PHOTONICS Research

Towards smart optical focusing: deep learning-empowered dynamic wavefront shaping through nonstationary scattering media

Yunqi Luo,[1,†] Suxia Yan,[1,†] Huanhao Li,[2,3,†] Puxiang Lai,[2,3,4] and Yuanjin Zheng[1,5]

[1]School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798, Singapore
[2]Department of Biomedical Engineering, The Hong Kong Polytechnic University, Hong Kong SAR, China
[3]The Hong Kong Polytechnic University Shenzhen Research Institute, Shenzhen 518034, China
[4]e-mail: puxiang.lai@polyu.edu.hk
[5]e-mail: yjzheng@ntu.edu.sg

Received 20 November 2020; revised 11 January 2021; accepted 21 January 2021; posted 22 January 2021 (Doc. ID 415590); published 16 July 2021

Optical focusing through scattering media is of great significance yet challenging in lots of scenarios, including biomedical imaging, optical communication, cybersecurity, three-dimensional displays, etc. Wavefront shaping is a promising approach to solve this problem, but most implementations thus far have only dealt with static media, which, however, deviates from realistic applications. Herein, we put forward a deep learning-empowered adaptive framework, which is specifically implemented by a proposed Timely-Focusing-Optical-Transformation-Net (TFOTNet), and it effectively tackles the grand challenge of real-time light focusing and refocusing through time-variant media without complicated computation. The introduction of recursive fine-tuning allows timely focusing recovery, and the adaptive adjustment of hyperparameters of TFOTNet on the basis of medium changing speed efficiently handles the spatiotemporal non-stationarity of the medium. Simulation and experimental results demonstrate that the adaptive recursive algorithm with the proposed network significantly improves light focusing and tracking performance over traditional methods, permitting rapid recovery of an optical focus from degradation. It is believed that the proposed deep learning-empowered framework delivers a promising platform towards smart optical focusing implementations requiring dynamic wavefront control. © 2021 Chinese Laser Press

https://doi.org/10.1364/PRJ.415590

1. INTRODUCTION

Light entering a disordered medium that is thicker than a few scattering mean free paths l (~0.1 mm for human skin) undergoes multiple scattering due to the mismatch of the refractive index [1], leading to pervasive obstacles in communication, astronomy, and high-resolution optical delivery and imaging through or within thick scattering media, such as biological tissues. If light is coherent, scattered light along different optical paths interferes randomly, forming optical speckles, whose intensity distribution can be recorded outside the medium using cameras. Although visually random, the way that light is scattered is actually deterministic within a certain time window (usually referred to as speckle correlation time) [2]. Built upon this property, various approaches have been inspired, such as time reversal [3–6], pre-compensated wavefront shaping [2,7–13], and memory effect [1,14–16], to obtain optical focusing and imaging through scattering media. Time reversal methods, such as time-reversed ultrasonically encoded (TRUE) method [17] and time reversal of variance encoded light (TROVE) [18], take advantage of guide stars (e.g., focused ultrasonic modulation) to encode diffused light; then, only the encoded light is time-reversed and focused inside the scattering medium. Pre-compensated wavefront shaping techniques modulate the phases of light incident into the scattering medium based on the measurement of the transmission matrix [8,10,11,19–21] or the maximization of feedback provided by the optical [7,22–25] or photoacoustic signal strength [2], with a goal to pre-compensate for the scattering-induced phase distortions. As for the memory effect, image information is encoded in the autocorrelation of the measured speckles as long as the imaging area is within the memory effect regime, and thus images can be reconstructed from speckles with iterative phase retrieval algorithms [1,26–29].

Each of the aforementioned approaches has its own advantages and limitations. For instance, pre-compensated wavefront shaping methods are attractive due to their plain working principle and experimental setup, but most reported approaches are inherently time consuming, as many iterations are required

regardless of the optimization algorithms [30,31], restricting most implementations reported thus far to static scenarios such as in fixed diffusers, which, however, scarcely exist in reality. Under the circumstance that scattering media are randomly changing or suffering from environmental disturbance that is inevitable, a focus will degrade or even vanish. To refocus light through/within time-variant media, the wavefront shaping iterations have to be repeated from the beginning each time the scattering medium changes, which is again a tedious and ineffective process [32]. This problem impedes the implementation of pre-compensated wavefront shaping from more general and realistic applications. Although imaging through non-static media has been explored with methods such as binary phase retrieval with optical phase conjugation [27,28,33,34], ghost imaging [35], shower-curtain effect [36], bispectrum analysis [37], advanced equipment [38], and memory effect [39], each has its limitations, such as the requirement for an ultrasound guide star, slow optimization, complex setup, and narrow effective regime.

Deep learning, which is a data-driven approach, has recently demonstrated wide uses to solve inverse problems like denoising [40], image reconstruction [41–46], and super-resolution imaging [47,48], owing to their superior ability in revealing complex relationships through transforming representations at one level to a higher and more abstract level [49]. The idea has also been exploited to focus light [50–52] and reconstruct images [53–55] through static scattering media. For example, Turpin *et al.* introduced neural networks for binary amplitude modulation and focused light through a single diffuser [50]; Li *et al.* trained U-Net with speckles generated by various objects with four diffusers [53]. The pre-trained network can be generalized to "unseen" objects or diffusers. All of these diffusers, however, are with the same macroscopic parameters. Sun *et al.* [56] trained five neural networks to model five different scattering conditions, and then blurred images are first classified to one of the five situations and then are fed into the pre-trained model for reconstruction. Note that, however, considering the computation time and memory budget, it is impractical to train hundreds of neural network models to cover all kinds of scattering conditions; considering only five conditions probably only get a rough classification and reconstruction.

In this paper, we aim to solve the problem comprehensively. We introduce a deep learning-empowered adaptive framework to tackle the challenge of optical focusing and refocusing through nonstationary scattering media by using wavefront shaping, which circumvents the dependency on classification or pre-trained models. A nonstationary process can be regarded as consisting of multiple piece-wise stationary stochastic processes, while the statistical properties of each stationary stochastic are analyzed to guide fine-tuning. The adaptive adjustment of hyperparameters of the proposed Timely-Focusing-Optical-Transformation-Net (TFOTNet), which is implemented by a multi-input-single-output deep convolutional long short-term memory (ConvLSTM) network, effectively circumvents the drawbacks of traditional long short-term memory (LSTM) that tends to remember only stationary variations [57]. The adaptive adjustment mechanism is non-trivial and depends on the statistical properties of a specific stationary stochastic process, which equivalently modifies the memory units in TFOTNet. Thus, modeling the spatiotemporal non-stationarity becomes possible. Another essential of the proposed framework is the recursive fine-tuning. It makes the best leverage of the correlation between medium statuses before and after the change, which is indicated by the speckle correlation [26,58–60]. Therefore, only a small amount of newly available samples are required to fine-tune the previous network, permitting fast recovery of the focusing performance. Note that during all of the phases, the medium is generally nonstationary; it keeps changing. Although recursive fine-tuning has already allowed timely focusing recovery, adaptive recursive estimation takes it one step further, efficiently balancing the trade-off between time cost and refocusing performance, allowing controllable light delivery through the time-variant scattering medium. It is worth highlighting here that the proposed adaptive framework becomes more attractive in circumstances with fast medium motion, considerable sudden disturbance, or low signal-to-noise ratio (SNR) regarding the light refocusing performance and time consuming in fine-tuning over traditional methods.

2. THEORETICAL ANALYSIS OF DEEP LEARNING FRAMEWORK FOR LIGHT FOCUSING AND REFOCUSING THROUGH NONSTATIONARY SCATTERING MEDIA

The scenario is that a monochromatic optical wave field propagates from the source to a randomly changing scattering layer at time t, and the transmitted scattered light is collected by a camera. Regular cameras only record the light intensity distribution of the speckle patterns on the receiving plane r_c (e.g., the camera plane in Fig. 1), and thus

$$I_c(t) = |E_c^{\text{out}}(t)|^2 = \left| \sum_a^N t_{ca}(t) E_a^{\text{in}}(t) \right|^2, \tag{1}$$

where $E_a^{\text{in}}(t)$ and $E_c^{\text{out}}(t)$ are the optical fields at $\boldsymbol{r}_a$ and $\boldsymbol{r}_c$ at time t, respectively [61], and $\boldsymbol{r}_a$ is the source plane [e.g., the spatial light modulator (SLM) plane in Fig. 1]. $t_{ca}(t)$ is a complex transmission coefficient describing light propagation from the source plane $\boldsymbol{r}_a$ to the receiving plane $\boldsymbol{r}_c$ at time t. The effects of absorption are neglected. To precisely compute the required incident complex optical field $E^{\text{in}}(t)$ with which light is focused to the position $\boldsymbol{r}_c$ through the current scattering medium, the inverse scattering model at time t has to be obtained based on the recorded transmitted light intensity distribution $I_c(t)$. As seen, the inverse scattering problem is nonlinear and ill-posed, which prohibits the adoption of direct inversion methods; iterative optimization with regularization is necessary [62] to resolve this problem. The objective is to find a desired reconstruction $E^{\text{in}}(t) = W(t)p(t)$ that minimizes the cost function formulated as

$$\arg\min_{E^{\text{in}}(t)} \|I_c(t) - |H(t)E^{\text{in}}(t)|^2\|^2 + \alpha(t)T[p(t)], \tag{2}$$

where $H(t)$ is the forward scattering model at time t, relating the transmitted light intensity $I_c(t)$ and the incident electrical field $E^{\text{in}}(t)$, $\alpha(t)T[p(t)]$ is the regulation, $\alpha(t)$ is the regulation coefficient, $W(t)$ is a convolutional transformation, and $p(t)$ is

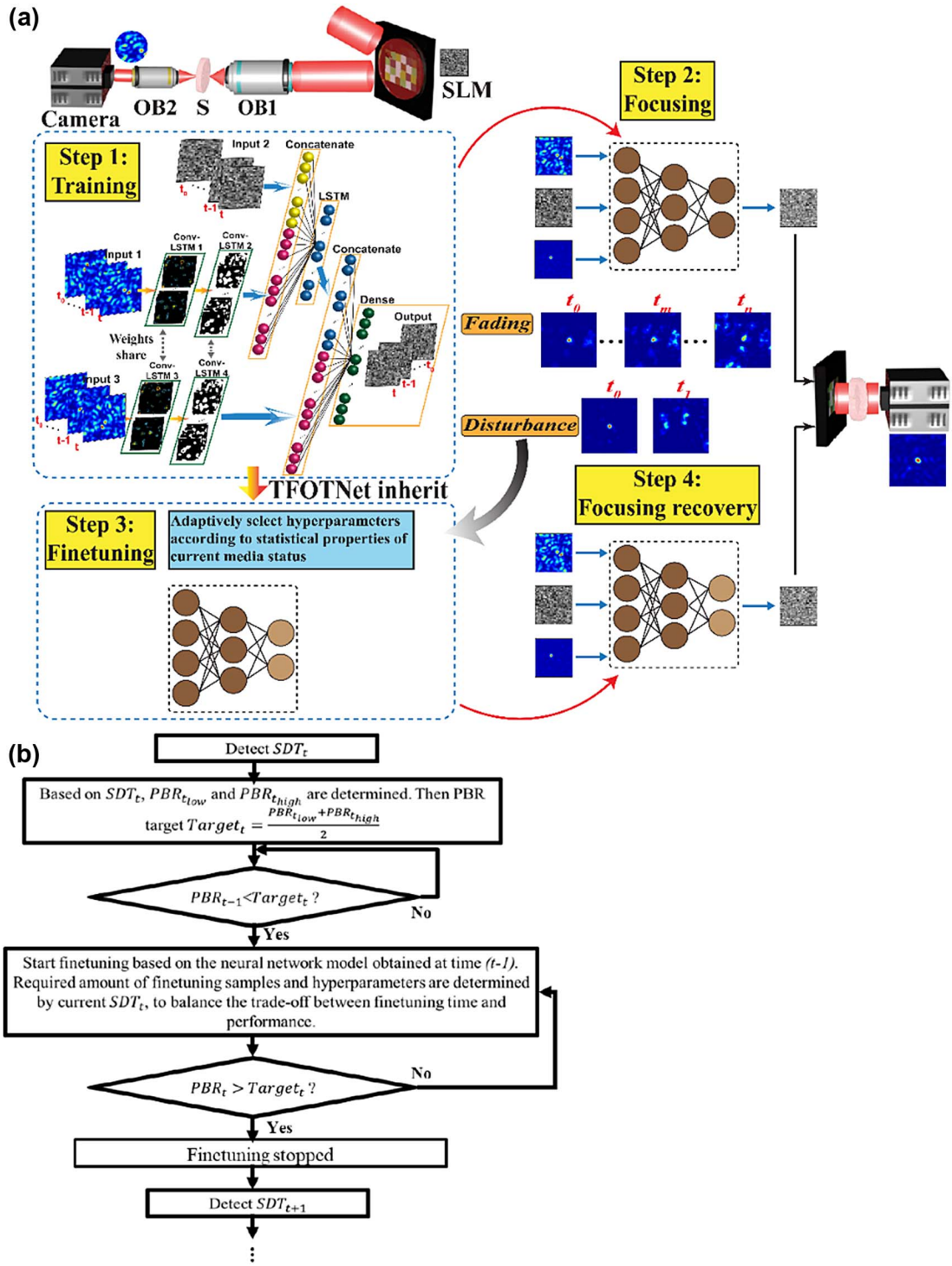


Fig. 1. Illustration of the proposed deep learning-empowered adaptive framework for wavefront shaping in nonstationary media. (a) General working principle of the proposed framework. In Step 1, samples are collected to train the TFOTNet. The structure of the proposed TFOTNet includes three inputs and one output. Input 1 is the speckle pattern, while the corresponding SLM pattern is noted as Input 2. Input 3 is the speckle pattern desired to be seen by the camera after light passes through the scattering medium in the experiment or simulation. TFOTNet output is the SLM pattern needed to get Input 3 through the present scattering medium. In Step 2, the well-trained TFOTNet can be applied to unseen speckles and output an SLM pattern that can obtain the target through the current medium. Inevitable environmental disturbance (disturbance) or non-stationary change in the medium (fading) results in degradation or even loss of the focal point. In Step 3, the pre-trained TFOTNet is fine-tuned with samples from the changing medium. Hyperparameters and fine-tuning sample amount are all adaptively chosen based on the medium status. After tuning, TFOTNet can adapt to the concurrent medium state and recover the optical focusing performance. (b) Flow chart of the proposed adaptive recursive algorithm for light focusing and refocusing in nonstationary media.

the transformation coefficients at time t. $E^{\text{in}}(t)$ consists of N input optical modes, $E^{\text{in}}(t) = [E_1^{\text{in}}(t)E_2^{\text{in}}(t)\cdots E_N^{\text{in}}(t)]$.

So far, a lot of iterative algorithms have been reported to solve the inverse problems in static situations, such as the distorted Born iterative method [63], subspace optimization method (SOM) [64], and iterative shrinkage and thresholding algorithm (ISTA) [65]. Most of them rely on a building block model [66], whereas, for dynamic media, medium statuses at time t and $t-1$ are correlated, indicating that $H(t)$ is not only determined by the current status but is also influenced by their previous values:

$$H(t) = g[\beta_1^{t-1}H(t-1) + \beta_2^t x(t)], \tag{3}$$

where β_1^{t-1} and β_2^t are time-dependent parameters, $g(\cdot)$ is a nonlinear function, $H(t-1)$ is the scattering model at time $t-1$, and $x(t)$ represents the information from the current scattering medium. Hence, in dynamic situations, Eq. (2) can still be solved using an iterative algorithm based on the building block model, but with temporal information included in it, and $p(t)$ at the $(m+1)$th iteration is given as

$$\begin{aligned} p^{m+1}(t) &= \left\{A_\theta\left\{\frac{1}{L}W(t) * H(t) * I_c(t) \right.\right. \\ &\quad \left.\left. + \left[I - \frac{1}{L}W(t) * H(t) * H(t)W(t)\right]p^m(t)\right\}\right\} \\ &= \left\{A_\theta\left\{\frac{1}{L}W(t) * g[\beta_1^{t-1}H(t-1) + \beta_2^t x(t)] * I_c(t)\right.\right. \\ &\quad + \left\{I - \frac{1}{L}W(t) * g[\beta_1^{t-1}H(t-1) + \beta_2^t x(t)]\right. \\ &\quad \left.\left.\left. * g[\beta_1^{t-1}H(t-1) + \beta_2^t x(t)]W(t)\right\}p^m(t)\right\}\right\}, \end{aligned} \tag{4}$$

where $L(t)$ is the Lipschitz constant at time t, $L(t) \le \text{eig}[W(t) * H(t) * H(t)W(t)]$. From Eq. (4), the iterative optimization process can be regarded as a sequence of linear filtering by kernel $I - [1/L(t)]W(t) * H(t) * H(t)W(t)$ and bias $[1/L(t)]W(t) * H(t) * I_c(t)$, followed by a point-wise nonlinear operation A_θ by value θ, and I is the identity matrix. Meanwhile, information from previous medium statuses is transmitted over time and imposes influence on the inverse scattering model at time t. Equations (2)–(4) suggest that TFOTNet, which is a multi-input-single-output ConvLSTM network, is suitable for solving the inverse scattering problems in dynamic situations. The inverse scattering problem is formulated as a regression task, while the learning process is evaluated by the mean integrated squared error [67,68]:

$$\text{MISE} = E\left\{\int_0^T \iint_{S_a} [y_t(r_a) - \hat{y}_t(r_a)]^2 \, \mathrm{d}^2 r_a \mathrm{d}t\right\}, \tag{5}$$

where $y_t(r_a)$ and $\hat{y}_t(r_a)$, respectively, denote the model prediction and true values at time t, r_a is a segment on the SLM plane S_a, and E denotes the expected value. To resolve wavefront shaping problems, $y_t(r_a)$ and $\hat{y}_t(r_a)$ are the predicted and true phase values of the incident optical mode $E_a^{\text{in}}(t)$, respectively.

The speckle correlation theory in random media suggests that, when the configurations of the scatterers are changed randomly, the scattering media before and after moderate change are correlated [69]. For dynamic media whose properties are time-variant, both spatial and temporal speckle correlations exist, and thus the speckle correlation is shown as [70]

$$\begin{aligned} C(t-t', r-r') &= \frac{\langle I(t,r)I(t',r')\rangle - \langle I(t,r)\rangle\langle I(t',r')\rangle}{\langle I(t,r)\rangle\langle I(t',r')\rangle} \\ &= C_1(t-t', r-r') + C_2(t-t', r-r') \\ &\quad + C_3(t-t', r-r'). \end{aligned} \tag{6}$$

According to the results reported by Feng *et al.* [69], the intensity correlation function $C(t-t', r-r')$ can be regarded as consisting of three contributions $C_1(t-t', r-r')$, $C_2(t-t', r-r')$, and $C_3(t-t', r-r')$, governing the short-range correlation, long-range correlation, and infinite-range correlation, respectively [71]. For most scattering media, the magnitude of $C_1(t-t', r-r')$, $C_2(t-t', r-r')$, and $C_3(t-t', r-r')$ decreases in sequence, but also decays more slowly with the increase of gap between t and t' or r and r' [70]. The proposed framework encodes the correlation between medium statuses, propagating the information over time; as a consequence, an accurate inverse model can be constructed.

3. RESULTS

A. Working Principle

The structure of the proposed TFOTNet is shown in Fig. 1(a). TFOTNet has three inputs and one output. Inputs 1 and 2 are paired, while Input 3 and the output are paired. Referring to Fig. 1(a), light is firstly reflected by the SLM, and its phase pattern is adjusted by the SLM; thus the optical phase patterns are represented by the SLM patterns. After the SLM, light will go through the diffuser and be scattered, forming speckles. The intensity distribution of the training speckle patterns will be recorded outside the diffuser by the camera, which is Input 1. The corresponding SLM pattern is Input 2. This forms a mapping from the training speckle pattern (Inputs 1) to the trained SLM pattern (Inputs 2), and it acts as a regularization term. Incorporating this regularization input into the TFOTNet, the targeted relationship from Input 3 to the output is obtained, and it is used to resolve the inverse scattering problems in real time based on the regularized cost function Eq. (2). Input 3 is the speckle pattern desired to be seen by the camera after light passes through the scattering medium in the experiment or simulation. The output of TFOTNet will be the corresponding SLM pattern that can lead to Input 3.

Inverse scattering problems are ill-posed, which may lead to difficulties in neural network training [72]. Offering prior information to regularize the inverse problem can mitigate the burden in training, which plays a significant role in successfully resolving inverse problems [73,74]. Besides setting analytic priors manually, it has also been reported that prior terms can be directly learned during the training of neural networks, which is tailored to the statistics of the training images, indicating a stronger regularization [75,76]. Chang *et al.* adopted an adversarial method to jointly train two networks, where one offers prior information, while the other one conducts inverse projection [76]. Inspired by these, the proposed TFOTNet consists of two parts: prior knowledge about scattering provided

by Inputs 1 and 2, and inverse mapping from Input 3 to the output. Through training, the network learns to extract suitable priors from Inputs 1 and 2, and they are passed to facilitate the resolving of the inverse problem represented by Input 3 and the output, alleviating the training burden and improving the modeling accuracy when compared with methods that directly learn the inverse mapping without any other knowledge.

Generally, in transfer learning, only the last few layers, rather than whole neural networks, are fine-tuned [77,78], as the last layers are task specific, while the earlier ones are modality specific [79]. Information learned by earlier layers can be shared among all inverse scattering problems, while the last few layers are customized for adapting to specialized changing conditions. Therefore, when the TFOTNet needs to be fine-tuned in the experiment, only the last layer in the TFOTNet is adjusted, while all other layers are frozen. By doing so, both time and computational resources can be saved without significant sacrifice of accuracy. The two ConvLSTM layers, ConvLSTM1 and ConvLSTM2, extract and abstract image features from Input 1; meanwhile, they pass the useful features from previous statuses throughout the network. Then, these features are flattened to concatenate with Input 2, which has also been flattened. The combination serves as the input to the first LSTM layer, followed by a dropout layer. The outputs of the LSTM layer concatenate with the features gathered from Input 3. The final TimeDistributed dense layer predicts the SLM pattern needed for Input 3 in the current situation. ConvLSTM1 and ConvLSTM2 consist of 16 and 32 filters, respectively, and the filter size of each layer is 7×7 and 5×5 with the stride setting as 3×3 and 2×2, respectively. ConvLSTM1 and ConvLSTM3, ConvLSTM2 and ConvLSTM4 share the same structure and weights, respectively. The number of neurons in the LSTM layer is 256 with a dropout rate set to 0.3. The number of neurons in the output layer is the same as the size of the SLM patterns. Kernel initializers of all layers are set as glorot normal. Mean squared error is employed as the loss function. Adam is used as the optimizer with alpha, beta1, beta2, and epsilon set as 0.0005, 0,9, 0.99, and 0.0001, respectively. It is worth noting that the proposed TFOTNet is a general network that can be applied to deal with speckle images and SLM patterns of arbitrary size. Herein, we just introduce a specific implementation for our typical setup as a proof-of-concept. The output size of TFOTNet is determined by the size of the SLM patterns, which is user defined. Kernel size is adjusted in the light of the relative size of speckle grains and recorded speckle images. In general, with smaller speckle grains, both the kernel size and stride have to be reduced accordingly. For larger speckle images or SLM patterns, naturally, more training and fine-tuning samples are required; meanwhile, the number of neurons will be increased, and the dropout rate is enlarged as well to avoid overfitting. The activation function of all layers is tanh, except for the last output layer whose activation function is sigmoid. The recurrent activation function of all ConvLSTM layers is set as hard sigmoid. The TensorFlow Keras library is used to construct the model.

It is worth noting here that the lynchpin of the proposed framework is the adaptive recursive fine-tuning system, rather than any other specific implementations. Despite this, TFOTNet includes information from Inputs 1 and 2 to facilitate the ill-posed inverse mapping from Input 3 to the output; thus, it not only allows for more efficient modeling over conventional single-input-single-output ConvLSTM network or convolutional neural network (CNN) (both simulation and experimental comparisons are shown in the following content), but is also a general network whose structure is scalable to accommodate various applications. Considering that SLMs are widely used to modulate incident optical wavefronts, as shown in Fig. 1(a), in this article, we employ SLM patterns to represent the incident optical phase patterns.

The working flow of the proposed adaptive deep learning framework for light focusing and refocusing in nonstationary media is illustrated in Fig. 1(a). First, samples are collected for TFOTNet training and initialization. After that, the well-trained TFOTNet is able to establish an inverse scattering model statistically that can accurately map the intensity distribution of speckles to their corresponding SLM patterns. Then, the desired speckle (a preset focused speckle pattern is used here) is sent to the TFOTNet through Input 3, and the TFOTNet outputs the SLM pattern that is required to restore the desired pattern for the current scattering system. Considering that the scattering media are nonstationary, and environmental perturbations with time are inevitable, an optical focus can be faded or even lost. To cope with it, ad hoc samples from the real-time medium are offered to recursively fine-tune the TFOTNet that is obtained previously. Meanwhile, hyperparameters are all adaptively chosen according to the instant status of the medium. During the fine-tuning phase, only the weights of the last layer in the TFOTNet are adjusted while all other layers are frozen. After the directed adjustment, the fine-tuned TFOTNet will be able to produce an SLM pattern to recover the focusing performance in a short period of time.

Figure 1(b) elaborates the proposed adaptive recursive algorithm to handle the spatiotemporal non-stationarity. In the whole article, light focusing performance is quantitatively evaluated by the peak-to-background-ratio (PBR), which is defined as the ratio between the intensity of the focal point and the mean intensity of the background [80]. Medium changing speed is characterized by speckle decorrelation time (SDT), which is defined as the time duration that the intensity autocorrelation function decreases from 1 to $1/e$ of its initial value [81]. Smaller SDT corresponds to faster medium altering speed. At time t, the SDT of the current medium state is computed, and the PBR target, Target_t, is determined based on the SDT (the method to calculate the PBR target is elaborated in Section 4). The PBR target indicates the pre-defined PBR that targets to be attained after light refocusing. The adaptive PBR target is employed to balance the trade-off between the fine-tuning cost and the focusing recovery performance. More samples are needed to enhance the PBR to a pre-defined level when the scattering medium changes faster, suggesting that longer time is required; thus focusing tracking performance will be affected. With an adaptive PBR target, a faster SDT accommodates to a relatively lower PBR target, which needs fewer fine-tuning samples, shortening the fine-tuning time. Instant PBR is

compared with Target_t, and fine-tuning will not be initialized until the instant PBR is lower than Target_t. Pairs of the SLM pattern and the corresponding speckle pattern are collected during the changing process of the scattering medium for fine-tuning, and the required fine-tuning sample amount and hyperparameters of the network are all chosen based on the SDT. The influence of hyperparameters on fine-tuning time is discussed in Section 4. The recursive algorithm indicates that the fine-tuning is based on the network obtained at time $t-1$, which can make the best use of the speckle correlation; thus, time cost in fine-tuning can be significantly reduced when compared with traditional iterative algorithms. Once the instant PBR after fine-tuning is higher than Target_t, fine-tuning will be ceased. With iterations of such an adaptive recursive fine-tuning process, optical focus can be recovered from deterioration in time, allowing for maintenance of a focal point with acceptable performance. Proof-of-concept simulation and experimental results with a typical setup are shown below as a verification. It should be highlighted here that all results demonstrated in the article, such as the PBR target and the amount of fine-tuning samples, are valid in all conditions with proper scaling in the light of specific implementations, not just limited to the setup used here. The methods of scaling will be discussed in the experiments part.

B. Simulation and Experimental Results

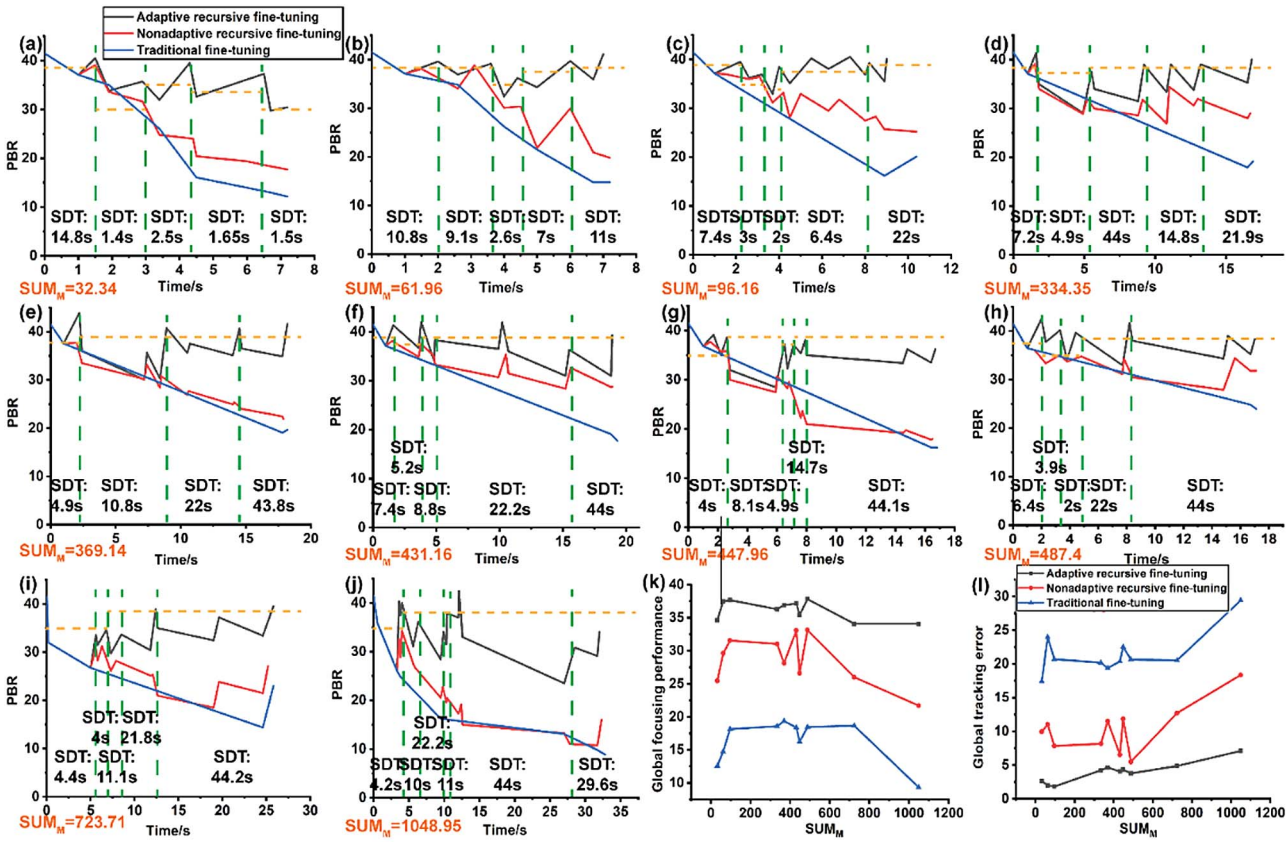
1. Simulation Results

Proof-of-concept continuous nonstationary processes are simulated to clearly manifest the effect of the adaptive recursive fine-tuning algorithm in dealing with spatiotemporal nonstationarity. One nonstationary course can be regarded as consisting of multiple piece-wise stationary stochastic sub-processes characterized by different SDT. Meanwhile, the time duration of each stationary sub-process also varies. To simulate the scattering process, a transmission matrix $\text{TX}(t)$ is used to describe a disordered medium at time t, following a circularly symmetric Gaussian distribution [23]. For a medium that is not static, the medium status at time $t+\Delta t$ is represented by $\text{TX}(t)+\Delta\text{TX}(\Delta t)$, and $\Delta\text{TX}(\Delta t)$ also follows a circularly symmetric Gaussian distribution. $\Delta\text{TX}(\Delta t)$ of different variances is employed to model media of various altering speeds. The size of the SLM patterns is set as 32×32, while the size of the speckle patterns is 64×64. First, in Step 1, a total of 10,000 samples are created for TFOTNet initialization and training. Sample collection time is estimated using the maximal frame rate of a commercial liquid crystal on silicon (LCoS) SLM, which is generally 60 Hz, and the SDT of the medium is as long as 10 min, resulting in correlation between medium statuses when the training sample collection initiates and ends reaching 0.8. After training, a desired speckle pattern [as shown in Fig. 1(a)] is sent to the well-trained TFOTNet, and a focused speckle can be obtained with the predicted SLM pattern. The PBR of the focused speckle obtained with the original trained model is 41.5. Since the proposed network is scalable, with the SLM pattern or speckle image size becoming larger, it can be expected that PBR will also increase. It is worth noting that the sample collection speed can be expedited nearly 400 times if faster modulators such as a digital micromirror device (DMD) are applied to conduct wavefront modulation whose frame rate can reach 23 kHz [82].

However, the PBR is decreasing, and the focal point is fading over time, as the scattering medium is altering over time, calling for the necessity of fine-tuning after a certain period of time. Herein, 10 nonstationary processes are randomly generated. Three fine-tuning algorithms, i.e., the adaptive recursive algorithm, nonadaptive recursive algorithm, and traditional algorithm, demonstrate significant differences in the PBR recovering performance, which is discussed separately in the following. Simulated focusing recovery results with three algorithms using TFOTNet in 10 nonstationary processes are shown in Figs. 2(a)–2(j), where the same legends are used. For each steadily altering sub-process, its PBR target (the ideal case) is indicated by the yellow dashed line. Six SDT intervals are used, and their corresponding PBR target is given in Table 1, serving as a criterion to evaluate the focusing recovery performance (details about the calculation of an adaptive PBR target are elucidated in Section 4). Although different setups result in different initial focused speckles, the ratio between the PBR target and PBR of the initial focused pattern remains unchanged as long as the SDT is the same. By doing so, the presented results can be safely scaled to any other implementation. With the adaptive recursive algorithm, the hyperparameters are selected on the basis of statistical properties of each stationary variation; with the nonadaptive recursive algorithm, the hyperparameters remain as the default values (shown in Fig. 2) all the time, regardless of the SDT changes. For a fair comparison, each time adaptive fine-tuning is conducted the same samples are offered to the nonadaptive recursive algorithm to do one fine-tuning as well. As for the traditional algorithm, it is not a recursive one; instead, traditional wisdoms only conduct fine-tuning once in the last sub-process from the original trained model due to the lack of engines of sensing intermediate processes, the hyperparameters remain as the default values, and total sample amount is the same as those in the recursive methods. It can be seen from Figs. 2(a)–2(j) that, qualitatively, PBR achieved by the adaptive recursive approach (gray line) is always the highest, while the traditional fine-tuning algorithm (blue line) demonstrates the worst performance. In view of the structure of TFOTNet, which consists of multiple ConvLSTM cells, adjusting those hyperparameters equivalently modifies the memory units in effect. By analyzing the statistical properties (including mean value and autocovariance) of diverse stationary variations, hyperparameters can be adjusted adaptively, compensating for the limitations of the conventional ConvLSTM network that it lacks the capability to encode high-order nonstationary spatiotemporal variations [57].

Table 1. SDT-Dependent PBR Target

SDT/s	SDT ≥ 7.4	7.4 > SDT ≥ 4.4	4.4 > SDT ≥ 2	2 > SDT ≥ 1.6	1.6 > SDT ≥ 1.4	1.4 > SDT
PBR target	38.75	37.5	35	33.75	30	27.5



	Initial learning rate	Batch size	Timestep
Default value	0.0025	64	4

Fig. 2. Fine-tuning results with ten random nonstationary processes using three different algorithms. The 10 nonstationary processes can be regarded as consisting of multiple piece-wise stochastically stationary sub-processes, while the SDT and time duration of each sub-process are different. (a)–(j) Fine-tuning results with the adaptive recursive algorithm (gray line), nonadaptive recursive algorithm (red line), and traditional fine-tuning algorithm (blue line) in the 10 nonstationary processes. Each process is characterized by SUM_M, which is the sum of the product of SDT of each sub-process and its time duration. Figures (a)–(j) use the same legend. (k) Global focusing performance of three fine-tuning algorithms in the ten random nonstationary processes. (l) Global tracking error of three fine-tuning algorithms in the ten random nonstationary processes. Figures (k) and (l) use the same legend. The inserted table lists the default values of these hyperparameters used in simulation. Specific hyperparameters and fine-tuning sample amounts used in the simulation are available in Ref. [83].

Quantitatively, one nonstationary process is characterized by the sum of the product of SDT of each sub-process and its time duration:

$$\mathrm{SUM}_M = \sum_{j=1}^{M} \mathrm{SDT}(j) \times \mathrm{Duration}(j), \tag{7}$$

where M represents the total number of stationary stochastic sub-processes contained in a nonstationary course. Lower SUM_M values suggest smaller SDT, shorter time duration, or both. To quantitatively evaluate the overall light focusing maintenance performance, the mean value of PBR over the whole nonstationary process is adopted:

$$\mathrm{GFP}\ (\text{global focusing performance}) = \sum_{i=1}^{N} \mathrm{PBR}(i)/N. \tag{8}$$

With the three algorithms, PBRs at the time when each fine-tuning starts and ends are recorded, and then all of the recorded PBR values are averaged, serving as the global focusing performance (GFP). Therefore, N is not a fixed value; instead, it is determined by how many times fine-tuning has been conducted in a nonstationary course. It is obvious that the higher the GFP, the better the focusing across time. The GFP of the three algorithms in these 10 nonstationary processes is illustrated in Fig. 2(k). Figures 2(k) and 2(l) use the same legend. In all cases, the adaptive recursive algorithm demonstrates the highest GFP, ranging from 34 to 38, while the traditional fine-tuning algorithm shows the worst results with GFP fluctuating between 9 and 19. The GFP enhancement percentage achieved by the adaptive and nonadaptive recursive algorithms over the traditional one is 82%–264% and 39%–132%, respectively, which is calculated by GFP enhancement percentage $= \left(\frac{\text{GFP with adaptive or nonadaptive recursive algorithm}}{\text{GFP with traditional algorithm}} - 1\right) \times 100\%$. These results manifest the merits of the adoption of the recursive approach, where they can make the best use of both temporal and spatial correlations of the medium for fine-tuning, allowing much better focusing recovery. Meanwhile, the adaptive

algorithm realizes 12%–57% GFP enhancement over the non-adaptive recursive algorithm, confirming that through the adaptive adjustments of hyperparameters, spatiotemporal non-stationarity can be better learned by neural networks. As for the traditional algorithm, medium status at the time when fine-tuning is conducted has little correlation with the status at the time when the initial model is obtained ($t = 0$), which will definitely lead to poor fine-tuning performance.

In addition, the root mean squared error with respect to the adaptive PBR target is employed to measure the tracking performance of an algorithm over the whole nonstationary processes:

$$\text{GTE (global tracking error)} = \sqrt{\frac{\sum_{i=1}^{N} [\text{PBR}(i) - \text{PBR target}(i)]^2}{N}}. \tag{9}$$

Small global tracking error (GTE) means that over the whole process, the PBR fluctuates mildly around the target, indicating an accurate and timely tracking of the optical focusing. The GTE of the three algorithms in these 10 non-stationary processes is illustrated in Fig. 2(l). As seen, the adaptive recursive algorithm shows the lowest values (1.8–7.1), while the traditional algorithm shows the largest values (17–29), as fine-tuning is conducted without considering process variations. The reduction percentage in GTE achieved by the two recursive algorithms over the traditional one is computed by $\text{GTE reduction percentage} = \left(1 - \frac{\text{GTE with adaptive or nonadaptive recursive algorithm}}{\text{GTE with traditional algorithm}}\right) \times 100\%$. As seen, the adaptive and nonadaptive recursive methods reduce the GTE over the traditional algorithm by 76%–91% and 37%–73%, respectively, suggesting that recursive algorithms demonstrate much better focusing tracking performance. Meanwhile, the adaptive recursive algorithm realizes 31%–81% reduction in GTE over nonadaptive recursive results. Due to the influence from environmental disturbance, PBR may drop sharply [as seen in Figs. 2(d)–2(j)], indicating that more time is required to recover the PBR. Therefore, the increase of GTE in these circumstances is inevitable, as shown in Fig. 2(l). In situations without perturbations [as shown in Figs. 2(a)–2(c)], PBR can be well maintained to be around the target with the proposed adaptive recursive algorithm, and the GTE is as low as 1.8–2.6. It should be noted that the conclusions deduced above are effective for all implementations of the proposed deep learning-empowered adaptive framework, not just limited to the setup and random nonstationary processes demonstrated in the article. For all realizations, similar performance improvements can be obtained.

It is worth noting that the fine-tuning time can be significantly reduced, and the SDT can be much smaller than values shown here if faster modulators and/or more powerful computation engines are adopted. As a proof-of-concept, in the simulation, the sample collection speed is estimated based on the maximal frame rate of commercially available LCoS-SLM, which is 60 Hz. As for computation, the TensorFlow Keras library is adopted, and the computing unit is an Acer Predator G9-792, 16 Gb RAM, and a GTX 980M graphics processing unit (GPU). However, if a DMD is applied to conduct wavefront modulation, whose frame rate can reach 23 kHz [82], together with onboard data acquisition, sample collection can be expedited by nearly 400 times. Furthermore, if a more powerful GPU or workstation such as the Nvidia Tesla series is employed, the computation speed will be improved by at least three times. Thus, the fine-tuning process can be speeded up by nearly 1000 times, indicating that the proposed framework becomes likely to achieve wavefront shaping in dynamic situations, such as *in vivo* tissues that decorrelate as fast as several milliseconds [84].

In Fig. 2(i), due to the sharp PBR drop at the beginning (from 41.5 to 26.7) and fast medium change (SDT = 4.4 s), even though recursive fine-tuning is conducted, the target PBR cannot be reached. Nevertheless, one attractive property of the adaptive method is that once a slower changing sub-process is detected, it is capable of making up the earlier PBR loss. As seen, with the fourth sub-process whose SDT has increased to 21.8 s, PBR is enhanced to meet the target, reaching 38.9. In contrast, the refocusing ability of nonadaptive algorithms is exacerbated, and they never meet the PBR target. Nonadaptive algorithms lack the ability to sense the current situation and make adjustments accordingly. Instead, it applies the same fine-tuning system to all processes regardless of their SDTs, which will definitely result in modeling deficiency.

2. Experimental Results

After the verification with simulations, experiments are conducted. The experimental setup is illustrated in Fig. 3. Light emitting from a He–Ne CW laser (633 nm, Melles Griot) is expanded by a telescope by 4.3 times. Then, a half-wave plate and a polarizer are followed to adjust the polarization of the incident light to be parallel to the long axis of an SLM (X13138-01, Hamamatsu). The light wavefront is modulated by the SLM, after which light passes through two successive lenses and is focused onto the surface of a diffuser (ground glass of 120 Grit, Edmund) by an objective lens (TU Plan Fluor

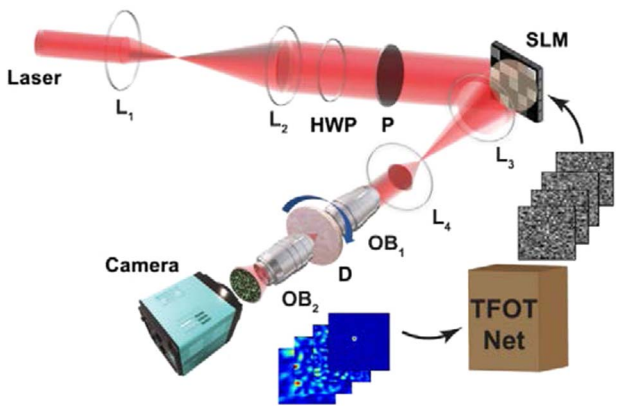


Fig. 3. Schematic of the experimental setup. Light is expanded by two lenses (L_1 and L_2), and then a half-waveplate (HW) and a polarizer (P) adjust the polarization state of the light incident onto the spatial light modulator (SLM). Light is modulated and reflected by the SLM, then passes through two lenses (L_3 and L_4), and is focused onto a diffuser (D) surface by an objective lens (OB_1). Scattered light is collected by another objective lens (OB_2) and recorded by a camera.

50×/0.80, Nikon). The light undergoes multiple scattering from the diffuser, and then the scattered light is collected by another objective lens (TU Plan Fluor 20×/0.45, Nikon) placed behind the diffuser. Finally, the speckles are recorded via a camera (Zyla s4.2, Andor). The resolution of the SLM screen is 1280 × 1024, and it is divided to 32 × 32 macropixels to display the SLM patterns, i.e., one macropixel contains 40 × 32 pixels. The dimensions of the speckle patterns recorded by the camera are 64 × 64 pixels. In experiment, we use 32 gray steps in the SLM to represent phase values from 0 to 2π. Due to the limitation of precision from the rotating stage (Motorized Precision Rotation Stage PRM1/MZ8, Thorlabs), the diffuser is rotated once as 100 samples are collected, and the equivalent rotating speed varies from 2.5 to 10 mdeg/s. Note that the nominal frame rate of the SLM is 60 Hz. However, due to the existence of the rising/falling transition time of the SLM, as well as limitations posed by the camera exposure time and the transmission speed between the laptop and the system, the frame rate achieved in operation is only ~6 Hz. This has restricted the medium change speed demonstrated in the current phase of the experiment. The diffuser is rotating at different speeds to create various SDTs. As the frame rate is 10 times slower than that used in simulation estimation, SDT in simulation has to be enlarged 10 times accordingly to be consistent with the experimental conditions. For instance, in the experiment, the ratio between the PBR target and the initial PBR in a stationary stochastic process whose SDT is 200 s should be the same as that in the stationary variation whose SDT is 20 s in simulation. However, it is worth noting that there is no fundamental limitation on the speed and performance of the proposed framework if faster modulators such as DMD can be applied, and hence the results demonstrated here are scalable.

Three experiments with and without environmental disturbance are conducted, respectively, and, in each experiment, the adaptive recursive algorithm and traditional algorithm are investigated for comparisons. The proposed framework consists of two parts: recursive fine-tuning and adaptive adjustments of hyperparameters. It has been proved by simulation that hybridizing these two systems will demonstrate better performance in handling non-stationarity rather than only employing recursive fine-tuning; thus the nonadaptive recursive algorithm is not applied to experiments. For a fair comparison, the total fine-tuning samples of the adaptive recursive algorithm and traditional algorithm during a nonstationary process are the same, and both algorithms are implemented with TFOTNet. Experimental results are shown in Figs. 4 and 5. Results without environmental perturbations are shown in Figs. 5(a)–5(c), while results with disturbance are given in Figs. 5(d)–5(f). The same legend is used in these figures. Figures 4(a) and 4(b) indicate the GFP and the GTE of the six experiments, respectively. The SDT of each sub-process is shown in the figures, while the PBR target (the ideal case) is indicated by the yellow dashed line. In all experiments, the first step is TFOTNet initialization and training using 10,000 samples to obtain a focused speckle, and the initial PBR is displayed in the figures at $t = 0$. As stated, the PBR target is determined by the SDT as well as the PBR of the initial focused speckle. For media of the same SDT, the ratio between the PBR target and the initial PBR always remains the same. Thus, the PBR target can be deduced from the typical simulation results demonstrated above.

As seen from Figs. 5(a)–5(c), under circumstances without sudden disturbance, the PBR target can always be reached after fine-tuning using the adaptive recursive algorithm (gray line), while the traditional algorithm (red line) never reaches the

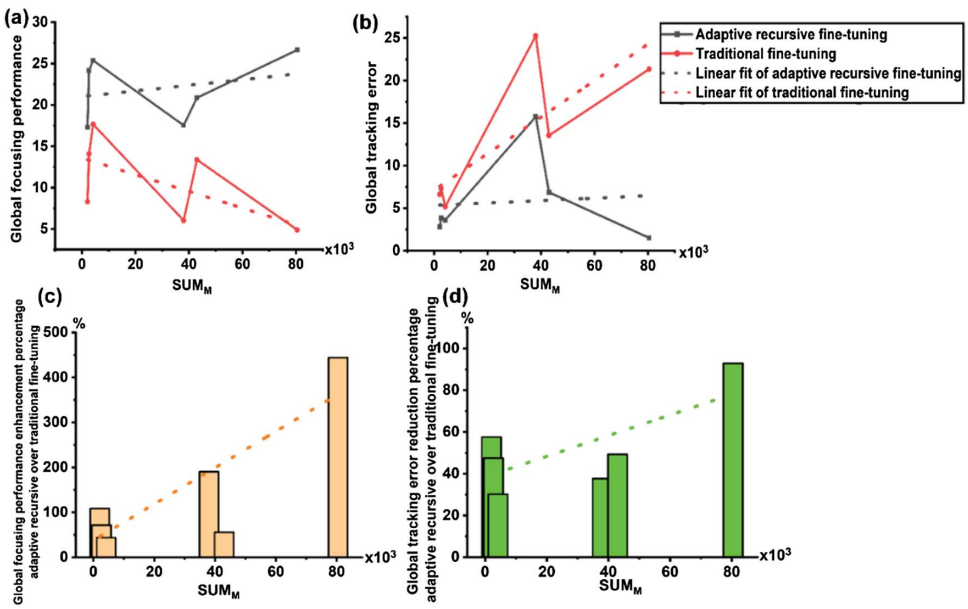


Fig. 4. Experimental results. (a) Global focusing performance in the six experiments with the adaptive recursive algorithm and traditional algorithm. (b) Global tracking error in the six experiments with adaptive recursive algorithm and traditional algorithm. Figures (a) and (b) use the same legend. (c) The enhancement percentage in global focusing performance achieved by the adaptive recursive algorithm over the traditional algorithm. (d) The reduction percentage in global tracking error achieved by the adaptive recursive algorithm over the traditional algorithm.

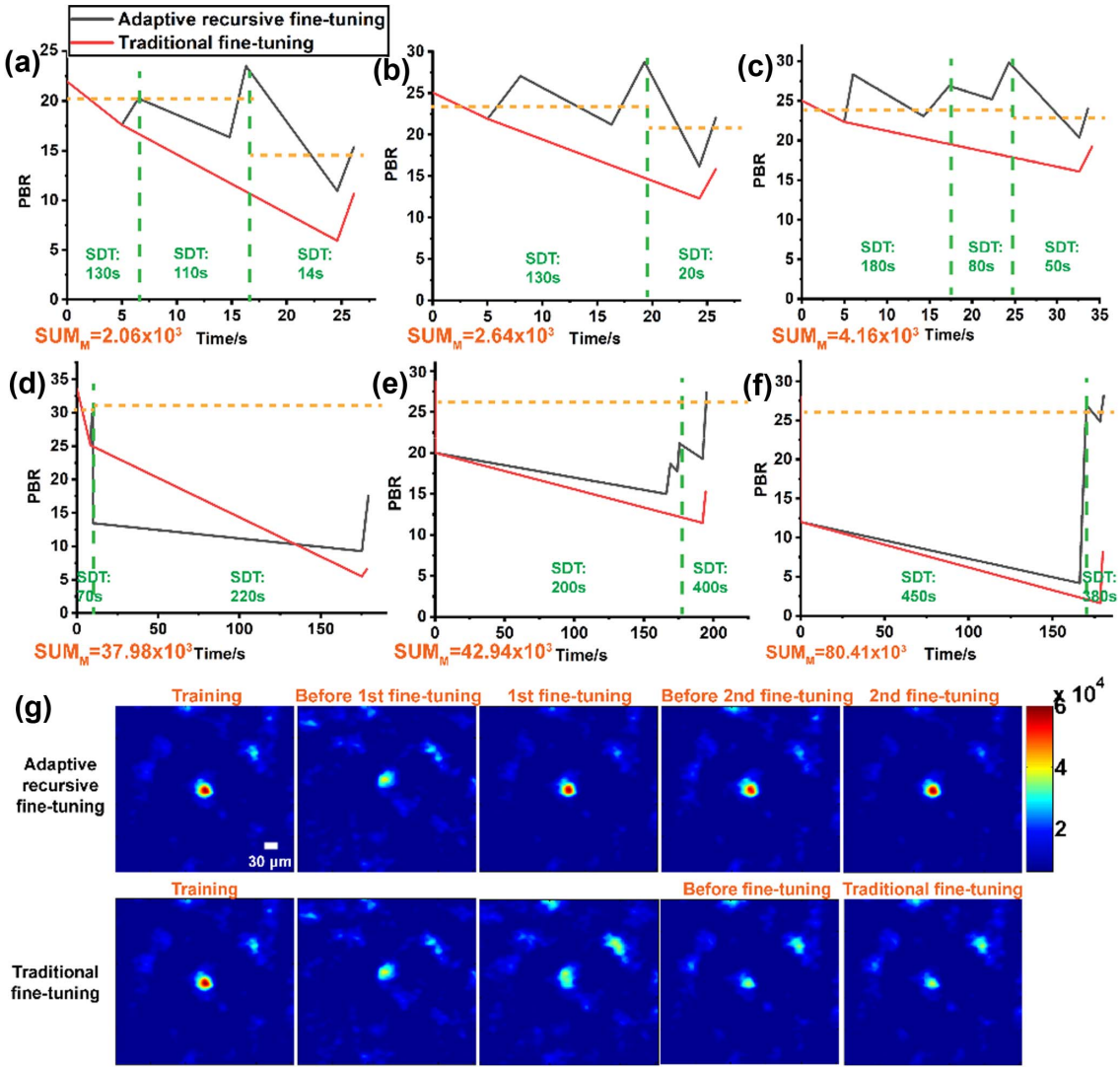


Fig. 5. (a)–(c) Experimental results of the three trials without environmental disturbance. The SDT of each stationary sub-process is shown in the figure, and the PBR target (ideal case) is indicated by yellow dashed lines. (d)–(f) Results of the three experiments with environmental disturbance. Figures (a)–(f) use the same legend. (g) Speckle images recorded during a nonstationary process with environmental perturbation using adaptive recursive and traditional algorithms. In (g), all speckle images use the same colormap and scale and are interpolated to 253 × 253 for a better view. The color bars indicate the detected light intensity in arbitrary units. The middle image in the bottom row (traditional fine-tuning) is an interpolated result using recorded speckles. Specific hyperparameters and fine-tuning sample amount used in experiments are available in Ref. [83].

target. In these circumstances, the adaptive recursive algorithm always shows much better GFP (17–25) than traditional algorithm results (8–17). As seen from Fig. 4(c), with adaptive recursive algorithms, the enhancement achieved in GFP over the traditional one is 43%–108%. Considering the influence from the environment, the enhancement percentage achieved in experiments not being as high as that in simulation is reasonable, but significant improvements in focusing performance are demonstrated by both simulation and experiment. The dotted lines in Fig. 4(c) indicate the trend that with the increase of SUM_M, i.e., longer nonstationary processes, the enhancement percentage realized by the adaptive recursive algorithm keeps rising, suggesting that the merits of the adaptive recursive algorithm are becoming notable as the nonstationary processes are becoming longer. With the traditional fine-tuning algorithm, when fine-tuning is conducted and when the original trained model is obtained ($t = 0$), the difference between the medium's status increases overall with SUM_M. The low statistical correlation between these two statuses induces degradation in the modeling accuracy, increasing the difficulty of focusing recovery. Moreover, sending all of the samples altogether to fine-tune the original model suggests that the whole process is regarded as a stationary variation by the network, which, actually, may be nonstationary and consists of multiple stationary stochastic sub-processes.

The reduction percentage in the GTE achieved by the adaptive recursive algorithm over traditional one is shown in Fig. 4(d), reaching 30%–57%, indicating that the adaptive recursive algorithm is much better than the traditional method in terms of focusing tracking, which is consistent with the

simulation results. Moreover, as SUM_M becomes larger, the focusing tracking performance of the traditional algorithm keeps being exacerbated due to the lack of ability in timely recovery. By contrast, the adaptive recursive algorithm conducts fine-tuning successively during the whole nonstationary process; thus, the reduction percentage achieved by the adaptive recursive algorithm increases as SUM_M becomes larger, as suggested by the dotted lines in Fig. 4(d).

In situations where environmental disturbance occurs, the experimental results are shown in Figs. 5(d)–5(f). The sudden perturbation can be regarded as a stationary stochastic sub-process, whose SDT is very small, and the time duration is extremely short. Although the occurrence of perturbations leads to inevitable increase in the GTE, as observed in both Figs. 4(b) and 2(l), the adaptive recursive algorithm still demonstrates much lower GTE in experiments (2–15) than that achieved by the traditional algorithm (13–25). The reduction percentage in the GTE realized by the adaptive recursive method over the traditional one is 38%–93%. As for the GFP, the adaptive recursive algorithm achieves larger values all the time, ranging from 17 to 27, while values obtained by the traditional method are much smaller, being from 4 to 13. The enhancement percentage of the adaptive recursive algorithm over the traditional performance is 56%–444%, achieving significant improvement in the focusing performance, which has also been indicated in simulation. These results suggest that the advantage of recursive fine-tuning becomes more outstanding in this situation, as timely tracking is more and more important in long nonstationary processes. Meanwhile, the merits of adaptive adjustments of hyperparameters also become more notable since the requirement of fewer fine-tuning samples expedites the focusing recovery and leads to better recovery performance as well. In addition, Fig. 5(e) demonstrates that once a slower changing sub-process is detected (SDT = 400 s), the adaptive recursive algorithm is able to make up the PBR loss caused by perturbations, which, again, is consistent with the simulation, as illustrated in Fig. 2(i).

Actually, all of the enhancement and reduction results obtained in the six experiments agree well with the simulations. As seen from Figs. 4(a) and 4(c), with the increase of SUM_M, the GFP of the adaptive recursive method keeps rising; meanwhile, the enhancement percentage it can achieve over the traditional performance also enlarges, regardless of the occurrence of disturbance. In all circumstances, the adaptive recursive algorithm always demonstrates the best results. As for the GTE, the reduction percentage over the traditional algorithm also increases with SUM_M, no matter whether perturbations take place or not. These results suggest that the proposed adaptive framework is robust and even more promisingly attractive when the nonstationary process lasts longer or significant sudden PBR degradation occurs.

The speckle images recorded in the nonstationary processes indicated by Fig. 5(f) with the adaptive recursive and traditional algorithms are shown in Fig. 5(g). All speckle images are interpolated to 253 × 253 for a better view, and the interpolation algorithm is based on splines. The diameter of the initial focused speckle is ∼30 μm. Speckle patterns before and after each fine-tuning are demonstrated. As seen in Fig. 5(g), the adaptive recursive algorithm can recover the focal point in time, and then the focus can remain over time. By contrast, with the traditional algorithm, due to the lack of ability of timely tracking, it cannot recover the focal point even though fine-tuning is conducted. It is worth noting that although we only report light focusing to a single position, the trained TFOTNet is capable of focusing light to an arbitrary position or multiple positions simultaneously on the image plane. As indicated above, during experiments, only the speckles before and after fine-tuning are recorded with the traditional algorithm, and the middle image in the bottom row in Fig. 5(g) is an interpolated result using recorded speckles.

3. Comparison of Light Focusing and Refocusing Performance with TFOTNet, Conventional ConvLSTM, and CNN

The ability of the conventional single-input-single-output ConvLSTM network [85,86] and single-input-single-output CNN in light focusing and refocusing through nonstationary scattering media is investigated using both simulation and experiments, and results are shown in Fig. 6. The structures of the conventional ConvLSTM and CNN are shown in Figs. 6(d) and 6(e), respectively. The conventional ConvLSTM network consists of two ConvLSTM layers, one LSTM layer, and one TimeDistributed dense layer working as the output layer. The input of the network is speckle patterns, while the output is their corresponding SLM patterns. All layers share the same parameters with their corresponding ones in TFOTNet, including kernel size, number of filters, activation function, etc. As for CNN, it consists of two convolutional layers, one fully connected layer, and the other fully connected layer serving as the output layer. Except timestep, which is not included in CNN, all of the other parameters are the same as the ConvLSTM network. With simulation, in the first step, the same 10,000 samples are used to train TFOTNet and CNN in order to obtain a focal point, and training results are shown in Fig. 6(a). All figures in Fig. 6(a) use the same colormap. The PBRs of the focused speckle achieved by TFOTNet and CNN are 41.5 and 15.6, respectively. As for the ConvLSTM network, 15,000 samples are used, which is an increase of 50% compared with that needed by TFOTNet. Nonetheless, the PBR of the focused speckle obtained with ConvLSTM is only 10.79, much lower than that achieved with the pre-trained TFOTNet (41.5). This phenomenon actually indicates a drawback of conventional ConvLSTM networks, where, as both temporal and spatial weights have to be learned during training, a large amount of samples are required. However, TFOTNet significantly enhances the modeling efficiency and effectively overcomes this drawback. After obtaining a focal point, during the fine-tuning phase with a nonstationary process, the same adaptive hyperparameters and fine-tuning samples are offered to TFOTNet, ConvLSTM, and CNN (except timestep). As seen in Figs. 6(a) and 6(f), with the same nonstationary process and fine-tuning algorithm, TFOTNet always exhibits the best performance in light focusing and refocusing. Nevertheless, with the ConvLSTM network or CNN, over time, the background becomes so bright that a single focal point is no longer able to be recovered, even though recursive fine-tuning is conducted. As for experimental results, 10,000 samples are sent to

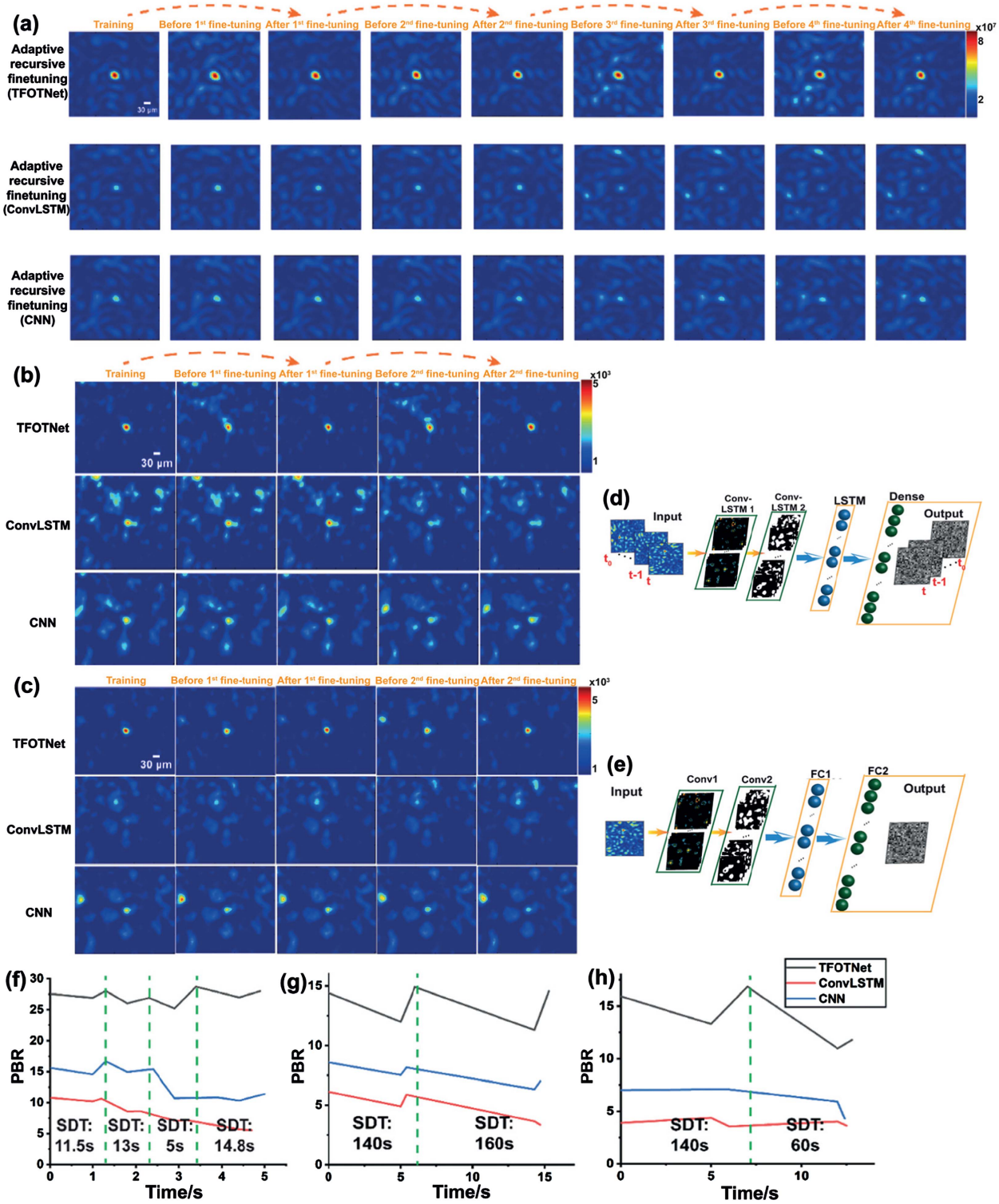


Fig. 6. Comparisons about fine-tuning ability in a nonstationary process using three different networks (see Visualization 1). (a) Simulation results. Light focusing and refocusing performance recorded at different times through a nonstationary process using the adaptive recursive algorithm with TFOTNet (the first row), ConvLSTM (the second row), and CNN (the third row) is shown. All images use the same colormap and scale, and the color bars indicate the light intensity in arbitrary units. (b) Experimental results. Light focusing and refocusing results using three networks with the adaptive recursive algorithm in the same nonstationary process are shown. All speckle images use the same colormap and scale and are interpolated to 253 × 253. The color bars indicate the detected light intensity in arbitrary units. (c) Experimental results in the situation of low SNR. All speckle images use the same colormap and scale and are interpolated to 253 × 253. The color bars indicate the detected light intensity in arbitrary units. (d) Structure of a conventional single-input-single-output ConvLSTM. (e) Structure of a conventional single-input-single-output CNN. (f) Details of the nonstationary process and PBR with the three networks in simulation (a). (g) Details of the nonstationary process and PBR with the three networks in experiment (b). (h) Details of the nonstationary process and PBR with the three networks in experiment (c). (f)–(h) use the same legend.

initialize and train TFOTNet, ConvLSTM, and CNN. With TFOTNet, after training, a focused speckle can be obtained while using the other two networks, clear background speckles are observed with PBR dropping to less than 60% of that achieved by TFOTNet, as shown in Figs. 6(b) and 6(g). With adaptive recursive fine-tuning, a focused speckle can always be retained using TFOTNet through a nonstationary scattering medium; by contrast, the focal point is submerged over time when ConvLSTM or CNN is used, which agrees well with simulation results. Interestingly, experimental results demonstrate that in situations of low SNR, as shown in Figs. 6(c) and 6(h), among the three networks, only the proposed TFOTNet is able to obtain focus after training, even though the same training samples and parameters are used. In the situations indicated by Figs. 6(a)–6(c), the SNR of the training results of TFOTNet is calculated as 14, 12, and 10, respectively, which is defined as the ratio of the mean value of the signal and the standard deviation of the noises [87,88]. As seen in Fig. 6(c), even if some fine-tuning samples are offered, the fine-tuned ConvLSTM or CNN still cannot focus light through a nonstationary scattering medium. As indicated by Figs. 6(a) and 6(b), the refocusing performance of conventional ConvLSTM or CNN degrades over time; thus, it can be deduced that under the circumstances of low SNR a focal point can hardly be obtained with these two networks. This phenomenon manifests that TFOTNet is more robust to noises than conventional single-input-single-output networks.

As mentioned above, the SLM used in the experiment limits the fine-tuning sample collection speed, which further restricts the allowed changing speed of the scattering medium. Nevertheless, it should be emphasized here that there is no fundamental limitation on the speed of the proposed adaptive deep learning framework, since much faster modulators can be employed. With a current commercially available DMD whose frame rate has reached 23 kHz [82], both sample collection speed and the SDT of the altering medium can be improved nearly 4000 times. Thus, for the experimental results shown in Fig. 5, the SDT can be shortened to 3.5–113 ms, indicating that the proposed framework can be potentially applied for wavefront shaping in dynamic media, for instance, optical focusing and imaging at depths *in vivo* that alters at the speed of milliseconds [84]. Therefore, the proposed framework opens up a potential pathway to resolve the high demand of wavefront shaping on responding time, taking a significant step towards practical realizations. The proposed adaptive recursive fine-tuning approach applying millisecond variation media *in vivo* will be further studied and will be reported elsewhere.

4. Comparison of Time Cost in Focusing Recovery with Various Algorithms

It should be noted that time cost in focusing recovery by the adaptive recursive fine-tuning algorithm and the two representative conventional wavefront shaping techniques, continuous sequential algorithm (CSA) and transmission matrix measurement, is discussed herein for comparison. Assuming a nonstationary process whose duration is t, on average, the medium status changes per Δt. For the adaptive recursive algorithm, on average, M samples are needed for each fine-tuning, and totally M_{total} samples are used during the whole nonstationary course. In comparison, if CSA or transmission matrix measurement is adopted to recover the focusing performance through the changed medium, the iterative optimization process or transmission matrix measurement has to be repeated from the beginning. Thus, the time cost should be $(KN^2 \times \frac{M_{\text{total}}}{M})/F$ and $(4N^2 \times \frac{M_{\text{total}}}{M})/F$, respectively, where N is the dimension of the SLM pattern, K is the pixel gray level, and F is the frame rate of the SLM. $\frac{M_{\text{total}}}{M}$ represents how many times fine-tuning has been done during the whole nonstationary process, and, each time fine-tuning is conducted, CSA or transmission matrix measurement will also be run once for focusing recovery. As for the adaptive recursive algorithm, the total time spent in one fine-tuning is $\frac{M}{F} + pt_p$. The fine-tuning time cost consists of two parts: sample collection time and computational time. The sample collection time is independent of the SLM dimension N; instead, it is determined by the sample amount M and the SLM frame rate F, written as $\frac{M}{F}$. The computational time is the product of epoch number p and time cost per epoch t_p. Considering that during the fine-tuning, only the last layer of the pre-trained network, which has N^2 neurons, is adjusted, t_p is the function of N^2, that is $t_p = g(N^2) = \lambda(N)N^2$. Besides N, $\lambda(N)$ is also influenced by various factors such as network structures, computational engine, and amount of fine-tuning samples. It should be noted that taking a powerful GPU will reduce t_p, thus reducing $\lambda(N)$. As an example, in our work, with the TFOTNet and reported computation platform (Acer Predator G9-792, 16 Gb RAM, and a GTX 980M GPU), in simulation, t_p is 0.38, 0.4, 0.45, 0.49, 0.54, and 0.61 s when N is set as 8, 16, 32, 64, 128, and 256, respectively. The fine-tuning sample amount used here is 1000, which is the largest sample amount used in the reported experiments, and the timestep and batch size are set as 2 and 64, respectively. It can be expected that t_p can be further reduced if the fine-tuning sample amount is smaller, or a more powerful computational unit is adopted. $\lambda(N) = t_p/N^2$ is calculated to vary from 9.2×10^{-6} to 5.9×10^{-3}. In our work, $N = 32$, and $\lambda(N)$ is calculated to be 4.4×10^{-4}. For an intuitive comparison, herein, time cost in one focusing recovery process using the adaptive recursive fine-tuning algorithm, CSA, and transmission matrix measurement is given below based on the setup reported in this article, with the SLM pattern size being 32×32 and frame rate of LCoS-SLM as 60 Hz. K varies with different setups, which can be set as 8 [22], 191 [89], or other values, and here we adopt $K = 32$ to be consistent with our experimental settings. Hence, more than 9 min is needed by CSA to complete an iterative optimization process. To measure a new transmission matrix that represents the changed medium status, nearly 70 s is required. As indicated by experimental results in Fig. 5, with the adaptive recursive algorithm, time spent in each fine-tuning varies from 6 to 170 s with the frame rate of the SLM being only ~6 Hz. Since 60 Hz is used for optimization time estimation, for a fair comparison, the fine-tuning time should be reduced by 10 times, varying from 0.6 to 17 s. Therefore, the proposed adaptive fine-tuning algorithm can improve the speed by 32–910 times and 4–113 times against CSA and transmission matrix measurement, respectively. In addition, to measure

a transmission matrix, interference between the modulated light and a reference light is required, which significantly increases the system complexity and reduces the utilization efficiency of the SLM, considering that part of the SLM pixels work as the reference. As CSA optimizes each pixel independently, the detected intensity improvement at the output plane is small, which may lead to errors in phase selection, especially when the SNR is low [90].

One promising application of the proposed framework is encryption. Recently, learning-based optical encryption has been reported [91] with parameters of the trained model as the security keys, achieving high security. In their study, static diffusers are used. With our framework, rotating diffusers can be applied to create much more complex scattering conditions, which enhances system security as more parameters are required to precisely model the process. More importantly, the introduction of a fine-tuning engine makes the system robust to attack. If it is sensed that current security key is partially eavesdropped, the diffuser can be rotated to create a new scattering situation to disable the leaked key. Meanwhile, with our adaptive fine-tuning system, security keys that fit for the new setting can be obtained rapidly, preventing loss due to attack. Demonstration of the idea is underway and will be reported elsewhere.

Lastly, but not the least, it should be admitted that since the fine-tuning cost is determined by the correlation between the medium statuses, it is natural that more samples and longer times are needed to recover a focal point in situations with a short SDT or dramatic disturbance. The proposed adaptive deep learning framework makes the best use of correlation between medium statuses to reduce the cost in focusing recovery as much as possible. However, in extreme cases where the medium decorrelates rapidly, or the persistence is low between iterations, a new training or optimization cycle may be required. The proposed adaptive recursive framework can achieve optimal tracking performance to a physical process change unless unpredictable.

4. INFLUENCE OF HYPERPARAMETERS ON FINE-TUNING AND THE IMPLEMENTATION OF ADAPTIVE PBR TARGET

Hyperparameters including timestep, batch size, initial learning rate, and numbers of fine-tuning samples are investigated, respectively, as they will influence light refocusing performance or fine-tuning time cost. Simulation is conducted to evaluate the fine-tuning time cost to reach a pre-defined PBR target, as the above listed relevant parameters are individually varied with the medium changing at different speeds; then these results can be scaled in light of specific implementations. In all cases, fine-tuning is conducted with the same time interval after the original focal point is obtained. This is to ensure the same degree of medium change when different hyperparameters are investigated in a steadily altering situation, since the fine-tuning cost is directly related with the correlation amongst medium statuses. In this simulation, serving as an example, the PBR target is set as 37, and the time interval is set as 1 s. Nevertheless, values are not fixed; instead, they are adjustable according to specific setups. When one hyperparameter is under test, all other hyperparameters remain as their default values, which are given in the inserted table in Fig. 2. From Figs. 7(a)–7(d), it can be seen that when medium altering is mild, and the SDT is larger than 10.8 s, varying one hyperparameter does not lead to significant differences in the fine-tuning time. With faster speed (SDT is smaller than 10.8 s), selecting suitable values for hyperparameters becomes more essential, as they are exerting growing influence on the fine-tuning time cost. Sample collection time is estimated using the maximal frame rate of commercial LCoS-SLM, which is generally 60 Hz. As for epochs, there is no doubt that more epochs require longer computation time and may lead to overfitting. On the other hand, more epochs may contribute to better light focusing performance.

Among all of the hyperparameters evaluated above, the amount of fine-tuning samples imposes the most significant influence on the fine-tuning time cost. A comparison of the required fine-tuning sample amount with and without adaptive adjustments of hyperparameters in the situations of different SDTs is shown in Fig. 7(e). As seen, without adaptive modifications of hyperparameters, the required sample amount is 2–3 times that needed by the adaptive algorithm to reach a pre-defined PBR target, resulting in much longer fine-tuning time cost. The result can be extended to other SDTs that are not tested here. During fine-tuning, if the number of newly collected samples is smaller than 100, then previously collected samples are included to concatenate with new samples; thus, a total of 1000 samples are used for fine-tuning to avoid overfitting. Increasing the fine-tuning sample amount theoretically leads to better focusing performance, which, however, also prolongs the fine-tuning process, as more time is spent at sample collection, indicating larger change to the medium status during this time period, which will degrade the fine-tuned PBR. With the goal of balancing the trade-off between the overall PBR and the fine-tuning time cost, we explore the relationship between PBR after fine-tuning and fine-tuning samples amounts using the adaptive algorithm when the medium changes at different speeds, with results shown in Figs. 7(f) and 7(g). In all cases, fine-tuning is conducted after a fixed time interval when the initial focused speckle is obtained, and it is chosen as 1 s as an example, which is scalable. With slow medium change (SDT larger than 2.8 s), less than 30 fine-tuning samples are sufficient to recover the PBR back to 37, which is regarded as an acceptable PBR threshold in this simulation. With faster medium changes, characterized by SDT ranging from 1.6 to 1.2 s, several hundred samples are needed to surpass the PBR threshold. With further expedition of the medium change (SDT lower than 1.2 s), up to several thousand samples are required. As longer time is required to collect more samples, the capability of tracking the medium change and keeping light focused will be affected. To mitigate this dilemma, an adaptive PBR target is employed; as a faster SDT accommodates a relatively lower PBR target, less time is needed. With instructive results shown in Fig. 7, for a certain SDT interval, the adaptive PBR target is defined as the mean value of the maximal and minimal PBR that can be achieved after fine-tuning, i.e., $\text{PBR target} = \frac{\max(\text{PBR})+\min(\text{PBR})}{2}$, serving as a criterion to evaluate the focusing recovery performance.

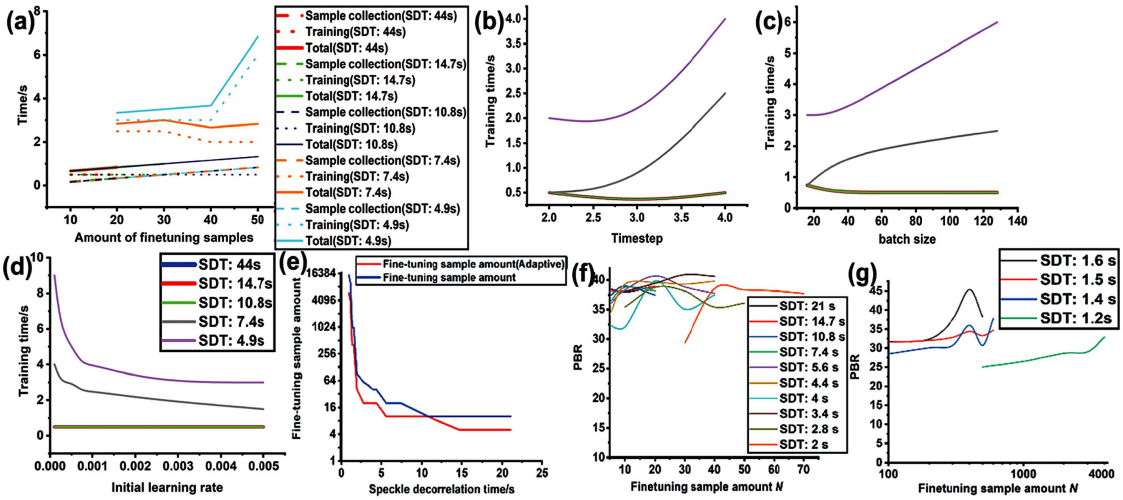


Fig. 7. Influence of hyperparameters on fine-tuning time cost and performance when the scattering medium changes at various speeds. (a)–(d) The effect of sample amount, timestep, batch size, and initial learning rate on fine-tuning time cost under five circumstances where the scattering medium is changing at different speeds (indicated by lines of different colors and quantified by speckle decorrelation time). (e) The required amount of fine-tuning samples with and without adaptive adjustments of hyperparameters as the medium changes at different speeds. (f), (g) The relationship between the PBR after fine-tuning and the fine-tuning sample amount using the adaptive algorithm as the medium changes at different speeds. The default values of these hyperparameters used in the simulation are listed in Fig. 2.

Although different setups result in different initial focused speckles, the ratio between the PBR target and PBR of the initial focused pattern remains unchanged as long as the SDT is the same. By doing so, the presented results can be safely scaled to any other implementations.

5. CONCLUSION

In summary, the proposed deep learning-empowered adaptive wavefront framework is able to achieve focusing and fast refocusing of light through time-variant scattering media, which now allows a complex nonstationary stochastic process for the kind of first time, to the best of our knowledge. With the proposed adaptive recursive fine-tuning of TFOTNet, optical focusing can be recovered in time from degradation, which is much more rapid (with fundamental potential to achieve real-time) than both the traditional fine-tuning algorithm and the representative conventional methods, which require a new time- and/or resource-demanding optimization process. Simulation and experimental results agree very well and manifest the merits of the proposed framework. The experimental results indicate that with the proposed adaptive recursive framework, for all SUM_M investigated here, the GFP can be enhanced by 43%–444% against the traditional algorithm, and the GTE is reduced by 30%–93%. Moreover, as the nonstationary process is prolonged, both the GFP enhancement percentage and the GTE reduction percentage over the traditional algorithm increase. It can be expected that similar performance improvement can be achieved by other implementations of the proposed framework. All results shown in the article are scalable according to the specific implementations; with the proposed framework, light focusing can be safely retained in all realizations. As stated, with DMD and more powerful GPU, the proposed framework has potential to deal with scattering media of SDT being several milliseconds; thus, it opens up a potential pathway to resolve the high demand of wavefront shaping on responding time, taking a significant step towards practical realizations.

Funding. Agency for Science, Technology and Research (A18A7b0058); National Natural Science Foundation of China (81627805, 81671726, 81930048); Guangdong Science and Technology Commission (2019A1515011374, 2019BT02X105); Hong Kong Innovation and Technology Commission (GHP/043/19SZ, GHP/044/19GD, ITS/022/18); Hong Kong Research Grant Council (25204416, R5029-19); Shenzhen Science and Technology Innovation Commission (JCYJ20170818104421564).

Disclosures. The authors declare no conflicts of interest.

†These authors contributed equally to this paper.

REFERENCES

1. J. Bertolotti, E. G. Van Putten, C. Blum, A. Lagendijk, W. L. Vos, and A. P. Mosk, "Non-invasive imaging through opaque scattering layers," Nature **491**, 232–234 (2012).
2. P. Lai, L. Wang, J. W. Tay, and L. V. Wang, "Photoacoustically guided wavefront shaping for enhanced optical focusing in scattering media," Nat. Photonics **9**, 126–132 (2015).
3. Z. Yu, M. Xia, H. Li, T. Zhong, F. Zhao, H. Deng, Z. Li, D. Li, D. Wang, and P. Lai, "Implementation of digital optical phase conjugation with embedded calibration and phase rectification," Sci. Rep. **9**, 1537 (2019).
4. J. Yang, Y. Shen, Y. Liu, A. S. Hemphill, and L. V. Wang, "Focusing light through scattering media by polarization modulation based generalized digital optical phase conjugation," Appl. Phys. Lett. **111**, 201108 (2017).

5. 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**, 250–256 (2019).
6. Y. Shen, Y. Liu, C. Ma, and L. V. Wang, "Sub-Nyquist sampling boosts targeted light transport through opaque scattering media," Optica **4**, 97–102 (2017).
7. I. M. Vellekoop and A. P. Mosk, "Focusing coherent light through opaque strongly scattering media," Opt. Lett. **32**, 2309–2311 (2007).
8. S. M. Popoff, G. Lerosey, R. Carminati, M. Fink, A. C. Boccara, and S. Gigan, "Measuring the transmission matrix in optics: an approach to the study and control of light propagation in disordered media," Phys. Rev. Lett. **104**, 100601 (2010).
9. O. Katz, E. Small, and Y. Silberberg, "Looking around corners and through thin turbid layers in real time with scattered incoherent light," Nat. Photonics **6**, 549–553 (2012).
10. H. Yu, T. R. Hillman, W. Choi, J. O. Lee, M. S. Feld, R. R. Dasari, and Y. Park, "Measuring large optical transmission matrices of disordered media," Phys. Rev. Lett. **111**, 153902 (2013).
11. T. Chaigne, O. Katz, A. C. Boccara, M. Fink, E. Bossy, and S. Gigan, "Controlling light in scattering media non-invasively using the photoacoustic transmission matrix," Nat. Photonics **8**, 58–64 (2014).
12. A. Sanjeev, Y. Kapellner, N. Shabairou, E. Gur, M. Sinvani, and Z. Zalevsky, "Non-invasive imaging through scattering medium by using a reverse response wavefront shaping technique," Sci. Rep. **9**, 12275 (2019).
13. A. Drémeau, A. Liutkus, D. Martina, O. Katz, C. Schülke, F. Krzakala, S. Gigan, and L. Daudet, "Reference-less measurement of the transmission matrix of a highly scattering material using a DMD and phase retrieval techniques," Opt. Express **23**, 11898–11911 (2015).
14. K. T. Takasaki and J. W. Fleischer, "Phase-space measurement for depth-resolved memory-effect imaging," Opt. Express **22**, 31426–31433 (2014).
15. O. Katz, E. Small, Y. Guan, and Y. Silberberg, "Noninvasive nonlinear focusing and imaging through strongly scattering turbid layers," Optica **1**, 170–174 (2014).
16. E. Edrei and G. Scarcelli, "Memory-effect based deconvolution microscopy for super-resolution imaging through scattering media," Sci. Rep. **6**, 33558 (2016).
17. X. Xu, H. Liu, and L. V. Wang, "Time-reversed ultrasonically encoded optical focusing into scattering media," Nat. Photonics **5**, 154–157 (2011).
18. B. Judkewitz, Y. M. Wang, R. Horstmeyer, A. Mathy, and C. Yang, "Speckle-scale focusing in the diffusive regime with time-reversal of variance-encoded light (TROVE)," Nat. Photonics **7**, 300–305 (2013).
19. S. Resisi, Y. Viernik, S. M. Popoff, and Y. Bromberg, "Wavefront shaping in multimode fibers by transmission matrix engineering," APL Photon. **5**, 036103 (2020).
20. X. Wei, Y. Shen, J. C. Jing, A. S. Hemphill, C. Yang, S. Xu, Z. Yang, and L. V. Wang, "Real-time frequency-encoded spatiotemporal focusing through scattering media using a programmable 2D ultrafine optical frequency comb," Sci. Adv. **6**, eaay1192 (2020).
21. G. Huang, D. Wu, J. Luo, Y. Huang, and Y. Shen, "Retrieving the optical transmission matrix of a multimode fiber using the extended Kalman filter," Opt. Express **28**, 9487–9500 (2020).
22. J. Thompson, B. Hokr, and V. Yakovlev, "Optimization of focusing through scattering media using the continuous sequential algorithm," J. Mod. Opt. **63**, 80–84 (2016).
23. D. B. Conkey, A. N. Brown, A. M. Caravaca-Aguirre, and R. Piestun, "Genetic algorithm optimization for focusing through turbid media in noisy environments," Opt. Express **20**, 4840–4849 (2012).
24. J. Luo, Z. Wu, D. Wu, Z. Liu, X. Wei, Y. Shen, and Z. Li, "Efficient glare suppression with Hadamard-encoding-algorithm-based wavefront shaping," Opt. Lett. **44**, 4067–4070 (2019).
25. Z. Wu, J. Luo, Y. Feng, X. Guo, Y. Shen, and Z. Li, "Controlling 1550-nm light through a multimode fiber using a Hadamard encoding algorithm," Opt. Express **27**, 5570–5580 (2019).
26. O. Katz, P. Heidmann, M. Fink, and S. Gigan, "Non-invasive single-shot imaging through scattering layers and around corners via speckle correlations," Nat. Photonics **8**, 784–790 (2014).
27. Y. Liu, C. Ma, Y. Shen, J. Shi, and L. V. Wang, "Focusing light inside dynamic scattering media with millisecond digital optical phase conjugation," Optica **4**, 280–288 (2017).
28. D. Wang, E. H. Zhou, J. Brake, H. Ruan, M. Jang, and C. Yang, "Focusing through dynamic tissue with millisecond digital optical phase conjugation," Optica **2**, 728–735 (2015).
29. M. Chen, H. Liu, Z. Liu, P. Lai, and S. Han, "Expansion of the FOV in speckle autocorrelation imaging by spatial filtering," Opt. Lett. **44**, 5997–6000 (2019).
30. J.-H. Park, Z. Yu, K. Lee, P. Lai, and Y. Park, "Perspective: wavefront shaping techniques for controlling multiple light scattering in biological tissues: toward *in vivo* applications," APL Photon. **3**, 100901 (2018).
31. Y. Luo, S. Yan, H. Li, P. Lai, and Y. Zheng, "Focusing light through scattering media by reinforced hybrid algorithms," APL Photon. **5**, 016109 (2020).
32. E. Bossy and S. Gigan, "Photoacoustics with coherent light," Photoacoustics **4**, 22–35 (2016).
33. Z. Li, Z. Yu, H. Hui, H. Li, T. Zhong, H. Liu, and P. Lai, "Edge enhancement through scattering media enabled by optical wavefront shaping," Photon. Res. **8**, 954–962 (2020).
34. Y. Shen, Y. Liu, C. Ma, and L. V. Wang, "Focusing light through scattering media by full-polarization digital optical phase conjugation," Opt. Lett. **41**, 1130–1133 (2016).
35. Y.-K. Xu, W.-T. Liu, E.-F. Zhang, Q. Li, H.-Y. Dai, and P.-X. Chen, "Is ghost imaging intrinsically more powerful against scattering?" Opt. Express **23**, 32993–33000 (2015).
36. E. Edrei and G. Scarcelli, "Optical imaging through dynamic turbid media using the Fourier-domain shower-curtain effect," Optica **3**, 71–74 (2016).
37. B. Hwang, T. Woo, and J.-H. Park, "Fast diffraction-limited image recovery through turbulence via subsampled bispectrum analysis," Opt. Lett. **44**, 5985–5988 (2019).
38. B. Blochet, L. Bourdieu, and S. Gigan, "Focusing light through dynamical samples using fast continuous wavefront optimization," Opt. Lett. **42**, 4994–4997 (2017).
39. R. H. Benjamin Judkewitz, I. M. Vellekoop, I. N. Papadopoulos, and Y. Changhuei, "Translation correlations in anisotropically scattering media," Nat. Phys. **11**, 684–689 (2015).
40. J. Xie, L. Xu, and E. Chen, "Image denoising and inpainting with deep neural networks," in *Advances in Neural Information Processing Systems* (2012), pp. 341–349.
41. G. Barbastathis, A. Ozcan, and G. Situ, "On the use of deep learning for computational imaging," Optica **6**, 921–943 (2019).
42. A. Sinha, J. Lee, S. Li, and G. Barbastathis, "Lensless computational imaging through deep learning," Optica **4**, 1117–1125 (2017).
43. L. Waller and L. Tian, "Computational imaging: machine learning for 3D microscopy," Nature **523**, 416–417 (2015).
44. A. Goy, K. Arthur, S. Li, and G. Barbastathis, "Low photon count phase retrieval using deep learning," Phys. Rev. Lett. **121**, 243902 (2018).
45. Y. Rivenson, Y. Wu, and A. Ozcan, "Deep learning in holography and coherent imaging," Light Sci. Appl. **8**, 1 (2019).
46. Y. Wu, Y. Rivenson, H. Wang, Y. Luo, E. Ben-David, L. A. Bentolila, C. Pritz, and A. Ozcan, "Three-dimensional virtual refocusing of fluorescence microscopy images using deep learning," Nat. Methods **16**, 1323–1331 (2019).
47. M. T. McCann, K. H. Jin, and M. Unser, "Convolutional neural networks for inverse problems in imaging: a review," IEEE Signal Process. Mag. **34**, 85–95 (2017).
48. C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," IEEE Trans. Pattern Anal. Mach. Intell. **38**, 295–307 (2015).
49. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature **521**, 436–444 (2015).
50. A. Turpin, I. Vishniakou, and J. D. Seelig, "Light scattering control in transmission and reflection with neural networks," Opt. Express **26**, 30911–30929 (2018).
51. Y. Zhang, C. Wu, Y. Song, K. Si, Y. Zheng, L. Hu, J. Chen, L. Tang, and W. Gong, "Machine learning based adaptive optics for doughnut-shaped beam," Opt. Express **27**, 16871–16881 (2019).

52. S. Cheng, H. Li, Y. Luo, Y. Zheng, and P. Lai, "Artificial intelligence-assisted light control and computational imaging through scattering media," J. Innov. Opt. Health Sci. **12**, 1930006 (2019).
53. Y. Li, Y. Xue, and L. Tian, "Deep speckle correlation: a deep learning approach toward scalable imaging through scattering media," Optica **5**, 1181–1190 (2018).
54. S. Li, M. Deng, J. Lee, A. Sinha, and G. Barbastathis, "Imaging through glass diffusers using densely connected convolutional networks," Optica **5**, 803–813 (2018).
55. B. Rahmani, D. Loterie, G. Konstantinou, D. Psaltis, and C. Moser, "Multimode optical fiber transmission with a deep learning network," Light Sci. Appl. **7**, 69 (2018).
56. Y. Sun, J. Shi, L. Sun, J. Fan, and G. Zeng, "Image reconstruction through dynamic scattering media based on deep learning," Opt. Express **27**, 16032–16046 (2019).
57. Y. Wang, J. Zhang, H. Zhu, M. Long, J. Wang, and P. S. Yu, "Memory in memory: a predictive neural network for learning higher-order non-stationarity from spatiotemporal dynamics," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2019), pp. 9154–9162.
58. Q. Luo, J. A. Newman, and K. J. Webb, "Motion-based coherent optical imaging in heavily scattering random media," Opt. Lett. **44**, 2716–2719 (2019).
59. H. Yilmaz, E. G. van Putten, J. Bertolotti, A. Lagendijk, W. L. Vos, and A. P. Mosk, "Speckle correlation resolution enhancement of wide-field fluorescence imaging," Optica **2**, 424–429 (2015).
60. A. Porat, E. R. Andresen, H. Rigneault, D. Oron, S. Gigan, and O. Katz, "Widefield lensless imaging through a fiber bundle via speckle correlations," Opt. Express **24**, 16835–16855 (2016).
61. I. M. Vellekoop, "Controlling the propagation of light in disordered scattering media," arXiv:0807.1087 (2008).
62. Z. Wei and X. Chen, "Deep-learning schemes for full-wave nonlinear inverse scattering problems," IEEE Trans. Geosci. Remote Sens. **57**, 1849–1860 (2018).
63. W. C. Chew and Y.-M. Wang, "Reconstruction of two-dimensional permittivity distribution using the distorted Born iterative method," IEEE Trans. Med. Imaging **9**, 218–225 (1990).
64. X. Chen, "Subspace-based optimization method for solving inverse-scattering problems," IEEE Trans. Geosci. Remote Sens. **48**, 42–49 (2009).
65. U. S. Kamilov and H. Mansour, "Learning optimal nonlinearities for iterative thresholding algorithms," IEEE Signal Process. Lett. **23**, 747–751 (2016).
66. K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," IEEE Trans. Image Process. **26**, 4509–4522 (2017).
67. E. Rueckert, M. Nakatenus, S. Tosatto, and J. Peters, "Learning inverse dynamics models in o(n) time with LSTM networks," in *IEEE-RAS 17th International Conference on Humanoid Robotics (Humanoids)* (2017), pp. 811–816.
68. J. S. Marron and M. P. Wand, "Exact mean integrated squared error," Ann. Stat. **20**, 712–736 (1992).
69. 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**, 834–837 (1988).
70. M. Breitkreiz and P. W. Brouwer, "Semiclassical theory of speckle correlations," Phys. Rev. E **88**, 062905 (2013).
71. P. Sebbah, *Waves and Imaging through Complex Media* (Springer, 2001).
72. L. Belfore, A. Arkadan, and B. Lenhardt, "ANN inverse mapping technique applied to electromagnetic design," IEEE Trans. Magn. **37**, 3584–3587 (2001).
73. L. Li, J. Pan, W.-S. Lai, C. Gao, N. Sang, and M.-H. Yang, "Learning a discriminative prior for blind image deblurring," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2018), pp. 6616–6625.
74. J. Adler and O. Öktem, "Solving ill-posed inverse problems using iterative deep neural networks," Inverse Prob. **33**, 124007 (2017).
75. A. Lucas, M. Iliadis, R. Molina, and A. K. Katsaggelos, "Using deep neural networks for inverse problems in imaging: beyond analytical methods," IEEE Signal Process. Mag. **35**, 20–36 (2018).
76. J. Rick Chang, C.-L. Li, B. Poczos, B. Vijaya Kumar, and A. C. Sankaranarayanan, "One network to solve them all–solving linear inverse problems using deep projection models," in *Proceedings of the IEEE International Conference on Computer Vision* (2017), pp. 5888–5897.
77. J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Advances in Neural Information Processing Systems* (2014), pp. 3320–3328.
78. H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," IEEE Trans. Med. Imaging **35**, 1285–1298 (2016).
79. L. Castrejon, Y. Aytar, C. Vondrick, H. Pirsiavash, and A. Torralba, "Learning aligned cross-modal representations from weakly aligned data," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016), pp. 2940–2949.
80. V. Tran, S. K. Sahoo, and C. Dang, "Fast 3D movement of a laser focusing spot behind scattering media by utilizing optical memory effect and optical conjugate planes," Sci. Rep. **9**, 1 (2019).
81. M. M. Qureshi, J. Brake, H.-J. Jeon, H. Ruan, Y. Liu, A. M. Safi, T. J. Eom, C. Yang, and E. Chung, "*In vivo* study of optical speckle decorrelation time across depths in the mouse brain," Biomed. Opt. Express **8**, 4855–4864 (2017).
82. Z. Yu, H. Li, and P. Lai, "Wavefront shaping and its application to enhance photoacoustic imaging," Appl. Sci. **7**, 1320 (2017).
83. Y. Luo, S. Yan, H. Li, P. Lai, and Y. Zheng, "Datafile_towards smart optical focusing," https://drive.google.com/drive/folders/1_jbotvdfKimgvNYVXqRMx5JykYpWXUZ?usp=sharing (2020).
84. I. Nissilä, T. Noponen, J. Heino, T. Kajava, and T. Katila, "Diffuse optical imaging," in *Advances in Electromagnetic Fields in Living Systems* (Springer, 2005), Vol. **4**.
85. A. Xavier, An Introduction to ConvLSTM, https://medium.com/neuronio/an-introduction-to-convlstm-55c9025563a7 (2019).
86. S. Xingjian, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional LSTM network: a machine learning approach for precipitation nowcasting," in *Advances in Neural Information Processing Systems* (2015), pp. 802–810.
87. L. Kaufman, D. M. Kramer, L. E. Crooks, and D. A. Ortendahl, "Measuring signal-to-noise ratios in MR imaging," Radiology **173**, 265–267 (1989).
88. B. M. Welsh, "Speckle imaging signal-to-noise ratio performance as a function of frame integration time," J. Opt. Soc. Am. A **12**, 1364–1374 (1995).
89. J. W. Tay, P. Lai, Y. Suzuki, and L. V. Wang, "Ultrasonically encoded wavefront shaping for focusing into random media," Sci. Rep. **4**, 3918 (2014).
90. Z. Fayyaz, N. Mohammadian, and M. R. Avanaki, "Comparative assessment of five algorithms to control an SLM for focusing coherent light through scattering media," Proc. SPIE **10494**, 104946I (2018).
91. L. Zhou, Y. Xiao, and W. Chen, "Learning complex scattering media for optical encryption," Opt. Lett. **45**, 5279–5282 (2020).