Physics-based neural network for non-invasive control of coherent light in scattering media

ALEXANDRA D'ARCO,[1,*] FEI XIA,[1] ANTOINE BONIFACE,[1,2]
JONATHAN DONG,[1,3] AND SYLVAIN GIGAN[1,4]

[1]*Laboratoire Kastler Brossel, ENS–Universite PSL, CNRS, Sorbonne Université, Collège de France, 24 Rue Lhomond, F-75005 Paris, France*
[2]*Laboratory of Applied Photonics Devices, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland*
[3]*Biomedical Imaging Group, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland*
[4]*sylvain.gigan@lkb.ens.fr*
[*]*alexandra.darco@lkb.ens.fr*

Abstract: Optical imaging through complex media, such as biological tissues or fog, is challenging due to light scattering. In the multiple scattering regime, wavefront shaping provides an effective method to retrieve information; it relies on measuring how the propagation of different optical wavefronts are impacted by scattering. Based on this principle, several wavefront shaping techniques were successfully developed, but most of them are highly invasive and limited to proof-of-principle experiments. Here, we propose to use a neural network approach to non-invasively characterize and control light scattering inside the medium and also to retrieve information of hidden objects buried within it. Unlike most of the recently-proposed approaches, the architecture of our neural network with its layers, connected nodes and activation functions has a true physical meaning as it mimics the propagation of light in our optical system. It is trained with an experimentally-measured input/output dataset built from a series of incident light patterns and corresponding camera snapshots. We apply our physics-based neural network to a fluorescence microscope in epi-configuration and demonstrate its performance through numerical simulations and experiments. This flexible method can include physical priors and we show that it can be applied to other systems as, for example, non-linear or coherent contrast mechanisms.

1. Introduction

Coherent light is subject to scattering when it propagates through optically heterogeneous samples such as biological tissues. Ballistic light is exponentially attenuated with depth, and the remaining light is scattered giving rise to a complex interference pattern, also called speckle [1]. Most imaging techniques rely on the use of ballistic light and are therefore limited to shallow depths. In the past few years, many advances have been made for imaging at depth, mostly by controlling the phase and/or amplitude of the light being scattered thanks to spatial light modulators (SLMs) [2,3].

The first wavefront shaping experiment consisted in iteratively optimizing the SLM phase pattern using feedback metrics, such as signal strength, to focus the transmitted scattered light onto a diffraction-limited spot [4]. This was a major step forward for optical imaging in turbid medium as scanning the spot across the sample offers an image. Later on, the transmission matrix (TM) that characterizes the linear and deterministic propagation of the optical field through the medium has been experimentally measured [5,6]. Knowing the TM gives more information than optimizing the input field and can also be used for imaging [7]. However in both cases, a feedback signal from the focal plane is needed. A detector (a single pixel detector or a camera) must then have a direct access to the focal plane which is in most scenarios highly invasive. This

problem can be overcome by using guidestars, such as nonlinear signals, to provide feedback in a non-invasive way [8,9], but some contrast mechanisms, such as linear fluorescence, are left aside as not easy to use. A few linear fluorescence settings were proposed for imaging but are only able to reconstruct the superficial layers of biological samples with high resolution [10–13]. Lately, a few approaches have been developed for non-invasive imaging in scattering media with fluorescence feedback, but they are specific and not very flexible [14,15].

Meanwhile, several techniques that utilize emerging computational tools, such as neural networks, have emerged in parallel for controlling light through scattering media such as diffusers and multimode fibres [16,17]. Although these methods offer the advantage of being physics-informed, they are primarily used in transmission and remain invasive. Moreover, neural networks approaches that are non invasive often include several layers, making it more difficult to interpret the neural network physics and thus add prior physical knowledge [18].

To overcome these issues, we propose a versatile neural network-based TM retrieval method for non-invasive light focusing through scattering media [19]. The approach is based on mapping the input/output information in a microscopy configuration thanks to a 2-layer neural network. It is simple, model-based, and applicable to a variety of imaging scenarios. We present experiments on non-invasive fluorescence imaging and simulations generalizing this concept to a different contrast mechanism (Second Harmonic Generation, SHG).

2.　Principle

In the general context of non-invasive imaging in scattering media, the propagation of light can be decomposed into two steps: the forward path (from the light source to the object of interest, corresponding to the illumination) and the backward path (from the object to the detector, corresponding to the signal re-emitted by the object). Due to scattering, the light waves in both paths are strongly distorted and as such no information can be retrieved directly. Inspired from the classical experiment to measure the TM, our strategy is to introduce an SLM to modulate the incident wavefront, and simply measure the corresponding re-emitted signal from the object with a camera. Most importantly, the latter captures the light that has traveled both the forward and backward paths, unlike in a conventional TM experiment where only the forward path is characterized. As this non-invasive geometry is more complicated to describe in terms of light propagation, it can be seen as a black box describing the relation between the input (on the SLM) and output (on the camera) patterns. It is then possible to make a "digital twin" of the system by describing this black box with a multi-layer neural network with the right constraints [20]. Conventional neural networks are built from an operational layer that involves a trainable matrix-vector multiplication followed by an element-wise nonlinear activation. The weights of the matrix-vector multiplication are adjusted during training in order for the neural network to implement a desired mathematical operation [21]. In the digital twin approach, the operation is linked to the physical properties of the system. By changing the network weights, one can alter the physical transformation performed on the input data in order to learn these parameters according to the physical system. Our approach is based on this representation (see Fig. 1).

We apply this concept to a microscopy imaging system. The physical system and its neural network interpretation are depicted in Fig. 2. The weights of the network connect all inputs to outputs of the system (from SLM to camera) through a fully connected two-layer network. The weights of each layer physically correspond to the coefficients of matrices T_1 and T_2 as we can see in Fig. 2(b). The first layer connects the input field E^{in} to the excitation field E^{exc} at the object plane. The weights of this layer correspond to the complex-valued coefficients of the matrix T_1. The size of this matrix is $N_{\text{target}} \times N_{\text{SLM}}$, where N_{target} is the number of emitters and N_{SLM} the dimension of SLM patterns. The weights of the second layer represent T_2, of size $N_{\text{CAM}} \times N_{\text{target}}$ with N_{CAM} the number of camera pixels. According to the physical nature of the process, the

(a) Physical system

Imaging system

Input modes

Output modes

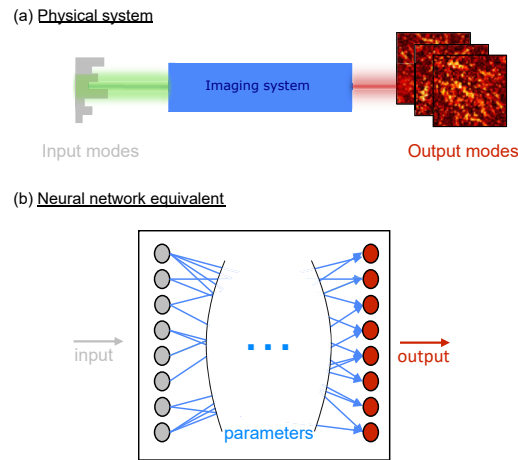(b) Neural network equivalent

input

· · ·

output

parameters

Fig. 1. Neural network representation of a physical system. (a) Physical system : from input and the features of the optical system (blue box), output modes such as images can be obtained. (b) Neural network representation of this physical system. The weight of the connection are the physical system parameters.

forward model can be written as:

$$x_2 = g_2(T_2 g_1(T_1 x_0)). \tag{1}$$

where $x_0 = E^{\text{in}}$ is the input layer, matrices T_1 and T_2 constitute the weights of the networks and g_1 and g_2 are the activation functions, defined according to the imaging process [22]. g_1 and g_2 are physical priors, they are chosen to mimic our optical system.

In order to find the weights, the network is trained with a dataset composed of p input-output pairs $\{E^{\text{in}}, I^{\text{out}}\}$. The gradient of the *Loss* function is computed, with respect to the weights of the network (i.e. back propagation [23]). As it is appropriate for regression predictive modeling problems, the loss function we used in the work is simply a L_2-norm loss:

$$Loss = ||I^{\text{out}} - x_2||^2. \tag{2}$$

After a given number of iteration corresponding to a significant decrease of the loss, the weights of the connection are trained. As a result, these weights correspond to the experimental transmission matrices of the system, T_1 and T_2. This result is experimentally validated by focusing on the fluorescence beads using the retrieved T_1 or reconstructing the image using T_2. Selective and non-invasive focusing onto beads is done by phase conjugating the retrieved T_1 [24,25]. If the focus can be obtained onto most beads, the network learned the physical transform at stake in the system with a given accuracy.

To validate the method, we implement it experimentally in a fluorescence imaging scenario also and also in simulation with a different contrast mechanism (SHG). A different contrast mechanism can be studied by simply modifying the non-linear activation functions of the network, without changing the overall 2-layer architecture. This shows the simplicity of our network and its resulting flexibility since it can be easily applied to other non-invasive model and non-linear imaging settings.

The experimental setup is shown in Fig. 2. A sparse fluorescent sample made of 1 µm beads is placed behind a scattering medium. In order for the 2-layer model to reconstruct the object, all beads are placed in the same plane. An SLM is used to modulate the phase of the incident optical field in order to send different illumination patterns to the sample. Experimentally, p random
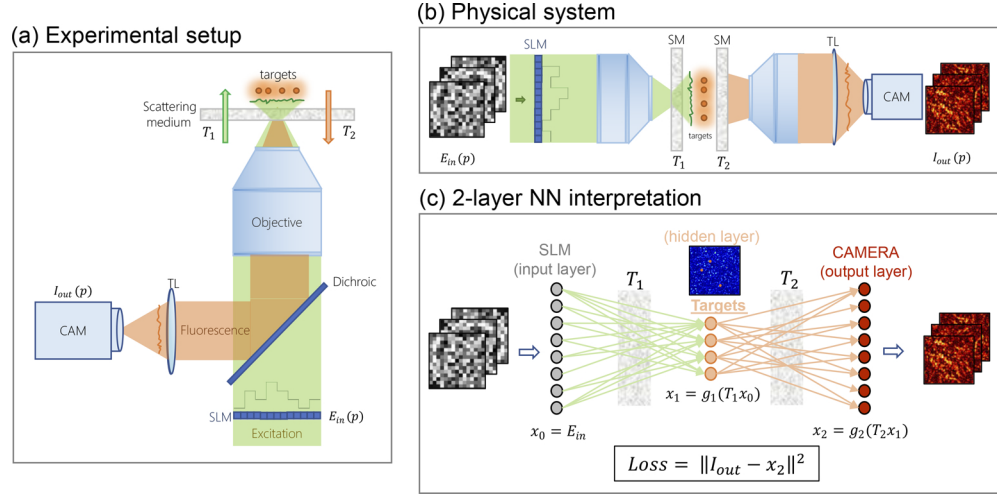
Fig. 2. Modeling fluorescence imaging setup as a Neural Network. (a) Schematic view of the non-invasive experimental setup. (b) Schematic view of the unfolded experimental setup. A randomly modulated speckle pattern illuminates a fluorescent object (beads or pollen seeds), hidden behind the scattering medium. $E^{\text{exc}} = T_1 E^{\text{in}}$ excite the object, which emits a fluorescence signal in return, that is backscattered by the medium and detected in epi-geometry on a camera : $I^{\text{out}} = T_2 |E^{\text{exc}}|^2$. TL : tube lens. (c) Neural network mimicking the experimental setup. The neural network is trained to regenerate fluorescence speckle $x_2 \approx I^{\text{out}}$ from the input pattern E^{in} displayed onto the SLM. x_2 is compared with the actual original image I^{out} (record in reflection on a camera) through the $Loss = ||I^{\text{out}} - x_2||^2$ function. Using gradient descent, back-propagation updates the weights (T_1 and T_2) to minimize the *Loss*

phase masks are displayed on the SLM, denoted $E^{\text{in}}(p)$. The phase of the N_{SLM} independent pixels is randomly chosen between 0 and 2π, following a uniform distribution. After its propagation through the scattering medium (which corresponds to a first matrix multiplication), a random speckle field $E^{\text{exc}}(p)$ is formed and illuminates the fluorescent object. During this first step, the transformation of the optical field between the SLM and the sample plane is linear and deterministic. It is represented by the matrix T_1 that maps the SLM input optical field E^{in} to the field on the sample in a linear way $E^{\text{exc}} = T_1 E^{\text{in}}$. The activation function of the first layer is then $g_1 = |.|^2$, corresponding to the fact that fluorescence signal is proportional to the intensity of the excitation. In the second step, the fluorescence signal is transmitted through the scattering medium and collected by the camera, a process that can be described by an incoherent and positive matrix T_2: as fluorescence is spatially incoherent, the captured camera image is the sum of the fluorescence speckles emitted by each fluorescent source : $I^{\text{out}} = T_2 |E^{\text{exc}}|^2$. In order to fit our physical system we ensure g_2 to be the activation function of the second layer defined by $g_2(x) = (x)$, the identity function. We define here *eigen patterns*, the speckles that are generated by each individual fluorescent emitter (of the focal spot size) on the camera. We made the assumption that the system (scattering medium and fluorescent sample) is static, therefore the *eigen patterns* are the same over time. Thus, the fluorescence signal $I^{\text{out}}(p)$ can be written as:

$$I^{\text{out}}(p) = T_2 |E^{\text{exc}}(p)|^2 = T_2 |T_1 E^{\text{in}}(p)|^2. \tag{3}$$

The experiment is completed by two sets of simulations, one to study our linear fluorescence setup, and another to simulate a different contrast mechanisms (SHG) to show how the activation functions and initial conditions on the TMs can be adjusted. Here, a numerical dataset consisting of pairs of input patterns displayed on the SLM, E^{in}, and corresponding speckle patterns

$I^{\text{out}} = g_2(T_2 g_1(E^{\text{in}}))$ are generated. The values of E^{in} are passed to the two fully connected layers which is equivalent to multiplications by the approximate matrices $\widetilde{T}_1$, $\widetilde{T}_2$. An approximate image is then obtained as $x_2 = g_2(\widetilde{T}_2 g_1(\widetilde{T}_1 x_0))$. Initially, $\widetilde{T}_1$ and $\widetilde{T}_2$ are random matrices, complex and real positive-valued, respectively. The derivatives of the *Loss* function, with respect to the elements of $\widetilde{T}_1$ and $\widetilde{T}_2$ are computed. A stochastic gradient descent approach (SGD or Adam optimizer in Pytorch) is then applied to gradually update the estimated matrices such that it effectively reduces the overall loss function, and the process is repeated for a fixed number of iterations or *epochs*, ensuring convergence of the loss function to a minimum value [26]. In this simulation, no noise is added neither on the input nor output data. We also notice that the noise does not significantly impact the reconstruction as long as enough training samples in the dataset can be obtained. (see Supplement 1 - Noise Influence to see the impact of noise on the reconstruction).

3. Results

3.1. Simulations

3.1.1. Discrete object

As a first step, we consider a linear fluorescence forward model based on the two-layer neural network architecture and a discrete object. In machine learning, a test set is typically used to evaluate the fit of a model on a training set [27]. The test set is a part of the original dataset which is set aside and used afterwards to assess the performance of the neural network. We choose a discrete fluorescent object composed of $N_{\text{target}} = 8$ targets, and set the dimension of the hidden layer to correspond to the number of targets. The dimensions of input layer and output layer are $N_{\text{SLM}} = 256$ and $N_{\text{CAM}} = 256$ respectively. Ground truths T_1 and T_2 are randomly generated following a standard normal distribution, we assume no correlation and no noise is added in the generated ground truth data. We use a training set up to $N_{\text{pat}} = 4900$ examples to estimate the weights of the matrices $\widetilde{T}_1$ and $\widetilde{T}_2$. Then a test set of 500 input examples on the SLM, E^{test} (unseen previously by the neural network) is passed to the trained network which generates 500 output images I^{test}. Correlations between I^{test} and I^{out}; T_1 and $\widetilde{T}_1$; T_2 and $\widetilde{T}_2$ are plotted with respect to $\alpha = P/N_{\text{pat}}$ on Fig. 3, where P is the training set size. This procedure is averaged over 10 repetitions for 10 different data sets (i.e different T_1 and T_2). See Supplement 1 for the choice of the rank, i.e., the size of the hidden layer.

As can be seen in Fig. 3, the loss function is minimized as the number of epochs grows. The correlations between ground truths and neural network guesses of fluorescent output speckle patterns reach a value close to 1 when the training set size increases, meaning that the transmission matrices were well retrieved by the training of the 2-layer neural network. Further, we visualize the reconstruction of both matrices, and observe that the dynamics are different based upon the size of the training set. With small training set sizes, $P<500$, $\alpha<0.09$, it seems that the best way to minimize the global loss function is to adjust the coefficients of $\widetilde{T}_2$. When backpropagating the gradient, it rapidly converges towards T_2 (correlation > 0.8 for $\alpha \simeq 0.1$). At this point ($\alpha \simeq 0.1$), $\widetilde{T}_1$ is still nearly completely random: its correlation with T_1 is around 0.12. Nevertheless, for larger training set sizes, this correlation significantly increases and reaches almost unity for $\alpha \simeq 0.5$. From the comparison between I^{test} and I^{out}, we show that this method is equivalent to performing the correlations between ground truths and predicted matrices (T_1 and $\widetilde{T}_1$, T_2 and $\widetilde{T}_2$). The use of training and testing sets allows us to implement the two-layer neural network method on experimental data since we can verify the reconstruction of the matrices without knowing the ground truth for T_1 and T_2.
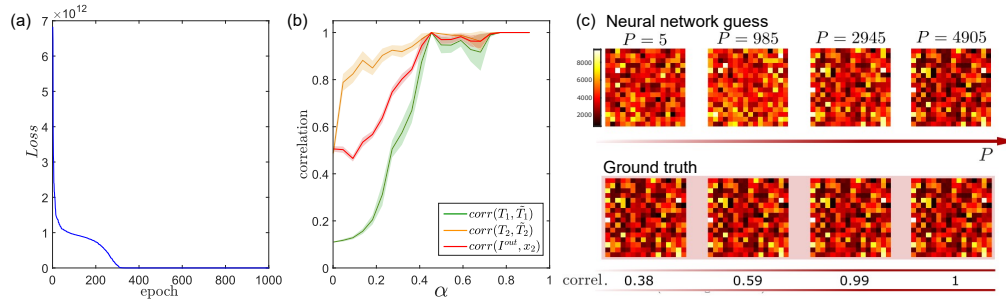
Fig. 3. Simulation results of transmission matrices retrieval. (a) Loss decrease after 1000 epoch of gradient descent for a given training set size. (b) Correlation between the ground truth I^{out} and the neural network guess I^{test}, and the ground truth T_1 and the neural network guess $\widetilde{T_1}$, and finally between the ground truth T_2 and the neural network guess $\widetilde{T_1}$, according to $\alpha = P/N_{\text{pat}}$, with P the training set size. Since the procedure is repeated 10 times for 10 different datasets, the plots represent the mean correlation and its standard deviation in shade. (c) Evolution of the reconstruction of the output image of the test set according to the training set size.

3.1.2. Continuous object

By increasing the dimension of the hidden layer, our physics-based neural network is able to reconstruct more complex continuous objects. We first study the case of continuous objects through simulations. To simulate the ground truth of the fluorescence imaging process, we use the code from [28]. It is simulated by an i.i.d. Gaussian complex with tunable speckle grain size (adjusted in the Fourier plane thanks to a pupil function). In this way, we can control the speckle grain size by varying the pupil size of the pupil function, and we found that the speckle grain size does not have an impact on the TM retrieval. (In the simulation in Fig. 4, we show a speckle grain size of 1 pixel and in Fig. 7, we show the case of larger speckle grain of 4 pixels). To construct T_1, T is further multiplied element-wise (Hadamard product) with the object vector, so that $T_1 \in \mathbb{C}^{N_{\text{SLM}} \times N_{\text{target}}}$. No correlations are embedded in T_1 generation. The intensity transmission matrix T_2 is simulated in a similar way, but by using real positive coefficients, corresponding to the intensity speckle patterns, $T_2 \in \mathbb{R}_+^{N_{\text{target}} \times N_{\text{CAM}}}$.

Here in our simulation, we chose as fluorescent continuous object a neuron with its connecting dendrites. We study the effect of the rank of the neural network (by changing the dimension of the hidden layer) on the final reconstruction. The training dataset is generated by randomly producing input patterns, and computing the corresponding output speckle images. These speckle images are obtained by propagating the input field with a field transmission matrix, and then multiply it with an intensity transmission matrix to simulate the final set of output fluorescence images. Once the neural network is trained, we phase conjugate T_1 and scan through each column to focus the images through the scattering media at all possible locations, and observe the images as they would appear at the object plane. Here we present the original target image without scattering and the focus combined image with phase conjugation $\widetilde{T_1}$, retrieved by the neural network [29]. Figure 4 shows the combined foci with the shape of the targeted neuron. Although the rank (i.e. the dimension of the hidden layer) is a required input parameter, it can be approximated, as it will impact the intensity distribution but not the overall shape. Note that there is no condition on the sparsity of the object in this section.
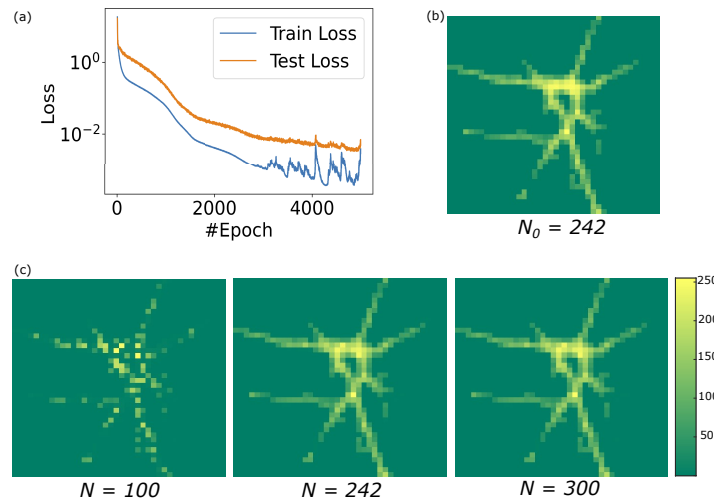
Fig. 4. Continuous fluorescent object reconstruction by the 2-layer neural network.
(a) Example of train and test losses during the neural network training (N =300 case). (b)
Ground truth image, simulated, N_0 is the number of targets. (c) Focus-combined images
under different middle layer size N (i.e rank) using phase conjugation of the retrieved T_1
from the 2-layer neural network showing an accurate prior information on the middle layer
is not needed for TM retrieval in continuous sample.

3.2. Experiment

3.2.1. Discrete object

Experimentally, the measurement was first applied to 8 fluorescent beads of diameter 1 μm,
placed on a holographic diffuser (Newport, 10DKIT-C1, 10° as the diffusion angle). A set of
known input random patterns (N_{pat} = 15360) are displayed onto the SLM and the corresponding
fluorescence speckles are recorded. With this input/output training set, the loss function of
the 2-layer neural network is minimized and two transmission matrices T_1 and T_2 are finally
obtained. In order to confirm the quality of this retrieval method, light is focused onto each bead
using phase conjugation of T_1 [29,30]. When light is successfully focused, it indicates that both
transmission matrices were well reconstructed. In Fig. 5, one can see the sum of all the control
camera snapshots (placed in transmission) of all the beads foci after phase conjugating T_1, and
the non-invasive reconstruction of the object using T_2. From T_2 transmission matrix exploitation
and the optical memory effect (OME), one can reconstruct the object shape [31–33]. The optical
memory effect is a type of wave correlation that is observed in coherent fields, allowing control
over scattered light through thin and diffusive materials [34]. Essentially, two beads in the
memory effect range should show translated fluorescent patterns with a displacement equal to
their relative distance. By correlating the fluorescence speckles recorded in epi-detection, while
displaying the phase of the SLM that allows focusing (using phase conjugation of T1), it is
possible to non-invasively obtain a distance map between all the emitters [31]. Another approach
is used to reconstruct the continuous object from OME [32]. Here only 16×16 macropixels
of the SLM are modulated for the input, so our focusing enhancement will be impacted by the
limited pixel counts. We expect to have an even higher signal to background ratio (SBR) with an
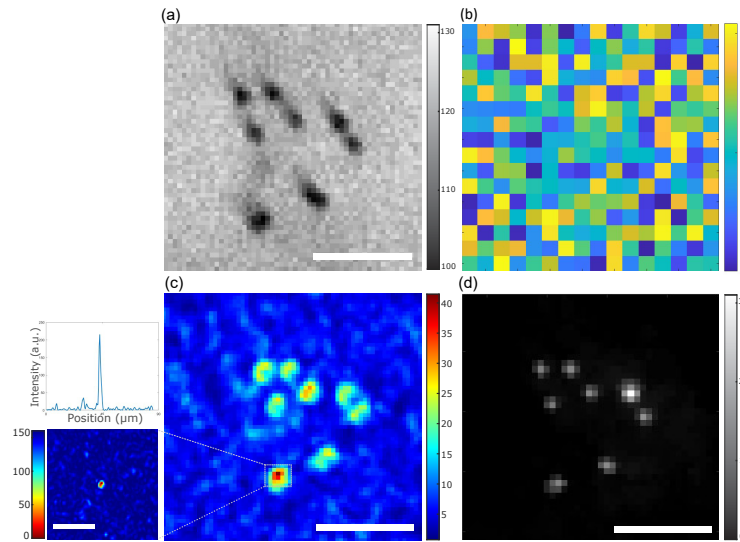increased number of SLM pixels.

Fig. 5. **Experimental reconstruction of the fluorescent beads sample in epi-detection**. The rank of the hidden layer in the algorithm is $N_{\text{target}} = 9$, to ensure the reconstruction of all the beads. The number of patterns send onto the SLM is $N_{\text{pat}} = 15360$. (a) Ground truth, 8 beads, brightfield illumination, in the control camera. (b) Example of the SLM phase pattern projected to focus on one bead, found by phase conjugation of $\widetilde{T_1}$. (c) Sum of the images of the focus on each bead in the control camera. From each focus, the SNR is computed, and only the ones with an SNR higher than 20 are added to the sum. On the left, one focus with its associated intensity profile. (d) Reconstruction of the object, through scattering medium, from T_2 and using the cross-correlation procedure shown in [31]. The number of pixels used on the epi-CAM is approximately 80x80. The white bar corresponds to 10 μm.

3.2.2. Continuous object

The measurement was also applied to continuous objects like pollen seeds to show that the method can be used with 3D continuous objects and biological ones. The whole approach of acquisition, measurement, and processing is similar to beads samples: *P* random incident wavefronts are produced by the SLM and sent onto the scattering medium, *P* fluorescent speckles are respectively measured on the camera; the neural network is used to retrieve transmission matrices, T_1 is used to focus onto the pollen seed and the *eigen patterns* are used to reconstruct the object thanks to OME. The dimension of the hidden layer is increased to match the number of equivalent discrete targets. The latter can be estimated from the minimization of the Frobenius norm of the Non Negative Matrix Factorization (NMF) residual is done [31,32]. To make this reconstruction work, we initialized the weight of the second layer with the result of the NMF algorithm over the output dataset [35], in order to help the training of the neural network. Only the result of the NMF over the same experimental data is used to initialize the weights of T2 in the 2-layer neural network. The 2-layer neural network then performs global optimization over the parameters in the forward model to further retrieve the transmission matrices under the minimized loss. Giving information over the elements of T_2 thanks to NMF results will help the convergence ratio of the neural network training. Once T_1 and T_2 are retrieved, the fingerprint of each target is computed by reshaping T_2 to a 2D image. In each subregion within the OME range, a pair-wise deconvolution produces a relative distance map that can be connected to reconstruct the final image, non invasively [32]. Fig. 6 shows the ground truth image of a pollen seed and the neural network's reconstruction. The dataset used is the same as in [31].
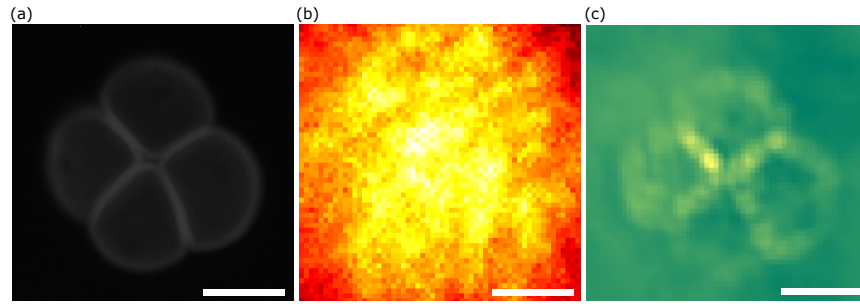
Fig. 6. Experimental reconstruction of a fluorescent biological object in epi-detection.
The rank in the algorithm is set to $N_{\text{target}} = 53$. The number of patterns send onto the SLM is
$N_{\text{pat}} = 5120$. (a) Ground truth, pollen seed image in transmission without scattering medium.
(b) Example of the speckle we can get in epi-detection for an arbitrary SLM pattern with the
pollen seed as the fluorescent object. (c) Reconstruction of the object, through the scattering
medium, from pairwise deconvolution over the fingerprints of T_1. Scalebar is 10 μm.

This shows that the structure of the 2-layer neural network can also be used to retrieve a
continuous object hidden in scattering medium in experiments since it can accept some priors
such as the NMF over the output data to initialize the second layer.

3.3. Other contrast mechanisms

To illustrate the versatility of this neural network approach, simulations for another contrast
mechanisms were performed. Here, we study numerically the case of imaging through a scattering
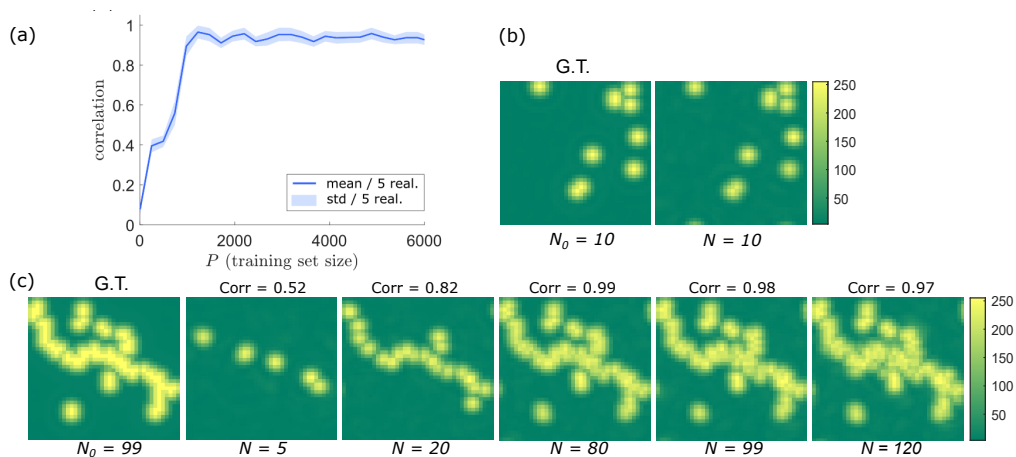medium with contrast from second harmonic generation (SHG). Essentially, this is a coherent

Fig. 7. Simulation on the 2-layer neural network for SHG imaging process. (a)
Correlation curve between generated images and ground truth according to the training
set size, over 5 realisations. At each realisation, the initialization of TMs is set to random
matrices. 200 examples used in the test set. (b) Ground truth (G.T.) and focus-combined
image, using phase conjugation of the retrieved $\widetilde{T_1}$ from the 2-layer neural network for a
simulated discrete object (like beads). (c) Ground truth and focus-combined images plus their
correlation with ground truth under different middle layer size N using phase conjugation
of the retrieved $\widetilde{T_1}$ from the 2-layer neural network for a simulated continuous object (like
collagen tissue fibers).

phenomenon which can be modelled by a pair of complex valued transmission matrices the same way as before, by simply changing the activation functions to $g_1 = .^2$ and $g_2 = |.|^2$. The forward model is then:

$$x_2 = g_2(T_2 g_1(T_1 x_0)) = |T_2(T_1 x_0)^2|^2. \tag{4}$$

The simulation procedure is the same as before, only the forward model has changed. It is more challenging to retrieve phase information correctly when compared to the fluorescent case. A correlation between the transmission matrices and the corresponding ground truth is shown in Fig. 7. Contrarily to previous results, the correlation does not reach exactly 1 and the training set size needed for high correlation is significantly higher than that in the case of fluorescence. This can be explained by the absence of a phase reference leaving some ambiguity for T_2. Despite this, the retrieved matrices should be sufficient for many applications, such as focusing, as can be seen on the simulation over a continuous object, using the same technique as described in the simulation section by simply changing the forward model over T_2. The focusing is again possible, as shown in Fig. 7. As simulated objects, we chose polystyrene beads and an image of tissue collagen fiber, which are common SHG contrast sources.

4. Discussion

Our proposed 2-layer physics-based neural network approach enables the efficient retrieval of the two transmission matrices non-invasively, meaning without direct access to the object plane. With phase conjugation over T_1, we are able to focus and thanks to T_2 we can successfully reconstruct the object, hidden behind the scattering medium, provided there is some optical memory effect. In order to recover these two TMs, experimentally, for fluorescence, a significant amount of measurements is required: approximately 15000 input/output pairs of experimental data are used for training our network. Through numerical study, we confirmed that it is not necessary to have an accurate estimation of the rank of the object to have two transmission matrices retrieved. In the experiment, the limitation to the reconstruction of these complex objects (for example the pollen seeds) is the contrast of the measured speckle which decreases with a $1/\sqrt{N_{target}}$ dependency [36]. In terms of transmission matrix reconstruction, our approach is limited by the simple architecture of the network, because of few freedom degrees. Even so, this still allows us to add physical priors, such as changing the initialization and the nature of the components of the two TMs as well as the activation functions. For example, using the result of the non-negative matrix factorization [35] over I^{out} to initialize T_2 helps find more targets than the previous 2-step approach based on NMF and phase retrieval [31] (see Supplement 1 - Influence of initialization). Since our training is performed with GPU acceleration with PyTorch, it takes less than 2 min for a training data size of 242000. The data acquisition time takes around tens of minutes due to the limited close-loop speed of our system. Also, our method currently is limited to static and structural imaging, it can be further optimized to image more dynamic systems. For example, the current training strategy can be further optimized in two aspects: 1) computationally, techniques such as online training or prior information can be incorporated to reduce the training/re-calibration time; 2) hardware-wise, the closed-loop speed of the imaging system can be improved by higher-sensitivity detectors and FPGA in the loop. Finally, this approach is versatile and applicable to other contrast mechanisms, such as SHG where the positivity constraint on T_2 does not hold, but also Raman signal or 2-photons processes. The forward model is simply changed according to the physical phenomenon at stake, as shown in Table 1. In experiment, we could have some practical constraints preventing us from directly applying the method on all non-linear phenomena, such as a weak signal. But these could potentially be addressed by increasing the number of training patterns, the excitation power or adding more priors to the model for example.

Table 1. Different phenomenon applicable to the 2-layer model with the right activation function.

Non-linearity	Coherent	Non-coherent
$n = 1$	Raman: $g_1 = Id$, $g_2 = \lvert . \rvert^2$	Linear Fluo.: $g_1 = \lvert . \rvert^2$, $g_2 = Id$
$n = 2$	SHG: $g_1 = .^2$, $g_2 = \lvert . \rvert^2$	2-photon Fluo.: $g_1 = \lvert . \rvert^4$, $g_2 = Id$
$n = 3$	THG: $g_1 = .^3$, $g_2 = \lvert . \rvert^2$	3-photon Fluo.: $g_1 = \lvert . \rvert^6$, $g_2 = Id$

5. Conclusion

Our study presented a physics-based machine learning method for characterizing light propagation through complex samples in order to focus on and reconstruct objects inside scattering media. Compared with other deep learning-based methods, our model is advantageous for data interpretation and incorporation of physical priors. Each node in the network has a physical meaning: it corresponds to the coefficients of the two transmission matrices. In order to test the quality of the neural network reconstruction, training and testing sets were used as is usually done in machine learning approaches. Compared to previous approach [31] which is limited to the single case of linear fluorescence imaging, in this work, we proposed a simple physics-based 2-layer neural network that can be adapted and generalized for imaging through scattering. Our new model can be used to retrieve transmission matrices in imaging systems with different contrast mechanisms. This new method could be more demanding in terms of measurements, because there are no priors on the algorithm. Furthermore, the phase information is not measured. However, the two-layer neural network is more general and less sensitive to noise (see Supplement 1 - Influence of noise). This approach can also be easily generalized and adapted to other contrast mechanisms such as 2-photon fluorescence or coherent processes where the previous method would not work. Moreover, the physics-based neural network approach is versatile, and additional physical priors may be added in the model to further enhance the capability of the method, such as memory effect [37] information or 3D composition of the sample.

Disclosures. The authors declare no conflicts of interest.

Data availability. Data and codes underlying the results presented in this paper are available in Ref. [38].

Supplemental document. See Supplement 1 for supporting content.

References

1. J. W. Goodman, "Some fundamental properties of speckle," J. Opt. Soc. Am. **66**(11), 1145–1150 (1976).
2. A. P. Mosk, A. Lagendijk, G. Lerosey, and M. Fink, "Controlling waves in space and time for imaging and focusing in complex media," Nat. Photonics **6**(5), 283–292 (2012).
3. S. Rotter and S. Gigan, "Light fields in complex media : Mesoscopic scattering meets wave control," Rev. Mod. Phys. **89**(1), 015005 (2017).
4. I. Vellekoop and A. P. Mosk, "Controlling waves in space and time for imaging and focusing in complex media," Opt. Lett. **32**(16), 2309 (2007).
5. A. Ishimaru, *Wave propagation and scattering in random media*, vol. 2 (Academic press New York, 1978).
6. S. Popoff, G. Lerosey, M. Fink, A. C. Boccara, and S. Gigan, "Controlling light through optical disordered media : transmission matrix approach," New J. Phys. **13**(12), 123021 (2011).
7. S. Popoff, G. Lerosey, M. Fink, A. C. Boccara, and S. Gigan, "Image transmission through an opaque material," Nat. Commun. **1**(1), 81 (2010).
8. R. Horstmeyer, H. Ruan, and C. Yang, "Guidestar-assisted wavefront-shaping methods for focusing light into biological tissue," Nat. Photonics **9**(9), 563–571 (2015).
9. T. Chaigne, O. Katz, A. Boccara, M. Fink, E. Bossy, and S. Gigan, "Controlling light in scattering media non-invasively using the photoacoustic transmission matrix," Nat. Photonics **8**(1), 58–64 (2014).

10. J. W. Lichtman and J.-A. Conchello, "Fluorescence microscopy," Nat. Methods **2**(12), 910–919 (2005).
11. D. J. Webb and C. M. Brown, "Epi-fluorescence microscopy," in *Cell imaging techniques*, (Springer, 2012), pp. 29–59.
12. G. Ghielmetti and C. M. Aegerter, "Direct imaging of fluorescent structures behind turbid layers," Opt. Express **22**(2), 1981–1989 (2014).
13. M. Hofer, C. Soeller, S. Brasselet, and J. Bertolotti, "Wide field fluorescence epi-microscopy behind a scattering medium enabled by speckle correlations," Opt. Express **26**(8), 9866–9881 (2018).
14. D. Li, S. K. Sahoo, H. Q. Lam, D. Wang, and C. Dang, "Non-invasive optical focusing inside strongly scattering media with linear fluorescence," Appl. Phys. Lett. **116**(24), 241104 (2020).
15. J. Bertolotti, E. van Putten, C. Blum, A. Lagendijk, W. Vos, and A. Mosk, "Non-invasive imaging through opaque scattering layers," Nature **491**(7423), 232–234 (2012).
16. P. Caramazza, O. Moran, R. Murray-Smith, and D. Faccio, "Transmission of natural scene images through a multimode fibre," Nat. Commun. **10**(1), 2029 (2019).
17. G. Barbastathis, A. Ozcan, and G. Situ, "On the use of deep learning for computational imaging," Optica **6**(8), 921–943 (2019).
18. A. Turpin, I. Vishniakou, and J. d. Seelig, "Light scattering control in transmission and reflection with neural networks," Opt. Express **26**(23), 30911 (2018).
19. M. Mounaix, D. M. Ta, and S. Gigan, "Transmission matrix approaches for nonlinear fluorescence excitation through multiple scattering media," Opt. Lett. **43**(12), 2831–2834 (2018).
20. M. W. Grieves, "Virtually intelligent product systems: digital and physical twins," (2019).
21. L. G. Wright, T. Onodera, M. M. Stein, T. Wang, D. T. Schachter, Z. Hu, and P. L. McMahon, "Deep physical neural networks enabled by a backpropagation algorithm for arbitrary physical systems," arXiv preprint arXiv:2104.13386 (2021).
22. S. Sharma, S. Sharma, and A. Athaiya, "Activation functions in neural networks," IJEAST **4**(12), 310–316 (2020).
23. Y. LeCun, D. Touresky, G. Hinton, and T. Sejnowski, "A theoretical framework for back-propagation," in *Proceedings of the 1988 connectionist models summer school*, vol. 1 (1988), pp. 21–28.
24. R. A. Fisher, *Optical phase conjugation* (Academic, 2012).
25. Z. Yaqoob, D. Psaltis, M. S. Feld, and C. Yang, "Optical phase conjugation for turbidity suppression in biological samples," Nat. Photonics **2**(2), 110–115 (2008).
26. L. Bottou, "Stochastic gradient learning in neural networks," Proceedings of Neuro-Nimes **91**, 12 (1991).
27. T. Mitchell, "Artificial neural networks," in *Machine Learning 10-701*, (Machine Learning Department Carnegie Mellon University, 2010).
28. A. Boniface, J. Dong, and S. Gigan, https://github.com/laboGigan/NMF_PR.
29. J. Yang, J. Li, S. He, and L. V. Wang, "Angular-spectrum modeling of focusing light inside scattering media by optical phase conjugation," Optica **6**(3), 250–256 (2019).
30. I. M. Vellekoop, M. Cui, and C. Yang, "Digital optical phase conjugation of fluorescence in turbid tissue," Appl. Phys. Lett. **101**(8), 081108 (2012).
31. A. Boniface, J. Dong, and S. Gigan, "Non-invasive focusing and imaging in scattering media with a fluorescence-based transmission matrix," Nat. Commun. **11**(1), 6154 (2020).
32. L. Zhu, F. Soldevila, C. Moretti, A. d'Arco, A. Boniface, X. Shao, H. B. de Aguiar, and S. Gigan, "Large field-of-view non-invasive imaging through scattering layers using fluctuating random illumination," Nat. Commun. **13**(1), 1447 (2022).
33. S. Feng, C. Kane, P. A. Lee, and A. D. Stone, "Correlations and fluctuations of coherent wave transmission through disordered media," Phys. Rev. Lett. **61**(7), 834–837 (1988).
34. G. Osnabrugge, R. Horstmeyer, I. N. Papadopoulos, B. Judkewitz, and I. M. Vellekoop, "Generalized optical memory effect," Optica **4**(8), 886–892 (2017).
35. Y.-X. Wang and Y.-J. Zhang, "Nonnegative matrix factorization: A comprehensive review," IEEE Trans. Knowl. Data Eng. **25**(6), 1336–1353 (2013).
36. A. Boniface, B. Blochet, J. Dong, and S. Gigan, "Noninvasive light focusing in scattering media using speckle variance optimization," Optica **6**(11), 1381–1385 (2019).
37. I. Freund, M. Rosenbluh, and S. Feng, "Memory effects in propagation of optical waves through disordered media," Phys. Rev. Lett. **61**(20), 2328–2331 (1988).
38. A. d'Arco, F. Xia, A. Boniface, J. Dong, and S. Gigan, https://github.com/laboGigan/2-layer-NN.