A physics informed neural network as a digital twin of scattering media

Mohammadrahim Kazemzadeh[1] Liam Collard[1,2] Linda Piscopo[1,3]
Massimo De Vittorio[1,2,3,4] Ferruccio Pisanello[1,2,4]

[1]Istituto Italiano di Tecnologia, Center for Biomolecular Nanotechnologies
[2]RAISE Ecosystem, Genova, Italy
[3]Dipartimento di Ingegneria dell'Innovazione, Università del Salento
[4]These two authors are the co-last authors.
emails: Mohammadrahim.kazemzadeh@iit.it, Massimo.DeVittorio@iit.it, Ferruccio.Pisanello@iit.it

Abstract

This paper presents a novel methodology for characterizing and creating a digital twin of turbid media without the need for reference beams, employing a unique architecture for the physics-informed neural network. Unlike previous approaches utilizing various deep neural network architectures that often function as black boxes, our method prioritizes interpretability, offering a clearer understanding of the underlying processes of light propagation through turbid media. Additionally, unlike classical solutions relying on reference beams, our approach eliminates the dependence on beam quality and associated issues prevalent in both internal and external reference-based methods. The possibility and use case of gradient calculation through this presented digital twin are showcased by solving the reverse problem to retrieve the initial wavefront shape of the light that passed through this medium, known as image transmission. Surprisingly, the results surpassed the accuracy of models directly optimized for this task, underscoring the precision of the proposed digital twin. This capability represents a pivotal advancement for future developments in neuromorphic and deep learning computation and training using such a photonic and optical systems.
Keywords: Digital Twin, Scattering Media, Image Transmission, Optical Computation

1 Introduction

Over the past decade, extensive research has been dedicated to unravelling the complex relationship between input and output light fields of a multimode optical fiber [1–9]..When coherent light is injected into an MMF, it couples to multiple guided modes, leading to complex interference patterns. This phenomenon, known as speckle patterns, poses challenges for image transmission through MMFs, as variations at the input facet strongly impact the entire speckle pattern at the output. Interferometric techniques have revolutionized imaging across these aberrating media by measuring their transmission matrix and have led to a new generation of optical endoscopses based on single multimode optical fibers [10] [11] .

Traditionally, measuring the complex valued transmission matrix (TM) of an optical fiber assumes the preservation of coherence properties during the scattering process and requires an interferomic measurement of transmitted light typically made via off-axis holography with an external reference [6, 12, 13]. Achieving stability in the interferometer, especially over long optical paths, is crucial. Alternatives like phase-stepping holography with a co-propagating internal reference field have emerged to address stability concerns [14] [15, 16]. However, due to phase discontinuities within the internal reference "blind spots" appear within the output field of view of the fiber[14] [15, 16].

Recently, convolutional neural networks (CNNs) [17–23] and vision transformers [24–27], have been employed to address the constraints associated with interferometric measurements. Noteworthy studies showcase the efficacy of these techniques in tasks such as image transmission, wavefront shaping, and holography at the distal end of multimodal optical fibers (MMFs). However, a common limitation in most of these methods is the opaque nature of the deep neural network's actual performance. It acts as a black box, offering limited insights into how precisely these methods translate speckle patterns into the

input electric field of the turbid media. Typically, these techniques work on recognising features in input output relation and are thus unable to transmit unfeatured datasets and thus underutilize the capacity of the MMF.

An alternative approach has recently emerged for characterizing the MMF fiber, employing a Physics-Informed Neural Network (PINN) [28]. In this innovative method, a single-layer Complex-Valued Neural Network (CVNN) has been implemented to map the modulated phase, facilitated by the spatial light modulator (SLM), onto the amplitude of the recorded polarized light originating from the proximal facet of the fiber. The versatility of this technique is demonstrated through its cascading capability, allowing the creation of a multi-layer CVNN for describing stacks of multiple turbid media. However, to fully harness this method for the digital twin purposes, an extension to all propagating polarizations of light is essential. Importantly, advancements in the field of deep physical neural networks and optical neural computation rely on having a digital twin with the capability of gradient calculation for the training of these systems [29–31]

Another significant challenge associated with the utilization of PINN is the requirement for an accurate understanding of the governing equation dictating light transmission through the media and its modulating elements. This task proves to be particularly formidable due to the inherently highly non-linear behavior exhibited by modulating elements, such as the spatial light modulator (SLM). Even with meticulous calibration and modulation, their capabilities remain confined to the modulation of only one polarization of light. This limitation introduces additional non-linearity, especially in the presence of other optical components like polarizers within the system. Thus, the intricate interplay of these elements further complicates the already challenging task of precisely characterizing and controlling light transmission through the media.

In this study, we propose a physics-informed neural network by introducing a novel architecture of neural network. This network not only characterizes the scattering media to indirectly deduce the transmission matrix for all polarizations but also provides a sufficiently descriptive understanding of the fiber's physics. This understanding extends to the point where the network can indirectly contribute to the complex task of image transmission through turbid media of unknown non-linear modulation, thus serving as a digital twin of the physical system.

The proposed method boasts substantial advantages over deep learning-based approaches. Firstly, it is not a black box; its performance is grounded in the physics of the fiber, allowing for direct association of all optimized weights and biases in the network with physical properties of the system. Furthermore, the possibility and use case of gradient calculation through this presented digital twin are showcased by solving the reverse problem. This reverse problem can be described as finding an SLM pattern that leads to a specific speckle pattern and retrieving the initial wavefront shape of the light that passed through this medium. In literature, this problem is known as image transmission. Remarkably, our model surpassed models directly optimized for image transmission, underscoring the level of accuracy achieved by this digital twin.

Despite the deep learning-based methods that often rely on data's features, limiting their generalization. For instance, a model developed for handwritten digit images may struggle to generalize effectively to reconstruct images from other datasets, such as those containing natural scenes [19]. In contrast, the digital twin solution for image reconstruction works based on pixel reconstruction, leading to context-free, universal image reconstruction. This distinctive feature enhances the adaptability and broader applicability of our method.

Finally, this digital twin concept enables the utilization of our method for a variety of tasks, particularly in calculating gradients for training neuromorphic photonic and optical systems [29].

2 Methodology

2.1 Optical system

The optical setup employed in this study is intricately detailed in Figure 1 (a) and (b), comprising three telescopic systems, a half-wave plate (HWP), a spatial light modulator (SLM), two reflecting flat mirrors, a beam splitter, a multimode optical fiber (MMF), with the core size of $50\mu m$ and numerical aperture of .22, and two microscope objectives (MO), each coupled with a CCD camera. For laser light coupling into the SLM screen, lenses denoted as L1 and L2 are positioned in a 4f configuration. The reflected light from the SLM is then guided into the first microscope objective (MO1), effectively conjugating the SLM screen onto the proximal side of the MMF.

Monitoring the proximal side involves CCD1 and a telescope setup (L4 and MO1). The generated speckle, along with the distal side of the fiber, is recorded using CCD2, facilitated by a telescopic arrangement (MO2 and L5). Throughout this setup, the SLM pattern is precisely commanded, and the resulting speckle pattern is recorded using a computer.

2.2 physical formalism of the optical setup

Taking into account the relatively low power of the continuous laser utilized in this study (less than 1 mw), we can treat the entire system as a linear time-invariant (LTI) system. Consequently, by applying the superposition principle, we can articulate the behavior of the electric field vector at the output facet of the fiber in response to the modulated phase from the SLM.

$$\tilde{\mathbf{E}}(x,y) = \sum_{w=1}^{W} \tilde{\mathbf{U}}_w(x,y) e^{j\phi_w} + \tilde{\mathbf{Z}}(x,y) \tag{1}$$

Here, $\tilde{\mathbf{E}}$ denotes the electric field vector, while $\tilde{\mathbf{U}}_w$ represents a complex vector corresponding to the electric field at the distal ends of the fiber, resulting from light emitted from a single pixel on the SLM screen. The Neper number is denoted by e, and ϕ_w signifies the modulated phase from the same SLM pixel. Additionally, $\tilde{\mathbf{Z}}$ characterizes the electric field caused by unmodulated light from the SLM. Finally, x and y specify the spatial location on the distal side of the MMF in the Cartesian system (see Figure 1 (b)).

Nevertheless, due to the camera's inherent limitation in recording electric field components or phase, our observations are confined to values proportional to the electric field's amplitude. Consequently, we formulate the camera reading as follows:

$$|\tilde{\mathbf{E}}|^2(x,y) = E_a^2(x,y) + E_b^2(x,y) = \sum_{k}^{a,b} \left| \left(\sum_{w=1}^{W} U_{k,w}(x,y) e^{j\phi_w} + Z_k(x,y) \right) \right|^2 \tag{2}$$

In this context, we introduce subscripts a and b, to signify components within an orthogonal Cartesian system established at the distal facet of the fiber. It's crucial to highlight that this Cartesian system may not align with the previously defined coordinates (x and y) for the distal fiber facet (see Figure 1 (b)). In this formulation of the electric field intensity for camera reading, there is no strict requirement for these coordinate systems to align.

In a more general case of the both phase and amplitude modulation we can formulate the electric field intensity as follows:

$$|\tilde{\mathbf{E}}|^2(x,y) = \sum_{k}^{a,b} \left| \left(\sum_{w=1}^{W} U_{k,w}(x,y) \left(f_r(\phi_w) + j f_i(\phi_w) \right) + Z_k(x,y) \right) \right|^2 \tag{3}$$

In this expression, the real functions f_r and f_i represent the real and imaginary components of the modulation. It is essential to note that in this generalized scenario, ϕ_w does not exclusively define the phase modulation; rather, it serves as a more comprehensive modulation parameter. This broader form of modulation will be employed later on for training the digital twin of the optical system. Notably, in this training process, we seek to discover the non-linear complex modulation, which remains unknown.

2.3 Data

By systematically transmitting multiple images through this turbid media, we construct an extensive dataset, affording us access to both the amplitude $|\tilde{\mathbf{E}}|$ and the phase or the modulation parameter (ϕ_w). In this context, our objective is to ascertain the unknown complex variables $U_{k,w}$, Z_k, and the real modulation functions, f_r and f_i, in case of dealing with unknown modulation.

For this study, we acquired a set of 150,000 Spatial Light Modulator (SLM) patterns and their associated speckle patterns. This collection comprises 120,000 randomly generated SLM patterns, 10,000 patterns showcasing celebrity faces from the CelebA dataset, 10,000 patterns of handwritten digits from the MNIST dataset, and finally, 10,000 patterns from the CIFAR dataset. Each grayscale image was employed to modulate light using the SLM, introducing phase shifts ranging from 0 to π. The corresponding speckle patterns resulting from these modulations were recorded.

Figure 1 (c) displays examples of images from each dataset along with their corresponding speckle patterns. Additionally, we conducted Principal Component Analysis (PCA) and Uniform Manifold

Approximation and Projection (UMAP), Figure 1 (d), on the speckle data to visualize the distribution of this dataset.

The PCA and UMAP projections unequivocally reveal significant distinctions in the statistical characteristics of speckle patterns from each dataset. The clear separation observed in both PCA and UMAP projections emphasizes the distinct and discernible nature of speckle patterns originating from the various datasets. One objective of this study is to demonstrate that the physics-informed model, once trained on one dataset, can be applied universally to other datasets, despite clear statistical differences between them. This is a challenge that conventional deep learning models cannot easily address, as their performance is restricted to the learned features from their training dataset, resulting in limited generalization capabilities.

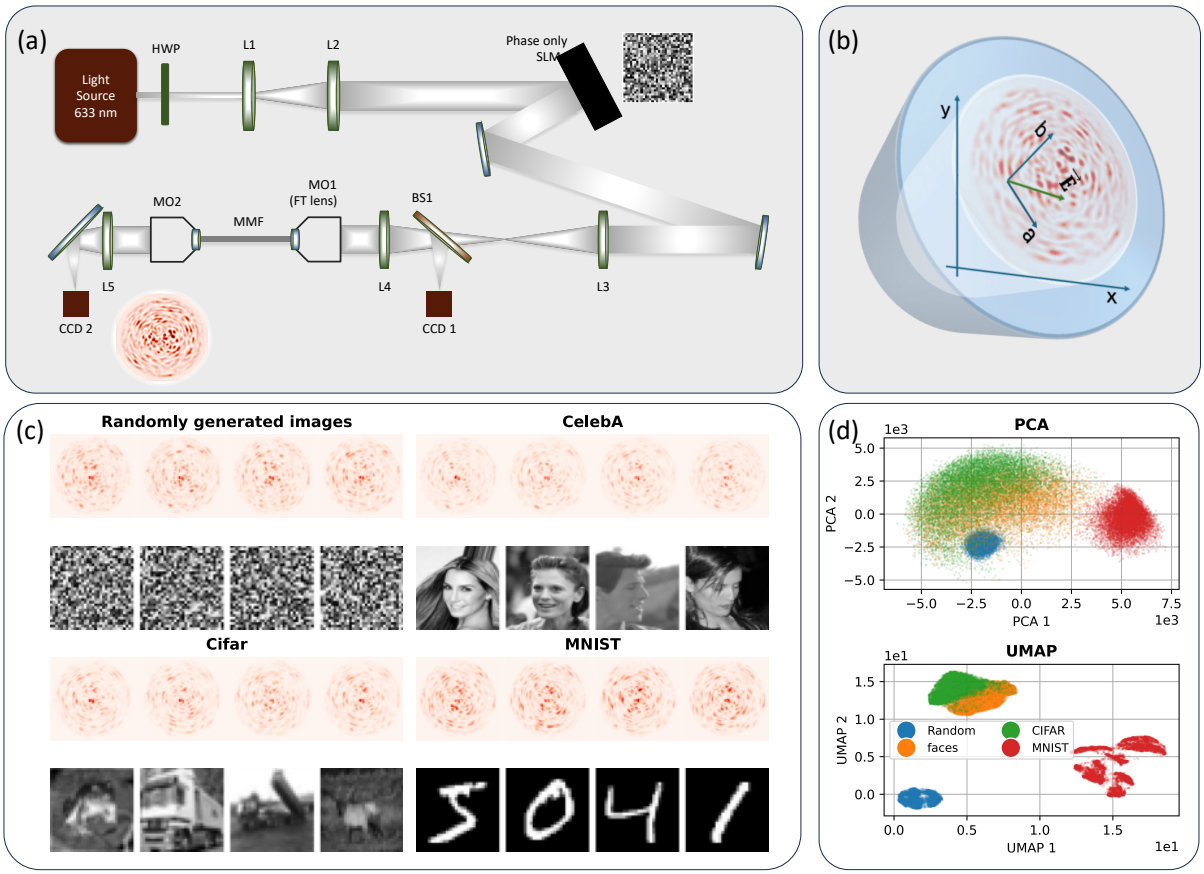

Figure 1: (a) The utilized optical setup, and (b) a schematic illustrating the tangential electric field vector on the distal facet of the fiber, along with the selected Cartesian coordinate systems for the speckle patterns and the polarization of the electric field, denoted by the axes pairs (x, y) and (a, b), respectively.(c) An example of the utilized dataset in this study along with their corresponding speckle patterns. (d) Principal Component Analysis (PCA) transformation of the speckle patterns used in this study.

We have also utilize the data presented in [19] to showcase the usability of the method for unknown modulation and different optical configuration.

2.4 Neural network architecture

Our proposed approach involves the utilization of a neural network as elucidated in Figure 2 to solve this probelem. Notably, the network is designed to accept one input which is the phase mask of the SLM. Accompanying this, we present a tensor operation emulating Equation 3. This is complemented by a cost function assessing the similarity between the generated speckle by the neural network and the ground truth, quantified through Mean Absolute Error (MAE).

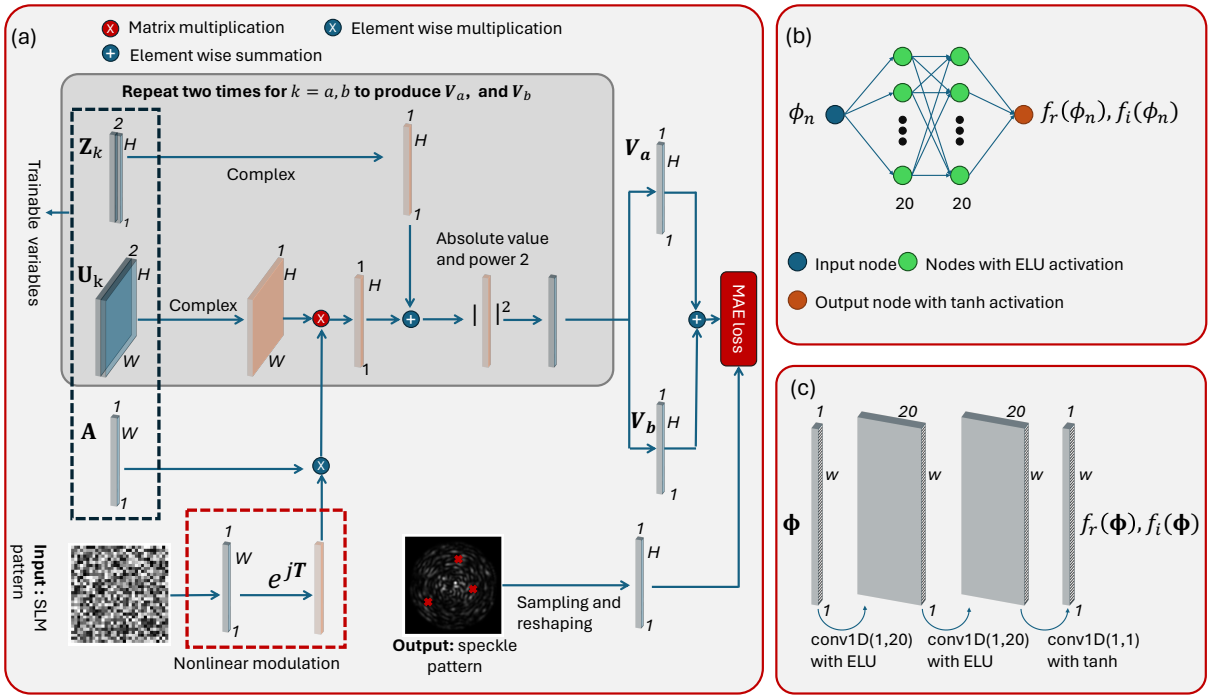

Figure 2: (a) Schematic of the utilized physics-informed neural network architecture. The trainable variables are marked with dashed blue squares, and the non-linear complex transformation of modulation for the case of known modulation is depicted inside the red square. (b) A simple architecture of a one-input and one-output deep neural network model that generates the real or imaginary part of the modulation. (c) A one-dimensional deep convolutional neural network that applies a similar transformation as in (b) across all input values.

To understand the utilized tensor operations, we should firstly notice that it is not necessary to investigate the entire (x, y) location of the output facet of the fiber. Instead, we can reduce the problem by considering only H selected points in that space. This matter wont effect the generality of the problem as the H can be taken large enough to cover all the pixels of the recording camera.

Therefore, by considering this selection, Equation 3 can be transformed into a set of H non-linear equations. Now, both variables with spatial dependencies, $U_{k,w}(x, y)$ and $Z_k(x, y)$, then only need to be determined at H points. As $U_{k,w}(x, y)$ variable has modulation element dependency in its subscript, we can represent these complex values in a two-dimensional array with the size $(W, H, 2)$. First axis covers the modulation parameter while the second one covers the sampled points from speckle. Finally the last axis is defined to accommodate the real and imaginary part of $U_{k,w}(x, y)$. On the other hand, for the $Z_k(x, y)$, we only have spatial dependency so a tensor with size of $(1, H, 2)$ suffices for its full description.

Moving forward, for the input, the modulated real-valued phase is reshaped into a tensor of size $(1, W, 1)$ and subjected to a complex element-wise operation denoted as e^{jT}, where T represents the elements of the replicated tensor. Please note that this has been done for the case of the known phase only modulation. Importantly, the output of this operation is a complex-valued tensor of the modulated phase. Then we multiply the resulted tensor to additional variable tensor (A) with the same shape $(1, W, 1)$. The reason behind the definition of this additional tensor is that if a pixel of the SLM is not contributing in the speckle generation (i.e. weak optical coupling such as being out of the accepted numerical aperture of any of the utilized optical elements) the corresponding number in that variable tensor can turn into zero. This effectively weights the computed U_k against the efficiency of the optical setup, in principle it is possible to skip this step however would result in a loss of information as an entire column of U_k would need to be discarded. Practically we found that addition of this tensor can not only proves helpful in the training process but also can deliver an insightful understanding of the optical system coupling behaviour. This will be discussed in details in the following sections.

The resulted complex-valued tensor of the phase modulation is then subjected to matrix multiplication with the complex-valued version of U_k tensor. Following this multiplication, the complex-valued tensor is taking the shape of $(1, H, 1)$.

The tensor is then added to the unmodulated part, complex-valued tensor (Z_k), followed by the calculation of the absolute value and powering to two, as indicated in Equation 3. This whole operation, except the part for the phase and its correction, needs to be repeated for all instances of 'k' to produce two tensors at the output, namely V_a and V_b. Based on Equation 3, then the summation of these two can be equated to the selected pixel's values from speckle. The MAE loss finds the difference between the calculated speckle value and the actual measured one.

2.4.1 Unknown Modulation

The presented model can also be applied to create an accurate digital twin of the optical system even when the modulation remains unknown or unclear. Through the following method, we can discover this non-linear complex modulation during the training process. To achieve this, we implement the mathematical model of the optical setup described in equations 3. Instead of relying on the complex modulation function e^{jT} used for phase-only modulation, we employ a non-linear activated deep neural network. This neural network, illustrated in Figure 2 (b) as a single-input single-output model, consists of two hidden layers with 20 nodes each, utilizing Exponential Linear Unit (ELU) activation functions. The output node is activated with a hyperbolic tangent function. The selection of these activation functions ensures that the neural network model mimics a continuous and derivable function at all points.

It's important to note that we need two versions of this neural network to produce the two independent real-valued functions f_r and f_i as per equation 3. Another crucial consideration is that these neural network functions must be applied to the entire modulation element. Therefore, we propose implementing these functions as a form of multi-layer one-dimensional convolutional neural network, as illustrated in Figure 2 (c). This implementation, equivalent to Figure 2 (b), ensures that the same function is applied to all modulating parameters.

2.5 Model training

We implemented the neural network using TensorFlow in Python. For optimization, we used the Adam optimizer with an initial learning rate of 0.001 for 100 epochs.

3 Results

3.1 "A" tensor and discovery of the non-linear modulation

The depiction of all 1024 randomly selected points from speckle, alongside the optimized A for both cases of known and unknown modulation functions, can be observed in Figure 3 (a)-(c), respectively. It's crucial to emphasize that, as previously mentioned, the A tensor adopts a shape of $(1, W, 1)$, where W denotes the number of modulated areas on the SLM screen. Consequently, it can be reshaped to match the SLM screen shape for visualization purposes.

An intriguing observation is that the A tensor exhibits almost identical shapes in these two distinct networks, with the sole difference lying in their amplitudes—a point we will delve into later. Additionally, it's worth noting that the numerical aperture (NA) shape of the coupling optical element, likely the fiber, is mirrored in this tensor as a circular curve at the top of the reshaped A tensor. It's understood that all SLM pixels falling outside the accepted NA of the fiber would result in very weak optical coupling into the fiber, thus playing a less significant role in speckle generation compared to those with better coupling.

Figures 3 (d)-(f) demonstrate the non-linear modulation uncovered by the designed convolutional network. It's crucial to note that while this discovered modulation may deviate from the theoretical one, it does so solely through a linear mapping. This is evident from Equation 3, where substituting $f_r + jf_i$ with $L(f_r + jf_i) + B$, where B and L are complex-valued numbers, maintains the equation's structure. However, these parameters solely affect the values within the U_k, A, and Z_k tensors, influencing the non-linear modulation behavior via linear complex scaling.

In Figure 3(e), by setting L to $0.019 - 0.1j$ and B to $0.48 - 2.11j$, the discovered non-linear modulation aligns with the theoretical prediction. Here, the AI-discovered modulation is depicted in orange, while the theoretical modulation and the linearly scaled AI-discovered modulation are represented in blue and red, respectively.

Similarly, Figure 3(f) presents the same dataset in a three-dimensional axis, where the third dimension corresponds to the modulation parameter, a scalar number ranging from 0 to π.

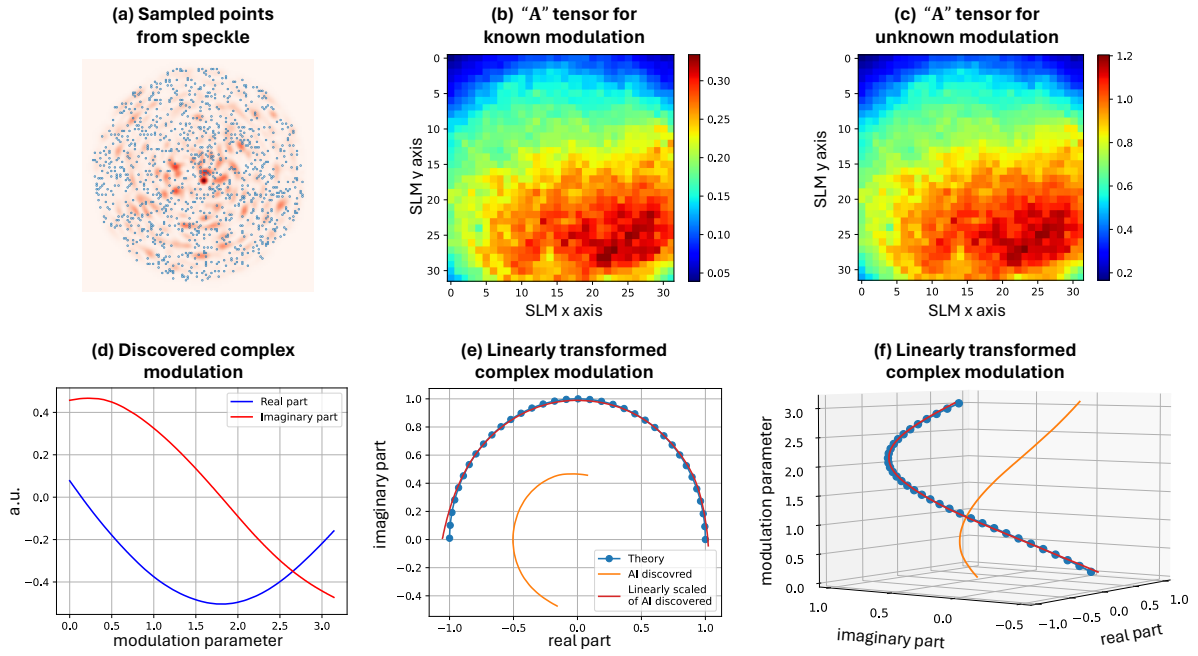

Figure 3: (a) Location of the randomly selected sampling points on the speckle. (b) and (c) Extracted tensor A for the cases of known and unknown modulation. (d) Discovered real (f_r) and imaginary parts(f_i) of the non-linear modulation. (e) Linearly scaled version of the discovered transformation and its comparison to the ground truth. (f) Same result as (e) with the modulation parameter also shown across the third dimension.

3.2 Polarization dependent transmission matrix

Figure 4 (a1) and (a2) shows the absolute value of the both discovred tesnors U_a amd U_b. In these figures the x axis is indicating the SLM pixel index while vertical axis shows the sampled point index from the speckle side. we can also calculate the cosine simality (CS) similarity matrix of these two tensor based on the follwoing eqution

$$CS_k = \hat{U}_k.\hat{U}_k^{\dagger} \tag{4}$$

In this analysis, $\hat{U}_k$ represents the normalized U_k tensor, where all rows have an amplitude equal to one, while $\hat{U}_k^{\dagger}$ denotes the complex conjugate transpose of $\hat{U}_k$. Figure 4 (b1) and (b2) display the absolute value of the calculated cosine similarity between U_a and U_b, with a closer examination provided in the zoomed versions (c1) and (c2).

An intriguing observation is the presence of parallel lines in the cosine similarity plot, particularly noticeable adjacent to the main diagonal elements. Remarkably, these lines are spaced at intervals of 32, indicating a relationship between pixels from the upper and lower rows of the SLM. This suggests that neighboring pixels on the SLM screen exhibit similarity in their transmission matrix rows, as illustrated in Figure 4 (a1) and (a2).

To further elucidate this phenomenon, we can extract a row from the cosine similarity matrix and reshape it back to match the SLM screen shape, as demonstrated in Figure 4 (d1) for the 500th row of Figure 4 (c1). Evidently, pixels close to the selected pixel display similar transmission matrix rows, whereas the similarity diminishes as we move further away. This observation cannot be attributed to the selected dataset alone, as there is no correlation between randomly generated pixel data. Instead, it stems from the underlying physics of the optical system, indicating that spatially close pixels on the SLM exhibit similar physical characteristics and effects in speckle generation.

Another intriguing observation arises when calculating the cosine similarity (CS) of each row in Figure 4 (a1) to the corresponding row in Figure 4 (a2). This yields 1024 values, each indicating the similarity between the governing equations for different polarizations. It's important to note that Figures 4 (a1) and (a2) depict the transmission matrix for perpendicular polarizations of light at the speckle side. These

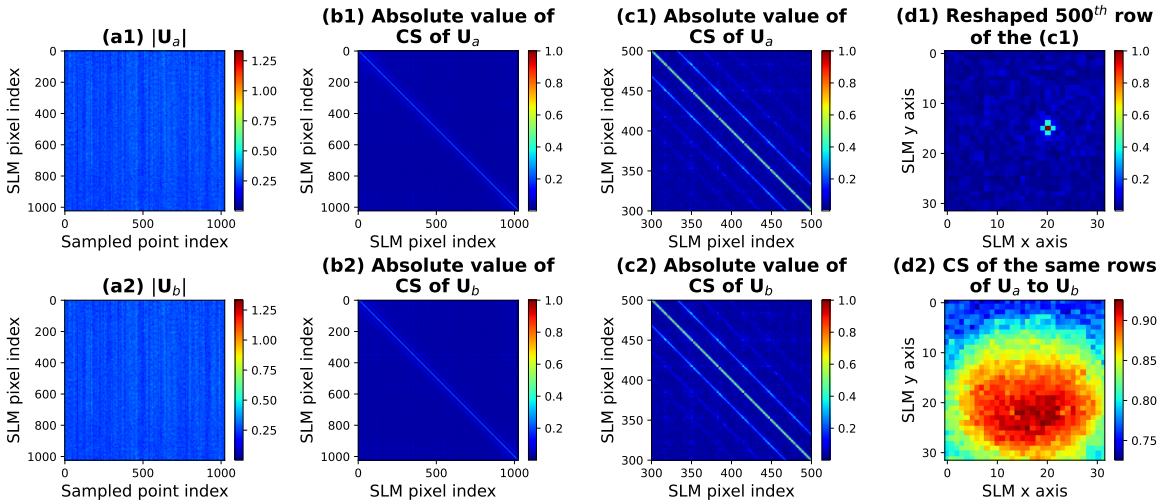

Figure 4: (a1) and (a2) The discovered transmission matrices for two perpendicular polarizations. (b1) and (b2) Absolute values of the cosine similarity matrices of (a1) and (a2), respectively. (c1) and (c2) Zoomed versions of (b1) and (b2), respectively. (d1) Reshaped version of the 500^{th} row of (b1). (d2) Absolute value of the cosine similarity of the same row from (a1) to (a2) reshaped back to the SLM's shape.

1024 values can then be reshaped to match the SLM screen shape and visualized as depicted in Figure 4 (d2).

Additionally, it's worth highlighting that each pixel of the SLM is coupled into the fiber at a different angle, based on the utilized optical system. Pixels located at the center of the numerical aperture are coupled perpendicularly to the distal fiber's facet axis, while pixels farther away are inserted at an angle. Figure 4 (d2) illustrates that when the insertion angle is perpendicular to the distal fiber facet, the output speckles are polarized, evident from the strong cosine similarity in the center of this figure (red color). Similarly, as the insertion angle increases (i.e., moving away from the center of the numerical aperture), the similarity between the two polarization transmission matrices decreases, resulting in polarization loss. This observation aligns well with the physics of the fiber, as perpendicular insertion angles lead to strong coupling of lower-order modes that maintain polarization better.

3.3 Digital-Twin of the scattering media

The primary function of this trained physics-informed model is to predict the speckle pattern based on the given SLM input. This capability is showcased across all testing datasets in Figures 5 and 6, where both 1024 and 10,000 sampling points from the speckle are analyzed.

In these figures, panels (a) to (d) illustrate the reconstruction of sampled speckle patterns and their disparities from the ground truth for SLM inputs derived from randomly generated images, CIFAR, celebrity faces, and MNIST, respectively. The associated reconstruction errors are graphically represented in panel (e1). Notably, the model exhibits superior performance with randomly generated images but faces more challenges with the MNIST dataset.

Moreover, Figures 5 and 6 provide UMAP projections of the sampled speckle from the MNIST dataset, with each projected speckle colored according to the handwritten digit it represents. Remarkably, both the ground truth (Figures 5 and 6 (e2)) and the reconstructed speckle (Figures 5 and 6 (e3)) exhibit identical topology and clustering of handwritten digits. This suggests that the reconstructed sampled speckle effectively captures the patterns and features of the MNIST dataset.

This promising outcome underscores the potential for reverse transformation using the trained model. In the next section, we will employ the model to predict the SLM pattern required to generate a given speckle pattern. Unlike previous deep learning models, this approach stands out as we leverage the physical equations of the fiber to address this problem.

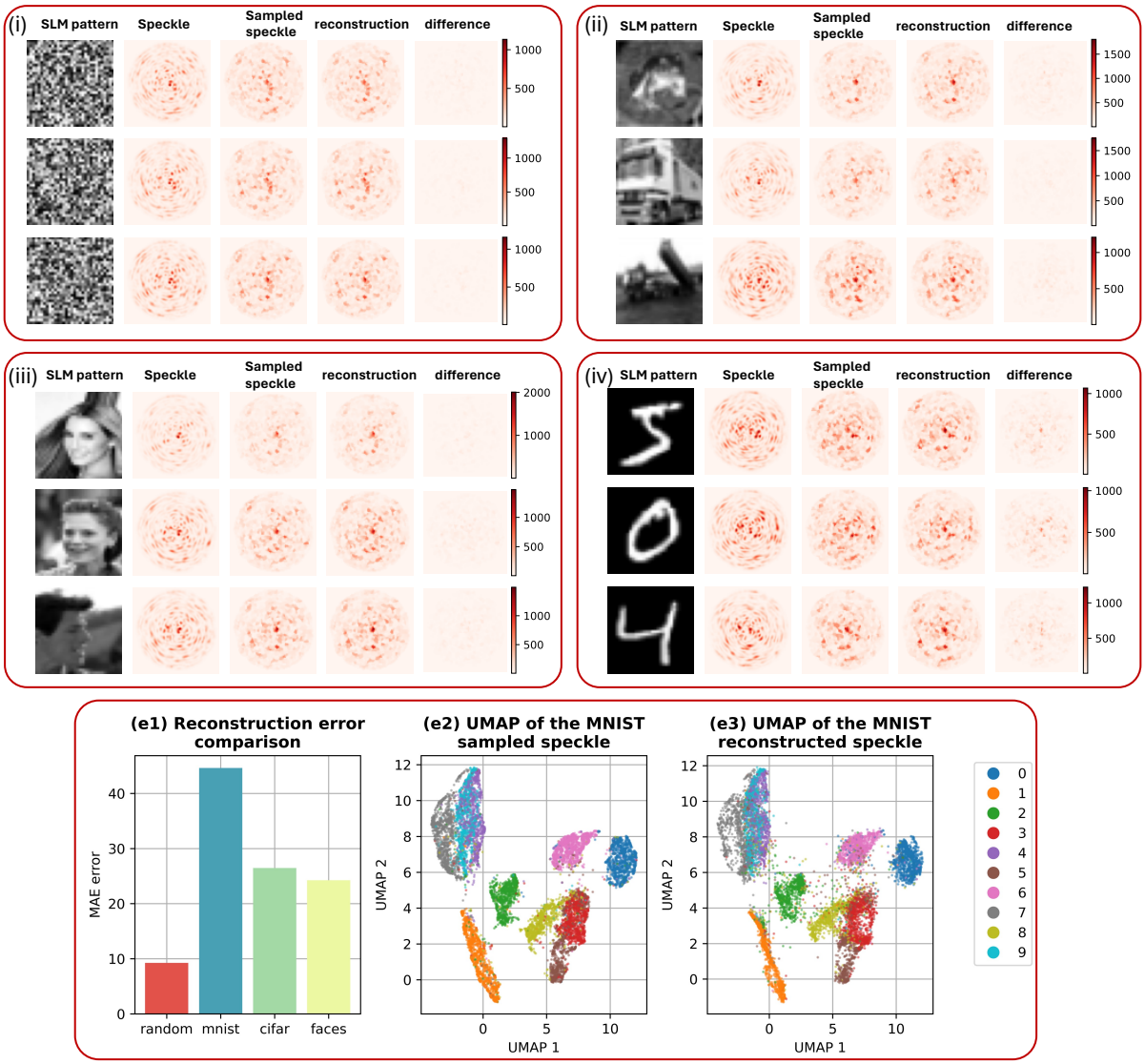

Figure 5: (a)-(d) Examples of speckle pattern reconstruction for randomly generated images, CIFAR, celebrity faces, and MNIST, respectively. This reconstruction was performed using only 1024 points of the speckle, and the plotted patterns are interpolated by the nearest neighbor method to produce an image. (e1) Mean absolute error (MAE) of the reconstructed speckle pattern. (e2) and (e3) UMAP (Uniform Manifold Approximation and Projection) of the ground truth speckle and the reconstructed one, respectively.

3.4 Universal pixel-wise image transmission through the fiber

As discussed previously, the presented digital twin of the optical system offers a significant advantage: the ability to compute gradients of its output, enabling various applications. One such application is image transmission through the fiber.

For this task, we propose utilizing the neural network depicted in Figure 7(a). This model comprises two elements. Firstly, a single-layer fully connected layer, consisting only of weights without biases, is connected to a constant input. The output of this layer is activated using a sigmoid activation function multiplied by the constant π. This ensures that the output of each node in the layer falls within the range of zero to π. This step aims to find the SLM pattern corresponding to the given speckle. Secondly, the output of this single layer is connected to the previously trained physics-informed model, which translates the given SLM pattern from the previous layer into speckle patterns (forward model). To prevent changes in the weights and biases of the physics-informed model during training, the model is frozen.

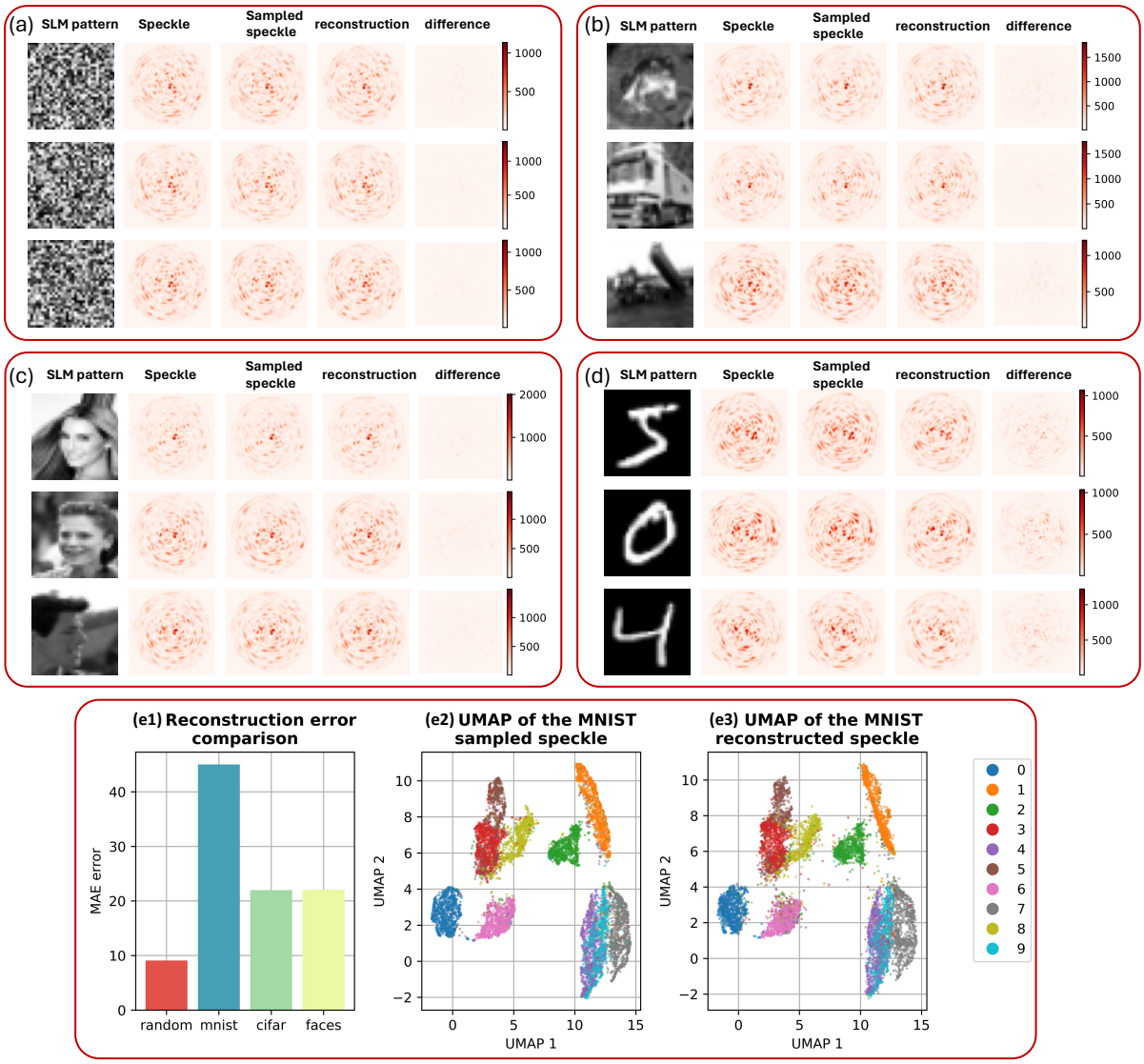

Figure 6: (a)-(d) Examples of speckle pattern reconstruction for randomly generated images, CIFAR, celebrity faces, and MNIST, respectively. This reconstruction was performed using only 10,000 points of the speckle, and the plotted patterns are interpolated by the nearest neighbor method to produce an image. (e1) Mean absolute error (MAE) of the reconstructed speckle pattern. (e2) and (e3) UMAP (Uniform Manifold Approximation and Projection) of the ground truth speckle and the reconstructed one, respectively.

Now that the entire model is drivable, we can apply gradient-based optimization to calculate the SLM pattern (through the single layer at the beginning) to produce a specific given speckle pattern (through the physics-informed model). However, this approach often leads to overfitting, posing a significant challenge. Although the physics-informed model accurately translates MNIST, CIFAR, or face images into their respective speckle patterns with an MAE loss exceeding 20, overfitting the single layer can yield a substantially lower loss, thereby producing inaccurate results.

To address this issue, we introduce a regularization method depicted in Figure 7(b). Leveraging the fact that the reconstructed image is stored in the weights of the single layer (due to the absence of biases), we extract these weights and reshape them into the expected image. Human-perceivable images exhibit grid topology and a sense of locality, meaning nearby pixels have higher correlation than those farther apart. Additionally, besides randomly generated data, images tend to be smooth, with variations in pixel values between adjacent pixels typically not strong.

This concept is operationalized through a kernel (weights) regularizer for the layer, where the spatial

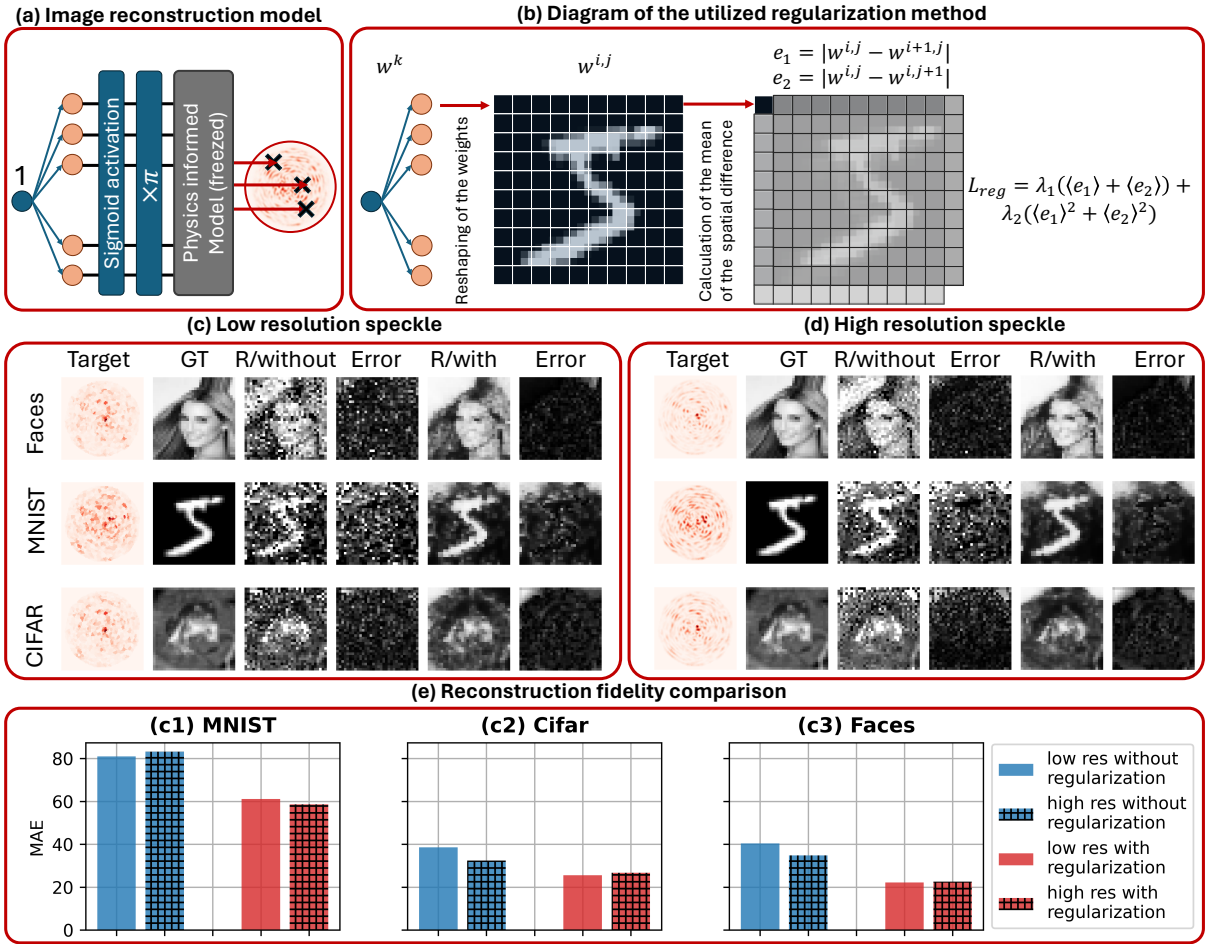

Figure 7: (a) Schematic of the model utilized for the image transmission task. (b) Schematic of the developed regularization technique. (c) An example of the reconstructed image of various datasets when only 1000 sampled points of the speckle have been used for the physics-informed model. This plot displays both image reconstruction and its corresponding spatial error when the proposed regularization has not been used (R/without) and the reconstruction when the proposed regularization has been used (R/with). (d) Same result as (c) but when the physics-informed model has been trained using 10,000 points of the speckle. (c1)-(c3) Image reconstruction fidelity in terms of the mean absolute error (MAE) for all the developed models for the MNIST, CIFAR, and celebrity faces datasets, respectively.

differences of the image in both horizontal and vertical directions are calculated. These differences, represented as e_1 and e_2 in Figure 7(b), are utilized to introduce another loss term during training, comprising weighted sums of the mean and square of these values. This combination forms a Lasso regularization, which should make sharp changes sparse in the resulting image, while the ridge (sum of squares) decreases the value of the differences.

Figures 7(c) and 7(d) demonstrate the advantage of this regularization in image transmission across all testing datasets in this study. It's evident that regularization enhances the performance of image reconstruction for all utilized testing sets. It's important to note that the physics-informed model is solely trained on randomly generated images, showcasing the universality of the image transmission model and its independence from dataset context.

3.5 Different optical setup

To assess the proposed method in a different optical configuration, we utilized data from [19]. Their experimental setup employed a one-meter-long step-index fiber with a core size of 105 μm, capable of carrying approximately 9000 optical modes, using a 532nm wavelength continuous laser. Their optical arrangement aimed to couple image intensity into a multimode optical fiber using a polarizing beam

splitter, a half-wave plate, and a microscope objective. The resulting speckle was captured using a 4f configuration and a digital camera. Their methodology involved a complex matrix correlating the

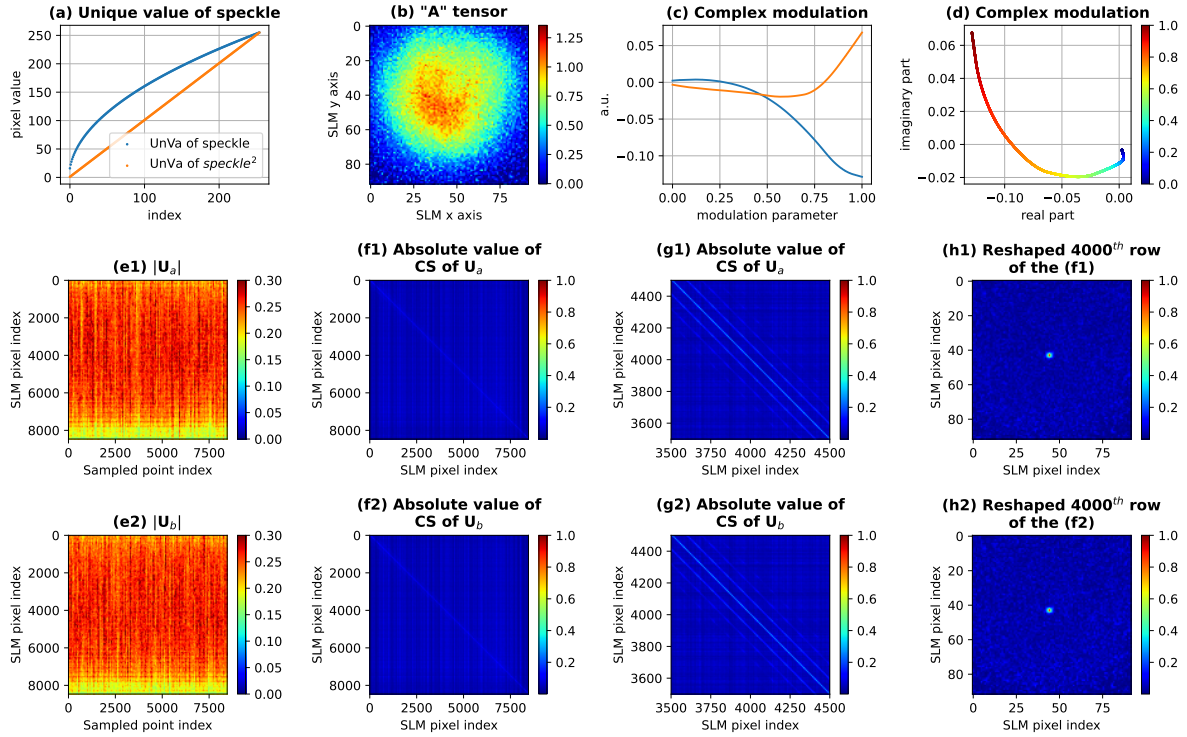

Figure 8: The discovered complex modulation and transmission matrices, along with their cosine similarity matrices, for the dataset used in the study by Caramazza et al. (2019) employing our proposed method.

image amplitude (square root of the intensity) at the two ends of the fiber, hence their provided data is amplitude-based. This is evident from the sorted unique values of the pixels in their recorded speckle, as illustrated in Figure 8(a) (in blue). The pattern formed by these blue values indicates the square root non-linear transformation over the data. However, when we calculate the square of the provided amplitude data and examine the sorted unique values, a straight line is formed (in orange), corresponding to intensity recording of the camera. It's important to note that our methodology requires speckle in intensity rather than amplitude (see Figure 2).

Regarding the input use case of amplitude or intensity for the SLM pattern in their dataset, the proposed convolutional network for an unknown modulation should be able to find the appropriate non-linear mapping during model training. For training, we randomly selected 10,000 samples from the speckle due to the higher number of modes that this optical setup can accommodate. Training was conducted over 100 epochs on their training dataset (45,000 natural science images from the ImageNet dataset), while testing was exclusively performed on their testing set (motion images in the Mybridge dataset). Here are the discovered characteristics of their system using our proposed physics-informed neural network.

Figure 8(b) depicts the discovered $\mathbf{A}$ tensor in the proposed algorithm. Similar to our example, this graph indicates lower coupling efficiencies due to the numerical aperture of the optical components, as the SLM pixels at the corners of the screen represent weaker coupling. Figure 8(c) shows the real and imaginary parts of the modulation in blue and orange, respectively, with the modulation parameter on the x-axis. Figure 8(d) displays complex modulation, with the real and imaginary parts of the modulation on the horizontal and vertical axes, respectively, while the color of each point represents the modulation parameter.

Figures 8(e1) and (e2) represent the transmission matrix for both polarizations, and similar to the previous optical setup, their cosine similarity matrices are calculated and represented in Figures 8(f1) and (f2), with a zoomed version in (g1) and (g2), respectively. As in the previous case, parallel lines appearing in the cosine similarity matrix are due to the grid topology of the SLM, with the distance

between lines now being 92 pixels, indicating the use of 92×92 pixel images in their experiment.

The reshaped versions of the 4000th row of the cosine similarity matrix to SLM screen shape for Figures 8(f1) and (f2) are also presented in (h1) and (h2), respectively. Similar to our setup, this demonstrates effective perpendicularity between the effects of the SLM pixels located far apart on the SLM screen.

3.5.1 Image transmission

We have also tested the developed physics-informed model for this optical setup in the intricate task of image transmission. The results and their comparison to the original article are illustrated in Figure 9.

From these results, it is evident that our proposed model exhibits a higher SSIM over their model, despite not being directly optimized for image transmission. It is important to note that all weights in our proposed model are optimized to achieve the best reconstruction of the speckle based on the given SLM pattern. However, image transmission represents the reverse task, mapping from speckle to the SLM pattern.

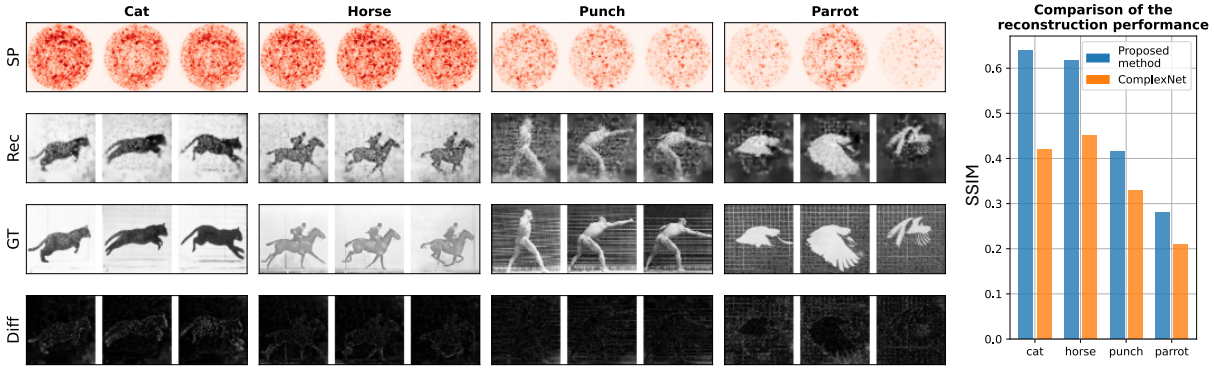

Figure 9: An example of image reconstruction using the proposed method for the testing dataset utilized in the study by Caramazza et al. (2019). A comparison with ComplexNet, developed by Caramazza et al., is also presented in the form of a bar chart.

The achieved improvement stems from a more accurate model of the physics of light scattering. In their method, they considered a complex matrix relating the absolute value of the input electric field amplitude to the output one, without considering phase information, which is not an accurate assumption. Moreover, our model considers different transmission matrices for different polarizations of light, unlike their model, which contributes significantly to the observed enhancement.

4 Conclusion

In this study, we introduce an innovative approach to characterizing turbid media without the need for a reference beam, thanks to the use of a physics-informed neural network. This method captures the polarization-based transmission matrix without requiring direct polarization measurement, providing a deeper understanding of light scattering and enabling the creation of an accurate digital twin for inverse problems such as image transmission.

Our approach diverges from traditional deep learning models, which often lack transparency and operate as black-box systems. These conventional models typically rely heavily on the data's features, limiting their generalizability and constraining them to their training datasets. In contrast, our method demonstrates the high precision with which transmission matrices can be extracted. These matrices enable not only the forward transmission of information from SLM pattern to speckle but also the precise reconstruction of the SLM pattern from a given speckle pattern, facilitated by our proposed regularization technique and the ability to compute gradients through the digital twin.

Consequently, our method offers a fully interpretable framework relevant to a range of light scattering applications, opening new opportunities for significant advancements in the field. Additionally, this approach is valuable for applications in physics-based neural networks where the system's gradient is required, providing a pathway to optical and photonics-based artificial intelligence.

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

References

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

[2] Di Leonardo, R. & Bianchi, S. Hologram transmission through multi-mode optical fibers. *Optics Express* **19**, 247–254 (2011). URL `http://www.opticsexpress.org/abstract.cfm?URI=oe-19-1-247`.

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

[4] Velsink, M. C., Lyu, Z. P., Pinkse, P. W. H. & Amitonova, L. V. Comparison of round- and square-core fibers for sensing, imaging, and spectroscopy. *Optics Express* **29**, 6523–6531 (2021).

[5] Li, S. H. *et al.* Compressively sampling the optical transmission matrix of a multimode fibre. *Light-Science & Applications* **10** (2021).

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

[7] Collard, L. *et al.* Wavefront engineering for controlled structuring of far-field intensity and phase patterns from multimodal optical fibers. *APL Photonics* **6**, 051301 (2021). URL `https://doi.org/10.1063/5.0044666`. `https://pubs.aip.org/aip/app/article-pdf/doi/10.1063/5.0044666/13869941/051301_1_online.pdf`.

[8] Collard, L. *et al.* Holographic Manipulation of Nanostructured Fiber Optics Enables Spatially-Resolved, Reconfigurable Optical Control of Plasmonic Local Field Enhancement and SERS. *Small* **n/a**, 2200975 (2022). URL `https://doi.org/10.1002/smll.202200975`.

[9] N'Gom, M., Norris, T. B., Michielssen, E. & Nadakuditi, R. R. Mode control in a multimode fiber through acquiring its transmission matrix from a reference-less optical system. *Optics Letters* **43**, 419–422 (2018). URL `https://opg.optica.org/ol/abstract.cfm?URI=ol-43-3-419`.

[10] Turtaev, S. *et al.* High-fidelity multimode fibre-based endoscopy for deep brain in vivo imaging. *Light: Science & Applications* **7**, 92 (2018).

[11] Wen, Z. *et al.* Single multimode fibre for in vivo light-field-encoded endoscopic imaging. *Nature Photonics* **17**, 679–687 (2023). URL `https://doi.org/10.1038/s41566-023-01240-x`.

[12] Cuche, E., Marquet, P. & Depeursinge, C. Spatial filtering for zero-order and twin-image elimination in digital off-axis holography. *Applied optics* **39**, 4070–4075 (2000).

[13] Plöschner, M., Tyc, T. & Čižmár, T. Seeing through chaos in multimode fibres. *Nature Photonics* **9**, 529–535 (2015).

[14] 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).

[15] Collard, L. *et al.* Optimizing the internal phase reference to shape the output of a multimode optical fiber. *PLOS ONE* **18**, 1–12 (2023). URL `https://doi.org/10.1371/journal.pone.0290300`.

[16] Jákl, P. *et al.* Endoscopic imaging using a multimode optical fibre calibrated with multiple internal references. *Photonics* **9** (2022). URL `https://www.mdpi.com/2304-6732/9/1/37`.

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

[18] Rahmani, B. *et al.* Actor neural networks for the robust control of partially measured nonlinear systems showcased for image propagation through diffuse media. *Nature Machine Intelligence* **2**, 403–410 (2020).

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

[20] Zhu, C. *et al.* Image reconstruction through a multimode fiber with a simple neural network architecture. *Scientific reports* **11**, 896 (2021).

[21] Xu, R., Zhang, L., Chen, Z., Wang, Z. & Zhang, D. High accuracy transmission and recognition of complex images through multimode fibers using deep learning. *Laser & Photonics Reviews* **17**, 2200339 (2023).

[22] Rahmani, B., Loterie, D., Konstantinou, G., Psaltis, D. & Moser, C. Multimode optical fiber transmission with a deep learning network. *Light: science & applications* **7**, 69 (2018).

[23] Collard, L., Kazemzadeh, M., Piscopo, L., De Vittorio, M. & Pisanello, F. Exploiting holographically encoded variance to transmit labelled images through a multimode optical fiber. *arXiv preprint arXiv:2309.15532* (2023).

[24] Wu, G., Sun, Y., Yin, L., Song, Z. & Yu, W. High-definition image transmission through dynamically perturbed multimode fiber by a self-attention based neural network. *Optics Letters* **48**, 2764–2767 (2023).

[25] Vaswani, A. *et al.* Attention is all you need. *Advances in neural information processing systems* **30** (2017).

[26] Chen, Y., Song, B., Wu, J., Lin, W. & Huang, W. Deep learning for efficiently imaging through the localized speckle field of a multimode fiber. *Applied Optics* **62**, 266–274 (2023).

[27] Liu, Z. *et al.* Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, 10012–10022 (2021).

[28] Goel, S., Conti, C., Leedumrongwatthanakun, S. & Malik, M. Referenceless characterisation of complex media using physics-informed neural networks. *arXiv preprint arXiv:2303.16041* (2023).

[29] Wright, L. G. *et al.* Deep physical neural networks trained with backpropagation. *Nature* **601**, 549–555 (2022).

[30] Teğin, U., Yıldırım, M., Oğuz, İ., Moser, C. & Psaltis, D. Scalable optical learning operator. *Nature Computational Science* **1**, 542–549 (2021).

[31] Oguz, I. *et al.* Forward–forward training of an optical neural network. *Optics Letters* **48**, 5249–5252 (2023).