating ultrafast spatiotemporal foci in space and time. *Light: Science and Applications* 7, 17117 (2018).

[75] X. A Xu, Liu, H. L. & Wang, L. V. Time-reversed ultrasonically encoded optical focusing into scattering media. *Nature Photonics* 5, 154–157 (2011).

[76] E. G. van Putten et al. Scattering lens resolves sub-100 nm structures with visible light. *Physical Review Letters* 106, 193905 (2011).

[77] Chollet, F. *et al.* Keras. *https://github.com/keras-team/keras* (2015).

[78] Abadi, M. *et al.* TensorFlow: Large-scale machine learning on heterogeneous systems (2015). *URL* https://tensorflow.org/
FIG. 1. (a) Approach for light control through scattering media with NNs. A NN is trained with pairs of illumination and speckle patterns (illustration, see supplementary Fig. 1 for examples of actual illumination and speckle patterns), using the speckle patterns as input of the network and the illumination as output. Once the NN is trained, it is used to predict the illumination necessary to generate a target pattern after the scattering material. The predicted illumination is subsequently sent through the material resulting in the desired light pattern. (b) Experimental setup. An expanded laser beam at 640 nm illuminates the SLM (in our case a DMD) allowing for binary modulation of the light beam. After filtering the maximum-intensity diffraction order with a pair of lenses and a pinhole (PH), the DMD pattern is sent to the scattering material (SM) with a 10× microscope objective (MO, NA = 0.25). An additional identical objective is used to collect the transmitted light and is imaged onto a CCD camera with an additional pair of lenses. Both objectives and the sample are mounted on XYZ translation stages (omitted in the figure). A non-polarizing beam splitter (NPBS), a pair of lenses and a CCD camera are used to retrieve speckle patterns reflected by the sample for experiments with combined transmission and reflection.
FIG. 2. Focusing with neural networks. Top: Illustration of SLNN (a) or CNN (b) with speckle and illumination pattern. Bottom: Intensity distributions (first row) and intensity profiles through the foci along horizontal (second row) and vertical (third row) directions of Gaussian beams obtained at different positions after training (a) the SLNN and (b) the CNN. Red-dashed lines (- -) are the targeted intensity distributions that enter the NN, normalized to the corresponding experimental result. Scale bars = 2 μm. FCL: fully-connected layer; Conv $n \times m \times p$: convolutional layer of $p$ kernels with dimensions $n \times m$; MP: max pooling operation reducing the previous element size. Color bars: intensity (a.u.).
FIG. 3. **Light control through a glass diffuser with a SLNN.** Normalized intensity patterns obtained after the glass diffuser with the SLNN: (a) five Gaussian foci; (b) a line at 45°; (c) a ring; (d) the letter "E"; (e) the number "5"; and (f) the letters "np". Insets show the desired light distribution. Scale bars = 2 µm. Color bars: intensity (a.u.) normalized for each image.

FIG. 4. **Neural networks focus light through multimode fibers.** Normalized transverse maximum intensity projection of the light field after a multimode fiber when (a) no correction is applied and when a single focus is scanned (maximum intensity projection) along (b) a circle, (c) a square, and (d) an array of 5 × 5 points. Scale bars = 2.3 µm. Color bar: intensity (a.u.) normalized for each image.
FIG. 5. Neural networks find functional relationships between transmitted and reflected speckle patterns. (a) Histogram of the Pearson correlation coefficient \( \rho_{r,p} \) between the measured reflected and transmitted speckle patterns (in blue) and the measured and predicted reflected speckle patterns (orange). Number of bins: 30. Total number of pairs of samples used: 2000. Bottom: Examples of the normalized transverse intensity distributions of the light field of the (b) measured reflected speckle pattern, (c) predicted reflected speckle pattern, and (d) measured reflected speckle pattern while focusing in transmission (e).
Sequential neural networks

(a)

SLNN1

SLNN2

NN input

NN output

Reflection

Transmission

Illumination

Beam splitter

Scattering material

FIG. 6. **Focusing and scanning in transmission using reflected light.** Illustration of network approach to control transmitted light using reflected light. A SLNN is trained to learn the relationship between simultaneously recorded transmitted and reflected speckle patterns. After training of the network, it is sufficient to train a second SLNN to relate reflection to illumination for controlling light through the combined network. Bottom: Normalized transverse maximum intensity projection of the light field after a sheet of paper when a single focus is scanned (maximum intensity projection) along (a) a circle, (b) a square, and (c) an array of $5 \times 5$ points. Scale bars = 2.7 $\mu$m. Color bar: intensity (a.u.) normalized for each image.
LIGHT CONTROL OVER DIFFERENT FIELDS OF VIEW

In the main text we have shown the ability of NNs to shape light through disordered media within a field of view of $20 \times 20 \mu m^2$ imaged onto $96 \times 96$ pixels on the CCD camera. However, the presented method performs similarly for other fields of view, as shown in Supplementary Figure 1 for the case of the glass diffuser, with $64 \times 64$ macropixels displayed on the DMD controlled with the SLNN. In all cases, the experimental setup is the same as the one shown in the main text and the only difference is the resolution of the images captured with the CCD camera and used to train the SLNN. Note that the illumination resolution additionally impacts the quality of light shaping through the diffuser (in agreement with previous reports, see e.g. Ref. [9]).

PATTERN GENERATION BY SCANNING A FOCUS AT HIGH FREQUENCY

Since neural networks can generate single foci with high fidelity, one can take advantage of the fast operation of the DMD (22.7 kHz) to obtain any transverse intensity distribution after the scatter by scanning a single spot (or multiple spots) at high speed. This is shown in Supplementary Fig. 1 for a single focus tracing out (a/d) a circle, (b/e) a square, and (c/f) a square grid consisting of 128/96, 256/256, and 25/25 scanning positions, respectively, through a glass diffuser both for the SLNN (first row) and the CNN (second row). Supplementary Movies 1–6 show projections of these sequences recorded at a speed of 500 Hz (the fastest allowed by our CCD camera). Note that for the DMD operating at full speed, i.e. without restrictions imposed by the camera, a sequence composed of 96 focus positions projecting a certain pattern will result in a pattern projection frame rate higher than 200 Hz. An analogous image for focusing through paper is shown in Supplementary Figure 2.

TRAINING PERFORMANCE OF SLNN

Although neural networks are capable of predicting the illumination patterns necessary for light control through scattering media, there are multiple variables affecting the efficiency of the training. Here, we discuss the impact of the size of the dataset used for training. In Supplementary Fig. 3 we plot the mean-square error (MSE) between the predicted illuminations after training the single-layer neural network and a set of 100 original illuminations that were not included in the training. The analysis is performed for different sizes of the dataset (ranging from 800 pairs to 8000 pairs) and different illumination sizes: $64 \times 64$ (red), $32 \times 32$ (green), and $16 \times 16$ (blue). As an illustration of the training performance, we have included an example of original illumination and corresponding predictions obtained when 1600, 3200, 4800, 6200, and 8000 pairs are used ($16 \times 16$ case). As expected, the lower the size of the illumination pattern, the higher the MSE.

In Figure 3 we further explore the training performance of the SLNN as the size of the dataset used in the training varies, in this case in terms of (a) training time, and (b) enhancement of the generated foci. While the training time shows a linearly increasing tendency, the enhancement tends to saturate beyond a certain amount of samples (typically for $N > 5000$).

SUPPLEMENTARY MOVIE CAPTIONS

Note: for all movies intensity is normalized for each frame.

Supplementary Movie 1

A single focus generated with a SLNN is scanned at 128 different positions forming a circle through a glass diffuser at a projection rate of 500 Hz.

Supplementary Movie 2

A single focus generated with a SLNN is scanned at 256 different positions forming a square through a glass diffuser at a projection rate of 500 Hz.

Supplementary Movie 3

A single focus generated with a SLNN is scanned at 25 different positions forming a grid through a glass diffuser at a projection rate of 500 Hz.

Supplementary Movie 4

A single focus generated with a CNN is scanned at 96 different positions forming a circle through a glass diffuser at a projection rate of 500 Hz.

Supplementary Movie 5

A single focus generated with a CNN is scanned at 256 different positions forming a square through a glass diffuser at a projection rate of 500 Hz.
Supplementary Movie 6

A single focus generated with a CNN is scanned at 25 different positions forming a grid through a glass diffuser at a projection rate of 500 Hz.

Supplementary Movie 7

A single focus generated with a SLNN is scanned at 96 different positions forming a circle through a multimode fiber at a projection rate of 500 Hz.

Supplementary Movie 8

A single focus generated with a SLNN is scanned at 256 different positions forming a square through a multimode fiber at a projection rate of 500 Hz.

Supplementary Movie 9

A single focus is scanned at 25 different positions forming a grid through a multimode fiber at a projection rate of 500 Hz.

Supplementary Movie 10

A single focus generated with a SLNN is scanned at 96 different positions forming a circle through paper at a projection rate of 500 Hz.

Supplementary Movie 11

A single focus generated with a SLNN is scanned at 256 different positions forming a square through paper at a projection rate of 500 Hz.
FIG. 1. **Light control over different fields of view.** Normalized transverse intensity distributions of an example of speckle pattern (first column), a single focus (second column), the number “1” (fourth column), and the number “5” (6th column) for different fields of view: 256 × 256 pixels (first row), 128 × 128 pixels (second row), 96 × 96 pixels (third row), and 64 × 64 pixels (last row). Columns number 3, 5, and 7 are the actual DMD patterns used to generate the light distributions from columns number 2, 4, and 6, respectively. Color bars: intensity (a.u.) normalized for each image.
FIG. 2. Single-focus scanning allows time-averaged pattern projection. Patterns obtained when a single focus is scanned following (a/d) a circle (128/96 scanning points), (b/e) a square (256/256 scanning points), and (c/f) a grid of $5 \times 5$ points with the SLNN (first row) and the CNN (second row). Color bars: intensity (a.u.) normalized for each image. Scale bars: 2.3 $\mu$m

FIG. 3. Single-focus scanning allows time-averaged pattern projection through paper. Maximum intensity projections of patterns obtained when a single focus is scanned following (a) a circle (128 scanning points), (b) a square (256 scanning points), and (c) a grid of $5 \times 5$ points with the SLNN and paper as scattering material. Color bars: intensity (a.u.) normalized for each image. Scale bars: 2.7 $\mu$m
FIG. 4. **Training performance of SLNN.** Mean-square error (MSE) between the predicted and original illumination patterns for the single-layer neural network for different sizes of the dataset. Red-, green-, and blue-dashed curves correspond to illumination sizes of $64 \times 64$, $32 \times 32$, and $16 \times 16$, respectively. Insets below show the the predicted illumination at different stages of the training for the $16 \times 16$ case.

FIG. 5. **Speed of SLNN training and quality of focus.** (a) Training time of the SLNN and (b) enhancement $\eta$ of the generated foci with paper as scattering sample for different values of the number of image pairs used during training. The enhancement is defined as $\eta \equiv I_{\text{focus}}/\langle I_{\text{speckle}} \rangle$, where $I_{\text{focus}}$ is the intensity at the generated foci and $\langle I_{\text{speckle}} \rangle$ is the mean value of the background speckle [73]. Image in the lower part of the figure are examples of focusing with $N$ samples used for training (with the number $N$ of samples used for training at the top of each image.